Generating Realistic Underwater Images Using Contrastive Learning

Kazeem SHAMBA[a], Asgeir SØRENSEN[a], Oscar PIZARRO[a], Ricard MARXER[b], Hacene KARRAD[c], Nicola LUMINARI[c]

[a]Department of Marine Technology - NTNU, Trondheim, Norway

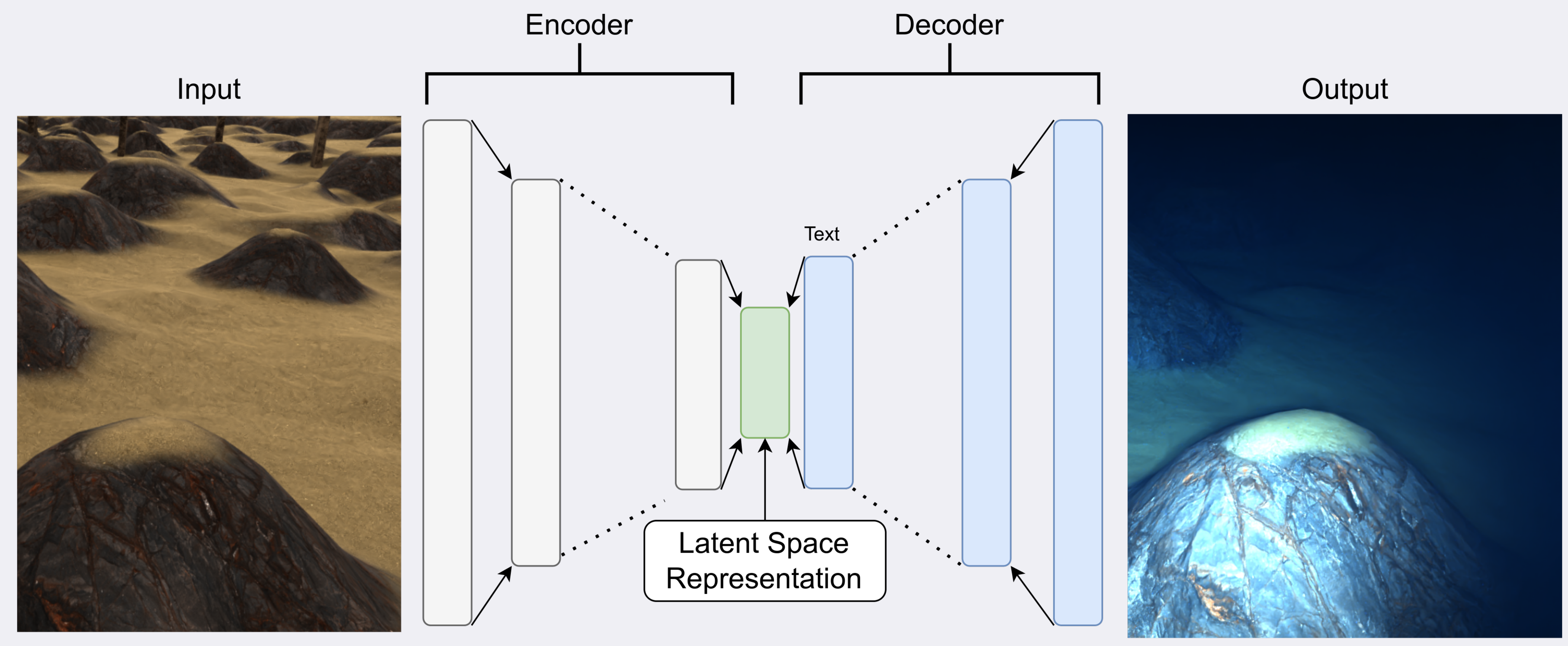
[b]Seatech - Marine Intelligent Robotics, University of Toulon, France [c]Data Science - Alteia, Toulouse, France

Co-funded by the Erasmus+ Programme of the European Union

Introduction

Most of the models we have today are supervised, therefore success or **good generalization** capability of the model, relies heavily on large amounts of labeled training data. Due to the cost, time, and complexity of providing annotations, as the natural world is **unannotated**, there is a need for a learning technique that can **improve the realism** of **synthetic data** from simulated environments. In this thesis, we investigate two major **image-to-image translation** techniques using autoencoders and **contrastive learning** for improving the realism of **synthetic underwater images** and compare our results with several baselines.

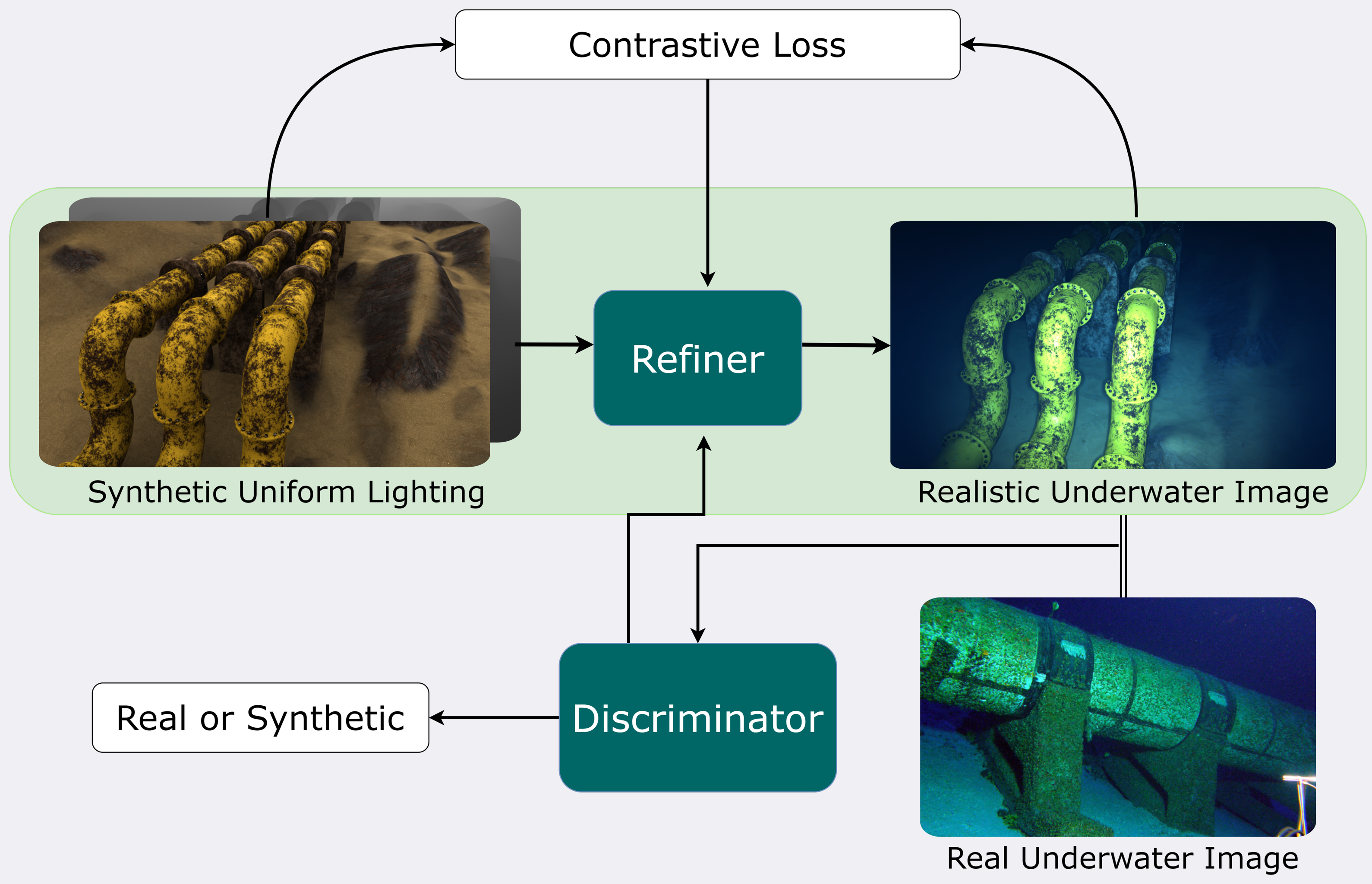
Autoencoder Network - Paired Image Translation



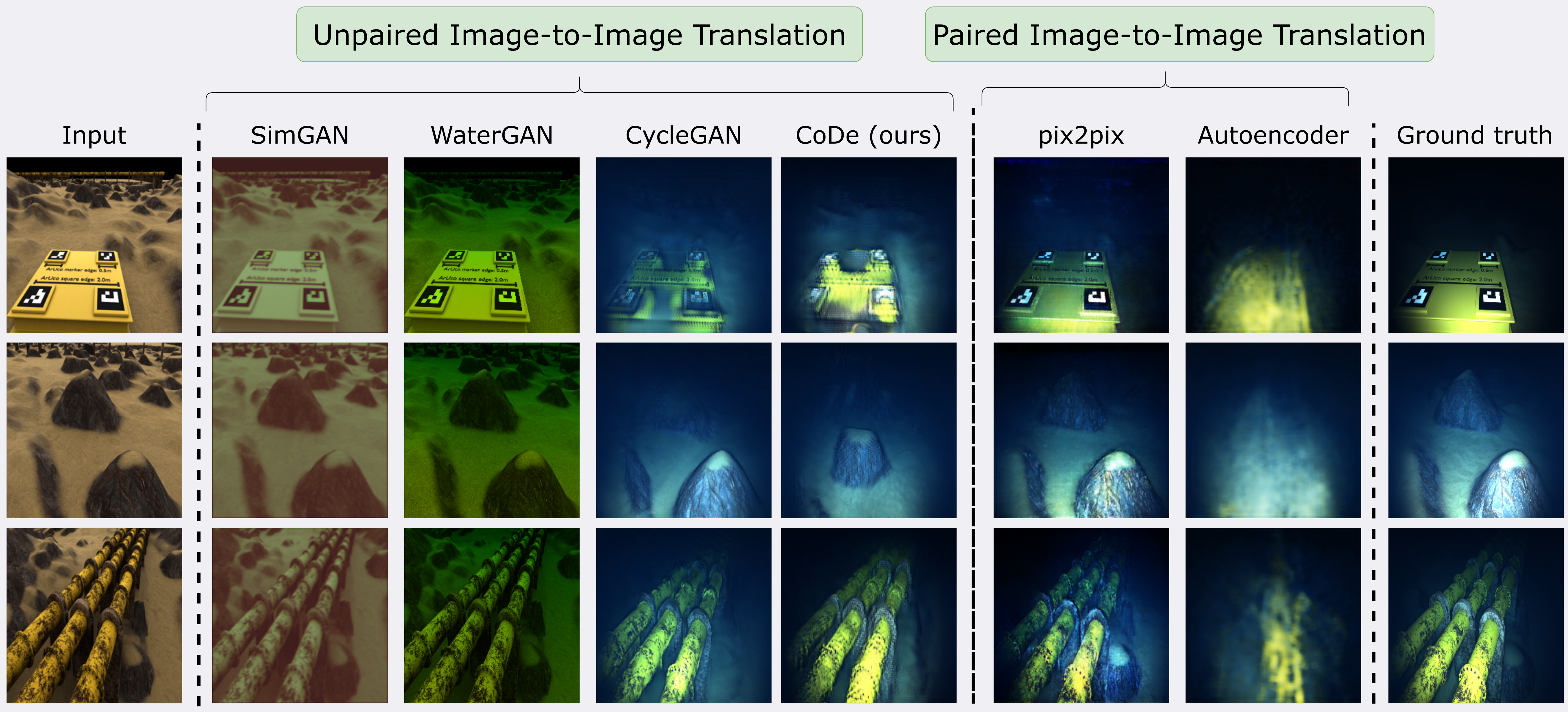
Methods

We explored both **paired** and **unpaired image translation**. First, we build a rather naive image translation model using an **autoencoder** and compare our results with those obtained using the SOTA **pix2pix** [1]. Second, we adapted the CUT [2] framework and adjust the input to account for an extra dimension of **depth information**. We call our method **CoDe (Contrastive+Depth)** for simplicity. The Figure in the next section shows an overview of our CoDe algorithm. We perform a comprehensive comparison with the state-of-the-art in unpaired image-to-image translation as shown in the Results section.

Contrastive Learning - Unpaired Image Translation



Results



Conclusion

In this work, we have explored the benefits of using contrastive learning to improve the realism of synthetic underwater images. We compared the performance with SOTA methods on image translation as shown above. From the above Figure, we derive the following conclusions

- **pix2pix:** This method produced the best results, however, it requires paired data which can be difficult to obtain
- **CycleGAN:** This unpaired image translation technique generated remarkable results
- **CoDe:** our CoDe model produced compelling results comparable with SOTA CycleGAN despite having lesser parameters and consequently lesser training time

References

[1] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2017, pp. 1125–1134.

[2] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," Springer, 2020, pp. 319–345.