Computational techniques for virtual reconstruction of fragmented archaeological textiles

Davit Gigilashvili[1*], Hana Lukesova[2], Casper Fabian Gulbrandsen[1], Akash Harijan[1] and Jon Yngve Hardeberg[1]

Abstract

Archaeological artifacts play important role in understanding the past developments of the humanity. However, the artifacts are often highly fragmented and degraded, with many details and parts missing due to centuries' long degradation. Archaeologists and conservators attempt to reconstruct the original state of the objects either physically or virtually. This process includes characterizing and matching fragments' features to identify which ones belong together. However, this process currently requires an extensive and tedious manual labor. Recent development in computational techniques gave rise to computer-assisted ways of virtual reconstruction, where the computer suggests solutions to the puzzle of scattered fragments and supplements or fully replaces manual labor. However, the capabilities of computational techniques remain limited in many aspects. This review summarizes the state-of-the-art computational techniques for puzzle and virtual reconstruction problems in cultural heritage applications, in general – with a particular interest in archaeological textiles. We overview existing computational methods, their applications and limitations. Afterward, based on the current knowledge gaps, we discuss where the field should go next.

Keywords Archaeological textiles, Puzzling solving, Virtual reconstruction

Introduction

Archaeological objects provide invaluable insights into the past of humanity. Textiles have been essential for society throughout history and have always played an important role in demonstrating social- and economic status [1, 2]. It has been suggested that textile crafts predate metallurgy and even pottery [3]. Preserved archaeological textiles constitute a rich source for cultural heritage research. The areas of use can be split into three main categories: clothing (as garments,

headcovers, shoes, and accessories), furnishing and textile art (as upholstery, curtains, bedding, carpets, tapestries, wall hangings, canvas for paintings) and functional and transport textiles (as sails, ropes, fishing nets, various packing) [4].

Even though textiles are often associated with organic materials, composite textiles containing inorganic materials, such as metal threads, are not an exception. There are also rare examples of pure inorganic textiles as those made of asbestos [5]. Archaeological textile materials vary broadly and represent material groups of plant origins such as bast- and leaf fibres, seed and fruit hairs, grasses and even moss; animal origins such as hairs, silks and tissues (i.e. rawhide, leathers, sinews, intestines and sea silk) and inorganic origins such as asbestos and metal [4]. Knowledge about materials used for heritage

*Correspondence:
Davit Gigilashvili
davit.gigilashvili@ntnu.no
[1] Department of Computer Science, Norwegian University of Science and Technology, Postboks 191, 2802 Gjøvik, Norway
[2] University Museum of Bergen, University of Bergen, Postboks 7800, 5020 Bergen, Norway

Gigilashvili *et al. Heritage Science*   (2023) 11:259

Page 2 of 16

textiles and their processing is essential for reconstruction purposes because different materials degrade in different ways [6].

Archaeological textiles belong to highly vulnerable objects due to their material characteristics and physicochemical properties. That is why they have often been recovered in a fragmentary state. For instance, materials of animal origin survive better in wet and slightly acidic environments than cellulose-based plant materials [6]. It is no exception that only a part of a composite textile survives, e.g., a weft made of animal material; in contrast, a warp of plant material can decompose, or certain parts of the weft pattern can be missing. Moreover, archaeological textiles can often be brittle and excessive physical manipulation can have severe consequences for their further preservation. Thus, searching for new ways to reduce their manipulation is crucial. Developing computational techniques for virtual reconstruction can be one solution.

The fragmentary state of archaeological textiles often requires reconstruction to interpret the stories they are telling. Archaeologists, conservators, and other experts must identify which fragments belong together to recover the objects' former shape and appearance. The manual reconstruction process often requires physical manipulation of the original material. Therefore, replacing at least some parts of the manual reconstruction workflow through virtual analysis can positively impact the preservation of original objects. We may compare a reconstruction process of a flat, fragmented textile object to solving

jigsaw puzzles (Fig. 1). Unlike the popular game that is usually provided with a solution image, many important fragments are often missing from a fragmented archaeological artifact and no solution or "correct answer" is known a priori in archaeological applications. Similarly, removing a solution image and a part of the jigsaw pieces may complicate solving the puzzle substantially. Interpreting motifs without knowing their original state may thus be more ambiguous and demanding.

Archaeological objects come with further challenges – they are highly damaged and sometimes even changed due to various post-excavation processes and past restoration treatments (see Fig. 2). Dealing with multiple fragmented objects presents another challenge. In that case, it is crucial to identify which fragments come from the same object in the first place. The tedious and time-consuming task is usually done manually by archaeologists and conservators, where they rely on their experience and expertise.

Automatic puzzle solving by a machine is an interesting computational problem that has been addressed by numerous researchers. While the majority of them use different computer vision and machine learning techniques to recover the photographs of different scenes, the use of machine learning for reconstructing cultural heritage artifacts has been less common. Manual reconstructions can be based on visual cues, such as identification of matching geometric and chromatic patterns while considering some constraints of plausibility: for instance,
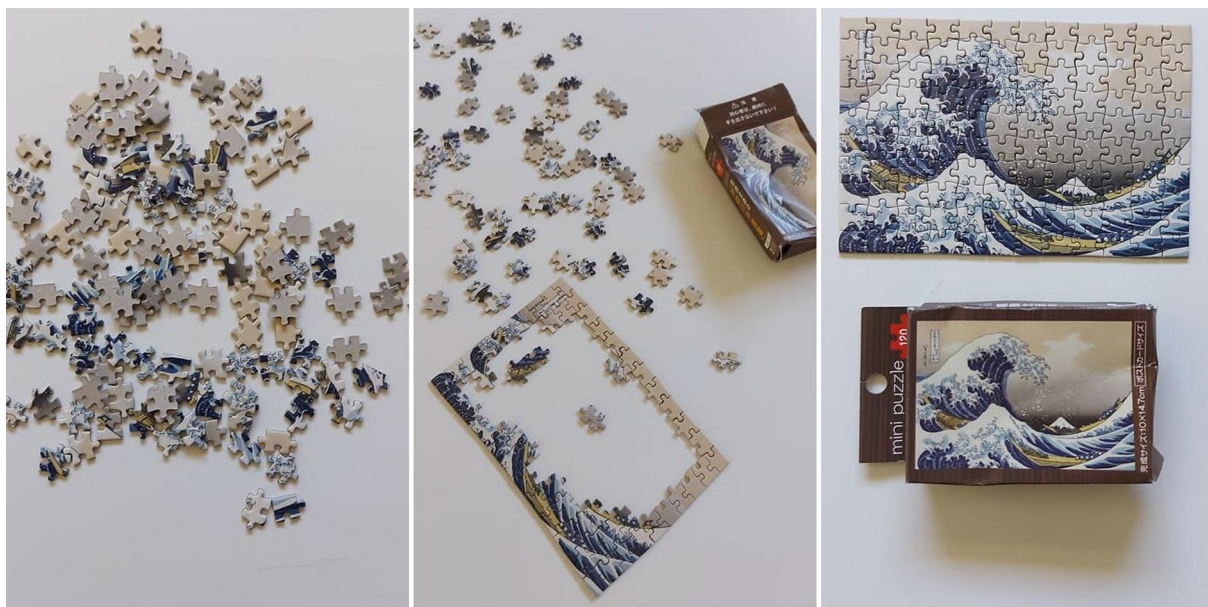

Fig. 1 The standard jigsaw puzzles we solve in daily lives. They usually have a ground truth image as a reference with only one possible solution, and well-preserved pieces make solving the puzzle feasible by comparing shapes, contour continuities, and colors among pairs of pieces. Photographs by Davit Gigilashvili
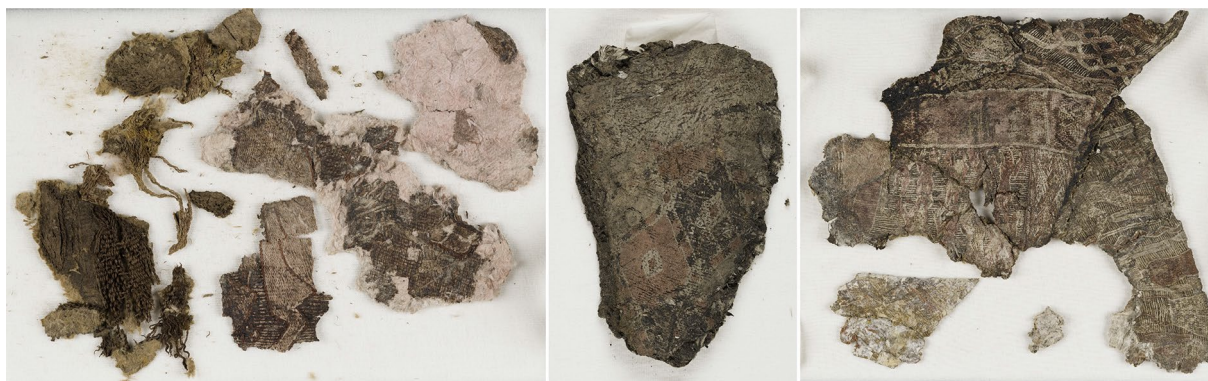
Fig. 2 The example of archaeological fragments of Oseberg tapestries [9]. Puzzle solving is extremely challenging, as the fragments are highly faded and degraded, they have irregular shapes, many fragments are missing. Moreover, some fragments, as the one on the very right, show former, post-excavation changes – here, for instance, sticking fragment edges by an adhesive, which may not reflect the arrangement in the original object. Photographs by George Alexis Pantos

ensuring that sky is above ground, human head is above its torso, or the sequence of events depicted in the artifact are consistent and meaningful, *"such as cooking being done before eating"* [7]. Another group of cues can be technical features given by the textile processing such as material and its quality, spin direction, warp and weft direction, thread count and others.

The first computational solution to the puzzle problem was conceived six decades ago [8]. A substantial progress in machine learning, and especially the emergence of deep learning within the past decade, has enabled solutions that were unimaginable before (see section *General Puzzle Solving Algorithms* below). Therefore, we want to overview the advances in computational techniques for puzzle solving and discuss how they could contribute to the specific problem of fragmented archaeological textile reconstruction. Virtual puzzle solving has another advantage, which is rarely discussed: the archaeological artifacts are usually very fragile, and development of virtual alternatives will limit physical interaction with them and, hence, facilitate their preservation. The contribution of this work is the following:

- We provide the overview of the different computational approaches to puzzle solving, both in general,

and more specifically, for cultural heritage reconstruction.
- We provide a comprehensive analysis of existing literature on puzzle solving specifically for textile artifacts.
- We analyze the existing knowledge gaps for virtual reconstruction of archaeological textiles and briefly discuss the potential future developments (see the bullet-point summary in the concluding part of the "Discussion" section).

The article is organized as follows: we start with a historical discourse on general puzzle solvers; afterward, we focus on puzzle solving applications in cultural heritage, and then, we discuss the works specifically on textiles. Finally, we discuss the results and draw conclusions.

Existing computational algorithms for puzzle solving

Puzzle solving has been a broadly explored topic in machine learning. However, most of the algorithms address natural images and only part of the works are intended for cultural heritage applications, out of which only handful are about archaeological textiles.

Table 1 Example queries and respective number of hits on Google Scholar search engine

Search query	Hits
"machine learning" AND "puzzle"	52600
("machine learning" OR "AI" OR "artificial intelligence") AND "puzzle solving"	8140
"machine learning" AND "puzzle" AND "cultural heritage"	1010
"machine learning" AND "puzzle" AND "cultural heritage" AND "textile"	77
"machine learning" AND "puzzle" AND "cultural heritage" AND "archaeological textile"	2

Gigilashvili *et al. Heritage Science* (2023) 11:259

Page 4 of 16

Table 1 illustrates the example queries used for article retrieval from Google Scholar search engine and the number of respective results (hits). While "machine learning" and "puzzle" return 52,600 articles, addition of "cultural heritage" to the filter reduces the number to 1010, "textile" to 77, and "archaeological textile" to only 2. Only the articles written in English were considered. The objective of this work was to provide an exhaustive review on puzzle solving for archaeological textiles, while also discussing non-exhaustive list of representative highly-cited articles introducing different, rather broad machine learning methods as well as those addressing specific sub-domains of cultural heritage to better understand the overall context, existing knowledge gaps as well as opportunities for archaeological textile reconstruction.

The fundamental problem in solving the puzzle is the fact that it is often impractical to try all combinations for reaching a global optimal solution, due to the immensity of required time and resources. Feasible methods, instead, should seek sub-optimal puzzling, and instead of exhaustive brute-force combinatorial approach, the solutions should utilize some exiting regularities to identify neighboring pieces. The first computer-based solution to jigsaw puzzles is a classic work by Freeman and Garder [8] from back in 1964. The work used apictorial, i.e. homogeneous gray pieces only, puzzled together solely based on contours of the pieces and considered the problem of up-to 9 pieces. The authors noted that false positives – matching contours that are ultimately false were one of the most significant challenges. The authors used a *chain-encoding* technique. A graph – called a *chain* – was constructed, which was a quantized version of piece's outline. From each node, there are 8 different ways to proceed to the next node (four orthogonal directions and four diagonals). Each of these directions were labeled with numbers, creating the chain code. Each chain was split into smaller *chainlets*. To consider two chainlets as matching, the feature vector of a given chainlet should have been as close as possible to the inverted version of another chainlet's one. Finally, the authors discuss two ways to assemble the puzzle: first is finding pairs of matching chainlets. This may create disjoint clusters and can mess up the whole process if there is wrong match at an early stage. The second, a more robust method is the algorithm that tries to find a matching pair for each junction, solving the puzzle incrementally.

The research has advanced significantly in the six decades since then. In addition to cultural heritage, it can have applications in broad range of problems, such as recovery of shredded documents [10–12], or fractured bones [13]. As well noted by Goldberg et al. [14], jigsaw puzzle is worth exploring on its own sake, as *"it is a natural and challenging problem that catches people's imaginations".*

General puzzle solving algorithms

The research has developed in multiple directions. According to Zhang et al. [15] fragmented image puzzle solvers can be classified in two distinct categories, geometry- and color-based ones. The former primarily relies on fragment contours, while the latter analyzes colors, to assess the likelihood of two given fragments being adjacent. While the majority of the modern works use both types of information [16–18], it took nearly three decades after the work by Freeman and Garder [19] to start utilizing color and appearance (texture, style) information. An alternative way to classify the approaches is whether these features in question are local or global. Some works rely on global features [20, 21] – fragment geometry and color distribution in the entire image, usually optimized by considering global geometric and color compatibility [14, 22]); while others rely on local features [23, 24] of a given piece, such as its shape, color, and texture that will be compared with its neighbors. State-of-the-art deep learning solvers may leverage both global as well as local features [25–27]. The examples of global-feature based methods are Growing Consensus, such as a work by Son et al. [21], and a Genetic Algorithm, such as the one by Sholomon et al. [20]. The former uses natural images and solves the puzzle of 432 pieces in 120 s. It deals with artificially created pieces as small as 7×7 pixels. It does not attempt to maximize the compatibility among all pairs of pieces, which may be misleading when pieces are small, because small pieces contain little information and may have little or no color variation. The method rather relies on geometric consistencies among different configurations of neighbors. The latter is designed to solve puzzles with very high number of pieces. It can solve a puzzle of 22 834 pieces in 13.19 h – which is the largest automatically solved puzzle to the best of our knowledge. The Genetic Algorithm approach is inspired from the evolutionary processes. The process starts with 1000 random pseudo-solutions – called chromosomes, i.e. random arrangements of the pieces. The evolutionary-inspired steps of selection, reproduction, and mutation are applied to iteratively improve the solutions from two "parent" solutions to a better "child" solution, which makes more sense in terms of expected regularities.

The examples of local methods are the works by Paikin and Tal [23] and Son et al. [24]. Both works use artificially fragmented photographs of natural images. The former is claimed to be robust even when the number and orientation of the pieces, as well as the number of original images, are unknown, and some pieces are missing. It identifies compatible neighboring pieces. The authors

Gigilashvili *et al. Heritage Science*   (2023) 11:259

Page 5 of 16

use a greedy algorithm. Greedy algorithm means that the algorithm usually makes decisions that are optimal at a given step at a given point in time, without considering whether this decision is optimal in the long term through the entire pipeline. Using a greedy algorithm creates a risk that mistakes at the initial stage may significantly compromise the eventual result. Therefore, only the matches of high confidence are accepted. They use the principle of *best buddies* introduced in [28]. This means that each piece independently identifies the other as the best match. The first piece is selected that has best buddies in all four directions; then comes the placement of other pieces with respect to the first piece and so on. Son et al. [24], on the other hand, start with a small loop of 4 pieces (2 × 2), created with dissimilarity metrics taken from previous works [29, 30]. The loop dimensionality gradually increases, eventually creating a N-dimensional ($N \times N$) loop in a bottom-up fashion using the dissimilarity metrics. In general, global features based techniques are viable when the jigsaw puzzle is very large or the size of pieces of puzzle is very small, because when the pieces of puzzle are very small, they don't contain enough spatial or local information that can be exploited to make a good prediction of their position; this is where the global geometry comes into handy, which is exploited by the global features based techniques. On the other hand, if the pieces are large and contain enough information, then local features based techniques work better in accuracy, and they are also more robust than global features based techniques. However, the methods that use local or global techniques work good with same source of pieces (same original image) in puzzle; they fail when the pieces of puzzles come from different sources with different objects and features, i.e. these techniques are not robust enough to work on every type of puzzles. And these techniques are also not robust to erosion and fragments loss that is quite common in real world. This is where we need techniques that use both local and global features to come up with more generalized techniques. Deep learning is one of them that leverages both type of features and creates more generalized puzzle solver models [25, 26].

Deep learning has revolutionized many applications of computer vision and image processing, and puzzle solving is no exception. Deep learning uses artificial neural networks that are inspired from the intelligence mechanisms in biological organisms. It has demonstrated an impressive performance in tasks such as identifying objects in the images [31]. Doersch et al. [25] used unsupervised learning for puzzle solving. Unsupervised means that the model automatically identifies patterns and arranges the data according to it without humans providing any labels. On the contrary, supervised methods depend on the labels, or "correct answers" provided by humans. For instance, if we want to teach a model to distinguish a cat image from dog images, we can show it a vast number of cat and dog images and tell which one is which, and the model will learn patterns characteristic to each of them. In unsupervised learning, we do not tell which image contains a cat and which once depicts a dog – the model captures itself that the patterns differ between the two groups. A neural network based model used by Doersch et al. [25] was trained to predict the position of a piece relative to the other piece when a pair of pieces was given to a neural network. They used images from Pascal VOC Dataset [32] and ImageNet Dataset [6] to create pieces to train a Convolutional Neural Network (CNN) based model [33]. To make the model more generalizable, they added jitter and gaps at the borders of the pieces to simulate missing parts.

Chen et al. [34] propose an interesting approach to puzzle solving – namely, puzzle solving is used as a Vision Transformer (ViT) component for image classification in natural images. Vision Transformers [35] are deep learning architecture that were primarily created for Natural Language Processing, but they have demonstrated the state-of-the-art performance in computer vision and image classification tasks [36]. Unlike traditional CNNs, ViTs do not contain a convolution layer and split images into fixed-size non-overlapping patches and use positional encodings to track their spatial locations. Since ViTs work on raw image patches, the authors anticipate it to work well for jigsaw puzzle solving. Their model, called Jigsaw-ViT, uses puzzle solving as a self-supervised auxillary loss during image classification. In other words, the task of the model is to solve a jigsaw puzzle with the available data, and its accuracy is used as a loss function. They removed positional embeddings not to have direct explicit cues to patch positions, and used patch masking, i.e. randomly dropping different patches to force the model to consider the global information. They conclude that inclusion of a jigsaw puzzle improved generalization and robustness of the image classifier.

Another deep-learning based alternative to traditional CNN architectures can be diffusion models that are primarily intended for generative tasks. Diffusion models gradually add noise to a sample and then learn to invert this process - after training, they are eventually able to generate samples from noise signals by denoising [37, 38]. Diffusion models turned out to be effective when puzzle solving is treated as a conditional generation process. In this case, the properties of a fragment are the conditions for generative process. Two solutions are of special interest: HouseDiffusion [39] and PuzzleFusion [40]. The objective of HouseDiffusion is generation of a floorplan from polygons constrained by graph where nodes

Gigilashvili *et al. Heritage Science*   (2023) 11:259

Page 6 of 16

correspond to rooms and edges correspond to links between rooms via doors. The room generation problem implies generation of polygons as neighboring rooms fitting with one another. Their coordinates are initialized with Gaussian noise and are gradually denoised. Here, only the outlines of the polygons are relevant. Unlike this approach, PuzzleFusion [40] uses diffusion models explicitly for jigsaw puzzle solving, where pictorial (considering the texture of the piece as in image reconstruction) and apictorial (considering just the outline of the piece, as in room generation) cases are addressed.

Finally, Chanda et al. [41] used clustering to group the fragments of shredded papers that belong together. They hypothesized that the fragments that come from the same original paper have similar color and texture. They converted the images from a device-dependent RGB to a device-independent CIELAB color space and supplemented color information by texture features extracted with Gabor filters. While colors are limited to specific pixel values, texture descriptors describe the patterns and regularities in the spatial distribution of colors (or intensities if it is grayscale). The authors constructed a feature vector that included both color and texture features, which was subsequently used for clustering. Clustering is an unsupervised machine learning methodology that groups data into clusters based on specific features. Clustering algorithms try to minimize intra-cluster differences and maximize inter-cluster differences, i.e. group samples with similar features together and place the samples with different features into different clusters. Chanda et al. [41] reported that the accuracy of their algorithm was more than 97%. We will return to this work later in the context of textiles. Two works by Kleber, Diem, and Sablatnig [42, 43] analyze the snippets of torn documents that are fragmented either intentionally by humans for privacy concerns or by historical degradation over time. They analyze different properties of a snippet to identify matching ones: skew or rotational analysis to identify the orientation of the fragment, and binary classification of the snippet into printed or handwritten categories – which is achieved by checking gradient orientations in each pixel; color analysis after segmenting the text from the background; line detection in segmented binary image using run lengths; and paper type analysis (checked, lined, blank) using Fast Fourier Transform. Ukovich and Ramponi [44] propose a clustering-based method for reconstructing pages of shredded documents cut in strips. The strips from the same page are grouped by broad range of features, such as line spacing, number of lines, presence of markers, paper color, ink color, text edge energy to capture the preferred directions of strikes etc. Some of those features, such as line spacing, is not suitable for handwritten text.

It is worth noting that when a general puzzle solver deals with natural images or photographs of well-preserved heritage objects, the images are usually fragmented artificially to create the puzzle [20, 23–27]. To test the robustness of the algorithm some pieces may be removed or eroded [23, 25, 26]. The availability of the ground truth images gives the authors possibility to evaluate the performance in a reliable manner. On the other hand, the real puzzle problems with already fragmented objects and no ground truth exist in the cultural heritage domain [45–51] or when dealing with shredded documents, as in [41]. Unlike natural images that are fragmented digitally, digitization of these real-world fragments is needed, which itself adds a complexity to the problem [49, 51, 52]. Although the intended final use is for real cases devoid of ground truth, some degree of artificial fragmenting of real artifacts [48, 49, 53] or that of simulated datasets [54, 55] are often still needed to evaluate the performance of the method. Human annotations can be also used [48, 49, 52]. The algorithms are tested in different case studies, which may target real artifacts [41, 52, 56], simulated ones [57], or both [48, 49, 54, 55].

Puzzle solving for cultural heritage applications

Virtual reconstruction of cultural heritage is an important research problem for computer scientists due to its aesthetic and scholarly value. Not all cultural heritage artifacts that require virtual reconstruction are fragmented. Even whole objects can have smaller areas that are damaged, for example, due to stains or scratches.

Hyperspectral images have been successfully used for virtual stain removal from paintings [58, 59]. For example, Zhou et al. [58] identified the spectral bands that were least affected by stains. These bands exposed the areas that were covered by the stains in the RGB photographs. They used the Poisson editing method to reconstruct the stained areas. The method uses image gradient, i.e. spatial variation of intensities and tries to match this variation between the source and reconstructed images. Hou et al. [59] used maximum noise fraction (MNF) transform to calculate principal components of the hyperspectral data. Then they identified which principal components included the most information about the stains and skipped them in the reverse transform to produce a stain-free image.

Another important research problem is filling the damaged areas, such as scratches, which usually involves a combination of inpainting and texture synthesis [60, 61]. Inpainting involves filling small gaps based on the information available in the rest of the image, while texture synthesis means producing large repetitive textured regions from a small texture pattern (the examples of inpainting and texture synthesis

algorithms can be found in [62] and [63], respectively). For instance, Yamauchi et al. [60] separated high frequency and low frequency components of an image using discrete cosine transform (DCT). They used inpaiting techniques to fill the gap in the low frequency image based on the information available in the non-damaged parts of the image. Afterward, they synthesized a texture similar to that in the high frequency component and added inpainted low frequency and synthesized high frequency components together to produce the final result. Criminisi et al. [61] also used inpainting and texture synthesis to remove foreground objects and replace them with a plausible texture that mimicked the rest of the background. Further important direction in image processing is denoising and segmentation of the images with missing data (e.g. when the pixel values contain substantial amount of noise), such as [64].

Sometimes, however, cultural heritage artifacts are fragmented into multiple pieces and instead of simply inpainting to fill the gaps, solving the entire puzzle is needed to put the respective pieces together. As shown above in a non-exhaustive overview of the general puzzle solving algorithms, a broad range of approaches can be taken, and a substantial amount of literature exists on general puzzle solving. However, we are primarily interested in puzzle solving for cultural heritage applications.

Leitão and Stolfi [65–67] suggested simple 2D fragment outline matching for pottery, murals, and other fragmented artifacts. The primary objective of their approach was to *"find any pairs of curves that have long sections with similar shapes"*. They filter the curvature-encoded contours at different layers of detail. They first find matches at a coarser scale, and then gradually try to identify the best candidates for match at the increasingly finer levels of detail. The limitation of the work is the fact that it identifies only adjacent pieces and not those that have large eroded parts in between. However, the authors argue that identification of potential pairs can substantially decrease the complexity of the puzzle when multiple thousands of pieces are present. They acknowledge that erosions at the edges are the primary source of noise and suggest that 3D approaches could mitigate the problem, where not only surface contours but also the depth information is considered. One example of such 3D-geometry based method can be that of Igwe and Knopf [68]. 3D models of the fragments and the target model are needed in this approach. Similar fragments are grouped by clustering, while self-organizing feature maps (SOFM) are used to position and orient the fragments. If no ground truth target model is available, as it is the case in many archaeological applications, the authors propose

to retrieve a similar shape from the database and use that one instead.

Toler-Franklin [69] highlights that standard puzzle solving methods that use fragment contour and color information fail to successfully re-assemble damaged artifacts with missing pieces. The author instead proposes a multi-channel RGBN image for that purpose, where each pixel contains RGB color and N – surface normal information. They demonstrate the pipeline from acquisition (shape-from-shading paradigm for surface normal estimation) to matching and rendering, and the approach has been deemed successful for matching the three different sets of fresco fragments.

Unlike previous works, where photographs of natural scenes were addressed, Paumard et al. [26] use paintings from the MET dataset [70]. They try to solve 3×3 puzzle, where the fragments come from multiple items. They used 9 artificially created fragments with 96×96 pixels each. To simulate erosions present in real-world artifacts, they picked the fragments that were 48 pixel away from one another. They first extracted features using deep learning based model inspired by VGG-Net [71]. They assume that a central fragment is known. Afterward, they ran a binary classification based on the extracted features to identify which fragments belong to the same image as the central fragment. Subsequently, from the pool of the 8 fragments that belong to the same painting, the position of each of them was defined relative to the central fragment using a graph of possible reassambly scenarios. The solution was perfect only in 44% of the cases, while many homogeneous background pieces were often misplaced. In another work [72], they also assume that the central piece is known and generate a 8×8 matrix, where each row corresponds to a candidate piece and each column is a potential location relative to the center. Then they use a greedy algorithm, picking locations with maximal probability. In the follow-up work [27], the authors presented a method called Deepzzle that was tested on a broad range of images, such as paintings, engravings (geometrical engravings or text documents), artifacts photographed on a homogeneous background (clothing, tableware, pottery plates, sculptures etc.). They shortened graph processing time 1000 times, which enabled them to accommodate 8 additional fragments in less than 60 min. The authors provided more comprehensive analysis of different scenarios, such as: robustness against missing fragments, fragments from other photos of the same object, more fragments from other objects, and the case where it was unknown which fragment was central. They decided to tolerate the errors, where homogeneous patches were misplaced and the difference was not visually very noticeable. This led to the best performance for artifacts that were on a homogeneous background, and

Gigilashvili *et al. Heritage Science*   (2023) 11:259

Page 8 of 16

the worst performance for content-rich paintings. They demonstrated that while many pieces can be placed correctly, the perfect re-assembly is an extremely challenging task. Adding fragments from other images, missing fragments, and unknown central pieces all compromise the performance and decrease accuracy, since the number of possible solutions increases. Many inaccuracies were present in text document images. For instance, the algorithm could not separate the French and Italian languages. The authors conclude that the resolution is too low to capture this kind of high-level semantical details, and furthermore, a convolutional architecture learns large visual features and has a limited ability with fine-grain details. In another work, they tried to use overall image semantics to identify relative positions to the central piece [73], and eventually propose that *"if no pertinent reassembly is found, a more robust solution that combines contours, patterns, and semantics should be considered".* The latest work Alphazzle is based on a single-player Monte Carlo Tree Search and relies on deep reinforcement learning to iteratively consider the relationship between all fragments [74].

Machine learning has various applications in archaeology, including but not limited to dating, identification, and classification of artifacts [75–78]. As for fragment classification and reconstruction, 3D fragments, such as pottery artifacts have received most attention. A comprehensive review of pottery reconstruction is given in seminal reviews by Rasheed and Nordin, for 2D images [79] and 3D data [80], respectively. The methods for classification in 2D images is based on features such as Gabor Wavelet Transformation [81], Scale Invariant Features Transform (SIFT) [82], or Local Binary Patterns (LBP) [83] – clustered by K-nearest neighbors (KNN) clustering [83, 84]. Reconstruction is based on contours [85, 86], color [87], or a combination of the both [88].

De Lima-Hernandez et al. proposed several interesting approaches to puzzling damaged 3D archaeological artifacts [45–47]. The 3D Puzzling Engine [45, 46] registers the 3D ancient wall-decorated fragments based on surface normal coherence, and offers elements of full-automation, as well as user-input-based semi-supervised solutions. In their recent study, de Lima-Hernandez et al. [47] propose using a Generative Adversarial Network (GAN) to predict the missing decoration traces on broken heritage fragments, extending the texture information of the fragments and allowing for a more accurate estimation of fragment alignment.

Further applications include tile panels, where Rika et al. [89] introduce a puzzle solver that is based on genetic algorithm. They use deep learning to measure compatibility among potentially neighboring tiles.

The method uses high-level color and texture statistics, and despite challenges related to homogeneous tiles and degraded edges, provides accuracy of 82%, which authors claim is the state-of-the-art performance. They propose that accuracy can be improved by additional training data. Another work on ceramic reconstruction is Ceramic Fragment Reassembly System (CFRS) developed by Lin et al. [56], where fragments are segmented and candidate matching pairs are generated using curve matching. For final matching, both curve and color similarities are considered. Finally, puzzling is optimized by considering additional factors, such as overlapping area (fragments whose contours partly match, but this match causes large areas to overlap, are unlikely to be matching). The authors mention that future work should incorporate the knowledge on motifs, cultural features, and other metadata and high-level expert knowledge. In-painting and texture synthesis have also been used to predict the content beyond the edge of a ceramic fragment [54], which is subsequently used to calculate features and find a match by FFT-based registration.

Another problem the reconstruction solutions have been proposed for is assembly of the culturally significant heritage documents, such as papyrus, which often is fragmented and puzzle solving is needed to recover the complete text. Pironne et al. [53] utilized deep Siamese network, dubbed "Papy-S-Net", that was trained and validated on 500 fragments, and yielded 79% accuracy. The approach suffers from high false negative rate and needs more sophisticated pre-processing. In another work, Abitbol et al. [90] hypothesize that papyrus plants that the papers are made of contain unique thread patterns, which can be utilized to find a match. They developed deep-learning-based method to identify local thread-based features.

A recent study [57] investigates fresco reconstruction, which is a complex problem due to missing and damaged areas (sometimes half of the fresco is missing due to erosion), as well as many and mostly irregularly shaped fragments. They try to match hand-crafted key-points and use fresco and fragment local color histograms. The authors point out the need for ground truth in their approach. The paper by Derech et al. [55] focuses on the reconstruction of highly degraded 3D statues and frescoes of an arbitrary shape. The authors extrapolate fragments to predict how they would continue, and then search for transformations where two fragments overlap in the extrapolated parts only. Although the authors claim cutting-edge performance, the algorithm failed when large parts were missing in the center of the fresco. Enayati et al. [52] also extract semantic information from fresco fragments that will facilitate classification and reassembly of the fragments.

Gigilashvili *et al. Heritage Science*   (2023) 11:259

Page 9 of 16

A semi-automatic solution based on geometric features for 3D fragments have been proposed by Mellado et al. [51]. This approach runs in loop and takes into consideration a feedback of a human expert. The expert specifies relative positions of the two fragments by user interface; the computer subsequently uses the Iterative Closest Point (ICP) to estimate the potential contact surface and propose the new visualization to the human user. The semantic expertise of the user is a fundamental component for increasing the performance.

Textiles

While the puzzle solving for other types of cultural heritage, such as ceramics, clay pots, papyrus, engravings, and paintings, attracted relatively more attention, the works on textiles are rather few. Although some of the images tested from the MET dataset by Paumard et al. [27] are clothes, they occupy a small part of the overall image, and as mentioned by the authors, the convolutional model just captures large visual variations, not the fine-grain details that are characteristic for textiles.

Kodrič et al. [50] present an interesting case study of virtual reconstruction of two damaged heritage textiles from 18th century with non-invasive methods. They analyzed fibers using microscopy to determine their type and surface morphology. Afterward, they conducted technical analysis of the weave (yarn thickness; ground warp and weft thickness in mm; pattern weft thickness in mm; warp and weft density (i.e. yarns per cm) etc.). Obtaining such detailed technical information is currently one of the major problems in the puzzle solving process. They identified that both fragments were made of silk and had nearly identical weave structure. The stylistic and motif analysis also provided additional insight. After analyzing the technical properties of the weave, they eventually reconstructed the damaged areas using Adobe Photoshop and ArahWeave software. The damaged areas were restored by repeating the patterns present in the well-preserved parts. Eventually, they highlight the need for a large database of historical textiles with associated stylistic and technical features that would facilitate attribution and identification of the artifacts and will foster digital reconstruction and reproduction.

The recent work by Huang et al. [91] proposed a non-invasive methodology to classify the fiber material. They use hyperspectral imaging to capture the spectra of 25 different samples of 11 different textile materials of plant, animal, and synthetic origin. They used part of the images for training different models. The authors compared traditional machine learning classification algorithms, namely, k-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), and partial least squares-discriminant analysis (PLS-DA),

with a one-dimensional convolutional neural network (1D-CNN). However, this method may not be applicable for puzzle solving where an object is made of the same materials. It can be instead used for sorting the materials that differ substantially. Furthermore, the authors do not discuss degradation, which may limit its applicability to archaeological textiles. Another example of using machine learning for fiber analysis is the recent work by Rippel et al. [92]. Automatic panoptic segmentation and identification of animal fibers can potentially contribute to identification of textile fragments made of similar fibers.

Thread counting of painting canvases is a well-known problem in art forensics. As the paints usually cover the threads, the proposed approach uses X-Ray images [93], which makes warps and wefts more visible to a machine. The approach models the canvas as a sum of two sinusoids with orthogonal spatial frequencies (warps and wefts), and uses Fourier analysis for vertical and horizontal thread counting.

In terms of reconstruction of textile heritage objects, a recent work by Stoean et al. [94] used deep learning for inpainting in parts missing from the costumes. Considering the structural complexity and variation of motifs, the approach leaves substantial room for improvement. The authors proposed Generative Adversarial Networks (GANs) as a future work. Sun et al. [95] discuss the challenges related to inpainting for silk artifacts, which relies on previous information in the overall object and often fails to success when the patterns are unique and irregular. They propose a three-step process for virtual reconstruction of silk artifacts and claim the state-of-the-art performance in terms of structural similarity metric (SSIM). First, they pre-process the image to unblur and remove noise. Besides, damaged areas are identified by human experts and marked with green masks; afterward, they first reconstruct the missing structure lines by adaptive curve fitting and inverse distance weighted interpolation. In other words, they identify clear line trends that are discontinued due to damages and reconstruct these continuous lines. Finally, they use inpainting to reconstruct the remaining parts guided by already recovered line structure. This way they avoid erroneous fillings and line breaks, as well as blocking artifacts due to randomness of the inpainting. However, the clear structural trends need to be visible for this method to be successful.

Liu et al. [96] discuss the challenges related to reconstructing archaeological textiles. They point out that unlike rigid materials, textiles are flexible and more prone to deformations that considerably complicates the restoration process. Further challenge is the lack of unified standard for result evaluation. The authors propose a workflow based on Human-Computer

Gigilashvili *et al. Heritage Science*   (2023) 11:259

Page 10 of 16

Interaction (HCI) and demonstrate it on the example of the Chinese archaeological silk gauze gown found in the Mawangdui Han tomb. They use sketches to create patterns and virtual simulation to generate a 3D object. They reconstruct a virtual 3D version of the silk gauze gown using 2D images and in the process, convert back and forth between 2D and 3D spaces. They analyze the structure, color, fabric, and pattern of the artifact, try the 2D patterns on a 3D model, then unfold 3D to 2D for virtual stitching and pattern arrangement. For simulation, first human body and posture are modeled. Then 2D garment fragments are imported and stitched virtually, and then color and fabric are rendered for the final display. The evaluation is a complex and multi-faced problem. For this, they use the Analytic Hierarchy Process (AHP), which outlines the hierarchical index system of factors that need to be evaluated, such as overall shape, garment structure and garment fabric are primary indicators, which contain clothing silhouette, hem structure, and fabric color as secondary indicators, respectively. Each indicator is assigned a weight and then aggregated for the overall evaluation. The evaluation is done against a historical prototype and the process implies high degree of human expert involvement.

To the best of our knowledge, there are only two studies [48, 49] that explicitly address the puzzle problem for highly fragmented archaeological textile artifacts. Both of these works argue that due to high degradation of the archaeological textiles, missing fragments, and the unknown number of original items that the fragments come from, fully automated puzzle solving for archaeological textiles is unlikely to be achieved. The authors also discuss the problems associated with the lack of ground truth. First, there are not enough databases of such artifacts with available ground truth that could be used for training the models; and second, the lack of ground truth makes it challenging to evaluate the results and identify which methods work well. The authors create ground truth by splitting existing fragments into smaller pieces for training and evaluation purposes but point out to take the results with care, because false positives, i.e. match between the pieces cropped from different fragments is not necessarily wrong and may instead indicate the compatibility of their respective fragments. Both of these works point out that the foremost task is to identify which fragments belong to the same original item and propose clustering method to group similar fragments. The puzzle solving process is supposed to be finished by human intervention after it becomes clear which ones belong together. These works take the approach similar to Chanda et al. [41], where color and texture features are extracted from the photographs to conduct clustering.

Gigilashvili et al. [48] use highly fragmented and degraded Oseberg Tapestry from Norway as a case study [9]. The authors split the ultra high resolution photographs of the Oseberg tapestry fragments into patches of 200×200 pixels and extracted features by Opponent Color Local Binary Patterns (OCLBP) [97], Co-occurrence Matrices (CoM) [98], and AlexNet [33] convolutional neural network, which they fed to three different clustering algorithms: K-means, Mean-Shift, and Agglomerative Hierarchical clustering. They found two major clusters of low spatial frequency and high spatial frequency textures by traditional texture descriptors, which they recommend to take with care, because more complex variations among the fragments were not captured with these measures. The CNN also produced two groups—textures with the majority of the fragments, and a small group of homogeneous, faded ones. They highlight that the approach suffers from low number of training samples available. Their primary research question was to identify from how many original items the fragments come. Although the accuracy of the clustering was above 90%, they measured the accuracy based on false negatives only. Due to the lack of ground truth, they did not penalize for false positives, because if the patches from different fragments ended up in the same cluster, it could be an indication that the two fragments belonged together. Finally, the authors asked archaeologists to assess the clustering results. However, archaeologists did not identify valid trends in the results that could have shed more light to their research hypotheses.

Gulbrandsen [49] used color histograms and color moments as color features, Local Binary Patterns (LBP) [99] as a texture feature, and also the features extracted with VGG19 pre-trained deep convolutional neural network [100]. LBP measures statistical co-occurrencies of pixel intensities in grayscale images and captures spatio-structural information. VGG19 is a deep neural network architecture, and its pre-trained version is trained on more than one million images to detect the features of different complexities, from simple edges to complex object characteristics. It is widely used for image classification, such as identifying objects and animals. Afterward, K-means and Hierarchical clustering were conducted. The author reports three case studies of different complexities, where he used the images of fragmented household textiles that are in a good condition, virtually fragmented photographs of well-preserved (non-archaeological) heritage textiles (such as Tingelstad cloth [101]), and Oseberg archaeological tapestry studied by Gigilashvili et al. [48]. The ground truth was available for the first two cases but not for the latter. The accuracy for the cases that involved textiles in a good condition was high. Color and texture features performed better

Gigilashvili *et al. Heritage Science* (2023) 11:259

Page 11 of 16

for heritage textiles, while VGG19 features worked better for household ones (with nearly 100% accuracy). However, the performance dropped drastically when similar method was tested on Oseberg tapestry. Although there is no reliable ground truth information available to evaluate the performance, the results were compared with the archaeologists' hypotheses and found not to be aligned with them.

Gulbrandsen [49] noticed that the algorithm usually clusters fragments by their overall color. This was robust enough for household textiles and well-preserved heritage textiles, since the samples differed substantially in color. This is also what explains high accuracy of the work by Chanda et al. [41], where paper fragments came into noticeably different colors and textures (it is worth mentioning that Chanda et al. [41] used CIELAB color space, which is better aligned with the human visual perception, while Gulbrandsen [49] and Gigilashvili et al. [48] worked in RGB). However, the color is not a reliable feature to separate degraded archaeological artifacts. Gulbrandsen [49] argues that general feature extractors are not tailored to this very specific task and advocates for development of textile-specific feature extractors.

Interestingly, Gulbrandsen [49] also developed a software solution called *Artifact Assembly* that enables virtual manual puzzle solving that is intended to manually reconstruct the archaeological textiles without disturbing the fragile physical samples and support their preservation. The software includes color enhancement tools for better visualization to human users. However, the author points out that proper algorithms need to be developed for machine-assisted assembly that would significantly speed up the process.

Discussion

The computational solutions to the puzzle problem have been developed substantially since its inception six decades ago, and the state-of-the-art includes broad range of different approaches, from classical edge descriptors [8] to cutting-edge deep learning technologies [7], from attempting a fully automatic puzzle solver [20] to clustering for manual assembly facilitation [49]. Table 2 summarizes the alternative ways to categorize these approaches.

The results demonstrate that archaeological textile reconstruction is in its infancy and merits a rigorous future research. Virtual restoration of heritage has developed in several different directions. While some works attempt to remove stains, or inpaint smaller missing areas, others attempt to reassemble the artifacts that are fragmented into many small pieces. The fragment assembly may be based on 2D (image) or 3D (geometric) input data. The classification of above discussed articles is given in Table 3. Jigsaw puzzles often contain all necessary pieces that are in good condition. However, to increase the complexity of the computational problem and robustness of the algorithm, sometimes the noise, such as gaps and missing pieces, is artificially introduced, such as in [23, 72]. This is especially crucial for the cultural heritage applications, where the noise due to degradation and aging is inherently present. However, some approaches may introduce the noise artificially and test the algorithm on the artificial data, while others work directly on the actual fragmented heritage objects, which offers more realistic "in-the-wild" challenge. The classification of the works according to the noise type can be found in Table 4.

Table 2 There are multiple ways to group computational techniques for puzzle solving

The ways to classify computational solutions to the puzzle problem
Classical image processing **OR** Deep learning-based **OR** Combining the both
Using geometric features **OR** Using color and appearance features **OR** Using both
Using local features **OR** Using global features **OR** Using both
Using 2D data **OR** Using 3D data
Fully automatic **OR** Semi-automatic

Table 3 Classification of the heritage reconstruction algorithms by the type of the input data and the problem solved

Fragment reassembly (2D)	Fragment reassembly (3D)	Inpainting & texture synthesis	Stain removal
[27, 72–74, 79, 81, 82]	[65–69]	[60–63]	[58, 59]
[83–89]	[45–47, 51, 80]	[50, 94–96]	
[50, 53–57, 90]	[94, 96]		
[48, 49, 94, 95]			

Gigilashvili *et al. Heritage Science*   (2023) 11:259

Page 12 of 16

Table 4 The classification of the heritage reconstruction algorithms by the type of noise/fragmentation

Artificially introduced noise	Real Fragments
[27, 65–68, 72, 73]	[45, 46, 69, 81–88]
[49, 51, 55, 57, 74, 96]	[47–49, 53, 54, 56, 89, 90]

Table 5 Deep learning methods for puzzle solving

CNN-based	Diffusion-based	Visual transformer
[25–27, 48, 49, 72–74]	[39, 40]	[34]

Puzzle solving has impactful applications in the cultural heritage domain. However, a broad range of challenges, such as high number of pieces, missing pieces, degraded areas, homogeneous pieces with little structural information, or mixing pieces from different originals, complicate the problem and compromise the robustness of the algorithms (e.g. [26, 27, 48]). These problems are especially severe for archaeological textile artifacts that have undergone centuries of degradation. The research on computational reassembly of archaeological textiles remains in its infancy and is limited to a mere attempts of clustering to identify the number of original items, which leaves room for a rigorous research effort in the future, given that there is a need for such solution in the archaeological and conservation communities.

There is a broad range of machine learning approaches for solving the puzzle of fragmented cultural heritage artifacts, where deep learning-based methods have demonstrated the best performance. Table 5 provides the classification for deep learning-based methods. While the methods based on diffusion and visual transformers have demonstrated promising performance, they are relatively novel and have not yet been tested on complex, highly degraded artifacts. On the other hand, CNN-based methods suffer from a very significant limitation: when the fragments are missing or come from multiple initial items, the number of potential solutions increases and the performance drops substantially. This is relevant for archaeological textiles, where both of these limitations are usually present. The example of the works by Paumard et al. [26, 27] show that the task is very complex even when we are trying to solve a 3×3 puzzle of square patches with high quality photographs of well-preserved heritage objects, where the erosion is simulated by a simple gap with a fixed width, where all fragments come from the same initial image and the central fragment is known. Now imagine the complexity of the task when fragments come in irregular shapes and from unknown number of initial items, they are highly degraded, and unknown number of fragments with unknown sizes are missing. While convolution-based neural networks may have a good enough performance when high-level visual characteristics need to be captured, as demonstrated by Paumard et al. [27], their ability to capture fine details is limited. Puzzle solving for archaeological textiles needs a substantial amount of specialized training, which will consist of a broad range of high resolution textile images. Current deep learning models are often trained on general databases of natural images, such as ImageNet, that are primarily intended for classification tasks, such as distinguishing dogs from cats. The features extracted with these models turned out to be of little use for technical analysis of textiles, as shown by Gulbrandsen [49] and Gigilashvili et al. [48]. On the other hand, Kodrič et al. [50] proposed a database specifically for heritage textile classification. Training machine learning models from scratch on this kind of specialized database may lead to a significant breakthrough in textile classification. An alternative avenue is to automatize the technical analysis conducted by Kodrič et al. [50].

In comparison with solving jigsaw puzzles and puzzles on well-preserved paintings, solving a puzzle of archaeological textile fragments is especially challenging due to multiple reasons: first and foremost, it is not known whether the fragments at all belong to the same initial item; secondly, many fragments are missing; and thirdly, the surviving fragments are highly degraded and have irregular shape – and the degree of degradation can vary substantially among them. Furthermore, no big databases with known ground truth are available to train the models and evaluate their performance. As pointed out by Gigilashvili et al. [48], even if we artificially introduce ground truth by further fragmenting existing artifacts, the lack of ground truth in real archaeological problems makes assessment of accuracy extremely challenging, because it is difficult to tell the real false positives and fake false positives apart—i.e. when patches from different fragments end up in the same cluster, we do not know whether this is a mistake, or this is an indication that their respective fragments belong together.

While the macroscopic high-level features, such as motifs, can be detected by more generic machine learning algorithms, much of the features that are possible to be extracted from the surviving fragments and that are being used by the experts, are more low-level texture variations. However, Enayati et al. [52] have shown that part of the semantic information can be extracted from individual fragments too that can facilitate reassembly. Semantic information has been used in the analysis by Kodrič et al. [50] as well.

Gigilashvili *et al. Heritage Science* (2023) 11:259

Page 13 of 16

Kodrič et al. [50] demonstrated the robustness of yarn properties and weave technical analysis for reconstruction. It is also proposed by Gigilashvili et al. [48] as well as Gulbrandsen [49] that the future work should focus on automatic measurement of weave – such as thread count, thread diameter, twist and spin direction, and technique. This can develop in two directions: classical image processing and texture descriptors, where the existing work on X-Ray-based thread counting in painting canvas can be of use [93]; and supervised CNNs, which would require a large database of manually-labeled dataset of images with manually measured properties. Finally, features extracted by machine learning can be manually supplemented by other high-level semantic metadata, such as context and chronology of the discovery.

Additional information can be obtained using more sophisticated imaging techniques. For instance, Gulbrandsen [49] and Gigilashvili et al. [48] propose using hyperspectral imaging for a deeper insight into material chemical composition, and reflectance transformation imaging (RTI) for 3D structures; whereas Toler-Franklin [69] also proposes multi-channel 3D geometry imaging, which includes surface normal information.

Considering the limited accuracy of the cutting edge machine learning algorithms even for few fragments (as few as 3×3 patches) of well-preserved paintings, fully automated puzzle solving of archaeological textiles is highly unlikely to be achieved in near future, if ever. Many works on puzzle solving for cultural heritage applications state very explicitly that those solutions are not intended as a full substitute to human expertise, but they are rather intended for semi-automatic reconstruction to assist human experts by providing reasonable suggestions [27, 45–47, 49, 51, 56]. Current puzzle solvers rely on apparent motifs and contours and suffer in homogeneous areas. Many archaeological textile fragments lack high level motifs and look homogeneous, which would make accurate spatial positioning impossible. On the other hand, computational solutions have potential to substantially speed-up the calculations and hence, the puzzle solving process. In comparison with manual human labor, which is often hard to coordinate, computers provide an opportunity for massive parallelization that can considerably decrease the solving time. Parallelization efforts for speeding up the process should be one of the directions for future research.

To summarize the article, the main findings and conclusions of the literature analysis are as follows:

- General puzzle solvers, especially those based on deep learning, demonstrate promising performance on solving puzzles of artificially fragmented natural images.

- Less amount of work has been done on puzzle solving for heritage artifacts and especially archaeological textiles. The puzzle solver performance is often compromised due to missing or highly damaged pieces that come in irregular shapes and from unknown number of original objects.
- The overall pipeline of puzzle solving for archaeological textiles can be divided into three major steps: digitization, clustering and matching similar fragments, and virtual reconstruction including inpainting and placement in space. Each of these steps are complex and merit rigorous amounts of future work.
- The lack of ground truth for real archaeological puzzle problems complicates the evaluation of the results.
- The lack of specialized textile datasets with ground truth makes it difficult to train machine learning models. Development of such databases can be a substantial contribution on its own, since models trained on general natural image datasets have not shown promising results for archaeological textiles.
- Human experts rely on high-level motifs, yarn properties, and technical analysis of the weave that are measured manually. Automatization of these measurements can speed up the process.
- Semantic metadata, such as context and chronology of discovery, that are also used by human experts, can be also fed into the computational algorithms.
- In addition to photographs, some works have successfully utilized information on chemical properties and 3D geometry. Some authors proposed using hyperspectral data and reflectance transformation imaging (RTI).
- Overall, fully automated solutions for highly degraded archaeological textiles are unlikely to be achieved. The primary objective of the research is to facilitate work for human experts instead of substituting them.
- Computational techniques have potential of large-scale parallelization that can substantially speed up the process, and that itself is an interesting direction for future research.

Conclusions

In this work, we reviewed the state-of-the-art machine learning techniques for solving puzzles when an original image needs to be recovered from pieces. A particular focus was on cultural heritage artifacts, and more specifically, archaeological textiles. While the cutting-edge deep learning techniques enable puzzle solving to the extent that was unthinkable before, the approaches still suffer from very significant limitations, and the reconstruction accuracy when many pieces are missing and fragments

Gigilashvili *et al. Heritage Science*   (2023) 11:259

Page 14 of 16

from multiple initial items are mixed, is far from perfect. While many works address paintings, ceramic, and papyrus artifacts, the knowledge on archaeological textiles is extremely limited. Human experts rely on manually measured features, such as thread count, fibre thickness, twist and spin direction, weave/binding technique, and fibre material, to solve the puzzle, which requires a substantial time and effort. While more generic machine learning models fail to capture those features, we propose that the research can develop in two different directions: first, automatic measurement of the above-mentioned features that are currently measured or digitized by hand; and second, creating big labeled databases of specific archaeological textiles that would enable training deep learning models. Finally, since archaeological textiles are highly fragmented and degraded, we want to emphasize that fully automated puzzle solvers are unlikely to emerge in the near future, if ever, and machine learning should be seen as an assistance to human professionals not as a substitute to their expertise.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Mannering U, Gleba M. Textiles and textile production in Europe: from prehistory to AD 400. Ancient textile series, vol. 11. London: Oxbow Books; 2012.
2. Schneider J, Weiner AB. Cloth and human experience. In: Ivan Karp L, Merrill William, editors. Smithsonian series in ethnographic inquiry. Washington: Smithsonian Institution Press; 1989.
3. Barber EJW. Prehistoric textiles: the development of cloth in the neolithic and bronze ages with special reference to the aegean. Princeton: Princeton University Press; 1991.
4. Lukesova H. Fibres in heritage objects: identification and characterisation by imaging techniques. Bergen: The University of Bergen; 2021.
5. Shams GP. Some minor textiles in antiquity. Studies in Mediterranean archaeology and literature, Pocket-book Göteborg: Paul Åströms förlag. 1987;48:3–11.
6. Tímár-Balázsy Á, Eastop D. Chemical principles of textile conservation. 1st ed. Oxford: Butterworth-Heinemann; 1998.
7. Paumard MM. Solving jigsaw puzzles with deep learning for heritage. Cergy-Pontoise: CY Cergy Paris Université; 2020.
8. Freeman H, Garder L. Apictorial jigsaw puzzles: the computer solution of a problem in pattern recognition. IEEE Trans Electron Comput. 1964;2:118–27.
9. Vedeler M. The oseberg tapestries. Oslo: Scandinavian Academic Press; 2019.
10. De Smet P. Reconstruction of ripped-up documents using fragment stack analysis procedures. Forensic Sci Int. 2008;176(2–3):124–36.
11. Deever A, Gallagher A, In: Semi-automatic assembly of real cross-cut shredded documents. 19Th IEEE international conference on image processing. IEEE. 2012;2012:233–6.
12. Zhu L, Zhou Z, Hu D. Globally consistent reconstruction of ripped-up documents. IEEE Trans Pattern Anal Mach Intell. 2007;30(1):1–13.
13. Liu P. A system for computational analysis and reconstruction of 3D comminuted bone fractures. Charlotte: The University of North Carolina at Charlotte; 2012.
14. Goldberg D, Malon C, Bern M. A global approach to automatic solution of jigsaw puzzles. In: Proceedings of the eighteenth annual symposium on Computational geometry; 2002. p. 82–87.
15. Zhang K, Li X. A graph-based optimization algorithm for fragmented image reassembly. Graph Models. 2014;76(5):484–95.
16. Kosiba DA, Devaux PM, Balasubramanian S, Gandhi TL, Kasturi K. An automatic jigsaw puzzle solver. In: Proceedings of 12th International conference on pattern recognition. vol. 1. IEEE; 1994. p. 616–618.
17. Chung MG, Fleck MM, Forsyth DA. Jigsaw puzzle solver using shape and color. In: ICSP'98. 1998 Fourth International Conference on Signal Processing (Cat. No. 98TH8344). vol. 2. IEEE; 1998. p. 877–880.
18. Toyama F, Fujiki Y, Shoji K, Miyamichi J. Assembly of puzzles using a genetic algorithm. In: 2002 International Conference on Pattern Recognition. vol. 4. IEEE; 2002. p. 389–392.
19. Freeman H, Garder L. Apictorial jigsaw puzzles: the computer solution of a problem in pattern recognition. IEEE Trans Electron Comput. 1964;Ec–13(2):118–27.
20. Sholomon D, David O, Netanyahu NS. A genetic algorithm-based solver for very large jigsaw puzzles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2013. p. 1767–1774.
21. Son K, Hays J, Cooper DB, et al. Solving small-piece jigsaw puzzles by growing consensus. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 1193–1201.
22. Andalo FA, Taubin G, Goldenstein S. PSQP: puzzle solving by quadratic programming. IEEE Trans Pattern Anal Mach Intell. 2016;39(2):385–96.
23. Paikin G, Tal A. Solving multiple square jigsaw puzzles with missing pieces. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition; 2015. p. 4832–4839.
24. Son K, Hays J, Cooper DB. Solving square jigsaw puzzles with loop constraints. In: Fleet David, Pajdla Tomas, Schiele Bernt, Tuytelaars Tinne, editors. Computer vision-ECCV 2014. Berlin: Springer; 2014.
25. Doersch C, Gupta A, Efros AA. Unsupervised visual representation learning by context prediction. In: Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 1422–1430.
26. Paumard MM, Picard D, Tabia H. Image reassembly combining deep learning and shortest path problem. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 153–167.
27. Paumard MM, Picard D, Tabia H. Deepzzle: solving visual jigsaw puzzles with deep learning and shortest path optimization. IEEE Trans Image Process. 2020;29:3569–81.
28. Pomeranz D, Shemesh M, Ben-Shahar O. A fully automated greedy square jigsaw puzzle solver. In: CVPR 2011. IEEE; 2011. p. 9–16.
29. Cho TS, Avidan S, Freeman WT. A probabilistic image jigsaw puzzle solver. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE; 2010. p. 183–190.

30. Gallagher AC. Jigsaw puzzles with pieces of unknown orientation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2012. p. 382–389.

31. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. The history began from AlexNet: a comprehensive survey on deep learning approaches. arXiv Preprint; 2018. https://doi.org/10.48550/arXiv.1803.01164.

32. Hoiem D, Divvala SK, Hays JH. Pascal VOC 2008 challenge. World Literature Today;2009:24.

33. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.

34. Chen Y, Shen X, Liu Y, Tao Q, Suykens JA. Jigsaw-ViT: learning jigsaw puzzles in vision transformer. Pattern Recognit Lett. 2023;166:53–60.

35. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:1–11.

36. Bazi Y, Bashmal L, Rahhal MMA, Dayil RA, Ajlan NA. Vision transformers for remote sensing image classification. Remote Sens. 2021;13(1):3:516-19.

37. Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. Adv Neural Inform Process Syst. 2021;34:8780–94.

38. Nichol AQ, Dhariwal P. Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning. PMLR; 2021. p. 8162–8171.

39. Shabani MA, Hosseini S, Furukawa Y. Housediffusion: Vector floorplan generation via a diffusion model with discrete and continuous denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023. p. 5466–5475.

40. Hosseini S, Shabani MA, Irandoust S, Furukawa Y.: PuzzleFusion: Unleashing the Power of Diffusion Models for Spatial Puzzle Solving. Available from: https://sepidsh.github.io/puzzlefusion/pdfs/puzzlefusion.pdf.

41. Chanda S, Franke K, Pal U. Clustering document fragments using background color and texture information. In: Viard-Gaudin C, Zanibbi R, editors. Document recognition and retrieval XIX, vol. 8297. SPIE; 2012. p. 230–7.

42. Diem M, Kleber F, Sablatnig R. Document analysis applied to fragments: feature set for the reconstruction of torn documents. In: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems; 2010. p. 393–400.

43. Kleber F, Diem M, Sablatnig R. Document reconstruction by layout analysis of snippets. In: Bentkowska-Kafel Anna, Stork David G, Coddington Jim, editors. Computer vision and image analysis of art. Bellingham: SPIE; 2010.

44. Ukovich A, Ramponi G. Feature extraction and clustering for the computer-aided reconstruction of strip-cut shredded documents. J Electron Imaging. 2008;17(1):013008–013008.

45. de Lima-Hernandez R, Vincke S, Bassier M, Mattheuwsen L, Derdeale J, Vergauwen M. Puzzling engine: a digital platform to aid the reassembling of fractured fragments. Int Arch Photogramm Remote Sens Spatial Inform Sci. 2019. https://doi.org/10.5194/isprs-archives-XLII-2-W15-563-2019.

46. de Lima-Hernandez R, Vergauwen M. A hybrid approach to reassemble ancient decorated block fragments through a 3D puzzling engine. Remote Sens. 2020;12(1):16:2526-19.

47. de Lima-Hernandez R, Vergauwen M. A generative and entropy-based registration approach for the reassembly of ancient inscriptions. Remote Sens. 2021;14(1:6):1–27.

48. Gigilashvili D, Nguyen HT, Gulbrandsen CF, Havgar M, Vedeler M, Hardeberg JY. Texture-based clustering of archaeological textile images. In: The Proceedings of Archiving 2023 Conference. vol. 20. IS &T; 2023. p. 139–142.

49. Gulbrandsen CF. Reconstructing the original: machine learning puzzle assembly for matching archaeological textile fragments. Trondheim: Norwegian University of Science and Technology; 2023.

50. Kodrič Kesovia MM, Penava Ž, Jemo D. The story of two historical textile fragments: technical analysis and reconstruction of the lost textile pattern. Text Res J. 2021;91(23–24):2859–71.

51. Mellado N, Reuter P, Schlick C. Semi-automatic geometry-driven reassembly of fractured archeological objects. In: VAST 2010: the 11th international symposium on virtual reality, archaeology and cultural heritage. Eurographics Association; 2010. p. 1–6.

52. Enayati A, Palmieri L, Vascon S, Pelillo M, Aslan S, et al. Semantic Motif Segmentation of Archaeological Fresco Fragments. In: IEEE Proceedings of the International Conference on Computer Vision-Workshop E-Heritage; 2023. .

53. Pirrone A, Aimar MB, Journet N. Papy-S-Net: A Siamese Network to match papyrus fragments. In: Proceedings of the 5th International Workshop on Historical Document Imaging and Processing; 2019. p. 78–83.

54. Sagiroglu MS, Erçil A. A texture based matching approach for automated assembly of puzzles. In: 18th International Conference on Pattern Recognition (ICPR'06). vol. 3. IEEE; 2006. p. 1036–1041.

55. Derech N, Tal A, Shimshoni I. Solving archaeological puzzles. Pattern Recognit. 2021;119(108065):1–12.

56. Lin SY, Ku HJ, Ting CL, Chang RI, Wang YC, Ho JM, Content-based feature matching for fragment reassembly of ceramic reconstruction. In: IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). vol. 3. IEEE. 2014;2014:344–51.

57. Lermé N, Hégarat-Mascle SL, Zhang B, et al. Fast and efficient reconstruction of digitized frescoes. Pattern Recognit Lett. 2020;138:417–23.

58. Zhou P, Hou M, Lv S, Zhao X, Wu W. Virtual restoration of stained chinese paintings using patch-based color constrained poisson editing with selected hyperspectral feature bands. Remote Sens. 2019;11(1384):1–18.

59. Hou M, Zhou P, Lv S, Hu Y, Zhao X, Wu W, et al. Virtual restoration of stains on ancient paintings with maximum noise fraction transformation based on the hyperspectral imaging. J Cult Herit. 2018;34:136–44.

60. Yamauchi H, Haber J, Seidel HP. Image restoration using multiresolution texture synthesis and image inpainting. In: Proceedings Computer Graphics International 2003. IEEE; 2003. p. 120–125.

61. Criminisi A, Pérez P, Toyama K. Region filling and object removal by exemplar-based image inpainting. IEEE Trans Image Process. 2004;13(9):1200–12.

62. Esedoglu S, Shen J. Digital inpainting based on the Mumford-Shah-Euler image model. Eur J Appl Math. 2002;13(4):353–70.

63. Lockerman YD, Sauvage B, Allègre R, Dischler JM, Dorsey J, Rushmeier H. Multi-scale label-maps extraction for texture synthesis. ACM Trans Graph. 2016;35(4):1–12.

64. Tsai A, Yezzi A, Willsky AS. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. IEEE Trans Image Process. 2001;10(8):1169–86.

65. da Gama Leitao HC, Stolfi J. A multiscale method for the reassembly of two-dimensional fragmented objects. IEEE Trans Pattern Anal Mach Intell. 2002;24(9):1239–51.

66. Leitão H, Stolfi J. Information contents of fracture lines. In: The 8th international conference in Central Europe on computers graphics, visualization and interactive digital media '2000; 2000. p. 389–95.

67. Leitão HCDG, Stolfi J. Digitization and reconstruction of archaeological artifacts. In: Proceedings XIV Brazilian Symposium on Computer Graphics and Image Processing. IEEE Computer Society; 2001. p. 382–382.

68. Igwe PC, Knopf GK. 3D object reconstruction using geometric computing. In: Sarfraz M, Banissi E, editors. Geometric modeling and imaging–new trends (GMAI'06). IEEE; 2006. p. 9–14.

69. Toler-Franklin C. Matching, archiving and visualizing cultural heritage artifacts using multi-channel images. Princeton: Princeton University; 2011.

70. Ypsilantis NA, Garcia N, Han G, Ibrahimi S, Noord NV, Tolias G. The Met Dataset: Instance-level Recognition for Artworks. In: 35th Conference on Neural Information Processing Systems Datasets and Benchmarks Track; 2021. p. 1–12.

71. Liu Z, Li J, Shen Z, Huang G, Yan S, Zhang C. Learning efficient convolutional networks through network slimming. In: Proceedings of the IEEE International Conference on Computer Vision; 2017. p. 2736–2744.

72. Paumard MM, Picard D, Tabia H. Jigsaw puzzle solving using local feature co-occurrences in deep neural networks. In: 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE; 2018. p. 1018–1022.

73. Paumard MM. Semantic-based automated reassembly of heritage fragments. In: CeROArt. Conservation, exposition, Restauration d'Art. 12. Association CeROArt asbl; 2020. p. n.p.

Gigilashvili *et al. Heritage Science*　　(2023) 11:259

Page 16 of 16

74. Paumard MM, Tabia H, Picard D. Alphazzle: jigsaw puzzle solver with deep monte-carlo tree search. arXiv Preprint; 2023. https://doi.org/10.48550/arXiv.2302.00384.

75. Resler A, Yeshurun R, Natalio F, Giryes R. A deep-learning model for predictive archaeology and archaeological community detection. Human Soc Sci Commun. 2021;8(1):1–10.

76. Bickler SH. Machine learning arrives in archaeology. Adv Archaeol Pract. 2021;9(2):186–91.

77. Davis DS. Object-based image analysis: a review of developments and future directions of automated feature detection in landscape archaeology. Archaeol Prospect. 2019;26(2):155–63.

78. Grosman L, Karasik A, Harush O, Smilansky U. Archaeology in three dimensions: computer-based methods in archaeological research. J Eastern Mediterr Archaeol Herit Stud. 2014;2(1):48–64.

79. Rasheed NA, Nordin MJ. A survey of classification and reconstruction methods for the 2D archaeological objects. In: International symposium on technology management and emerging technologies (ISTMET) 2015. IEEE; 2015. p.142–7.

80. Rasheed NA, Nordin MJ. A survey of computer methods in reconstruction of 3D archaeological pottery objects. Int J Adv Res. 2015;3(3):712–4.

81. Li-Ying Q, Ke-Gang W. Kernel fuzzy clustering based classification of ancient-ceramic fragments. In: 2010 2nd IEEE International Conference on Information Management and Engineering. IEEE; 2010. p. 348–350.

82. Smith P, Bespalov D, Shokoufandeh A, Jeppson P. Classification of archaeological ceramic fragments using texture and color descriptors. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. IEEE; 2010. p. 49–54.

83. Makridis M, Daras P. Automatic classification of archaeological pottery sherds. J Comput Cult Herit. 2013;5(4):1–21.

84. Zhou P, Wang K, Shui W. Ancient porcelain shards classifications based on color features. In: 2011 Sixth International Conference on Image and Graphics. IEEE; 2011. p. 566–569.

85. Papaodysseus C, Panagopoulos T, Exarhos M, Triantafillou C, Fragoulis D, Doumas C. Contour-shape based reconstruction of fragmented, 1600 BC wall paintings. IEEE Trans Signal Process. 2002;50(6):1277–88.

86. McBride JC, Kimia BB. Archaeological fragment reconstruction using curve-matching. In: 2003 Conference on Computer Vision and Pattern Recognition Workshop. vol. 1. IEEE; 2003. p. 1–8.

87. Papaodysseus C, Exarhos M, Panagopoulos M, Rousopoulos P, Triantafillou C, Panagopoulos T. Image and pattern analysis of 1650 BC wall paintings and reconstruction. IEEE Trans Syst Man Cybern-Part A: Syst Humans. 2008;38(4):958–65.

88. Yasuhara S, Kato S, Kato S, Itoh H. An earthenware reconstruction system considering color pattern and shape similarities. Forma. 2005;19(4):413–25.

89. Rika D, Sholomon D, David E, Netanyahu NS. A novel hybrid scheme using genetic algorithms and deep learning for the reconstruction of portuguese tile panels. In: Proceedings of the Genetic and Evolutionary Computation Conference; 2019. p. 1319–1327.

90. Abitbol R, Shimshoni I, Ben-Dov J. Machine learning based assembly of fragments of Ancient papyrus. J Comput Cult Herit. 2021;14(3):33:1-33:21.

91. Huang J, He H, Lv R, Zhang G, Zhou Z, Wang X. Non-destructive detection and classification of textile fibres based on hyperspectral imaging and 1D-CNN. Anal Chim Acta. 2022;1224(340238):1–10.

92. Rippel O, Schönfelder N, Rahimi K, Kurniadi J, Herrmann A, Merhof D, Panoptic segmentation of animal fibers. In: IEEE International Instrumentation and Measurement Technology Conference (I2MTC). IEEE. 2022;2022:1–6.

93. Johnson DH, Johnson CR, Klein AG, Sethares WA, Lee H, Hendriks EA, thread counting algorithm for art forensics. In: IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop. IEEE. 2009;2009:679–84.

94. Stoean C, Bacanin N, Stoean R, Ionescu L, Alecsa C, Hotoleanu M, et al. On using perceptual loss within the U-net architecture for the semantic inpainting of textile artefacts with traditional motifs. In: 24th International symposium on symbolic and numeric algorithms for scientific computing (SYNASC). IEEE; 2022. p. 276–83.

95. Sun X, Jia J, Xu P, Ni J, Shi W, Li B. Structure-guided virtual restoration for defective silk cultural relics. J Cult Herit. 2023;62:78–89.

96. Liu K, Zhao J, Zhu C. Research on digital restoration of plain unlined silk gauze gown of Mawangdui Han Dynasty Tomb based on AHP and human-computer interaction technology. Sustainability. 2022;14(1):14:8713-19.

97. Mäenpää T. The local binary pattern approach to texture analysis: extensions and applications. Oulu: University of Oulu; 2004.

98. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Trans Syst Man Cybern. 1973;6:610–21.

99. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit. 1996;29(1):51–9.

100. Mascarenhas S, Agarwal M. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In: 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON). vol. 1. IEEE; 2021. p. 96–99.

101. Pedersen KA, Vedeler M. Kledet fra Tingelstad-En bibelsk tegneserie fra 1500-tallet. In: Stige M, editor. Gamle Tingelstad -liten kirke med stort innhold. Oslo: Gran kirkelige fellesråd / Instituttet for sammenlignende kulturforskning; 2020.

Publisher's Note

Davit Gigilashvili　　received his BSc degree in Information Technologies from the University of Georgia (2015), MSc in Color and Applied Computer Science from Jean Monnet University, University of Granada, and NTNU (2017), and PhD in Computer Science also from NTNU (2021). Currently he is a postdoctoral researcher at NTNU's Department of Computer Science. His research interests include material appearance and virtual reconstruction of cultural heritage.

Hana Lukesova　　is a Senior Textile Conservator at the University Museum of Bergen. She has earned a doctoral degree from the Faculty of Natural Sciences at the University of Bergen in the field of fibre analysis on heritage textiles. She has long time worked on the cutting edge of conservation, art history and archaeology. She promotes interdisciplinary collaboration and contributes to the networking of textile heritage researchers across Europe as a member of a Management Committee of the international project COST EuroWeb. She co-established the community ComTex attached to the European Association of Archaeologists joining archaeologists and conservators.

Casper Fabian Gulbrandsen　　is a master's degree student in Computer Science at the Department of Computer Science at the Norwegian University of Science and Technology. He received his bachelor's degree in Computer Science in 2021. His research interests include the Internet of Things, machine learning, and web and mobile technologies.

Akash Harijan　　received his BSc degree in Computer Science from the National University of Computer and Emerging Sciences Pakistan (2018), MSc in Computational Color and Spectral Imaging COSI from NTNU, Jean Monnet University, and the University of Eastern Finland (2023). Currently, he is writing his master's thesis at NTNU Gjøvik. His research interests include color science and machine learning.

Jon Yngve Hardeberg　　is a Professor of Colour Imaging at the Colourlab at NTNU - Norwegian University of Science and Technology, Gjøvik, Norway. His current research interests include spectral imaging, image quality, colour management, material appearance, and cultural heritage imaging; he has co-authored more than 300 publications within the field.