An Incremental improvement to underwater scene prior inspired deep underwater image and video enhancement: UWCNN++

Max Midwinter
University of Waterloo
200 University Ave., Waterloo, ON, Canada
mxxmidwi@uwaterloo.ca

Abstract

Images/videos collected from underwater are often severely degraded (e.g., colour distortion, low contrast, haziness etc.) This has a significant negative impact on the effectiveness of computer vision applications. In this paper the author proposes an incremental improvement to UWCNN+ proposed in Underwater scene prior inspired deep underwater image and videos enhancement by Li et al., 2020. The author proposes UWCNN++, which incorporates a novel stochastic image de-saturation augmentation step, inspired by image reconstruction methods, which makes UWCNN++ robust against artificial illuminations in images/videos. UWCNN++, like its predecessor, is a lightweight CNN architecture trained using synthetic underwater images generated from RGBD dataset and light degradation model based on water turbidity. To utilize UWCNN++ on real-underwater images a secondary hue-saturation-intensity (HSI) normalization is also applied. In this paper the author will compare the results of UWCNN++ against UWCNN+, and attempt to show improvement for real-underwater images with artificial illumination.

1. Introduction

Images or videos collected from underwater are often severely degraded (e.g., colour distortion, low contrast, haziness etc.) [1] This can have a significant negative impact on the effectiveness of computer vision (CV) applications such as object detection, recognition, feature extraction and feature matching algorithms. The performance of these computer vision methods is key to inspection, wildlife survey, mapping, re-localization etc. Therefore, an image correction method is desired to "de-water" (i.e., transform the underwater image such that it appears it was taken in air) underwater images.

Underwater image correction is a non-trivial task since light attenuates differently depending on its wavelength (e.g., the reason underwater images are blue green is that red light attenuates faster than blue and green wavelengths), suspended particles in the water absorb and scatter light resulting in haziness (which can vary based on the water's turbidity) and artificial illumination (e.g., flashlight) can greatly complicate the lighting of the image. Adding to the complication, is that there is very little ground-truth data available due to the Herculean task of draining the ocean. There are many methods that have attempted to address the image correction problem, however Li et. al, also identified the need for a real-time underwater correction method. Real-time processing is desirable since this will aid for the control of unmanned underwater vehicles (UUVs).

In the field of underwater image correction there are three popular approaches: hardware, image enhancement and image restoration. In this literature review various impactful methods of these approaches will be explored. Greater focus will be given to image enhancement methods and restoration due to limitations of the hardware approaches discussed in the next subsection.

1.1. Hardware Approaches

Generally, hardware approaches utilize various additional hardware beyond a single monocular camera to gather images from underwater. Some popular methods in this approach include using polarizing filters [2], range-gated imaging [3] and florescence imaging [4]. These methods exploit some physical property of the imaging hardware to extract additional information from the scene. For example, polarizing filters can reduce haziness in images, which then can be merged to recreate the underwater scenes [2]. The biggest drawback of hardware approaches for underwater image correction is the addition expense of the equipment. Due to the great expense of the equipment and data collection there have been few works that have collected real-world underwater images.

1.2. Image Enhancement

Image enhancement or non-physical model-based approach directly aims to process image pixel values to enhance specific image characteristics (e.g., colour, contrast and brightness) [12]. The intuition of image

enhancement is that underwater images, from similar environments assuming uniform ambient light, look the same: dim, low-contrast, blue-green and burry. This means that some image augmentation should be able change the underwater image as if it were taken in atmosphere. There have been many image enhancement methods proposed including a combination of traditional CV methods (e.g., filtering, histogram normalization, dehazing etc.) [7, 8], deep learning methods [10, 11] and recently hybrid traditional CV and deep learning methods [12].

Prior to the mass adoption of machine learning, various CV methods were proposed for underwater image enhancement. In [8] the authors proposed a weighted fusion method, which applies, in series, colour correction, denoising and contrast enhancement to the underwater image, which is then fused with weight measures at each pixel location. In [7], the authors propose a three-step method: dehazing, colour correction and bilateral filtering for noise reduction.

Deep learning methods have taken over the state of the art for many CV tasks. The most common deep learning approach being to use a convolutional neural network (CNN) to generate a result, that is compared to a ground truth to generate a loss, which is backpropagated through the network. (i.e., supervised learning) Then the challenge to deep learning methods in underwater image enhancement is the lack of ground truth data. (i.e., what are the restored underwater image supposed to look like?) To address the lack of ground truth data, deep learning underwater image enhancement methods are dominated by generative adversarial networks (GANs), specifically the CycleGAN inspired architectures. [10, 11] The CycleGAN architecture utilizes two sets of generators and discriminators, which simultaneously learns air to water and water to air image mapping. Deep neural networks may be slow to process an image, therefore deep neural network methods are sometimes used in conjunction with traditional CV methods. Examples of this include [12] and [21], upon which this work is based. In [21] the authors proposed UWCNN+ which utilizes a light neural network followed by HSI normalization. This 2-step process allows [21] to achieve a processing speed of 4.2 Hz, which the authors claim to be an order of magnitude faster than other algorithms in the underwater image correction domain.

1.3. Image Restoration

Image restoration methods or physical model methods are premised that it is possible using image cues to recover the model of light transmission, thus posing image restoration as an inverse problem. [12] This allows the user of the image restoration method to deterministically recover the undistorted/in-air image. The two popular methods in image restoration are dark-channel prior based [14, 15, 16] and medium transmission map estimation using the minimum information loss principle [17, 18]. The dark-channel prior is based on haze removal, where haziness is a function of depth to the camera, allowing for haze removal and depth map estimation. This property is exploited by [14, 15, 16], where the red-channel is chosen as the dark-channel as red light degrades the quickest underwater. Using the depth recovered from dark-channel prior, combined with estimation of ambient background light various methods determine the medium transmission map. The minimum loss principal method uses similar techniques to dark channel prior to estimate the transmission map but modifies it such that as many pixel values are preserved as possible. [17, 18] (i.e., modify the transmission map such that the most pixels fall between the range [0-255])

Deep learning methods have been proposed for image restoration of underwater images. As stated, prior, the main challenge for deep learning in the underwater domain is the lack of ground truth data. Since image restoration methods assume some degradation function, a popular technique is to use this function to generate artificial underwater images. [20] utilizes RGBD datasets to generate synthetic underwater images using a GAN framework, which are then used to train a U-Net style deep neural network. Like [20], [21] uses RGBD data to generate synthetic underwater images for training.

2. Methodology

This work heavily borrows from [21] and can be seen as an extension of the previous method. Like [21], there are 3 main components of the image enhancement framework: synthetic image generation, light-weight CNN and post-processing. In the following section these three components will be discussed in detail and how they have been changed relative to [21].

2.1. Synthetic Underwater Images

[21] utilizes the underwater image formulation model proposed in [13]. This popular underwater image degradation model is expressed as:

$$U_\lambda(x) = I_\lambda(x) \cdot T_\lambda(x) + B_\lambda \cdot (1 - T_\lambda(x)), \quad (1)$$

where, U is the underwater image, I is our target in air image, B is the homogenous global background light and T is the medium energy ratio, which represents the amount of light that will reach the camera. Finally, because image degradation is wavelength dependent, the subscript lambda denotes the wavelength of the light for which (1) is applied.

For some light wavelength lambda, the medium transmission map is defined as:

$$T_\lambda(x) = 10^{\beta_\lambda d(x)} = N_\lambda(d(x)) \quad (2)$$

where, β_λ is the wavelength dependent light attenuation coefficient for various water types, d(x) is the distance of a pixel to the camera and N_λ is the transmission map (ratio) in terms of the distance d(x). [1] The medium transmission map for the Blue Green and Red Channels are shown in Table 1, below, where Roman numeric are ocean waters and Arabic numeric are coastal waters of increasing turbidity.

Table 1: N_λ values for synthesizing underwater image types [21]

Types	I	IA	IB	II	III
Blue	0.982	0.975	0.968	0.940	0.890
Green	0.961	0.955	0.950	0.925	0.885
Red	0.805	0.804	0.830	0.800	0.750
Types	1	3	5	7	9
Blue	0.875	0.800	0.670	0.500	0.290
Green	0.885	0.820	0.730	0.610	0.460
Red	0.750	0.710	0.670	0.620	0.550

Applying equations (1) and (2), a random global background light on RGBD dataset, [21] were able augment in-air images to underwater images.

However, due to the uniform background light assumption, a model trained on this synthetic dataset is unable to handle more complex illumination schemes. (i.e., artificial illumination) To address this issue the author seeks a way to add artificial illumination to synthetic underwater images. The author takes inspiration from image restoration methods, where artificial illumination is masked by thresholding the saturation value of the image. [14] Specially in [14] the authors proposed a threshold of 0.4, where the normalized saturation below the threshold is segmented as artificial illumination. To replicate artificial illumination, random patches (circles) of low saturation are added to the synthetic underwater image. The detail implementation is such and RGBD image (I) is transformed to underwater image (I_{UW}). A mask (M) of the same dimension as image (I) is created for areas to augment to artificial illumination. I_{sat_norm} is saturation normalized and all saturation values are clipped to [0, 0.4]. Finally, the mask (M) is utilized to replace the appropriate sections of I_{UW} with I_{sat_norm}.

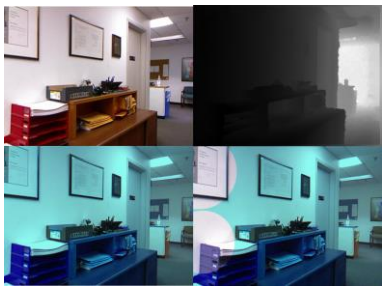


Figure 1: Synthetic UW Image. GT (top left), Depth Map (top right), Type1 UW Image (bottom left), UW Image with Proposed Random Artificial Illumination Augmentation (bottom right)

Figure 1 is an example of generating artificial training data for the low turbidity costal water (type 1). In the first row, are the RGB-D dataset, the second row shows the augmented image [21] and the proposed additional artificial illumination augmentation step. The following figure 2, shows what each turbidity augmentation looks like. It can be seen the image enhancement task is more challenging as the turbidity of the water increases.

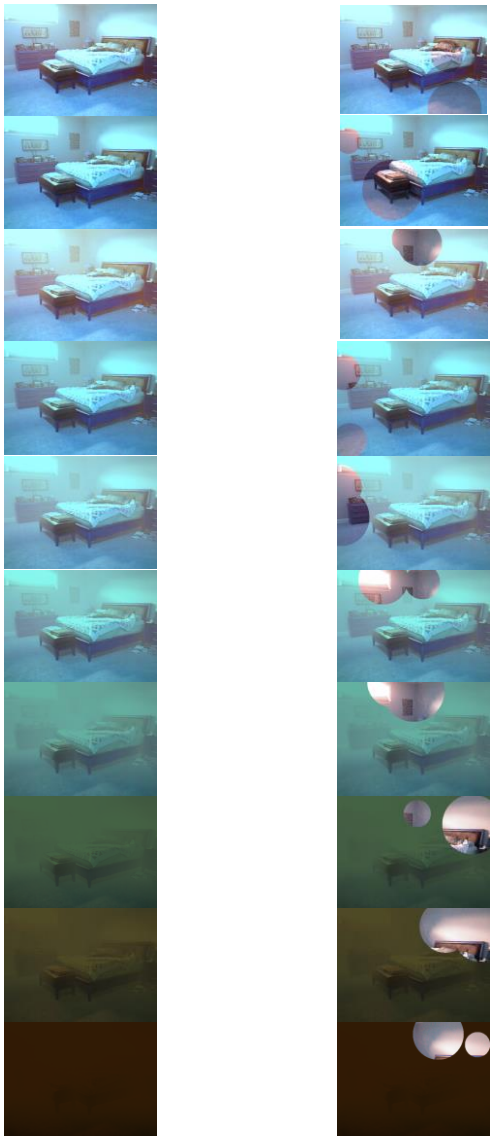


Figure 2: From Top (Type I, Type IA, Type IB, Type II, Type III, Type 1, Type 3, Type 5, Type 7, Type 9) From Left (UW Augmented, UW Augmented w/ Artificial Illumination)

2.2. Lightweight Enhancement Network

The UWCNN has the same architecture as was proposed in [21].

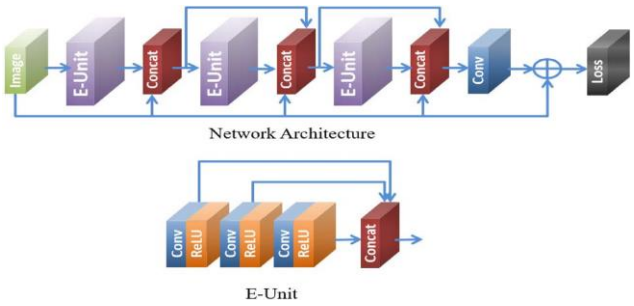


Figure 3: UWCNN architecture [21]

The architecture of UWCNN is a simple straightforward convolutional neural network featuring residual connections. The UWCNN network is composed of three pair of E-units (enhancement units) and concatenation layer. Each enhancement unit features 3 pair of 2D-convolution layers followed by ReLU activation, followed by a concatenation layer. The original image is concatenated in the concatenation layer along with filters from the previous concatenation layer. Finally, a 2D-convolutional layer is used to reduce the output filters to 3 and the input image is added to it, creating the final output.

The network generates a loss by comparing the output of the network with the un-augmented image. In [21] the authors utilize the combination of the ℓ_2 loss and SSIM loss. The ℓ_2 loss is defined as:

$$l_2 = \frac{1}{M}\sum_{i=1}^{M}\left|\left[U(x_i) + \Delta\left(U(x_i),\theta(x_i)\right)\right] - I^*(x_i)\right|^2 \quad (3)$$

where, inside the square brackets is the output from the CNN model and I* is the GT image. The SSIM loss is defined as the following:

$$L_{SSIM} = 1 - \frac{1}{M}\sum_{i=1}^{M} SSIM(x_i) \quad (4)$$

where SSIM is the structural similarity index measure, which measures the similarity of two black and white (BW) images at each pixel location for a given size patch ground the pixel. [21] claims the model is insensitive to c_1 and c_2 parameters and the patch size was set to a default of 13.

The total network loss is:

$$L = \alpha\, l_2 + \beta\, L_{SSIM} \quad (5)$$

where, α and β are coefficients of the loss. It was found that the ℓ_2 loss was an order of magnitude lower than L_{SSIM}, therefore in this study α = 10 and β = 1.

2.3. Post Processing

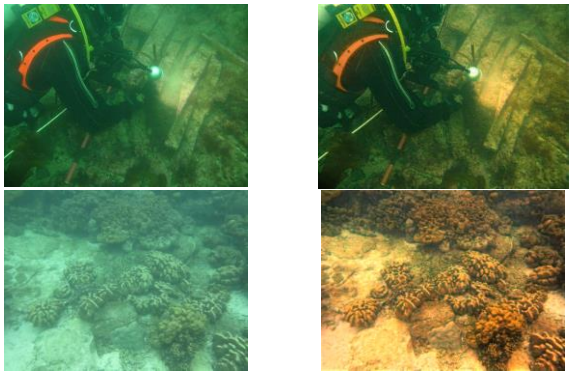
The UWCNN model is trained on synthetic images created from some RGBD dataset. The domain shift between the RGBD dataset and real-underwater images should be accounted for. In [21] this is accomplished with a saturation normalization step. This is accomplished by transforming the image to HSI colour space and min-max normalizing the saturation channel [0-1] and transforming the image back to the RGB colour space. The figure 4, below shows the difference normalization makes to the image.

3. Experiment

A UWCNN model was created for each turbidity β coefficient. The NYU-v2 RGBD dataset was used to create synthetic underwater images. Each image in NYU-v2 (1449 images) is used to create 3 pairs of underwater and underwater artificial illumination augmented images. This provides a training set of 8694 images. The model is trained for 40 epochs with the ADAM optimizer. UWCNN was reproduced for this work in TensorFlow 2 and was trained/validated run a custom desktop with Nvidia Titan V. Computational resources are provided by Computer Vision and Smart Structures (CViSS) lab, Civil and Environmental Engineering Department, University of Waterloo.

3.1. UWCNN++ Results

Due to time constraints, it was not possible for this paper to conduct an in-depth evaluation of UWCNN++ against its counterparts. In this paper the author will compare the enhancement results of UWCNN+ [21] with the here proposed UWCNN++. A series of challenging images, with and without artificial illumination, will be used to compare the two methods. This collection of real-underwater images is sourced from ImageNet dataset.



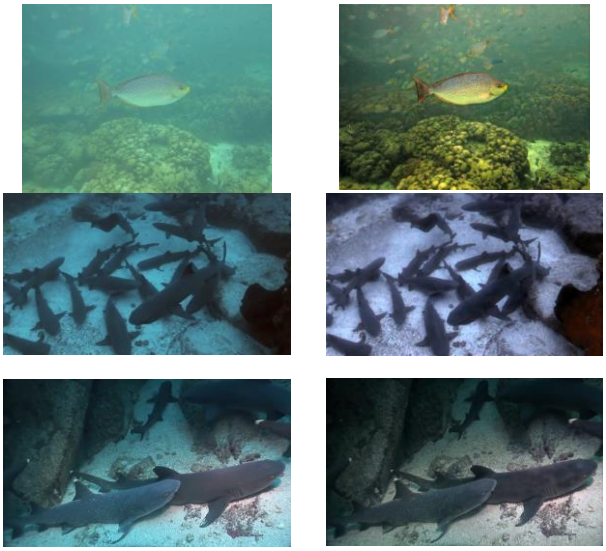


Figure 4: Raw Underwater Images (left), UWCNN++ Enhanced Underwater Images (right)

From figure 4, it can be seen that the proposed UWCNN++ model works well for enhancement of underwater images for both images with and without artificial illumination.

3.2. Comparison with UWCNN+

To compare the performance of the proposed UWCNN++ against UWCNN+, models were trained per instruction in [21]. The results are shown in Figure 5, below, for comparison.

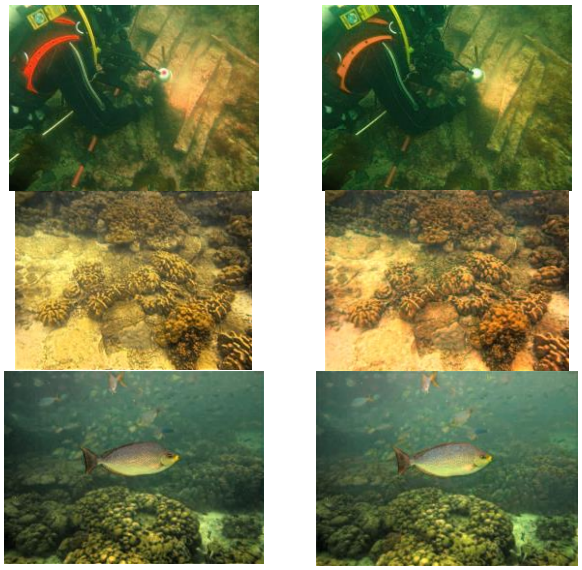


Figure 5: UWCNN+ Image (left), UWCNN++ Image (right)

In figure 5, there it is seen that for images without artificial illumination the performance of UWCNN+ and UWCNN++ is comparable. The biggest difference between UWCNN+ and the proposed UWCNN++ is how it handles artificial illumination, see diver and two sharks. In the UWCNN+ images on the diver's lamp and around the leading shark a pink artifact is observed. Compared to UWCNN++ these pink artifacts are not observed.

4. Conclusions

In this work, an improvement to UWCNN+ [21] was proposed: UWCNN++. UWCNN++ was created to address the need to handle images with artificial illumination, while maintaining the main speed benefits of UWCNN+. In this paper it is shown the UWCNN++ method performs comparably to UWCNN+ for underwater images without artificial illumination and that it produces better results for underwater images with artificial illumination.

UWCNN++ is different from UWCNN+ mainly in the data augmentation step. This augmentation step is not perfect, and such there are many parameters that can be modified to achieve better results. A comprehensive study was not possible in this paper. The author hopes this can be explored further in a follow-up paper.

References

[1] C.D. Mobley, Light and Water: Radiative Transfer in Natural Waters, Academic press, 1994

[2] Yemelyanov K, Lin S, Pugh E, and Engheta N (2006) Adaptive algorithms for two-channel polarization sensing under various polarization statistics with nonuniform distributions. Applied Optics, 45(22):5504-5520

[3] Schechner Y, and Averbuch Y (2007) Regularized image recovery in scattered media. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(9):1655-1660

[4] Tan C, Sluzek A, Jian T (2007) Range gated imaging systems for underwater robotic vehicle. In Proceedings of IEEE International Symposium on the Applications of Ferroeletrics, pp 1-7.

[5] Li H, Wang X, Bai T, Jian W, Huan Y, Ding K (2009) Speckle noise supression of range gated underwater imaging system. Applied Optics, 38(18):3937-3944
[6] Treibitz T, and Schechner Y (2012) Turbid scene enhancement using multi-directional illumination fusion. IEEE Trans. On Image Processing, 21)11):462-4667
[7] C. Li, J. Guo, Underwater image enhancement by dehazing and color correction, J. Electron. Imag. 24 (3) (2015) 033023-1 033023-10
[8] C. Ancuti, C.O. Ancuti, Enhancing underwater images and videos by fusion, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), IEEE, 2012, pp. 81-88
[9] C. Ancuti , C.O. Ancuti , C. Vleeschouwer , Color balance and fusion for underwater image enhancement, IEEE Trans. Image Process. 27 (1) (2018) 379–393 .
[10] C. Li , J. Guo , C. Guo , Emerging from water: underwater image color correc- tion based on weakly supervised color transfer, IEEE Signal Process. Lett. 25 (3) (2018) 323–327 .
[11] Y. Guo , H. Li , P. Zhuang , Underwater image enhancement using a multiscale dense generative adversarial network, IEEE J. Ocean. Engineer. (2019) 1–9 .
[12] Fu, X., & Cao, X. (2020). Underwater image enhancement with global–local networks and compressed-histogram equalization. Signal Processing: Image Communication, 86, 115892.
[13] J. Chiang , Y. Chen , Underwater image enhancement by wavelength compensa- tion and dehazing, IEEE Trans. Image Process. 21 (4) (2012) 1756–1769 .
[14] A. Galdran , D. Pardo , A. Picn , A. Alvarez-Gila , Automatic red-channel underwa- ter image restoration, Vis. Commun. Image Rep. 26 (2015) 132–145 .
[15] K. He , J. Sun , X. Tang , Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2011) 2341–2343 .
[16] P. Drews , E. Nascimento , S. Botelho , M. Campos , Underwater depth estimation and image restoration based on single images, IEEE Comput. Graph. Appl. 36 (2) (2016) 24–35 .
[17] C. Li , J. Guo , S. Chen , Y. Tang , Y. Pang , J. Wang , Underwater image restoration based on minimum information loss principle and optical properties of un- derwater imaging, in: Proc. IEEE Int. Conf. Image Process. (ICIP), IEEE, 2016, pp. 1993–1997 .
[18] C. Li , J. Guo , R. Cong , Y. Pang , B. Wang , Underwater image enhancement by de- hazing with minimum information loss and histogram distribution prior, IEEE Trans. Image Process. 25 (12) (2016) 5664–5677 .
[19] Y. Peng , P. Cosman , Underwater image restoration based on image blurriness and light absorption, IEEE Trans. Image Process. 26 (4) (2017) 1579–1594 .
[20] J. Li , K. Skinner , R. Eustice , M. Roberson , Watergan: unsupervised generative network to enable real-time color correction of monocular underwater images, IEEE Robot. Autom. Lett. 3 (1) (2017) 387–394 .
[21] C. Li, S. Anwar, F. Porikli , Underwater scene prior inspired deep underwater image and video enhancement, Pattern Recognition 98 (2020), 107038