Doctoral theses at NTNU, 2024:114

Muhammad Sarmad

Applications of Computer Vision and Deep Learning for Digital Rock Analysis

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Muhammad Sarmad

Applications of Computer Vision and Deep Learning for Digital Rock Analysis

Thesis for the Degree of Philosophiae Doctor

Trondheim, April 2024

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Computer Science

ISBN 978-82-326-7822-8 (printed ver.)
ISBN 978-82-326-7821-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2024:114

Printed by NTNU Grafisk senter

Abstract

Digital rock analysis utilizes various tools available in computer science to represent and process rock data in a way that enhances our understanding of its geological properties. It has become an essential tool for understanding rock samples' physical and chemical properties, which is crucial for exploration and production. It is a multi-step process that includes image acquisition, registration, super-resolution and segmentation. This thesis proposes various methods to improve the individual steps in this process to improve the overall digital rock analysis workflow.

One of the most critical steps in digital rock analysis is obtaining representative and high-quality rock images. This step is necessary for high accuracy in downstream tasks such as fluid flow simulations. The imaging of rock samples is done via micro-computed tomography (micro-CT) or electron microscopy. The resolution of the images obtained from micro-CT scanning can often be limited for a specific task requiring electron microscopes. However, using an electron microscope presents its own challenges, such as high cost and limited field of view.

First of all, this thesis addresses the limitations in the image acquisition process with the help of upsampling the low-resolution rock images. This process is called image super-resolution. Two deep learning-based super-resolution methods have been presented in the first two papers in this work that can potentially improve the digital rock workflow by enhancing image quality.

A critical and time-consuming aspect of digital rock analysis is image registration, which aligns multiple images of the same rock sample. Without registration, correlating different images of the same rock samples is impossible. The algorithm developed in the third paper, a rigid 3D-3D registration algorithm, is a tool that can finish a registration job in seconds instead of hours taken by current industrial image registration tools.

A rock sample can contain multiple mineral types and regions with different properties. Accurately identifying those regions is the first step in determining the properties of the individual parts and, eventually, the whole sample. Rock typing is the process of identifying those regions. It is an essential step in digital rock workflow, commonly performed manually. The second last paper in this thesis presents a deep learning-based method that takes the first steps towards improving the rock typing of laminar rocks.

Scanning the rock samples is costly and time-consuming. Therefore, it is attractive to use deep generative models to generate a representative sample of digital rocks that can be utilized in the workflow. In the fifth and final paper, this thesis presents a novel Diffusion model-based 2D to 3D image generation method. Using the proposed method, a complete 3D image of a rock can be generated using only a single 2D slice, thus addressing the scarcity of 3D data.

In summary, the contributions of this thesis improve the various steps involved in the digital rock analysis workflow using deep learning and conventional computer vision-based methods.

Acknowledgements

This journey towards obtaining my doctoral degree has been a collaborative effort. I have been privileged to be surrounded by numerous individuals whose support and guidance have been instrumental in completing my dissertation.

First, I express my utmost gratitude to my primary supervisor, Frank Lindseth. Your mentorship, motivation, and technical support throughout this research process have been invaluable. You have consistently provided encouragement and direction to keep this project on track.

I am deeply indebted to my co-supervisor Leonardo Ruspini from Petricore. Your generosity in providing this study's data and technical infrastructure has not gone unnoticed. Your insights into industry practices and willingness to support this project have contributed significantly to its success.

A special thank you goes to my third co-supervisor, Gabriel Kiss, from NTNU. Your invaluable feedback on my papers and assistance in navigating new research directions have significantly enriched the quality of this dissertation.

I am sincerely grateful for the companionship, technical support, and co-authorship of my colleague and fellow PhD student, Johan Phan. Your friendship, camaraderie, and academic engagement have made this journey enjoyable and intellectually stimulating.

I extend my deepest gratitude to my parents and siblings for their unwavering support and encouragement throughout my academic journey, without which this thesis would not have been possible. To my wife, Mamoona Birkhez Shami, your love and understanding have been my pillars of strength. Your belief in my abilities and patience throughout this process has been heartening. For my beloved twins, Saad and Ibrahim, your endless joy and zest for life have been the light guiding me through this journey. Your laughter, curiosity, and innocence have reminded me of the wonders and simplicity of life, making every challenge seem less daunting.

I would also like to express my appreciation to the HR and administrative staff at NTNU. Your efficient and prompt resolution of my concerns greatly facilitated my research process.

Finally, I am grateful for the beautiful country of Norway. The tranquil beauty, warm people, and conducive academic environment have made my research journey a delightful experience.

Contents

Abstract **i**

Acknowledgements **iii**

I Thesis **1**

1 Introduction **3**

1.1 Significance of Digitization for Geological Analysis 3

1.2 What is Digital Rock Analysis? 4

1.3 How to Improve Digital Rock Analysis? 4

1.4 Overall Goal and Research Questions 5

1.5 List of Papers 8

1.6 Research Contributions 9

1.7 Other Notable Works 12

1.8 Dissertation Outline 13

2 Background **15**

2.1 Rock Analysis in Oil and Gas Industry 15

2.2 Imaging Techniques 16

2.3 Image Analysis 18

2.4 Machine Learning and Computer Vision 20

3 Methods 31

3.1 Motivation for Method to Achieve RG 1 31

3.2 Motivation for Method to Achieve RG 2 36

3.3 Motivation for Method to Achieve RG 3 39

3.4 Motivation for Method to Achieve RG 4 41

4 Results 45

4.1 RQ 1: Realistic Enhancement of Image Quality via Generative Models 47

4.2 RQ 2: Reduced Latency and Increased Robustness of Image Registration 51

4.3 RQ 3: Automation of Rock Typing via Deep Learning 53

4.4 RQ 4: Synthetic 3D Rock Generation from 2D Image via Diffusion Models 56

5 Discussion 59

5.1 Plan Revisited 59

5.2 RQ 1: Enhance the Quality of Images from CT 61

5.3 RQ 2: Improve Current Dry-Wet Image Registration Methods . . 63

5.4 RQ 3: Automate Rock Typing Pipeline 64

5.5 RQ 4: Improve Methods to Generate High Fidelity Synthetic Rock Samples 65

6 Conclusion 67

6.1 Final Reflection 67

6.2 Limitations 68

6.3 Future Works 69

Bibliography **71**

List of Tables **85**

List of Figures **87**

Acronyms **91**

II List of Papers **93**

Appendix A Paper A: Photo-Realistic Continuous Image Super-Resolution with Implicit Neural Networks and Generative Adversarial Networks **95**

Appendix B Paper B: SIT-SR 3D: Self-supervised slice interpolation via transfer learning for 3D volume super-resolution **105**

Appendix C Paper C: GPU Assisted Fast and Robust 3D Image Registration of Large Wet and Dry Rock Images Under Extreme Rotations **115**

Appendix D Paper D: Core-Scale Rock Typing using Convolutional Neural Networks For Reservoir Characterization in the Petroleum Industry **127**

Appendix E Paper E: Generating 3D Images of Material Microstructures from a Single 2D Image: A Denoising Diffusion Approach **141**

Part I

Thesis

Chapter 1

Introduction

1.1 Significance of Digitization for Geological Analysis

Digitization is impacting a vast spectrum of industries. The primary reason for this exponential progress in the industry's digitization rate is the availability of digital forms of data and the ever-lowering cost of computing power and storage. The oil and gas industry has also been positively impacted due to this wave of modernization. It is a crucial industry responsible for a boon in the modern economy. This industry's role as the primary energy source for transport and electricity generation can not be denied. It has impacted all aspects of our lives by contributing significantly to gross domestic product (GDP) and creating vast employment opportunities. Using digital data and computing to digitize this industry can have a massive impact on the economy. Technological advancements in the field of computer science and imaging acquisition methods have been a major driving force for progress in this industry by improving the tools available for analysis and operations. Advanced technologies like 3D imaging and simulations have led to more efficient and cost-effective analysis of geological samples [9, 14, 39, 5].

The rock samples from a reservoir contain essential information that can reveal the properties of the reservoir. The determination of these properties is crucial for safe operation and maximizing production from the reservoir. In a typical lab analysis, the sample is subjected to various physical and chemical processes to extract important information about the reservoir. This conventional analysis is highly effective at determining the desired properties. However, it has many disadvantages, e.g. the sample is sometimes destroyed, and the analysis can be time-consuming and costly.

1.2 What is Digital Rock Analysis?

Recently, digital rock analysis has emerged as a prominent method. Its main goal is to develop a digital representation of a rock sample. The process typically begins with acquiring images through X-ray or electron microscope technology, capturing either 3D or 2D representations of the sample. These images are subsequently refined through appropriate image processing techniques, preparing them for further analysis. The final step involves utilizing the processed images to construct a digital model of the rock. The digital rock model is finally used in the simulations to determine various properties of the rock sample. The properties that are determined during various steps of digital rock analysis include porosity permeability, electrical conductance, fluid flow, etc. This analysis can be cheaper and faster compared to traditional lab-based analysis [9, 41, 14]. In this thesis, we focus on improving the image processing techniques that are used in the digital rock analysis pipeline.

1.3 How to Improve Digital Rock Analysis?

The digital rock workflow has multiple steps; it starts from image acquisition or generation, followed by multiple steps such as image segmentation, registration, and optional enhancement by super-resolving the image. Finally, a digital model of the rock is created, which is employed in simulations to determine various properties such as fluid flow, permeability and porosity. Improving any individual component in the pipeline will result in the improvement of the overall workflow.

Computer vision and machine learning have rapidly advanced in recent decades. The progress in this field has evolved from using hand-crafted features for solving various problems to using deep features [65, 67, 23]. The digital model of the rock sample is also a 3D image. Therefore, it seems pertinent that digital rock analysis benefits from the recent advancements in computer vision and machine learning. This connection between machine learning and digital rock analysis forms the basis for this thesis, i.e., the topics of this thesis lie at the intersection of machine learning, computer vision, and digital rock analysis.

This thesis develops various methods that benefit or have the potential to help multiple aspects of digital rock analysis directly. In particular, we target the image processing steps of the digital rock analysis pipeline. By making a better image processing method, the aim is to provide the digital rock simulation with a better digital model of the rock since the whole simulation quality depends on the image provided.

1.4 Overall Goal and Research Questions

This thesis aims to develop novel tools and methods for improving digital rock analysis. Within the broad area of digital rock analysis, this work formulates several research questions (RQs) related to the digital rock pipeline. For each RQ, the thesis formulates goals and strategies (RGs) to answer them based on the research gap in the literature. The RQs are stated as follows:

1.4.1 RQ 1: Can we enhance the image quality from sensors such as micro-CT for digital rock analysis?

The images from sensors such as micro-CT used in digital rock analysis have several limitations [6]. These limitations lead to several issues in image quality, such as noise, artefacts, and poor image quality. In some cases, sensor costs prohibit the usage of a higher-quality sensor. The output quality can be enhanced by various computer vision and image analysis methods, such as noise removal and creating a higher-resolution image. However, the current methods have much room for improved output image quality. This thesis seeks to address the question of image quality enhancement that looks more realistic compared to currently available methods. A more representative and high-resolution image can lead to a more accurate simulation for property determination.

RG 1: Explore the use of generative models for realistic enhancement of digital rock images To answer RQ 1, we formulate an objective that limits the scope of multiple means of image quality enhancement to super-resolution. Super-resolution is converting an image of a given resolution to another image of a higher resolution. The resulting image should contain more details compared to the original image. Super-resolution is performed using many different methods; however, deep learning methods have recently become famous due to their effectiveness [65, 68, 121]. Within deep learning, deep generative models have been identified to be good at crafting images with high-quality details that look realistic to human observers.

For digital rock analysis, multiple super-resolution methods have been created [123]. It can be observed from the works that most methods are a direct application of methods already developed for natural images. Therefore, any method developed for natural images could also benefit digital rock analysis.

1.4.2 RQ 2: Can we improve the existing registration methods of dry and wet imaging of the rock samples?

In digital rocks, image registration is a task of spatially aligning two images obtained from a sensor at different times and with variable conditions. A typical scenario is obtaining the wet and dry images from scanning rock samples using micro-CT. A rock sample is first scanned in its dry state during this process. Then, the same rock sample is scanned in the wet state. The resulting images are registered to remove misalignment. The registered images can be subtracted to obtain images that can be used to find properties such as microporosity [14].

Registration of wet and dry images takes much time, depending on the sample size. Usually, it can take hours to register a sample using current industrial tools. In addition, these tools might fail, which requires an expert to register the sample manually. This thesis addresses the question of whether the current industrial methods can be improved such that the resulting registration pipeline has a lower latency while being accurate.

RG 2: Reduce the latency of wet-dry image registration methods and ensure robustness The image registration of natural images has been performed using both conventional methods, such as the intensity-based and feature-based method [95, 76, 118]. In addition, deep learning-based methods have also been developed for similar data types like medical imaging [44]; however, they need to be more robust to be deployed in the industry. Previous works of image registration in the digital rock industry focus on using conventional methods [66]. However, they take substantial time to solve the registration task, especially for large 3D images.

Since the current pipelines in the industry have already shown success using conventional methods. This thesis aims to improve these methods by reducing the latency of the current methods. Specifically, the thesis aims to answer RQ 2 by formulating a method to speed up the computations of conventional methods for wet-dry image registration so that they can provide high-quality and robust registration in the shortest possible time.

1.4.3 RQ 3: Can we replace the manual procedure of rock typing of samples with an automated pipeline?

A given rock sample can contain multiple rock types in a single sample. Identifying the different types is called the task of rock typing. The different types of rocks in a single sample have different properties. Identifying various samples in the rock sample is beneficial since we can determine the properties of the individual rock types. These properties can then be propagated to find the properties of the whole sample [102]. Rock typing is also a part of the digital rock analysis pipeline.

Rock typing is performed manually since the decision boundary between two rocks is subjective based on expert opinions. Therefore, in this thesis, we seek to address automating this task.

RG 3: Explore automation of rock typing of laminar rocks using deep learning This thesis aims to answer RQ 3 by automating the rock typing of rock samples that contain lamination only. This formulation means that the layers of rock types are deposited over one another, thus marking clear horizontal boundaries between various rock types. These boundaries that exist can be subjective to expert opinion. Therefore, this work aimed to get an expert-labeled dataset and explore a supervised deep-learning method to create a pipeline that can distinguish between various rock types. This strategy addressed the subjective nature of boundaries.

1.4.4 RQ 4: Can we improve the fidelity of current synthetic sample generation methods for rock simulation?

The image of the rock sample in the digital rock pipelines is used to create a digital rock. This digital rock is ultimately used in simulations [14]. These simulations reveal various properties of the rock. The step of obtaining the original rock image can be costly, and sometimes only 2D images are available. Whereas the simulations require a 3D image. Therefore, various synthetic sample generation methods are used to create 3D samples that can be used in simulation in place of real images [59]

The fidelity of samples generated by current works is low [1, 113, 17, 114, 115, 59]. This leads to samples that do not represent the underlying statistical distribution of samples in the real image. Therefore, to improve the digital rock pipeline from this aspect, RQ 4 seeks to ask the question about improving the fidelity of current synthetic sample generation methods.

RG 4: Explore diffusion models for high fidelity 3D sample creation from 2D rock images Currently, the most promising methods for synthetic sample generation are based on deep learning-based generative model [59]. In addition, the challenging task of 3D image generation from 2D samples has also been targeted for rock image generation. However, previously generative adversarial networks (GANs) have been utilized for this task, which are prone to various problems that lead to low-fidelity samples. Recently, diffusion models have become a popular alternative to GANs for high-fidelity generation [47]. However, they have not been explored for the task of 2D to 3D image generation in digital rock analysis.

Therefore, to address RQ 4, we formulate a goal focusing on 2D to 3D generation of synthetic rock samples using diffusion models.

1.5 List of Papers

This thesis is being delivered as a collection of papers. Three papers have been accepted, respectively. Two are under review at the time of submission of this thesis. The work consists of the following articles:

- Paper A: Muhammad Sarmad, Leonardo Ruspini, and Frank Lindseth, ”Photo-Realistic Continuous Image Super-Resolution with Implicit Neural Networks and Generative Adversarial Networks” Accepted to the Proceedings of the Northern Lights Deep Learning Workshop. Vol. 3. 2022 (Oral Presentation).

- Paper B: Muhammad Sarmad, Leonardo Carlos Ruspini, Frank Lindseth ”SIT-SR 3D: Self-supervised slice interpolation via transfer learning for 3D volume super-resolution”, Accepted to Pattern Recognition Letters (Journal).

- Paper C: Muhammad Sarmad, Johan Phan, Leonardo Ruspini, Gabriel Kiss, Frank Lindseth, ”GPU Assisted Fast and Robust 3D Image Registration of Large Wet and Dry Rock Images Under Extreme Rotations”, Under Review in Journal of Computers and Geosciences.

- Paper C: Muhammad Sarmad, Johan Phan, Leonardo Ruspini, Gabriel Kiss, Frank Lindseth, ”Core-Scale Rock Typing using Convolutional Neural Networks for Reservoir Characterization in the Petroleum Industry”, Accepted in 23rd International Multidisciplinary Scientific GeoConference SGEM 2023.

- Paper E: Johan Phan*, Muhammad Sarmad*, Leonardo Ruspini, Gabriel Kiss, Frank Lindseth, ”Generating 3D Images of Material Microstructures from a Single 2D Image: A Denoising Diffusion Approach”, Under Review in Nature Machine Intelligence Journal. ($*$ indicates equal contribution)

The Figure. 1.1 demonstrates how various papers are related to research questions. Here, it is relevant to inform the reader about the overlap between RQ 1 and RQ 4, as shown in Figure. 1.1. The method created while addressing RQ 1 produces high-resolution images given low-resolution images. Meanwhile, the methods that were developed to answer RQ 4 create synthetic images from scratch. Therefore, the methods from RQ 1 can be used to enhance the synthetic image resolution from RQ 4 further. However, testing this is not within the scope of this work.

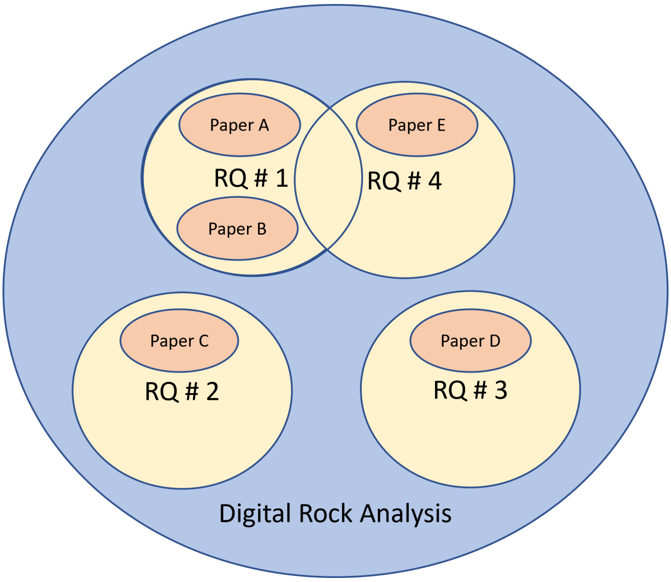


Figure 1.1: Research Questions (RQs) formulated in the scope of digital rock analysis. This diagram shows the link of the list of Papers from A to E to their respective RQ.

1.6 Research Contributions

Table. 1.1 provides an overview of all the research questions, goals formulated to answer those questions, publications and contributions made within the publications.

1.6.1 Contribution 1

RQ 1 seeks to find the answer to the question of image quality enhancement of sensors for digital rock analysis as shown in Table. 1.1. According to the formulated goal, the scope of the investigations was limited to deep generative models for realistic image enhancement. In pursuit of this goal, it is demonstrated in the table that the first contribution was made in Paper A.

Paper A contributes by formulating a novel method for deep learning-based super-resolution. This work focuses on the use of a particular type of deep network called the implicit neural network [107]. This neural network is capable of outputting better-quality images compared to convolutional neural networks (CNNs). Previous works use the standard loss function L_1, which promotes blurry image output [68]. The main contribution is that we propose to train implicit neural networks using generative loss functions that promote realistic image generation.

Table 1.1: Thesis Overview

Research Questions	Research Goals	Paper	Contributions
RQ 1: Enhance the quality of images from CT.	RG 1: Explore use of generative models for realistic enhancement of digital rock image	A	Proposed a generative model for realistic image super-resolution of 2D images
		B	Proposed a method for utilizing 2D generative model for 3D image super resolution for rock images
RQ 2: Improve current dry-wet image registration methods	RG 2: Reduce latency and improve robustness of conventional methods	C	Formulated a registration tool that utilizes graphical processing units to reduce method latency from hours to minutes
RQ 3: Automate rock typing pipeline	RG 3: Explore deep learning based methods for rock typing of laminar rocks	D	Proposed a supervised deep learning method for rock typing of laminar rocks and performed explainability analysis on model
RQ 4: Improve methods to generate high fidelity synthetic rock samples	RG 4: Explore deep generative models for 2D to 3D generation	E	Proposed a novel method based on deep generative diffusion models for 2D to 3D image generation

This contribution addresses RQ 1 as the improved model for 2D image super-resolution that can be used to enhance the quality of images from sensors such as micro-CT and SEM.

1.6.2 Contribution 2

Paper B contributes a novel algorithm that performs 3D image super-resolution. Previous methods for digital rock analysis mostly learn 2D image super-resolution [123]. The method suggested in this work focuses on realistic 3D image super-resolution, which has not been addressed in previous work. In addition, it is capable of learning 3D image super-resolution using only 2D image ground truth.

Contributions 1 and 2 provide novel methods for image enhancement that address RQ 1. The resulting pipelines are capable of enhancing the image quality of the images obtained from sensors such as micro-CT and SEM. The focus was that images are enhanced such that they look more realistic to the human expert observer. These enhanced images lead to more accurate simulations due to higher quality images.

1.6.3 Contribution 3

It can be observed from Table. 1.1 that Paper C contains the contribution that addresses RQ 2. More specifically, Paper C provides a registration tool that uses graphical processing units (GPUs) to parallelize the computations of a conventional image registration algorithm. A novel aspect of this paper is that the conventional base method is augmented with a novel algorithm that can handle extreme rotations. For example, if someone accidentally places the wet sample upside down before imaging, then even in that case, the algorithm can successfully perform registration. Since the computations are based on GPUs, the algorithm completes the task in under minutes compared to the hours used by the current standard method. Addressing RQ 2 with the contribution in Paper C will thus lead to a time-efficient digital rock pipeline.

1.6.4 Contribution 4

Semantic segmentation is the task of pixel-wise classification. This means that each pixel in an image is classified and assigned a distinct class. To address RQ 3, paper D contributes a method that treats the rock typing problem as a semantic segmentation problem. The method assigns each pixel in the image a class label corresponding to the rock type. The method used in the paper is based on supervised learning. Therefore, it utilizes an expert-labelled dataset to learn the boundary between various rock types. The output of the network is a segmentation map of all the rock types that exist in the sample. This method positively contributes to automating the rock typing and digital rock workflow by addressing RQ 3.

1.6.5 Contribution 5

The Table. 1.1 shows that Paper E contributes a novel diffusion model that is capable of generating 3D rock images from 2D samples. This is exactly in line with the goals that were set out to answer RQ 4.

The diffusion model is capable of generating 3D high-fidelity synthetic samples from 2D images, overtaking the previous state-of-the-art in performance [59]. This directly contributes to making a more accurate digital rock pipeline. This will lead to the availability of high-quality synthetic samples for the digital rock simulation, thus leading to accurate determination of the properties of the underlying rock.

1.7 Other Notable Works

In addition to the list of papers included in this thesis, the author was also involved in the following works during the PhD program related to computer vision and deep learning, as well as their applications in the automation industry.

- Paper F: Muhammad Sarmad*, Mishal Fatima, Jawad Tayyub*, ”Reducing Energy Consumption of Pressure Sensor Calibration Using Polynomial HyperNetworks with Fourier Features” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. 2022. [105]

- Paper G: Rabia Ali*, Muhammad Sarmad*, Jawad Tayyub*, Alexander Vogel, ”Accurate Detection of Weld Seams for Laser Welding in Real-World Manufacturing”, Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37. 2023. [3]

- Paper H: Jawad Tayyub*, Muhammad Sarmad*, Nicolas Schönborn*, ”Explaining deep neural networks for point clouds using gradient-based visualisations”, Proceedings of the Asian Conference on Computer Vision. 2022. [116]

- Paper I: Tejaswini Medi*, Jawad Tayyub*, Muhammad Sarmad*, Frank Lindseth, Margret Keuper, ”FullFormer: Generating Shapes Inside Shapes”, Accepted to DAGM German Conference on Pattern Recognition 2023. [79]

* indicates equal contribution.

1.8 Dissertation Outline

This section shows the outline of the thesis. The various topics in this thesis have been divided into chapters as follows:

- Chapter 1 contains the introduction and overview of the thesis.
- Chapter 2 provides the necessary background knowledge and related work for the contributions.
- Chapter 3 summarises the key points in the methodology of each paper.
- Chapter 4 discusses the salient results of respective methods.
- Chapter 5 discusses the results on a higher level.
- Chapter 6 concludes with the limitations and possible future works.
- Part II of the thesis includes the list of articles related to this thesis.

Chapter 2

Background

2.1 Rock Analysis in Oil and Gas Industry

2.1.1 Conventional Core Analysis

Conventional core analysis (CCA) is a prevalent technique in the oil and gas industry to ascertain the properties of rock samples extracted from petroleum reservoirs. The comprehensive evaluation of these samples' physical and chemical properties, obtained through drilling, facilitates a deeper understanding of reservoir characteristics such as porosity, permeability, and fluid saturation. Such parameters are critical in determining the quantity and quality of production from the reservoir. [39, 5]

In conjunction with CCA, special core analysis (SCAL) is also utilized to provide more complex reservoir properties, such as capillary pressure, relative permeability, and electrical properties. SCAL often complements CCA by offering a more detailed understanding of fluid flow behaviour and rock-fluid interactions within the reservoir. [41, 78]

CCA and SCAL necessitate using specialized equipment and skilled personnel for collecting and subsequent laboratory analysis of rock samples. Although these methods yield accurate results, they are time-consuming and resource-intensive, incurring substantial costs. Nevertheless, despite the inherent challenges, CCA and SCAL maintain their positions as essential techniques within the oil and gas industry, offering invaluable insights into reservoir properties and informing decisions on reservoir management strategies.

2.1.2 Digital Rock Analysis

Digital rock technology (DRT) is an umbrella term encompassing two related but distinct methodologies employed within the oil and gas industry: digital rock analysis (DRA) and digital rock physics (DRP). These methodologies contribute to a deeper understanding of the reservoir's complete properties, such as porosity and permeability. [9, 41, 14]

DRA, the initial step in DRT, focuses on characterizing rock properties using advanced digital imaging methods, such as X-ray tomography, rather than CCA. This approach allows for a more accurate and detailed characterization of rock properties and their distribution throughout the reservoir. [19]

DRP, the subsequent step in DRT, integrates DRA with numerical modelling to derive critical rock physics. This process involves solving transport equations on digital representations of rock samples to simulate various rock physics phenomena. [9, 41, 14]

Both DRA and DRP contribute to complete reservoir understanding. However, the scope of the present thesis is limited to DRA. Implementing DRA offers several advantages over CCA. These advantages include being non-destructive for the rock sample, faster property determination, and more cost-effective analysis. As a result, DRA is becoming an increasingly indispensable tool for rock analysis. The fast pace of development in computer science is benefiting the development of novel DRA tools.

2.2 Imaging Techniques

The starting point of DRA is obtaining the rock sample data in digital form, mostly in the form of an image of the rock sample. This image can be obtained using various modalities and instruments such as electron microscopes and X-ray computed tomography.

2.2.1 3D Images

Employing a cutting-edge imaging method, X-ray computed tomography (CT) reconstructs three-dimensional depictions of samples by capturing X-ray images from an assortment of viewpoints. This innovative technology has profoundly influenced numerous domains, including the medical and oil industries, which have utilized it extensively for several decades. The data obtained from CT is mostly 3D data, which is advantageous for analysis since it provides a complete 3D view of the rock sample [30].

Common Lab CT Currently, laboratory-based micro-CT instruments are the go-to solution for procuring 3D depictions of rock samples in digital rock physics studies. The X-ray source and detector remain stationary in most laboratory micro-CT systems while the sample is in motion [126, 125, 48, 42, 32]. However, micro-CT gantry scanners with moving parts are used in some instances. This setting prevents undesirable movements of the sample and fluid flow phenomena from happening within the sample, thus helping in accurate evaluations of properties. [21]

Use of Medical CT in Rock Analysis The oil industry has also leveraged medical CT imaging for various applications, such as the inspection of cores, the undertaking of core flooding experiments, and the appraisal of rock characteristics [126].

Synchrotron Imaging Determining effective rock properties from digital images of pore structures demands adequate resolution to discern all pertinent structural elements. Consequently, DRP calls for a greater image resolution than that offered by conventional medical CTs. Initially, high-resolution images of rock samples were acquired by employing synchrotron facilities' beam lines as the X-ray source in previous works. However, currently, synchrotron imaging mainly focuses on recording rapid transient phenomena at the pore level, highlighting its crucial role in enhancing our understanding of intricate rock formations [38, 15].

In Table. 2.1, the comparison of different CT imaging techniques mentioned above is provided. The table provides information such as the movement of source and sample and an estimate of the relative expense of various methods.

Table 2.1: Comparison of different CT imaging techniques

Imaging Technique	Source Movement	Sample Movement	Cost	Citation
Gantry	Rotates w/ detector around sample	Stationary	$$	[21]
Synchrotron	Fixed	Rotates on a stage	$$$$$	[15, 38]
Medical CT	Rotates w/ detector around patient	Stationary	$$$	[126]
Lab CT	Stationary	Rotates	$	[126]

2.2.2 2D Images

The most desirable form of image of a 3D sample is also a 3D image. However, sometimes, the imaging method can only provide a high-resolution sample at the cost of providing a 2D image of a thin slice part of the 3D sample.

Although medical CT provides 3D images, it trades off resolution for the ability to cover a more comprehensive volume. In contrast, scanning electron microscopes (SEM) images can be stitched together through an automated process, generating comprehensive, large-scale images, albeit remaining two-dimensional [20].

Backscattered electron (BSE) imaging, a critical mode in SEM, enables high-resolution, sub-micron examination of rock samples. While BSE imaging excels in providing detailed two-dimensional information, it cannot produce the three-dimensional images required for specific applications [20].

Focused ion beam scanning electron microscopy (FIB-SEM) presents an alternative approach to acquiring 3D images [69]. This technique employs an ion beam to remove thin layers from the sample sequentially, enabling the SEM to image each exposed surface layer by layer, ultimately creating a 3D representation. However, FIB-SEM has its limitations. Its primary constraint lies in its localized focus, which results in images covering a relatively small volume of approximately five μm^3. Consequently, this technique may not provide a comprehensive sample view [58]. Additionally, FIB-SEM's destructive nature may limit its applicability in specific contexts, mainly when sample preservation is crucial.

When considering the strengths and limitations of BSE, FIB-SEM, and medical CT imaging techniques, it becomes apparent that each method serves a distinct purpose. BSE imaging delivers high-resolution, two-dimensional information on rock microstructure and mineralogy but lacks the three-dimensional capabilities of FIB-SEM and medical CT. Conversely, FIB-SEM's destructive nature and localized focus may impede its effectiveness in certain situations, particularly when a more comprehensive view is required. Medical CT offers a broader 3D perspective at the expense of reduced resolution compared to SEM-based approaches.

2.3 Image Analysis

The image acquisition process is a critical first step, followed by the core image analysis component. Issues often arise in CT images, such as beam hardening, ring artefacts, and partial volume effects, necessitating pre-processing techniques to mitigate noise and enhance image quality. Such techniques include histogram equalization, spatial filtering, and adaptive thresholding [126]. Furthermore, sophisticated approaches, such as super-resolution, may be employed to upsample the image for improved analysis.

Subsequent digital rock image analysis stages involve determining properties such as mineral composition, porosity, water saturation, and grain size distribution. Segmentation algorithms are employed to delineate distinct regions within the image, allowing for extracting these properties. Techniques such as region growing, watershed, and level-set segmentation have been widely adopted in the literature.

Analytical techniques are further applied to evaluate the interconnectivity and geometry of the pore space to understand fluid flow behaviour and transport phenomena [7, 70].

2.3.1 Mineralogy

The application of image analysis techniques enables the elucidation of mineralogical composition within geological samples. Traditional X-ray diffraction analysis methodologies can allow for determining mineral constituents; however, they lead to the mechanical disintegration of the specimen. In addition, this approach can not provide insight into the spatial distribution and geometric relationships among the constituent minerals within the sample [112].

In contrast, micro-CT and SEM offer the advantage of non-destructive, high-resolution visualization of geological specimens' internal structure and composition. These techniques yield grayscale images, from which quantitative information on mineral phases, porosity, and other microstructural attributes can be extracted. The grayscale values in these images facilitate the classification of the constituent materials within the specimen, including solids, pores, and clay minerals, among others. This comprehensive characterization of mineralogical composition, in conjunction with spatial and geometric information, enables the differentiation of various rock types based on their unique attributes.[9, 41, 14]

2.3.2 Grain Size

The grain size distribution within a geological material is a critical parameter that significantly influences its microstructural and macroscopic properties. As such, the arrangement and packing of grains within a sample are vital in determining crucial characteristics such as porosity, permeability, and mechanical strength as shown in Fig. 2.1. Grain sizes may vary across a spectrum from fine to coarse, and samples may exhibit homogeneous or heterogeneous grain size distributions. The variability in grain size and packing configurations within a single geological sample can yield distinct rock types with divergent properties, thus making grain size analysis an essential tool in classifying and differentiating geological materials. Understanding grain size distribution and packing patterns is crucial for developing accurate models.

2.3.3 Simulation

Simulation is central to DRP, encompassing the final stage after digital rock analysis. By employing simulations, crucial properties such as permeability and fluid flow can be effectively calculated, enriching our understanding of the behaviour of porous media. Permeability calculations can be performed using two distinct

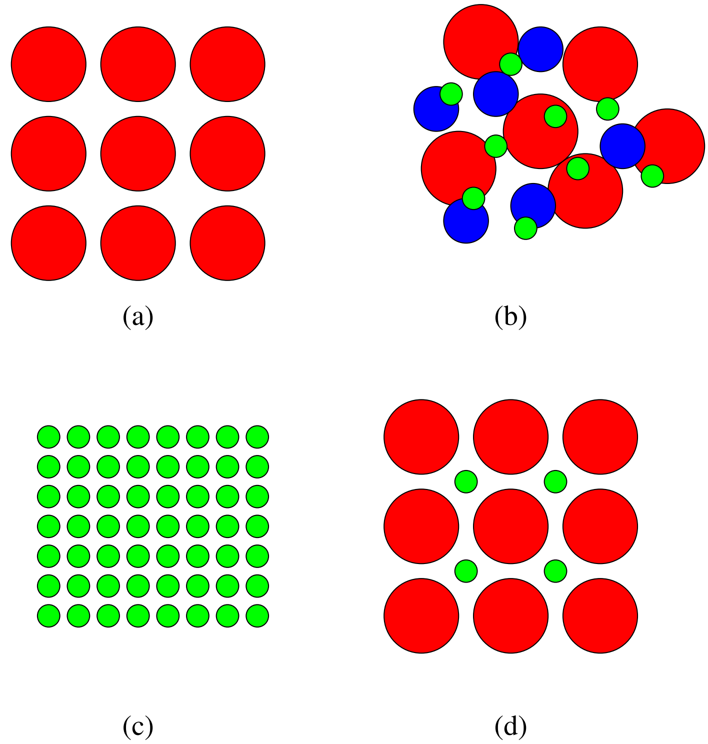


Figure 2.1: Examples of different grain packing settings in rock samples: (a) large grains, (b) random grains, (c) small grains, and (d) a mix of large and small grains. The arrangement and size of grains impact the porosity and permeability of the rock.

approaches: direct modelling based on binarized images of the pore structure or network modelling that considers the overall topology of the pore network.[19, 102]

Additional properties that can be derived through DRP simulations include electrical conductance, diffusion, and elastic properties. However, it is interesting to note that the quality of these simulations is heavily affected by the first step, i.e. image analysis. If artefacts and noise are present in the image, they propagate and lead to poor results in the simulation step.[126]

2.4 Machine Learning and Computer Vision

The genesis of computer vision research dates back to the 1960s, with initial efforts concentrating on detecting simple shapes and patterns. The development of foundational edge detection algorithms, such as Sobel and Canny operators, marked a significant milestone in the field [35]. Popularizing scale-invariant feature transform (SIFT) and other feature descriptors for object recognition and tracking during the 1980s and 1990s opened up new avenues for computer vision applications [73].

Machine learning and computer vision have evolved in tandem, with techniques like support vector machines (SVM) being employed for object classification and recognition [85]. In the early stages, these machine learning approaches for computer vision relied on hand-crafted features. However, the introduction of deep learning by Geoffrey Hinton and his collaborators revolutionized the field, enabling significant advancements in computer vision applications without depending on hand-crafted features [65].

It has been discovered through further research that a combination of hand-crafted and deep features can be even more effective at solving various computer vision problems, showcasing the continued relevance of traditional techniques alongside modern deep learning methods.

2.4.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a specialized class of deep learning models designed explicitly for handling grid-like data structures, such as images [65, 67, 23]. The primary function of CNNs is to automatically and adaptively learn spatial hierarchies of features from the input data. The architecture of a CNN consists of a series of layers, including convolutional layers, pooling layers, and fully connected layers. In the convolutional layer, a set of filters or kernels slide across the input image, performing element-wise multiplication and aggregation to generate feature maps. These filters are adept at detecting local patterns, such as edges and textures, while preserving the spatial relationships within the image. The pooling layer, which typically follows the convolutional layer, reduces the spatial dimensions of the feature maps, thereby reducing computational complexity and promoting translation invariance. This process is repeated across multiple layers, with higher layers learning increasingly complex and abstract features. Finally, the extracted features are flattened and fed into fully connected layers, which enable the network to make predictions or classifications based on the learned hierarchical feature representations. CNNs have demonstrated exceptional performance in various computer vision tasks, such as image classification, object detection, and semantic segmentation, revolutionizing the field and paving the way for numerous applications.

2.4.2 Generative Adversarial Neural Networks

Generative Adversarial Networks (GANs) were presented by Goodfellow et al. in 2014 as a novel approach to generative modelling. The core idea behind these models is to train two neural networks in tandem. The first network is a generator network, which generates samples from a given distribution. The second network is a discriminator network, which distinguishes between real samples from the true data distribution and the samples produced by the generator. The training process can be seen as a two-player minimax game, where the generator seeks to create samples that can fool the discriminator. In contrast, the discriminator tries to improve its ability to detect fake samples [43].

Deep Convolutional GANs (DCGANs) by Radford et al. [97] significantly improved these models, providing architectural guidelines for stable training using convolutional layers. This work paved the way for further improvements in GANs, such as the introduction of Wasserstein GANs (WGANs) by Arjovsky et al. [8], which presented a new training objective that improves stability and convergence.

2.4.3 Implicit Neural Networks

Implicit Neural Representations or Implicit Neural Networks (INNs) refer to a class of neural networks that learn to represent data implicitly without explicitly defining the dimension of output. These networks can model complex, high-dimensional data such as images, 3D shapes, or functions. INNs map from a lower-dimensional input space (e.g., coordinates) to the desired output (e.g., colour or density values) [108].

One popular approach to implicit neural representations is using continuous Signed Distance Functions (SDFs) for 3D shape representation, as Park et al. [89] proposed in their work on DeepSDF. DeepSDF is a deep neural network that learns to represent a 3D shape by mapping 3D coordinates to signed distance values, which are positive outside the shape and negative inside. This continuous representation allows for high-resolution shape generation and can be easily manipulated or combined with other shapes.

Another prominent example of implicit neural networks is Neural Radiance Fields (NeRF), introduced by Mildenhall et al. [81]. NeRF is a method for representing complex 3D scenes as a continuous function, which maps 3D coordinates and viewing directions to RGB colour values and densities. This technique allows for reconstructing high-quality, novel scene views from a sparse set of input images by optimizing the neural network to minimize the discrepancy between rendered and observed images.

2.4.4 Progress in Super-Resolution

Reducing Distortion: Many previous works in the field of super-resolution focus on minimizing distortion by improving the Peak Signal-to-Noise Ratio (PSNR). These approaches aim to minimize the discrepancy between the reconstructed high-resolution image and the ground truth, often measured by pixel-wise metrics such as Mean Squared Error (MSE) or PSNR. Early efforts include the work of Dong et al. [36], who proposed an end-to-end CNN-based super-resolution network that does not require handcrafted features. Later, Kim et al. [61] incorporated residual connections with recursive CNNs for the same purpose. The advancements in CNN architectures, such as skip connections [46], have also been applied to Single Image Super-Resolution (SISR) [68]. Furthermore, the DenseNet architecture [50] was adapted by Zhang et al. [132] for SISR tasks. RCAN [131] demonstrated even better performance using a deeper network, reducing distortion in the super-resolved images.

Improving Perceptual Quality: The other SISR methods focus on perceptual quality-based metrics [130, 18] to generate more visually appealing and realistic outputs. The introduction of VGG features as a perceptual loss function [56] has shifted the research focus towards improving perceptual quality. Adversarial training using GANs [43, 68] has further contributed to this direction. The ESRGAN [121] model uses a relativistic discriminator [57] to enhance the quality of the output images. Edge and gradient-based methods additionally focus on geometry and perceptual quality [84], producing sharp details; however, they may sometimes induce artefacts. To preserve fine details while achieving state-of-the-art performance on benchmarks [129, 16, 51, 77] based on the LPIPS metric [130], Ma et al. [75] propose a gradient branch and loss. This approach combines distortion reduction and perceptual quality improvement, producing a more balanced and visually appealing super-resolution output.

Applications to Digital Rocks

The literature on super-resolution techniques for 3D data falls into two primary categories. In the first category, *3D images are considered as sets of 2D slices*, and conventional 2D SISR methods designed for coloured images are employed. This approach allows for super-resolution in two dimensions; however, it neglects the lower resolution of the third dimension. To effectively upsample 3D images, it is crucial to account for all three dimensions. There have been numerous instances of using CT data for training 2D SISR networks to super-resolve 2D slices within a 3D image, both in the medical and digital rock fields [123, 25, 128].

The second category involves *training end-to-end 3D networks* for the super-resolution of volumes. Although these methods are more challenging to develop and thus relatively scarce, they provide a more comprehensive solution [29, 93]. Chen et al. [29] proposed the mDSCRN method for 3D volume super-resolution, inspired by DenseNet [50]. On the other hand, Peng et al. [93] presented SAINT, arguing that mDCSRN yields suboptimal results and has a higher memory and computational burden. However, the SAINT method requires ground truth high-resolution data for supervised training. While this technique has been applied to medical CT data that only require one-dimensional upsampling via frame interpolation, it is ill-suited for data necessitating three-dimensional upsampling.

2.4.5 Progress in Registration

Conventional Methods Correlation methods have been employed for unimodal image registration methods since they gained prominence due to their efficacy, as highlighted by Pratt et al. [95]. However, these methods often require supplementary preprocessing and cleaning steps to guarantee successful cross-correlation. As demonstrated by Althof et al. [4], employing correlation-based approaches can result in reasonably accurate solutions for various registration challenges. In the context of multimodal images, Viola et al. [118] and Maes et al. [76] proposed utilizing mutual information, offering enhanced robustness for such image types.

An alternative approach involves Fourier-based methods, which operate on the Fourier representation of images and boast faster performance compared to cross-correlation techniques De et al. [34]. Focusing on the optimization algorithm involved in the process, Jenkinson et al. [54] developed a global optimization algorithm tailored explicitly for image registration tasks, resulting in more efficient registration processes.

An appropriate similarity measure (e.g., correlation versus mutual information) is critical to registration success. In order to facilitate this decision-making process, Roche et al. [100] provided valuable insights and guidance on selecting the optimal measure for the best possible results. Despite the effectiveness of these traditional methods, they often require multiple iterations to converge to an acceptable solution. Open-source libraries Pytorch [91] and AIRLab [104] allows to harness the power of graphical processing units (GPUs) to expedite gradient computation necessary for optimization.

Deep Learning Methods With the advent of deep learning, novel solutions for various registration challenges have emerged, particularly leveraging the capabilities of CNN [65, 67]. Learning-based approaches have been incorporated into every stage of the registration process, leading to innovative methods and improved results. For instance, Haskins et al. [45] proposed using a CNN to learn the similarity metric while preserving the traditional optimization process. In contrast, Miao et al. [80] and Chee et al. [24] utilized synthetic transform-based data generation techniques to train a CNN model capable of predicting the transformation matrix in a single attempt, significantly reducing computation time.

Taking a different approach, Liao et al. [71] employed a reinforcement learning-based technique to train an agent for robust image registration, opening up new possibilities for adaptive registration strategies. While these methods primarily utilize medical image datasets and exhibit impressive speed during inference, they face limitations when applied to datasets that lack sufficient distinguishing features. Additionally, unsupervised approaches for deformable image registration have been explored by Hu et al. [49]; however, this thesis is limited to rigid transformations. For a more comprehensive understanding of learning-based approaches, consult the surveys presented by Haskins et al. [44] and Fu et al. [40].

Application towards Digital Rocks

Image registration represents a crucial preliminary phase in numerous investigations on the estimation of rock properties, as demonstrated in various studies [64, 10, 88, 96]. Despite the ubiquity of image registration solutions in the medical imaging domain, there need to be more well-developed approaches for wet and dry image registration. This is true due to the lack of well-defined matching features and key points in some wet and dry rock samples. The seminal work of Latham et al. [66] utilized a correlation-based technique with an optimization algorithm for 3D dry-to-wet image registration. However, the iterative nature of this optimization process renders the method significantly time-consuming. Research that harnesses GPU technology's capabilities is essential as it can substantially reduce the computation duration from hours to mere seconds.

The AIRLab open-source library, founded on Pytorch and employing GPU technology, has proven highly effective for image registration [104]. Although alternative toolboxes such as ITK [127], Elastix [63], and ANTs [11] are available, they do not exploit GPU capabilities as transparently and efficiently as AIRLab. As a result, the prototyping process with these methods is time-intensive and error-prone, potentially creating obstacles in industrial applications. Various other toolboxes are also available for registration. Each of them has their weakness and strengths. For an in-depth examination of other toolboxes for image registration tasks, we refer the reader to the comprehensive survey conducted by Keszei et al. [60].

2.4.6 Progress in Segmentation

Segmentation is a process in computer vision that involves dissecting an image into different regions, each representing a specific object or category. It goes beyond simply recognizing objects in an image by determining the boundaries and extent of each object, effectively assigning a label to every pixel. This pixel-wise classification allows for a deeper understanding of the visual content. It comprehensively represents the scene, enabling more advanced analysis and decision-making in various applications.

Fully Convolutional Networks (FCNs) demonstrated remarkable performance in semantic segmentation by replacing fully connected layers with convolutional layers to preserve spatial information [72, 86]. Following this, encoder-decoder architectures such as SegNet [12] and U-Net [101] emerged, leveraging the hierarchical structure of CNNs to generate high-resolution segmentation maps.

U-Net proved to be highly effective, especially for biomedical image segmentation [101]. Its architecture incorporated a contracting path for capturing context and a symmetric expanding path for precise localization, improving segmentation quality. The success of U-Net led to its widespread adoption and further adaptations in various domains.

Application towards Rock Typing

In digital rock analysis, the objective is to ascertain the comprehensive properties of reservoirs. It is beneficial to partition the reservoir into distinct sub-regions, or rock types, that share similar properties. This process, referred to as rock typing, involves classifying sub-regions based on criteria such as their physicochemical properties.

Ismail et al. [53] employed regional Minkowski measures, which characterize the geometric properties of objects, in conjunction with a multivariate Gaussian mixture model for rock-type classification. Wang et al. [124] proposed an image analysis approach for rock typing that utilizes the iterative Chan-Vese segmentation model to segment binary images into distinct rock types. However, this model is subject to hyperparameter selection, sensitivity to initial condition selection, and slow convergence and is not adequate when handling noisy images or images with poorly defined boundaries. These manual methods fall under the category of feature engineering.

Nuclear Magnetic Resonance (NMR) is crucial for rock typing, particularly when assessing fluid diffusion. NMR is effective in such situations because it relies on the presence of hydrogen nuclei, which are prevalent in oil or water samples. Despite its utility, NMR has several drawbacks, including requiring a fluid sample, low spatial resolution, significant sample preparation efforts, high cost, and slow image acquisition. Cui et al. [33] investigated the impact of diffusional coupling on NMR measurements of saturated laminated sandstone at the layer scale to evaluate the feasibility of NMR-based rock-typing approaches. In contrast, Jiang et al. [55] utilized more cost-effective X-ray images and introduced efficient numerical techniques based on Minkowski functionals for deriving regional Minkowski measure fields for large-scale three-dimensional X-ray tomography (3D) datasets. Although still reliant on manual intervention, the researchers demonstrated the applicability of these 3D feature fields to microstructure classification by implementing a multivariate Gaussian mixture model and thin-bedded sandstone.

Alhwety el al. [2] examined the conventional interpretation of NMR measurements on fluid-saturated reservoir rocks, with the fluid being a crucial component of most NMR analyses. They emphasized that the transverse relaxation time and pore size distributions, expected to be most directly related, are often not directly correlated in many multi-scale porosity systems due to diffusion coupling between varying pores. This framework presents yet another approach to developing rock typing methodologies.

Existing research employing deep learning for rock typing has primarily concentrated on classifying rock types within image patches based on the rock's elemental composition rather than segmenting full images according to the significant physical properties of regions [13]. Consequently, this thesis focuses on a deep learning based segmentation method to generate more accurate boundaries between distinct rock types based on essential properties such as porosity while utilizing more affordable and prevalent image modalities like CT/X-ray.

2.4.7 Progress in Image Generation

Implicit Generative Models Generative adversarial networks (GAN) have been phenomenal in achieving realistic-looking image generation [43]. However, they are prone to multiple issues, due to which they are not always the preferred model for the task of image generation. Some issues that commonly affect GANs are susceptibility to model collapse and catastrophic forgetting in the case of conditional generation [117]. In a conditional generation, the model can generate an image based on a condition, e.g. to generate a 3D image; we can input a 2D image as a conditional, and then the model will generate a 3D image that is conditioned on the 2D image.

Likelihood-based Models A way to model data is to model the distribution of the data using a likelihood function. Such models are known as likelihood models. The Likelihood Function for a model M with some parameters θ, and data D, then the likelihood function can be represented as $L(\theta|D)$. Variational auto-encoders are a class of likelihood models that are very fast at inference and do not suffer from mode collapse. However, generally have a lower quality than GAN [62, 99].

Similarly, Auto-regressive models can also accurately capture the data distribution with superior sample quality. However, they are slow at inference, making them impractical for multiple applications [87, 98, 90, 26]. Various hybrid models have recently become very popular, capable of generating high-resolution images by combining transformer architecture in the framework of likelihood-based model and generative adversarial networks [37]. Such models attempt to use the strengths of each framework.

Score-Based and Diffusion Models A third category of the model is the denoising diffusion probabilistic model which is also known as score matching models [52, 47, 109, 110]. The score function $S(\theta|D)$ is the gradient (derivative) of the log-likelihood with respect to the parameters θ. This can be written as $S(\theta|D) = \nabla_\theta \log L(\theta|D)$. Score matching is a technique to find the best parameters θ that makes the model's score function match the true score function of the data. They have accomplished exceptional results in multiple downstream tasks such as super-resolution, generation, and diverse computer vision problems [103, 22, 27]. Our work also adopts these models for 3D generation of rock samples from 2D data.

Applications to Digital Rocks

The generation of 3D microstructure is a beneficial application of generation models for digital rock workflow. However, this task presents unique challenges. Many previous works explore the solution.

Rules based Models Before the advent of data-driven and deep learning models, conventional methods based on stochastic modelling were used to generate 3D models. One such stochastic model is based on modelling the sedimentation process, which can model a wide variety of rock configurations [1, 113, 17, 114, 115]. However, each minor component needs to be modelled in this simulation-based tool, due to which they are cumbersome to develop and maintain. In addition, they are very limited, so they cannot capture the complexities of actual rock samples.

Implicit Generative Models The advent of deep learning and generative adversarial networks has dramatically improved the generation of rock data for digital rock workflow. Previous work has utilized GANs for unconditionally generating a singular rock type [82, 83]. The conditional generation provides more control over the output, essential for utilizing these models for digital rock workflow. Previous works have also explored such models [59, 119, 133, 31]. However, they are negatively affected by certain limitations due to inherent limitations of the GANs, such as mode collapse. Kench et al.[59] demonstrated the ability to learn from 2D images to generate 3D microstructures similar to our method; however, we utilize score-matching / diffusion-based models.

Likelihood-based Models Phan et al. [94] have utilized a likelihood-based model to generate high-quality 3D rock samples from 2D conditional images. But they always require a 3D ground truth. On the other hand, we utilize score based model and only need the 2D sample to synthesize 3D images.

Chapter 3

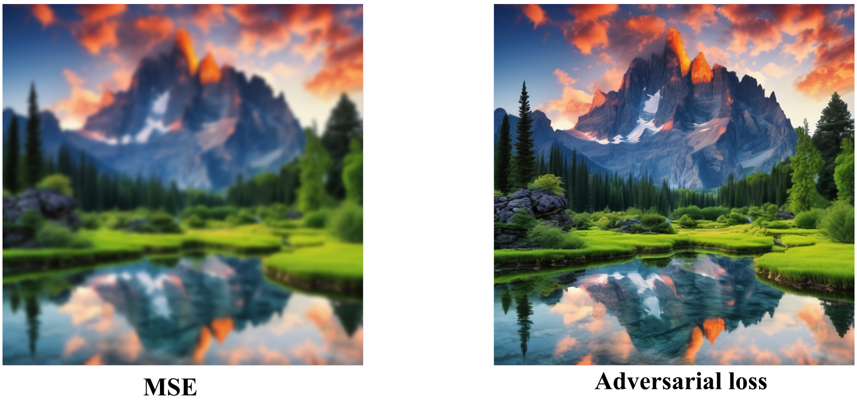
Methods

The motivation for the choice of method for each paper in this thesis is primarily based on the research goals set out in the introduction section. In this section, the thesis will aim to first motivate the choice of the method based on the research goals already established, followed by more details on the implementation and algorithm. The method section aims to provide a gentle introduction to the reader about the method developed in each paper. The complete details of the method developed for each paper are provided in the respective article found in the part II.

3.1 Motivation for Method to Achieve RG 1

The RG 1 was to explore the use of deep generative models for the realistic enhancement of digital rock images. A famous deep generative model is the generative adversarial network (GAN)s [43]. GANs are primarily used for image generation. GANs consist of two networks, a generator that generates fake samples and a discriminator capable of classifying real data and fake samples. However, we want to use them to enhance the quality of rock images. One way to enhance the quality is super-resolution, which is the task of increasing the resolution of a given image such that more details are visible in the super-resolved image. The details added by the task of super-resolution are vital for accurate simulation in digital rock analysis. In addition, it is an observation from previous works that most super-resolution techniques developed for natural images can be adapted for digital rock images with appropriate changes.

Super Resolution using Generative Loss Functions Super-resolution can be performed using a neural network in a self-supervised manner. Usually, given a high-resolution image of size h x w, a low-resolution image $\frac{h}{s}$ x $\frac{w}{s}$ is obtained by downsampling by scale factor s. The image obtained is then fed into a neural network. The neural network output is the super-resolved image of size h x w. The super-resolved image is then compared to the high-resolution image using a loss function to update the weight of the neural network using the backpropagation algorithm. The most common form of loss function used for super-resolution is the L_1 loss function, also known as the mean absolute error (MAE), is given as $L_1(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$ where y represents the actual values, $\hat{y}$ represents the predicted values, and n is the number of samples. However, this loss function can lead to blurry images [68]. Previous work has demonstrated that GANs can be used as a loss function in addition to the $L1$ loss. In this setting, the output of the super-resolution from the neural network is sharp and photo-realistic [68] as shown in Fig. 3.1. The generative adversarial loss function derived from GANs is sometimes known in the literature as adversarial loss. In an adversarial loss setting, the generator is the CNN for super-resolution, and the discriminator is a CNN used to create the output of the adversarial loss. Note that both generator and discriminator are neural networks. However, the discriminator is a neural network with fewer parameters than the generator.



MSE

Adversarial loss

Figure 3.1: Adversarial loss based output (right) vs MSE loss based model output (left)

Another type of loss function that provides photo-realistic outputs is perceptual loss [56]. The perceptual loss calculates the style and perceptual differences between the predicted and target images. This loss is calculated by separately passing both the input image and the target image to a CNN that has been pre-trained on a large dataset such as ImageNet. Then, the distance between the feature maps of the neural network is calculated using Euclidean distance.

Therefore, the primary motivation for choosing the methods developed in papers A and B is to use adversarial and perceptual losses such that the networks trained as a result can enhance rock data and reveal crisp details that can aid digital rock analysis. Paper A focuses on super-resolving 2D images, while paper B focuses on super-resolving 3D images.

3.1.1 Method: Paper A

It has been noted from previous works that super-resolution methods developed for colored image datasets such as ImageNet can be applied to super-resolve related datatypes like medical CT and digital rock images [123, 25, 128]. Therefore, the methods developed for paper A are primarily tested on colored images. This method focuses on developing a new method to utilize adversarial and perceptual losses with neural networks for super-resolution.

There are many different types of neural network layers. The most common ones are fully connected layers and CNN [65]. In a CNN, an *discrete* array of the image is the output of the neural network. There is another type of neural network that is know as the implicit neural network (INN) as shown in Fig. 3.2. The implicit neural network takes input coordinates and outputs a color value (RGB) corresponding to the coordinates. It can represent an image in a continuous space as we can input any floating point number as the input coordinates while querying the neural network. This property makes it superior to CNNs as they are discrete. For super-resolution, this means that when we train a CNN to upsample a low-resolution image, then it can be trained and tested for that particular discrete scale. However, in the case of INN, we can use an arbitrary scale. For example, consider a CNN trained to upsample an image by four times. Then, it can not upsample an image up to 10 times as the output of CNN is a discrete array. However, in the case of INN, it is possible due to a continuous representation. The INN is usually parameterized using a multiple fully connected layer as shown in Fig. 3.3.

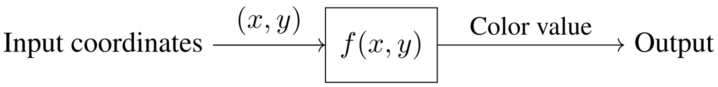


Figure 3.2: Illustration of an implicit function that takes coordinates as input and provides colour values as output.

Adversarial loss and perceptual loss have been previously used with CNN for the task of super-resolution. This combination has resulted in super-resolution output that looks realistic and sharp [68, 121, 122]. However, whether INN can benefit from adversarial and perceptual losses has not been investigated. Therefore, in Paper A, we use adversarial and perceptual losses with INN for super-resolution in pursuit of RG 1 as shown in Fig. 3.4. The rest of the details about the method in paper A can be found in appendix A.

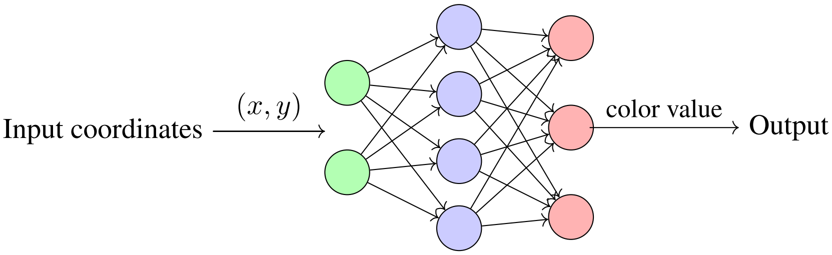


Figure 3.3: Illustration of a neural network-based implicit function that takes coordinates as input and provides color values as output.

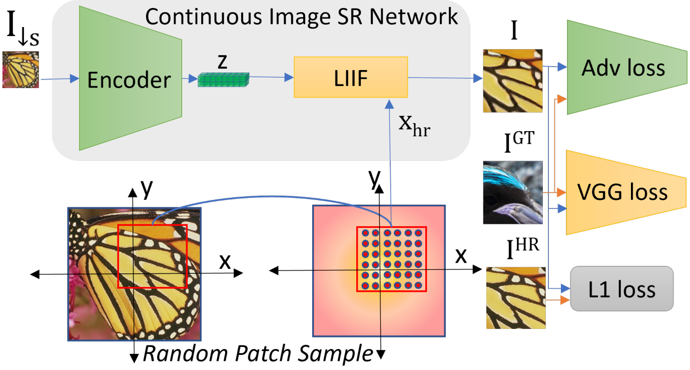


Figure 3.4: Training Method: The low-resolution image $I_{\downarrow s}$ is passed through CNN encoder to get feature vector z. A random patch is selected from the coordinate space of the desired high-resolution image to obtain high-resolution coordinates x_{hr}. z and x_{hr} are passed through the Local implicit function image (LIIF) generator to obtain the super-resolved output image I. This I is compared with I^{GT} using adversarial loss ('Adv loss'), perceptual loss ('VGG loss') and with I^{HR} using pixel loss L_1.

3.1.2 Method: Paper B

Paper B proposes a method to super-resolve a digital rock image using generative models to achieve RG 1. Most samples of rock images are 3D images, e.g. from image sources such as micro-CT. The 3D image is a three-dimensional variant of a 2D image. The 2D images consist of pixels, whereas the basic building block of a 3D image is called a voxel. The CNNs that consume 3D images are called 3D CNNs. 3D CNNs and 2D CNNs are very similar, with 3D CNNs having an extra dimension in the weight matrices. CNNs are very popular for super-resolution of both 2D and 3D images of rocks. Previously, many efforts have been made by previous works to perform super-resolution of digital rock images. However, most of the previous works utilize 2D CNNs [123]. The adversarial and perceptual losses have been used to train 2D CNN to produce realistic rock images. However, it is not possible to realize a 3D CNN equivalent. This is due to the behavior of adversarial and perceptual loss functions with 3D images.

The generator and discriminator are 3D CNNs in the adversarial setting. However, since the 3D equivalent of a 2D CNN increases the GPU memory consumption, therefore training this setting can be prohibitively expensive. Therefore, we developed a method in Paper B to ensure that we can utilize the 2D variant of the GANs since they can be trained more efficiently.

As mentioned before, the perceptual loss uses the weights of a pre-trained neural network that is always a 2D CNN. A 2D CNN can not consume a 3D image directly. Therefore, we can not use a perceptual loss to train 3D CNN directly. However, some approaches that use perceptual loss by calculating it slice by slice on 3D image. However, this does not lead to a better result than a 2D CNN trained with GANs. Therefore, we propose a more elegant method to utilize the perceptual loss for 3D CNN.

The summary of the proposed method is shown in Fig. 3.5. The idea behind the method is to super-resolve a low-resolution 3D image into a high-resolution 3D image through slice-by-slice super-resolution and then fuse the results into a single 3D cube. We propose a method that consists of two parts. The first part is a super-resolution by slice pipeline whereas the second part is a fusing module.

The first part uses a 2D CNN. The 2D CNN is used to upsample the 3D low-resolution image by a certain scale factor. A 2D CNN can not process a 3D image completely. Therefore, we apply a 2D CNN slice by slice along a given dimension. We apply the 2D CNN slice by slice along the x, y, and z axes to obtain three cubes of high resolution as shown in the figure. These cubes are blurry on all faces of the cube except for the two faces along the direction of the slice-by-slice super-resolution. So the next step is to combine the results from all of the cubes into one cube. It is interesting to note that a pre-trained 2D model can also be used for slice-by-slice super-resolution in this first part. In this work, we specifically train the 2D model with adversarial and perceptual loss to achieve RG 1.

In the second part of the pipeline, we use a 3D CNN as a fusion module. That fuses the three cubes as shown in the figure. The output is a 3D high-resolution image. To train this network, we can use the L_1 loss with a 3D ground truth image. In addition, we have also used a loss which is named the consistency loss L_c. This is shown below:

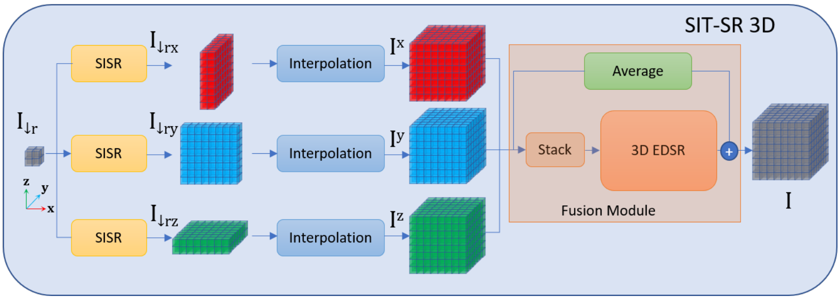


Figure 3.5: SIT-SR 3D: The architecture of our proposed method. The low-resolution image $I_{\downarrow r}$ is upsampled along x, y and z respectively with the pre-trained 2D SISR, as a result, the volumes $I_{\downarrow r_x}$, $I_{\downarrow r_y}$ and $I_{\downarrow r_z}$ are obtained. I^x, I^y, and I^z are the corresponding volumes obtained after the interpolation operation. I^x, I^y and I^z are passed through the Fusion module to get the output volume I.

$$\mathcal{L}_c = \|I^x - I\|_1 + \|I^y - I\|_1 + \|I^z - I\|_1 \tag{3.1}$$

This loss L_c uses the output image from the neural network I and compares it with the three intermediate cubes ($I^x, I^y and I^z$) formed by the first part that uses the 2D CNN. This means that we do not need a ground truth image. We train the whole pipeline end to end in three settings i.e. first with L_1 loss, consistency loss, and finally with a combination of the two. We use rock images to train this pipeline. The details about the algorithm and hyperparameters are given in the paper.

As demonstrated, the method achieves RG 1 by developing a method that can perform 3D super-resolution while using 2D super-resolution networks that have been trained by an adversarial and perceptual loss. The intricate technical details of the method can be found in the appendix B.

3.2 Motivation for Method to Achieve RG 2

The RG 2 was formulated to reduce the conventional dry-wet rock image registration latency. The main conventional method for micro-CT images are intensity-based methods and feature-based methods [40]. Dry and wet rock images are both obtained from the same sensor. Therefore, they are unimodal images. A popular method to perform registration of unimodal images rock images is cross-correlation based image registration [66]. Therefore, we also focus on improving the inference speed of image registration using cross-correlation.

3.2.1 Method: Paper C:

Maximizing the cross-correlation through optimization can determine the transformation between two images. The transformation between the dry and the wet rock image can be in translation, rotation or scaling. When transformation is applied to either wet or dry image, then the cross-correlation between the two is maximized and the images are in perfect alignment and thus registered as shown in Fig. 3.6. The process of optimization using cross-correlation can be time-consuming since it takes several interactions to reach the optimal solution. Therefore, we use an open-source image registration library called AIRLAB [104]. It uses Pytorch [92] to calculate the the analytic gradients of the objective functions such as cross-correlation automatically. This enables the calculations on the graphical processing unit. The result is the the operation can be performed on the GPU.

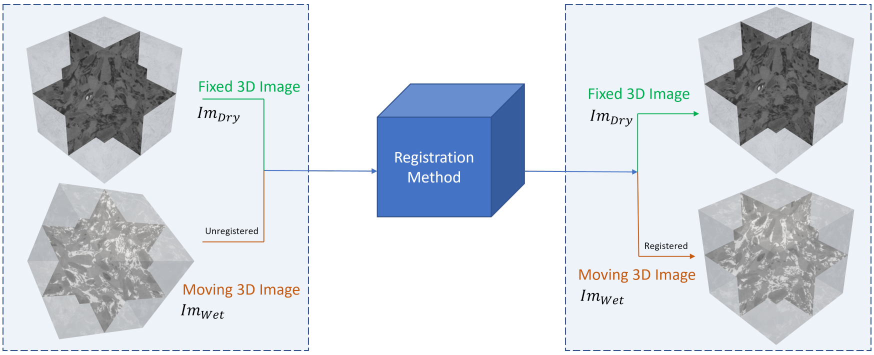


Figure 3.6: We solve the problem of Dry-Wet rock Image registration. Our Registration Method (Blue box) can register Image pairs efficiently and accurately. In this figure, the Fixed 3D Image is the Dry Image, whereas the Moving 3D Image is the Wet Image. The registration method finds the transformation required to warp the Moving 3D image to register it to the Fixed 3D Image.

We develop a GPU-based robust algorithm for image registration that registers the two wet and dry images. Transformations between the two samples can be significant, as demonstrated in Fig. 3.7. For instance, the wet image may rotate around the z-axis by ±180 degrees and around the x and y-axes by ±5 degrees. The sample can be inadvertently placed upside down during handling, resulting in a complete inversion around the x and y axes. The cross-correlation measure is highly dependent on the visuals of the images on which it is applied, and the wet and dry images can appear drastically different. We employ a histogram normalization technique to ensure the images look as similar as possible before cross-correlation optimization. The complete details of the algorithm are given in the paper in the appendix C. However, here we summarize the main steps:

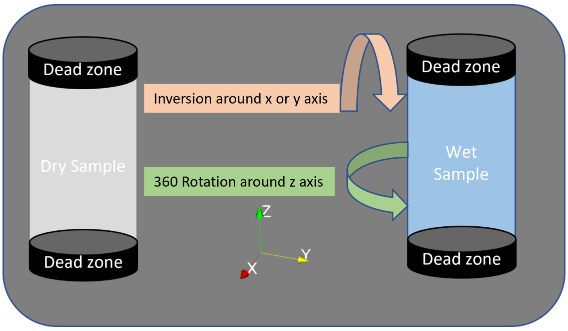


Figure 3.7: Extreme Rotation: A micro-CT image of a wet and dry sample are shown. Note that The wet sample can have extreme rotation of ±180° along z-axis and a 180° inversion around the x or y axes.

Summary of the Proposed Registration Algorithm The dry and wet images may exhibit extreme mismatches, such as a 180-degree inversion or rotation about the z-axis, as illustrated in Fig. 3.7. The dry image serves as the source image, while the wet image is the target image for registration.

The dry and wet images may exhibit extreme mismatches, such as a 180-degree inversion or rotation about the z-axis, as illustrated in Fig. 3.7. The dry image serves as the source image, while the wet image is the target image for registration. The dry and wet images are prepared for registration by performing the necessary pre-processing step of histogram normalization to reveal details in each image. Initially, the method conducts an initial registration step to obtain an approximate match, providing coarse transformation parameters.

This initial registration involves transforming the wet image by angles along the z-axis, checking the registration match for each angle in a complete 360-degree rotation. This process repeats until the full rotation is completed. The registration match is assessed for each angle using cross-correlation, and the highest cross-correlation value is recorded.

Subsequently, the wet image is rotated by 180 degrees about the y-axis, and the rotation with delta angles along the z-axis is repeated. Again, the highest cross-correlation value for each delta angle is noted for both the original wet image and its inverted counterpart. The maximum cross-correlation value between the original wet image and the inverted one is compared, and the highest cross-correlation value is retained.

This highest cross-correlation value serves to determine the initial transformation. Finally, another step of image registration is performed, involving a larger number of steps to ascertain the precise transformation between the wet and dry images.

3.3 Motivation for Method to Achieve RG 3

The third goal, RG 3, was to explore the deep learning-based models for rock typing of rocks that contain laminations, i.e. the samples formed such as one layer is deposited on another, thus creating clear boundaries. Our study focuses on core sample scale and employs porosity trends as the properties for grouping laminar rock regions to determine properties. These properties are then propagated to the whole sample as shown in Fig. 3.8. This choice is motivated by recent works that utilize these trends to calculate region-specific properties. The process can be repeated at different scales until the desired properties are propagated to the desired scale [102].

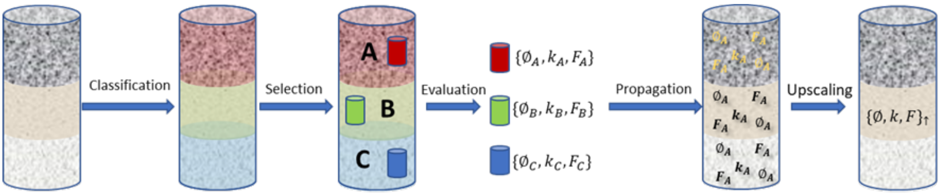

Classification
Selection
A
B
C
Evaluation
$\{\emptyset_A, k_A, F_A\}$
$\{\emptyset_B, k_B, F_B\}$
$\{\emptyset_C, k_C, F_C\}$
Propagation
$\emptyset_A$ F_A k_A F_A $\emptyset_A$
Upscaling
$\{\emptyset, k, F\}_\uparrow$

Figure 3.8: Rock typing and Upscaling Workflow: There are multiple steps involved in upscaling.

Deep learning was chosen to achieve this goal primarily due to the subjective nature of the boundaries that separate the two rocktypes. The boundaries for one expert can be slightly different from the boundaries of another. Therefore methods that are unsupervised and based on clustering can lead to imperfect results. A good solution is to label a dataset using experts and then learn the task of rock typing using a deep neural network. An example of expert labelled image is shown in Fig. 3.9. Apart from rocks that contain lamination, other type of rock types also exist with more complex patterns. However, they are out of the scope of current work.

3.3.1 Method: Paper D

We formulate the rock typing as a semantic segmentation problem [72]. Semantic segmentation is a task where each pixel in an image is assigned a class. In the case of rocks, the class corresponds to the type of the rock, e.g. 0, 1, 2 etc.. The image of the rock sample is processed by a class of neural networks called UNet [101]. UNet has been previously shown in the literature to work well for the task of semantic segmentation. Therefore, we also adapt this neural network for rock typing via semantic segmentation. The Unet is trained by a cross-entropy loss using the images with labels provided by the experts.

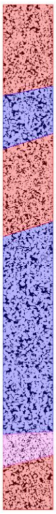


Figure 3.9: Example of labelling of a rock image for rock types

The UNet like all deep learning networks is a black box, therefore the paper utilizes an explainability technique GradCAM [106] to use gradients activations logged from the last layers of the UNet to get an insight into the decision-making process of the neural network. The details of the algorithm and hyperparameter are provided in the appendix D.

3.4 Motivation for Method to Achieve RG 4

This thesis's final goal, RG 4, was to explore using deep generative models called diffusion models for high-fidelity 3D sample creation from 2D rock images [47]. The digital rock workflow is dependent on simulations of a digital rock image. This digital rock image is 3D. However, in certain situations, the 3D image of the rock sample is not available, and only a 2D image is available. In this scenario, it is desirable to use a 2D to 3D synthetic sample generation method. However, creating a 3D image from a 2D slice is an inverse problem. The famous method includes Process-based modelling that is based on the simulation of the physical mechanism of rock formulation through deposition [1, 113, 17, 114, 115]. However, simulation cannot model all the uncertainties of the real samples. Another type of approach is machine learning based models that learn to to generate 3D synthetic samples from 2D data. Previously 2D to 3D generation has been demonstrated to work well using neural networks, but require 3D training data [94]. However, we are interested in an even more difficult task of learning from 2D data to generate 3D synthetic samples. Previously this has been achieved by GANs [59]. However, GANs are known to have issues such as mode collapse that adversely affect the diversity of the generated samples [117]. On the other hand, the diffusion model is a deep generative model that can generate high-quality, diverse samples. However, they are very slow at generating new samples as they iteratively generate a sample by removing noise from a random noise image. To achieve this goal, we combine the GANs with diffusion models to reduce the generation latency while maintaining high-quality sample generation.

3.4.1 Method: Paper E

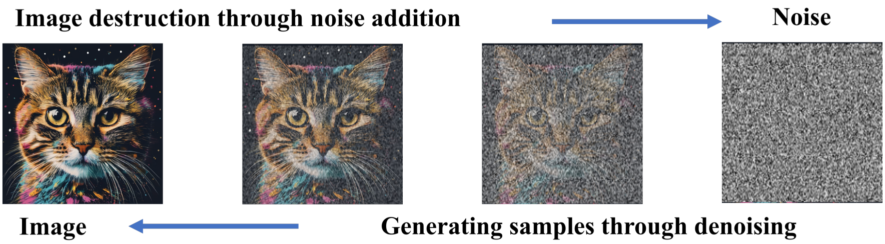


Image destruction through noise addition
Noise
Image
Generating samples through denoising

Figure 3.10: Understanding the diffusion process

Diffusion Model Sampling Procedure Starting from a 2D image, diffusion adds a small noise to the image in multiple steps, as shown in Fig. 3.10. After a sufficient number of steps, the image is destroyed, and now the image is an array of noise. This is the forward process. The reverse process is the opposite, i.e. we start from a pure noise vector and remove noise from the image step by step until we reach an image. In denoising diffusion models using deep learning, the reverse process is modelled by a neural network, e.g. a UNet. This means that starting from a noise vector, the noise is removed step by step using a trained UNet until we obtain an image. This process is also known as sampling. As sampling requires several steps of using the neural network to estimate the noise for each step, it is time-consuming. From the theory of diffusion models, it is known that the more steps there are, the better the quality of the generated sample. In practice, the number of steps is 1000 to 2000 steps.

Diffusion Model Training Procedure Until now, we have only discussed sampling, the inference step. However, we first need to train the UNet before it can be used for sampling. We need pairs of input and output data to train the neural network. The input of the UNet is the noisy image, and the output of the UNet is the noise present in the image. To formulate the pairs, we return to the forward diffusion process that generates several noisy image chains by adding noise to the clean image. For each step of the forward process, we know how much noise we have added to create a noise image. Therefore, we form pairs of input and output training data using the noisy image and the noise added to obtain the image during the forward process. We do this for all the images in the training dataset to obtain a large data set of images and noise to train the UNet.

GANs for Diffusion Model GAN consists of a generator and a discriminator. In a GAN, the generator creates synthetic samples, and the discriminator learns to classify the generator samples as fake. In this adversarial game, both the generator and discriminator get better. The generator network generates the sample in one single shot from noise. This setting is also the source of GAN problems like mode collapse [117]. The generator is forced to generate the sample in several steps to convert this GANs to a diffusion GAN. This ensures stability.

2D to 3D Generation We use the diffusion GAN in the proposed method. However, we modify it to learn to generate 3D samples by learning from 2D data. The key idea is to start from a noise cube and denoise it using a 3D UNet, as shown in Fig. 3.11. To train the 3D UNet, we need paired 3D data from the forward diffusion process. However, we do not have it; we only have real 2D data. Therefore, the sampling/ inference process was used to generate the simulated 3D data using 3D UNet. This is not ideal as 3D UNet is untrained at the beginning of training. However, this works mainly because following the diffusion GAN, we use the output of the 3D UNet generator to obtain 2D slices along the x, y, and z-axis of the output cube and feed it to a 3D discriminator. In addition, the discriminator also receives the real samples from 2D real image data. This means we use the adversarial loss training signal for both the generator and the discriminator. This provides the necessary training for the 3D UNet and allows it to learn to generate noise-free synthetic 3D images from 2D real data.

Since we do not have any real 3D data, we store all the samples generated by the 3D UNet in a data bank and use them as training data to train the 3D UNet. This strategy slowly teaches the 3D UNet to denoise the cube. When the training finishes the fully trained 3D Unet can generate novel 3D rocks by denoising a random cube of noise. A new novel sample can be obtained by initializing a random cube of noise and denoising using the sampling procedure with 3D UNet. The intricate details of the method are provided in the appendix E.

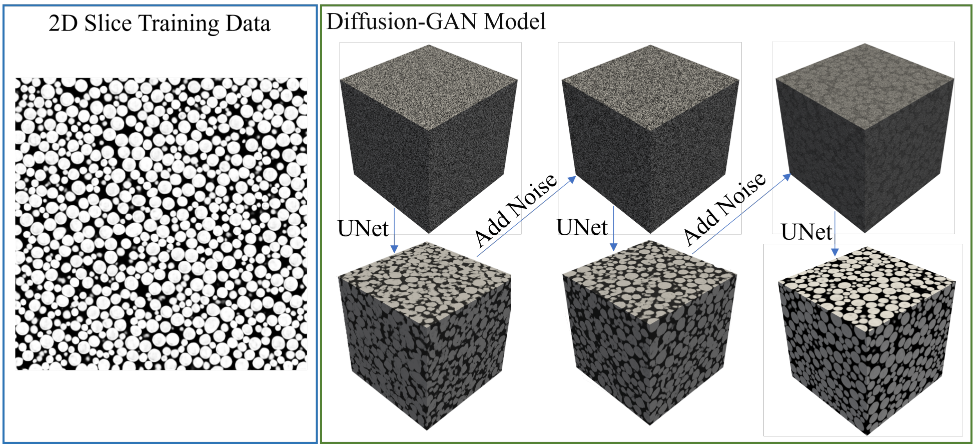


Figure 3.11: Diffusion-GAN Model: The proposed method is based on a denoising diffusion process combined with a generative adversarial framework. In this setting, at test time, starting from a cube of noise, the noise is iteratively estimated using a Unet, removed, and then added back to the sample. This process is repeated to obtain a noise-free representative sample. Our method can learn from only a few 2D slices of the training image, as shown on the left.

Chapter 4

Results

This chapter presents the results obtained from the new digital rock imaging and analysis pipeline due to the novel methods developed in this work. Each method affects the digital rock analysis pipeline positively. The overall effects of the methods can be one of the following:

- Increases the efficiency by reducing the time it requires to complete the rock analysis.
- Improves the accuracy of the analysis by enabling more accurate simulation and property determination.
- Improves both the accuracy and efficiency of analysis.
- Provides a novel alternative method to perform digital rock analysis.

The research objectives were structured around four key research questions to improve different aspects of rock imaging and analysis. These questions focused on enhancing image quality, improving image registration methods, automating rock typing processes, and refining synthetic rock sample generation techniques. The digital rock analysis and imaging pipeline improvements obtained from achieving each research goal are provided below. In addition, Table. 4.1 summarises the improvements due to each method.

Table 4.1: Result of the New Analysis Pipeline

Research Questions	Research Goals	Paper	Contributions	Improvement in Pipeline
RQ 1: Enhance the quality of images from CT	RG 1: Explore use of generative models for realistic enhancement of digital rock image	A	Proposed a generative model for realistic image super-resolution of 2D images	Increased accuracy of analysis
		B	Proposed a method for utilizing 2D generative model for 3D image super resolution for rock images	Increased accuracy of analysis
RQ 2: Improve current dry-wet image registration methods	RG 2: Reduce latency and improve robustness of conventional methods	C	Formulated a registration tool that utilizes graphical processing units to reduce method latency from hours to minutes	Reduced time requirement for analysis
RQ 3: Automate rock typing pipeline	RG 3: Explore deep learning-based methods for rock typing of laminar rocks	D	Proposed a supervised deep learning method for rock typing of laminar rocks and performed explainability analysis on model	Increased accuracy and reduced time for analysis
RQ 4: Improve methods to generate high fidelity synthetic rock samples	RG 4: Explore deep generative models for 2D to 3D generation	E	Proposed a novel method based on deep generative diffusion models for 2D to 3D image generation	Increased accuracy for existing analysis and novel analysis possible

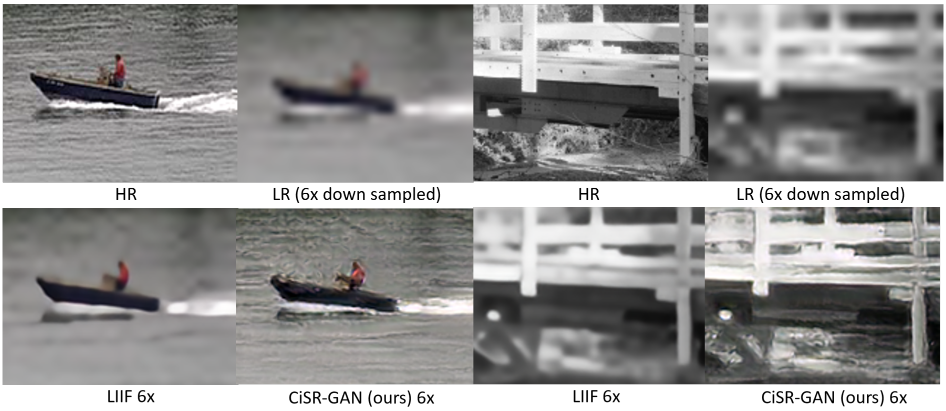
4.1 RQ 1: Realistic Enhancement of Image Quality via Generative Models

The first research goal aimed to enhance the quality of images obtained from CT scans. Two papers, labelled A and B, proposed novel generative models for realistic image super-resolution of 2D and 3D images, respectively. Implementing these models significantly increased the chosen image quality metrics on famous distortion and perception-based metrics such as PSNR and LPIPS metrics [74, 130]. In addition, the methods also demonstrated superior performance due to improved image clarity and photo-realistic outputs from models.

After applying these methods to the related rock image data, the resolution of the image quality will be better. This provides a better image for the digital rock pipeline, thus potentially increasing the accuracy of the rock properties determined through analysis and simulation.

4.1.1 Salient Results from Paper A

By employing perceptual and adversarial loss functions with implicit neural networks known for generating photorealistic images, the method in paper A has taken a step towards resolving the challenge of upsampling micro-CT images to achieve SEM-level detail. Our qualitative and quantitative experiments on coloured images showcase the potential of this approach as shown in Fig. 4.1 and Table. 4.2. The results in the figure show that the proposed method generates photo-realistic output compared to mean square error trained only implicit neural network-based method [28]. In addition, the comparison in Table. 4.2 with other state-of-the-art CNN-based neural networks demonstrates that our method surpasses other networks on perception-based metrics such as LPIPS on benchmark datasets. Previous works [123] have demonstrated that performance on generic datasets correlates with the performance on digital rock datasets. So, even though digital rock datasets were not tested with this pipeline, the method will demonstrate similar results on rock data.



HR

LR (6x down sampled)

HR

LR (6x down sampled)

LIIF 6x

CiSR-GAN (ours) 6x

LIIF 6x

CiSR-GAN (ours) 6x

Figure 4.1: Qualitative Comparison on Set 14[129]:. This figure shows the high-resolution ground truth image (HR), the low-resolution image (LR), the super-resolved image using LIIF model [28] and our model's output. All input images are 6x down-sampled from ground-truth images and super-resolved to 6x. All models were trained for 1x-4x only. We observe the same smoothing effect for LIIF outputs where the high-level details such as water waves and texture in the fence has been blurred, while our model retains the high-level details and the image produced is much more realistic than LIIF.

Dataset	Metric	SFTGAN [122]	SRGAN [68]	ESRGAN [121]	SPSR [75]	CiSR-GAN (ours)
Set5	LPIPS	0.0890	0.0882	0.0748	0.0644	**0.0604**
	PSNR	29.932	29.168	**30.454**	30.400	30.05
Set14	LPIPS	0.4393	0.1663	0.1329	0.1318	**0.1160**
	PSNR	26.100	26.171	26.276	**26.640**	26.62
B100	LPIPS	0.5249	0.1980	0.1614	0.1611	**0.1436**
	PSNR	25.961	25.459	25.317	25.505	**25.72**
Urban100	LPIPS	0.4726	0.1551	0.1229	0.1184	**0.1179**
	PSNR	23.145	24.397	24.360	**24.799**	24.36

Table 4.2: Quantitative comparison with CNNs on benchmark datasets This table shows our method with other perceptual quality-focused methods. The best results are in **bold**. All models have been trained and tested on 4x down-sampled images.

4.1.2 Salient Results from Paper B

As depicted in the associated Fig. 4.2, the results compare the performance of our model (SIT-SR 3D) vs other state-of-the-art, i.e. 3D ESRGAN [121]. RRDBNet is the ESRGAN trained without the adversarial losses. The figure shows that the proposed method produces photorealistic and sharp results. ESRGAN also produces sharp results. However, it has been trained with the 3D ground truth using 3D GANs and perceptual losses, requiring many more parameters and a much larger computing budget. On the other hand, our method required only a 2D ground truth to train and much fewer compute resources while producing comparable results. The proposed method enhances the resolution and reveals the intricate details in the rock image used for analysis, thus promoting accurate determination of properties through simulation and analysis. These results contribute to answering RQ1 by enhancing the image quality of the 3D digital rock image.

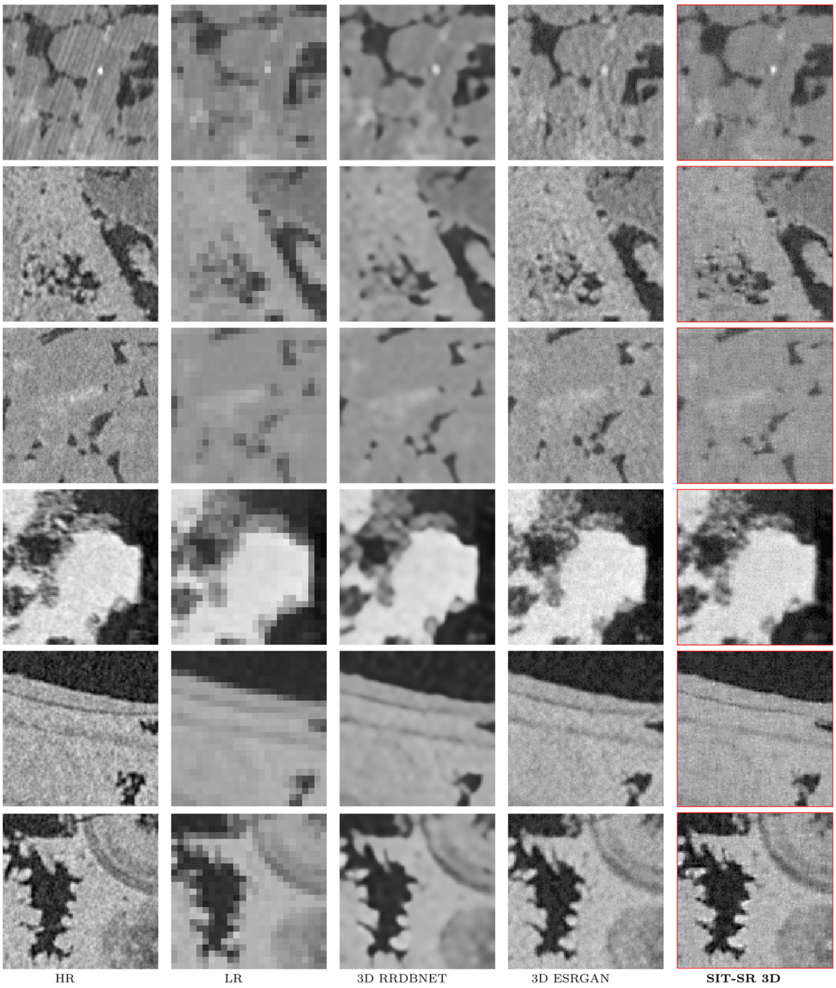


Figure 4.2: This figure shows a visual comparison of different methods. HR, LR indicates high-resolution and low-resolution images. 3D RRDBNet is a 3D convolution-based network supervised with L1 loss. 3D ESRGAN is trained with GAN and VGG loss using pre-trained weights of 3D RRDBNet. SIT-SR 3D is trained in a self-supervised setting using only the consistency loss.

4.2 RQ 2: Reduced Latency and Increased Robustness of Image Registration

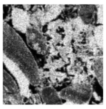
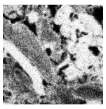
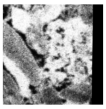
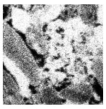
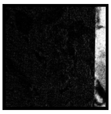
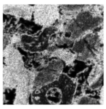
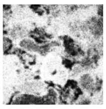
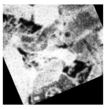
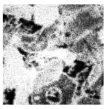
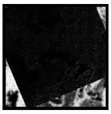
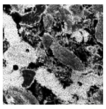
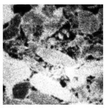
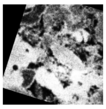
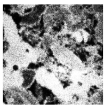
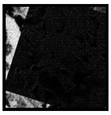
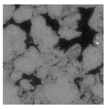
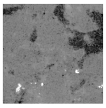
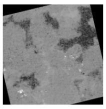
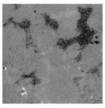
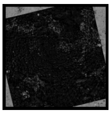
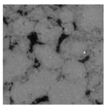
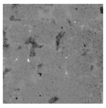
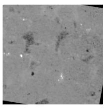
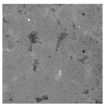
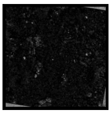
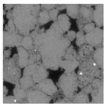
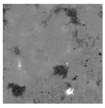
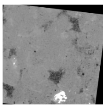
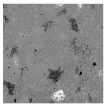
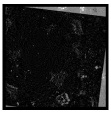
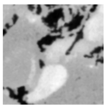
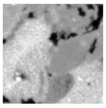
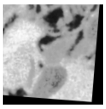
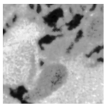
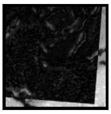
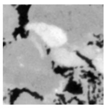
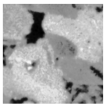
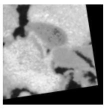
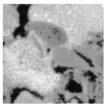
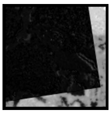
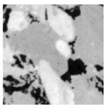
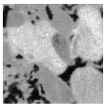
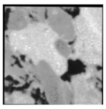
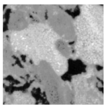
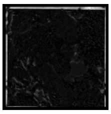
To achieve the second research goal, efforts were made to improve existing dry-wet image registration methods. Paper C focuses on reducing the processing time of the image registration task while maintaining accuracy.

The image registration is a part of the digital rock analysis pipelines. By reducing the time it takes to perform wet-dry rock image registration from hours to minutes, this method has successfully reduced the time required for completing the analysis.

4.2.1 Salient Results from Paper C

The proposed image registration method's performance on dry-wet image sample pairs from each rock dataset considered in paper C is shown in the Table. 4.3. The algorithm calculates the transformation needed to transform the 'Moving Image', which is the wet image and align it with the 'Fixed Image', which is the Dry image. The table compares the warped moving image with the 'Ground Truth (GT) Moving Image' and displays the quantitative parameters of the transformation: translation in x, y, and z in pixels (Trans X, Trans Y, Trans Z), rotation in x, y, and z respectively in degrees (Angle X, Angle Y, Angle Z), and scaling in x, y, and z respectively (Scale X, Scale Y, Scale Z). The figure demonstrates that even when the original and moving images are extremely misaligned, our method can successfully find the transformation between the dry and wet images, and it can do so in under a minute, thus answering RQ2. Our method has been developed as a tool for use in industry.

Table 4.3: Dry Wet Image Registration

Image Dataset	Fixed Image	Moving Image	Warped Moving Image	GT Moving Image	Difference Image	Parameter	Prediction	Ground Truth
[120]						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-27.73 1.04 -16.20 -4.83 177.09 80.91 1.05 1.05 1.04	-28.00 1.00 -16.00 -4.82 177.06 80.88 1.04 1.04 1.04
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	10.22 -27.92 -16.92 -1.93 1.15 -57.06 1.03 1.03 1.03	10.00 -28.00 -17.00 -1.93 1.13 -57.00 1.02 1.02 1.02
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	27.89 -12.35 3.15 -4.74 -3.36 162.23 1.02 1.02 1.03	28.00 -12.0 3.00 -4.74 -3.36 162.23 1.02 1.02 1.02
ST C14						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-6.10 3.80 -9.88 -3.34 0.16 93.23 1.08 1.01 1.00	-5.00 3.00 -10.00 -4.38 -0.78 94.17 0.97 0.97 0.97
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-4.49 -0.50 2.06 -0.69 179.69 -90.27 1.03 1.07 1.14	-3.00 -1.00 3.00 -1.59 178.57 -87.77 1.00 1.00 1.00
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-7.92435 10.66 -2.86 3.33 0.34 -7.11 1.00 1.00 1.03	-10.0 11.00 -2.00 3.94 0.85 -7.87 0.98 0.98 0.98
[111]						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-10.71 -8.17 9.45 2.70 184.28 -35.39 1.06 1.02 1.02	-10.00 -9.00 10.00 2.23 184.14 -35.77 0.99 0.99 0.99
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-13.29754 -16.31 -5.65 -1.14 -1.25 128.61 1.02 0.98 0.99	-13.0 -16.00 -6.00 -1.76 -1.75 128.53 0.98 0.98 0.98
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	0.70478 1.92 10.57 -2.82 4.28 -20.30 0.97 0.97 0.97	1.0 2.00 11.00 -2.83 4.44 -20.25 0.96 0.96 0.96

4.3 RQ 3: Automation of Rock Typing via Deep Learning

The third research goal focused on automating the rock typing pipeline. Implementing the method in Paper D resulted in a substantial increase in the accuracy and a significant reduction in the time required for rock typing analysis.

Experts predominantly do rock typing manually since the boundaries between various rock types in the case of lamination can be subjective. Therefore, the analysis takes substantial time. It is also possible that human interaction can generate errors. Therefore, the use of the model automates this laborious process. The process of rock typing is important for the accurate determination of the properties of the sample. Therefore, the introduction of the method has resulted in an improvement in both the accuracy and efficiency of digital rock analysis and imaging.

4.3.1 Salient Results from Paper D

The results of our deep learning-based rock typing method, as demonstrated in Fig. 4.3, exhibit strong performance. The output of the model is a segmentation mask. Compared to the ground-truth mask, the output of the results shows good segmentation capability of the proposed deep learning model. After dividing the rocks into rock types, the digital rock analysis can determine the properties of each rock type segment separately instead of treating the whole sample. This will lead to more accurate property determination. Additionally, the neural network takes only one to two seconds to perform this compared to much more time required by a human expert, thus addressing RQ 3.

Additionally, we examine the decision-making process of the neural network using a famous explainability method, gradient-weighted class activation mapping [106]. This analysis provides valuable insights into the black box model's decision-making process as shown in Fig. 4.4. In particular, for each chosen rock type, we check which pixels in the image were responsible for the decision made by the neural network. The figure clearly shows that the chosen rock-type pixels in the input image were highlighted by the explainability method. These pixels are responsible for the decisions of the neural network. This analysis increases our confidence in the black box model by demonstrating that the model is indeed looking at the relevant regions to make a decision and not overfitting the dataset.

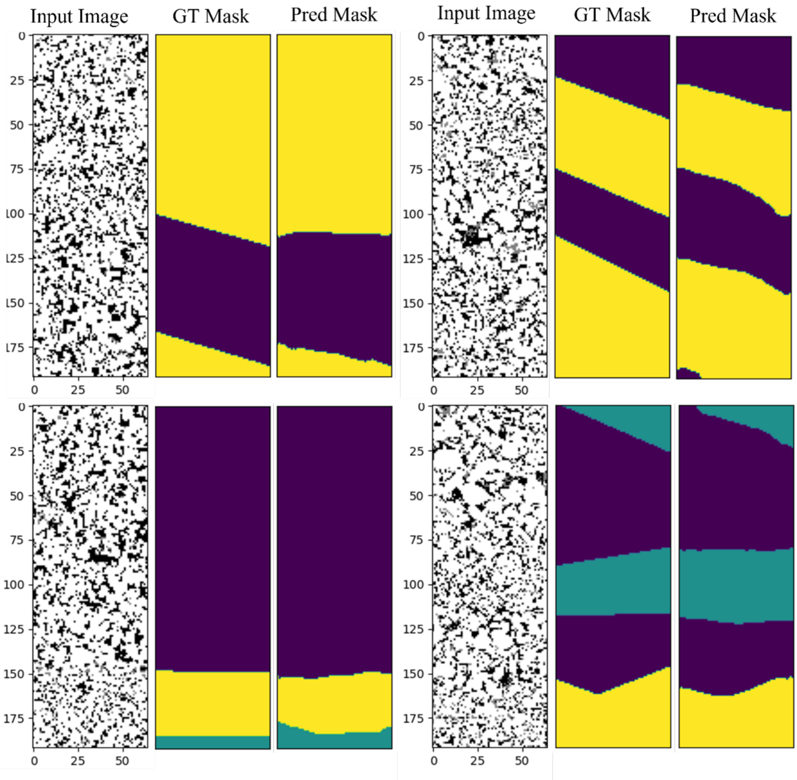


Figure 4.3: The Qualitative results of our model are shown. From left to right, we have the input image, ground truth segmentation mask and predicted segmentation mask.

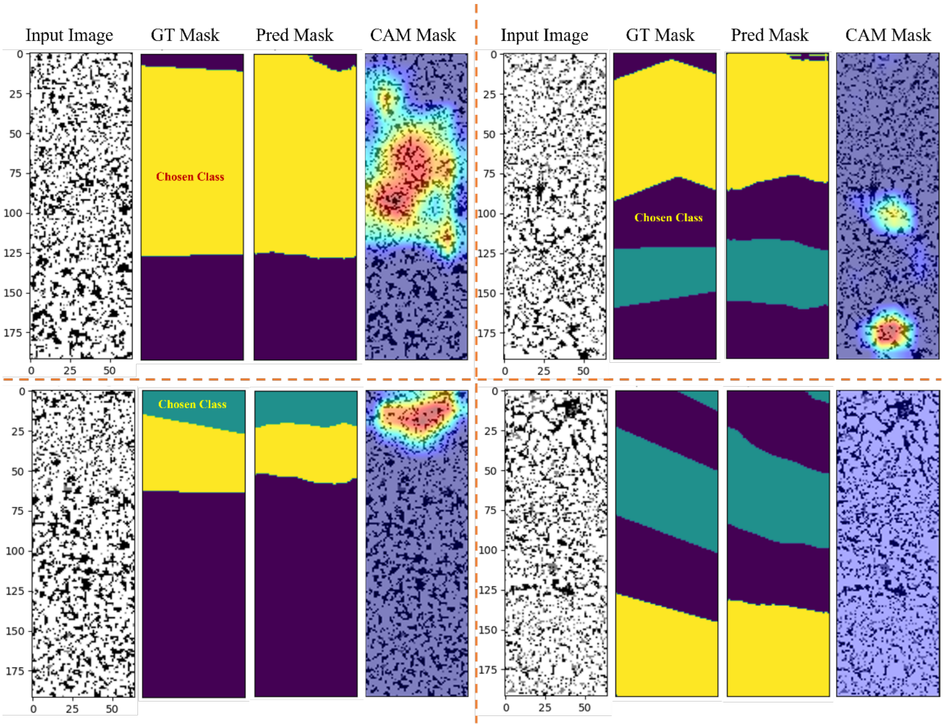


Figure 4.4: Grad Cam Explain ability Analysis: We query each class as shown by the tag 'Chosen Class' in the GT masks. It can be seen that the CAM mask image shows the area where the neural network is paying attention for obtaining the Pred Mask.

4.4 RQ 4: Synthetic 3D Rock Generation from 2D Image via Diffusion Models

Finally, the fourth research goal aimed to improve methods for generating high-fidelity synthetic rock samples. Paper E introduced a novel method for 3D image generation from a single 2D slice of a micro-CT image of a rock. The generation of 3D rock sample using the proposed method produces high-quality, diverse synthetic samples that represent the grain distribution in the real sample accurately.

Multiple analyses on digital rocks require a 3D image. e.g. determining the fluid flow using simulation. Given a single slice of a 2D rock, the proposed method determines the synthetic 3D image of the rock sample, which can be used to run simulations for fluid flow and other properties. Therefore, this method addresses RQ 4 and opens avenues for novel analysis by providing more realistic synthetic samples.

4.4.1 Salient Results from Paper E

The results of our deep learning-based generation method, as demonstrated in Fig. 4.5, exhibit strong performance compared to previous state-of-the-art [59]. Our model can generate 3D images from only a few 2D image ground truth slices. The most challenging structures to generate have been shown in this figure, where our model excels at capturing the distribution of the rock structure much more than SliceGAN. Along with capturing the distribution, we notice that our model can also capture the colour distribution of the original 2D image in a better manner.

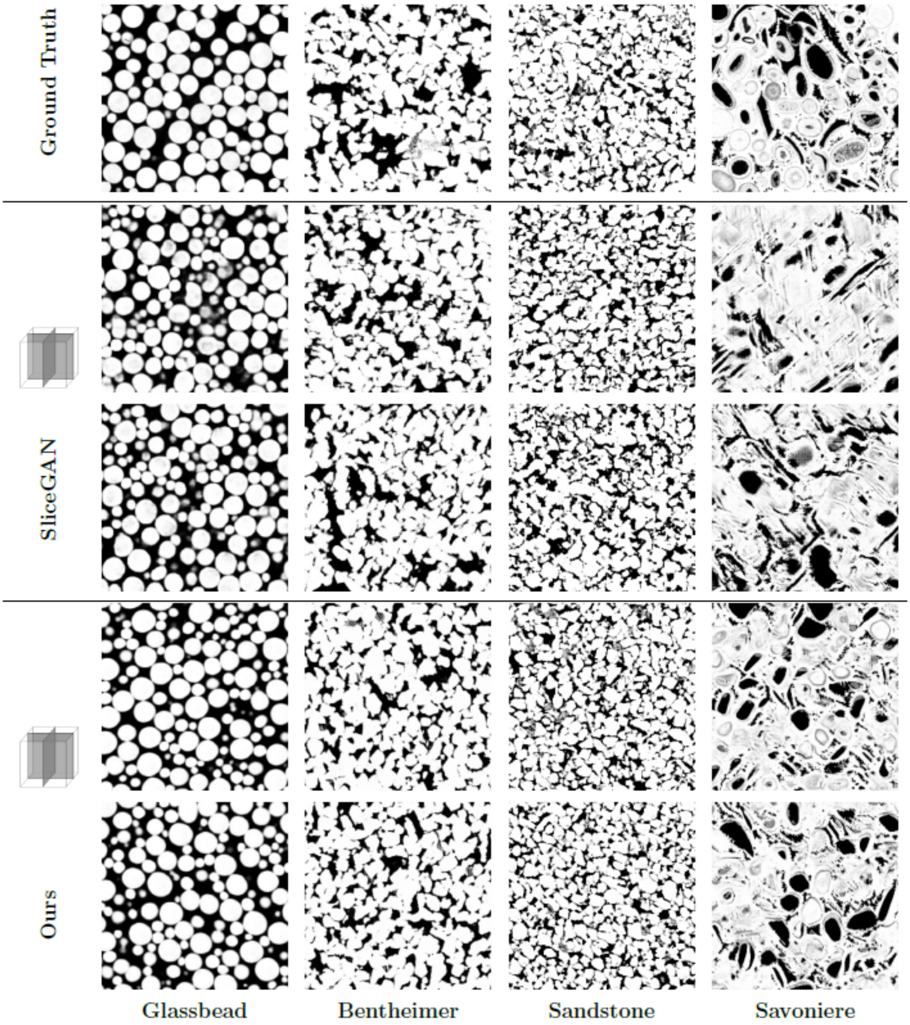


Figure 4.5: Visual comparison with micro-CT images: Cross-sections of 3D images generated by our method and SliceGAN [59], alongside their respective ground truth or training data. The ground truth images are 3D X-ray micro-CT scans obtained at varying resolutions. The Glassbeads case showcases our method's superior performance over SliceGAN. Our model can capture the spherical shape of the object, even though it only sees circles at the 2D input. In more challenging cases like the Savoniere - a carbonate of fossilized microorganism - our method proves its robustness by generating images that bear a higher resemblance to reality, despite the heterogeneous nature of the original image.

Overall, the performance assessment of the new analysis pipeline resulted in a considerable improvement in the efficiency and performance of rock imaging and analysis processes. The contributions made by each method addressed specific challenges in the field, leading to enhanced accuracy, reduced processing time, and the facilitation of automated analysis tasks. The combined impact of these advancements demonstrates the efficacy of the proposed pipeline in advancing the state-of-the-art in rock imaging and analysis.

Chapter 5

Discussion

In this chapter, the overarching contributions of the research work are discussed, aligning the results from the different papers with the identified research questions and addressing the gaps in the field of digital rock analysis. A comprehensive overview of each paper's key outcomes is provided, linking them to the overall research plan and the identified research goals.

5.1 Plan Revisited

Before delving into the detailed discussion of novel contributions of each paper, it is pertinent to revisit the summary of research questions and the improvement to the digital rock pipeline outlined in Table. 5.1. We will refer to this Table to summarize the novel contributions of each paper and identify the gaps addressed by each of the papers for digital rock analysis. The Table demonstrates that the research contributes novel methods that improve the pipeline of digital rock analysis in terms of efficiency by saving time and improving the accuracy of the properties determined by the digital rock pipeline.

Table 5.1: Discussions of the New Analysis Pipeline

Research Questions	Paper	Contributions	Improvement in Pipeline
RQ 1: Enhance the quality of images from CT	A	Proposed a generative model for realistic image super-resolution of 2D images	Increased accuracy of analysis
	B	Proposed a method for utilizing 2D generative model for 3D image super resolution for rock images	Increased accuracy of analysis
RQ 2: Improve current dry-wet image registration methods	C	Formulated a registration tool that utilizes graphical processing units to reduce method latency from hours to minutes	Reduced time requirement for analysis
RQ 3: Automate rock typing pipeline	D	Proposed a supervised deep learning method for rock typing of laminar rocks and performed explainability analysis on model	Increased accuracy and reduced time for analysis
RQ 4: Improve methods to generate high fidelity synthetic rock samples	E	Proposed a novel method based on deep generative diffusion models for 2D to 3D image generation	Increased accuracy for existing analysis and novel analysis possible

5.2 RQ 1: Enhance the Quality of Images from CT

5.2.1 Paper A: Enhancing Image Quality with Generative Models

Novel Contributions Paper A addresses RQ 1 by proposing a generative model for realistic image super-resolution of 2D images. The novel contributions of the work are given below:

- The paper proposes to use an INN with adversarial and perceptual losses.
- The adversarial loss produces photo-realistic results.
- The INN ensures continuous super-resolution i.e. using a single INN model we can upsample 10x even though the network was only trained for 1x to 4x.

Addressed Research Question The above contributions address the research question RQ 1 by providing a pipeline that can enhance the quality of all types of images photo-realistically. Even though the paper demonstrates the results on generic images, this is not a limitation of the work, as superior performance has been demonstrated on multiple famous benchmark datasets. Therefore, it is simple to extend the results of the pipeline to a rock dataset. The results on benchmarks indicate the generalization capability of the method. It indicates that the quality of the digital rock analysis will enhanced by using the method contributed in Paper A.

Targeted Research Gaps The research gap identified at the beginning of the study was that the images coming from micro-CT could be enhanced to reveal important, intricate details crucial for property determination. However, the current state-of-the-art deep learning methods based on CNNs do not generalize well for a scale (e.g. 10 x) outside the training scale (e.g. 1 x to 4x). In addition, the methods based on (INNs) that generalize well for a scale outside the training scale do not produce sharp results. Therefore, there was a gap that needed to be addressed by combining INNs) with adversarial losses. To address this gap, the paper formulated a pipeline that trained INNs with adversarial losses. The resulting pipeline produced super-resolved images with enhanced details. The enhanced details can lead to superior simulation performance and finding accurate rock properties. When we combined INNs with adversarial loss, we obtained the benefits of both INNs and adversarial loss. This means the method obtained sharp images that could be zoomed into much more than the training scale.

5.2.2 Paper B: Utilizing Generative Models for 3D Image Super Resolution

This paper also addresses RQ 1 by enhancing the quality of images from CT. The proposed model provides a very high-resolution (4x) rock image. This high-resolution image contains important and intricate details that can be used for accurate property determination and simulations. The focus of this method was also on photo-realistic quality thus the produced images are crisp. This addresses the research questions. The novel contributions of the work are given below:

- The method is a self-supervised method for 3D image super-resolution which we train using a proposed consistency loss with only 2D ground truth.
- The method can adopt state-of-the-art 2D pre-trained models out of the box, thus being more flexible than end to end 3D super-resolution method. As a direct consequence, the method developed in Paper A can be adopted in the 2D super-resolution pipeline.
- The approach is data and compute efficient compared to end-to-end deep 3D super-resolution alternative. It requires fewer parameters (one-third) compared to the 3D super-resolution models.

Addressed Research Question The above contributions address the research question RQ 1 by providing a pipeline that can enhance the quality of 3D rock images photo-realistically. This work answers the question by specifically demonstrating the results of the method on the micro-CT rock images.

Targeted Research Gaps The work addressed important gaps that existed in the literature. The current photo-realistic 3D super-resolution method requires 3D GANs and 3D perceptual losses. 3D GANs are computationally heavy to train, and 3D perceptual losses did not exist. In addition, many state-of-the-art 2D super-resolution pipelines exist. In the literature, there was no way to utilize these 2D super-resolution pipelines to perform 3D super-resolution or train a compute-efficient 3D super-resolution pipeline. The method proposed in this work provides a compute-efficient way of training a 3D super-resolution pipeline from scratch using only 2D data (self-supervised). It also supports using the pre-trained weights from any existing state-of-the-art 2D super-resolution pipeline. Thus adequately addressing the research gap that existed.

5.3 RQ 2: Improve Current Dry-Wet Image Registration Methods

5.3.1 Paper C: Accelerating Dry-Wet Image Registration with GPU

Addressing RQ 2, Paper C introduced a novel registration tool utilizing GPUs to reduce the latency of conventional dry-wet image registration methods. By harnessing the computational power of GPUs, the proposed tool significantly reduces the registration time from hours to minutes, enhancing the efficiency of digital rock analysis workflows. This advancement is crucial for expediting data processing and analysis in the field of digital rock characterization.

- The paper proposes a robust pipeline for image registration under extreme rotation and transformations of wet and dry image registration
- The proposed method can complete the task in minutes due to the usage of parallel processing of the optimization problem.
- The method was validated on synthetic transformation on three different rocks to demonstrate the effectiveness of the method

Addressed Research Question The RQ 2 was to improve the current dry-wet image registration method. This method was developed for industrial applications. Therefore it was best to avoid the deep learning-based method due to lack of enough data. The classical image registration was the best candidate since they were tried and tested in the industry. Therefore, we focused on improving the latency of the current image registration method thus addressing the research question.

Targeted Research Gaps The gap existed in the literature to reduce the latency of the image registration algorithm and utilize the increased speed. The latency of current dry-wet registration methods was reduced by improving the implementation to a Pytorch-based framework[104]. This new implementation reduced the latency drastically. The paper utilized the increased speed by proposing a novel wet-dry image registration algorithm with enhanced robustness. The algorithm proposed is capable of being robust under extreme rotation. Therefore, it will work under extreme requirements that are accompanied by industrial deployment and fill the existing gap for a tool.

5.4 RQ 3: Automate Rock Typing Pipeline

5.4.1 Paper D: Deep Learning-Based Rock Typing of Laminar Rocks

Paper D contributes to addressing RQ 3 by proposing a supervised deep learning method for rock typing of laminar rocks. Through extensive experimentation and explainability analysis, the paper demonstrates the efficacy of deep learning-based approaches in automating the rock typing pipeline. This not only streamlines the characterization process but also provides insights into the underlying geological features contributing to rock typing decisions. The contributions of the model are given as follows:

- The paper formulates the rock typing problem as a supervised semantic segmentation problem.
- A dataset labeled by industry experts was collected for training.
- The explainability analysis of the proposed model was performed to enhance the trustworthiness of the model.

Addressed Research Question The research question of automating the rock typing pipeline was adequately addressed based on the above contributions. The choice of supervised method was important for this problem as the boundary between rock types was subjective. Therefore, successful automation depends on getting labels from an expert instead of an unsupervised method. However, the adoption of a deep learning-based method introduces a black box method. It was important to perform an explainability analysis to ensure that the outputs of the model can be trusted.

Targeted Research Gaps The gap in research was based on the industry's need for an automated rock typing pipeline. In this regard, the network (UnNet) and explainability analysis tools (GradCAM) already existed. However, there was a gap in terms of data collection and then training the model with the appropriate neural network architecture with this data. The proposed method not only proposed an optimal training formulation but also analyzed the neural network with an explainability method to enhance confidence in the black box tool. Therefore, an important industrial application was addressed in this paper.

5.5 RQ 4: Improve Methods to Generate High Fidelity Synthetic Rock Samples

5.5.1 Paper E: Deep Generative Models for Synthetic Rock Sample Generation

Lastly, Paper E tackles RQ 4 by introducing a novel method based on deep generative diffusion models for 2D to 3D image generation of synthetic rock samples. By leveraging deep generative models, the proposed method enables the generation of high-fidelity synthetic rock samples with realistic structures and textures. This contributes to advancing the field of digital rock analysis by providing researchers with synthetic datasets for training and validation purposes, ultimately improving the accuracy and reliability of rock characterization algorithms. The contribution of this method are given below:

- The proposed method is based on diffusion GAN that ensures the stable and high-quality generation of 3D images by training only on 2D samples.
- The proposed method can learn efficiently from very few 2D images
- The proposed method was tested for the generation of many different 3D structures including geological rock data.

Addressed Research Question The research question addressed the need to improve generative methods for 3D high-fidelity data from 2D samples. By addressing the question in Paper E, 3D data can now be produced using only 2D data using better generative models. This ensures that simulations can be performed in low or no 3D data situations addressing RQ 4.

Targeted Research Gaps The gap existed in the literature since the previous state of the art was mainly based on GANs, which had qualities like one-shot generation and some drawbacks like mode collapse [117]. The work identified that diffusion models are better than GANs at generation quality. However, they had their drawbacks, like slow inference speed. Therefore, to get the best of both worlds, the models were combined to get diffusion GAN. Equipped with this tool, a novel 2D to 3D generation pipeline was proposed based on the diffusion-GAN pipeline that excelled at both inference speed and high quality of image generation, thus addressing the gap.

Chapter 6

Conclusion

6.1 Final Reflection

This thesis has presented a comprehensive suite of innovative solutions to add value to the digital rock workflow in an industrial environment by implementing advanced concepts of machine learning, deep learning, and computer vision. This body of work targets the key component of digital rock workflow, which is image processing. The thesis made a case for the importance of image processing by observing that the complete pipeline is dependent on the image. The better the image provided at the image processing step, the better the analysis that can be carried out in the rest of the pipeline. The image processing part of the digital rock analysis can be improved in several ways. The thesis formulated research questions related to the improvement of digital rock pipelines. However, the research goals were fine-tuned based on gaps in the literature as well as industrial requirements. To achieve the goal, deep learning methodologies were found to be the best candidate to achieve three out of the four goals set out in this work; however, a non-deep learning-based method was used for one of the goals. However, the parallel processing capability made available due to the advent of deep learning was used to speed up conventional methods. In this way, all the goals achieved in this thesis benefit from the latest developments in deep learning and computer vision.

Each goal achieved in this work resulted in academic contributions in the form of new knowledge and industrial value. The contributions affected the digital rock workflow by either improving the accuracy and efficiency of the analysis or both. In the case of the generation of a new rock sample from 2D images, the methods developed made novel analysis feasible. In short, the work demonstrates the application of deep learning and computer vision for digital rock analysis.

6.2 Limitations

The work presented in this work has several limitations.

6.2.1 Size of 3D Rock Images

The images of the rock samples obtained are enormous and can even be several gigabytes. This leads to several delays in the pipeline due to the time it takes to visualize these images.

These huge images can quickly fill up a computer's entire Random Access Memory (RAM). Loading these images into the RAM can take several minutes. This presents a significant challenge and an essential choice over the method for loading these images, as loading and unloading data can take several minutes. The first choice is to load all the 3D data into the CPU's (Central Processing Unit's) RAM and perform training. The second option is to read the data from the hard disk as required. In the first option, loading the data on the RAM can take time; however, once the data is loaded, the RAM reading is fast. In the second option, the data is read from the hard disk. This can be very slow if a huge image has to be read repeatedly from the disk. A hybrid approach based on the number of files in the complete dataset and system specification works best.

The training using famous scripting languages such as PyTorch can quickly reach the computing limits on two fronts: the CPU RAM and the GPU RAM. We have already addressed the issue of CPU RAM. The GPU RAM requirement is dependent on the size of the neural network model. The parameter count of neural networks that handle 3D data, such as 3D micro-CT images, requires much more GPU RAM than those that operate on 2D data. Therefore, digital rock technologies utilizing 3D data and models require more GPU RAM. This limitation can be tackled by designing efficient models or using GPUs with higher RAM.

6.2.2 Paper A and B: Super Resolution

The most critical limitations identified in the super-resolution tasks are the lack of real low-resolution and high-resolution ground truth image pairs. The low-resolution images used in training this model and all related works are created artificially; therefore, they do not represent the actual low-resolution images. This leads to unpredictable performance in the real-world domain. This lack of data needs to be addressed in the future by collecting registered low-resolution and high-resolution ground truth data using suitable micro-CT images.

Another challenge is that sometimes, the high-resolution ground truth does not exist. Therefore, the super-resolution model's performance depends on experts' subjective opinions. This opinion can vary considerably between experts.

6.2.3 Paper C: Image Registration

Transitioning into the field of image registration, the third paper addressed a persistent issue in industry-level applications: the registration of wet and dry images. The proposed solution delivered significant time savings through GPU-accelerated processes, reducing the standard image registration time from at least an hour to just a minute.

The main limitation in the image registration work is that sometimes, the pair of dry and wet images do not contain significant distinguishing features that can be used to correlate and register images. This was mitigated somewhat in this work by using appropriate normalization techniques.

6.2.4 Paper D: Rock Typing

The fourth paper tackled the challenge of rock typing using a UNet-based segmentation network, capitalizing on the visual characteristics of rock types to enable more accurate and efficient segmentation. The proposed method offered a supervised learning approach to the problem by leveraging labelled in-house data.

The fundamental limitation of the current method is the need for more training data. A very vast amount of labelled data is needed to make sure that the current pipeline generalizes well to unseen data. However, this is a limitation of supervised deep learning methods.

6.2.5 Paper E: 2D to 3D Generation

Finally, the fifth paper ventured into 3D microstructure generation, using only 2D images for both testing and training. The proposed solution, a unique diffusion GAN model, opens up new possibilities for creating 3D synthetic image representations from limited 2D data. The general limitation of generative models is that they are not perfect at reflecting the statistical properties of the real data.

The fundamental limitation specific to this work is that the diffusion-GAN model can be slow in generating the desired sample. The multistep inference leads to slow speed compared to the shot inference of GANs. The incorporation of GANs improves the speed somewhat; however, it is still slower than the one-shot generation. The slow generation performance might be undesirable in some scenarios.

6.3 Future Works

Future works can address various limitations of the current work, as follows:

6.3.1 Training with SEM Data

The methods in papers A and B can be trained on large amounts of SEM data and tested with micro-CT data. The input to the model can be the micro-CT images, and the expected output can be an image with intricate details enhanced to the resolution of an SEM image. However, this would require future work to address the challenge of inter-sensor differences and domain differences between SEM and micro-CT images.

6.3.2 Rock Typing for Diverse Rock Topologies

The scope of current rock typing automation was limited to rock samples that contain laminations. However, future work can extend this pipeline to samples that contain rock types and layers that exist in other patterns. This would require more data collection of various rock-type topologies.

6.3.3 Image Registration of Dry-Wet Images with Few Common Features

The proposed pipeline struggles in the case of a lack of common features between the wet and dry images. This can often happen when the wet rock sample undergoes complete transformation due to conversion from a dry to a wet sample. As long as there are a few common features, image registration can be performed successfully. Future methods can develop new methods to perform registration under a few standard features.

6.3.4 Generating 3D SEM Images from 2D SEM Images

The current method in Paper E produces 3D micro-CT images from 2D micro-CT images. However, a reasonable extension would be to generate 3D SEM images given 2D SEM images. However, SEM images have different physics in image formation compared to micro-CT images. Therefore, future work can study the feasibility of 3D SEM image generation.

Bibliography

[1] P.M. Adler, C.G. Jacquin and J.A. Quiblier. 'Flow in simulated porous media'. In: *International Journal of Multiphase Flow* 16.4 (1990), pp. 691–712. ISSN: 0301-9322. DOI: https://doi.org/10.1016/0301-9322(90)90025-E. URL: https://www.sciencedirect.com/science/article/pii/030193229090025E.

[2] Nader H Alhwety et al. 'Rock-typing of thin-bedded reservoir rock by NMR in the presence of diffusion coupling'. In: *SPWLA 57th Annual Logging Symposium*. OnePetro. 2016.

[3] Rabia Ali et al. 'Accurate detection of weld seams for laser welding in real-world manufacturing'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 13. 2023, pp. 15468–15475.

[4] Raymond J Althof, Marco GJ Wind and James T Dobbins. 'A rapid and automatic image registration algorithm with subpixel accuracy'. In: *IEEE transactions on medical imaging* 16.3 (1997), pp. 308–316.

[5] Mark A. Andersen, Brent Duncan and Ryan McLin. 'Core truth in formation evaluation'. In: *Oilfield Review* 25.2 (2013). Cited by: 28, pp. 16–25.

[6] Heiko Andrä et al. 'Digital rock physics benchmarks—Part I: Imaging and segmentation'. In: *Computers & Geosciences* 50 (2013). Benchmark problems, datasets and methodologies for the computational geosciences, pp. 25–32. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2012.09.005. URL: https://www.sciencedirect.com/science/article/pii/S0098300412003147.

[7] Heiko Andrä et al. 'Digital rock physics benchmarks—Part I: Imaging and segmentation'. In: *Computers & Geosciences* 50 (2013). Benchmark problems, datasets and methodologies for the computational geosciences, pp. 25–32. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2012.09.005. URL: https://www.sciencedirect.com/science/article/pii/S0098300412003147.

[8] Martin Arjovsky, Soumith Chintala and Léon Bottou. 'Wasserstein generative adversarial networks'. In: *International conference on machine learning*. PMLR. 2017, pp. 214–223.

[9] CH Arns et al. 'Digital core laboratory: Petrophysical analysis from 3D imaging of reservoir core fragments'. In: *Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description* 46.04 (2005).

[10] Christoph H Arns et al. 'Computation of linear elastic properties from microtomographic images: Methodology and agreement between theory and experiment'. In: *Geophysics* 67.5 (2002), pp. 1396–1405.

[11] Brian B Avants et al. 'A reproducible evaluation of ANTs similarity metric performance in brain image registration'. In: *Neuroimage* 54.3 (2011), pp. 2033–2044.

[12] Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla. 'SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation'. In: *arXiv preprint arXiv:1511.00561* (2015).

[13] Evgeny E Baraboshkin et al. 'Deep convolutions for in-depth automated rock typing'. In: *Computers & Geosciences* 135 (2020), p. 104330.

[14] Carl Fredrik Berg, Olivier Lopez and Håvard Berland. 'Industrial applications of digital rock technology'. In: *Journal of Petroleum Science and Engineering* 157 (2017), pp. 131–147.

[15] Steffen Berg et al. 'Real-time 3D imaging of Haines jumps in porous media flow'. In: *Proceedings of the National Academy of Sciences* 110.10 (2013), pp. 3755–3759. DOI: 10.1073/pnas.1221373110. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1221373110.

[16] Marco Bevilacqua et al. 'Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding'. In: *Proceedings of the British Machine Vision Conference*. DOI: http://dx.doi.org/10.5244/C.26.135. BMVA Press, 2012, pp. 135.1–135.10. ISBN: 1-901725-46-4.

[17] Stephen C. Blair, Patricia A. Berge and James G. Berryman. 'Using two-point correlation functions to characterize microgeometry and estimate permeabilities of sandstones and porous glass'. In: *Journal of Geophysical Research: Solid Earth* 101.B9 (1996), pp. 20359–20375. DOI: https://doi.org/10.1029/96JB00879. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/96JB00879. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/96JB00879.

[18] Yochai Blau et al. '2018 PIRM Challenge on Perceptual Image Super-resolution'. In: *CoRR* abs/1809.07517 (2018). arXiv: 1809.07517. URL: http://arxiv.org/abs/1809.07517.

[19] Martin J Blunt et al. 'Pore-scale imaging and modelling'. In: *Advances in Water resources* 51 (2013), pp. 197–216.

[20] A. Bogner et al. 'A history of scanning electron microscopy developments: Towards "wet-STEM" imaging'. In: *Micron* 38.4 (2007). Microscopy of Nanostructures, pp. 390–401. ISSN: 0968-4328. DOI: https://doi.org/10.1016/j.micron.2006.06.008. URL: http://www.sciencedirect.com/science/article/pii/S0968432806001016.

[21] Tom Bultreys et al. 'Fast laboratory-based micro-computed tomography for pore-scale research: Illustrative experiments and perspectives on the future'. In: *Advances in Water Resources* 95 (2016). Pore scale modeling and experiments, pp. 341–351. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2015.05.012. URL: https://www.sciencedirect.com/science/article/pii/S0309170815001062.

[22] Ruojin Cai et al. 'Learning gradient fields for shape generation'. In: *European Conference on Computer Vision*. Springer. 2020, pp. 364–381.

[23] Alfredo Canziani, Adam Paszke and Eugenio Culurciello. 'An Analysis of Deep Neural Network Models for Practical Applications'. In: *CoRR* abs/1605.07678 (2016). arXiv: 1605.07678. URL: http://arxiv.org/abs/1605.07678.

[24] Evelyn Chee and Zhenzhou Wu. 'Airnet: Self-supervised affine registration for 3d medical images using neural networks'. In: *arXiv preprint arXiv:1810.02583* (2018).

[25] Honggang Chen et al. 'Super-resolution of real-world rock microcomputed tomography images using cycle-consistent generative adversarial networks'. In: *Physical Review E* 101 (2 2020). ISSN: 24700053. DOI: 10.1103/PhysRevE.101.023305.

[26] Mark Chen et al. 'Generative pretraining from pixels'. In: *International Conference on Machine Learning*. PMLR. 2020, pp. 1691–1703.

[27] Nanxin Chen et al. 'WaveGrad: Estimating gradients for waveform generation'. In: *arXiv preprint arXiv:2009.00713* (2020).

[28] Yinbo Chen, Sifei Liu and Xiaolong Wang. 'Learning Continuous Image Representation with Local Implicit Image Function'. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR46437.2021.00852. 2021, pp. 8624–8634.

[29] Yuhua Chen et al. *Efficient and Accurate MRI Super-Resolution using a Generative Adversarial Network and 3D Multi-Level Densely Connected Network*. 2018. arXiv: 1803.01417 [cs.CV].

[30] Veerle Cnudde and Matthieu Nicolaas Boone. 'High-resolution X-ray computed tomography in geosciences: A review of the current technology and applications'. In: *Earth-Science Reviews* 123 (2013), pp. 1–17.

[31] Guillaume Coiffier, Philippe Renard and Sylvain Lefebvre. '3D Geological Image Synthesis From 2D Examples Using Generative Adversarial Networks'. In: *Frontiers in Water* 2 (2020), p. 30. ISSN: 2624-9375. DOI: 10.3389/frwa.2020.560598. URL: https://www.frontiersin.org/article/10.3389/frwa.2020.560598.

[32] *The Use of a Medical Computer Tomography (CT) System To Observe Multiphase Flow in Porous Media*. Vol. All Days. SPE Annual Technical Conference and Exhibition. SPE-13098-MS. Sept. 1984. DOI: 10.2118/13098-MS. eprint: https://onepetro.org/SPEATCE/proceedings-pdf/84SPE/All-84SPE/SPE-13098-MS/2038129/spe-13098-ms.pdf. URL: https://doi.org/10.2118/13098-MS.

[33] Yingzhi Cui et al. 'A numerical study of field strength and clay morphology impact on NMR transverse relaxation in sandstones'. In: *Journal of Petroleum Science and Engineering* 202 (2021), p. 108521. ISSN: 0920-4105. DOI: https://doi.org/10.1016/j.petrol.2021.108521. URL: https://www.sciencedirect.com/science/article/pii/S0920410521001807.

[34] E De Castro and CJIT Morandi. 'Registration of translated and rotated images using finite Fourier transforms'. In: *IEEE Transactions on pattern analysis and machine intelligence* 5 (1987), pp. 700–703.

[35] Lijun Ding and Ardeshir Goshtasby. 'On the Canny edge detector'. In: *Pattern recognition* 34.3 (2001), pp. 721–725.

[36] Chao Dong et al. 'Image Super-Resolution Using Deep Convolutional Networks'. In: *CoRR* abs/1501.00092 (2015). arXiv: 1501.00092. URL: http://arxiv.org/abs/1501.00092.

[37] Patrick Esser, Robin Rombach and Bjorn Ommer. 'Taming transformers for high-resolution image synthesis'. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pp. 12873–12883.

[38] Brian P. Flannery et al. 'Three-Dimensional X-Ray Microtomography'. In: *Science* 237.4821 (1987), pp. 1439–1444. DOI: 10.1126/science.237.4821.1439. eprint: https://www.science.org/doi/pdf/10.1126/science.237.4821.1439. URL: https://www.science.org/doi/abs/10.1126/science.237.4821.1439.

[39] Pierre Forbes. 'The status of core analysis'. In: *Journal of Petroleum Science and Engineering* 19.1 (1998), pp. 1–6. ISSN: 0920-4105. DOI: https://doi.org/10.1016/S0920-4105(97)00030-2.

[40] Yabo Fu et al. 'Deep learning in medical image registration: a review'. In: *Physics in Medicine & Biology* 65.20 (2020), 20TR01.

[41] *Overview of Advancement in Core Analysis and Its Importance in Reservoir Characterisation for Maximising Recovery*. Vol. Day 1 Tue, August 11, 2015. SPE Asia Pacific Enhanced Oil Recovery Conference. Aug. 2015. DOI: 10.2118/174583-MS. eprint: https://onepetro.org/SPEEORC/proceedings-pdf/15EORC/1-15EORC/D011S003R009/2344016/spe-174583-ms.pdf. URL: https://doi.org/10.2118/174583-MS.

[42] *Use of CT Scanning in the Investigation of Damage to Unconsolidated Cores*. Vol. All Days. SPE International Conference and Exhibition on Formation Damage Control. SPE-19408-MS. Feb. 1990. DOI: 10.2118/19408-MS. eprint: https://onepetro.org/SPEFD/proceedings-pdf/90FD/All-90FD/SPE-19408-MS/2007548/spe-19408-ms.pdf. URL: https://doi.org/10.2118/19408-MS.

[43] Ian Goodfellow et al. 'Generative Adversarial Nets'. In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

[44] Grant Haskins, Uwe Kruger and Pingkun Yan. 'Deep learning in medical image registration: a survey'. In: *Machine Vision and Applications* 31.1 (2020), pp. 1–18.

[45] Grant Haskins et al. ‘Learning deep similarity metric for 3D MR–TRUS image registration’. In: *International journal of computer assisted radiology and surgery* 14.3 (2019), pp. 417–425.

[46] Kaiming He et al. ‘Deep Residual Learning for Image Recognition’. In: *CoRR* abs/1512.03385 (2015). arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385.

[47] Jonathan Ho, Ajay Jain and Pieter Abbeel. ‘Denoising diffusion probabilistic models’. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 6840–6851.

[48] *Reservoir Rock Descriptions Using Computed Tomography (CT)*. Vol. All Days. SPE Annual Technical Conference and Exhibition. SPE-14272-MS. Sept. 1985. DOI: 10.2118/14272-MS. eprint: https://onepetro.org/SPEATCE/proceedings-pdf/85SPE/All-85SPE/SPE-14272-MS/2075528/spe-14272-ms.pdf. URL: https://doi.org/10.2118/14272-MS.

[49] Yipeng Hu et al. ‘Label-driven weakly-supervised learning for multimodal deformable image registration’. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE. 2018, pp. 1070–1074.

[50] Gao Huang, Zhuang Liu and Kilian Q. Weinberger. ‘Densely Connected Convolutional Networks’. In: *CoRR* abs/1608.06993 (2016). arXiv: 1608.06993. URL: http://arxiv.org/abs/1608.06993.

[51] Jia-Bin Huang, Abhishek Singh and Narendra Ahuja. ‘Single image super-resolution from transformed self-exemplars’. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR.2015.7299156. 2015, pp. 5197–5206.

[52] Aapo Hyvärinen and Peter Dayan. ‘Estimation of non-normalized statistical models by score matching.’ In: *Journal of Machine Learning Research* 6.4 (2005).

[53] *Rock-typing using the complete set of additive morphological descriptors*. Vol. All Days. SPE Reservoir Characterisation and Simulation Conference and Exhibition. SPE-165989-MS. Sept. 2013. DOI: 10.2118/165989-MS. eprint: https://onepetro.org/SPERCSC/proceedings-pdf/13RCSC/All-13RCSC/SPE-165989-MS/1530124/spe-165989-ms.pdf. URL: https://doi.org/10.2118/165989-MS.

[54] Mark Jenkinson and Stephen Smith. ‘A global optimisation method for robust affine registration of brain images’. In: *Medical image analysis* 5.2 (2001), pp. 143–156.

[55] Han Jiang and Christoph H Arns. 'Fast Fourier transform and support-shift techniques for pore-scale microstructure classification using additive morphological measures'. In: *Physical Review E* 101.3 (2020), p. 033302.

[56] Justin Johnson, Alexandre Alahi and Li Fei-Fei. 'Perceptual losses for real-time style transfer and super-resolution'. In: *European conference on computer vision*. Springer. 2016, pp. 694–711.

[57] Alexia Jolicoeur-Martineau. 'The relativistic discriminator: a key element missing from standard GAN'. In: *International Conference on Learning Representations*. 2019. URL: https://openreview.net/forum?id=S1erHoR5t7.

[58] Shaina Kelly et al. 'Assessing the utility of FIB-SEM images for shale digital rock physics'. In: *Advances in Water Resources* 95 (2016). Pore scale modeling and experiments, pp. 302–316. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2015.06.010.

[59] Steve Kench and Samuel J Cooper. 'Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion'. In: *Nature Machine Intelligence* 3.4 (2021), pp. 299–305.

[60] András P Keszei, Benjamin Berkels and Thomas M Deserno. 'Survey of non-rigid registration tools in medicine'. In: *Journal of digital imaging* 30.1 (2017), pp. 102–116.

[61] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee. 'Deeply-Recursive Convolutional Network for Image Super-Resolution'. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR.2016.181. 2016, pp. 1637–1645.

[62] Diederik P Kingma and Max Welling. 'Auto-encoding variational bayes'. In: *arXiv preprint arXiv:1312.6114* (2013).

[63] Stefan Klein et al. 'Elastix: a toolbox for intensity-based medical image registration'. In: *IEEE transactions on medical imaging* 29.1 (2009), pp. 196–205.

[64] MA Knackstedt et al. 'Digital Core Laboratory: Properties of reservoir core derived from 3D images'. In: *SPE Asia Pacific conference on integrated modelling for asset management*. OnePetro. 2004.

[65] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. 'ImageNet Classification with Deep Convolutional Neural Networks'. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. NIPS'12. Lake Tahoe, Nevada: Curran Associates Inc., 2012, pp. 1097–1105. URL: http://dl.acm.org/citation.cfm?id=2999134.2999257.

[66] Shane Latham, Trond Varslot, Adrian Sheppard et al. 'Image registration: enhancing and calibrating X-ray micro-CT imaging'. In: *Proc. of the Soc. Core Analysts, Abu Dhabi, UAE* (2008), pp. 1–12.

[67] Yann LeCun et al. 'Handwritten Digit Recognition with a Back-Propagation Network'. In: *Advances in Neural Information Processing Systems 2*. Ed. by D. S. Touretzky. Morgan-Kaufmann, 1990, pp. 396–404. URL: http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network.pdf.

[68] Christian Ledig et al. 'Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network'. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR.2017.19. 2017, pp. 105–114.

[69] *FIB/SEM and Automated Mineralogy for Core and Cuttings Analysis*. Vol. All Days. SPE Russian Petroleum Technology Conference. SPE-136327-MS. Oct. 2010. DOI: 10.2118/136327-MS. eprint: https://onepetro.org/SPERPTC/proceedings-pdf/10ROGC/All-10ROGC/SPE-136327-MS/1748429/spe-136327-ms.pdf. URL: https://doi.org/10.2118/136327-MS.

[70] Leon Leu et al. 'Fast X-ray micro-tomography of multiphase flow in berea sandstone: A sensitivity study on image processing'. In: *Transport in Porous Media* 105.2 (2014), pp. 451–469.

[71] Rui Liao et al. 'An artificial agent for robust image registration'. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1. 2017.

[72] J. Long, E. Shelhamer and T. Darrell. 'Fully convolutional networks for semantic segmentation'. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

[73] G Lowe. 'Sift-the scale invariant feature transform'. In: *Int. J* 2.91-110 (2004), p. 2.

[74] Yingjing Lu. *The Level Weighted Structural Similarity Loss: A Step Away from the MSE*. 2019. arXiv: 1904.13362 [cs.CV].

[75] Cheng Ma et al. 'Structure-Preserving Super Resolution With Gradient Guidance'. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR42600.2020.00779. 2020, pp. 7766–7775.

[76] Frederik Maes et al. 'Multimodality image registration by maximization of mutual information'. In: *IEEE transactions on Medical Imaging* 16.2 (1997), pp. 187–198.

[77] D. Martin et al. 'A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics'. In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Vol. 2. DOI: 10.1109/ICCV.2001.937655. 2001, 416–423 vol.2.

[78] Colin McPhee, Jules Reed and Izaskun Zubizarreta. *Core analysis: A best practice guide*. Vol. 60. Elsevier, 2015. ISBN: 978-0-444-63533-4.

[79] Tejaswini Medi et al. 'FullFormer: Generating Shapes Inside Shapes'. In: *arXiv preprint arXiv:2303.11235* (2023).

[80] Shun Miao, Z Jane Wang and Rui Liao. 'A CNN regression approach for real-time 2D/3D registration'. In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1352–1363.

[81] Ben Mildenhall et al. 'Nerf: Representing scenes as neural radiance fields for view synthesis'. In: *European conference on computer vision*. Springer. 2020, pp. 405–421.

[82] Lukas Mosser, Olivier Dubrule and Martin J Blunt. 'Reconstruction of three-dimensional porous media using generative adversarial neural networks'. In: *Phys. Rev. E* 96.4 (2017).

[83] Lukas Mosser, Olivier Dubrule and Martin J Blunt. 'Stochastic reconstruction of an oolitic limestone by generative adversarial networks'. In: *Transport in Porous Media* 125.1 (2018), pp. 81–103.

[84] Kamyar Nazeri, Harrish Thasarathan and Mehran Ebrahimi. 'Edge-Informed Single Image Super-Resolution'. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. DOI: 10.1109/ICCVW.2019.00409. 2019, pp. 3275–3284.

[85] William S Noble. 'What is a support vector machine?' In: *Nature biotechnology* 24.12 (2006), pp. 1565–1567.

[86] Hyeonwoo Noh, Seunghoon Hong and Bohyung Han. 'Learning Deconvolution Network for Semantic Segmentation'. In: *CoRR* abs/1505.04366 (2015). arXiv: 1505.04366. URL: http://arxiv.org/abs/1505.04366.

[87] Aaron van den Oord et al. 'Conditional image generation with pixelcnn decoders'. In: *arXiv preprint arXiv:1606.05328* (2016).

[88] GS Padhy et al. 'Pore size distribution in multiscale porous media as revealed by DDIF–NMR, mercury porosimetry and statistical image analysis'. In: *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 300.1-2 (2007), pp. 222–234.

[89] Jeong Joon Park et al. 'DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation'. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR.2019.00025. 2019, pp. 165–174.

[90] Niki Parmar et al. 'Image transformer'. In: *International Conference on Machine Learning* (2018), pp. 4055–4064.

[91] Adam Paszke et al. 'Automatic differentiation in PyTorch'. In: (2017).

[92] Adam Paszke et al. 'PyTorch: An Imperative Style, High-Performance Deep Learning Library'. In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[93] Cheng Peng et al. 'SAINT: Spatially Aware Interpolation NeTwork for Medical Slice Synthesis'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2020.

[94] Johan Phan et al. 'Size-invariant 3D generation from a single 2D rock image'. In: *Journal of Petroleum Science and Engineering* 215 (2022), p. 110648. ISSN: 0920-4105. DOI: https://doi.org/10.1016/j.petrol.2022.110648. URL: https://www.sciencedirect.com/science/article/pii/S0920410522005174.

[95] William K Pratt. 'Correlation techniques of image registration'. In: *IEEE transactions on Aerospace and Electronic Systems* 3 (1974), pp. 353–358.

[96] M Prodanović, WB Lindquist and RS Seright. '3D image-based characterization of fluid displacement in a Berea core'. In: *Advances in Water Resources* 30.2 (2007), pp. 214–226.

[97] Alec Radford, Luke Metz and Soumith Chintala. 'Unsupervised representation learning with deep convolutional generative adversarial networks'. In: *arXiv preprint arXiv:1511.06434* (2015).

[98] Ali Razavi, Aaron van den Oord and Oriol Vinyals. 'Generating diverse high-fidelity images with vq-vae-2'. In: *Advances in neural information processing systems* (2019), pp. 14866–14876.

[99] Danilo Jimenez Rezende, Shakir Mohamed and Daan Wierstra. 'Stochastic Backpropagation and Approximate Inference in Deep Generative Models'. In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Eric P. Xing and Tony Jebara. Vol. 32. Proceedings of Machine Learning Research 2. Bejing, China: PMLR, 22–24 Jun 2014, pp. 1278–1286. URL: https://proceedings.mlr.press/v32/rezende14.html.

[100] Alexis Roche, Gregoire Malandain and Nicholas Ayache. 'Unifying maximum likelihood approaches in medical image registration'. In: *International Journal of Imaging Systems and Technology* 11.1 (2000), pp. 71–80.

[101] Olaf Ronneberger, Philipp Fischer and Thomas Brox. 'U-Net: Convolutional Networks for Biomedical Image Segmentation'. In: *arXiv preprint arXiv:1505.04597* (2015).

[102] Leonardo Ruspini et al. 'Multiscale Digital Rock Analysis for Complex Rocks'. In: *Transport in Porous Media* 139 (Sept. 2021), pp. 1–25. DOI: 10.1007/s11242-021-01667-2.

[103] Chitwan Saharia et al. 'Image super-resolution via iterative refinement'. In: *arXiv preprint arXiv:2104.07636* (2021).

[104] Robin Sandkühler et al. 'AirLab: autograd image registration laboratory'. In: *arXiv preprint arXiv:1806.09907* (2018).

[105] Muhammad Sarmad, Mishal Fatima and Jawad Tayyub. 'Reducing Energy Consumption of Pressure Sensor Calibration Using Polynomial HyperNetworks with Fourier Features'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 11. 2022, pp. 12145–12153.

[106] Ramprasaath R Selvaraju et al. 'Grad-cam: Visual explanations from deep networks via gradient-based localization'. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.

[107] Vincent Sitzmann et al. 'Implicit Neural Representations with Periodic Activation Functions'. In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 7462–7473. URL: https://proceedings.neurips.cc/paper/2020/file/53c04118df112c13a8c34b38343b9c10-Paper.pdf.

[108] Vincent Sitzmann et al. 'Implicit neural representations with periodic activation functions'. In: *Advances in neural information processing systems* 33 (2020), pp. 7462–7473.

[109] Yang Song and Stefano Ermon. 'Generative modeling by estimating gradients of the data distribution'. In: *Advances in Neural Information Processing Systems* 32 (2019).

[110] Yang Song et al. 'Score-based generative modeling through stochastic differential equations'. In: *arXiv preprint arXiv:2011.13456* (2020).

[111] Catherine Spurin et al. *Decane and brine injected into Estaillades carbonate - steady-state experiments*. http://www.digitalrocksportal.org/projects/344. 2021. DOI: 10.17612/cd7a-y955.

[112] Jan Srodon et al. 'Quantitative X-ray diffraction analysis of clay-bearing rocks from random preparations'. In: *Clays and Clay Minerals* 49.6 (2001), pp. 514–528.

[113] S. Strebelle. 'Conditional simulation of complex geological structures using multiple-point statistics'. In: *Mathematical geology* 34 (2002), pp. 1–21.

[114] P. Tahmasebi, A. Hezarkhani and M. Sahimi. 'Multiple-point geostatistical modeling based on the cross-correlation functions'. In: *Computat. Geosci.* 16 (2012), pp. 779–797.

[115] P. Tahmasebi, M. Sahimi and J. Caers. 'MS-CCSIM: accelerating pattern-based geostatistical simulation of categorical variables using a multi-scale search in Fourier space'. In: *Comput. Geosci.* 67 (2014), pp. 75–88.

[116] Jawad Tayyub, Muhammad Sarmad and Nicolas Schönborn. 'Explaining deep neural networks for point clouds using gradient-based visualisations'. In: *Proceedings of the Asian Conference on Computer Vision*. 2022, pp. 2123–2138.

[117] Hoang Thanh-Tung and Truyen Tran. 'Catastrophic forgetting and mode collapse in GANs'. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2020, pp. 1–10.

[118] Paul Viola and William M Wells III. 'Alignment by maximization of mutual information'. In: *International journal of computer vision* 24.2 (1997), pp. 137–154.

[119] D. Volkhonskiy et al. 'Reconstruction of 3D Porous Media From 2D Slices'. In: *arcXiv:1901.1023v1* (2019).

[120] Shan Wang et al. *Dataset for unsteady-state capillary drainage experiment on Estaillades carbonate*. http://www.digitalrocksportal.org/projects/363. 2021. DOI: 10.17612/6rtt-5w16.

[121] Xintao Wang et al. 'ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks'. In: *CoRR* abs/1809.00219 (2018). arXiv: 1809.00219. URL: http://arxiv.org/abs/1809.00219.

[122] Xintao Wang et al. 'Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform'. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. DOI: 10.1109/CVPR.2018.00070. 2018, pp. 606–615.

[123] Ying Da Wang, Ryan T. Armstrong and Peyman Mostaghimi. 'Enhancing Resolution of Digital Rock Images with Super Resolution Convolutional Neural Networks'. In: *Journal of Petroleum Science and Engineering* 182 (2019). ISSN: 09204105. DOI: 10.1016/j.petrol.2019.106261.

[124] Yuzhu Wang et al. 'Image-based rock typing using local homogeneity filter and Chan-Vese model'. In: *Computers & Geosciences* 150 (2021), p. 104712. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2021.104712. URL: https://www.sciencedirect.com/science/article/pii/S0098300421000261.

[125] S.L. Wellington and H.J. Vinegar. 'X-Ray Computerized Tomography'. In: *Journal of Petroleum Technology* 39.08 (Aug. 1987), pp. 885–898. ISSN: 0149-2136. DOI: 10.2118/16983-PA. eprint: https://onepetro.org/JPT/article-pdf/39/08/885/2225391/spe-16983-pa.pdf. URL: https://doi.org/10.2118/16983-PA.

[126] Dorthe Wildenschild and Adrian P. Sheppard. 'X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems'. In: *Advances in Water Resources* 51 (2013). 35th Year Anniversary Issue, pp. 217–246. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2012.07.018. URL: https://www.sciencedirect.com/science/article/pii/S0309170812002060.

[127] Terry S Yoo et al. 'Engineering and algorithm design for an image processing API: a technical report on ITK-the insight toolkit'. In: *Medicine Meets Virtual Reality 02/10*. IOS press, 2002, pp. 586–592.

[128] Chenyu You et al. 'CT Super-Resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE)'. In: *IEEE Transactions on Medical Imaging* 39.1 (Jan. 2020), pp. 188–203. ISSN: 1558-254X. DOI: 10.1109/tmi.2019.2922960. URL: http://dx.doi.org/10.1109/TMI.2019.2922960.

[129] Roman Zeyde, Michael Elad and Matan Protter. 'On Single Image Scale-Up Using Sparse-Representations'. In: *Curves and Surfaces*. Ed. by Jean-Daniel Boissonnat et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 711–730. ISBN: 978-3-642-27413-8.

[130] Richard Zhang et al. 'The Unreasonable Effectiveness of Deep Features as a Perceptual Metric'. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. DOI: 10.1109/CVPR.2018.00068. 2018, pp. 586–595.

[131] Yulun Zhang et al. 'Image Super-Resolution Using Very Deep Residual Channel Attention Networks'. In: *ECCV*. 2018.

[132] Yulun Zhang et al. 'Residual Dense Network for Image Super-Resolution'. In: *CoRR* abs/1802.08797 (2018). arXiv: 1802.08797. URL: http://arxiv.org/abs/1802.08797.

[133] Jiuyu Zhao, Fuyong Wang and Jianchao Cai. '3D tight sandstone digital rock reconstruction with deep learning'. In: *Journal of Petroleum Science and Engineering* 207 (2021), p. 109020. ISSN: 0920-4105. DOI: https://doi.org/10.1016/j.petrol.2021.109020.

List of Tables

1.1 Thesis Overview . . . 10

2.1 Comparison of different CT imaging techniques . . . 17

4.1 Result of the New Analysis Pipeline . . . 46

4.2 **Quantitative comparison with CNNs on benchmark datasets** This table shows our method with other perceptual quality-focused methods. The best results are in **bold**. All models have been trained and tested on 4x down-sampled images. . . . 48

4.3 Dry Wet Image Registration . . . 52

5.1 Discussions of the New Analysis Pipeline . . . 60

List of Figures

1.1 Research Questions (RQs) formulated in the scope of digital rock analysis. This diagram shows the link of the list of Papers from A to E to their respective RQ. 9

2.1 Examples of different grain packing settings in rock samples: (a) large grains, (b) random grains, (c) small grains, and (d) a mix of large and small grains. The arrangement and size of grains impact the porosity and permeability of the rock. 20

3.1 Adversarial loss based output (right) vs MSE loss based model output (left) 32

3.2 Illustration of an implicit function that takes coordinates as input and provides colour values as output. 33

3.3 Illustration of a neural network-based implicit function that takes coordinates as input and provides color values as output. 34

3.4 **Training Method:** The low-resolution image $I_{\downarrow s}$ is passed through CNN encoder to get feature vector z. A random patch is selected from the coordinate space of the desired high-resolution image to obtain high-resolution coordinates x_{hr}. z and x_{hr} are passed through the Local implicit function image (LIIF) generator to obtain the super-resolved output image I. This I is compared with I^{GT} using adversarial loss ('Adv loss'), perceptual loss ('VGG loss') and with I^{HR} using pixel loss L_1. 34

3.5 **SIT-SR 3D:** The architecture of our proposed method. The low-resolution image $I_{\downarrow r}$ is upsampled along x, y and z respectively with the pre-trained 2D SISR, as a result, the volumes $I_{\downarrow r_x}$, $I_{\downarrow r_y}$ and $I_{\downarrow r_z}$ are obtained. I^x, I^y, and I^z are the corresponding volumes obtained after the interpolation operation. I^x, I^y and I^z are passed through the Fusion module to get the output volume I. 36

3.6 We solve the problem of Dry-Wet rock Image registration. Our Registration Method (Blue box) can register Image pairs efficiently and accurately. In this figure, the Fixed 3D Image is the Dry Image, whereas the Moving 3D Image is the Wet Image. The registration method finds the transformation required to warp the Moving 3D image to register it to the Fixed 3D Image. 37

3.7 Extreme Rotation: A micro-CT image of a wet and dry sample are shown. Note that The wet sample can have extreme rotation of ±180° along z-axis and a 180° inversion around the x or y axes. . 38

3.8 Rock typing and Upscaling Workflow: There are multiple steps involved in upscaling. 39

3.9 Example of labelling of a rock image for rock types 40

3.10 Understanding the diffusion process 41

4.1 **Qualitative Comparison on Set 14[129]:**. This figure shows the high-resolution ground truth image (HR), the low-resolution image (LR), the super-resolved image using LIIF model [28] and our model's output. All input images are 6x down-sampled from ground-truth images and super-resolved to 6x. All models were trained for 1x-4x only. We observe the same smoothing effect for LIIF outputs where the high-level details such as water waves and texture in the fence has been blurred, while our model retains the high-level details and the image produced is much more realistic than LIIF. 48

4.2 This figure shows a visual comparison of different methods. HR, LR indicates high-resolution and low-resolution images. 3D RRDB-Net is a 3D convolution-based network supervised with L1 loss. 3D ESRGAN is trained with GAN and VGG loss using pre-trained weights of 3D RRDBNet. SIT-SR 3D is trained in a self-supervised setting using only the consistency loss. 50

4.3 The Qualitative results of our model are shown. From left to right, we have the input image, ground truth segmentation mask and predicted segmentation mask. 54

4.4 Grad Cam Explain ability Analysis: We query each class as shown by the tag 'Chosen Class' in the GT masks. It can be seen that the CAM mask image shows the area where the neural network is paying attention for obtaining the Pred Mask. 55

4.5 **Visual comparison with micro-CT images**: Cross-sections of 3D images generated by our method and SliceGAN [59], alongside their respective ground truth or training data. The ground truth images are 3D X-ray micro-CT scans obtained at varying resolutions. The Glassbeads case showcases our method's superior performance over SliceGAN. Our model can capture the spherical shape of the object, even though it only sees circles at the 2D input. In more challenging cases like the Savoniere - a carbonate of fossilized microorganism - our method proves its robustness by generating images that bear a higher resemblance to reality, despite the heterogeneous nature of the original image. 57

Acronyms

BSE Back-scattered electron. 18

CCA Conventional core analysis. 15, 16

CNN Convolutional Neural Network. 9, 21, 23–26, 32–36, 47, 48, 61, 85, 87

CT Computed Tomography. vi, 10, 16–18, 23, 24, 27, 33, 46, 47, 60, 61, 85

DRA Digital rock analysis. 16

DRP Digital rock physics. 16, 17, 19, 20

DRT Digital rock technology. 16

ESRGAN Enhanced Super-Resolution Generative Adversarial Network. 23, 48–50, 88

GAN Generative Adversarial Network. 7, 22, 23, 27, 28, 31, 32, 35, 41–43, 49, 50, 62, 65, 69, 88

GDP Gross Domestic Product. 3

INN Implicit Neural Networks. 22, 33, 61

LPIPS Learned Perceptual Image Patch Similarity. 23, 47, 48

MAE Mean Absolute Error. 32

micro-CT Micro computed Tomography. i, 5, 6, 11, 17, 19, 34, 36, 38, 47, 56, 57, 61, 62, 68, 70, 88, 89

MSE Mean Squared Error. 23, 32, 87

NMR Nuclear Magnetic Resonance. 26, 27

PSNR Peak Signal-to-Noise Ratio. 23, 47, 48

SCAL Special core analysis. 15

SEM Scanning Electron Microscope. 11, 18, 19, 47, 70

SISR Single Image Super-Resolution. 23, 36, 88

Part II

List of Papers

Appendix A

Paper A: Photo-Realistic Continuous Image Super-Resolution with Implicit Neural Networks and Generative Adversarial Networks

Photo-Realistic Continuous Image Super-Resolution with Implicit Neural Networks and Generative Adversarial Networks

Muhammad Sarmad*[1], Leonardo Ruspini[2], and Frank Lindseth[1]

[1]Norwegian University of Science and Technology, Trondheim, Norway
[2]Petricore, Norway

Abstract

The implicit neural networks (INNs) can represent images in the continuous domain. They consume raw (X, Y) coordinates and output a color value. Therefore they can represent and generate images at arbitrarily high resolutions in contrast to convolutional neural networks (CNNs) that output a constant-sized array of pixels. In this work, we show how to super-resolve a single image using an INN to produce sharp and photo-realistic images. We employ a random patch-based coordinate sampling method to obtain patches with context and structure; we use these patches to train the INN in an adversarial setting. We demonstrate that the trained network retains the desirable properties of INNs while the output is sharper compared to previous work. We also show qualitative and quantitative comparisons with INN and CNN baselines on benchmark datasets of DIV2K, Set5, Set14, Urban100, and B100. Our code will be made public at `https://github.com/iSarmad/CiSRGan`.

1 Introduction

Image enhancement and super-resolution find applications in various consumer products such as smartphone photography, TV and video, etc. The advent of deep learning and neural networks has enabled advancements in single-image super-resolution (SISR). Convolutional neural networks (CNNs) are the most popular method for SISR [11]. However, the output of CNNs is an array of pixels with a fixed size. Therefore, we need to train a new network for different scaling factors. This strategy can be very inconvenient and time-consuming.

Recently a class of neural networks called implicit neural networks (INNs) has gained attention [33, 25, 28]. These networks can represent an image by storing the color value of each pixel corresponding to a given pixel coordinate [26, 31]. This image representation leads to a continuous model where one can zoom in to a single image arbitrarily by changing the discretization level of the input coordinates.

Chen et al. [8] proposed an INN based method called local implicit image function (LIIF) for SISR. They used a single INN to perform SISR for any scale and achieved arbitrary zooming capability i.e. given a neural network that was trained for scales in the range of 1x to 4x (*we refer to this range as in-scale*), their model can perform super-resolution on 6x and 8x etc (*out-of-scale*). This ability to extrapolate makes LIIF very beneficial for super-resolution. Furthermore, LIIF is on par with CNNs in terms of distortion metrics such as the PSNR [22]. Despite these advantages, LIIF suffers from blurry outputs for *out-of-scale* super-resolution due to the use of pixel-wise loss function. In this work, we propose continuous image super-resolution generative adversarial network (CiSR-GAN) that trains INNs in an adversarial setting for super-resolution, thus improving the perceptual quality and photo-realism of output for out-of-scale SISR. To the best of our knowledge, training implicit network for the task of out-of-scale single image super-resolution in an adversarial setting has not been proposed before.

We compare our method with previous state of the art in INN and CNN based super-resolution methods.

*Corresponding Author: muhammad.sarmad@ntnu.no

`https://doi.org/10.7557/18.6285`

2 Related Works

Convolutional Neural Network based SISR Before convolutional neural networks (CNNs) [18, 19, 13], handcrafted algorithms were used to perform single image super-resolution (SISR); e.g., Yang et al. [39] used sparse coding to solve this task. Recently, SISR using CNN has become main stream [20, 27, 23, 37]. SISR can be divided into algorithms that either focus on lowering distortion or improving perceptual quality [6]. Our work focuses on improving the perceptual quality.

Implicit Neural Networks for SISR Implicit neural networks (INNs) have recently become popular as a way to represent continuous images and shapes [26, 38, 4, 9, 3, 10]. Occupancy Networks [25] and Deep SDF [28] used INNs for 3D shape representation. Then Sitzman et al. [31], and Tancik et al.[34] showed that the INNs could also be used to represent images with high fidelity. Later works learned GANs using INNs [7, 32, 30, 2]. Local implicit image function (LIIF) [8] recently showed that continuous representation could also be used to perform SISR. The resulting SISR model is agnostic to resolution, and a single model can be used to super-resolve images to any required resolution. LIIF [8] uses the L_1 loss to train the network, which renders the output blurry. However, we train our model in the adversarial setting to perform photo-realistic SISR and achieve a better result.

3 Method

Consider a low-resolution 2D Image $I_{\downarrow s}$ that consists of arrays of pixels. The high resolution 2D image corresponding to $I_{\downarrow s}$ is given as $I \rightarrow I(x, y) \in \mathbb{R}^{X \times Y}$. Where $I_{\downarrow s}(x, y) \in \mathbb{R}^{\frac{X \times Y}{s}}$, and s is the scaling factor. Each pixel in I has coordinates x and y. Let's assume that a continuous image can be represented by a function f_θ. Then the discrete image I can be represented as:

$$I = f_\theta(c, z), \tag{1}$$

z is the latent vector of the features of low-resolution image $I_{\downarrow s}$. Note that $c = x_{hr} - v$, x_{hr} are the pixel coordinates of image I and v are the coordinates of the feature vector z in the image domain. In this work, f_θ is the implicit neural (INN).

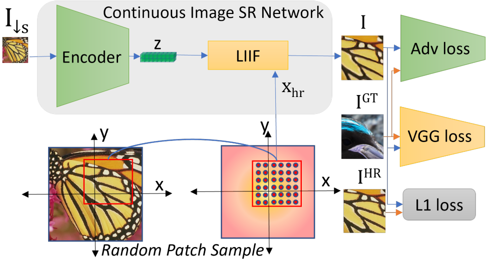


Figure 1: **Training Method:** The low-resolution image $I_{\downarrow s}$ is passed through CNN encoder to get feature vector z. A random patch is selected from the coordinate space of desired high resolution image to obtain high resolution coordinates x_{hr}. z and x_{hr} are passed through Local implicit function image (LIIF) generator to obtain the super-resolved output image I. This I is compared with I^{GT} using adversarial loss ('Adv loss'), perceptual loss ('VGG loss') and with I^{HR} using pixel loss L_1.

More specifically, for f_θ we employ the local implicit image function (LIIF) with default configurations. For details, we refer to the paper [8].

Training LIIF in an Adversarial Setting An overview of our approach is shown in Figure. 1. The input image is passed through a convolutional encoder to obtain a latent vector z. This latent vector z and the image I coordinates x_{hr} are used to obtain the color values of the pixels at input coordinates x_{hr} using LIIF block [8]. Note that the INN consists of a few multilayer perceptron (MLP) layers that are present inside the LIIF block. We need an output image patch to train the INN using adversarial and perceptual loss. The previous method [8] uses a random set of coordinates from the image. This sampling method works well when the objective is to minimize the pixel-wise loss, e.g., L_1. However, looking at only pixels means the contextual information is lost. Therefore, we propose a random patch-based sampling procedure instead of a random point-based sampling method to retain contextual information. We first train LIIF [8] with random patches instead of random points with only a pixel-wise loss. We notice that this random patch-based sampling method performs similar to a random coordinate-based sampling method in terms of performance.

We use the L_1 loss following previous work [8], which trains with only the L_1 objective leading to smooth images which blur the textural information for *out-of-scale* super-resolution.

The use of a patch-based sampling procedure permits the use of adversarial loss that is based on generative adversarial network (GAN) [12]. The GAN consists of a generator and a discriminator that compete against each other. The goal of the generator is to generate realistic images, whereas the goal of the discriminator is to get good at classifying generated images as fake. In this joint training, both get better, resulting in realistic image generation. However, instead of using a standard GAN formulation, we use a relativistic GAN formulation instead [16]. This formulation is different from the standard discriminator, which estimates the probability that an input image is real. Instead, the discriminator predicts the probability that a real image is relatively more realistic than a fake one. We define a discriminator network D_{θ_D}, which is optimized in an alternating manner along with generator network G_{θ_G} to solve the adversarial min-max problem. The relativistic GAN solves the following min-max problem:

$$\begin{aligned}\min_{\theta_G}\max_{\theta_D}\ & \mathbb{E}_X[\log D_{\theta_D}(I^{GT}, G_{\theta_G}(I_{\downarrow s}))]+ \\ & \mathbb{E}_X[\log(1 - D_{\theta_D}(G_{\theta_G}(I_{\downarrow s}), I^{GT}))]\end{aligned} \qquad (2)$$

Note that, $X = (I^{GT}, I_{\downarrow s}) \sim (p_{\text{train}}(I^{GT}), p_G(I_{\downarrow s}))$ and $D_{\theta_D}(I^{GT}, G_{\theta_G}(I_{\downarrow s})) = \sigma(\mathcal{C}(I^{GT}) - \mathbb{E}_{G_\theta(I_{\downarrow s})}[\mathcal{C}(G_{\theta_G}(I_{\downarrow s}))])$. Where $\mathbb{E}_{G_\theta(I_{\downarrow s})}[.]$ is mean over the generated data in the mini-batch. σ is the sigmoid activation function and $\mathcal{C}$ is the output of discriminator before the activation function. For details, we refer to [16].

We also use the perceptual loss that is the distance between the features of a pre-trained VGG network between the predicted image I and the ground-truth image I^{GT} [15]. The complete training objective for the generator is as follows:

$$\mathcal{L}_t = \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}_G + \lambda_3\mathcal{L}_{VGG} \qquad (3)$$

Where $\mathcal{L}_1, \mathcal{L}_G$ and $\mathcal{L}_{VGG}$ are the content, adversarial and perceptual losses respectively. The λ_1, λ_2 and λ_3 are weighting hyperparameters terms for each of the objectives respectively. We set them following guidelines from previous work [37].

4 Experiments

We employed Pytorch for the implementation of all our models [29]. We trained all the networks on an NVIDIA RTX Titan GPU. The code is built on the open-source implementations [8, 35].

Dataset and Metrics Like [8], we use the DIV2K dataset with standard split for training and validation [1] for fair comparison. Testing is performed on multiple test datasets including Set5, Set14, Urban100 and B100 [5, 40, 14, 24]. The results for the related works were generated for comparison using pre-trained models provided by Chen et al. [8], and SPSR [23]. We use peak signal-to-noise ration (PSNR) as a metric for comparison. PSNR (measured in dB) is a measure of quality between super-resolved image and ground truth. Even though it is a good measure of distortion, however, it is a poor indicator of perceptual quality [6]. Therefore we additionally report perceptual similarity metric (LPIPS) [41] for comparison with previous works. LPIPS measures the distance in VGG [15] feature space between the super-resolved and the ground-truth image. The lower the distance, the more perceptually similar the super-resolved image is to the ground truth.

Training Details Similar to LIIF [8], we use RDN [42] as the encoder, where a feature map z is generated with the same size as the input image. The INN f_θ is a 5-layer MLP with ReLU activation and hidden dimensions of 256. Encoder and INN act as the generator in our model. The discriminator is based on the architecture used by ESRGAN [37]. We use input patches of 64 x 64 during training. The generator's output is the same as the input patch size, i.e., 64 x 64; therefore, the discriminator is adjusted to cater to an image patch of this size. We use transfer learning and initialize the weights of our generator from a pre-trained RDN-LIIF [8]. We train all models for 75 epochs with batch size 16 on the DIV2K training set. We utilize the Adam [17] optimizer for both generator and discriminator with a learning rate of 1^{-4}. The weights for λ_1, λ_2 and λ_3 are set to 1^{-2}, 5^{-3} and 1 [37]. For a fair comparison with LIIF, we also train the models from the 1x-4x scale range.

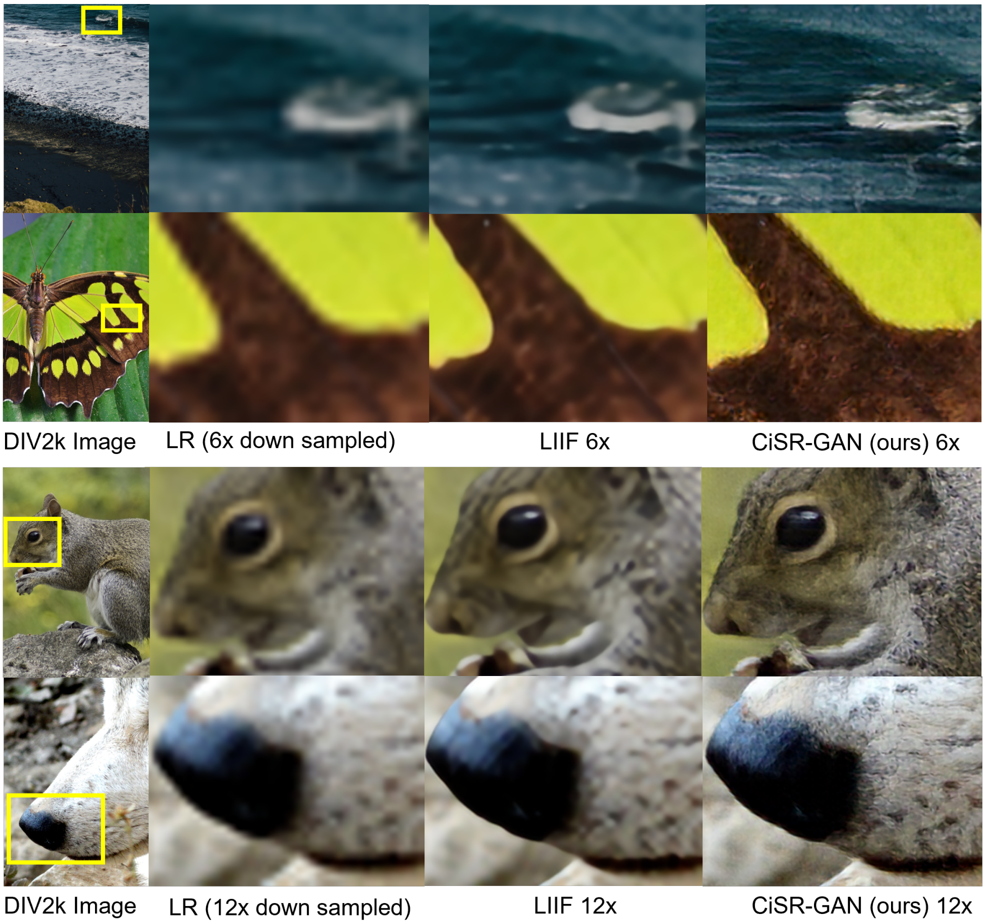


Figure 2: **Out-of-Scale Qualitative Comparison on DIV2K:**. This figure shows the reference image from DIV2k, the low-resolution input image (LR), super-resolved image using LIIF [8] and finally our model's output (CiSR-GAN). LR images are 6x and 12x down-sampled from ground-truth HR images and super-resolved to 6x and 12x in the top 2 and bottom 2 rows respectively demonstrating out-of-scale performance. All models were trained for 1x-4x only therefore we refer to 6x and 12x as *out-of-scale*. From the images we can see that LIIF has a smoothing effect where it blurs out the high-level detail in the images. Comparatively, our models clearly produces sharper results retaining textural details like waves of water, texture in butterfly wings and fine hair of animals.

Qualitative Analysis

Out-of-Scale: The qualitative results on DIV2K validation set [1] and Set14 [40] test set are shown in Figure. 2 and Figure. 3 respectively. The proposed CiSR-GAN produces realistic images containing textures due to the adversarial and perceptual nature of the objective as compared to the LIIF [8]. LIIF's output is always blurry for *out-of-scale* super-resolution smoothing out

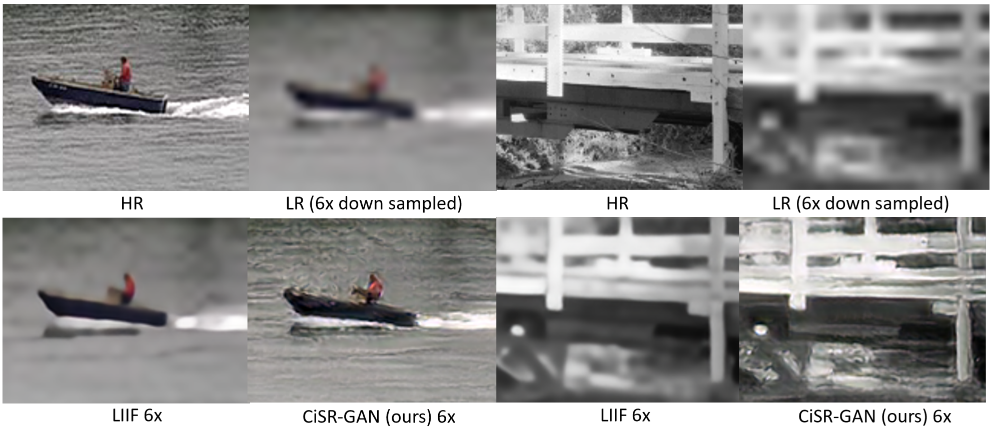


Figure 3: **Out-of-Scale Qualitative Comparison on Set 14:**. This figure shows the high resolution ground truth image (HR), the low-resolution image (LR), super-resolved image using LIIF model [8] and our model's output (CiSR-GAN's). All input images are 6x down-sampled from ground-truth images and super-resolved to 6x. All models were trained for 1x-4x only. We observe the same smoothing effect for LIIF outputs where the high level details such as water waves and texture in the fence has been blurred, while our model retains the high-level details and the image produced is much more realistic than LIIF.

the textural information. At the same time, we also maintain all the desired properties of an implicit network, e.g., a single model can perform super-resolution at higher scales even if the model is not trained for it. All the results presented in the qualitative comparison are for 6x or 12x upsampling to compare with LIIF, whereas we train our models on 1x-4x down-sampled images.

In-Scale: Please note that CNN decoder based models [37, 27, 21] are not a direct competitor of our method since they can not perform *out-of-scale* super-resolution. However, we test their performance for *in-scale* super-resolution i.e. for 4x scaling factor for the sake of comprehensiveness. We compare with the best performing recent CNN-based method Structure-Preserving Super Resolution (SPSR) [23], that recently showed great results in retrieving sharp lines and geometry. All images are 4x down-sampled from the ground truth HR images and super-resolved to 4x. The performance is shown in Figure. 4. SPSR model adds edge artifacts like lines or texture to the super-resolved image whereas CiSR-GAN produces more realistic results.

Quantitative Results

CiSR-GAN vs LIIF We compare our model (CiSR-GAN) with previous work on the DIV2k dataset, as shown in Table. 1. The perceptual similarity metric (LPIPS) is a distance metric; therefore, the lower the value, the better. Whereas the higher the peak signal-to-noise ratio (PSNR), the better. Blau et al. [6] have previously shown that there is a trade-off between distortion and perception, and this can also be observed for our model. CiSR-GAN formulation has lower PSNR values than local implicit image function LIIF [8] as it is trained on the adversarial and perceptual loss. However, it consistently performs better than LIIF in terms of LPIPS metric. Lower LPIPS means that we can expect aesthetically pleasing results from CiSR-GAN. CiSR-GAN can also be evaluated for *out-of-scale* models easily since it is based on an INN. It maintains the edge over LIIF in terms of perceptual metrics for all scales evaluated.

In-Scale: We further compare the performance with state-of-the-art methods, including SRGAN, ESRGAN, and SPSR [23, 37, 20]. We notice that CiSR-GAN outperforms all in LPIPS while main-

Method	Metric	In-Scale			Out-of-Scale			
		×2	×3	×4	×6	×12	×24	×30
RDN-LIIF [8]	PSNR	**34.99**	**31.26**	**29.27**	**26.99**	**23.89**	**21.31**	**20.59**
	LPIPS	0.0558	0.1344	0.1947	0.2760	0.4163	0.5506	0.5845
CiSR-GAN (ours)	PSNR	32.01	27.95	26.30	24.27	21.67	19.52	18.92
	LPIPS	**0.0254**	**0.0641**	**0.1016**	**0.1642**	**0.3409**	**0.4839**	**0.5319**

Table 1: **Distortion vs Perception.** Scaling factor for training is in range ×1–×4. Best values are bold.

Dataset	Metric	SFTGAN [36]	SRGAN [20]	ESRGAN [37]	SPSR [23]	CiSR-GAN (ours)
Set5	LPIPS	0.0890	0.0882	0.0748	0.0644	**0.0604**
	PSNR	29.932	29.168	**30.454**	30.400	30.05
Set14	LPIPS	0.4393	0.1663	0.1329	0.1318	**0.1160**
	PSNR	26.100	26.171	26.276	**26.640**	26.62
B100	LPIPS	0.5249	0.1980	0.1614	0.1611	**0.1436**
	PSNR	25.961	25.459	25.317	25.505	**25.72**
Urban100	LPIPS	0.4726	0.1551	0.1229	0.1184	**0.1179**
	PSNR	23.145	24.397	24.360	**24.799**	24.36

Table 2: **In-Scale Quantitative comparison with CNNs on benchmark datasets** This table shows CiSR-GAN with other perceptual quality focused methods. Best results are in **bold**. All models have been trained and tested on 4x down-sampled images.

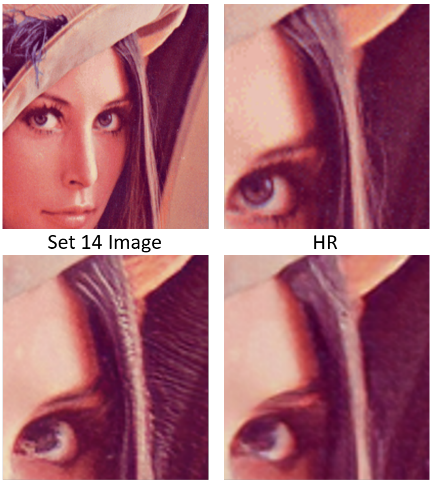


Figure 4: **In-Scale Qualitative Comparison with CNN:**. This figure shows the reference image, the high resolution image (HR), the 4x super-resolved image using Structure-Preserving Super Resolution (SPSR) [23] and our model's output (CiSR-GAN). In the SPSR output, we see lines in the background and artifacts in the eye and the hair whereas CiSR-GAN produces more realistic result.

taining comparable PSNR, as shown in Table. 2. Generally there is large gap between the SPSR and CiSR-GAN based on LPIPS metric, however, the difference is small in the test set Urban100 [14]. This behavior is expected as the gradient guidance based structure priors used in their model encourage the retrieval of lines and geometry that are commonly found in that dataset.

5 Conclusion

In this work, we improved the perceptual quality of the implicit neural network based single image super-resolution. The main hindrance in utilizing adversarial losses for continuous image representation models was the random co-ordinate-based sampling procedure adopted by previous works. We proposed to use a patch-based sampling method. Then we trained the implicit neural network with additional objectives based on adversarial and perceptual losses. We demonstrated that the resulting network produces sharp and photo-realistic images while maintaining the desirable properties of the implicit neural networks i.e out-of-scale super-resolution. As future work, our method can also be trained with gradient guidance based structure prior to improve PSNR.

Acknowledgements. This work was partially supported by the Norwegian Research Council (grant number 296093) and the members of the SmartRocks joint industry project (ENI AS, Repsol AS, and Chevron Corporation).

References

[1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[2] I. Anokhin, K. Demochkin, T. Khakhulin, G. Sterkin, V. Lempitsky, and D. Korzhenkov. Image generators with conditionally-independent pixel synthesis. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14273–14282, 2021. DOI: `10.1109/CVPR46437.2021.01405`.

[3] M. Atzmon and Y. Lipman. Sal: Sign agnostic learning of shapes from raw data. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2562–2571, 2020. DOI: `10.1109/CVPR42600.2020.00264`.

[4] A. Basher, M. Sarmad, and J. Boutellier. Lightsal: Lightweight sign agnostic learning for implicit surface representation. *CoRR*, abs/2103.14273, 2021.

[5] M. Bevilacqua, A. Roumy, C. Guillemot, and M. line Alberi Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *Proceedings of the British Machine Vision Conference*, pages 135.1–135.10. BMVA Press, 2012. DOI: `http://dx.doi.org/10.5244/C.26.135`.

[6] Y. Blau and T. Michaeli. The perception-distortion tradeoff. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6228–6237, 2018. DOI: `10.1109/CVPR.2018.00652`.

[7] E. R. Chan, M. Monteiro, P. Kellnhofer, J. Wu, and G. Wetzstein. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5795–5805, 2021. DOI: `10.1109/CVPR46437.2021.00574`.

[8] Y. Chen, S. Liu, and X. Wang. Learning continuous image representation with local implicit image function. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8624–8634, 2021. DOI: `10.1109/CVPR46437.2021.00852`.

[9] Z. Chen and H. Zhang. Learning implicit fields for generative shape modeling. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5932–5941, 2019. DOI: `10.1109/CVPR.2019.00609`.

[10] J. Chibane, M. A. mir, and G. Pons-Moll. Neural unsigned distance fields for implicit function learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 21638–21652. Curran Associates, Inc., 2020.

[11] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016. DOI: `10.1109/TPAMI.2015.2439281`.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[14] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015. DOI: `10.1109/CVPR.2015.7299156`.

[15] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[16] A. Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard GAN. In *International Conference on Learning Representations*, 2019.

[17] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 1097–1105, USA, 2012. Curran Associates Inc.

[19] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems 2*, pages 396–404. Morgan-Kaufmann, 1990.

[20] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017. DOI: `10.1109/CVPR.2017.19`.

[21] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017. DOI: `10.1109/CVPRW.2017.151`.

[22] Y. Lu. The level weighted structural similarity loss: A step away from the mse, 2019.

[23] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou. Structure-preserving super resolution with gradient guidance. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7766–7775, 2020. DOI: `10.1109/CVPR42600.2020.00779`.

[24] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423 vol.2, 2001. DOI: `10.1109/ICCV.2001.937655`.

[25] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4455–4465, 2019. DOI: `10.1109/CVPR.2019.00459`.

[26] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020.

[27] K. Nazeri, H. Thasarathan, and M. Ebrahimi. Edge-informed single image super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3275–3284, 2019. DOI: `10.1109/ICCVW.2019.00409`.

[28] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 165–174, 2019. DOI: `10.1109/CVPR.2019.00025`.

[29] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[30] K. Schwarz, Y. Liao, M. Niemeyer, and A. Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. In H. Larochelle,

M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 20154–20166. Curran Associates, Inc., 2020.

[31] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 7462–7473. Curran Associates, Inc., 2020.

[32] I. Skorokhodov, S. Ignatyev, and M. Elhoseiny. Adversarial generation of continuous images. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10748–10759, 2021. DOI: `10.1109/CVPR46437.2021.01061`.

[33] K. O. Stanley. Compositional pattern producing networks: A novel abstraction of development. *Genetic programming and evolvable machines*, 8(2):131–162, 2007.

[34] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *arXiv preprint arXiv:2006.10739*, 2020.

[35] X. Wang, K. Yu, K. C. Chan, C. Dong, and C. C. Loy. Basicsr, 2020.

[36] X. Wang, K. Yu, C. Dong, and C. Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 606–615, 2018. DOI: `10.1109/CVPR.2018.00070`.

[37] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In L. Leal-Taixé and S. Roth, editors, *Computer Vision – ECCV 2018 Workshops*, pages 63–79, Cham, 2019. Springer International Publishing.

[38] X. Xu, Z. Wang, and H. Shi. Ultrasr: Spatial encoding is a missing key for implicit image function-based arbitrary-scale super-resolution. *CoRR*, abs/2103.12716, 2021.

[39] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010. DOI: `10.1109/TIP.2010.2050625`.

[40] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In J.-D. Boissonnat, P. Chenin, A. Cohen, C. Gout, T. Lyche, M.-L. Mazure, and L. Schumaker, editors, *Curves and Surfaces*, pages 711–730, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[41] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. DOI: `10.1109/CVPR.2018.00068`.

[42] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. *CoRR*, abs/1802.08797, 2018.

Appendix B

Paper B: SIT-SR 3D: Self-supervised slice interpolation via transfer learning for 3D volume super-resolution

Pattern Recognition Letters 166 (2023) 97–104

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

SIT-SR 3D: Self-supervised slice interpolation via transfer learning for 3D volume super-resolution

Muhammad Sarmad [a,*], Leonardo Carlos Ruspini [b], Frank Lindseth [a]

[a] *Department of Computer Science, NTNU, Trondheim, Norway*
[b] *Petricore, Norway*

ARTICLE INFO

Article history:
Received 18 December 2021
Revised 29 November 2022
Accepted 10 January 2023
Available online 12 January 2023

Edited by: Jiwen Lu

Keywords:
Super-resolution
Digital rock analysis
Self-supervised learning

ABSTRACT

We present SIT-SR 3D, a novel self-supervised method for 3D single image super-resolution (SISR). Scaling 2D SISR networks to 3D SISR requires code redesign, high computing resources, and 3D ground-truth. However, we circumvent this by (1) using a pre-trained 2D SISR for indirect supervision and (2) using a novel consistency loss to learn frame interpolation. Any pre-trained state of the art 2D SISR method can replace the 2D SISR used in SIT-SR 3D, thus transferring the merits of 2D to 3D and ensuring modularity. We trained two end-to-end 3D baselines in a supervised setting; a 3D RRDBNet trained only with L1 loss and a 3D ESRGAN trained with adversarial and perceptual loss. We show that the proposed pipeline's self-supervised version is qualitatively better than the baselines. When trained in a supervised setting, SIT-SR 3D achieves better PSNR than its counterparts. Furthermore, our pipeline uses fewer parameters compared to the baselines. We demonstrate our results on an open-source digital rock CT dataset. Our code and pre-trained models will be made publicly available.

1. Introduction

The future depends on our capacity to minimize the emissions of CO2 in the atmosphere while keeping the economy's wheel in motion to reduce poverty and improve the quality of life in developing countries. Upcoming technologies such as Carbon dioxide Capture and Storage (CCS) and more efficient oil and gas production will play a significant role in achieving carbon neutrality. The derivation of rock properties from high-resolution CT images (Digital Rock) is a disruptive technology in that it can fundamentally change the way we characterize rocks. High-resolution photos can help characterize the properties of rocks and minerals, such as porosity, permeability, and flow [1]. Often it is beneficial to enhance or super-resolve a 3D image before usage in other domains, e.g., medical CT [2,3].

Super-resolving a 3D image presents unique challenges compared to a 2D image because of three prominent reasons. The first reason is the lack of high and low-resolution ground-truth image pairs in 3D. It is often costly and time-consuming to obtain such images. However, recent advances in deep learning [4,5] and single image super-resolution (SISR) [6] rely heavily on such paired data, which is not always available. While 2D image pairs are available in abundance, they cannot be used directly to train 3D pipelines.

Secondly, the advances in 2D image super-resolution techniques are not always scalable to 3D images. e.g., a 3D equivalent of perceptual loss does not exist to the best of our knowledge [7]. Similarly, a 3D version of SRGAN requires a 3D convolutional discriminator, which adds many training parameters [8]. Some recent advances in 2D SISR have custom operators and layers [9] that are not easily adaptable to the 3D domain and require low-level code redesign. Similarly, transfer learning from well-known architectures, e.g., ResNet50 [10] is not possible as the state-of-the-art encoders consist of 2D convolution layers. The only option to train 3D variants of deep networks on 3D data is to design from scratch.

Thirdly, if we design the 3D variants of novel operators from the 2D domain by writing custom code, the training cost increases cubically with the resolution. For example, the cost of training increases when enhancing the input size of the 3D convolutional discriminator in an adversarial setting. Some of the custom loss functions, such as perceptual loss, need to be applied to individual 2D slices of the 3D image and cause additional computation and latency overhead. To overcome these challenges, we propose an architecture that transfers the benefits of advances in 2D SISR to 3D SISR without engineering a 3D version of 2D SISR. We also do not require any paired 3D ground-truth data.

[*] Corresponding author.
E-mail addresses: muhammad.sarmad@ntnu.no (M. Sarmad), Leonardo.ruspini@petricore.com (L.C. Ruspini), frankl@ntnu.no (F. Lindseth).

https://doi.org/10.1016/j.patrec.2023.01.008

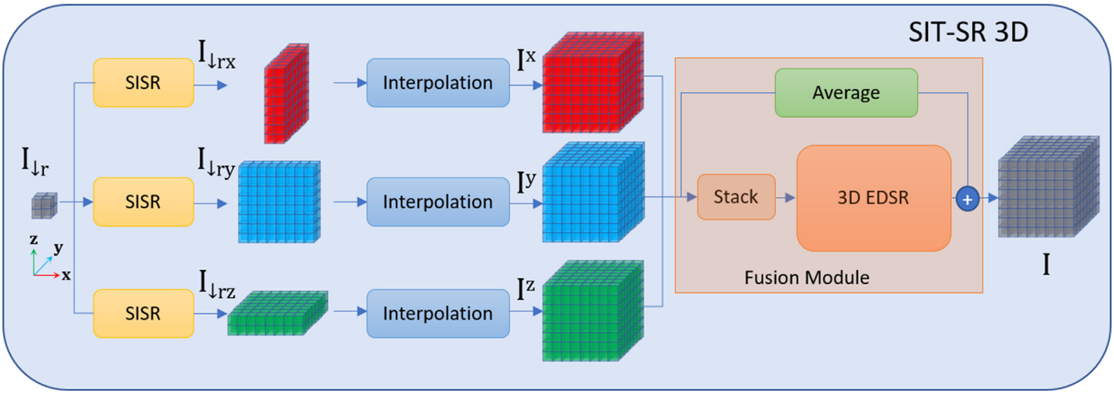


Fig. 1. SIT-SR 3D: The architecture of our proposed method. The low-resolution image $I_{\downarrow\ r}$ is upsampled along *x*, *y* and *z* respectively with the pre-trained 2D SISR, as a result, the volumes $I_{\downarrow\ r_x}$, $I_{\downarrow\ r_y}$and $I_{\downarrow\ r_z}$are obtained. I^x, I^y, and I^z are the corresponding volumes obtained after the interpolation operation. I^x, I^y, and I^z are passed through the Fusion module to get the output volume I.

In this work, we present SIT-SR 3D, a novel self-supervised 3D volume super-resolution technique. Our architecture is shown in Fig. 1. We employ transfer learning by utilizing a 2D SISR model trained on 2D feature-rich networks. We use this pre-trained 2D SISR to train the 3D SISR model on low-resolution 3D data only. We super-resolve the 3D image using this 2D SISR model along each of the three possible dimensions. We obtain the final volume by merging the three images using a Fusion module. This slice interpolation network's weights can be learned in an entirely self-supervised manner or with high-resolution ground-truth, if available. We introduce a consistency loss to train our method in a self-supervised manner. Our pipeline's self-supervised version compares well to the supervised baseline. We perform ablation studies and also compare the qualitative and quantitative results with multiple baselines. Our key contributions are the following:

- We present SIT-SR 3D, a novel self-supervised interpolation and transfer learning framework for 3D volume super-resolution.
- Our method can use any pre-trained 2D SISR model with desired qualities to transfer the merits of 2D SISR to the 3D pipeline.
- We propose a novel consistency loss for training SIT-SR 3D without 3D ground-truth.
- The approach is data-efficient, uses fewer parameters. Moreover, training converges fast and does not require 3D high-resolution ground truth.

2. Related works

2.1. Super resolution

The advent of convolutional neural networks (CNNs) [4] led to applications of deep learning for various computer vision problems. Single image super-resolution (SISR) is one such task that has benefited from the progress in deep learning. Initially, Dong et al. [11] proposed to perform end-to-end SISR using SR-CNN. They established state of the art by outperforming conventional methods such as the sparse coding-based method by Yang et al. [12]. Kim et al. [13] added skip connections and designed a lightweight recursive CNN architecture. Generative adversarial networks [14] can produce photo-realistic images but are hard to train due to the adversarial nature of training. Ledig et al. [15] utilized adversarial training in conjunction with the content loss to obtain photo-realistic SISR. Lim et al. [16] improved a ResNet based architecture [10] by removing batch normalization layers and introduced a multi-scale architecture to further enhance performance. Zhang et al. [17] applied an improved DenseNet [18] for SISR by removing batch normalization layers, pooling layers, and introduced a global feature fusion.

Various super-resolution quality measures have been developed, such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). There is no single metric that is considered better than the other, and both have their merits as shown in various work [19,20]. Both of these metrics have their shortcomings since they do not model human perception. Therefore, some works have employed human-based perceptual evaluation [15]. However, such measures are costly and time-consuming to obtain. We can extend state of the art 2D SISR methods for 3D images with significant customized code changes. However, we propose to utilize 2D SISR without extending them to 3D.

2.2. 3D super resolution

Previous works which super resolve 3D data are of two distinct types. The first one treats the entire *3D image as a collection of 2D slices* and then performs SISR on individual slices. Therefore, we can utilize any traditional 2D SISR methods applicable to colored images. However, such methods can only super-resolve a 3D image along two dimensions, and the third dimension is still of lower resolution. 3D images contain contextual information in all dimensions, and we must consider all three dimensions to upsample such images. There are many examples of using CT data to train 2D SISR networks for super resolving 2D slices of a 3D image in both the medical and digital rock domain [21–23].

The other type trains *end-to-end 3D networks* to super-resolve volumes. These methods are more challenging to design; hence only a few works exist, but they provide a complete solution [2,3]. Chen et al. [2] proposed mDSCRN for 3D volume super-resolution inspired by DenseNet [18]. Peng et al. [3] proposed SAINT and demonstrated that mDCSRN suffers from sub-optimal results and also has a higher memory and compute footprint. However, their method needs ground truth high-resolution data for supervised training. The approach has been applied to medical CT data that requires upsampling in one dimension. Therefore, they perform a frame interpolation method. However, our data requires upsampling in all three dimensions. They do not provide any code for comparison.

3. Method

3.1. Problem formulation

The proposed work provides a solution to 3D volumetric super-resolution. We demonstrate our method on 3D CT Images. Consider a 3D image $\mathrm{I} \rightarrow \mathrm{I}(\mathrm{x}, \mathrm{y}, \mathrm{z}) \in R^{X \times Y \times Z}$ which represents a densely sampled CT image. For *I*, the corresponding sparsely sampled volume $\mathrm{I}_{\downarrow \mathrm{r}}$ is defined as:

$$\mathrm{I}_{\downarrow r} = \mathrm{I}(r \cdot x, r \cdot y, r \cdot z) \tag{1}$$

where $\mathrm{I}_{\downarrow r} \in R^{\frac{X \times Y \times Z}{r}}$, and *r* is the sparsity factor along the *x*, *y* and *z* axis from I to $\mathrm{I}_{\downarrow \mathrm{r}}$ and the up-sampling factor from $\mathrm{I}_{\downarrow \mathrm{r}}$ to I.

Along each axis, there can be three kinds of slices which are referred to as follows:

- The down sampled slices along x-axis are given as: $\mathrm{I}_{\downarrow \mathrm{r_x}} = I(\mathrm{r} \cdot \mathrm{x}, \mathrm{y}, \mathrm{z}), \sim \forall \mathrm{x}$. Interpolated version of $\mathrm{I}_{\downarrow \mathrm{r_x}}$ is given as I^{x}.
- The down sampled slices along y-axis are given as: $\mathrm{I}_{\downarrow \mathrm{r_y}} = I(\mathrm{x}, \mathrm{r} \cdot \mathrm{y}, \mathrm{z}), \sim \forall \mathrm{y}$. Interpolated version of $\mathrm{I}_{\downarrow \mathrm{r_y}}$ is given as I^{y}.
- The down sampled slice along z-axis are given as: $\mathrm{I}_{\downarrow \mathrm{r_z}} = I(\mathrm{x}, \mathrm{y}, \mathrm{r} \cdot \mathrm{z}), \sim \forall \mathrm{z}$. Interpolated version of $\mathrm{I}_{\downarrow \mathrm{r_z}}$ is given as I^{z}.

The *objective of SIT-SR 3D* is to find a mapping $\mathcal{F} : \mathrm{R}^{\frac{X \times Y \times Z}{r}} \rightarrow R^{X \times Y \times Z}$ that can convert $\mathrm{I}_{\downarrow \mathrm{r}}$ back to I for a given resolution factor r.

3.2. Overview of the proposed approach

Fig. 1: shows the overview of SIT-SR 3D. The low-resolution image $\mathrm{I}_{\downarrow \mathrm{r}}$ is upsampled along x, y, and z respectively, using a pre-trained SISR network trained on 2D images. After this operation, we obtain $\mathrm{I}_{\downarrow \mathrm{r_x}}$, $\mathrm{I}_{\downarrow \mathrm{r_y}}$ and $\mathrm{I}_{\downarrow \mathrm{r_z}}$. These anisotropic 3D volumes are converted to isotropic volumes using an interpolation operation. After the interpolation, we obtain the volumes I^{x}, I^{y}, and I^{z}, which are stacked to form a single volume with three channels. We then process this volume by a 3D convolution-based 3D EDSR [16]. The average of I^{x}, I^{y}, and I^{z} is also added to the output of the 3D EDSR to form I. We can supervise the output of 3D EDSR using the high-resolution ground-truth I_{GT} with L_1 loss to learn how to combine I^{x}, I^{y} and I^{z}. We can also train it in a self-supervised manner using I^{x}, I^{y} and I^{z} and a novel consistency loss. One can also combine both loss functions to obtain a hybrid loss formulation that provides control over the network's output. We demonstrate results with all three loss function settings. Next, we will describe each of the modules in SIT-SR 3D in detail.

4. Single image super-resolution

This module's objective is to super-resolve 3D images efficiently using transfer learning based on models trained on 2D single image super-resolution (SISR). The proposed pipeline is modular, which means we can train it with any domain-specific dataset, e.g., 2D CT images, or it can also be trained on colored image datasets to learn useful features depending on the objective. Recent work of Asano et al. [24] demonstrated that a single image contains enough information to train the initial layers of a neural network given enough data augmentation. Our fundamental motivation is to enable rapid prototyping by selecting any 2D SISR trained on a custom dataset.

We process the low-resolution image $\mathrm{I}_{\downarrow \mathrm{r}}$ using a state-of-the-art 2D SISR pipeline. The selection of a 2D SISR pipeline is crucial for transfer learning and self-supervised learning of the subsequent module, i.e., the Fusion module. The Fusion module relies on the pre-trained 2D SISR model in the self-supervised case. We use ESRGAN [25] trained on 2D CT images to demonstrate that the properties such as realism and sharpness, inherited from a pre-trained 2D SISR, can be transferred to the Fusion module. ESRGAN is an encoder-decoder-based architecture trained using L_1, adversarial and perceptual loss.

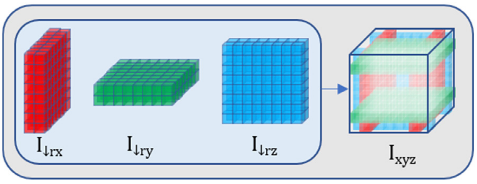

$\mathrm{I}_{\downarrow \mathrm{rx}}$
$\mathrm{I}_{\downarrow \mathrm{ry}}$
$\mathrm{I}_{\downarrow \mathrm{rz}}$
$\mathrm{I}_{\mathrm{xyz}}$

Fig. 2. 2D SISR slices overlap: $\mathrm{I}_{\downarrow r_x}$, $\mathrm{I}_{\downarrow r_y}$and $\mathrm{I}_{\downarrow r_z}$ can be superimposed to obtain I_{xyz}. The empty regions are missing information that SIT-SR 3D learns.

The input image $\mathrm{I}_{\downarrow \mathrm{r}}$ is processed along three axes x, y, and z using a pre-trained 2D SISR. Consider the dimensions of $\mathrm{I}_{\downarrow \mathrm{r}}$ given by Eq. (2).

$$\mathrm{I}_{\downarrow r} : c \times X \times Y \times Z \tag{2}$$

Where c are the number of channels of greyscale or RGB image. We need to process $\mathrm{I}_{\downarrow \mathrm{r}}$ along each dimension x, y, and z. However, 2D SISR can process images of the form $\mathrm{I}_{2\mathrm{D}} : B \times c \times H \times \mathrm{W}$ only, where B is the batch size, H and W are the dimensions of the 2D image. Therefore, in order to process $\mathrm{I}_{\downarrow \mathrm{r}}$, we first need to transpose it to obtain three copies $\mathrm{I}_{\downarrow \mathrm{rcx}}$, $\mathrm{I}_{\downarrow \mathrm{rcy}}$ and $\mathrm{I}_{\downarrow \mathrm{rcz}}$ as shown in Eqs. (3), (4) and (5).

$$\mathrm{I}_{\downarrow \mathrm{rcx}} : X \times c \times Y \times \mathrm{Z} \tag{3}$$

$$\mathrm{I}_{\downarrow \mathrm{rcy}} : Y \times c \times X \times \mathrm{Z} \tag{4}$$

$$\mathrm{I}_{\downarrow rcz} : Z \times c \times X \times Y \tag{5}$$

We transpose $\mathrm{I}_{\downarrow \mathrm{r}}$ such that the considered axis is along the first dimension, whereas the remaining axes are on the third and the fourth dimension. We then process images $\mathrm{I}_{\downarrow \mathrm{rcx}}$, $\mathrm{I}_{\downarrow \mathrm{rcy}}$ and $\mathrm{I}_{\downarrow \mathrm{rcz}}$ by 2D SISR by considering the first dimension as the batch size. We again transpose the output of 2D SISR to restore the axes to normal for all three inputs. This operation's output gives us three upsampled anisotropic volumes $\mathrm{I}_{\downarrow \mathrm{r_x}}$, $\mathrm{I}_{\downarrow \mathrm{r_y}}$ and $\mathrm{I}_{\downarrow \mathrm{r_z}}$. All three inputs $\mathrm{I}_{\downarrow \mathrm{rcx}}$, $\mathrm{I}_{\downarrow \mathrm{rcy}}$ and $\mathrm{I}_{\downarrow \mathrm{rcz}}$ share the weights of 2D SISR, i.e., we use the same network for upsampling along axes x, y, and z.

4.1. Anisotropic volume interpolation

Three anisotropic volumes $\mathrm{I}_{\downarrow \mathrm{r_x}}$, $\mathrm{I}_{\downarrow \mathrm{r_y}}$ and $\mathrm{I}_{\downarrow \mathrm{r_z}}$ contain unique information. If they are overlayed on each other to fit in a isotropic cube, they result in a sparsely populated cube $\mathrm{I}_{\mathrm{xyz}}$ which contains several empty and filled regions as shown in Fig. 2. One can treat the transformation from $\mathrm{I}_{\mathrm{xyz}}$ to I as an inpainting problem. However, We first interpolate $\mathrm{I}_{\downarrow \mathrm{r_x}}$, $\mathrm{I}_{\downarrow \mathrm{r_y}}$ and $I_{\downarrow \mathrm{r_z}}$ to form I^{x}, I^{y} and I^{z}. To achieve this, we employ a trilinear interpolation operation. The interpolated volumes, I^{y} and I^{z} are now isotropic and still contain all the useful information which was present in $\mathrm{I}_{\downarrow \mathrm{r_x}}$, $\mathrm{I}_{\downarrow \mathrm{r_y}}$and $\mathrm{I}_{\downarrow \mathrm{r_z}}$. We feed these three volumes to the Fusion module as input.

4.2. Fusion module

We combine the information contained in the isotropic volumes I^{x}, I^{y}, and I^{z} into a single volume I using the Fusion module. The Fusion module's goal is to learn how to combine this information

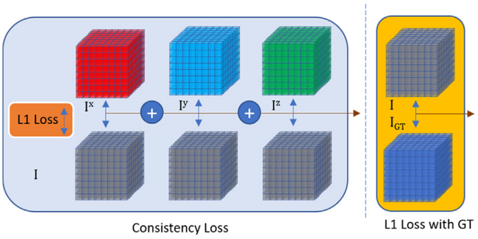


Fig. 3. Consistency Loss: I^x, I^y, and I^z are compared with I based on L_1 distance. This leads to unsupervised training compared to L_1 loss between I and I_{GT}.

in the best possible way. The Fusion module stacks the three inputs on the channel dimension such that we obtain a three-channel volume $I_{stack} : 3{*}c \times X \times Y \times Z$. This volume passes through a 3D convolutional neural network. We adapt EDSR [16] architecture to 3D for our purposes by converting all operations such as convolution and batch normalization to 3D to form 3D EDSR. We also take the average of the three input cubes and add it to the output of the 3D EDSR as shown in Eq. (6).

$$I_{avg} = \frac{I^x + I^y + I^z}{3} \tag{6}$$

Note that the input to the 3D EDSR has three channels, whereas this module's output is a 3D volume with a single channel. The input and output of the average module is a 3D volume with a single channel. The Fusion module's output is I with a single channel, which is the upsampled version of $I_{\downarrow r}$ by a factor r.

4.3. Loss function formulation

We train the Fusion module in a supervised or self-supervised manner. We achieve this by employing various loss functions. Next, we describe these loss functions in detail.

4.4.1. Supervised training with L_1 loss

If ground truth paired images are available, Fusion module can be trained in a supervised setting by using L_1 loss. We calculate the L_1 distance between I and the ground-truth high-resolution image I_{GT} by using the Eq. (7):

$$L_1 = \|I - I_{GT}\|_1 \tag{7}$$

The L_1 loss has a smoothing effect on the output, which is a known property of L_1 loss [15,25]. If we train 2D SISR on the adversarial loss, training the Fusion module with L_1 loss alone can lead to over smoothing of the output image I. This formulation leads to an inadequate transfer of properties of 2D SISR to I, but this method achieves the highest possible PSNR. This method also requires ground truth that is either not available or is expensive to obtain in medical CT or digital rock domains.

4.4.2. Self-Supervised training with consistency loss

We propose a unique formulation of loss to transfer the 2D SISR properties to the Fusion model. We call this formulation the consistency loss. The loss is given in Fig. 3 and Eq. (8):

$$L_c = \|I^x - I\|_1 + \|I^y - I\|_1 + \|I^z - I\|_1 \tag{8}$$

This loss calculates the L_1 distance of I from each of the three interpolated volumes I^x, I^y and I^z. It does not require ground truth image I_{GT}. In our experiments, we note that consistency loss maintains the desirable properties of preceding 2D SISR in the output image I. It can also filter out some high-frequency noise due to the L_1 nature of the three terms. The resulting image is sharp and also has lower noise levels than the real image. This noise filtering is an added advantage as CT images often suffer from high-frequency noise.

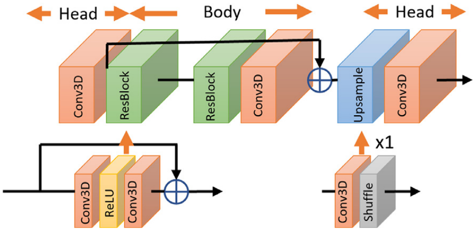


Fig. 4. 3D EDSR Architecture: We modify EDSR [16] and add 3D convolutional layers instead of 2D. This 3D EDSR is the part of the fusion module of our SIT-SR 3D model.

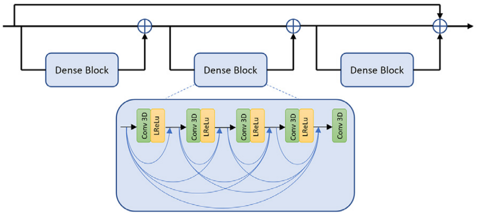


Fig. 5. 3D RRDB: The 3D residual in residual dense block contain 3D convolutional layers instead of 2D. We use this as the building block of our 3D RRDB-Net and 3D ESRGAN baselines [25].

4.4.3. Hybrid loss

We also propose to use L_1 loss in tandem with consistency loss that allows controlling the quality of the output. We, therefore, introduce a hybrid loss. The hybrid loss is given in Eq. (9).

$$L_{\langle} = \alpha L_1 + (1 - \alpha) L_{\rfloor} \tag{9}$$

The parameter α serves as a tuning parameter that can control the contribution of each loss. By changing the parameter α, we can obtain the desired quality in the output image I. We can set the α value closer to 1 to achieve a higher PSNR. On the other hand, if the output needs to be closer to the 2D SISR model, a lower value can be used.

5. Experiments

5.1. Data

We utilize the digital rock dataset provided by Wang et al. [21] since it is the largest dataset available for this study. This work uses the default train, validation, and test split. This dataset consists of paired 2D (12,000 images) and 3D (3000 images) low and high-resolution images of various rock types. We use the input (x4 downsampled) and output image pairs for this study while using all rock types they provide. We used the 2D image pairs to train 2D ESRGAN while using 3D image pairs to train the 3D baselines and the Fusion module in supervised mode. They provide low-resolution images by downsampling high-resolution images using various downsampling operations such as box, triangle, lanczos2, lanczos3, and Lanczos. We use this dataset to ensure that the model is robust to the type of down-sampling operation used.

5.2. Training and testing details

We used PyTorch for all our models, and experiments. We utilize two RTX Titan GPUs for training our models. We build upon open-source Github repositories [26,27].

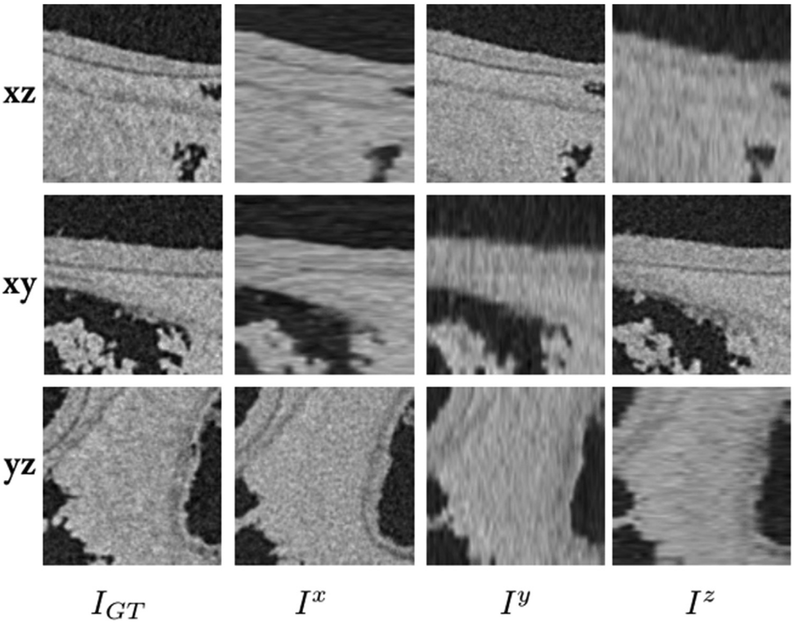


Fig. 6. Why the Fusion module is necessary? The first slice of cubes I_{GT}, I^x, I^y, I^z from all three faces, i.e., xz, xy, and yz, is shown. Only one of the frames for each I^x, I^y, and I^z contains sharp details. To form I, all three frames should be super-resolved. The Fusion module helps to combine the best of I^x, I^y, and I^z.

5.2.1. 2D SR image pipeline

We train a 2D ESRGAN [25] for the image super-resolution using 2D paired images provided by Wang et al. [21]. We modify the generator of ESRGAN to have 23 residual blocks. We use a batch size of 16. The image size used for training was 128 × 128. We use rotation and flipping to augment the training data. We first train the generator for 50 epochs with L_1 loss. Then we use the weights of this generator to train it further for 15 epochs with adversarial and VGG loss to obtain the final 2D ESRGAN configuration. More training details are in the supplementary section.

5.2.2. Interpolation of anisotropic 3D volumes

The anisotropic volumes produced after the 2D operation are converted to isotropic volumes using an interpolation operation. We used trilinear interpolation for this purpose due to its low computational cost. We also considered the consistency loss variant, which works with anisotropic volumes directly without the interpolation. However, this leads to checkerboard artifacts since the network is not motivated to learn anything meaningful or spatial consistent in the empty 3D regions.

5.2.3. Fusion module

Fig. 6 shows the importance of the Fusion module to combine I^x, I^y and I^z to formulate I. We train SIT-SR 3D in three different configurations (supervised, self-supervised and hybrid) as described in the method section. We can train the hybrid configuration with various values of the parameter α. We document the result with α set to 0.5, but in our experiments, we found that this parameter can control the output quality. Choosing a high value for α leads to smooth output and high PSNR, while low values drive the results closer to the underlying pipeline's properties. The training converges in about 12 h with a batch size of 2. We use a 3D EDSR architecture for the Fusion module as shown in Fig. 4. This custom 3D EDSR contains 16 residual blocks. 3D EDSR is the only component in the Fusion module that needs to be learned.

5.2.4. 3D baseline

To perform a comparison, we create our own 3D baseline networks. The baseline of choice is ESRGAN [25] due to its good performance in 2D domain. We designed the baseline network by converting all 2D operations such as convolution, batch normalization, etc., into 3D versions. The basic building block of our baselines is the 3D residual in residual dense block (3D RRDB) as shown in Fig. 5. Similarly, the 3D version of adversarial loss is obtained by converting the 2D convolution-based discriminator model to a 3D convolution-based model. However, a 3D version of VGG loss [7] does not exist. Therefore, we apply a 2D VGG loss frame by frame on the complete 3D volume along all three dimensions such that we obtain three-loss values along each dimension. We take the average of the three VGG loss terms to ensure the best possible training for the network.

3D ESRGAN uses 3D RRDBNet as the generator architecture. We first train 3D RRDBNet with L_1 loss. We then use this 3D RRDBNet as one of the baselines. We train 3D RRDBNet with the L_1 loss as 3D RRDBNet for reference in the quantitative and qualitative results. We then train the 3D ESRGAN by using weights of pretrained 3D RRDBNet as initial weights. This network formulates the second baseline, and we call it 3D ESRGAN in results. This work uses the default test, and train splits provided by [21]. 3D RRDBNet was trained with the batch size 6, while we train the 3D ESRGAN with the batch size two due to the GPU memory constraint caused by an additional discriminator network. The input channel size for both networks was one instead of three due to the training set's greyscale nature. We flip and rotate the training images for data augmentation. The size of the high-resolution image used for training was 92^3.

5.3. Quantitative results

Table 1 summarizes the quantitative results. SIT-SR 3D performs well in both supervised and self-supervised settings. It out-

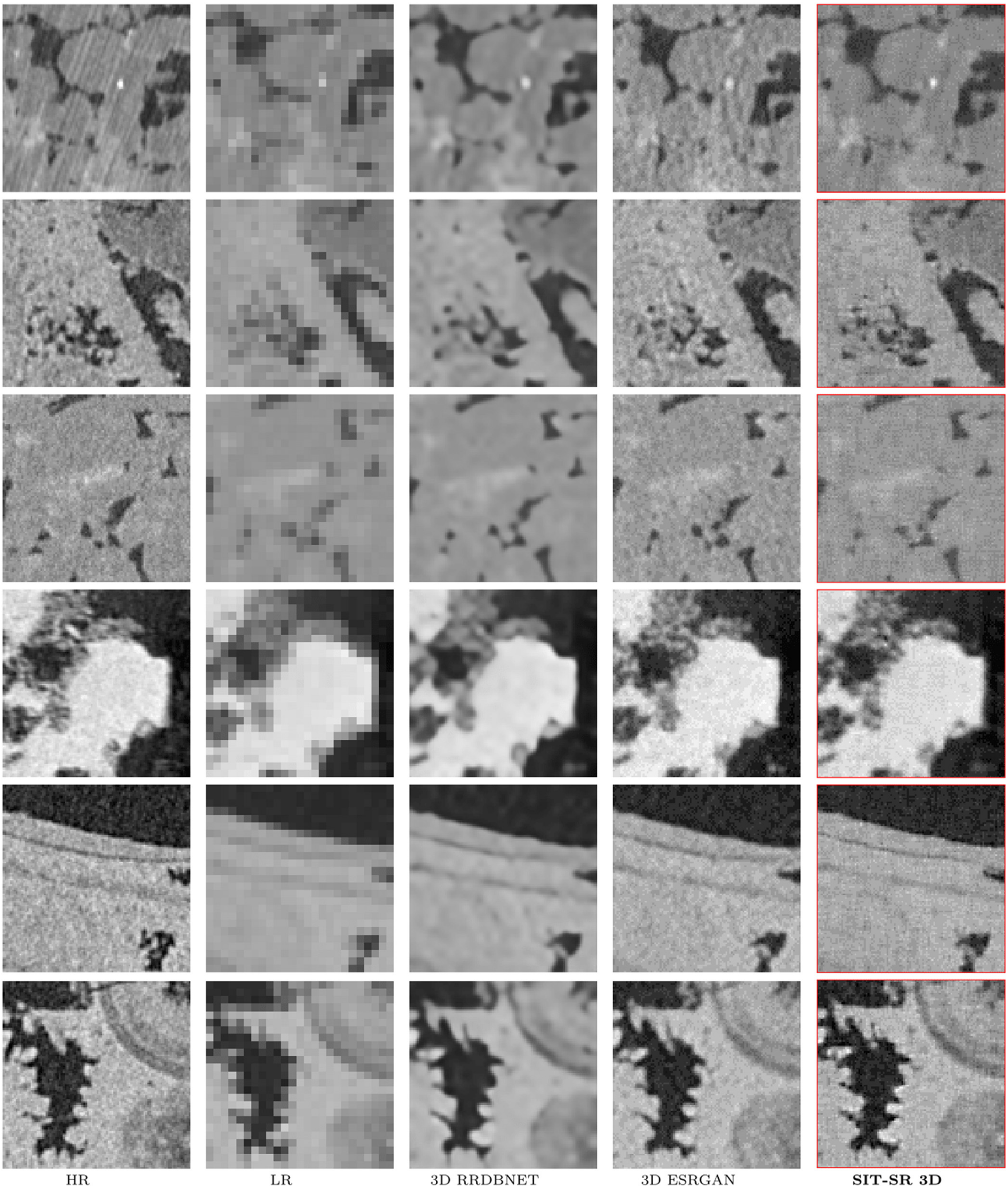


Fig. 7. This figure shows a visual comparison of different methods. HR, LR indicates high-resolution and low-resolution images. 3D RRDBNet is a 3D convolution-based network supervised with L_1 loss. 3D ESRGAN is trained with GAN and VGG loss using pre-trained weights of 3D RRDBNet. SIT-SR 3D is trained in an utterly self-supervised setting using only the consistency loss.

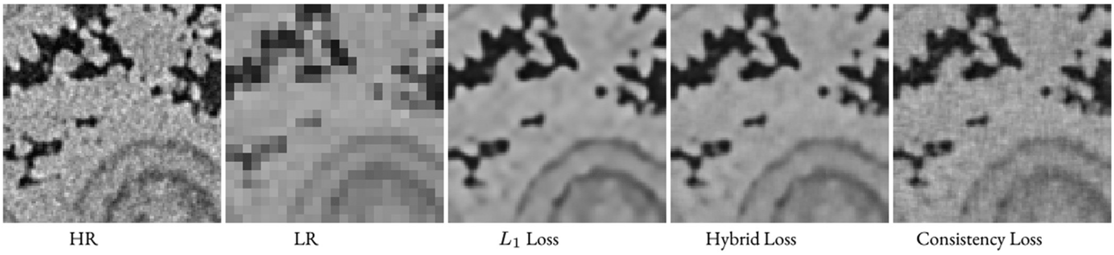


Fig. 8. Effect of α: This figure shows the ablation study with various values of α in Eq. (9). The values are 1, 0.5, and 0 for supervised (L_1 Loss), hybrid, and self-supervised (consistency loss) respectively of SIT-SR 3D.

Table 1
Quantitative comparison of SIT-SR 3D with 3D baselines. The best supervised case is underlined. The self-supervised case is shown in **bold**. We perform all experiments with x4 down-sampled image as input.

Arch	Loss	Params	PSNR (dB)	SSIM
3D RRDBNet	L_1	50.05 M	30.84	<u>0.71</u>
3D ESRGAN	GAN	101.15 M	28.41	0.60
SIT-SR 3D$_{sup}$	L_1	31.44 M	<u>30.98</u>	0.69
SIT-SR 3D$_{self}$	L_c	31.44 M	**29.78**	**0.64**

Table 2
Effect of α and Average Module: We performed an ablation study to see the effect of α and average module on the PSNR and SSIM. The value of α is 1, 0.5 and 0. The best supervised case is underlined, whereas the best self-supervised case is in **bold**.

Scale	Loss	Average Module	PSNR (dB)	SSIM
x4	L_1	Yes	<u>30.987</u>	0.6921
	L_c	Yes	**29.789**	**0.6441**
	L_h	Yes	30.932	0.6846
	L_1	No	30.932	<u>0.6935</u>
	L_c	No	29.788	0.6440
	L_h	No	30.90	0.6844

performs the end-to-end 3D learning methods such as 3D RRDBNet in purely supervised settings. In the self-supervised setting, it has a higher PSNR than the 3D ESRGAN. The number of training parameters used by SIT-SR 3D are lesser than both 3D RRDBNet and 3D ESRGAN, demonstrating the merit of this approach.

5.4. Qualitative results

Fig. 7 shows the qualitative performance of SIT-SR 3D on the test set of 3D image pairs provided by Wang et al. [21]. The figure shows that SIT-SR 3D learns sharp details in the self-supervised setting well, and it also transfers the properties of the underlying 2D ESRGAN to the 3D SISR. The output of SIT-SR 3D is sharper than the one produced in a wholly supervised 3D RRDBNet (with L_1 loss). The produced result has less noise and artifact than the GT data and 3D ESRGAN's output due to the L_1 loss formulation of the consistency loss. Unfortunately, all the slices of the 3D image cannot be shown here due to presentation constraints. We have attached 3D images in the supplementary with the section and instructions on how to view them.

5.5. Ablation study

We performed ablations as shown in Fig. 8 and Table 2 to see the effect of α and the average module on the performance of SIT-SR 3D. The hybrid loss allows us to control the quality of the output, i.e., higher α leads to more blurry images with high PSNR results, and lower α favors the consistency loss and produces sharper results with lower PSNR. In this work, we only used α value of 0.5, which can be changed based on the required output. We also note that the average module has a relatively small effect on the PSNR, but since it consistently improves PSNR values in all experiments, we choose to keep it in the pipeline.

6. Discussion and future work

We have presented SIT-SR 3D, a modular and efficient network for super-resolution of 3D images. The proposed approach learns to super-resolve 3D low-resolution images in a self-supervised manner. We achieved this task by utilizing a 2D SISR pipeline trained with adversarial and VGG loss on 2D image pairs. We applied this 2D SISR along three dimensions of the 3D image to obtain three asymmetric cubes. We interpolated these anisotropic volumes using trilinear interpolation to obtain isotropic volumes. Then these symmetric cubes were fused to form a single cube. The fusion operation was learned by a 3D CNN using a novel consistency loss. The proposed approach outperformed the end-to-end 3D baseline when trained in a supervised manner in quantitative and qualitative metrics while using fewer parameters. We can use any 2D SISR pipeline depending on the desired output characteristics. SIT SR-3D is especially useful when 3D ground-truth is not available, but 2D ground-truth is available. In the case of digital rocks, it is often the case that 3D high-resolution ground truth is not available. We can go beyond a particular resolution using backscattered electrons in a scanning electron microscope (BSEM) [28] to obtain 2D images. We can then train a 2D SISR on 2D SEM images in these scenarios and transfer the knowledge to upsample low-resolution 3D CT images. However, this will be a subject of future study.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was partially supported by the Norwegian Research Council (grant number 296093) and the members of the SmartRocks joint industry project (ENI AS, Repsol AS, and Chevron Corporation).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patrec.2023.01.008.

References

[1] H. Andrä, N. Combaret, J. Dvorkin, E. Glatt, J. Han, M. Kabel, Y. Keehm, F. Krzikalla, M. Lee, C. Madonna, M. Marsh, T. Mukerji, E.H. Saenger, R. Sain, N. Saxena, Digital rock physics benchmarks—Part I: imaging and segmentation, Comput. Geosci. 50 (2013) 25–32.

[2] Y. Chen, F. Shi, A.G. Christodoulou, Z. Zhou, Y. Xie and D. Li, *Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network*, arXiv, 2018.

[3] C. Peng, W.-A. Lin, H. Liao, R. Chellappa, S.K. Zhou, SAINT: spatially aware interpolation network for medical slice synthesis, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6, 2020.

[4] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems, 1, 2012, pp. 1097–1105. -.

[5] Y. LeCun, B.E. Boser, J.S. Denker, D. Henderson, R.E. Howard, W.E. Hubbard, L.D. Jackel, Handwritten digit recognition with a back-propagation network, Adv. Neural Inf. Process. Syst. 2 (1990) 396–404.

[6] W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient, CoRR abs/1609.05158 (2016).

[7] J. Johnson, A. Alahi, F.-F. Li, Perceptual losses for real-time style transfer and super-resolution, CoRR abs/1603.08155 (2016).

[8] Y. Wang, Q. Teng, X. He, J. Feng, T. Zhang, CT-image of rock samples super resolution using 3D convolutional neural network, Comput. Geosci. 133 (2019) 104314 12.

[9] K. Nazeri, H. Thasarathan, M. Ebrahimi, Edge-Informed Single Image Super-Resolution, arXiv, 2019.

[10] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, CoRR abs/1512.03385 (2015).

[11] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, CoRR abs/1501.00092 (2015).

[12] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.

[13] J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network for image super-resolution, CoRR abs/1511.04491 (2015).

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014) 2672–2680.

[15] C. Ledig, L. Theis, F. Huszar, J. Caballero, A.P. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial, CoRR abs/1609.04802 (2016).

[16] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced deep residual networks for single image super-resolution, CoRR abs/1707.02921 (2017).

[17] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, CoRR abs/1802.08797 (2018).

[18] G. Huang, Z. Liu, K.Q. Weinberger, Densely connected convolutional networks, CoRR abs/1608.06993 (2016).

[19] Y. Lu, The Level Weighted Structural Similarity Loss: A Step Away from the MSE, arXiv, 2019.

[20] A. Horé, D. Ziou, Image quality metrics: PSNR vs. SSIM, in: 2010 20th International Conference on Pattern Recognition, 2010, pp. 2366–2369.

[21] Y.D. Wang, R.T. Armstrong, P. Mostaghimi, Enhancing resolution of digital rock images with super resolution convolutional neural networks, J. Pet. Sci. Eng. 182 (2019).

[22] H. Chen, X. He, Q. Teng, R.E. Sheriff, J. Feng, S. Xiong, Super-resolution of real–world rock micro-computed tomography images using cycle-consistent generative adversarial networks, Phys. Rev. E 101 (2) (2020).

[23] C. You, W. Cong, M.W. Vannier, P.K. Saha, E.A. Hoffman, G. Wang, G. Li, Y. Zhang, X. Zhang, H. Shan and, et al., CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE), IEEE Trans. Med. Imaging 39 (1) (2020) 188–203 1.

[24] Y. Asano, C. Rupprecht, A. Vedaldi, A critical analysis of self-supervision, or what we can learn from a single image, International Conference on Learning Representations, 2020.

[25] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C.C. Loy, Y. Qiao, X. Tang, ESRGAN: enhanced super-resolution generative adversarial networks, CoRR abs/1809.00219 (2018).

[26] X. Wang, K. Yu, K.C.K. Chan, C. Dong and C.C. Loy, *BasicSR*, https://github.com/xinntao/BasicSR, 2020.

[27] H.T. Kamyar Nazeri and M. Ebrahimi, Edge-informed-sisr, https://github.com/knazeri/edge-informed-sisr, 2019.

[28] P. Wandrol, J. Matějková, A. Rek, High resolution imaging by means of backscattered electrons in the scanning electron microscope, Mater. Struct. Micromech. Fract. V 567 (2008) 313–316 11.

Appendix C

Paper C: GPU Assisted Fast and Robust 3D Image Registration of Large Wet and Dry Rock Images Under Extreme Rotations

Highlights

GPU Assisted Fast and Robust 3D Image Registration of Large Wet and Dry Rock Images Under Extreme Rotations

Muhammad Sarmad, Johan Phan, Leonardo Ruspini, Gabriel Kiss, Frank Lindseth

- Our method accurately performs image registration for dry-wet image pairs of various textures
- Our algorithm can find a solution in under one minute compared to 1 hour taken by human expert by utilizing a graphical processing unit (GPU) for optimization.
- The algorithm can handle extreme rotation along the vertical and horizontal axes of the sample.

GPU Assisted Fast and Robust 3D Image Registration of Large Wet and Dry Rock Images Under Extreme Rotations

Muhammad Sarmad[a,*], Johan Phan[a,b], Leonardo Ruspini[b], Gabriel Kiss[a] and Frank Lindseth[a]

[a] *NTNU, Høgskoleringen 1, Trondheim, 7491, Norway*
[b] *Petricore , Stiklestadveien 1, Trondheim, 7041, Norway*

ARTICLE INFO

Keywords:
Dry-wet image registration
Digital rock analysis
Pytorch
AirLAB

ABSTRACT

Image registration is a process used to align or register multiple images or volumes to facilitate comparison or combination of the data. In the context of 3D wet and dry images of rock samples, it is essential to accurately align these images to analyze and utilize the data in various experiments. These images can be huge and contain minimal corresponding key points in the wet and dry images. A lack of these key points makes the registration problem extremely difficult. Traditional intensity-based optimization-based methods for image registration can be slow, while deep learning-based "one shot" methods fail due to a lack of key points and training data.

We propose a new optimization-based algorithm for image registration of large 3D wet and dry images of rock samples to address these issues. Our algorithm can handle extreme rotations of the samples and even complete inversion along horizontal axes. Additionally, it is optimized for speed while maintaining high accuracy by utilizing a graphical processing unit (GPU). We have demonstrated that our algorithm can provide a solution in under a minute for samples of size 1000^3 cube, compared to the several hours of expert time needed by the current industrial practice. We provide quantitative and qualitative results and compare our algorithm to the solution time of a human expert.

1. Introduction

Digital rock analysis is a sub-field in the geology field that involves using advanced imaging techniques to analyze rocks and other geological materials at a microscopic level. These techniques allow geologists to study rocks' internal structure and composition in great detail, which can provide valuable insights into the physical and chemical processes that have shaped the Earth's crust over time. Digital rock analysis can be used to study various rocks, including sedimentary, metamorphic, and igneous rocks. It can be applied to various research areas, including hydrocarbon reservoirs, environmental geology, and geomechanics. Some of the main techniques used in digital rock analysis include X-ray computed tomography (CT), scanning electron microscopy (SEM), and micro-computerized tomography (micro-CT).

Digital rock analysis often begins with creating an accurate 3D porosity model based on porous media X-ray images. This model can calculate various physical and fluid flow properties through image analysis techniques and flow simulations. This approach can be faster and less destructive than traditional laboratory measurements. However, when working with complex porous materials such as reservoir rocks, it is often necessary to use multiple X-ray images due to the wide range of pore sizes present in the sample. Additionally, it is typical for micro-CT images of rocks to have a significant portion of the percolating porosity below the resolution limit, making it challenging to model the pore network accurately Aarnes et al. (2007). Using a dry-wet image pair is common to obtain a more accurate porosity image in such cases. It can explain how the pore network changes as the rock absorb fluids. Feali et al. (2012); Long et al. (2013); Bhattad et al. (2014); Ruspini et al. (2021).

In dry-wet imaging, a sample is first scanned using X-rays while it is dry and then re-scanned after it has been saturated with a high X-ray attenuation fluid, such as brine. Comparing the images taken at these two states makes it possible to create a map showing the porosity level at each voxel. However, one of the main challenges of this imaging technique is the need for image registration, which is the process of aligning the images taken in the dry and wet states. This procedure is necessary because the wet sample must be aligned with the dry sample to compare the two states accurately.

Dry-wet imaging is a powerful technique in digital rock analysis, but it also presents several challenges in terms of image registration. The significant differences between the dry and wet images, as well as the relatively homogeneous nature of rock images at the texture level, can make it difficult to use feature point matching techniques for alignment. Additionally, the large size of 3D images obtained from micro CT, which can contain billions of voxels and multiple gigabytes of data, makes it resource-intensive to load and manipulate these images. This can hinder the performance of automated registration algorithms and require a significant amount of expert labor. Improving the image registration process for digital rock analysis is therefore important for streamlining the digital rock workflow and reducing the amount of labor required.

In this work, we aim to address the unique challenges associated with image registration in dry-wet imaging for digital rock analysis. Specifically, we recognize that the

[*] Corresponding author
muhammad.sarmad@ntnu.no (M. Sarmad); johan.phan@ntnu.no (J. Phan); leonardo.ruspini@petricore.com (L. Ruspini)
ORCID(s): 0000-0002-8635-9000 (M. Sarmad)

significant textural differences between the dry-wet images, extreme misalignments, and large image sizes can make it difficult to align these images quickly using traditional techniques. To address these challenges, we propose a solution based on a GPU-accelerated implementation of an intensity-based registration algorithm. This approach allows for fast, robust and accurate image registration in a practical setting. Our contributions are summarized as follows:

- Our method accurately performs image registration for dry-wet image pairs of various textures.
- Our algorithm can find a solution in under one minute by utilizing a graphical processing unit (GPU) for optimization.
- The algorithm can handle extreme rotation along the vertical axis of cylindrical samples, as well as inversion along the horizontal axes of the sample.
- Evaluations on various datasets demonstrate our algorithm's performance. We also compare the method to a human expert who takes more than one hour for the same task.

2. Related Work

Common techniques for unimodal images include correlation methods, as described in Pratt (1974). These methods often require additional cleaning steps to ensure the success of cross-correlation. Althof et al. (1997) demonstrated that automatic and fast solutions with reasonable accuracy can be achieved through correlation-based methods. For multimodal images, Viola and Wells III (1997) and Maes et al. (1997) proposed the use of mutual information, which is more robust to such images. Fourier-based methods, which work with the Fourier representation of images, are faster than cross-correlation methods De Castro and Morandi (1987). Jenkinson and Smith (2001) focused on the optimization algorithm itself and tailored a global optimization algorithm specifically for registration. The choice of similarity measure (e.g., correlation vs mutual information) is crucial for the success of registration, and Roche et al. (2000) provided guidance on how to select the correct measure for the best results. While these methods are effective, they can require a number of iterations to reach the final result. In this work, we also utilize optimization-based methods, but leverage the power of graphical processing units (GPUs) through the use of open-source libraries Pytorch Paszke et al. (2017) and AIRLab Sandkühler et al. (2018) for gradient calculation required for optimization.

Deep learning has been a popular choice for solving many registration problems since convolutional neural networks (CNN) have become popular Krizhevsky et al. (2012); LeCun et al. (1990). Learning-based methods have been used for every stage of the registration process. Haskins et al. (2019) propose to learn the similarity metric using a CNN while keeping the classical optimization process. On the other hand, Miao et al. (2016) and Chee and Wu (2018) use a synthetic transform-based data generation method to train a CNN model that predicts the transformation matrix in one shot.Liao et al. (2017) utilize a reinforcement learning-based method to train an agent for robust image registration. These methods mostly use medical image datasets and are impressive as they are fast at inference time. However, they do not work well with our dataset due to the lack of enough distinguishing features in the dry and wet images. There are unsupervised approaches for deformable image registration as well Hu et al. (2018). However, we are limited to rigid transform. Haskins et al. (2020) and Fu et al. (2020) present a through survey on learning based approaches.

Image registration is the first step in many problems where the properties of the rock need to be estimated Knackstedt et al. (2004); Arns et al. (2002); Padhy et al. (2007); Prodanović et al. (2007). Even though the problem of image registration has been solved in the medical image domain by many methods, that is not the case for wet and dry image registration. The seminal work of Latham et al. (2008) uses a correlation-based method coupled with an optimizer to perform 3D dry image to wet image registration which is slow due to the iterative nature of the optimization process. We bring the time from hours to seconds due to using a graphical processing unit (GPU).

Our work is based on the open-source librarry AIRLab Sandkühler et al. (2018). However, other toolboxes are also available, e.g. ITK Yoo et al. (2002), Elastix Klein et al. (2009), and ANTs Avants et al. (2011). However, these toolboxes and libraries do not utilise the GPU transparently and efficiently as AIRlab. Therefore, prototyping in these methods is time-consuming and error-prone, which can be problematic in an industrial setting. For a detailed review of other toolboxes available for the task of registration, we refer to Keszei et al. (2017).

3. Method

In this section, we provide the detailed working of our registration algorithm. The problem overview is shown in Fig. 1. We solve the problem of Dry-Wet Image registration. Consider a dry image Im_{Dry}. This image is a micro CT scan of a rock sample. This sample can then be imbibed with liquid, e.g. brine or mercury, to obtain a wet image Im_{Wet} of the same rock sample. The injection of liquid in this sample changes the visual characteristics of this sample to a large extent under a micro CT. This phenomenon can be seen in the figure.

Fig. 1 shows a perfectly registered dry and wet image pair as the output of the registration process on the right side of the image. A close examination of the registered images reveals that the regions in the Dry-Wet image pair that correspond well are sparse. Secondly, the regions that correspond to each other in both samples reside in different colour spaces. This input setting can lead to a problem for intensity-based methods since they rely on these visual cues to determine a feasible solution through optimization. Therefore, we utilize pre-processing steps before registration

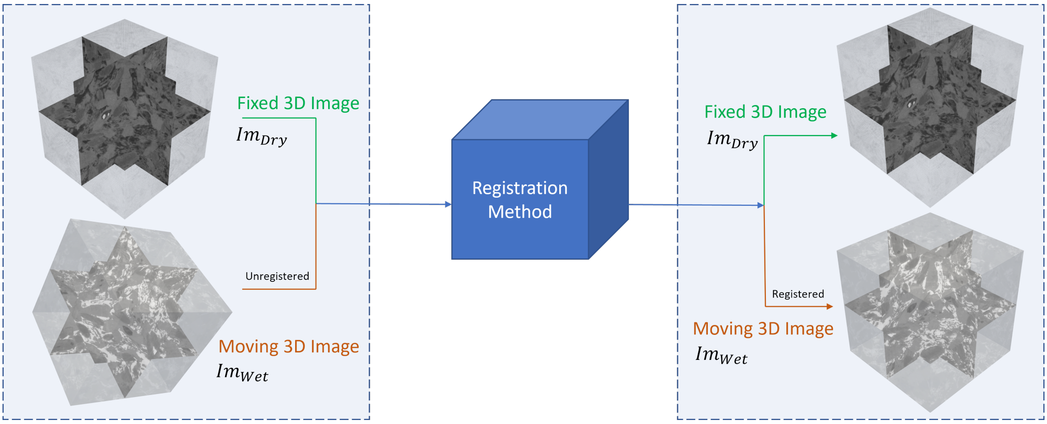


Figure 1: We solve the problem of Dry-Wet rock Image registration. Our Registration Method (Blue box) can register Image pairs efficiently and accurately. In this figure, the Fixed 3D Image is the Dry Image, whereas the Moving 3D Image is the Wet Image. The registration method finds the transformation required to warp the Moving 3D image to register it to the Fixed 3D Image.

to ensure that the optimization-based registration algorithm works robustly. The pre-processing steps are gives as follows:

Color Inversion and Histogram Matching We observe from Fig. 1 that the wet image and dry image correspond with each other. However, the colours seem inverted. Therefore, we invert the colours of the wet image to ensure that it corresponds well with the dry image. We note that a simple inversion of the images is not sufficient. The histogram of the wet-dry image pair must be matched prior to the execution of the registration algorithm. The raw images from the CT scan are provided in a 16-bit unsigned raw format. This format means that the possible values can range from 0 to 65535. Metallic regions with a high atomic number have high pixel values associated with them in the 3D image, whereas the material with a low atomic number has a lower value. In the case of the rock sample of concern, metal objects often shine very bright and distort the colour of the remaining vital regions, such as pores and solids in the rock samples. Therefore to get the best possible match between the dry-wet images, we match the histogram of the wet image to the dry image.

Table 1 shows a thorough comparison between dry and wet images before and after the operation of colour inversion and histogram matching. This comparison has been performed on all datasets used in this work. The details of each dataset will be provided in the experiment section. It can be seen that the colour shift between dry-wet pairs is a colour inversion in all cases. The matching regions are made more prominent after an inversion of colours. We always perform this procedure for the Wet image only. After the inversion, the results of this operation are shown in the column ‘Inverted W’ in Table 1. It can be seen that in the row ‘ST C14’, the ‘Wet Image (W)’ seems washed out after inversion. This is because of the metallic regions (white spots) in the ‘Dry Image (D)’. Therefore, we apply histogram matching to solve this problem. After applying histogram matching, we observed that the ‘Histogram Matched (W)’ image corresponds well with the ‘Dry Image (D)’.

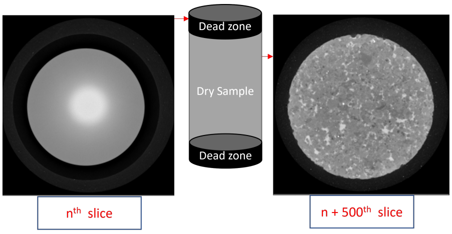
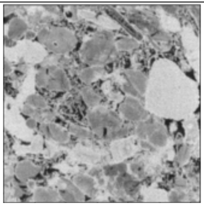
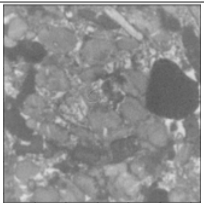
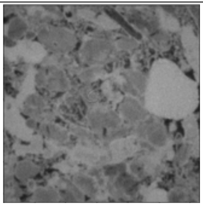
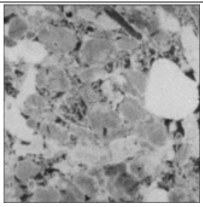
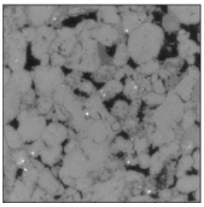
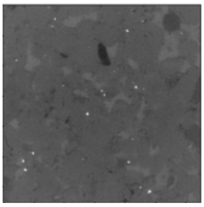
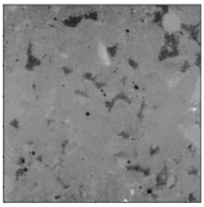
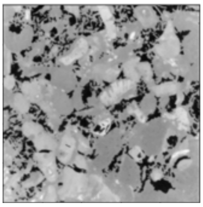
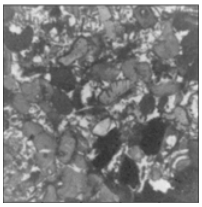
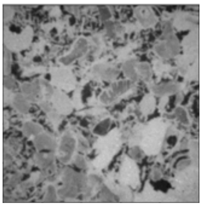
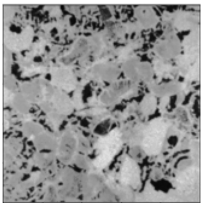


Figure 2: Deadzone: A microCT image of a sample contains dead-zones of variable length due to padding material placed at the end to hold sample in place.

Dead Zones A special consideration for image registration in digital rock analysis is the existence of so-called ‘dead zones’ as shown in Fig. 2. These exist since the cylindrical sample is padded on top and bottom before putting the sample inside a CT scan. This extra padding leads to blank regions included in the final image. These regions must be considered to ensure that they do not negatively affect the automatic registration process.

Extreme Transformation Another aspect that makes our problem different from typical 3D image registration is the characteristic of the transformation. The transformation is a similarity transformation in nature. This setting means that there are nine elements in the transformation. These include translation, rotation and scaling for their respective axes. Of particular importance are the rotations as shown in Fig. 3. Along z-axis, this rotation can be ±180 ° Since the wet sample can be placed at any angle. The second important

Table 1
Image Samples from each of the dataset used have been shown in this table. The Dry Image (D) and Wet Image (W) do not correspond with eachother. The Colors of Wet Image are inverted and shown in Column 'Inverted W'. The Inverted W image's histogram is matched to Dry Image (D) to create 'Histogram Matched W' Image.

Image / Dataset	Dry Image (D)	Wet Image (W)	Inverted W	Histogram Matched W
Wang et al. (2021)				
ST C14				
Spurin et al. (2021)				

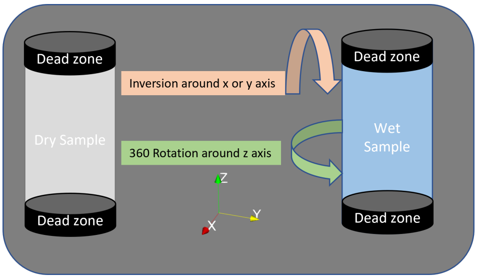


Figure 3: Extreme Rotation: A microCT image of a wet and dry sample are shown. Note that The wet sample can have extreme rotation of ±180° along z-axis and a 180° inversion around the x or y axes.

aspect is the possible inversion of the sample along the x or y axes. This means that a 180° inversion of the sample is also possible.

3.1. Registration Algorithm

After the pre-processing steps mentioned before, the samples are ready for registration. We use the normalized cross correlation between the dry and wet image to find the perfect matching. The details are given in algorithm 1. We utilize the similarity transformation along with the normalized cross-correlation similarity objective which is given as follows:

$$\mathcal{L}_{\text{NCC}} := \frac{\sum Im_{Dry} \cdot (Im_{Wet} \circ f) - \sum \mathrm{E}(Im_{Dry})\mathrm{E}(Im_{Wet} \circ f)}{|\mathcal{Y}| \cdot \sum \mathrm{Var}(Im_{Dry})\mathrm{Var}(Im_{Wet} \circ f)} \quad (1)$$

The sum in the above mentioned equation is over over the image domain $\mathcal{Y}$, E is the expectation value (or mean) and Var is the variance of the respective image. The registration algorithm aims to find the best match between the two 3D images. For the scope of this work, the dry and we image can contain translation of about 40 pixels in the x and y axes. Only 5 pixels in the z axis. A uniform scaling factor of 0.95 to 1.05 can also be present along any of the three axes. Rotation along x-axis and y axis can be about ± 5°. If the samples are inverted about x-axis or y-axis, the rotation can still be about ±5°. The rotation along z axis can be a ±180°. We design our method to be suitable for the worst case scenario. It assumes two worst case scenarios i.e. it contains a rotation of ±180° about the z axis and the second is that the sample to be registered is inverted about the x or y axis as shown in Fig. 3.

Algorithm 1: How to write algorithms

```
$L_{best} = \infty$ and $\delta_\theta = 20°$ and $S_{best} = 0$ ;
$Iter_{init} = 25$ , $Iter_{final} = 1000$ ;
$Lr_{init} = 0.01$ , $Lr_{final} = 0.0005$;
Fetch $Im_{Dry}$ and $\tilde{Im}_{Wet}$ ;
$L, S_j, F = LocalRegistration(F_{Invert} = False)$;
$\tilde{L}, \tilde{S}_j, F = LocalRegistration(F_{Invert} = True)$;
if $L \leqq \tilde{L}$ then
    $L_{best} = L$;
    $S_{best} = S_j$ ;
else
    $L_{best} = \tilde{L}$;
    $S_{best} = \tilde{S}_j$ ;
end
if $F$ is False then
    Rotate $\tilde{Im}_{wet}$ by $\delta_\theta \times S_{best}$ along z-axis ;
else
    Rotate $\tilde{Im}_{wet}$ by $\delta_\theta \times S_{best}$ along z-axis and
      invert along x-axis ;
end
Define $Loss_{NCC}$ and Optimizer($Lr_{final}$);
Start Similarity Transformation
  Registration($Iter_{final}$,$Im_{Dry}$,$\tilde{Im}_{Wet}$);
Function $L_{best}$, $S_{best}$, $F_{best}$ =
  LocalRegistration($F_{Invert}$):
    $L_{current} = 0$
    $S_j = 0$
    for $S_j \leqq \frac{360}{\delta_\theta}$ do
        if $F_{Invert}$ is False then
            Rotate $\tilde{Im}_{wet}$ by $\delta_\theta \times S_j$ along z-axis ;
        else
            Rotate $\tilde{Im}_{wet}$ by $\delta_\theta \times S_j$ along z-axis
              with 180° inversion along x-axis ;
        end
        $Loss_{NCC}$ and Optimizer($Lr_{init}$);
        Rigid Transformation
          Registration($Iter_{init}$,$Im_{Dry}$,$\tilde{Im}_{Wet}$);
        Update $L_{current}$ ;
        if $L_{current}$ is less than $L_{best}$ then
            if $F_{Invert}$ is False then
                $L_{best} \longleftarrow L_{current}$ ;
                $S_{best} \longleftarrow S_j$;
                $F_{best} \longleftarrow$ False;
            else
                $L_{best} \longleftarrow L_{current}$ ;
                $S_{best} \longleftarrow S_j$;
                $F_{best} \longleftarrow$ True;
            end
        else
            pass;
        end
        $S_j = S_j + 1$
    end
```

Sector Search For Rotation about Vertical Axis We first describe how to deal with the first scenario i.e. rotation about the z axis. . This scenario can occur if the operator places the sample after rotating it in the CT chamber for obtaining the Wet image. The rotation about the z-axis can be determined by the NCC loss based optimization. However, we notice that if the rotation in the z-axis is more than ± 10° then the algorithm starts to fail. Therefore we propose a localized registration check, where we divide the image to be registered into sectors. The chunk of each sector corresponds to a angle δ_θ. We set this angle empirically to a value that does not break the algorithm. We then check all sectors until a 360° rotation is complete about the z axis. We also log the scores of matching of each sector using the NCC loss. At the end of this process, we get a best sector. We use the best sector ID to rotate the image to be registered to the best sector for a more fine registration for $Iter_{final}$ iteration. Note that since this is an initial step therefore we do not perform a lot of iteration and only use $Iter_{init}$. Where as in the final iteration step we use $Iter_{final}$ such that $Iter_{init}$ are significantly less than $Iter_{final}$. We call this step a 'sector search'.

Sector Search Inversion about Horizontal Axis The second extreme scenario is that the image to be registered is inverted along the horizontal axis (i.e. x or y axis). This scenario can occur if the operator places the sample upside down in the CT chamber for obtaining the Wet image. To handle this we first perform sector search without inverting the image to be registered and log the best loss value achieved. Later we perform a sector search after inverting the image to be registered and again log the best loss. Finally we compare the loss value achieved by both the inverted and original sector search to determine the correct sector and if inversion is needed or not.

Once the registered image is catered for extreme rotations. We perform a final set of iterations $Iter_{final}$ to find all elements of the similarity transformation namely 3 translation, 3 rotations and 3 scaling factors. The detailed steps of our method are given in the algorithm 1. Please note that L_{best} is the best loss value achieved, δ_θ is the angle in degree that , S_{best} is the best sector, $Iter_{init}$ are the initial iterations used for sector search and inversion, Lr is the learning rate used for optimization algorithm.

4. Experiments and Results

4.1. Open Source Libraries

We utilize AIRLab Sandkühler et al. (2018) for the registration algorithm. It allows for fast prototyping and utilization of GPUs since it utilizes Pytorch Paszke et al. (2017). Our program's execution is finished in under a minute due to the utilization of GPU. We also use Pytorch Paszke et al. (2017) and MONAI Consortium (2020) libraries for creating transformations in the dataloader on the moving Image for validation of our approach.

Table 2
DataSet: Dimensions (in Pixels) and Properties Of the Dataset used

Data-set	x	y	z
Wang et al. (2021)	630	630	1087
ST C14	1300	1300	2500
Spurin et al. (2021)	445	445	445

4.2. Dataset

This work considers various types of dry and wet 3D images of rocks. We use two open-source datasets freely available from Wang et al. (2021) and Spurin et al. (2021). In addition, we utilize an in-house dataset of rock samples which we call the ST C14 dataset. The dimensions of images in these datasets are shown in the Table. 2. All samples are imbibed with fluids such as water or brine to obtain the wet image. For each sample, we have two perfectly registered dry and wet 3D image pairs of the same dimension.

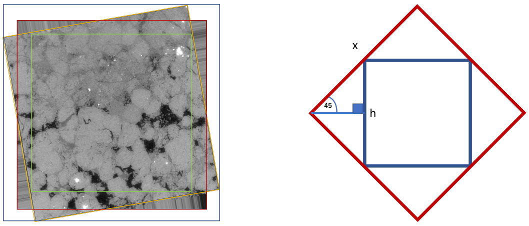
Table Table. 1 shows the detailed pre-processing steps for images. As shown in the Table, we only process the Wet Image by first inverting the colour of the Wet Image. However, this is not enough, and we additionally perform Histogram matching by matching the histogram of the Wet Image with the Dry Image. The corresponding image is shown in the column 'Histogram Matched W'.

4.3. Dataset Generation for Evaluation

To evaluate the effectiveness of our algorithm, we need to create many possible combinations of three translations, three rotations and one scaling. Please note that the three rotations about x and y can include a possible inversion of the image. At the same time, the rotation about the z-axis can be a 360-degree rotation. We use the three perfectly registered samples in our dataset and transform the wet image randomly on the fly using the a custom Pytorch data loader Paszke et al. (2017). This provides us with a large number of transformations to evaluate our dataset. We limit the randomly generated data transformations to 200 for each of the three datasets.

Padding Requirement: Padding the moving image is essential before transforming the image with the data loader since every transformation causes some amount of image information loss near the image border. The padding requirement for synthetic translation and scaling is linearly related to each parameter's value. However, a unique challenge from a rotation can be observed in Fig. 4a. Consider an original image as shown in red lines. Upon rotation, the image becomes an image shown in yellow lines. However, the image we can utilize after rotation is in green lines due to information loss due to rotation. This image is smaller than the intended size of the image. In order to get an image of the size shown in red lines. We need to start with an image of the size shown in blue lines.

Consider the Fig. 4b, where the image with red lines of size x has been rotated by 45; then we can extract the image with blue lines of size h from this square. It can be seen that when we rotate an image by a certain angle, some parts of the image are permanently lost due to the rotation operation. Therefore, we must calculate the padding value p needed to add to the original image of size h such that when starting with an image of size $x = h + p$, we obtain a rotated image which can be cropped by p to obtain a final image of size h such that no loss of information occurs. The calculation of this pad factor p will be maximum when θ is 45° as shown in Fig. 4b. It can be seen that $h = x * \sin(45°)$. Then since $x = h + p$, the pad factor p can be given as follows:



(a) Why padding is needed?

(b) Padding factor

Figure 4: Fig.4a demonstrates that loss of image information occurs due to rotation. Fig.4b shows how to derive a padding factor p such that starting from image of size $x = h + p$ we obtain h after cropping with p without loss of any information

$$p = h \times (\frac{1}{\sin(45°)} - 1) \tag{2}$$

To scale this value to be dynamic from 0° to 45°. We use a scale factor based on the angle θ as follows:

$$p = h \times (\frac{1}{\sin(45°)} - 1) \times \sin(2\theta) \tag{3}$$

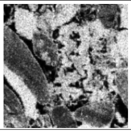
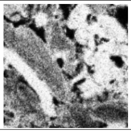
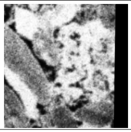
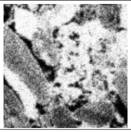
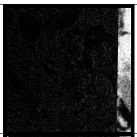
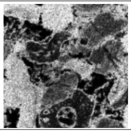
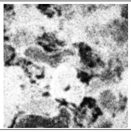
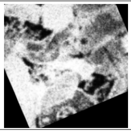
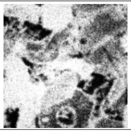
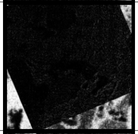
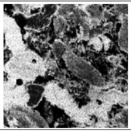
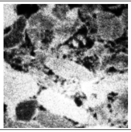
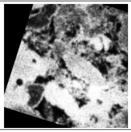
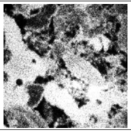
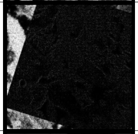
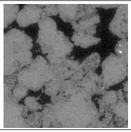
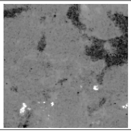
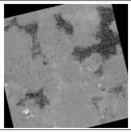
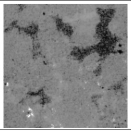
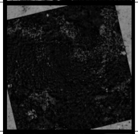
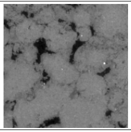
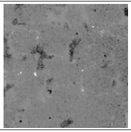
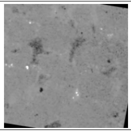
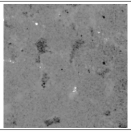
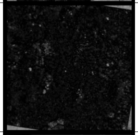
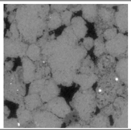
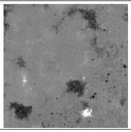
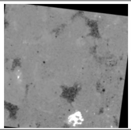
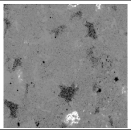
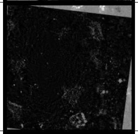
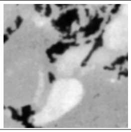
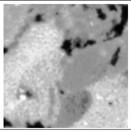
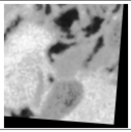
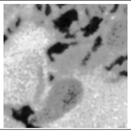
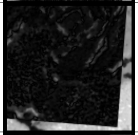
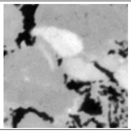
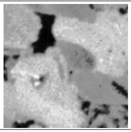
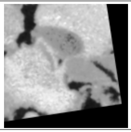
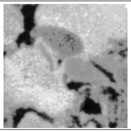
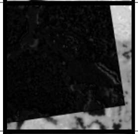
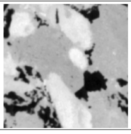
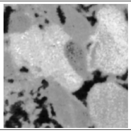
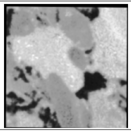
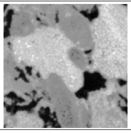
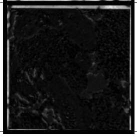
4.4. Metric

We use the root mean square error (RMSE) metric between the ground truth and the predicted transformations to report our algorithm's performance in registering various dry and wet rock images on the dataset. The formula of RMSE is $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$.

4.5. Quantitative Results

We used the synthetic data described in the previous section to generate various transformations, including extreme transformations such as rotation about the z-axis and complete inversion about the x-axis and y-axis. We compare each transformation prediction's root mean square error (RMSE) and ground truth transformation. We use 200 random transformations for each rock dataset to find the predicted transformation parameters for the registration task. The results in Table 4 demonstrate the RMSE value for each transformation quantity. It can be noted that ST C14 is a relatively tricky sample. It is not surprising, as it can be seen from Table 1 that ST C14 contains fewer correspondences and therefore is naturally a complex sample to register. From

Table 3
The performance of our method on image samples from each dataset is shown in the table. The algorithm calculates the transformation needed to warp the 'Moving Image' and align it with the 'Fixed Image'. The table compares the warped moving image with the 'Ground Truth (GT) Moving Image' and displays the quantitative parameters of the transformation: translation in x, y, and z in pixels (Trans X, Trans Y, Trans Z), rotation in x, y, and z respectively in degrees (Angle X, Angle Y, Angle Z), and scaling in x, y, and z respectively (Scale X, Scale Y, Scale Z).

Image Dataset	Fixed Image	Moving Image	Warped Moving Image	GT Moving Image	Difference Image	Parameter	Prediction	Ground Truth
Wang et al. (2021)						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-27.73 1.04 -16.20 -4.83 177.09 80.91 1.05 1.05 1.04	-28.00 1.00 -16.00 -4.82 177.06 80.88 1.04 1.04 1.04
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	10.22 -27.92 -16.92 -1.93 1.15 -57.06 1.03 1.03 1.03	10.00 -28.00 -17.00 -1.93 1.13 -57.00 1.02 1.02 1.02
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	27.89 -12.35 3.15 -4.74 -3.36 162.23 1.02 1.02 1.03	28.00 -12.0 3.00 -4.74 -3.36 162.23 1.02 1.02 1.02
ST C14						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-6.10 3.80 -9.88 -3.34 0.16 93.23 1.08 1.01 1.00	-5.00 3.00 -10.00 -4.38 -0.78 94.17 0.97 0.97 0.97
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-4.49 -0.50 2.06 -0.69 179.69 -90.27 1.03 1.07 1.14	-3.00 -1.00 3.00 -1.59 178.57 -87.77 1.00 1.00 1.00
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-7.92435 10.66 -2.86 3.33 0.34 -7.11 1.00 1.00 1.03	-10.0 11.00 -2.00 3.94 0.85 -7.87 0.98 0.98 0.98
Spurin et al. (2021)						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-10.71 -8.17 9.45 2.70 184.28 -35.39 1.06 1.02 1.02	-10.00 -9.00 10.00 2.23 184.14 -35.77 0.99 0.99 0.99
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	-13.29754 -16.31 -5.65 -1.14 -1.25 128.61 1.02 0.98 0.99	-13.0 -16.00 -6.00 -1.76 -1.75 128.53 0.98 0.98 0.98
						Trans X : Trans Y : Trans Z : Angle X : Angle Y : Angle Z : Scale X : Scale Y : Scale Z :	0.70478 1.92 10.57 -2.82 4.28 -20.30 0.97 0.97 0.97	1.0 2.00 11.00 -2.83 4.44 -20.25 0.96 0.96 0.96

the table, it can be observed that our algorithm performs robustly. This method is robust enough to be deployed in the industry to solve the registration of wet and dry images.

4.6. Qualitative Results

Table 3 shows the qualitative results of our method. It should be noted that both qualitative and quantitative values

Table 4
Root mean square error (RMSE) of the Registration Algorithm

Error Dataset	Angle x	Angle y	Anlge z	Translation x	Translation y	Translation z	*Scale x*	*Scale y*	*Scale z*
Wang et al. (2021)	0.989	0.473	0.432	0.876	0.415	0.670	0.016	0.030	0.046
ST C14	1.411	1.201	1.297	1.501	1.136	1.826	0.102	0.075	0.051
Spurin et al. (2021)	0.345	0.337	0.466	0.429	0.218	0.179	0.039	0.017	0.014

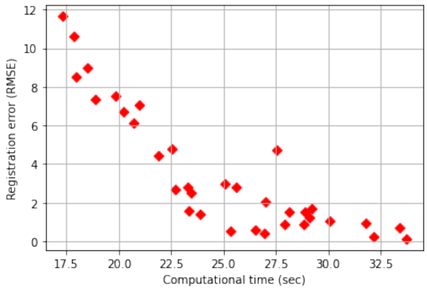


Figure 5: The trade-off between Computational time (sec) vs Registration Error (RMSE) is shown.

for each example are given. This gives a good idea of real-term performance. e.g. note that for ST C14, the error seems more significant for specific transformation parameters compared to the rest of the samples from other data set. However, visual inspection reveals that the results are very accurate in reality.

4.7. Timing and Efficiency Analysis

Computational Time We compare the effect of computational time (linearly related to optimization steps) algorithm in the form of a scatter plot. From the scatter plot in Fig. 5, we can observe that increasing the computational time generally leads to a better registration error up to a certain point, whereas the computation times needed also increases. This experiment was performed for the dataset provided by Wang et al. (2021) et al.. For each data point in the figure, the average RMSE is calculated for all transformation parameters (translation, rotations and scalings) and multiple runs (50 runs). Using multiple runs ensures that noisy data points are removed.

Human Expert Time Comparison: We compare the speed of our method with those of a human expert in terms of time. The human performance data was only available for ST C14 dataset. Our method takes under 1 minute for most examples. On the other hand, the human expert uses a combination of manual labour and expert tools to achieve the task in 120 minutes. It is also pertinent to note that the human expert performed this transformation for just one case. It was relatively easy, as no inversion of the sample was present in the unregistered image.

5. Conclusion

In this work, we have presented an algorithm for fast and accurate registration of 3D dry and wet digital rock images. We have used a cross-correlation-based optimization process for this task. The process of 3D registration based on optimization is considered slow and usually shunned in many cases due to the advent of one-shot methods. However, one-shot methods either do not provide a robust enough solution or sometimes do not provide a solution at all. Therefore, this work demonstrated that the optimization process could be sped up considerably using GPU. In addition, we provide an algorithm that is robust enough to ensure the high accuracy of the registration solution. We demonstrate the effectiveness of our results under plausible industrial scenarios, such as extreme rotation along the vertical axis and inversion of the sample.

6. Acknowledgment

This work was partially supported by the Norwegian Research Council (grant number 296093) and the members of the SmartRocks joint industry project (ENI AS, Repsol AS, and Chevron Corporation).

References

Aarnes, J.E., Kippe, V., Lie, K.A., Rustad, A.B., 2007. Modelling of multiscale structures in flow simulations for petroleum reservoirs, in: Geometric Modelling, Numerical Simulation, and Optimization. Springer, pp. 307–360.

Althof, R.J., Wind, M.G., Dobbins, J.T., 1997. A rapid and automatic image registration algorithm with subpixel accuracy. IEEE transactions on medical imaging 16, 308–316.

Arns, C.H., Knackstedt, M.A., Pinczewski, W.V., Garboczi, E.J., 2002. Computation of linear elastic properties from microtomographic images: Methodology and agreement between theory and experiment. Geophysics 67, 1396–1405.

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage 54, 2033–2044.

Bhattad, P., Young, B., Berg, C.F., Rustad, A.B., Lopez, O., 2014. X-ray micro-ct as-sisted drainage rock typing for characterization of flow behaviour of laminated sandstone reservoirs .

Chee, E., Wu, Z., 2018. Airnet: Self-supervised affine registration for 3d medical images using neural networks. arXiv preprint arXiv:1810.02583 .

Consortium, T.M., 2020. Project monai. URL: https://doi.org/10.5281/zenodo.4323059, doi:10.5281/zenodo.4323059.

De Castro, E., Morandi, C., 1987. Registration of translated and rotated images using finite fourier transforms. IEEE Transactions on pattern analysis and machine intelligence , 700–703.

Feali, M., Pinczewski, W.V., Cinar, Y., Arns, C.H., Arns, J.Y., Turner, M., Senden, T., Francois, N., Knackstedt, M., et al., 2012. Qualitative and quantitative analyses of the three-phase distribution of oil, water, and gas in bentheimer sandstone by use of micro-ct imaging. SPE Reservoir Evaluation & Engineering 15, 706–711.

Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X., 2020. Deep learning in medical image registration: a review. Physics in Medicine & Biology 65, 20TR01.

Haskins, G., Kruecker, J., Kruger, U., Xu, S., Pinto, P.A., Wood, B.J., Yan, P., 2019. Learning deep similarity metric for 3d mr–trus image registration. International journal of computer assisted radiology and surgery 14, 417–425.

Haskins, G., Kruger, U., Yan, P., 2020. Deep learning in medical image registration: a survey. Machine Vision and Applications 31, 1–18.

Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Noble, J.A., Barratt, D.C., Vercauteren, T., 2018. Label-driven weakly-supervised learning for multimodal deformable image registration, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 1070–1074.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Medical image analysis 5, 143–156.

Keszei, A.P., Berkels, B., Deserno, T.M., 2017. Survey of non-rigid registration tools in medicine. Journal of digital imaging 30, 102–116.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2009. Elastix: a toolbox for intensity-based medical image registration. IEEE transactions on medical imaging 29, 196–205.

Knackstedt, M., Arns, C., Limaye, A., Sakellariou, A., Senden, T., Sheppard, A., Sok, R., Pinczewski, W.V., Bunn, G., 2004. Digital core laboratory: Properties of reservoir core derived from 3d images, in: SPE Asia Pacific conference on integrated modelling for asset management, OnePetro.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Curran Associates Inc., USA. pp. 1097–1105. URL: http://dl.acm.org/citation.cfm?id=2999134.2999257.

Latham, S., Varslot, T., Sheppard, A., et al., 2008. Image registration: enhancing and calibrating x-ray micro-ct imaging. Proc. of the Soc. Core Analysts, Abu Dhabi, UAE , 1–12.

LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E., Jackel, L.D., 1990. Handwritten digit recognition with a back-propagation network, in: Touretzky, D.S. (Ed.), Advances in Neural Information Processing Systems 2. Morgan-Kaufmann, pp. 396–404. URL: http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network.pdf.

Liao, R., Miao, S., de Tournemire, P., Grbic, S., Kamen, A., Mansi, T., Comaniciu, D., 2017. An artificial agent for robust image registration, in: Proceedings of the AAAI conference on artificial intelligence.

Long, H., Nardi, C., Idowu, N., Carnerup, A., Øren, P., Knackstedt, M., Varslot, T., Sok, R., Lithicon, A., 2013. Multi-scale imaging and modeling workflow to capture and characterize microporosity in sandstone 13.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE transactions on Medical Imaging 16, 187–198.

Miao, S., Wang, Z.J., Liao, R., 2016. A cnn regression approach for real-time 2d/3d registration. IEEE transactions on medical imaging 35, 1352–1363.

Padhy, G., Lemaire, C., Amirtharaj, E., Ioannidis, M., 2007. Pore size distribution in multiscale porous media as revealed by ddif–nmr, mercury porosimetry and statistical image analysis. Colloids and Surfaces A: Physicochemical and Engineering Aspects 300, 222–234.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch .

Pratt, W.K., 1974. Correlation techniques of image registration. IEEE transactions on Aerospace and Electronic Systems , 353–358.

Prodanović, M., Lindquist, W., Seright, R., 2007. 3d image-based characterization of fluid displacement in a berea core. Advances in Water Resources 30, 214–226.

Roche, A., Malandain, G., Ayache, N., 2000. Unifying maximum likelihood approaches in medical image registration. International Journal of Imaging Systems and Technology 11, 71–80.

Ruspini, L., Oeren, P.E., Berg, S., Masalmeh, S., Bultreys, T., Taberner, C., Sorop, T., Marcelis, F., Appel, M., Freeman, J., Wilson, O., 2021. Multiscale digital rock analysis for complex rocks. Transport in Porous Media 139, 1–25. doi:10.1007/s11242-021-01667-2.

Sandkühler, R., Jud, C., Andermatt, S., Cattin, P.C., 2018. Airlab: autograd image registration laboratory. arXiv preprint arXiv:1806.09907 .

Spurin, C., Krevor, S., Blunt, M., Bultreys, T., 2021. Decane and brine injected into estaillades carbonate - steady-state experiments. http://www.digitalrocksportal.org/projects/344. doi:10.17612/cd7a-y955.

Viola, P., Wells III, W.M., 1997. Alignment by maximization of mutual information. International journal of computer vision 24, 137–154.

Wang, S., Bultreys, T., Van Offenwert, S., Ruspini, L., 2021. Dataset for unsteady-state capillary drainage experiment on estaillades carbonate. http://www.digitalrocksportal.org/projects/363. doi:10.17612/6rtt-5w16.

Yoo, T.S., Ackerman, M.J., Lorensen, W.E., Schroeder, W., Chalana, V., Aylward, S., Metaxas, D., Whitaker, R., 2002. Engineering and algorithm design for an image processing api: a technical report on itk-the insight toolkit, in: Medicine Meets Virtual Reality 02/10. IOS press, pp. 586–592.

Appendix D

Paper D: Core-Scale Rock Typing using Convolutional Neural Networks For Reservoir Characterization in the Petroleum Industry

CORE-SCALE ROCK TYPING USING CONVOLUTIONAL NEURAL NETWORKS FOR RESERVOIR CHARACTERIZATION IN THE PETROLEUM INDUSTRY

Muhammad Sarmad[1]

Johan Phan[1]

Dr. Leonardo Ruspini[2]

Assoc. Prof. Dr. Gabriel Kiss[1]

Prof. Dr. Frank Lindseth[1]

[1] Norwegian University of Science and Technology, **Norway**
[2] Petricore, **Norway**

ABSTRACT

Rock typing is an essential tool for reservoir characterization and management in the petroleum industry. It is the process of grouping portions of a rock sample based on their physical and chemical properties. This process is currently done by experts in the industry, which consumes valuable industry resources. Precise and efficient rock typing can build accurate geological models, optimize exploration and production strategies, and reduce exploration and production risks. This work proposes a deep learning method to identify and classify rocks based on their pore geometry, mineralogy, and other characteristics. The proposed technique segments a micro-CT image into different rock types using a neural network for automated rock typing. We suggest using a UNet architecture for the neural network for this task. The network has been trained in a supervised manner on expert-labelled images. The method's performance has been evaluated using qualitative and quantitative metrics. The neural network takes less than 200 milliseconds to provide the rock types, which is much faster than a human expert. We perform an explainability analysis of the neural network using class activation heatmaps approach to get insight into the learned weights. Rock typing using deep learning can improve the petroleum industry's workflow.

Keywords: rock typing, digital rock analysis, deep learning, segmentation.

INTRODUCTION

Rock analysis is critical in the oil and gas industry, yielding vital information about reservoir properties for efficient production and effective reservoir management. The traditional method for rock analysis is called conventional core analysis. This analysis involves using actual rock samples and various laboratory equipment to determine physical and chemical properties. However, this process is costly and time-consuming since experiments need to be conducted on actual samples. This process also sometimes leads to the destruction of the sample because of the analysis [1]. With the advent of advanced imaging techniques, such as micro-computed tomography (micro-CT) and scanning electron microscopy (SEM), a new approach to rock analysis has emerged. This approach, known as digital rock analysis, leverages high-resolution images of core samples obtained digitally, preventing the need for physical testing and consequently preventing sample destruction. In addition to preserving the sample's integrity, digital rock analysis offers the advantage of performing multiple analyses on the same sample [2].

Digital rock analysis allows for the application of various numerical simulations and algorithms to scrutinize the microstructures at the pore scale level. The transition to the digital domain is particularly advantageous as it facilitates using data-driven and machine-learning methods for analyzing rock samples and determining significant properties. The properties deduced from digital rock analysis include but are not limited to, fluid flow under various transport scenarios through the rock's pore spaces, permeability, and porosity. Through this process, we can ascertain the rock's behaviour under different conditions and predict its response to various operational strategies, aiding in optimizing production and reservoir management.

Rock typing groups rock samples from reservoirs into categories based on their physical and chemical properties. Rock typing is an essential tool for predicting reservoir behaviour in the petroleum industry. Rock typing typically involves analyzing core samples taken from the subsurface reservoir and characterizing the rock based on various parameters such as mineralogy, texture, porosity, permeability, and fluid saturation. By grouping rocks with similar properties, geoscientists can create geological models that accurately represent the subsurface reservoir and help to predict its behaviour.

Utilizing Digital rock analysis for rock typing can provide valuable insights into the microstructure and properties of rocks. It can make the process of rock characterization more efficient by allowing us to use data-driven and machine-learning methods to perform rock typing efficiently. Using a semantic segmentation model, we utilize convolutional neural networks (CNNs) to segment different regions of a rock sample into rock types. This work uses images of rock samples obtained through micro-CT scanning to train a CNN model to identify and classify different rock types. The model is trained on an annotated image dataset, where each pixel is labelled with the corresponding rock type.

Figure 1 shows how rock typing builds accurate geological models. Once classification is performed, we obtain the rock types. Each rock type can be used to calculate and propagate various properties, e.g., porosity permeability etc., for each type. This information can then be combined in an upscaling step to obtain the properties of the complete sample.

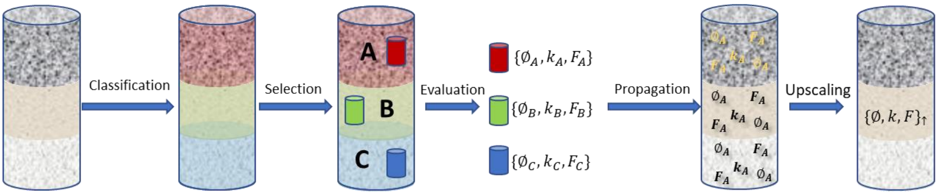

Classification
Selection
A
B
C
Evaluation
$\{\emptyset_A, k_A, F_A\}$
$\{\emptyset_B, k_B, F_B\}$
$\{\emptyset_C, k_C, F_C\}$
Propagation
$\emptyset_A$ F_A k_A F_A $\emptyset_A$
Upscaling
$\{\emptyset, k, F\}_\uparrow$

Figure 1 : Rock typing and Upscaling Workflow: There are multiple steps involved from classification to upscaling.

In this work, we are dealing with the first step in the figure, i.e. Classification of the rock sample in various rock types. One of the most common approaches is segmenting the rock types based on porosity, a key parameter in rock typing. Porosity is the percentage of the total rock volume that is occupied by pores, and it can distinguish between different rock types. We train a neural network to identify and segment various porosity regions within the rock sample and then use this information to classify the rock type. Once the neural network is trained, we segment and classify new unseen rock images from the test set. In this work, we use an expert annotated dataset of rock images to train a neural network to identify rock types. Neural networks are black boxes; therefore, to explain the network decision, we use the explainability technique to get insight into the decision-making process of the neural networks.

Our contributions in this work are as follows:

1. We propose a fast and efficient deep learning-based method for rock typing.
2. A new expert annotated dataset is used to train our model in a supervised manner.
3. We provide insight into the network's decisions using class activation heatmaps.

RELATED WORKS

Convolutional neural networks (CNNs) have significantly developed over the past few decades. The first CNN architecture, LeNet-5, was proposed by Yann LeCun in 1998 and was designed for handwritten digit recognition [3]. However, it was only in the ImageNet challenge in 2012 that CNNs gained widespread attention. AlexNet achieved state-of-the-art results on the ImageNet dataset [4]. This breakthrough sparked a wave of research in CNNs, developing many new architectures such as VGGNet, GoogLeNet (Inception), and ResNet. VGGNet uses tiny convolutional filters to achieve high accuracy on ImageNet [5]. GoogLeNet (Inception), introduced the concept of "inception modules" for efficient computation [6]. ResNet allowed for very deep neural networks using residual connections [7].

There has been much research on using CNNs for semantic segmentation, which involves labelling each pixel in an image with its corresponding class. FCN uses a fully convolutional network to produce dense predictions [8]. SegNet introduced a novel pooling method to handle upsampling [9]. U-Net was specifically designed for biomedical image segmentation tasks and has since been widely used in other domains [10]. U-Net uses an encoder-decoder architecture with skip connections to produce highly accurate segmentations even with limited training data. Overall, the development of CNNs and their applications in semantic segmentation has led to significant improvements in many areas of computer vision. We also use U-Net since we segment each rock type in an image.

Rock typing is a fundamental aspect of reservoir characterization and management in the petroleum industry. Various approaches have been developed to identify and classify different rock types based on their physical and chemical properties. Most works use conventional techniques while recently deep learning is also becoming popular to solve this problem. [11] used regional Minkowski measures and a multivariate Gaussian mixture model to classify rock types. [12] proposed an image-based rock typing method using local homogeneity filtering and the Chan-Vese model to segment binary images into different rock types. [13] assessed the impact of diffusional coupling on Nuclear Magnetic Resonance (NMR) measurements of saturated laminated sandstone at the layer scale to evaluate the feasibility of NMR rock-typing approaches. On the other hand, [14] introduced fast numerical techniques based on the Minkowski functionals to derive fields of regional Minkowski measures over large regional support for large 3D data sets as generated from x-ray tomography techniques. They demonstrated the application of these

3D feature fields to microstructure classification for a set of heterogeneous microstructures using a multivariate Gaussian mixture model and thin-bedded sandstone. Finally, [15] explored the conventional interpretation of NMR measurements on fluid-saturated reservoir rocks, showing that the T2 and pore size distributions are not directly related in many multi-scale porosity systems due to diffusion coupling between different pores. Some previous works also deal with rock typing using deep learning, but they classify the rock type in an image patch instead of segmenting the full image [16]. Therefore, our method can obtain better boundaries between two different rock types.

MATERIALS & METHOD

Our research employs a deep learning method to automatically detect laminar segments within rock types using 2D slices of 3D micro-CT rock sample images. The labelled images from this process inform the determination of segment-scale properties, which can later be upscaled to the whole sample scale.

A. Model Architecture

As depicted in Figure 2, we utilize the U-Net architecture in our deep learning-based rock typing method. This network processes each input image (Height: 192, Width: 64 pixels), potentially containing multiple rock types, to produce a similarly sized output segmentation image. The U-Net architecture, favored for image segmentation tasks, features an encoder-decoder structure. The encoder, comprised of convolutional and max-pooling layers, extracts and condenses input image features. The decoder reconstructs the segmented image from these condensed features through up sampling and additional convolutional layers. Skip connections between the encoder and decoder preserve spatial information throughout the process. The outcome is a segmentation map that assigns a rock type to each pixel, providing an efficient, scalable methodology for rock typing and enhancing our understanding of various geological formations.

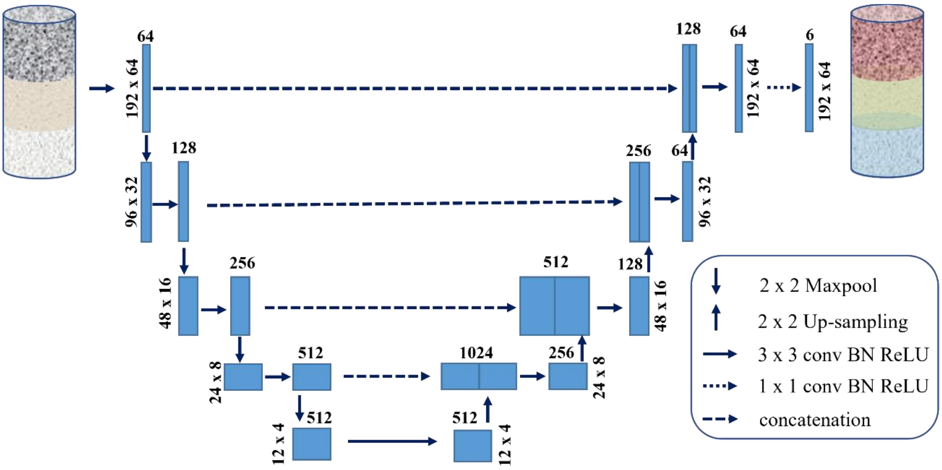


Figure 2 Neural Network architecture: We use the U-Net architecture, the input image is of size 192 x 64 and the output image is of the same size

B. *Binary Cross-Entropy Loss:*

The binary cross-entropy loss is used to train the U-Net model for binary image segmentation tasks. For a given pixel, the binary cross-entropy loss compares the predicted probability of the pixel belonging to the target class (rock type) to the true probability of the pixel belonging to the target class. The loss function is defined as:

$$L(y, \hat{y}) = x = -\frac{1}{\mathrm{N}}\sum_{i=1}^{N}[y_i log(\widehat{y_i}) + (1 - y_i)log(1 - \widehat{y_i})]$$

Where y is the ground truth segmentation map, $\hat{y}$ is the predicted segmentation map, and N is the total number of pixels in the image. The loss function penalizes the model for incorrect predictions and encourages it to produce accurate segmentation maps.

C. *Explainability analysis using Grad-CAM*

Due to their opacity, deep learning models, such as the UNet used herein, are frequently labelled as "black box" models. To illuminate the decision-making process within our UNet model, we employed Gradient-weighted Class Activation Mapping (Grad-CAM), a technique for interpreting deep neural network predictions [17].

Grad-CAM illuminates the regions of an input image that contribute significantly to a specific prediction, offering visual insight into model decisions. The technique utilizes the gradients of a target segmentation class, flowing from the final layers of the neural network, to generate a coarse heatmap highlighting the crucial regions in the input image.

Our work applied Grad-CAM to the UNet model's output to determine the input image regions crucial for predicting distinct rock types. Precisely, the gradient of the predicted class score relative to the feature maps of the penultimate convolutional layer in our UNet model's encoder was calculated. These gradients were then deployed to weigh the feature maps, which were subsequently aggregated to produce a class activation map. The resulting map emphasizes the input image regions most vital for predicting the associated rock type, allowing for a more transparent interpretation of the model's decision-making process.

D. *Data*

1) *Sample Collection and Labelling*

We use two samples, Rock Sample 1 (dimensions: 1300 x 7880) and Rock Sample 2 (dimensions: 1405 x 13940), for training and evaluating our network, as shown in Figure 3. We label the data with the help of in-house experts. A 2D sample of the labelled data is shown in the figure. All the 2D images were labelled by hand. We categorized the images into different rock types based on the size of their visual properties that affect flow properties. These properties include porosity and grain size. Since labelling is a complex and time-consuming task, we only labelled a limited number of images, 8 and 13 images for Rock Sample 1 and Rock Sample 2, respectively, which were used for training and evaluation. The total number of segmentation classes in our data is six. It is important to note that unsupervised methods do not work well for the segmenting laminar rock types. Therefore, we use a supervised approach. By laminar, we mean that each rock type is horizontally laid upon each other instead of having complex topologies of rock types.

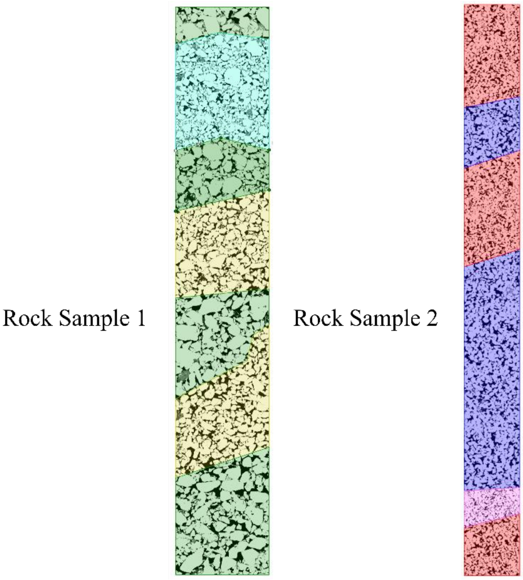


Figure 3 Training data consists of two different rock samples which have diverse pores and solid regions.

2) Data Split

To train and evaluate the performance of the U-Net model, we split the labeled dataset into a training set and a test set. We used 70% of the data for training and 30% for testing. The split was done randomly, while ensuring that both the training and test sets had similar proportions of different rock types.

3) Image Pre-processing:

We start with a high-resolution 3D image of a rock sample obtained through micro-CT scanning. We fix the size of the input image to be 192 x 64 so that input to U-Net is always fixed in size. Therefore, we crop and resize the original images for training and inference. We normalize the image to have pixel values in the range [0,1].

4) Data Augmentation:

We apply data augmentation techniques to increase the training dataset's diversity. The first data augmentation technique is the random cropping of the images vertically. The horizontal direction is not cropped. To ensure a good learning process, we select the size of the vertical crop to be three times the width of the horizontal image. This setting provides sufficient rock types encountered in the training image. At the same time, random crops are chosen in the vertical direction to ensure that the data augmentation effect is achieved to avoid overfitting. In addition to cropping, we randomly flipped along the x and y directions of the input image, improving the training process. The final image is resized by downsampling to 192 x 64 to ensure that input to the U-Net is always standardized.

5) Training Parameter detail:

We use the Adam optimizer with a learning rate of 0.001 and a batch size of 64. We train the model for 150 epochs.

6) Evaluation Metrics:

Dice coefficient, mean Intersection over Union (mIoU), and mean Average Precision (mAP) are widely recognized evaluation metrics for semantic segmentation tasks.

The Dice coefficient, or Sørensen–Dice index, quantifies the similarity between two sets of pixels, specifically the predicted segmentation mask and the actual, or ground truth, segmentation mask. This metric is computed as twice the area of overlap between the two-pixel sets, divided by their total pixel count.

The mIoU, another prevalent evaluation metric for semantic segmentation, gauges the overlap extent between the predicted and ground truth segmentation masks. It calculates the intersection over union (IoU) between these two masks for each class and subsequently averages these values across all classes.

The Average Precision (AP) metric assesses the algorithm's precision and recall at varying confidence thresholds. To compute the mean Average Precision (mAP), the AP is initially calculated for each class at different decision thresholds, after which the mean of these AP values over all classes is determined.

In essence, while the Dice coefficient measures pixel-wise agreement between the predicted and ground truth masks, the IoU provides a measure of region-wise agreement. The AP, conversely, focuses on the trade-off between precision and recall for each class. Together, these metrics provide a comprehensive evaluation of the performance of semantic segmentation tasks.

RESULTS

A. Qualitative Results

The performance evaluation of our U-Net model involved a thorough visual analysis of the model-generated segmentation maps, which were then compared to the ground truth labels. Depicted in Figure 4 The Qualitative results of our model are shown. From left to right we have the input image, ground truth segmentation mask and predicted segmentation mask., are the qualitative results of our methodology. This figure features the original 2D image, the corresponding ground truth label, and the U-Net model-produced segmentation map for a selection of images.

As seen in the figure, the model's output confirms the U-Net model's proficiency in delineating the laminar rock types in our dataset. The segmentation maps generated by our model demonstrate an impressive alignment with the ground truth labels, particularly in discerning the boundaries differentiating various rock types.

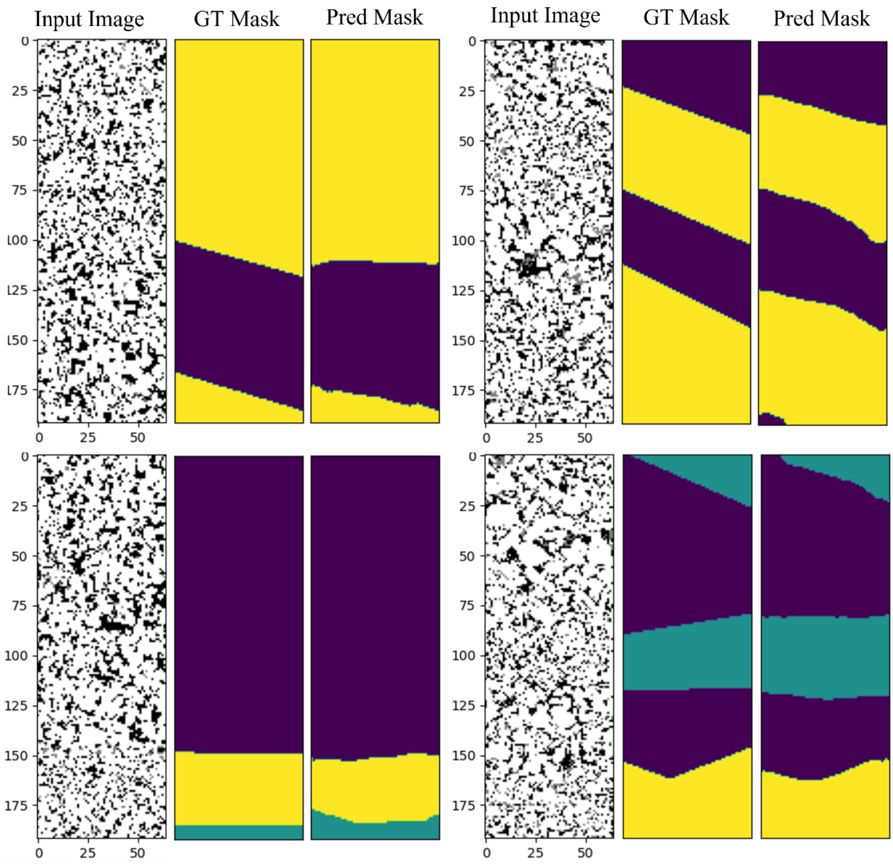


Figure 4 The Qualitative results of our model are shown. From left to right we have the input image, ground truth segmentation mask and predicted segmentation mask.

B. Quantitative results

For a quantitative assessment of our proposed method, we calculated the Dice score, mAP, and mIOU on the test set.

Table 1 summarises these quantitative results, encompassing the Dice score, mIOU, and AP metric for the test set. For the algorithm to be effective, each metric should score above 50, with higher scores indicating better performance.

The results affirm the effectiveness of our method in segmenting the laminar rock types within the 2D images. Given its success, this method holds promise for addressing similar challenges in geology and materials science.

Table 1 Summary of Dice Score, mean Intersection over union score (mIOU) and Average precion (AP) of our model.

Model	**Dice Score**	**mIOU**	**AP**
U-Net	94.71	85.18	85.93

C. Explainability Analysis

Using the Grad-CAM [17] technique, we analyzed our model's decision-making and identified regions vital for rock-type predictions.

The technique involves overlaying a heatmap onto the original image, highlighting regions most crucial for rock-type predictions. This process, demonstrated in Figure 5, enables visual examination of class activation maps and improves understanding of our model's predictions.

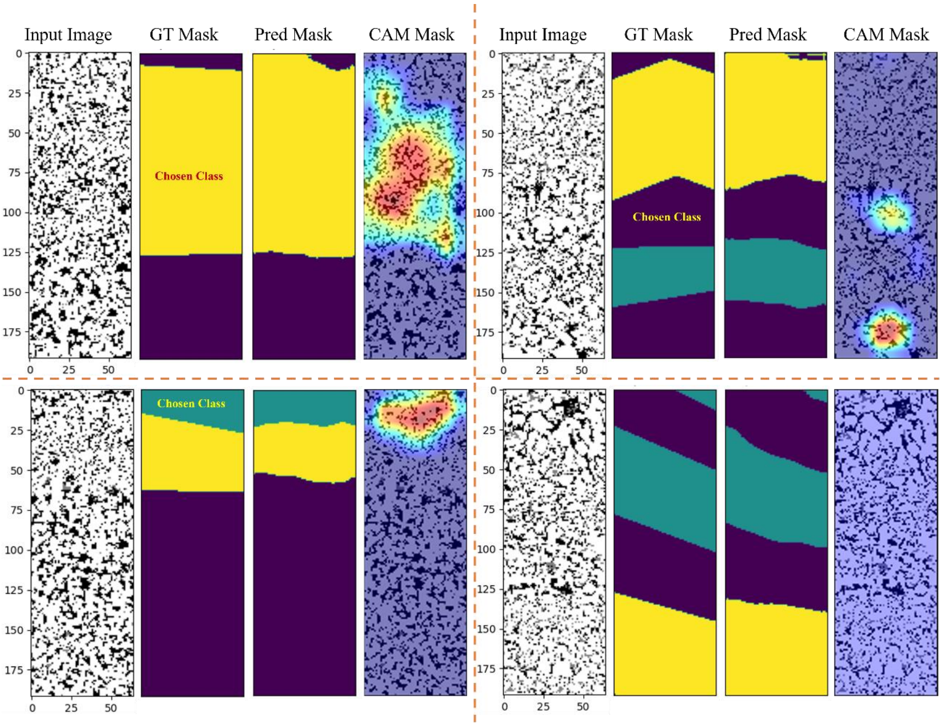


Figure 5 Grad Cam Explain ability Analysis: We query each class as shown by tag 'Chosen Class' in the GT masks. It can be seen that the CAM mask image shows the area where the neural network is paying attention for obtaining the Pred Mask.

In our Grad-CAM analysis, we focused on the neural network's decision-making by identifying pixels responsible for detecting the 'Chosen Class'. This is depicted in each sub-figure, excluding the bottom right one, where the chosen class for each figure is shown. Observations show that the neural network accurately concentrates on the significant regions for decision-making. As a sanity check, we queried the network to highlight pixels responsible for a class that is not present in the input image. As seen from the bottom right sub-figure, the network correctly does not highlight any region, suggesting a sound decision-making process.

Overall, the Grad-CAM analysis provides qualitative evidence that our U-Net method is performing well in rock-type segmentation tasks and has a logical basis for its predictions.

DISCUSSION

In this study, we proposed a U-Net-based approach for rock-type segmentation from 2D micro-CT images. Our method achieved promising results on the test set, with good scores in evaluation metrics. We demonstrate that learning-based methods can produce accurate rock types. These maps can be used in the digital rock workflow, making the overall workflow more efficient.

One limitation of our approach is that it only deals with laminar rock types, where each rock type is horizontally laid upon the other. This limitation could be addressed in the future by increasing the amount of training data that is also diverse. Another line to explore is the unsupervised methods for rock-type segmentation. In our observation, these methods fail on laminar rock types.

CONCLUSION

In conclusion, we have demonstrated that U-Net is a powerful tool for rock-type segmentation in 2D images. Our model achieved high accuracy on a test dataset and provided visually meaningful results. The data set we used was limited to laminar rock types. Experts labelled the data and images. Due to the black-box nature of deep learning models, we utilized Grad-CAM to visualize the features that contributed to the segmentation map, providing insights into the model's decision-making process. Our study showcases the potential of deep learning methods for rock-type segmentation and offers a method to make the overall digital rock workflow more efficient.

ACKNOWLEDGEMENTS

This work was partially supported by the Norwegian Research Council (grant number 296093) and the members of the SmartRocks joint industry project (ENI AS, Repsol AS, and Chevron Corporation). (Ismail, 2013)

REFERENCES /no more than 15-18 references for 8-page article/

[1] P. Forbes, "The status of core analysis," *Journal of Petroleum Science and Engineering,* vol. 19, pp. 1-6, 1998.

[2] C. Arns, F. Bauget, A. Sakellariou, T. Senden, A. Sheppard, R. Sok, A. Ghous, W. Pinczewski, M. Knackstedt and J. Kelly, "Digital core laboratory: Petrophysical analysis from 3D imaging of reservoir core fragments," *Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description,* vol. 46, 2005.

[3] Y. Lecun, Bottou, L., Bengio, Y. and Haffne, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, pp. 2278-2324, 1998.

[4] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM,* vol. 60, no. 6, pp. 84-90, 2017.

[5] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition,* arXiv preprint arXiv:1409.1556, 2014.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.

[7] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[8] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.

[9] V. Badrinarayanan, A. Kendall and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, pp. 2481--2495, 2017.

[10] O. Ronneberger, P. Fischer and T. Brox, U-net: Convolutional networks for biomedical image segmentation, Springer, Ed., Medical Image Computing and Computer-Assisted Intervention--MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 2015, pp. 234-241.

[11] N. I. Ismail, S. Latham and C. H. Arns, "Rock-typing using the complete set of additive morphological descriptors," in *SPE Reservoir Characterization and Simulation Conference and Exhibition*, OnePetro, 2013.

[12] Y. Wang, A. Alzaben, C. H. Arns and S. Sun, "Image-based rock typing using local homogeneity filter and Chan-Vese model," *Computers & Geosciences,* p. 104712, 2021.

[13] Y. Cui, I. Shikhov, R. Li, S. Liu and C. H. Arns, "A numerical study of field strength and clay morphology impact on NMR transverse relaxation in sandstones," *Journal of Petroleum Science and Engineering,* vol. 202, p. 108521, 2021.

[14] H. Jiang and C. H. Arns, "Fast Fourier transform and support-shift techniques for pore-scale microstructure classification using additive morphological measures," *Physical Review E,* vol. 101, p. 033302, 2020.

[15] N. H. Alhwety, I. Shikhov, J.-Y. Arns and C. H. Arns, "Rock-typing of thin-bedded reservoir rock by NMR in the presence of diffusion coupling," in *SPWLA 57th Annual Logging Symposium*, OnePetro, 2016.

[16] E. E. Baraboshkin, L. S. Ismailova, D. M. Orlov, E. A. Zhukovskaya, G. A. Kalmykov, O. V. Khotylev, E. Y. Baraboshkin and D. A. Koroteev, "Deep convolutions for in-depth automated rock typing," *Computers & Geosciences,* vol. 135, no. 0098-3004, p. 104330, 2020.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017.

Appendix E

Paper E: Generating 3D Images of Material Microstructures from a Single 2D Image: A Denoising Diffusion Approach

Generating 3D images of material microstructures from a single 2D image: A denoising diffusion approach

Johan Phan[1,2*†], Muhammad Sarmad[1†], Leonardo Ruspini[2], Gabriel Kiss[1], Frank Lindseth[1]

[1]Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway.
[2]Petricore Norway, Trondheim, Norway.

*Corresponding author(s). E-mail(s): johan.phan@ntnu.no;
Contributing authors: muhammad.sarmad@ntnu.no;
†These authors contributed equally to this work.

Abstract

Three-dimensional (3D) images provide a comprehensive view of material microstructures, enabling numerical simulations unachievable with two-dimensional (2D) imaging alone. However, obtaining these 3D images can be costly and constrained by resolution limitations. We introduce a novel method capable of generating large-scale 3D images of material microstructures, such as metal or rock, from a single 2D image. Our approach circumvents the need for 3D image data while offering a cost-effective, high-resolution alternative to existing imaging techniques.

Our method combines a denoising diffusion probabilistic model (DDPM) with a generative adversarial network (GAN) framework. To compensate for the lack of 3D training data, we implement chain sampling, a technique that utilizes the 3D intermediate outputs obtained by reversing the diffusion process. During the training phase, these intermediate outputs are guided by a 2D discriminator. This technique facilitates our method's ability to gradually generate 3D images that accurately capture the geometric properties and statistical characteristics of the original 2D input.

This study features a comparative analysis of the 3D images generated by our method, *SliceGAN* (the current state-of-the-art method), and actual 3D micro-CT images, spanning a diverse set of rock and metal types. The results shown an improvement of up to three times in the FID (Frechet Inception Distance) score, a typical metric for evaluating the performance of image generative models, and enhanced accuracy in derived properties compared to *SliceGAN*. The potential of our method to produce high-resolution and statistically representative 3D images paves the way for new applications in material characterization and analysis domains.

Keywords: 3D image generation, Material microstructures, Denoising diffusion probabilistic models, Single-image-based generation

1 Introduction

Three-dimensional (3D) volumetric images are a critical resource in various disciplines, including geophysics, petroleum, and materials science, due to their role in the numerical analysis and computational modeling of materials' internal structures. These data, represented as a 3D grid of voxels (volumetric pixels), provide an intricate view of the internal structure of diverse materials, which is vital for deriving physical properties in different industries.

The acquisition of 3D images often poses significant challenges. Traditional image acquisition methods, such as computed tomography (CT) scanners, require substantial financial investment and skilled operators. Conventional techniques like micro-CT are often limited by the resolution capabilities of the imaging equipment. Practical issues related to sample preparation can further complicate acquiring high-resolution 3D images for specific materials or structures. Micro-CT scanners, which utilize X-rays, may also encounter difficulties penetrating radiodense materials, particularly metals, creating additional challenges in capturing comprehensive 3D images of such materials.

In contrast, sub-micrometer 3D scanning solutions such as nano-CT [1] and FIB-SEM (Focused Ion Beam Scanning Electron Microscope) [2] may offer higher resolution capabilities. However, these technologies come with a significantly higher price tag compared to micro-CT, have limited scanning sample sizes, and face limitations related to image quality, inhibiting their widespread application across various sectors within academia and industry [3] [4].

Given these challenges, there has been growing interest in developing techniques that can generate 3D volumetric data using 2D images. This approach offers a promising alternative, as 2D imaging methods such as optical microscopes or Scanning Electron Microscopes (SEM) are in many cases more cost-effective and flexible in terms of resolution capabilities, including at the nanoscale [5]. Currently, 2D images are primarily used alongside 3D images for quality control or as supplementary material when the resolution of 3D imaging fails to fully capture the studied sample's structure. Consequently, the capability to create 3D images from 2D images could reduce costs associated with the imaging process, enhance accessibility, and improve the efficiency of 3D image generation across diverse scientific fields.

Most previous approaches for generating 3D voxelized data from 2D input require learning from 3D image data. Our work belongs to the category of methods that generate 3D images from a single 2D input, as shown in Figure 1. In addition, they can do so by learning from 2D images. The uniqueness of our work is that we propose a framework that utilizes a denoising diffusion-based probabilistic model (DDPM) [6]. Since DDPM requires ground truth (GT) data for training, we propose modifications to enable it to operate without needing 3D GT. This is achieved through utilizing the reverse diffusion process (chain sampling). In addition, we utilize the generative adversarial network (GAN) loss [7] since it helps to generate realistic samples. GANs can be unstable when used in a standalone setting to learn from 2D images generating 3D images. However, we propose to stabilize training by combining GANs with diffusion models.

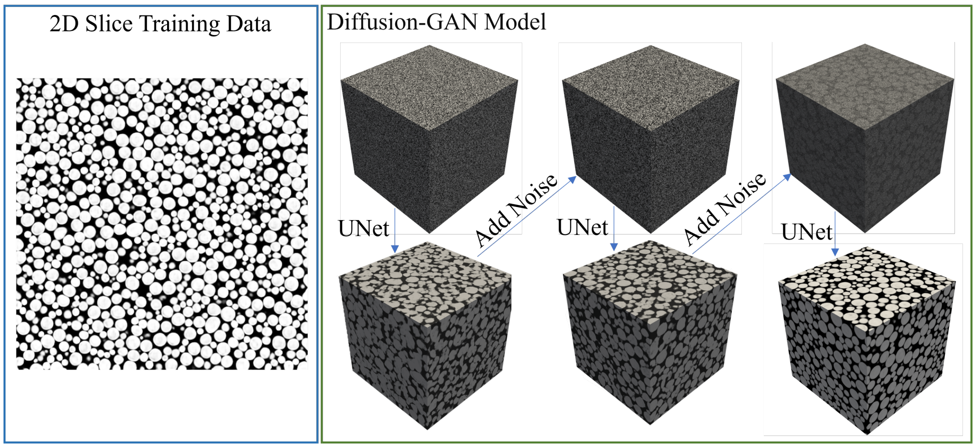


Fig. 1: Diffusion-GAN Model: The proposed method is based on a denoising diffusion process combined with a generative adversarial framework. In this setting, at test time, starting from a cube of noise, the noise is iteratively estimated using a Unet, removed and then added back to the sample. This process is repeated to obtain a noise-free representative sample. Our method can learn from only a few 2D slices of the training image, as shown on the left.

Additionally, addressing the practical requirements of 3D image generation for industrial applications, we have adapted the diffusion process to converge into the final image with just a few denoising steps, specifically 11 in this study, as opposed to the thousands of steps typically used in 2D image generation. This modification is crucial for reducing computation time when working with 3D data, where a large image with over a billion voxels ($1000 \times 1000 \times 1000$) is often necessary for comprehensive material characterization.

Our results prove to be more accurate than previous works in both visual quality and physical/statistical properties. We further showed that our model can successfully learn to generate 3D images from a single 2D input across a wide variety of cases, ranging from rocks to metal alloys.

The contributions of our work can be summarized as follows:

- We propose a method based on the Denoising Diffusion Probabilistic Model (DDPM) for generating 3D microstructures from a single 2D image.
- We demonstrated the feasibility of applying DDPM without the need for training on GT data.
- Our method significantly outperforms existing approaches in terms of visual quality and statistical properties. Moreover, it demonstrates robust performance even on complex images with high heterogeneity, where current state-of-the-art methods fail.

The ability to generate 3D images from a single 2D image would allow us to perform characterizations and analyses that require the availability of 3D data. This technique would be suitable for application in cases where micro-CT imaging is not feasible, such as when capturing features at sub-micrometer resolution or for materials without any density contrast, or in the case of high-density materials like metals [5].

1.1 Related Works

Creating 3D models or images of specific porous structures or materials has been a long-standing research challenge since the advent of image-based numerical analysis. Existing methodologies for tackling this problem can be broadly classified into three main categories: process-based modeling, properties-based generation, and machine learning-based generation.

Process-based modeling

Process-based modeling approaches aim to emulate the mechanisms underlying the natural formation of materials. In these models, the physical and chemical processes that occur during material formation, such as deposition, compaction, cementation, dissolution, and fracturing, are translated into mathematical and computational algorithms. By closely imitating these natural processes, process-based models allow for extensive control over the properties and characteristics of the generated samples, making them useful for hypothesis testing or simulating a wide array of possible scenarios [8–13].

Despite their benefits, process-based models also have certain limitations. Simulating natural processes with satisfactory accuracy is computationally intensive, time-consuming, and challenging due to the complex interplay of numerous factors and the stochastic nature of many processes. Furthermore, operating these models requires a comprehensive understanding of the processes being replicated and the simplification of actual phenomena. As a result, there can be substantial discrepancies between the structures produced by these models and the real materials.

Properties-based generation

Generating 3D images can also be achieved through an iterative generation process that aims to converge toward a structure with desired statistical properties. This approach encompasses both stochastic-based modeling and optimization-based modeling [14–16] where statistical descriptors such as the Minkowski functional and the n-point correlation function are commonly used. One of the main advantages of this method is its capability to generate models with specific desired properties. However, this approach is restricted to binary segmented images since most statistical descriptors for images are specifically designed for binary data. In more complex scenarios, especially for heterogeneous material, accurately capturing non-statistical representative features becomes challenging, potentially leading to the generation of unrealistic images even when the material's statistical properties are matched.

Machine learning-based generation

The recent advancements in 3D image generation have predominantly focused on utilizing machine learning techniques with existing 3D data to generate new images. One prominent approach in this field is the use of Generative Adversarial Networks (GANs) [7]. GANs have been applied for unconditional generation [17, 18] and conditional generation [19**?** −23] of 3D images for micro-CT data. However, training a GAN-based model requires careful attention to ensure stability [24, 25]. Challenges such as mode collapse and catastrophic forgetting can arise when using GANs for conditional generation tasks, necessitating the incorporation of additional consistency loss [26, 27].

To overcome the limitations of GAN-based methods, hybrid models that combine transformers and VQ-VAEs have emerged as an alternative solution. These models offer stable training and the ability to generate high-fidelity 3D rock samples from 2D conditional images [28].

However, all of the mentioned works rely on the availability of 3D GT data for training, which can pose limitations in terms of accessibility, particularly when dealing with samples that contain a significant number of sub-micrometer features. In a recent study, Kench et al. [29] showcased the capability of generating 3D microstructures with only 2D images as training data. Nevertheless, their approach relied on GANs, which are prone to common issues like unstable training and mode collapse.

In contrast to existing approaches, we propose a novel and stable diffusion-based method that achieves 3D image generation of material microstructures using only a single 2D image.

1.2 Background

Denoising Diffusion Probabilistic Models

This section provides a basic understanding of Denoising Diffusion Probabilistic Models (DDPM), also known as diffusion models, which serve as the foundation for our proposed method. DDPM consists of two main processes: the forward and reverse processes [6].

In the forward process, noise, typically Gaussian noise, is gradually added to the data distribution $q(\mathbf{x}_0)$, where $\mathbf{x}_0$ is the noise-free target. This process proceeds step by step, with the variance of the added noise changing according to a predefined schedule β_t $(\beta_1, \ldots, \beta_T)$. The forward process can be expressed as follows:

$$
\begin{aligned}
q(\mathbf{x}_{1:T}|\mathbf{x}_0) &= \prod_{t\geq 1} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \\
&= \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}),
\end{aligned}
\tag{1}
$$

In the reverse process, the aim is to recover the data from noise in steps. A diffusion model is required, which is parameterized by θ with mean $\boldsymbol{\mu}\theta(\mathbf{x}t, t)$ and variance σ_t^2. The reverse denoising process is given as:

$$
\begin{aligned}
p_\theta(\mathbf{x}_{0:T}) &= p(\mathbf{x}_T) \prod_{t\geq 1} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \\
&= \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \sigma_t^2\mathbf{I}),
\end{aligned}
\tag{2}
$$

To train this model, the variational bound on the negative log-likelihood objective $p_\theta(\mathbf{x}_0)$ is optimized, defined as $\int p_\theta(\mathbf{x}_{0:T})d\mathbf{x}_{1:T}$. The variational lower bound is equivalent to matching the true denoising distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ with the parameterized denoising model $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ using the loss function:

$$
\mathcal{L} = -\sum_{t\geq 1} \mathbb{E}_{q(\mathbf{x}_t)} \left[D_{\mathrm{KL}} \left(q(\mathbf{x}_{t-1}|\mathbf{x}_t) \| p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \right) \right] + C \tag{3}
$$

where D_{KL} represents the Kullback-Leibler (KL) divergence between the two distributions, i.e., the true denoising distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ and the parameterized denoising model $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$. C is a constant.

Two fundamental assumptions are commonly made in diffusion models: First, the denoising distribution $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ is modeled as a Gaussian distribution. Second, the number of denoising steps T is assumed to be large.

Denoising Diffusion GANs

To address the challenge of requiring a large number of denoising steps in diffusion models, Xiao et al. [30] proposed a combination of Denoising Diffusion Probabilistic Models (DDPM) and Generative Adversarial Networks (GANs). Their work introduced two major modifications to the original diffusion process:

- Adversarial Loss: Instead of using typical loss functions like Mean Squared Error (MSE) or Mean Absolute Error (MAE), this method used an adversarial loss from a conditional discriminator.
- Direct Output of Noise-Free Images: Rather than training the DDPM to output the noise for a given image, which is then subtracted to obtain the noise-free image, this method directly generates the noise-free image.

By combining DDPM with GANs and implementing the mentioned modifications, the method proposed by Xiao et al. achieved a drastic reduction in the number of denoising steps required by a factor of 10^3, while also improving the quality of the generated images and maintaining stable training.

2 Method

Denoising diffusion models alone are inherently unsuitable for solving the problem of generating 3D images from a single 2D image, as they require a noise-free GT image $\mathbf{x}_0$ for training. The absence of $\mathbf{x}0$ renders the forward process $q(\mathbf{x}1 : T|\mathbf{x}_0)$ mentioned in Equation 1 invalid. Consequently, to adapt DDPM to our specific problem, we changed the original DDPM pipeline to cater to learning with only 2D GT data.

Inspired by the *SliceGAN* architecture introduced by Kench et al. [29], which utilized a 2D discriminator as the adversarial loss for a 3D generator, we have adapted the method proposed in the Denoising Diffusion GANs work by Xiao et al. [30] to a similar setting as *SliceGAN*. An overview of our method is shown in Fig. 2. The main difference between SliceGAN and our method is that SliceGAN uses only generative adversarial networks. However, we use the diffusion model with GANs to get a more stable and robust training pipeline. However, using the diffusion model for this task is not trivial. Therefore, we provide a novel method to deploy diffusion models for 3D image generation from 2D slices.

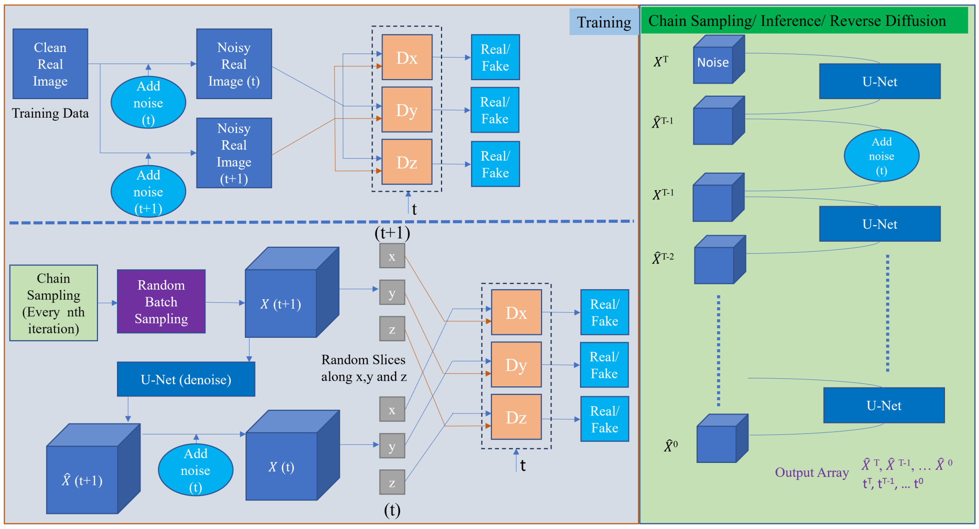


Fig. 2: Our Method: This figure shows the overview of our diffusion GAN based model for 3D generation using only 2D data for training.

Chain Sampling as a means of Data generation

The lack of a noise-free GT $\mathbf{x}_0$ presents a significant challenge when adapting the Denoising Diffusion Probabilistic Models (DDPM) framework to the task of generating 3D images from a single 2D input. To overcome this challenge, we propose a novel approach called **chain sampling**, which involves leveraging the reverse diffusion process during training alongside the forward diffusion process. This departure from the conventional usage of the reverse process solely for inference or testing purposes is a key distinction in our method. By employing chain sampling, we can utilize intermediate results during training as a substitute for the missing noise-free 3D GT $\mathbf{x}_0$, under the assumption that the denoising model is still progressing in the correct direction. The chain sampling process, illustrated in Figure 2, involves adding the corresponding level of noise using a simplified noise addition process as shown in Equation 4. The denoising model G then performs the denoising operation on the image, as described in Equation 5.

$$\mathbf{x}_t = q(\hat{\mathbf{x}}_{t+1}) = \beta_t * \hat{\mathbf{x}}_{t+1} + (1 - \beta_t) * \mathcal{N}(\boldsymbol{\mu}_\theta(\hat{\mathbf{x}}_{t+1}, t+1), \sigma_t^2) \tag{4}$$

$$\hat{\mathbf{x}}_t = G(x_t, t) \tag{5}$$

Discriminator setting

To ensure the training of our denoising model in the absence of noise-free GT $\mathbf{x}_0$, we employ a 2D discriminator trained on both the 2D image and the intermediate results from the chain sampling process. Similar to any GANs-based architecture, our discriminator requires both real data and fake (generated) data to train:

Fake Data: In the lower part of Figure 2, we illustrate the process for generating fake data. During each training iteration, we sample an image from the output array of the chain sampling process, which corresponds to a denoised generated image $\hat{\mathbf{x}}_t$. From this image, we randomly select a slice from each axis (X, Y, Z) and feed these slices to their respective discriminators D_x, D_y, D_z (Eq. 6). While it is possible to use a single discriminator, we have found that utilizing three separate discriminators leads to more stable training and enables us to handle asymmetrical images effectively.

$$\begin{gathered} D_\phi(\mathbf{x}_{t-1}^{3D}, \mathbf{x}_t^{3D}, t) = D_x(\mathbf{x}_{t-1}^{3Dx}, \mathbf{x}_t^{3Dx}, t) + \\ D_y(\mathbf{x}_{t-1}^{3Dy}, \mathbf{x}_t^{3Dy}, t) + D_z(\mathbf{x}_{t-1}^{3Dz}, \mathbf{x}_t^{3Dz}, t) \end{gathered} \tag{6}$$

Real data: In the upper part of Figure 2, we depict the process of generating real data for training the discriminator. Since our discriminator consists of 2D convolutional layers, we can easily add different levels of noise into the 2D image to retrieve the corresponding $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$. These noisy images, with noise added according to predefined β_t values, serve as the real data inputs for training the discriminator.

$$\begin{aligned} D_\phi(\mathbf{x}^{2D}_{t-1}, \mathbf{x}^{2D}_t, t) = D_x(\mathbf{x}^{2D}_{t-1}, \mathbf{x}^{2D}_t, t) + \\ D_y(\mathbf{x}^{2D}_{t-1}, \mathbf{x}^{2D}_t, t) + D_z(\mathbf{x}^{2D}_{t-1}, \mathbf{x}^{2D}_t, t) \end{aligned} \tag{7}$$

$$\begin{aligned} \min_\theta \sum_{t\geq 1} \mathbb{E}_{q(\mathbf{x}_t)} \big[& D_{\mathrm{x}}\big(q(\mathbf{x}^{2D}_{t-1}|\mathbf{x}^{2D}_t)\|q_(\mathbf{x}^{3Dx}_{t-1}|\mathbf{x}^{3Dx}_t)\big) \\ & +D_{\mathrm{y}}\Big(q(\mathbf{x}^{2D}_{t-1}|\mathbf{x}^{2D}_t)\|q_(\mathbf{x}^{3Dy}_{t-1}|\mathbf{x}^{3Dy}_t)\Big) \\ & +D_{\mathrm{z}}\big(q(\mathbf{x}^{2D}_{t-1}|\mathbf{x}^{2D}_t)\|q_(\mathbf{x}^{3Dz}_{t-1}|\mathbf{x}^{3Dz}_t)\big)\big], \end{aligned} \tag{8}$$

To manage the computational intensity of generating large 3D images, we had to restrict the number of denoising timesteps to a smaller value, specifically $T = 11$. Consequently, this resulted in larger β_t values for each diffusion step. Since our approach involved a significantly reduced number of denoising timesteps compared to the original Denoising Diffusion GANs, we paid close attention to selecting suitable β_t values. The aim was to maintain a similar level of denoising complexity for each step, despite the reduced overall number of steps.

For the adversarial training, we define the sum of the three time-dependent discriminators as $D_\phi(\mathbf{x}_{t-1}, \mathbf{x}_t, t) : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R} \to [0, 1]$, with parameters ϕ_x, ϕ_y, and ϕ_z, as shown in Equation 8. This discriminator takes the N-dimensional 2D slices $\mathbf{x}^{2D}_{t-1}$ and $\mathbf{x}^{2D}_t$ of $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$ as inputs and determines whether the input is a plausible denoised version of $\mathbf{x}^{2D}_t$ or not.

3 Experiments

3.1 Data

In this study, we used a combination of internal data and publicly available data to evaluate the performance of our model. In the first case, we validated the quality of the generated images compared to 3D GT (Table 1) using four micro-CT 3D images: a Glass Bead image, two sandstone images with different resolutions, and a Savoniere carbonate image from the digital rock portal [31]. For each 3D image, we randomly selected five unrelated 2D slices to train our model.

In the second case (Table 2), we considered scenarios where only a single 2D image was available for both training and evaluation. The images used in this case include a Cast Iron with magnesium-induced spheroidized graphite and a Brass (Cu 70%, Zn 30%) with recrystallized annealing twins from Microlib [32]. Both images were captured using reflected light microscopy [33]. In addition, we also used an SEM image of kaolinite clay minerals.

3.2 Evaluation Metric

We utilize Fréchet Inception Distance (FID) as our evaluation metric [34]. FID is a popular choice for assessing the quality of generated images in tasks such as GAN evaluations. It can serve as a measure of similarity between two datasets of images. The FID metric calculates the Fréchet distance between two multivariate Gaussian distributions that are fitted to feature representations of the Inception network. One distribution represents real images, while the other represents generated images. The lower the FID score, the more similar the two datasets of images are in terms of their distribution in the high-dimensional space defined by the Inception network. Hence, a lower FID signifies a higher quality of generated images. It captures how well the generated images mimic the real ones.

3.3 Resources and Hyper-parameters

Each experiment in this study is conducted using PyTorch on a single Nvidia RTX 3090 GPU. The training time for both *SliceGAN* and our method was set to 24 hours. In our experiments, we used 11 denoising time steps ($T = 11$) with corresponding β_t values ranging from 0.9100 to 0.0000, as follows: [0.9100, 0.8109, 0.7058, 0.5985, 0.4929, 0.3931, 0.3025, 0.2238, 0.1586, 0.1070, 0.0685, 0.0000].

4 Results and Discussions

4.1 Glass beads generation

Our study introduces a novel 3D image generation method, the efficacy of which we assessed through a comparative analysis with glass bead pack images. These images, which naturally depict spherical formations in a densely packed array, are reduced to varying sizes of 2D circles in their planar representations. Our validation approach involved contrasting our method's output with results from similar studies, focusing specifically on the fidelity of reconstructing 3D spherical shapes from these 2D circular projections. For this purpose, we selected benchmark studies by [29], [35], [36],[22] and [18] for comparison.

Figure 3 showcases the cross-sectional views of structures synthesized using our method, alongside those generated by an authentic glass bead image, the SliceGAN algorithm, and the methods employed in the aforementioned studies. This highlights our method's unique capability in accurately rendering spherical shapes in 3D from 2D inputs, a feature distinctively absent in the comparative methods, especially in terms of artifact-free shape generation.

4.2 Comparison with 3D GT

Visual comparison and FID score

Our first case study used four different 3D micro-CT images to evaluate both the visual quality and the accuracy of the characterized properties of our generated 3D images against the GT. For each image, five 2D slices of the xy plane, taken from different locations along the z-axis, were used to train our model and *SliceGAN*. We chose to use five 2D images since a single 2D slice might not fully capture the range of structural variation present in the 3D image. A step length of 11 was selected for our model to ensure fast generation times for large images. During each training iteration,

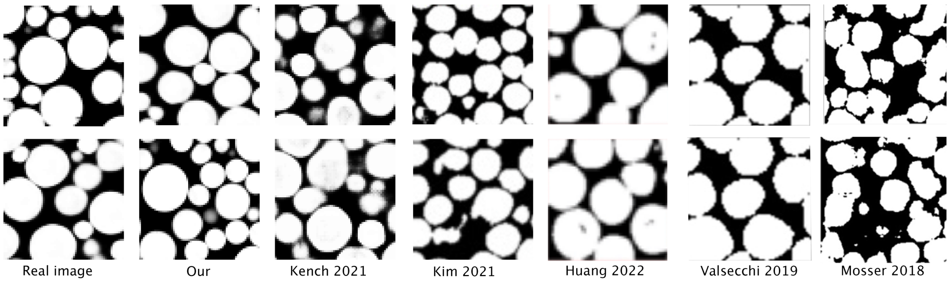


Fig. 3: Visual comparison of spherical shape generation from 2D circular inputs in glass bead packs. This study contrasts our method's results with previous deep learning-based 3D image generation techniques, highlighting our approach's enhanced accuracy in generating spherical shapes.

we used a batch of random 64x64 pixel crops as input, which subsequently produced outputs of 64x64x64. The cross-sections of the images used in the training, as well as the images generated by both methods, are shown in Figure 4.

In assessing the performance of our method compared to *SliceGAN*, we used the Frechet Inception Distance (FID) scores as a measure of visual quality [37]. To compute the FID score, the original requirement was for 2D images as input. To adapt this calculation to 3D images, we treated them as stacks of 2D images and computed the FID score across three dimensions (x, y, z). In comparison to other studies that used the FID score for image generation evaluation, the FID scores presented in Table 1 are notably higher. These higher FID scores are due to the few slices from the 3D image used for training not being able to cover the real data distribution of the 3D GT, especially for heterogeneous materials like the Savoniere Carbonate.

Rock type	Dimension	Our (x, y ,z)			SliceGAN (x, y ,z)		
Glassbead	200x200x200	54.78	60.95	59.67	87.37	99.02	72.33
Bentheimer Sandstone	256x256x256	35.62	49.13	40.99	46.91	61.59	54.65
Sandstone	500x500x500	23.58	23.81	20.46	25.25	29.82	21.76
Savoniere	250x250x250	171.57	186.60	172.20	476.33	436.95	435.58

Table 1: Measured FID score across three dimensions (x, y, z) between the generated image and the GT, the closer to 0 the better

Comparison of Porous Media Properties

In the context of porous media, it is crucial to evaluate our model's performance in terms of physical properties. To calculate these properties, we used Porespy [38], an open-source tool specifically designed to analyze 3D images of porous materials. With Porespy, we calculated local porosity, the two-point correlation function, and pore size distribution of the images depicted in Figure 4.

Porosity (ϕ) – Porosity, representing the volume fraction of void spaces, is a fundamental characteristic of porous media. To calculate porosity, we first convert the

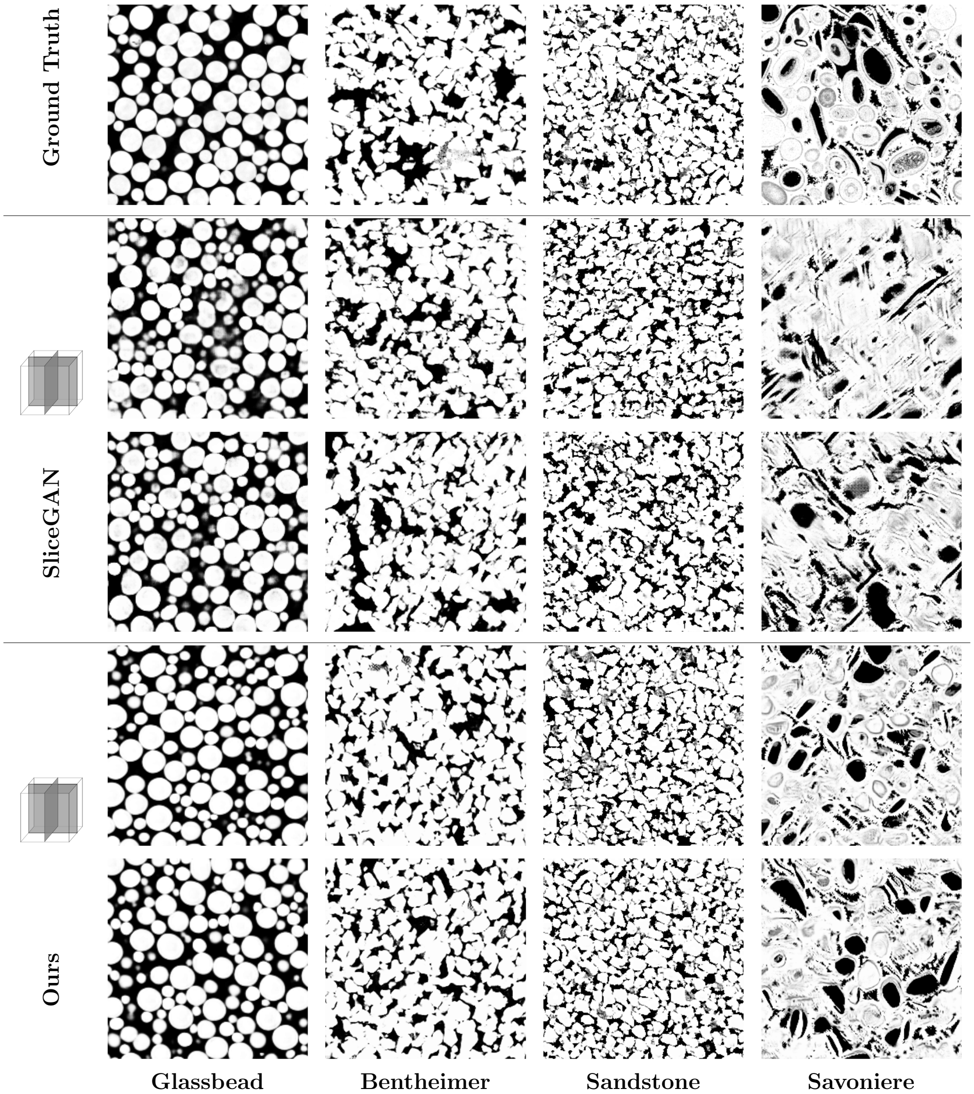
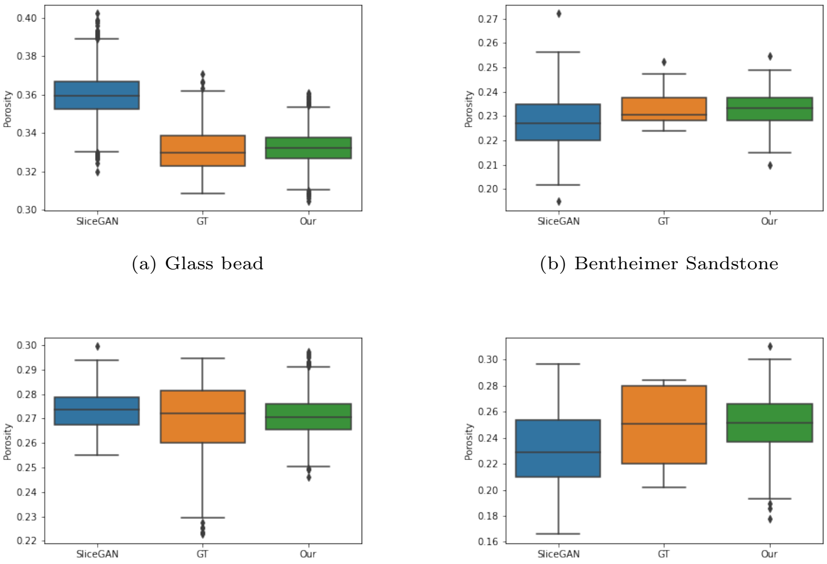


Fig. 4: **Visual comparison with micro-CT images**: Cross-sections of 3D images generated by our method and SliceGAN, alongside their respective ground truth or training data. The GT images are 3D X-ray microCT scans obtained at varying resolutions. The Glassbeads case showcases our method's superior performance over SliceGAN. Our model can capture the spherical shape of the object, even though it only sees circles at the 2D input. In more challenging cases like the Savoniere - a carbonate of fossilized microorganism - our method proves its robustness by generating images that bear a higher resemblance to reality, despite the heterogeneous nature of the original image.

images into binary format through thresholding. Subsequently, we divide the generated images into overlapping cubes with a side length of 128 voxels. The process of calculating porosity is then applied to these cubes, and the results are visualized using box plots shown in Figure 5.



(a) Glass bead

(b) Bentheimer Sandstone

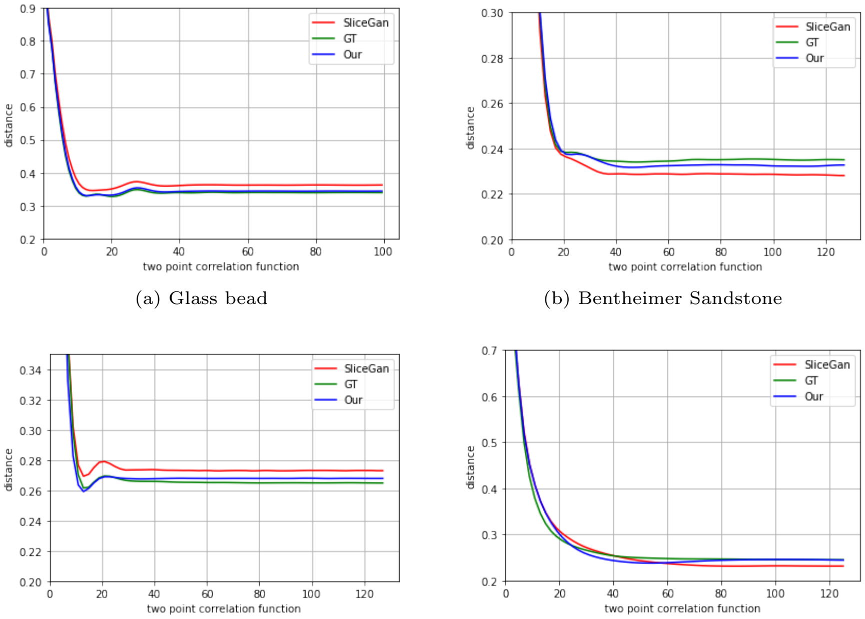
(c) Sandstone

(d) Savoniere Carbonate

Fig. 5: **Porosity** – These box plots show the comparison of porosity between the ground truth, our model and SliceGAN.

Two-Point Correlation Function (ξ) – The two-point correlation function is a significant metric in image analysis, utilized to describe the spatial arrangement and connectivity of the porous structure. In this study, we calculated the probability that a pair of points, separated by a certain distance, both reside within the pore space. This statistical measure is sensitive to the image's degree of homogeneity and isotropy, thus allowing us to capture subtle geometric features of the pore network. The two-point correlation function plots are shown in Figure 6.

Pore Size Distribution – Pore size distribution is a metric that characterizes the range of pore sizes within a porous material, and it plays a vital role in determining how fluids flow and permeate through the material. Ensuring the accuracy of pore size distribution in our 3D generation from 2D images is important because the sizes and arrangement of pores define the transport properties of the porous medium. Accurate pore size distribution in the generated 3D images is essential for maintaining physical accuracy and predictive usefulness in representing the actual material. We calculated the pore size distribution through a process known as porosimetry, which interprets



(a) Glass bead

(b) Bentheimer Sandstone

(c) Sandstone

(d) Savoniere Carbonate

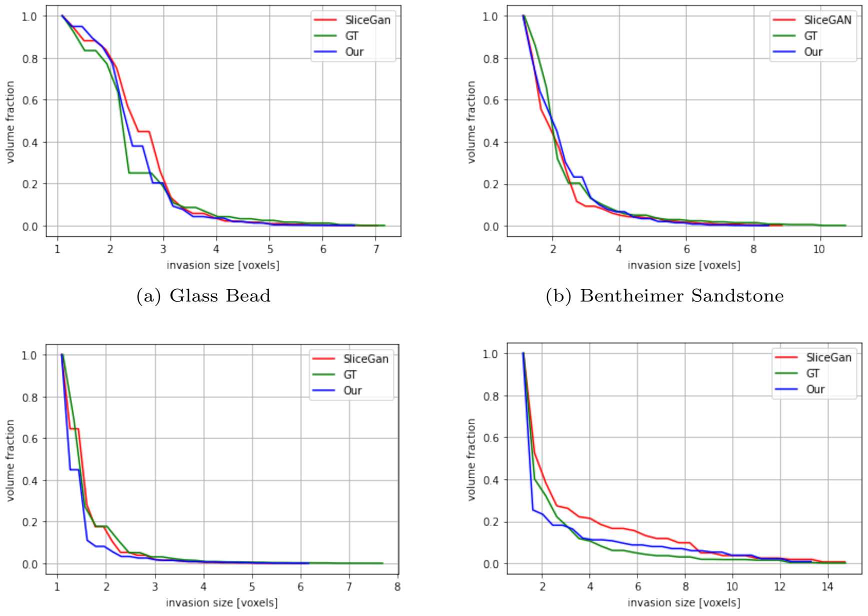
Fig. 6: **Two-Point Correlation** – These plots depict the two-point correlation function in 3D for the Ground Truth (GT) and the images generated by both SliceGAN and our method. They display the relationship between distance and the probability of a given pixel appearing in a binarized segmented image.

each voxel in the image as the radius of the largest sphere that would overlap it. For this, we used the Porespy library and the result is shown in Figure 7.

4.3 3D Generation Experiments from 2D Image of a sample

In the final part of our evaluation, we went beyond testing our method's performance with 3D GT data and tested it in scenarios where only a single 2D image was available, as shown in Figure 8. We used two metal images (Cast Iron and Brass) from Microlibs [32], an online database for images generated by *SliceGAN*, as our test data. We trained our model on the downloaded 2D image and subsequently compared our output with both the original training image and the 3D image created by *SliceGAN*, also sourced from the same Microlibs platform. Our study also included an SEM-acquired image of kaolinite clay, recognized for its complex nanostructure that necessitates capturing in 2D. This is due to the high resolution required to capture the sub-microscopic structure of kaolinite, which is beyond the capabilities of current 3D CT scanners.

These cases represent the scenarios where our algorithm may find its most practical use - situations where acquiring 3D images is infeasible, hence the necessity to generate a 3D model from a 2D image.



(a) Glass Bead

(b) Bentheimer Sandstone

(c) Sandstone

(d) Savoniere Carbonate

Fig. 7: **Pore Size Distribution** – These plots shown the comparison of pore size distribution between the generated images and the ground truth. They aid in understanding our model's effectiveness and SliceGAN's matching the pore structures to the original 3D micro-CT images.

The comparison in terms of FID score is shown in Table 2.

Material type	Dimension	Our (x, y ,z)			SliceGAN (x, y ,z)		
Magnesium treated Cast Iron	800x528	60.03	62.90	63.37	63.81	67.89	66.04
Brass (Cu 70%, Zn 30%)	542x800	107.18	97.37	109.18	266.78	252.17	216.54
Kaolinite clay mineral	256x256	177.78	177.21	183.78.65	360.52	335.13	317.98

Table 2: Measured FID score across three dimensions (x, y, z) between the generated image and the GT

5 Discussion

Our method demonstrates a significant improvement over *SliceGAN*, evident at a visual level, as shown in Figures 4 and 8. For example, in the glass bead case, our method successfully manages to generate 3D spherical structures from training with 2D circular input, while *SliceGAN* and other machine learning based method failed. In the Sandstone cases, our method demonstrated its capability to handle various resolutions and grain sizes. The Savoniere case, owing to the image's heterogeneity, presents a

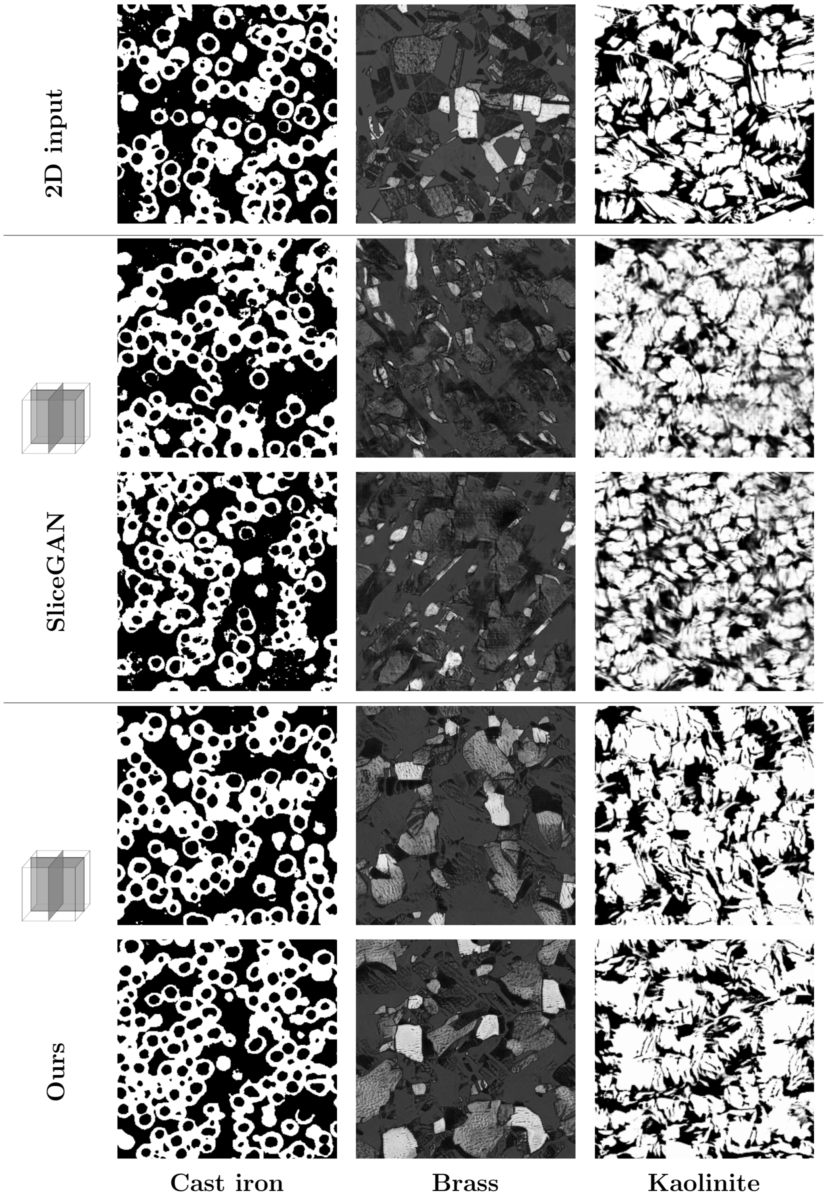


Fig. 8: **Visual comparison with non-CT data sources**: Cross-sections of 3D images produced by our method and SliceGAN, presented alongside their respective 2D training images. This figure includes images of cast iron, brass, and kaolinite clay mineral.

challenge in generating a representative 3D image solely from 2D information. Despite this, our method manages to produce a more visually accurate output compared to *SliceGAN* with a significantly lower FID score (Table 1).

Our evaluation of porosity (Figure 5), the two-point correlation function (Figure 6), and pore size distribution (Figure 7), further affirms our method's efficacy in representing the statistical properties of the 3D GT, even when trained only on five 2D slices. In the final study case where only a single 2D image is available for training and evaluation, we downloaded the 2D image for training and the 3D image created by *SliceGAN* from Microlib. As seen in Figure 8, our method exhibits versatility by generating high-quality images across different material types. Due to the absence of 3D GT images for these cases, we relied on visual inspection and FID scores for evaluation. Despite the insignificant difference in the visual quality and FID score for the cast iron case, the FID score measured in the Brass case clearly favors our method over *SliceGAN* (see Table 2).

In all study cases included in this work, our method has shown comparable or better performance compared to *SliceGAN*. Additionally, the generated 3D images for materials with simple and homogeneous structures closely match the real images, exhibiting both comparable visual quality and measured properties. However, for more complex and heterogeneous materials, especially those with asymmetrical 3D features, there are areas that indicate potential for improvement. Nevertheless, the results of this study lay a promising foundation for future exploration in the domain of 3D porous media image generation from 2D inputs.

6 Conclusion

In this study, we introduced a novel approach to 3D image generation using denoising diffusion probabilistic models (DDPMs) with only a single 2D slice as training data. While DDPMs are not inherently designed for learning from 2D data to represent 3D structures, we introduced a modified reverse diffusion step that effectively denoises a 3D noise vector using a 2D GAN-based discriminator. Our method outperforms state-of-the-art techniques in terms of key physical validation metrics for various types of materials.

Our work marks a significant advancement in the domain of 3D material microstructure generation from 2D inputs. By reducing the dependency on extensive 3D image data and offering a cost-effective, high-resolution alternative to prevailing imaging techniques, our approach paves the way for novel research and practical applications in material characterization and analysis.

Acknowledgments

This work was partially supported by the Norwegian Research Council (grant number 296093) and the members of the SmartRocks joint industry project (Repsol AS, and Chevron Corporation).

Data availability

The 2D training datasets presented in this study is publicly available at [32]. The 3D datasets presented in this study available from the corresponding author on reasonable request

References

[1] Kampschulte, M., Langheinirch, A., Sender, J., Litzlbauer, H., Althöhn, U., Schwab, J., Alejandre-Lafont, E., Martels, G., Krombach, G.: Nano-computed tomography: technique and applications **188**(02), 146–154 (2016). © Georg Thieme Verlag KG

[2] Groeber, M.A., Haley, B., Uchic, M.D., Dimiduk, D.M., Ghosh, S.: 3d reconstruction and characterization of polycrystalline microstructures using a fib–sem system. Materials characterization **57**(4-5), 259–273 (2006)

[3] Xu, C.S., Hayworth, K.J., Lu, Z., Grob, P., Hassan, A.M., García-Cerdán, J.G., Niyogi, K.K., Nogales, E., Weinberg, R.J., Hess, H.F.: Enhanced fib-sem systems for large-volume 3d imaging. elife **6**, 25916 (2017)

[4] Ahmed, H.M.A.: Nano-computed tomography: current and future perspectives. Restorative Dentistry & Endodontics **41**(3), 236–238 (2016)

[5] CT vs. SEM: Which Is Better? — imaging.rigaku.com. https://imaging.rigaku.com/blog/ct-vs-sem-which-is-better. [Accessed 27-12-2023]

[6] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)

[7] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)

[8] Adler, P.M., Jacquin, C.G., Quiblier, J.A.: Flow in simulated porous media. International Journal of Multiphase Flow **16**(4), 691–712 (1990) https://doi.org/10.1016/0301-9322(90)90025-E

[9] Strebelle, S.: Conditional simulation of complex geological structures using multiple-point statistics. Mathematical geology **34**, 1–21 (2002)

[10] Blair, S.C., Berge, P.A., Berryman, J.G.: Using two-point correlation functions to characterize microgeometry and estimate permeabilities of sandstones and porous glass. Journal of Geophysical Research: Solid Earth **101**(B9), 20359–20375 (1996) https://doi.org/10.1029/96JB00879 https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/96JB00879

[11] Tahmasebi, P., Hezarkhani, A., Sahimi, M.: Multiple-point geostatistical modeling based on the cross-correlation functions. Computat. Geosci. **16**, 779–797 (2012)

[12] Tahmasebi, P., Sahimi, M., Caers, J.: Ms-ccsim: accelerating pattern-based geostatistical simulation of categorical variables using a multi-scale search in fourier space. Comput. Geosci. **67**, 75–88 (2014)

[13] Feng, J., Teng, Q., He, X., Qing, L., Li, Y.: Reconstruction of three-dimensional heterogeneous media from a single two-dimensional section via co-occurrence correlation function. Computational Materials Science **144**, 181–192 (2018)

[14] Seibert, P., Raßloff, A., Ambati, M., Kästner, M.: Descriptor-based reconstruction of three-dimensional microstructures through gradient-based optimization. Acta Materialia **227**, 117667 (2022)

[15] Scheunemann, L., Balzani, D., Brands, D., Schröder, J.: Design of 3d statistically similar representative volume elements based on minkowski functionals. Mechanics of Materials **90**, 185–201 (2015)

[16] Lu, B., Torquato, S.: Lineal-path function for random heterogeneous materials. Physical Review A **45**(2), 922 (1992)

[17] Mosser, L., Dubrule, O., Blunt, M.J.: Reconstruction of three-dimensional porous media using generative adversarial neural networks. Phys. Rev. E **96**(4) (2017)

[18] Mosser, L., Dubrule, O., Blunt, M.J.: Stochastic reconstruction of an oolitic limestone by generative adversarial networks. Transport in Porous Media **125**(1), 81–103 (2018)

[19] Volkhonskiy, D., Muravleva1, E., Sudakov, O., Orlov, D., Belozerov, B., Burnaev, E., Koroteev, D.: Reconstruction of 3d porous media from 2d slices. arcXiv:1901.1023v1 (2019)

[20] Zhao, J., Wang, F., Cai, J.: 3d tight sandstone digital rock reconstruction with deep learning. Journal of Petroleum Science and Engineering **207**, 109020 (2021) https://doi.org/10.1016/j.petrol.2021.109020

[21] Coiffier, G., Renard, P., Lefebvre, S.: 3d geological image synthesis from 2d examples using generative adversarial networks. Frontiers in Water **2**, 30 (2020) https://doi.org/10.3389/frwa.2020.560598

[22] Valsecchi, A., Damas, S., Tubilleja, C., Arechalde, J.: Stochastic reconstruction of 3d porous media from 2d images using generative adversarial networks. Neurocomputing **399**, 227–236 (2020)

[23] Shams, R., Masihi, M., Boozarjomehry, R.B., Blunt, M.J.: A hybrid of statistical

and conditional generative adversarial neural network approaches for reconstruction of 3d porous media (st-cgan). Advances in Water Resources **158**, 104064 (2021)

[24] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.: Improved training of wasserstein gans. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, pp. 5769–5779. Curran Associates Inc., Red Hook, NY, USA (2017)

[25] Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 214–223 (2017)

[26] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., *et al.*: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690 (2017)

[27] Thanh-Tung, H., Tran, T.: Catastrophic forgetting and mode collapse in gans. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–10 (2020). IEEE

[28] Phan, J., Ruspini, L., Kiss, G., Lindseth, F.: Size-invariant 3d generation from a single 2d rock image. Journal of Petroleum Science and Engineering **215**, 110648 (2022) https://doi.org/10.1016/j.petrol.2022.110648

[29] Kench, S., Cooper, S.J.: Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion. Nature Machine Intelligence **3**(4), 299–305 (2021)

[30] Xiao, Z., Kreis, K., Vahdat, A.: Tackling the generative learning trilemma with denoising diffusion gans. arXiv preprint arXiv:2112.07804 (2021)

[31] Bultreys, T.: Savonnières carbonate. Digital Rocks Portal (2016). https://doi.org/10.17612/P7W88K

[32] Kench, S., Squires, I., Dahari, A., Cooper, S.J.: Microlib: A library of 3d microstructures generated from 2d micrographs using slicegan. Scientific Data **9**(1), 645 (2022)

[33] Ryan, J., Gerhold, A.R., Boudreau, V., Smith, L., Maddox, P.S.: Introduction to modern methods in light microscopy. Light Microscopy: Methods and Protocols, 1–15 (2017)

[34] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium.

Advances in neural information processing systems **30** (2017)

[35] Kim, S.E., Yoon, H., Lee, J.: Fast and scalable earth texture synthesis using spatially assembled generative adversarial neural networks. Journal of Contaminant Hydrology **243**, 103867 (2021) https://doi.org/10.1016/j.jconhyd.2021.103867

[36] Huang, Y., Xiang, Z., Qian, M.: Deep-learning-based porous media microstructure quantitative characterization and reconstruction method. Physical Review E **105**(1), 015308 (2022)

[37] Seitzer, M.: pytorch-fid: FID Score for PyTorch. https://github.com/mseitzer/pytorch-fid. Version 0.3.0 (2020)

[38] Gostick, J.T., Khan, Z.A., Tranter, T.G., Kok, M.D., Agnaou, M., Sadeghi, M., Jervis, R.: Porespy: A python toolkit for quantitative analysis of porous media images. Journal of Open Source Software **4**(37), 1296 (2019)

ISBN 978-82-326-7822-8 (printed ver.)
ISBN 978-82-326-7821-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology