Underwater Image Enhancement

Sonia Cromp (cromp)

11 March 2022

1 Introduction and problem importance

Underwater image enhancement is the task of improving the visual quality of images which were captured underwater. Faithfully capturing the underwater environment is notoriously difficult. For example, a remote operated deep-sea robot may explore the seafloor, combating a strong current which blows the robot around and stirs up sand or mud, while only diffuse light is able to reach the camera lens. Image enhancements may take the form of color correction, defogging, compensating for low light conditions and backscattering, or many other changes which allow the image to more accurately capture the appearance and details of the underwater environment[6]. For examples of various kinds of underwater image corrections, see figure at the end of this proposal from [6].

There are many reasons that we may wish to see underwater, from archaeology and marine biology research to piloting military submarines. A large proportion of the ocean is as yet unexplored, and documenting these areas along with the effects of mankind and climate change on the deep sea is crucial before any such information may be permanently lost. Further, from a computer science research perspective, underwater image enhancement is a challenging but interesting problem because it allows for the application of image style translation techniques from elsewhere in computer vision, as well as techniques from other areas of artificial intelligence more broadly such as natural language processing. These connections will be further explored in the following section.

2 Related work

A large body of work is centered on improving underwater image quality, which may be divided into two groups[1]. First, many approaches attempt to model and counteract the physical processes by which images become distorted. There are a variety of physical processes which degrade image quality, including the current blowing the camera around and the ways in which light is distorted in water. This first research direction contains many of the earlier contributions to underwater image enhancement, but it continues to generate interest and progress in recent times as well. For instance, to correct image coloring, Li and Cavallaro[7] estimate the non-uniform background illumination of an image's scene in order to compensate color wastage caused by the scene-to-camera ratio. To de-haze images, Chiang et al.[4] designed a framework that involves simulating the addition of an artificial light source shining on the contents of an image.

While these approaches achieve noticeably improved image quality in many situations, modeling underwater distortion faces several roadblocks. Accurately modeling distortion is challenging even for above-ground images, so it is no surprise that doing so for underwater images in varying current strengths that cause the camera to move, lighting conditions and proximity to the water floor (where sand and mud may be blowing around) can prove unreliable in many circumstances. According to

[1], the best quality models of underwater distortion processes are mathematically complicated and difficult to employ in image enhancement systems, while still failing to reliably model what is occurring in real life.

The second major group of approaches relies on machine learning to model the underwater distortion processes automatically. As in many areas of computer science, the presence of machine learning has steadily increased over time in underwater image enhancement. [1] provides a summary of dozens of ML underwater image enhancement architectures, which commonly consist of an encoder-decoder, multi-branch or generative adversarial network (GAN) structure. As one of the more approachable architectures, Sun et al.[11] enhance underwater images using an encoder-decoder pixel-to-pixel network, where the encoder and decoder contain three convolutional and three deconvolutional layers respectively. Interestingly, many ML approaches heavily borrow wisdom and domain knowledge from the first group of more physics-based works. UIR-Net from Cao et al.[3] estimate the transmission map and background lighting in an image via two independent networks, then use this information to restore the image and improve its visual quality.

Another interesting direction within ML approaches, which is highly related to the present project's aim, is dedicated to relaxing the restrictions on the type of required training data. Typically, ML training data consists of input examples with features X and labels Y. In the image enhancement domain, this would imply using un-enhanced images as the input features X and training a model to convert these inputs as closely as possible to their corresponding cleaned images Y. However, it is exceedingly difficult to create large datasets of paired unclean-clean underwater images. Humans are typically required to manually edit and restore the images, which requires time and expert knowledge that prevent the dataset generation from being handed off to crowd-sourcing platforms. This issue inspired Li et al.[8] to create UWGAN, which does not require unclean-clean image pairs. Instead, UWGAN uses a cycle structure comprised of a forward and a backward network, which learn the mapping functions between source (water) and target (air) domains. Then, the water-to-air function may be applied to an underwater image in order to minimize the underwater distortions. Another approach which does not require image pairs, MCycleGAN from Lu et al.[9] also aims to transfer image style from an underwater to a recovered style. The estimated depth map of a turbid underwater image serves as input to a generator, which then outputs a cleaned image. Then, after some processing, a discriminator attempts to differentiate between generated and manually-cleaned images.

Style transfer and unpaired training data are not unique to underwater image enhancement. Other work has applied style transfer to general image datasets[2], and uses of unpaired training data have even been explored in the area of machine translation for natural languages. For instance, [5] trains a machine translation network using two large, monolingual datasets (one for source and one for target language), along with a small parallel dataset of sentences presented in each language.

3 Project plan

Initially, efforts will focus on re-implementing prior state-of-the-art approaches, particularly those relating to style transfer[1] such as MCycleGAN[9]. Time allowing, it seems promising to also explore approaches from other areas such as machine translation. Borrowing from [5], a potential underwater image enhancement architecture may consist of an encoder and decoder, which are trained using a large dataset of unclean underwater images, a large dataset of clean images and a smaller dataset

[1] Unfortunately, the publisher for UWGAN[8] has not made the PDF available to the public, so it is necessary to rely on surveys and other works which reference it. For this reason, although the model is highly relevant, it may be challenging to gather enough information to re-implement it.

of images which have both clean and unclean versions.

4 Why there is need for a new approach

As previously stated, it is difficult to find large amounts of paired clean-unclean images in the domain of underwater image enhancement. As such, further exploration of unpaired approaches will enable training with more data and designing larger models which may capture the underwater distortion process more accurately. The two approaches from [8, 9] for unpaired data are quite recent and there has been little borrowing from other domains with more mature bodies of work on unpaired training. For this reason, the model described in the previous section merits exploration.

In addition, many models still rely on human domain knowledge to design architectures which specifically target and model particular distortion processes such as hazing and backscattering[1]. This may be partially necessary in underwater image enhancement due to the continued difficulties in training large models as a result of the small dataset sizes. So, training a larger model may lessen the need for human domain knowledge and allow a more general encoder-decoder architecture to model the distortion processes automatically. Aside from being more convenient for those lacking expert domain knowledge, it also allows one model to be trained to improve a wide range of distortions globally on an image instead of training one model to target hazing, another to perform color correction, etc. According to [1], this more general model is a worthy goal for future work.

5 Evaluation and timeline

For ease of comparison, evaluation will rely on the same metrics and datasets as some key papers evaluating UWGAN and MCycleGAN[9] used. [1] surveyed both of these models, along with many others, on the metrics of underwater color image quality evaluation (UCIQE)[12], which focuses on chroma, contrast and saturation, and underwater image quality measure (UIQM)[10] which measures image colorfulness, sharpness and contrast. Of course, qualitative evaluation of output images will also be performed.

The ideal timeline would look as follows:

- Replication (March): Training and testing pre-existing underwater image enhancement models
- Innovation (first two weeks of April): Implementing the proposed model
- Evaluation (last two weeks of April): Detailed experiments
- Summarization (May): Reviewing experimental results, preparing the website and presentation

References

[1] Saeed Anwar and Chongyi Li. "Diving deeper into underwater image enhancement: A survey". In: *Signal Processing: Image Communication* 89 (2020), p. 115978.

[2] Mu Cai et al. "Frequency domain image translation: More photo-realistic, better identity-preserving". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 2021, pp. 13930–13940.

[3] Keming Cao, Yan-Tsung Peng, and Pamela C Cosman. "Underwater image restoration using deep networks to estimate background light and scene depth". In: *2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI).* IEEE. 2018, pp. 1–4.

[4] John Y Chiang and Ying-Ching Chen. "Underwater image enhancement by wavelength compensation and dehazing". In: *IEEE transactions on image processing* 21.4 (2011), pp. 1756–1769.

[5] Vu Cong Duy Hoang et al. "Iterative back-translation for neural machine translation". In: *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*. 2018, pp. 18–24.

[6] Muwei Jian et al. "Underwater image processing and analysis: A review". In: *Signal Processing: Image Communication* 91 (2021), p. 116088.

[7] Chau Yi Li and Andrea Cavallaro. "Background light estimation for depth-dependent underwater image restoration". In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE. 2018, pp. 1528–1532.

[8] Chongyi Li, Jichang Guo, and Chunle Guo. "Emerging from water: Underwater image color correction based on weakly supervised color transfer". In: *IEEE Signal processing letters* 25.3 (2018), pp. 323–327.

[9] Jingyu Lu et al. "Multi-scale adversarial network for underwater image restoration". In: *Optics & Laser Technology* 110 (2019), pp. 105–113.

[10] Karen Panetta, Chen Gao, and Sos Agaian. "Human-visual-system-inspired underwater image quality measures". In: *IEEE Journal of Oceanic Engineering* 41.3 (2015), pp. 541–551.

[11] Xin Sun et al. "Deep pixel-to-pixel network for underwater image enhancement and restoration". In: *IET Image Processing* 13.3 (2019), pp. 469–474.

[12] Miao Yang and Arcot Sowmya. "An underwater color image quality evaluation metric". In: *IEEE Transactions on Image Processing* 24.12 (2015), pp. 6062–6071.

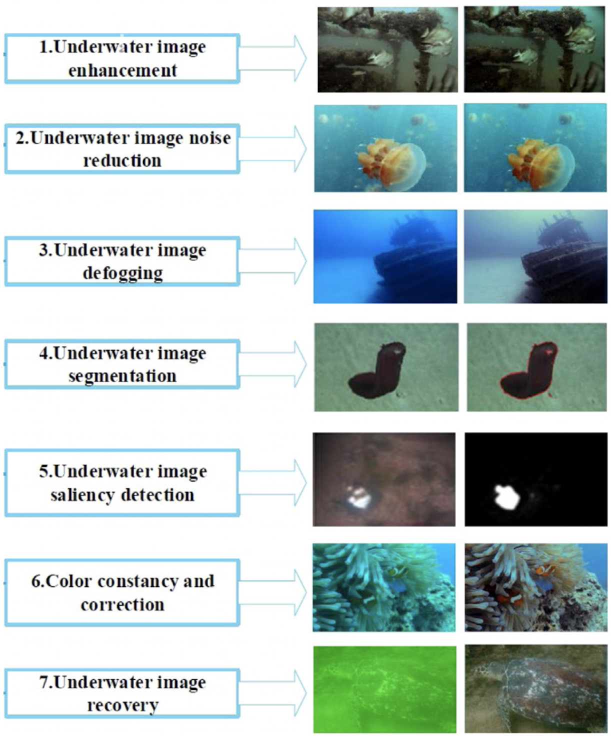

Figure 1: Categories of underwater image enhancement, from [6].