An Efficient Approach for Underwater Image Improvement: Deblurring, Dehazing, and Color Correction

Alejandro Rico Espinosa
University of Victoria
Victoria, Canada
arico@uvic.ca

Declan McIntosh
University of Victoria
Victoria, Canada
declanmcintosh@uvic.ca

Alexandra Branzan Albu
University of Victoria
Victoria, Canada
aalbu@uvic.ca

Abstract

As remotely operated underwater vehicles (ROV) and static underwater video and image collection platforms become more prevalent, there is a significant need for effective ways to increase the quality of underwater images at faster than real-time speeds. To this end, we present a novel state-of-the-art end-to-end deep learning architecture for underwater image enhancement focused on solving key image degradations related to blur, haze, and color casts and inference efficiency. Our proposed architecture builds from a minimal encoder-decoder structure to address these main underwater image degradations while maintaining efficiency. We use the discrete wavelet transform skip connections and channel attention modules to address haze and color corrections while preserving model efficiency. Our minimal architecture operates at 40 frames per second while scoring a structural similarity index (SSIM) of 0.8703 on the underwater image enhancement benchmark (UIEDB) dataset. These results show our method to be twice as fast as the previous state-of-the-art. We also present a variation of our proposed method with a second parallel deblurring branch for even more significant image improvement, which achieves an improved SSIM of 0.8802 while operating more efficiently than almost all comparable methods. The source code is available at `https://github.com/alejorico98/underwater_ddc`

1. Introduction

Underwater imaging is a growing domain with unique challenges related to underwater atmospheric conditions and lighting which decrease the quality of these images [24, 26, 2, 10, 19, 18]. Despite this, large institutions worldwide have seen the utility of collecting underwater video data from fixed locations for ecological and biological research [24, 14, 25]. This data enables better species counts, studying organism behaviors, and estimating population stress levels [25, 24, 14]. These contribute to better conservation efforts in oceans and freshwater bodies [24, 14]. From this data collection, institutions have amassed huge datasets quickly, which continue to accelerate due to camera deployment and decreasing storage costs [24, 14, 6]. Despite this abundance of data, underwater images are limited in their utility as they often struggle with atmospheric effects on image quality, especially affecting color accuracy and blurring the image [27, 28, 2, 1]. For instance, scattering in underwater images from water turbidity and algae blooms can completely obscure objects, especially in shallow or coastal waters [2, 1]. In addition, the attenuation of light in water image color channels is not consistent, making species identification challenging. Due to light attenuation and absorption in water, images suffer from low contrast and haze [7, 23]. These degradations often severely limit the utility of this imaging data. Therefore, there is a need for image enhancement methods that address light scattering, color distortion, and blur. These methods would also need to operate efficiently with greater than real-time speeds to handle enormous dataset backlogs and live data streams like those used for remotely operated underwater vehicle (ROV) exploration.

There has been a large body of recent research into practical ways to improve underwater images [27, 28, 2, 1]. These methods are primarily divisible into traditional computer vision and deep learning [2, 10, 19, 18]. Each group has significant trade-offs between speed, quantitative improvement, and generalization [27, 26, 2, 10, 19, 18]. For example, traditional computer vision methods are non-generalizable and might require considerable expertise to apply to specific domains, but they can be highly efficient [10]. On the other hand, deep learning-based statistical methods are limited to the domain of their training data and are generally slow, but have excellent performance in image improvement [10, 15]. We introduce a deep learning-based method that uses several key architectural features encouraging good performance in blur and color corruptions while remaining extremely efficient.

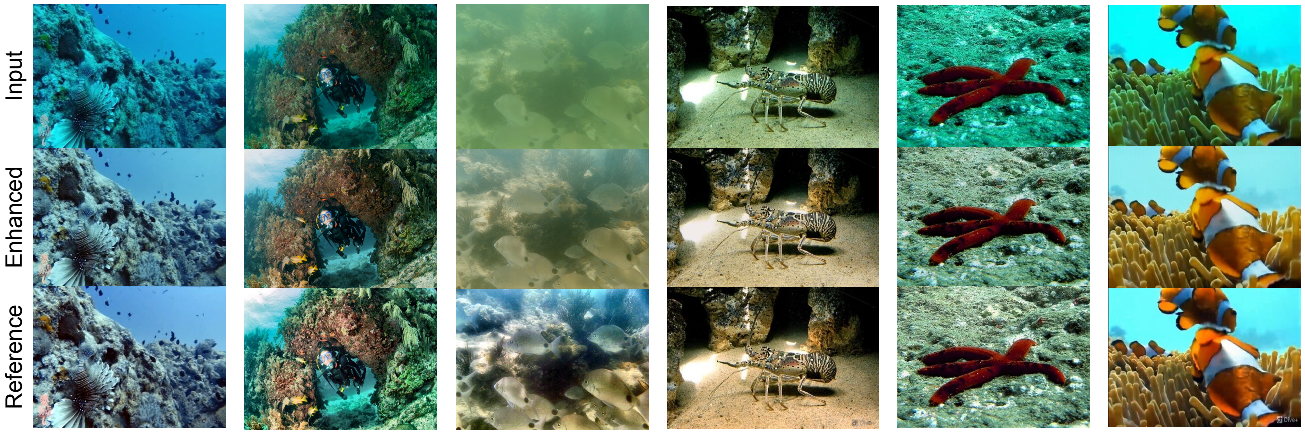

Figure 1. Visual samples of underwater image enhancement using our proposed efficient architecture. More qualitative results with our extended larger architecture can be seen in Section 4.

Our proposed method uses a U-Net style encoder-decoder architecture which is trained in a GAN structure with a second adversarial network [13, 29]. Our initial network is designed to be efficient and small [29]. We then present specific modifications of this base training regimen and architecture, which are well motivated by the image corruptions we consider more damaging to underwater images: atmospheric scattering, blur, and color absorption. The modifications minimally degrade the model's efficiency as a core goal of our work is to present greater than real-time performance, basing its effectiveness primarily on the structural similarity index (SSIM) score. First, we include Discrete Wavelet Transform (DWT) skip connections [9]. We apply the DWT to all channels of our feature maps before down-sampling and skip connections to increase the accessible information for the decoder. The DWT increases the number of feature channels, separating them into multiple representations at different frequency bands [9]. We also propose to use the channel and spatial attention blocks (CBAM) as an effective way to increase the color recovery of our model. This is because CBAM helps to represent multiple feature channels into relative color channels [34]. Finally, we add a gradient penalty to our GAN discriminator to enforce a Lipschitz constraint; this makes the training of our reconstruction network more stable [13].

Our proposed method shows state-of-the-art results in SSIM and inference speed on the UIEB test set [19]. Also, our model is competitive with PSNR with the current state-of-the-art methods. In this paper, we compare our method with existing classical computer vision, statistical deep learning methods, and hybrid methods. Section 2 describes the related works, Section 3 detail our proposed approach and each of its principal components, Section 4 shows the quantitative and qualitative results compared with other previous approaches, and Section 5 finally state our conclusions.

2. Related Works

There is an expanding number of applications of underwater cameras in ever more challenging visual environments, especially shallow waters. As a result, there has been an increase in underwater image enhancement research [24, 14]. Attention has been directed to two important and challenging components of enhancement; dehazing, a particular case of non-homogeneous deblurring, and color correction. These image corruptions exist in all underwater imaging but are especially prevalent in naturally lit shallow water applications. As deblurring and dehazing corruptions have terrestrial analogs, we include terrestrial and underwater deblurring methods in this Section [17, 5]. Furthermore, with the proliferation of deep learning-based statistical methods, the classical approaches based on visual priors and physical models have become less prevalent while still being used in applications lacking ground truth data.

Ancuti *et al.* proposed a fusion-based method only relying on the information gained from the degraded image [2]. They separately enhance the color and contrast of the image and then incorporate several other weight maps to account for the non-linear image corruption from long-distance objects [2]. These weight maps and improved images are then fused to generate an enhanced image. The method is a classical computer vision method and it is agnostic to image scene structure or the specific underwater conditions [2].

The method with significant success in deblurring using a terrestrial dataset was DeblurGAN from Kupyn *et al.* [17]. They present a conditional GAN loss in conjunction with a reconstruction loss [17]. This loss improves perceptual losses by penalizing the model for generating reconstruc-

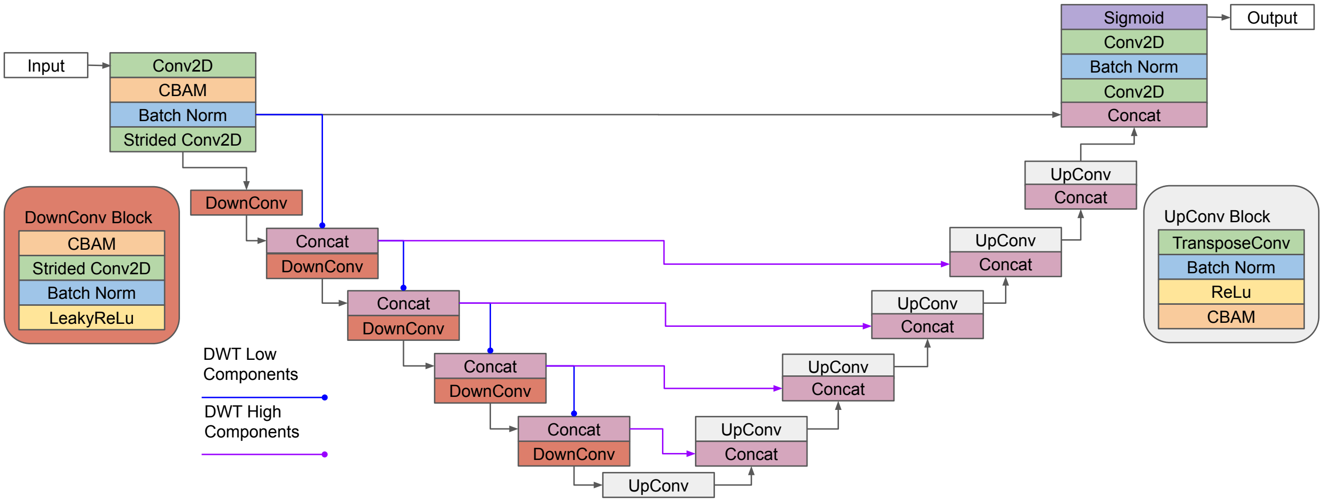

Figure 2. Proposed architecture based on the U-Net with multiple skip connections, DWT algorithm and CBAM blocks. The low frequency DWT information paths are highlighted in blue and the corresponding high frequency components are highlighted in purple.

tions that can score well on perceptual metrics but do not accurately sharpen the image. This improves the overall sharpness of the final images [17]. They also propose using a secondary detector's performance on the enhanced images as a metric or real-world applicability over previously used metrics like PSNR or SSIM [17].

Previous studies proposed the use of discrete wavelet transformation (DWT) feature spaces in deep learning models. With the use of a wavelet residual network W. Bae *et al.* [5] discovered the benefit of learning on subbands. Wavelet residual networks are able to learn on additional frequency subbands, increasing model representational power. Also, Fu *et al.* propose a discrete wavelet transform (DWT) based GAN for dehazing: DW-GAN [9]. DW-GAN consists of two parallel networks seeking to improve details in small datasets. More specifically, using non-homogeneous hazed images. The first network method uses DWTs to downsample the feature maps generated by the CNN while maintaining more of the high-frequency information in additional skip connections [9]. This provides the decoder with extra high-frequency feature map information. These additional features allow the network to better reconstruct the deblurred image, specifically by creating a sharper reconstruction. Then the two parallel CNNs responses predictions are averaged [9]. The second shallower knowledge adaptation branch uses pre-trained weights from a classification problem to increase model performance on smaller datasets [9]. This is motivated by the relative rarity of ground-truth information for dehazing problems.

Peng *et al.* present a U-shaped transformer for the underwater image enhancement task [22]. Their proposed method uses a channel-wise multi-scale feature fusion transformer and a novel spatial-wise global feature modeling transformer [22]. The latter is designed to increase the model's attention to regions of the image with significant attenuation and the relationship between color channels in the overall image [22]. The authors also propose a multi-space loss function that considers the reconstructed image in multiple color spaces to improve the qualitative [22].

3. Proposed Approach

3.1. Model Architecture

Our proposed approach consists of an end-to-end CNN based image enhancement method. We build a variant of the classical U-Net encoder-decoder structure incorporating several key modifications, each grounded in its contribution to specific corruptions of our target underwater images. This ground-up design from a lightweight architecture allows us to have greater than real-time efficiency beyond comparable previous works. First, we incorporate Discrete Wavelet Transform (DWT) enhanced skip connections while down-sampling in our encoder [9] to maintain high-frequency feature map components, allowing to preserve texture details for sharper output images during the dehazing task. Next, we adjust the design of our architecture using spatial and channel attention blocks between convolutions. We incorporate Channel Based Attention Module (CBAM) blocks which better exploit relationships between feature maps to create significantly improved color accuracy. Hence, we use them in a novel way to address the absorption effect of the red channel found in the underwater images. The entire model is presented in Figure 2.

Additionally, we also present a modification of our model that incorporates a second parallel deblurring

Res2Net architecture with an additional global skip connection resulting in better generalization and faster training than the conventional Res2Net [17]. This modification of our proposed approach does come at the cost of model efficiency. Figure 3 shows this variation.

Finally, we formulate the training of our network against a second adversarial network in a GAN based structure. The specific training structure we select is the Wasserstein variation because it helps mitigate issues with Jenssen-Shannon divergence. More details of these issues in the context of image reconstruction can be seen in Section 3.5. We modify this variation further by including a gradient penalty in the discriminator rather than clipping the values to enforce a Lipschitz constraint [13]. This modification increases the sharpness and color restoration in the final images while making the model more robust when the second deblurring branch is added. The following sections will discuss each of these contributions to our final novel architecture in detail.

3.2. Discrete Wavelet Transform Skip Connections

We use the same formulation of the DWT as a method for preserving high frequency features in skip connections [9]. This formulation is used to maintain more information from the feature maps in the high frequency ranges which are of particular concern for crisp image reconstruction. The DWT decomposes the image into high and low frequency sub-images which as a collection retain more of the original information than standard down-sampling.

Specifically, the 2D DWT is implemented by performing the convolution operation with four different filters f_{LL} low-low-pass, f_{LH} low-high-pass, f_{HL} high-low-pass, and f_{HH} high-high-pass. The low frequency components are comprised of the downsampling feature outputs obtained through the convolutional layers. The high frequency components are derived using the 1D Haar filter bank. The 2D problem is then formulated as a sequential repetition of the 1D procedure, first applying the filter in rows and then in columns [8]. Then, equations can be formulated as $f_{LL}(x, y) = f_L(x)f_L(y)$, $f_{LH}(x, y) = f_L(x)f_H(y)$, $f_{HL}(x, y) = f_H(x)f_L(y)$, $f_{HH}(x, y) = f_H(x)f_H(y)$. Where the second filter is applied after downsampling the signal by 2. Figure 4 illustrates this process. Despite testing our architecture with biorthogonal losses DWT filters (LeGall 5-3), we find that conventional Haar filters performed better. Then, we use $f_L = \frac{1}{\sqrt{2}}(1 + z^{-1})$ and $f_H = \frac{1}{\sqrt{2}}(1 - z^{-1})$ with z representing the z transform.

Low frequency components generated from the strided convolutions are passed down in the encoder. The results of upsampling from our decoder transpose convolutions are concatenated with the high frequency components retrieved from the skip connections. Therefore, the network is forced to learn from both the spatial and the frequency domain, retaining abundant high frequency feature content improving the sharpness and contrast in the reconstructed images while learning the color mapping from hazy to haze-free images [9].

3.3. CBAM

The convolution block attention module was initially proposed by [34] for the domains of image classification and object detection. Their main purpose is to infer attention maps along the channel and spatial dimensions without modifying the input dimension. This helps the network to improve the representation of interest by focusing on the important features and suppressing the unnecessary ones. The process is described in Equation 1, where F is the input feature map, M_c is a 1D channel attention map, M_s is a 2D spatial attention map, F' is the feature map defined after the channel attention mapping, F'' is the final refined output, and $*$ denotes the convolution operation.

$$\begin{aligned} F' &= M_c(F) * F \\ F'' &= M_s(F') * F' \end{aligned} \tag{1}$$

We propose the use of CBAM modules, especially channel attention, to improve color correction. Strong color correction in the CNNs requires good integration of multiple feature channels together. This integration is highly dependent on the weighting of these feature channels throughout the network. Hence, using the CBAM helps to learn the most effective relative weighting of these channels for reconstruction. One could even say CBAM modules are better suited for strongly channel-dependent applications like underwater image improvement.

Instead of using the CBAM in the U-Net throughout the skip connections like suggested by [35], we decide to use a similar approach to the one suggested by [21]. Specifically, we place these blocks at the encoder while down-sampling before each convolution layer, to ensure a more optimal feature extraction and make the network focus its attention on the main characteristics before losing information during down-sampling. Additionally, we use the CBAM in the decoder as well. We apply the attention module after we concatenate all input skip connections to ensure we have a good synthesis of these channels from multiple sources (DTW skip connections, high frequency skip connections, and upsampled feature maps) for the transpose convolution.

3.4. Deblurring branch (Model variant)

Inspired by [20] and [9] we propose an augmentation to our initial structure to improve performance at the detriment of some efficiency. We use a second parallel CNN in conjunction with our proposed architecture and then synthesize our final reconstruction from both of these models. The idea for this is to predict a residual image, similarly to methods previously used in physical models, for the dehazing task

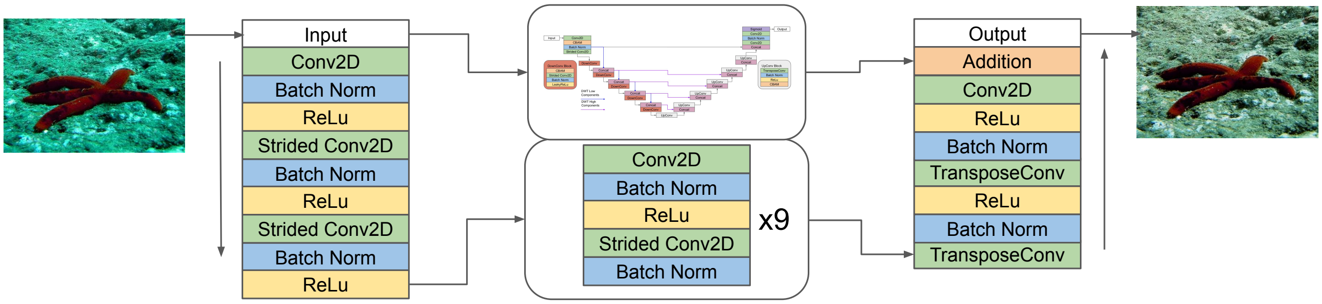

Figure 3. Proposed architecture variant with Res2Net second branch in parallel. The parallel Branch is simply added with the final activation function or Sigmoid for final image reconstruction. Notably the processing only down-samples the input image by a factor of 4, then symmetrically up-samples to the original resolution.

and input this knowledge to the DWT branch with the simple summation, see Figure 3. Due to the deblurring nature, the output image quality might be reduced. Then, we found that including the gradient penalty will compensate for this effect making the model more robust and getting sharper images. This approach increases network complexity and processing time per image but overall achieves better performance in the underwater image enhancement metrics, as shown in Table 2.

For this branch, we use the lightweight CNN proposed by [17] which is similar to [16] initially presented for the style transfer task. With this architecture, we aim to integrate the residual image knowledge and shallower encoder features to enhance the underwater images, specifically their sharpness by only processing the image at a higher resolution. It contains nine residual blocks (convolutional layer, instance normalization layer, ReLu activation function, and Dropout with probability of 0.5), two stride convolution blocks with $\frac{1}{2}$ stride, and two transposed convolution blocks. This architecture was found to train faster and generalize better due to a unique global skip connection [17]. Furthermore, since back scattering is usually homogeneous in our domain, we avoid unnecessary pixel shuffling layers previously suggested by [9] for terrestrial dehazing.

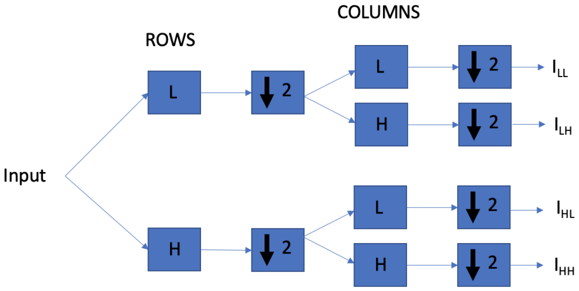

Figure 4. Two dimension DWT algorithm illustrated as a repetition of one dimension DWT first applied into rows and then into columns. L makes reference to the low-pass filter, H to the high-pass filter and 2 means down-sampling by two

3.5. Gradient penalty (Wasserstein GAN variation)

GANs (generative adversarial networks) were proposed by [12]. The idea of using GAN is to define a game between two competitive networks, the discriminator and the generator [17]. In the context of image enhancement, the generator receives a noise image as input and attempts to generate a clean output. Meanwhile, the discriminator receives generator output (improved image) as input along with a real clean image and tries to distinguish between them. The reward for the generator is associated with fooling the discriminator and the reward for the discriminator is associated with correctly distinguishing the improved image against the target goal image.

Motivated by issues with GAN convergence in this formulation we utilize the contributions of [4] who described GANs training difficulties caused by the Jenssen-Shannon divergence. To solve this divergence [13] proposed the addition of a gradient penalty into the discriminator to enforce the Lipschitz constraint. Compared to other methods to solve this training instability (like gradient clipping) this variation makes the training more robust and requires almost no hyperparameter tuning [17].

3.6. Loss Function

We denoted the recovered enhanced image as I^e, the underwater original input image as I^u, the groundtruth image as I^g, and the GAN generator and discriminator as G and D respectively.

3.6.1 Smooth L1 Loss

Let $I_c(i)$ be the intensity in pixel i of the c-th channel and N the total number of pixels.

$$L1_{smooth} = \frac{1}{3N}\sum_{i=1}^{N}\sum_{c=1}^{3}\alpha(e) \qquad (2)$$

Where e is the error $(I_c^e(i) - I_c^g(i))$, and $\alpha(e)$ is a function of it defined as:

$$\alpha(e) = \begin{cases} 0.5e^2, & \text{if } |e| < 1 \\ |e| - 0.5, & \text{otherwise} \end{cases} \tag{3}$$

3.6.2 MS-SSIM Loss

As explained by [9], denotes two windows of common size E and U centered at pixel i in the recovered enhanced image and the underwater input image respectively. Then, a Gaussian filter is applied to each window, and means μ_E,μ_U standard deviation σ_E,σ_U, and covariance σ_{EU} are computed. The SSIM for pixel i is defined as:

$$SSIM(i) = \frac{2\mu_E\mu_U + C_1}{\mu_E^2 + \mu_U^2 + C_1} * \frac{2\sigma_{EU} + C_2}{\sigma_E + \sigma_U + C_2} = l(i)*cs(i) \tag{4}$$

where $l(i)$ represents luminance and $cs(i)$ represent contract and structure measures. Also, C_1 and C_2 are constraints to stabilize division with a weak denominator. The MS-SSIM loss uses M levels of SSIM (1-MS-SSIM).

$$MS(i) - SSIM(i) = l_M^{\alpha}(i) * \prod_{m=1}^{M} cs_m^{\beta_m}(i) \tag{5}$$

with α and β default parameters suggested in the original study [33].

3.6.3 Adversarial loss

Since we are using WGAN-GP for training, the adversarial loss is computed as:

$$L_{GAN} = \sum_{n=1}^{N} -D(G(I^u)) \tag{6}$$

3.6.4 Total Loss

The total loss is the combination of losses defined as:

$$L_{total} = L1_{smooth} + R1 * L_{Ms-SSIM} + R2 * L_{GAN} \tag{7}$$

Where R1=0.2 and R2=0.005 are hyperparameters inspired by optimal performance showed in [9].

4. Results and Discussion

We present state of the art results on the UIEBD dataset [19] while operating at a greater than real time inference speed.

4.1. Dataset

For our experiments we used a real-world underwater dataset: UIEBD [19]. This dataset contains 890 underwater images and their corresponding groundtruth. These groundtruth images are based on the human-subjective ratings and were obtained after studying 12 different underwater enhanced images. We use standard splits for direct comparison [11]. The first 700 images are used for training or validation and the remaining 190 for testing. Also, for the qualitative color comparison, we used the Color Checker dataset [3].

4.2. Training Details

We use Adam optimizer with an initial learning rate 1e-4 for 400 epochs to train the generator (our model) and the discriminator. We decrease the learning rate by half on epochs 250 and 350. Our models were trained on a GPU NVIDIA GeForce RTX 3060 with 12GB of memory and used a batch size of 2. For the dataset, each image was randomly cropped in patches of 256x256. Then, augmentation was performed by randomly rotating it 90, 180, or 270 degrees and horizontal flipping. No other augmentations were performed to ensure the model was specifically tailored to underwater image corruption.

4.3. Ablation

We performed different experiments to show that each module of our proposed architecture helps to improve the enhancement of underwater images. The multiple variations of the model are shown in Table 1. It is possible to appreciate that using a different DWT filter increases problem complexity and does not improve the scores. Also, note that the deblurring branch introduces noise (PSNR decreases). Then, it is necessary to use the gradient penalty while training to compensate for the sharpness and color quality, leading to higher scores.

4.4. Quantitative Results

Our proposed method shows state-of-the-art results on this dataset in both SSIM and inference speed. The details of our results are shown in Table 2.

We present two configurations of our method: one minimal configuration and a configuration with a second parallel deblurring branch. Our minimal implementation shows state-of-the-art SSIM results on UIEBD task with a score of 0.8616. This is only a slight improvement over other methods such as the methods proposed by DWG but we show a nearly 18x speed-up over their method while improving the SSIM scores. Further, when we introduce the second deblurring branch we show a larger gap to other methods achieving an SSIM score of 0.8802, almost 2 percentage points higher than the next comparable method. This mod-

Component				Average SSIM	Red SSIM	PSNR	Inference time (s)
CBAM	**DWT**	**Deblurring**	**Grad Penalty**				
✓	Haar	✓	✓	0.8802	0.8465	20.9226	0.3266
✓	Haar	✓		0.8651	0.8264	19.8227	0.3267
✓	Haar			0.8703	0.8434	20.8694	0.0248
✓				0.8683	0.8265	20.3843	0.02324
				0.8637	0.8126	19.4034	0.01457
✓	Biorthogonal			0.8606	0.8046	19.4747	0.3094

Table 1. Ablation study showing individual improvement of each module of the proposed architecture using UIEBD dataset.

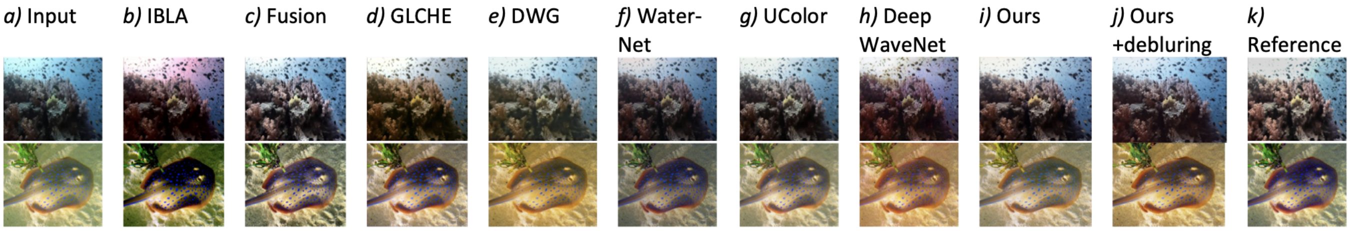

Figure 5. Results of different models for UIEBD dataset for qualitative comparison. These same models are used during the quantitative analysis.

Method	PSNR↑	SSIM↑	Inference Time↓ (s)	Type
IBLA [26]	14.3856	0.4299	38.71	Classic
Fusion [2]	21.1849	0.8222	6.58	Classic
GLCHE [10]	21.027	0.8487	0.05	Mix
DWG [9]	19.6727	0.8614	0.4487	Deep
Water-Net [19]	19.3134	0.8303	0.61	Deep
Ucolor [18]	20.63	0.77	2.75	Deep
SCNet [11]	**22.08**	0.8625	0.4495	Deep
Deep WaveNet [15]	21.57	0.8	1.16	Deep
Ours	20.8694	0.8703	**0.025**	Deep
Ours + deblurring branch	20.9226	**0.8802**	0.3266	Deep

Table 2. Quantitative results on the UIEDB [32] test set. We show state-of-the-art results in SSIM and inference speed while having competitive results in Peak Signal to Noise Ration (PSNR). Best scores shown in bold, second best underlined.

erate increase in scores over our minimal method does come with significant costs with regards to speed.

Notably, our method does not gain state of the art results in PSNR in either configuration, only achieving 20.8694 and 20.9226 with our minimal and deblurring branch configurations respectively. We argue this is not a significant blemishing, because PSNR is not as good of a tool for image quality especially in this domain of underwater image improvement and on this dataset [30][31][19]. The goal of this dataset and our method and this dataset was qualitative improvements on images, even the ’ground truths’ in this dataset are based purely on subjective studies [19]. For this qualitative goal, SSIM is a much better metric than PSNR as it is based on luminance, contrast, and structure as a human perception analog rather than an unnormalized absolute divergence metric like PSNR [30][31].

It is possible to observe that prior-based methods (classic) are computationally expensive and do not achieve the best scores in the evaluation metrics. There are several causes for this. First, the classic methods generally require more complicated pipelines to achieve strong performance as they are required to generalize for many possible underwater image degradation cases. Secondly these methods, including and beyond those presented in Table 2, use histogram based approaches, leading to parallelization on specialized hardware impractical (i.e. GPUs). In general, as statistical deep learning methods use CNNs to process the underwater images they are able to leverage this GPU parallelization to be faster than existing classic methods. The fastest method outside of the one we propose for image enhancement is a hybrid method which preformed local and global statistical methods, synthesizing them together to increase their efficiency.

We present the fastest high performance method for underwater image enhancement on this dataset. Our proposed method (minimal configuration) is able to process images at a rate of 40/s on low-end consumer hardware. This is two times faster than any comparable previous method in this domain, and the only to achieve the real-time performance necessary for being deployed in applications like on controlled Unmanned Underwater Vehicles (UUV)s. We achieve this speed through a minimalist design philosophy starting simply with a well tested encoder-decoder struc-

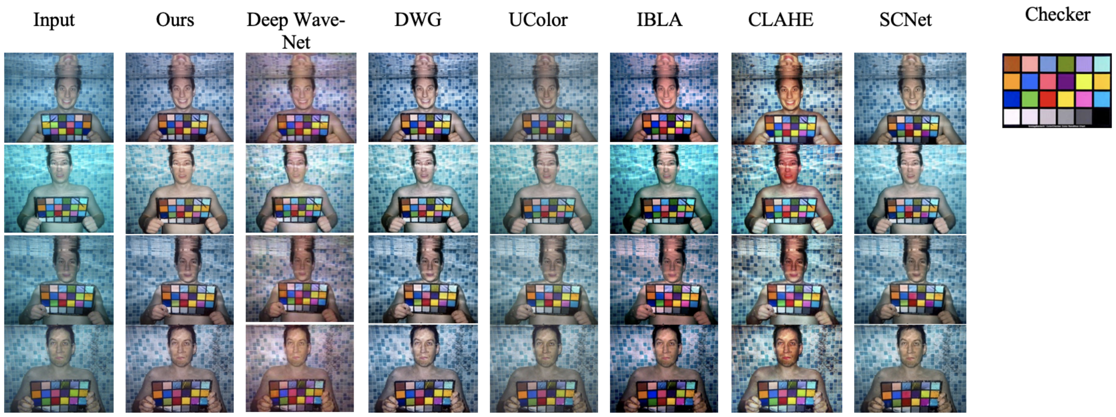

Figure 6. Results of different models for the ColorChecker dataset for qualitative comparison. These same models are used during the quantitative analysis.

ture and only adding efficient modifications which address specific challenges in underwater images. Our deblurring branch configuration does not perform at real time speeds but still shows competitive results compared to other methods with an inference speed of 0.3266s/image.

4.5. Qualitative Results

Figure 1 shows qualitative results of the proposed model in different types of water. It is possible to appreciate that our method recovers color efficiently by addressing the red channel absorption effect. Also, it retrieves information close to the truthiness colors of objects, especially in the red channel. On the other side, our model still creates brightness environment artifacts on images with objects in different depths when the green channel is more representative, visible through some examples presented in Figure 1. This might be related to the nature of the dehazing design. Nevertheless, details are retrieved avoiding unnatural looks with sharp results.

Additionally, Figure 5 presents a qualitative subjective comparison of the results achieved for each model. We appreciate that classical methods create unnatural color artifacts while our results show a closer real recovery of colors, maintaining a balance between red, green, and blue channels to create natural looking images with minimal brightness artifacts. The coral example in Figure 5 clearly shows that our model handles the red channel absorption effect efficiently and still performs well on recovering the blue, white and yellow objects. This means that our model effectively learns the weighting of color channels, justifying that attention (CBAM) is a fundamental component of the architecture. Furthermore, we might see a reduction of the brightness artifacts of our model when we add the deblurring branch, which leads to an increase in the image sharpness as shown in the second sample of Figure 5.

Furthermore, the ColorChecker dataset [3] was used for additional qualitative analysis. Some of the results are presented in Figure 6. We appreciate that classical models do not provide the best color recovery and introduce unnatural brightness artifacts. DWG and SCNet provide high quality results, however, some colors such as blue and pink are not correctly recovered. Our model seems to avoid unnatural looking artifacts and recover colors close to the true ones.

5. Conclusion

We propose an efficient real-time method for underwater image enhancement which achieves state-of-the-art results. Our quantitative analysis is primarily focused on the SSIM index since it is based on luminance, contrast, and structure as human perception. We achieved optimal results in inference time and SSIM on the UIEB dataset. Additionally to setting comparison points with other models, the UIEB dataset is ideal to demonstrate the model behavior in real underwater images and not only in synthetic data. Our proposed method effectively builds up from a minimal U-Net based encoder-decoder architecture with DWT skip connections, CBAM blocks for better color casting, domain-specific training loss for GAN training, and an optional deblurring branch configuration. Our method debuts the only real-time (40frames/second) method on this dataset while presenting state-of-the-art results on SSIM of 0.8703. Our model variant increases this improvement over other methods to an SSIM of 0.8802, outperforming state of the art results.

For future work, it will be important to aboard the underwater image enhancement requirements from an unsupervised learning perspective, since results will not depend on visual qualitative analysis for groundtruth construction. Also, this will lead to more realistic models that could act accurately in different types of water.

References

[1] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1682–1691, 2019.

[2] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 81–88, 2012.

[3] Codruta O. Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Transactions on Image Processing*, 27(1):379–393, 2018.

[4] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017. cite arxiv:1701.07875.

[5] Woong Bae, Jaejun Yoo, and Jong Chul Ye. Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1141–1149, 2017.

[6] Jonathan A Bergshoeff, Nicola Zargarpour, George Legge, and Brett Favaro. How to build a low-cost underwater camera housing for aquatic research. *Facets*, 2(1):150–159, 2017.

[7] Jian Chen, Hao-Tian Wu, Lu Lu, Xiangyang Luo, and Jiankun Hu. Single underwater image haze removal with a learning-based approach to blurriness estimation. *Journal of Visual Communication and Image Representation*, 89:103656, 2022.

[8] Fayaz Ali Dharejo, Yuanchun Zhou, Farah Deeba, Munsif Ali Jatoi, Muhammad Ashfaq Khan, Ghulam Ali Mallah, Abdul Ghaffar, Muhammad Chhattal, Yi Du, and Xuezhi Wang. A deep hybrid neural network for single image dehazing via wavelet transform. *Optik*, 231:166462, 2021.

[9] Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. Dw-gan: A discrete wavelet transform gan for non-homogeneous dehazing. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 203–212, 2021.

[10] Xueyang Fu and Xiangyong Cao. Underwater image enhancement with global–local networks and compressed-histogram equalization. *Signal Processing: Image Communication*, 86:115892, 2020.

[11] Zhenqi Fu, Xiaopeng Lin, Wu Wang, Yue Huang, and Xinghao Ding. Underwater image enhancement via learning water type desensitized representations, 2021.

[12] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. June 2014.

[13] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[14] Martin Heesemann, Tania L Insua, Martin Scherwath, Kim S Juniper, and Kate Moran. Ocean networks canada: from geohazards research laboratories to smart ocean systems. *Oceanography*, 27(2):151–153, 2014.

[15] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters (RA-L)*, 5(2):3227–3234, 2020.

[16] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, 2016.

[17] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8183–8192, 2018.

[18] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, 30:4985–5000, 2021.

[19] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2020.

[20] Jinjiang Li, Guihui Li, and Hui Fan. Image dehazing using residual-based deep cnn. *IEEE Access*, 6:26831–26842, 2018.

[21] Wenmei Li, Jiaqi Wu, Huaihuai Chen, Yu Wang, Yan Jia, and Guan Gui. Unet combined with attention mechanism method for extracting flood submerged range. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:6588–6597, 2022.

[22] Yi Li, Yang Sun, and Syed Naqvi. U-shaped transformer with frequency-band aware attention for speech enhancement, 12 2021.

[23] Yongbin Liu, Shenghui Rong, Xueting Cao, Tengyue Li, and Bo He. Underwater image dehazing using the color space dimensionality reduction prior. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1013–1017, 2020.

[24] Delphine Mallet and Dominique Pelletier. Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012). *Fisheries Research*, 154:44–62, 2014.

[25] Declan GD McIntosh, Tunai Porto Marques, Alexandra Branzan Albu, Rodney Rountree, and Fabio De Leo Cabrera. Movement tracks for the automatic detection of fish behavior in videos. In *NeurIPS 2020 Workshop on Tackling Climate Change with Machine Learning*, 2020.

[26] Yan-Tsung Peng and Pamela Cosman. Underwater image restoration based on image blurriness and light absorption. *IEEE Transactions on Image Processing*, PP:1–1, 02 2017.

[27] Tunai Porto Marques and Alexandra Branzan Albu. L2uwe: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion. In *Proceedings of the IEEE/CVF Conference on Computer*

Vision and Pattern Recognition Workshops, pages 538–539, 2020.

[28] Tunai Porto Marques, Alexandra Branzan Albu, and Maia Hoeberechts. A contrast-guided approach for the enhancement of low-lighting underwater images. *Journal of Imaging*, 5(10):79, 2019.

[29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[30] Umme Sara, Morium Akter, and Mohammad Shorif Uddin. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, 7(3):8–18, 2019.

[31] De Rosal Igantius Moses Setiadi. Psnr vs ssim: imperceptibility quality assessment for image steganography. *Multimedia Tools and Applications*, 80(6):8423–8444, 2021.

[32] Muhammad Aldila Syariz, Chao-Hung Lin, Manh Van Nguyen, Lalu Muhamad Jaelani, and Ariel C. Blanco. Waternet: A convolutional neural network for chlorophyll-a concentration retrieval. *Remote Sensing*, 12(12), 2020.

[33] Z. Wang, Eero Simoncelli, and Alan Bovik. Multiscale structural similarity for image quality assessment. volume 2, pages 1398 – 1402 Vol.2, 12 2003.

[34] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module, 2018.

[35] Zhengxuan Zhao, Kaixu Chen, and Satoshi Yamane. Cbam-unet++:easier to find the target with the attention module ”cbam”. In *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, pages 655–657, 2021.