Seeing through Water: From Underwater Image Synthesis to Generative Adversarial Networks based Underwater Single Image Enhancement

Tomasz Łuczyński, Yvan Petillot, Sen Wang
Institute of Sensors, Signals and Systems
Heriot-Watt University
Edinburgh, UK
T.Luczynski@hw.ac.uk

Abstract—Underwater images notoriously suffer from exceptionally bad imaging conditions. Low light, attenuation, scattering and refraction-based distortions limit the space of applications of the underwater machine vision. Although solutions based on traditional image processing and deep learning have been proposed, they tend to be unreliable when facing various kinds of underwater images degraded by different lighting propagation conditions. The main contribution of this paper is three-fold: 1) A novel method is proposed to generate underwater-like images. It utilizes images taken in air and physics-based underwater image degradation model to create synthetic underwater datasets. 2) Preliminary results on the application of a Generative Adversarial Network architecture for single image haze and blur removal is reported. It is trained purely on the simulated data, yet demonstrates good generalization to the real underwater data. 3) A qualitative evaluation of underwater image enhancement algorithms covering contrast-based methods, advanced image processing algorithms and recent advances utilizing deep learning. This benchmarks state-of-the-art methods and outlines the strong and weak points of each approach.

Index Terms—Underwater vision, haze removal, deep learning, physics-based vision, simulation

I. INTRODUCTION

Underwater inspection missions are very demanding due to the hostile environment, lack of key sensors (e.g. GPS) and poor quality of data. Acoustic sensors work reliably but provide low resolution, grey scale, noisy data with limited update rate. Often visual inspection with cameras would be the best solution, as it provides colour data with high frame rate and resolution. This is important in tasks such as 3D reconstruction, manipulation and navigation in tight spaces. Furthermore images are easily interpreted by humans and there many algorithms for automated object detection, tracking etc. Unfortunately, optical underwater imaging suffer from exceptionally challenging conditions. There is usually little natural light and artificial illumination causes significant vignetting and requires recording in the high range of the intensity of light - shadows tend to be very dark and reflective surfaces can get very bright. Furthermore, light is attenuated and scattered (hazed). Generation of the haze and its influence on the image

is very similar to the effect that fog has on the images in air. However, underwater this effect is much stronger, more common and wavelength dependent. In multiple papers [10], [17], [18], [21] haze was identified as the main source of the underwater image degradation and therefore for the best results of any vision-based tasks, images should be enhanced to remove the veiling light. A deep learning based method was proposed to transform in-air images to underwater like [9]. In this paper a solution based on physicals model is proposed instead to generate the training dataset.

Underwater image enhancement has been studied in multiple papers before. Very common approach utilizes contrast enhancement techniques to reduce the influence of the haze [3], [23]. Other methods are based on more sophisticated approaches and often rely on some priors. Most notably, the Dark Channel Prior (DCP) [5] is known to give good results in air. Underwater adaptation of this method has also been proposed [10]. Another haze removal technique, known as haze lines [1], [2], is based on the assumption that the colour of the haze is known and no object in the scene is of that colour. Similarly to the DCP this method gives good results but can lead to significant image degradation when these assumptions are not fulfilled.

Finally, there are methods based on deep learning. These recently gained on popularity, but the success of these methods in the underwater image correction has been limited. Among others, [4], [15] have been proposed for haze removal. However these networks were designed and trained for removing haze (fog) for in-air images, which may not work for underwater. In [8] a network was proposed to enhance underwater images using colour transfer.

II. IMAGE FORMATION MODEL

Common mathematical description of the underwater image formation was introduced by Jaffe [6] and McGlamery [12] and is shortly summarized here for the sake of completeness. The signal forming an image is a sum of three components
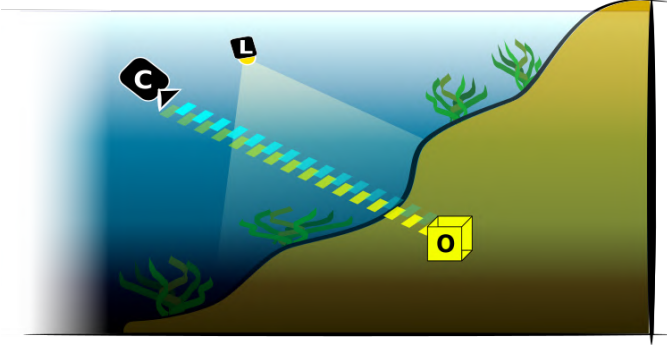
$$E_T = E_d + E_f + E_b \tag{1}$$

Fig. 1. Underwater image formation: the light is attenuated when traveling through the body of water, leading to the shift in colour. It is also scattered, causing blur and creating the foggy appearance. (L-light, C-camera)

E_T is a total radiance sensed by the camera. E_d is the direct transmission - attenuated radiance E_o reflected from the object. Attenuation is caused by the absorption of the energy by water as well as other particles suspended in it. Attenuation coefficients depend on the wavelength which causes the change of colour. This effect gets stronger with distance as more energy is absorbed through water. E_f is the forward scatter signal coming from the light reflected from the object and scattered on its way back to the camera and E_b is the backscatter caused by the ambient light scattered back to the camera. All of the components in this sum depend on the distance r and water parameters:

$$E_d = E_o e^{-c_\lambda r} \tag{2}$$

Where c is a total attenuation coefficient. Subscript denotes that c depends on the wavelength.

Scattering is a physical process where some form of radiation, in this case the light, is forced to deviate from the straight trajectory. In case of the underwater image formation, it is relevant in two ways. So called forward scattering occurs, when the light reflected from the object is scattered on its way to the camera. Reflected light ray influences the light rays reflected from the nearby points. This results in an effect very similar to the Gaussian blur. The main difference is that it's not uniform - the strength of the smoothing depends on the distance. The other effect is usually called backscatter. In principle this is the same physical phenomenon, only here the source of scattered light is different. The ambient light coming from the surface or the artificial lights is scattered. Part of this light is reflected to the camera. The more ambient light there is and the bigger the distance, the stronger the backscatter is.

$$E_f = E_d * g_r \tag{3}$$

Where g_r is a point spread function (PSF) and is parameterized with distance r. Multiple models of the underwater PSF have been presented in the literature [22].

$$E_b = B_{\text{inf}}(1 - e^{-c_\lambda r}) \tag{4}$$

Where B_{inf} is a background signal (a.k.a. global atmospheric light), i.e., what would be seen if the camera was looking into open water with no objects in front of it. Similar equations are true in air, only attenuation and forward scattering are much less prominent or often negligible. Backscatter may be quite significant, e.g. in the fog, but it is usually quite uniform. Underwater B_{inf} can vary significantly, with the rapid changes of the illumination in the field of view.

III. TRAINING DATASET

Training a deep neural network requires a large amount of data. Access to ground truth to compare it to the the output that the network is producing is also necessary. For this reason using real underwater images is difficult or impossible. Underwater datasets are very limited, they usually cover relatively small locations and therefore cannot be treated as proper representation of the full variety of conditions that can be found underwater. Moreover, images within these datasets tend to be very similar and therefore may cause over-fitting during training. For these reasons it was decided to use simulated images based on the NYU dataset V2 [14]. It's relatively big, versatile and, most importantly, includes the ground-truth depth for each image. This is important for the method described later on in this section. This dataset was modified to match the requirements of the underwater domain. First, an attenuation of colour was simulated. Even though colour will not be corrected by the neural network it is important to simulate this effect as well. Shift in hue can cause some colours, originally different, to look very similar, effectively reducing the contrast. Colours are attenuated according to the equation 2. Distance is known thanks to the depth image corresponding to the RGB image, provided in the NYU dataset. Once the colours are attenuated, the forward scatter is simulated by blurring the image with the Gaussian blur multiple times. It's equivalent to multiple blurs with increasing standard deviation of the Gaussian kernel. Then these images are fused: the further away from the camera, the more blurred image is used. Finally, the backscatter can be added. Following the equation 4, the global atmospheric light B_{inf} needs to be estimated. Underwater this might change significantly in the field of view and cannot be assumed to be uniform. To estimate the B_{inf} image two factors need to be considered: colour and light intensity. The colour of haze is strictly connected to the attenuation coefficients c_λ. Since these are selected randomly, the colour of the veiling light needs to be derived from these values. Firstly, the hue of the veiling light is specified: a white point is attenuated using the equation 2 and c_λ. An arbitrary distance $d = 10m$ is used, as assumed maximum visibility range underwater. In the next step RGB coordinates of the attenuated value are transformed to the HSV colour space and the H channel is stored as the hue of the veiling light. To get the saturation and value of the global atmospheric light image, an original image is heavily blurred with a Gaussian filter. This leaves just the bright and dark regions in the image. The blurred image is transformed to the HSV colour space and the H channel is set to the previously found hue. S and V
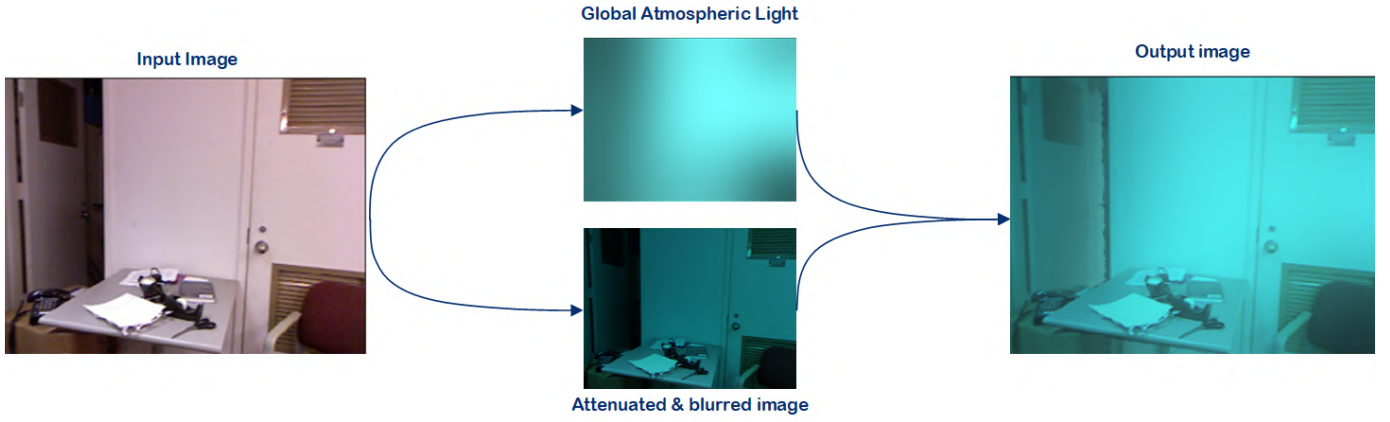
Fig. 2. Generation of the underwater-like images. Using the depth image the colours are attenuated and blurred. The non-uniform global atmospheric light is estimated using the input image and the attenuation coefficients.

Fig. 3. Created dataset. For each input image different versions of the underwater haze are created (top row). For each of those, the ground truth image is also computed (bottom row), so that the training may be focused on haze and blur removal, ignoring the shift in colour.

channels can be scaled for increasing or decreasing the amount of haze in the final image (this corresponds to increasing or decreasing the amount of ambient light in the water), compare Fig. 2). Once the B_{inf} image is specified the veiling light can be added, following the equation 4.

However, using these images for training with an original input image as ground truth, would require not only to remove haze and increase sharpness, but also to correct colour. Unfortunately usually there is not enough residual colour information and therefore this cannot be optimized accurately. Furthermore, the shift in colour appears (according to our preliminary tests) as the dominant image degradation factor in the optimization process, even though it is not the one that needs to be corrected the most. Therefore additional steps are performed to produce the ground truth images for training. Both original and produced, underwater-like, images are transformed into CIE L*a*b colour space. This separates the colour and information about the clarity/contrast. Then the

L channel in the simulated underwater image is substituted with the one coming from the original image. Afterwards it is converted back to the RGB colour space. This way the image retains the overall brightness level and colour but it is also fully dehazed and clear (see Fig. 3). This way a pair of hazed and ground truth images is created and can be used for training.

IV. PROPOSED SOLUTION

After some initial tests including various architectures (CNN, GAN, U-Net, ResNet) it was decided to use the GAN architecture for training. Generator takes a hazy image as an input and attempts to produce a haze-free version of this image. Both, hazy and dehazed images are passed to the discriminator, which processes both and assess, whether the image was correctly dehazed or not. Discriminator is also trained on the ground truth pairs of images. Please note that even though it is a GAN architecture, there is no "generative" component to this network. In a way it could be treated as a

case of supervised learning, where the discriminator serves as an adaptive loss function.
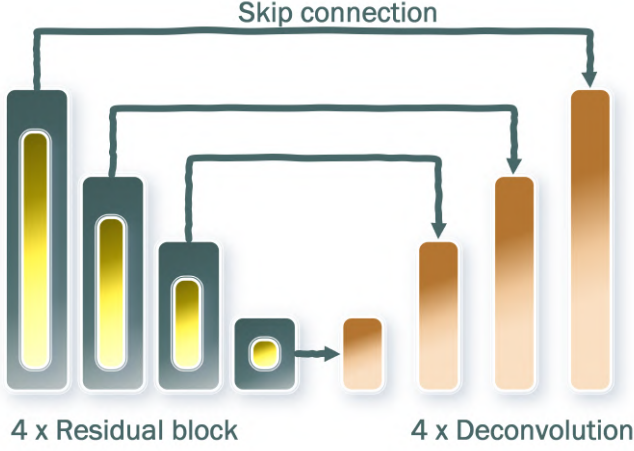
A. Generator

Fig. 4. Architecture of the generator: the classic U-Net architecture was modified to use residual blocks instead of just convolution layers.

The main task of the generator is to remove blur and haze from the image. In the case of the underwater images both are distance dependent their spacial influence on the image can vary, depending on the patch that is being analysed. Therefore Features extracted at different scales are important. A U-Net architecture [16] is known to preform well in such tasks. Furthermore, the skip connections guarantee that no high-frequency signal is lost in the processing. This is crucial, as the goal is not the synthesis of data but image enhancement. On the other hand, the convolutional networks based on the residual blocks were shown to perform well in the case of blur removal [20] and increasing the resolution [7]. We therefore decided to merge these two, proposing the Residual U-Net architecture. At the encored part of the standard U-Net architecture residual blocks are used (compare Fig. 4 and 5).
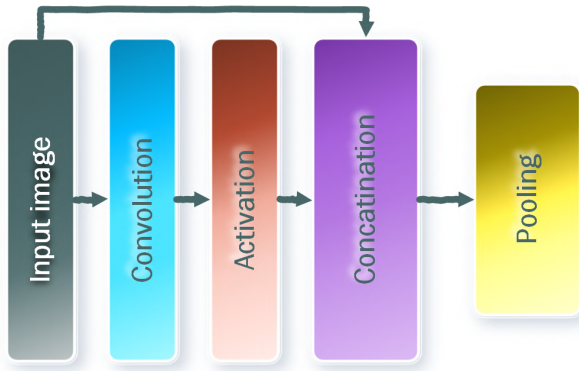

Fig. 5. Residual block used in the generator.

B. Discriminator

The discriminator has a relatively simple, convolutional architecture. Hazy and dehazed images are the two inputs. These are concatenated and processed through four convolutional layers. Finally, two fully connected layers are used to produce the verdict: weather the image was enhanced correctly or not.

V. EXPERIMENTS

The goal of the experiment is to compare the existing methods and evaluate their performance in a wide range of conditions. When selecting the methods used in the comparison, we ensured, that even though direct comparison between all available methods is not feasible, the best performing methods from the following groups were used. First, we included methods based on contrast enhancement and belonging to what could be called classic image processing techniques. The second group contains advanced image processing techniques using algorithms published recently and especially tailored to underwater haze removal. Finally, methods based on deep learning are presented. The last method in the comparison is the one proposed in this paper:

- CLAHE [23] - contrast limited adaptive histogram equalization - is a very well known method that is commonly used to enhance contrast in the images.
- Screened Poisson [13] - this method is based on a high pass filter. It amplifies the high frequencies in the image, also increasing the visibility of the details.
- UDCP [10] - this method represents a wide range of methods based on the Dark Channel Prior [5]. The method selected was modified to work underwater.
- Haze Lines [2] - similarly to the DCP, this method is based on a colour prior. However, the underlying assumptions made here are slightly different and do not rely on the local properties of the image within a given patch. The underwater version [2] of the original method [1] was used in the comparison.
- DehazeNet [4] - this algorithm is based on deep learning. It was developed to remove haze from images taken in air, which may influence its performance on underwater datasets.
- MSCNN [15] - another algorithm based on the deep learning and convolutional neural networks. Both methods focus on retrieving the transmission map rather than direct haze removal.
- Our preliminary results.

Results of the experiment are presented in Fig. 6. First of all it should be appreciated, that, with the exception of the last one, all images in the dataset were selected to pose a significant challenge to the algorithms and therefore the results presented are highlighting the limitations of the methods. Selected images have significant shif in colour towards both blue and green as well as heavy forward and backscatter components. Image J is, arguably, not hazed and is included to see how each method performs on the haze-free image and therefore test its robustness. All of these algorithms could produce much better
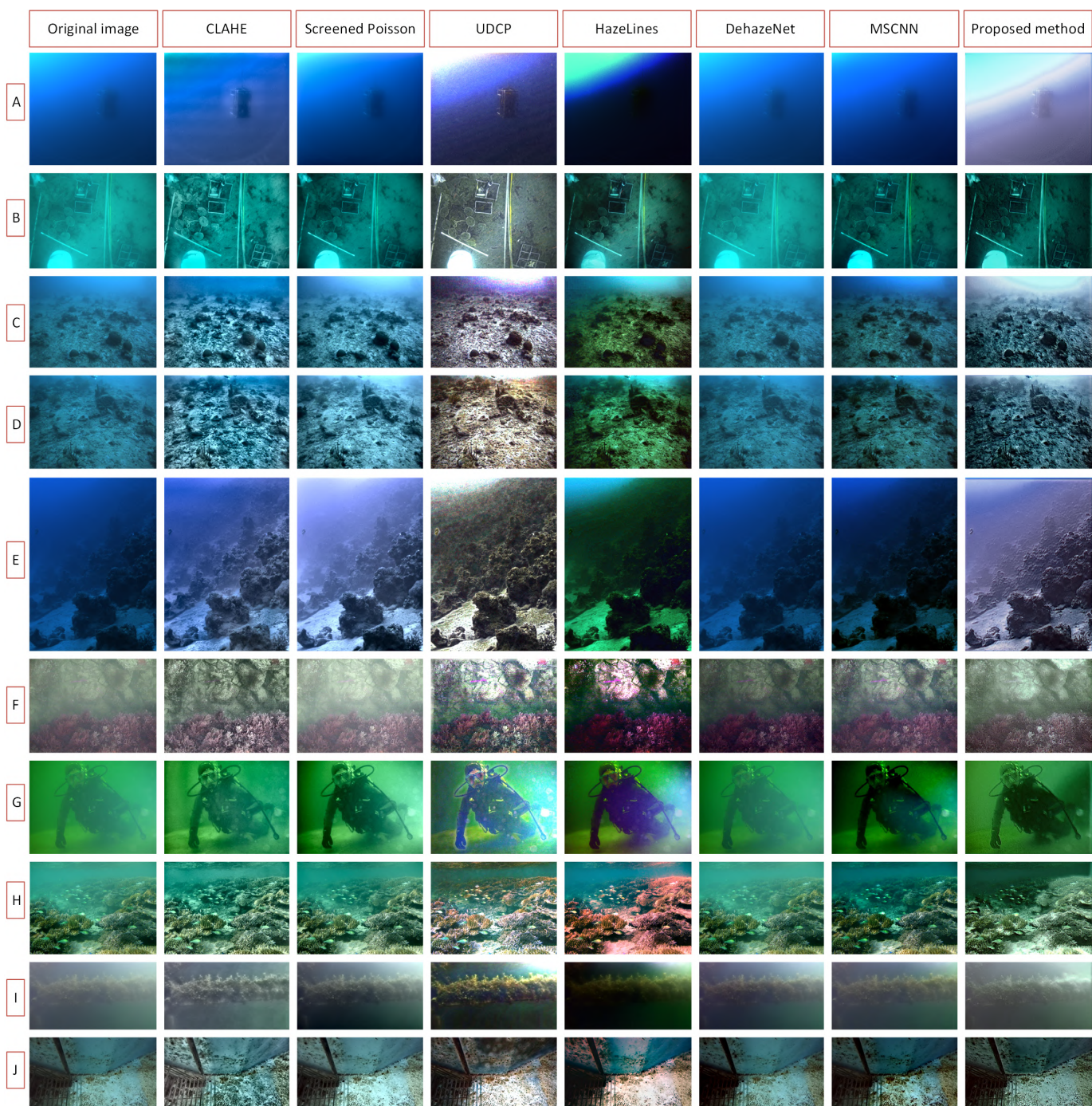
Fig. 6. Evaluation of the different methods for underwater single image haze removal. Competing results presented in columns, images come from the following sources: A and B - used in [10], shared by the authors; C and D - images from the publicly available dataset [19] (https://github.com/Breakend/AquaBoxDataset); E - image from [18], available on the project's website (http://webee.technion.ac.il/people/yoav/research/underwater.html); F - image from the underwater caves sonar data set [11] (https://cirs.udg.edu/caves-dataset/); G and H - images published as test data with the Matlab code of [2]; I and J - images extracted from one of our datasets. Image J was included as haze free to evaluate the robustness of the compared methods. Viewing with zoom on the screen is recommended.

results on more favourable conditions. However the goal of this experiment is to evaluate the performance and robustness of these methods, hence working on worst-case-scenario is desirable. CLAHE performed surprisingly well, given that it's used relatively rarely for underwater haze removal. It was tuned to enhance the images rather aggressively, which resulted in good recovery of the details. On the other hand some artifacts from the tiles used by this method remain in some images (e.g. Fig. 6, image A). Furthermore it may lead to amplifying temporary features, like flickering from the surface, especially when the haze is less prominent (see Fig. 6, image J).

Screened Poisson also aims at increasing the contrast, however the results are rather different. The effect is often much less visible than when using CLAHE. On the other hand it also never leaves any undesirable effects on the image. The downside of this method is that it relies on the processing in the frequency domain and calculation of the Fourier Transform may limit the speed of processing, depending on the hardware being used.

UDCP is the only method in this comparison, that (as part of its processing pipeline) also influences the colour of the image, hence the results presented here may appear to be significantly different from other methods. The haze removal is in this case very aggressive and often leads to very significant improvement in the images. However it may also lead to amplification of the noise or compression artifacts in the image. Finally this method shares the common weaknesses of the DCP-based methods. It relies heavily on proper estimation of the global atmospheric light. When this step fails or does not perform well the image can be significantly altered, to the point where it looks worse, than the input image. In the less significant case, this may lead to flickering in the video stream - see Fig. 6, images C and D. Both images were taken close to each other, both spatially and in time. Yet the first one has light purple hue and the second one is rather orange. This behaviour might be an issue in some cases. It is also important to mention that this method may require hardware acceleration to process images in real time, depending on the resolution of the image.

Haze Lines performed surprisingly poorly on this dataset. Similarly to the UDCP, this method depends on estimating light and water parameters. Errors in these estimations could explain the worse than expected performance observed. The original implementation published by the authors of this method was used. With some minor modifications the results could possibly be improved, but this lies outside the scope of this paper.

DehazeNet and MSCNN are two methods based on neural networks selected for this comparison. Even though these methods are inherently different, they share some background and performed very similarly. Both networks were developed and trained for removing haze from the images taken in air. This is probably the reason for the very poor performance on the underwater images. Some improvement may be noticed in the images taken close to the surface or close to the light source (e.g. Fig. 6, image F). This result also confirms that even though the basic hazing process is similar in air and underwater, its magnitude and influence on underwater images make it significantly more difficult to remove.

The proposed method presented overall good performance. The improvement was often not as significant of when using other methods, but it was very consistent across the test set and never caused unwanted image degradation. Given that this model was trained purely on the simulated data, it is a very successful result.

VI. Conclusions

In this paper three main contributions have been made. A new method for simulating underwater-like images has been proposed. Unlike many other methods it takes into account the forward scatter and non-uniform global atmospheric light. Base on this, a GAN architecture was trained to remove haze from the underwater images. Finally, the results of this method were evaluated in a qualitative tests, involving a wide range or competing methods using images from different datasets representing various challenging scenarios. Future work will focus on tuning the network and training parameters, to improve the results even further, as well as introducing a quantitative measures to assess the quality of achieved enhancement.

References

[1] D. Berman, T. Treibitz, and S. Avidan. Non-local image dehazing. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1674–1682, June 2016.

[2] D. Berman, T. Treibitz, and S. Avidan. Diving into haze-lines: Color restoration of underwater images. In *Proceedings of the British Machine Vision Conference*. BMVA Press, 2017.

[3] G. Bianco and L. Neumann. A fast enhancing method for non-uniformly illuminated underwater images. In *OCEANS 2017 - Anchorage*, pages 1–6, Sept 2017.

[4] K. J. C. Q. Bolun Cai, Xiangmin Xu and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[5] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, Dec 2011.

[6] J. S. Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2):101–111, Apr 1990.

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

[8] C. Li, J. Guo, and C. Guo. Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Processing Letters*, 25(3):323–327, March 2018.

[9] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation Letters*, 3(1):387–394, Jan 2018.

[10] T. Luczynski and A. Birk. Underwater image haze removal and color correction with an underwater-ready dark channel prior. *CoRR*, abs/1807.04169, 2018.

[11] A. Mallios, E. Vidal, R. Campos, and M. Carreras. Underwater caves sonar data set. *The International Journal of Robotics Research*, 36(12):1247–1251, 2017.

[12] B. L. McGlamery. A computer model for underwater camera systems, 1980.

[13] J.-M. Morel, A.-B. Petro, and C. Sbert. Screened Poisson Equation for Image Contrast Enhancement. *Image Processing On Line*, 4:16–29, 2014.

[14] P. K. Nathan Silberman, Derek Hoiem and R. Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012.

[15] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, 2016.

[16] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).

[17] Y. Y. Schechner and Y. Averbuch. Regularized image recovery in scattering media. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1655–1660, Sept 2007.

[18] Y. Y. Schechner and N. Karpel. Clear underwater vision. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I, June 2004.

[19] F. Shkurti, W.-D. Chang, P. Henderson, M. J. Islam, J. Camilo Gamboa Higuera, J. Li, T. Manderson, A. Xu, G. Dudek, and J. Sattar. Underwater multi-robot convoying using visual tracking by detection. In *Proc. of The IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2017.

[20] H. Son and S. Lee. Fast non-blind deconvolution via regularized residual networks with long/short skip-connections. In *2017 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10, May 2017.

[21] T. Treibitz and Y. Y. Schechner. Active polarization descattering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):385–399, March 2009.

[22] K. J. Voss. Simple empirical model of the oceanic point spread function. *Appl. Opt.*, 30(18):2647–2651, Jun 1991.

[23] K. Zuiderveld. Graphics gems iv. chapter Contrast Limited Adaptive Histogram Equalization, pages 474–485. Academic Press Professional, Inc., San Diego, CA, USA, 1994.