

Underwater Image Enhancement using Masked MSE Loss

Animesh Devendra Chourey

Student Number - 210765551

MSc Artificial Intelligence, Queen Mary University of London

a.chourey@se21.qmul.ac.uk

Project Supervisor : Chau Yi Li

chauyi.li@qmul.ac.uk

Abstract—The underwater image processing and enhancement field have been garnering significant attention because of the emphasis put on aquatic and marine life by the researchers. Aquatic robotics also play a key role in the exploration and exploitation of underwater life along with ecological research. New potential sources for food, renewable energy generation, and developing medicinal drugs can be unlocked. This is achievable if exceptional algorithms and models are developed for enhancing the underwater imaging environment. The goal of this task is for the model to assert its focus on the color chart added to the images taken in the underwater environment by introducing corresponding masked images. Masks, which indicate the position of the color chart, and the underwater images are both applied together to train the network. color chart serve as reference standards to estimate quality degradation under varying lighting conditions. The loss function has been modified so that the algorithm acts accordingly. This served as a more robust and agile way possible to enhance the images.

Index Terms—Image Enhancement, Underwater Imaging, Transfer Learning, Masked Image Modelling, Convolutional Neural Networks.

I. INTRODUCTION

Capturing images in an underwater environment has been one of the most daunting tasks in the computer vision field as it poses quite unique challenges. Undistorted images are hard to acquire the deeper we go. Constraints such as light penetration and the underwater environment hinder the image quality captured. Deteriorated images impact hugely on feature extraction as well as object recognition. Images underwater gets degraded due to color cast, mainly blue and green color cast because blue color and green color possess longer wavelengths compared to others and can travel deeper resulting in selective attenuation with greenish and bluish hues [1], due to wavelength-dependent attenuation and scattering, due to haze because of suspended particles, and the marine snow also affecting in the form of noise. The ambient light and the vertical depth from which the images have been taken, resulting in varying water colors observed at various depths. The intensity is reduced along each scene in the image to camera distance (range) because the objects are illuminated by the attenuation of the ambient lights and also because the reflected lights are scattered and absorbed [2]. Thus the more the objects are farther from the camera the more they are blurred and lack contrast. Enhancing these distorted

and degraded images into better quality will be crucial in many applications of underwater images and videos in aquatic life. It will also be beneficial and of great significance in speeding up the research and exploration in marine biology and archaeology [3] and marine ecology [4] and numerous other domains. Utilization of deep-sea resources can lead to technological advancements and can also serve as new source of food [5], [6], [7].

In order to build models for the tasks required, one needs a huge amount of high-quality data to learn. Getting those high-quality underwater images is an arduous task and is extremely hard to acquire. Obtaining those ground-truth images is difficult due to the constraints mentioned before. Despite the advancements in image enhancement algorithms in underwater environment, they remain inadequate because of the unavailability of the real-world dataset of underwater images that correctly mimics the underwater environment. These underwater images need to accurately encapsulate most essential information so that they can be effectively incorporated into myriad applications. This issue results in the image enhancement algorithms in deep learning not performing up to the standard matched by the recent deep learning based low-level and high-level vision problems [8], [9], [10], [11], [12]. To overcome the said barrier and boost advancement in the progress in image enhancement in respect to underwater field, various datasets have been proposed. *Underwater Image Enhancement Benchmark* (UIEB) [13] is one of those datasets and it encompasses the underwater environment most accurately to my knowledge. This UIEB dataset has been used in this project. UIEB has contributed by providing a platform with well-designed pairwise comparison generally needed in image enhancement algorithms, as well as allowing the comprehensive study of state-of-the-art single underwater image enhancement algorithms (quantitative and qualitative evaluation) thus helping by offering insights into the strengths and shortcomings of current algorithms while also suggesting new possible directions for research areas. As already mentioned, because of the punishing underwater environment, there is severe lighting conditions. Using color chart allows even in different illuminations, how the actual color is getting degraded and serves as a better reference to make any adjustments needed. Masked image modeling is

implemented to alter the loss function in this project. Masking is done here so that only some portions are revealed and other remaining portion is concealed and completely untouched. Masking basically ensures selective adjustment is possible. This allows us to estimate the quality of image degradation occurring by referring to certain patches equivalent to those of color chart. Masking makes image manipulation more dynamic and also adds productivity to the algorithm.

II. RELATED WORK

Research will only be valuable and plausible in the domains of aquatic and marine life by attaining good quality images. There are numerous methods present functioning on enhancing the underwater images. Existing methods for enhancing underwater images can be classified into two categories: traditional methods and deep-learning-based methods.

A. Traditional Methods

To improve the visual quality, the traditional approach has been to find ways to tune the pixel values. Contrast enhancement method [14] focuses on optimizing the contrast of underwater images, as our vision is more sensitive towards contrast as compared to luminance. Another algorithm enhances contrast using maximum likelihood estimation of scale parameter followed by correcting the loss of energy in signal due to contrast enhancement [15]. Dynamic pixel range stretching [16] tries to equalize contrast and lighting concerns by implementing two algorithms on RGB and HSI color models. Pixel distribution adjustment [17] model improves over histogram equalization approach by averaging RGB and HSV color spaces while extending on Rayleigh histogram approach. Image fusion model [18] improved on fusion-based strategy for underwater image enhancement by blending the images obtained from a white balanced single input image. Another framework presented is using particle a swarm optimization algorithm for enhancing underwater images [19]. Adjustments were made in the RGB by the algorithm which showed refinement in the illumination and true colors of images. A color correction technique [20] published is implemented in such a way that a threshold conditional to the histogram is used to adjust low light distributed regions. Another color correction technique [21] enhances diminishing colors in underwater images by evaluating manual and automatic color correction techniques. It does so with the use of stretched histograms by acquiring their mean values. Ultimately it was ascertained that due to the significance level, the manual correction technique surpasses the automatic enhancing approach. These models mentioned before are the physical model-free based algorithms that omit prior knowledge of environmental conditions needed and therefore producing either under-enhanced results or over-enhanced results, or they introduce artificial colors. For instance, the image fusion model [18] does not produce results when dealing with diverse and challenging underwater scenarios.

Physical model-based methods are more extensively used as they incorporate underwater imaging mechanisms by ex-

tracting features/parameters according to prior information. Prior methods such as Red Channel Prior [22] and Dual Dark Channel Prior [13] aids in the rectification of problems concerning non-uniform illumination and artificial lighting by assisting in detecting the artificially illuminated regions. Blurriness Prior [23] adopts the image formation model (IFM) and integrates it with image blurriness, making it possible to evaluate the distance between the camera and scene points and thus ultimately enabling to recover and further enhance those underwater images. General Dark channel prior [24] proposed by Peng deals with generalizing dark channel prior in order to handle images taken in extreme weather conditions. There are certain issues with using methods extending physical-based models. Firstly, they are either time-consuming or are sensitive toward particular image types [25]. Secondly, obtaining the parameters for the medium transmission depends discontinuously upon the initial data. Moreover, the algorithms based on current physical models [26], [27], [22], [28], [24] faces challenges, as getting valuation of complex underwater imaging parameters is not definite i.e. underwater imaging models necessarily struggle to have authority and influence.

B. Deep Learning Models

There has been explosive growth and development in the fields of artificial intelligence and machine learning in recent decades many of which have been fuelled by the research breakthrough in Deep Learning. The fierce study and application of Convolutional Neural Networks from 2011-2016 of last decade for the challenge on massive image datasets such as ImageNet [30] in annually held competition *ILSVRC*, ignited the boom in the rapid development of the computer vision tasks and innovations in the architecture of the CNN models.

Deep learning allows to identify and extract features from images without relying on the need to break down the images into individual pixels. In recent years, the researchers have managed to achieve extreme preciseness in solving image recognition and classification tasks with deep neural networks (convolutional neural networks), and especially there is substantial boost in low-level visual tasks [31], [32], [33], [12], [34]. Due to breakthroughs and significant performance improvements, deep learning is also being tried upon improving the tasks of performing underwater image enhancement. Novel ideas have been proposed in the underwater imaging field. A GAN (Generative Adversarial Network) has been implemented with an image formation model to produce clean images for the task of supervised learning [35]. A deep residual network is implemented to deal with the underwater image enhancement task in one of the paper [36]. In this paper, convolutional neural network models are fed input training data by generating synthetic underwater images using Cycle consistent adversarial network (CycleGAN). Then a reconstruction model i.e. very-deep super-resolution reconstruction model (VDSR) is implemented which acts as an enhancement model. Thus a residual network model was proposed for enhancing the underwater image. A new multiscale dense generative adversarial network (GAN) was proposed for underwater image enancement [37].

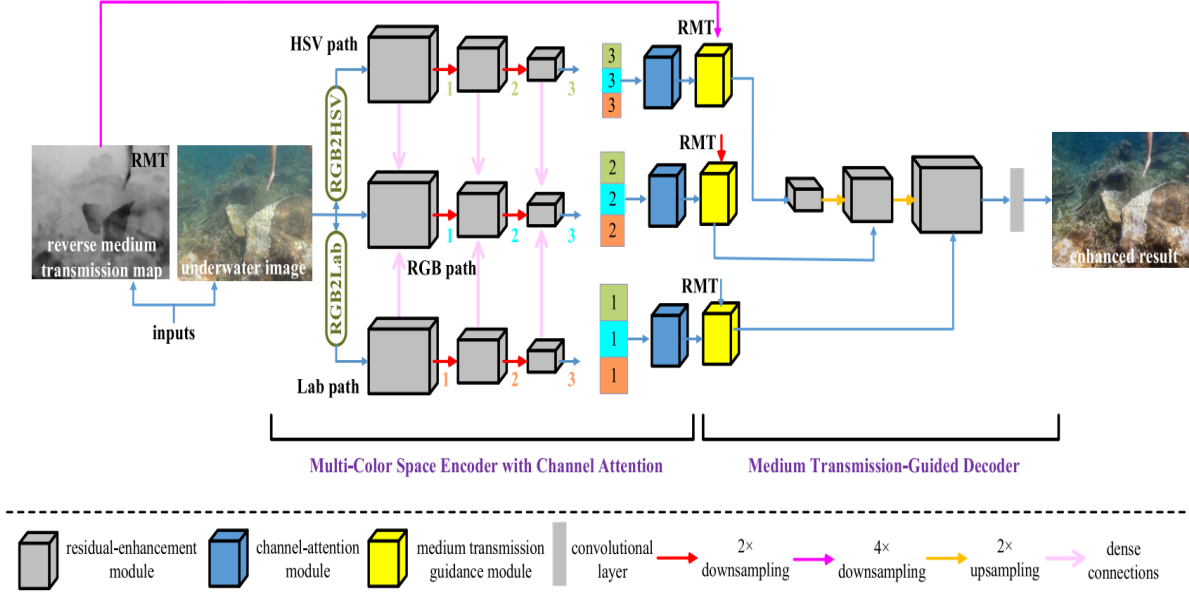


Fig. 1: Outline of the implemented model Ucolor can be seen here. The architecture consists of an encoder and a decoder network. Max-pooling is used to implement *downsampling*, and bilinear interpolation is used to implement *upsampling*. kernel size is set to 3x3 with stride fixed to 1 for each convolutional layer. The fig used here is taken from the Ucolor paper [29] and can be found in Sec. III of that paper [29].

Here, for generator, a dense-block performing residual multi-scale is used because of which there is a boost in performance, as it renders more details and also exploits preceding features. As for the discriminator, to stabilize its training, a computationally light spectral normalization is used. A conditional generative adversarial network (cGAN) [38], where a dual discriminator has been developed whose task is to determine the authenticity of the generated results from multiple views has also been proposed.

III. PROPOSED METHOD

In this section, we discuss the architecture used in the project, along with the training dataset used here to implement the model. Also, in the end, the loss function is described as used in the project to evaluate the outcome presented by the model. This project is based on a pre-built model called Underwater image enhancement via medium transmission-guided multi-color space embedding (Ucolor) created by *Li and Anwar* [29].

A. Architecture

The unique characteristics implemented in their method's architecture distinctive to the already functioning deep learning models are: a) As attention mechanism has been combined with multi-color space encoder network, this allowed distinctive features representation due to multi-color space, while the representative information is also determined in an adaptive manner; b) Attention mechanism has been tailored by incorporating the existing expertise of underwater image behavior

into the neural network so that medium transmission guided decoder network can emphasize on quality-degraded areas; c) No pre-processing step is needed before the model training and since it imitates supervised learning, more stable results are generated; d) Supports end-to-end learning, thus the method operates the underwater scenes in a unified structure; e) The model is able to attain outstanding results and marvellous performance on various datasets.

The Ucolor model's architecture is shown in Fig. 1. The following are the key components of the model's architecture.

1) *Multi-Color Space Encoder*: Each image is represented in the form of three different color spaces i.e. features are extracted by the algorithm in three color spaces (HSV, RGB, Lab). Every degradation-related component (color, saturation, hue, luminance, and intensity) gets incorporated into a unified structure when using all these color spaces. There might be contrasting color difference values between two points in one color space compared to the other color space. Therefore, the measurement of color deviations of underwater images is facilitated by multiple color spaces embedding. Also, more non-linear operations are brought by a multi-color space encoder while color space transformation occurs. These non-linear transformations ultimately improve deep models performance [39].

2) *Residual-Enhancement Module*: In order to retain data fidelity and tackle vanishing gradient problem, residual enhancement module is used, with convolutional layers in each module having exact number of filters. In the encoder network,

number of filters are raised progressively with a factor of 2 from 128 to 512, whereas, in the decoder network, number of filters are reduced from with the same factor from 512 to 128. Kernel size was set to 3x3 with a stride of 1 for every convolutional layer. Fig. 2 shows the construction of each residual enhancement module.

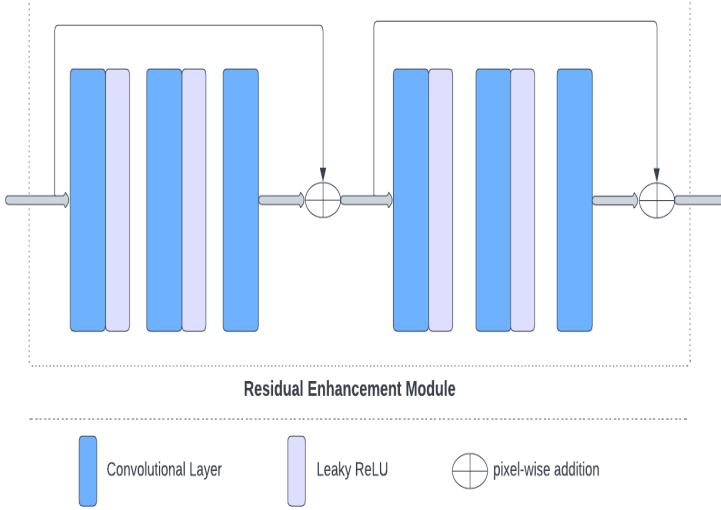


Fig. 2: Every residual-enhancement module used here in the model consists of two residual blocks, and three stacks of convolutional layers followed by activation function (Leaky ReLU) are constructed in each of these residual blocks, with only last convolutional layer not having one activation function after it. For identity connection, after every residual block there is pixel-wise addition.

3) *Channel Attention Module*: Features extracted from each of the three color spaces have disparate contributions and thus interdependencies are exploited by the channel attention module explicitly between the channel features from these three color spaces used. Identical mapping fashion is used to handle the channel attention weights because of which it is able to retain the fine properties of original features and avert the vanishing gradient problem. Mathematical notations used by them can be seen in Sec. III-C [29].

4) *Medium Transmission Guidance Module*: With the medium transmission guidance module, the decoder network incorporates the medium transmission map. To be more precise, a pixel-wise attention map is used by reversing the medium transmission map i.e. reverse medium transmission (RMT). These RMT maps are utilized as feature selectors to highlight the significance of various spatial positions of these features. One thing to note here is that there is no availability of the ground truth medium transmission maps of the corresponding input underwater images in reality. To tackle this issue and collect corresponding ground truth medium transmission maps, they utilized a prior-based estimation algorithm. More details about the mathematical notations used

for this module can be obtained in Sec. III-D [29].

B. Training Images

To obtain the input training images and their corresponding ground truth reference images, Seas-thru [40] dataset is referred. The model proposed in the Sea-thru successfully attains to remove the water from the underwater images making it possible for datasets to be analyzed with higher efficiency. Reference images are generated using this.

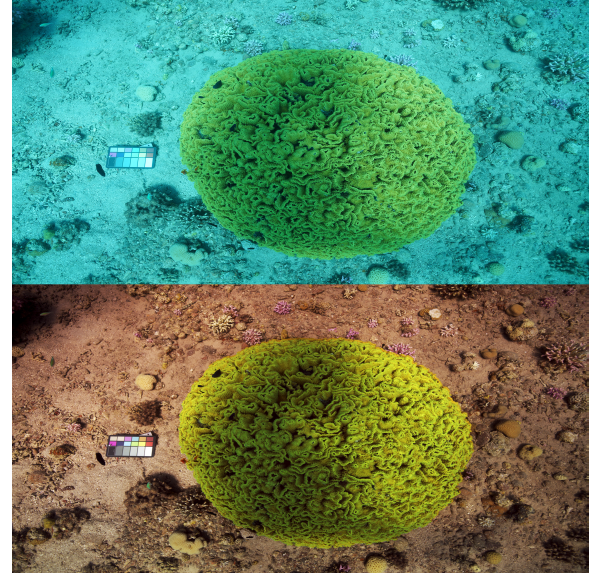


Fig. 3: The figure shows the image taken underwater along with its corresponding result showing as if every object present is not in an underwater environment. The images showed here are also presented in the original paper [40].

The method implemented by them, when fed an RGBD image, in a similar fashion Dark Channel Prior (DCP) works with haze [41], model firstly calculates backscatter with only difference being employing the known range map. Then, to calculate the range-dependent attenuation coefficient, the optimization framework uses an illumination map. This illumination map is procured by adopting local space average color [42] as inputs. Unknowns in the optimization step can be substantially scaled down and can be modeled as 2 term exponential because of the distance dependent attenuation coefficient. They are the first to employ revised image formation model which surpasses every method using the previous models, by doing quantitative and qualitative evaluation. More information regarding the working of the model's method can be found in the Sec. IV [40].

The image shown in fig. 4 is one example of the set of 40 training and reference images used for the training purpose. The images trained are of size 620x420. The masked images used in the project are grayscale images. In each image, the pixel intensity is maximum at those pixel positions where there are color charts present. One such example of the masked images used is shown in the Fig. 5.

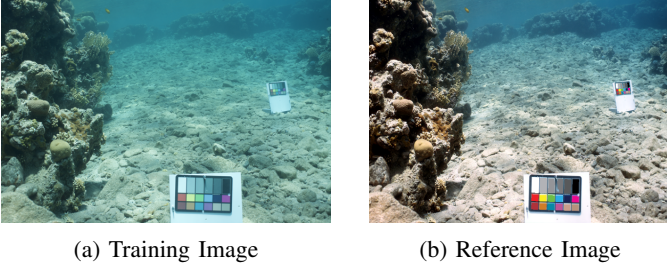


Fig. 4: This figure shows one of the input image used for training and it's corresponding reference image. This reference image is generated by using [40].

C. Loss Function

The aim is to enhance the underwater image by removing the color casts. Continuing the previous work performed [29], changes have been made to the loss function in such a way that it focuses mainly towards the color chart introduced into the images. To ensure this, corresponding masked images are introduced.

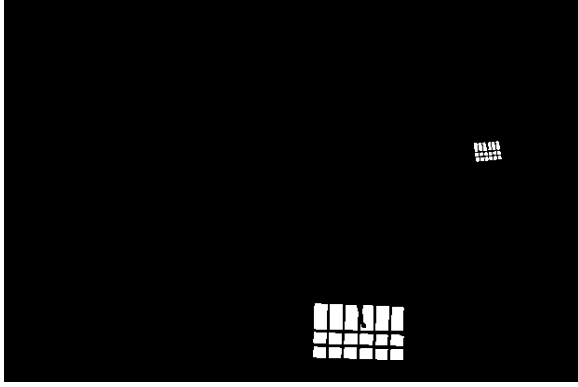


Fig. 5: The figure shows one of the masked image corresponding to the training and reference image shown in Fig. 4. Every masked image used is of grayscale in nature.

To perform the masking procedure, binary mask are added. These binary masks are nothing but an array of binary values of grayscale images holding values. The values that masked image contains are either 0 or 1. Value 1 here corresponds that at that pixel position there is a presence of color chart. Masked images are loaded first in a similar fashion as other images are loaded. Every set of images are then converted into an array of float values. For the training procedure, image patches are selected after randomly cropping them into a shape of 128x128. Patches selected here from the input images, depth images and the masked images are always picked from same x and y position. Once this is done, the image patches are finally sent to the loss function for evaluation. The loss function consists of combination of MSE loss (L_{MSE}) and VGG loss (L_{vgg}). MSE loss here is mean squared error while VGG loss corresponds to the network's loss.

Let's suppose, $I(x,y)$ is an ground truth image, and $N(x,y)$

is representing the enhanced result of the feeded input image generated by the network. $M(x,y)$ represents the corresponding masked image. Firstly, pixel-wise difference is taken in between the underwater input image and the enhanced result. Let the difference be denoted by $D(x,y)$.

$$D(x,y) = I(x,y) - N(x,y) \quad (1)$$

Once the difference is calculated i.e. $D(x,y)$, element-wise multiplication is executed between $D(x,y)$ and $M(x,y)$. This multiplication with masked image enforces that only the desired region gets calculated for the loss. The loss function is only able to focus on that window, and forces training accordingly. The resulting value is the error in pixel values corresponding to only the color chart patch that we are trying to get. It is denoted by me (masked error).

$$me = D(x,y) * M(x,y) \quad (2)$$

After this, the square of the masked error i.e. me is calculated. Element-wise square is computed here. It is denoted by sme (squared masked error).

$$sme = (me)^2 \quad (3)$$

Then the mean is taken for the pixels having non-zero values. This is denoted by nzv (non zero values) and calculated by taking the result with the help from (2) i.e. me .

One thing to note here is that there could be instances where the image patches getting selected from the masked images only contains the pixels corresponding to value 0. While taking the mean of the sme , only those pixel values are considered that are non-zero. If this constraint is considered, and if the image patch getting cropped from masked image only contains zero values, then it could lead to unwanted results in the form of error getting divided by the 0 as there are no non-zero values present at the moment. This could ultimately results in us getting undefined value. This is an undesirable result for the loss. To tackle this issue, maximum between the non-zero values and the constant value 1 is taken. This is denoted by m (non zero values).

$$m = \max(1, nzv) \quad (4)$$

Finally, the mean is taken by combining (3) and (4). Squared masked error sme is divided by m obtained in (4). This results in getting Mean Squared Error denoted by L_{MSE} .

$$L_{MSE} = sme/m \quad (5)$$

VGG loss (L_{VGG}) is the network's loss on pre-trained VGG-19 network on ImageNet dataset [30]. From this pre-trained network, the relu5_4 layer is used to compute the vgg loss by calculating the distance between the feature representation of the ground truth image and the corresponding reconstructed result. This is done for every convolutional layer present in the network. Let $\phi_j(.)$ be the vgg networks loss for the j th convolutional layer [39].

$$\sum_{x=1}^H \sum_{y=1}^W |\phi_j(N)(x,y) - \phi_j(I)(x,y)| \quad (6)$$

As mentioned before, total loss is a linear combination of MSE loss (L_{MSE}) and VGG loss (L_{VGG}) calculated. This is done so as to maintain reliable and steady equilibrium in between quantitative scores and visual quality. Total loss is denoted by (L_{total}) and is expressed as:

$$L_{total} = L_{MSE} + \lambda L_{VGG} \quad (7)$$

Here, as both the losses are obtained in different scales, in order to stabilize the range of these varying losses, from observational analysis, λ is set to 0.05.

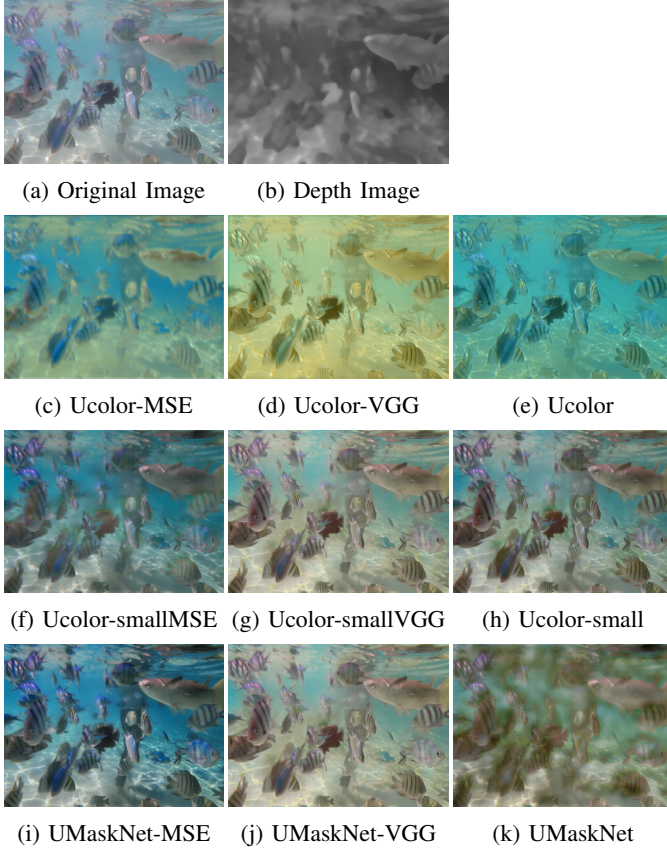


Fig. 6: This figure shows the visual comparison done on input image having minute blurriness and blue color deviation. The input image contains several objects in it and of different sizes. Results produced by each model is shown here. UMaskNet-MSE enhances almost every object without deviating from the original color.

IV. EXPERIMENT

This section comprises of implementation details, experiment setting is then explained and finally in the end visual comparison is shown.

A. Implementation Details

To train the network, 40 images are taken. Each image used has different illumination setting and the position of the color chart used is also varying from image to image. This ensures

various characteristics regarding the quality degradation occurring gets recorded. The model is trained with a batch size of 10 while the learning rate is set to $1e^{-6}$ i.e. 0.000001. The training for the model is done for 100 epochs with the learning rate is fixed with same value as it was initialized with throughout the whole process. Gaussian distribution is used for initializing the filter weights for every layer. Initially, the bias has been set to a constant value. As for the network's optimization, ADAM optimizer is used.

B. Experiment Setting

To evaluate the model, 90 sets of images are taken from the UIEB dataset [13] for the testing purpose. Various sets of combinations have been used to evaluate the testing images. Various combinations have been used to test out our model UMaskNet and is compared with the original Ucolor model. UMaskNet contains the modified loss function which adapts to the masked images added as mentioned in Sec. III-C. Two variants of UMaskNet have been used: one where only MSE loss (L_{MSE}) is being calculated i.e. *UMaskNet-MSE* and another one where only VGG loss (L_{VGG}) is being used towards contributing to total loss i.e. *UMaskNet-VGG*. Then, with the learning rate being $1e^{-6}$ as used for UMaskNet instead of $1e^{-4}$ (used for the original Ucolor model), the Ucolor model has been used (let it be denoted by *Ucolor-small*). It does not take into account the masked images whatsoever. However, here also two variants of *Ucolor-small* are used. One calculating only MSE loss (*Ucolor-smallMSE*) and other time to compute VGG loss only (*Ucolor-smallVGG*). Finally, with keeping the learning rate $1e^{-4}$ as done in the original Ucolor model has been used. Similarly, original Ucolor has also been tweaked twice, once to calculate VGG loss only (denoted by *Ucolor-VGG*) and the other time to calculate MSE loss only (denoted by *Ucolor-MSE*).

The model has been trained using NVIDIA Tesla P100-PCIE with the CUDA version 11.2. Training each set of model took 3 hours.

C. Visual Comparison

This section comprises of visual results obtained after executing the different experiment setting mentioned before.

Fig. 6 shows the results on a forward looking image. The model *UMaskNet* seems to be adding unwanted noise. This could be because of VGG loss. Since the VGG loss function uses L2 distance in between the activations used in the hidden layers by training on particular image datasets (here ImageNet dataset [30]). One drawback for using it could be that this perceptual loss needs hyper-parameter tuning and requires regularization for unrelated tasks. Therefore, as can be seen from variants where only VGG loss is calculated, the results are worsening. However, with *UMaskNet-MSE* the results are quite pleasing and seems to be effectively eliminating haze and adds more sharpness to the objects making them more identifiable.

Fig. 7 shows results on a downward looking image which has distinct green color cast present in it. As we can see due

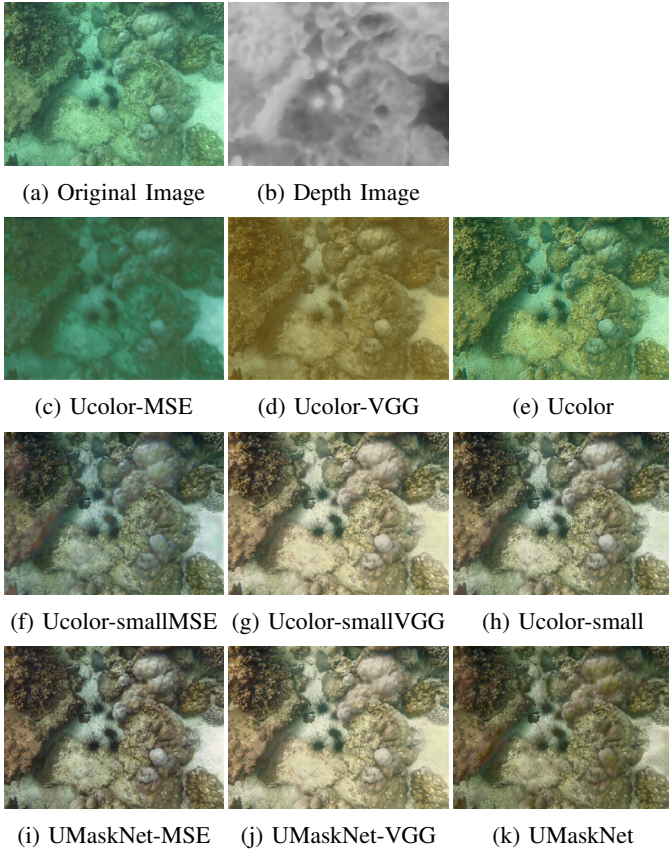


Fig. 7: This figure shows the visual comparison done on input image having high green color deviation. There is also low color contrast in the input image. Results produced by each model is shown here.

to the green color cast, structural details in the underwater image are not quite comprehensible and makes it demanding to visualize. Color artifacts adds up in the case of *Ucolor-VGG*, while in *Ucolor-MSE* and *Ucolor* there is still some green color cast left. In terms of visual quality, other models either under-enhances or adds over-saturation to the image, whereas, *UMaskNet-MSE* again gives best result by managing to sharpening the contrast and not adding obvious over-saturation.

Fig. 8 shows the results on image with high-backscatter. Backscatter occurs because in between the object and the camera i.e. distance in between the scene to be viewed and the camera, the particles present gets illuminated. This results in more fogginess in the image and the contrast also shrinking. Our model on the images having low backscatter as can be seen in Fig. 6 and Fig. 7 is able to solve the issue extremely effectively. However, on the images having high-backscatter, every model seems to be struggling to solve the issue. As can be seen from the images shown in Fig. 8, the results are not quite desirable. Almost every model seems to be adding some color cast into the result. The results are also quite blurred. *UMaskNet-MSE* seems to be producing decent result out of all

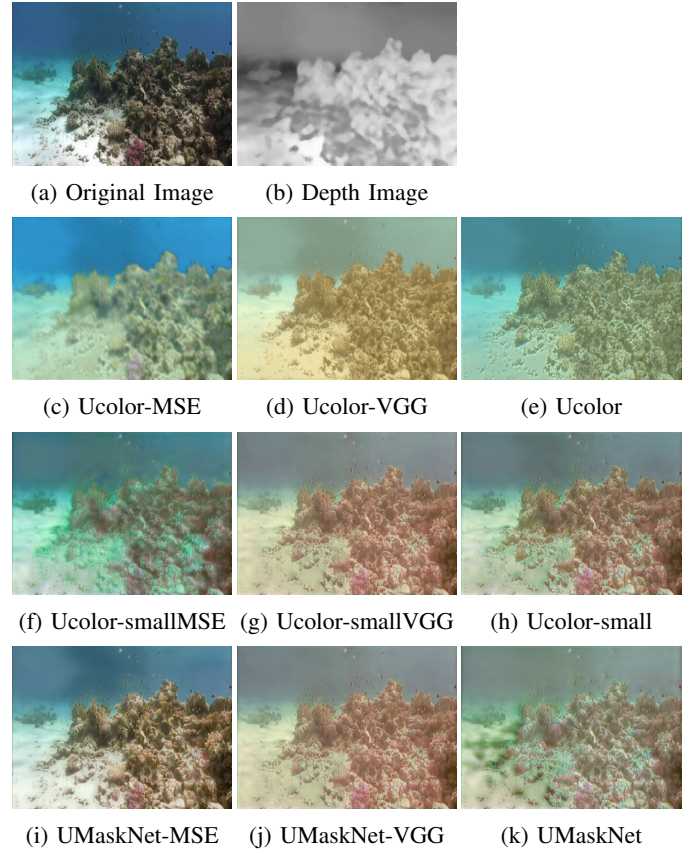


Fig. 8: This figure shows the visual comparison done on input image which has high-backscatter scene. Results produced by each model is shown here. Mixed results are produced by each model where almost every resulting output seems to be under-enhanced.

other models as the color correction seems to be original and a bit realistic but even here the result seems under-enhanced and it is not a desirable result. *UMaskNet-MSE* is slightly better at solving backscatter issue. Some modifications are needed to tackle this issue more efficiently and robustly.

Fig. 9 shows the progress made by the *UMaskNet-MSE* model after every 10 consecutive epochs. Since the model is trained for 100 epochs, there are 10 images shown in the figure. Though there does not seem much difference in the changes occurring at each epoch shown, as images presented here are quite smaller and it is not plausible to show them in their original size. However, there is evident changes present. The contrast is slightly rising in each epoch result. The images in the front are getting more highlighted. The reason for that is because the color charts used in the images are closer to the camera and that's why the network is forcing the foreground objects in getting more enhanced.

V. CONCLUSION

The main focus of this paper was to implement loss function that adjusts according to the color charts introduced in the underwater images so that it is able to estimate quality

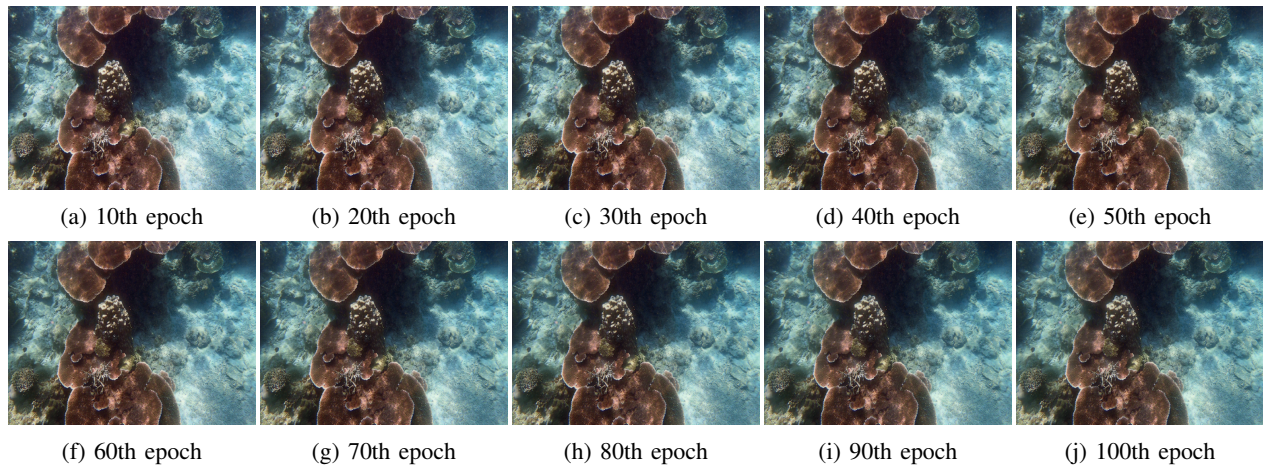


Fig. 9: This figure shows the output enhanced results by the *UMaskNet-MSE* model recorded at different consecutive stages of epochs. The images show the progress made by the model at different iterations.

degradation more efficiently. As can be seen from the results obtained, the proposed model is enhancing the images as intended to a better extent. The results are quite pleasing to the eye and no color is getting attenuated. Superior results are being generated. In certain scenarios, such as images having limited lighting i.e. low illumination, the output results are decent and almost all models generates results with improved visibility, even though images do not get enhanced to much extent, our model is not adding color casts to the output image. For the future work, the color charts can be more evenly distributed amongst the underwater image to tackle the issues of high back-scatter and low illumination.

VI. ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my supervisor, Ms Chau Yi Li for providing me great opportunity to work on the project. Thank you for giving me constant guidance and help needed throughout the project. Your invaluable suggestions and timely feedback, provided me much needed direction in order for me to fulfill everything needed.

REFERENCES

- [1] D. Akkaynak, T. Treibitz, T. Shlesinger, Y. Loya, R. Tamir, and D. Iluz, "What is the space of attenuation coefficients in underwater computer vision?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4931–4940, 2017.
- [2] C. Y. Li and A. Cavallaro, "Cast-gan: Learning to remove colour cast from underwater images," in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 1083–1087, IEEE, 2020.
- [3] M. Ludvigsen, B. Sortland, G. Johnsen, and H. Singh, "Applications of geo-referenced underwater photo mosaics in marine biology and archaeology," *Oceanography*, vol. 20, no. 4, pp. 140–149, 2007.
- [4] N. Strachan, "Recognition of fish species by colour and shape," *Image and vision computing*, vol. 11, no. 1, pp. 2–10, 1993.
- [5] Z. Li, G. Li, B. Niu, and F. Peng, "Sea cucumber image dehazing method by fusion of retinex and dark channel," *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 796–801, 2018.
- [6] J. Ahn, S. Yasukawa, T. Sonoda, T. Ura, and K. Ishii, "Enhancement of deep-sea floor images obtained by an underwater vehicle and its evaluation by crab recognition," *Journal of Marine Science and Technology*, vol. 22, no. 4, pp. 758–770, 2017.
- [7] B. J. Boom, P. X. Huang, C. Beyan, C. Spampinato, S. Palazzo, J. He, E. Beauxis-Aussalet, S.-I. Lin, H.-M. Chou, G. Nadarajan, *et al.*, "Long-term underwater camera surveillance for monitoring and analysis of fish populations," *VAIB12*, 2012.
- [8] C. Li, C. Guo, J. Guo, P. Han, H. Fu, and R. Cong, "Pdr-net: Perception-inspired single image dehazing network with refinement," *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 704–716, 2019.
- [9] W. Wang, Q. Lai, H. Fu, J. Shen, H. Ling, and R. Yang, "Salient object detection in the deep learning era: An in-depth survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [10] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, "Nested network with two-stream pyramid for salient object detection in optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9156–9166, 2019.
- [11] R. Cong, J. Lei, H. Fu, W. Lin, Q. Huang, X. Cao, and C. Hou, "An iterative co-saliency framework for rgbd images," *IEEE transactions on cybernetics*, vol. 49, no. 1, pp. 233–246, 2017.
- [12] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, "Hierarchical features driven residual learning for depth map super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2545–2557, 2018.
- [13] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.
- [14] H. Liu and L.-P. Chau, "Underwater image restoration based on contrast enhancement," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, pp. 584–588, IEEE, 2016.
- [15] S. Sankpal and S. Deshpande, "Underwater image enhancement by rayleigh stretching with adaptive scale parameter and energy correction," in *Computing, Communication and Signal Processing*, pp. 935–947, Springer, 2019.
- [16] K. Iqbal, M. Odetayo, A. James, R. A. Salam, and A. Z. H. Talib, "Enhancing the low quality images using unsupervised colour correction method," in *2010 IEEE International Conference on Systems, Man and Cybernetics*, pp. 1703–1709, IEEE, 2010.
- [17] A. S. A. Ghani and N. A. M. Isa, "Underwater image quality enhancement through integrated color model with rayleigh distribution," *Applied soft computing*, vol. 27, pp. 219–230, 2015.
- [18] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Transactions on image processing*, vol. 27, no. 1, pp. 379–393, 2017.
- [19] A. AbuNaser, I. A. Doush, N. Mansour, and S. Alshattawi, "Underwater image enhancement using particle swarm optimization," *Journal of Intelligent Systems*, vol. 24, no. 1, pp. 99–115, 2015.
- [20] J. Ao and C. Ma, "Adaptive stretching method for underwater image color correction," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 02, p. 1854001, 2018.
- [21] N. bt Shamsuddin, B. b Baharudin, M. Kushairi, M. Rajuddin, F. bt Mohd, *et al.*, "Significance level of image enhancement techniques

- for underwater images,” in *2012 International Conference on Computer & Information Science (ICCIS)*, vol. 1, pp. 490–494, IEEE, 2012.
- [22] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, “Automatic red-channel underwater image restoration,” *Journal of Visual Communication and Image Representation*, vol. 26, pp. 132–145, 2015.
 - [23] Y.-T. Peng, X. Zhao, and P. C. Cosman, “Single underwater image enhancement using depth estimation based on blurriness,” in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4952–4956, IEEE, 2015.
 - [24] Y.-T. Peng, K. Cao, and P. C. Cosman, “Generalization of the dark channel prior for single image restoration,” *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2856–2868, 2018.
 - [25] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, “Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4861–4875, 2020.
 - [26] P. L. Drews, E. R. Nascimento, S. S. Botelho, and M. F. M. Campos, “Underwater depth estimation and image restoration based on single images,” *IEEE computer graphics and applications*, vol. 36, no. 2, pp. 24–35, 2016.
 - [27] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, “Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior,” *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.
 - [28] Y.-T. Peng and P. C. Cosman, “Underwater image restoration based on image blurriness and light absorption,” *IEEE transactions on image processing*, vol. 26, no. 4, pp. 1579–1594, 2017.
 - [29] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, “Underwater image enhancement via medium transmission-guided multi-color space embedding,” *IEEE Transactions on Image Processing*, vol. 30, pp. 4985–5000, 2021.
 - [30] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.
 - [31] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.
 - [32] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, “Deblurring images via dark channel prior,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 10, pp. 2315–2328, 2017.
 - [33] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.
 - [34] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, “Low-light image enhancement via a deep hybrid network,” *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, 2019.
 - [35] 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*, vol. 3, no. 1, pp. 387–394, 2017.
 - [36] P. Liu, G. Wang, H. Qi, C. Zhang, H. Zheng, and Z. Yu, “Underwater image enhancement with a deep residual framework,” *IEEE Access*, vol. 7, pp. 94614–94629, 2019.
 - [37] Y. Guo, H. Li, and P. Zhuang, “Underwater image enhancement using a multiscale dense generative adversarial network,” *IEEE Journal of Oceanic Engineering*, vol. 45, no. 3, pp. 862–870, 2019.
 - [38] M. Yang, K. Hu, Y. Du, Z. Wei, Z. Sheng, and J. Hu, “Underwater image enhancement based on conditional generative adversarial network,” *Signal Processing: Image Communication*, vol. 81, p. 115723, 2020.
 - [39] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
 - [40]
 - [41] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.
 - [42] M. Ebner, “Color constancy based on local space average color,” *Machine Vision and Applications*, vol. 20, no. 5, pp. 283–301, 2009.