Uncertainty Inspired Underwater Image Enhancement

Zhenqi Fu[1], Wu Wang[1], Yue Huang[1], Xinghao Ding[1]*, and Kai-Kuang Ma[2]

[1] Xiamen University, Fujian, 361005, China
{fuzhenqi,23320170155546}@stu.xmu.edu.cn, {yhuang2010,dxh}@xmu.edu.cn
[2] Nanyang Technological University, 639798, Singapore
ekkma@ntu.edu.sg

Abstract. A main challenge faced in the deep learning-based Underwater Image Enhancement (UIE) is that the ground truth high-quality image is unavailable. Most of the existing methods first generate approximate reference maps and then train an enhancement network with certainty. This kind of method fails to handle the ambiguity of the reference map. In this paper, we resolve UIE into distribution estimation and consensus process. We present a novel probabilistic network to learn the enhancement distribution of degraded underwater images. Specifically, we combine conditional variational autoencoder with adaptive instance normalization to construct the enhancement distribution. After that, we adopt a consensus process to predict a deterministic result based on a set of samples from the distribution. By learning the enhancement distribution, our method can cope with the bias introduced in the reference map labeling to some extent. Additionally, the consensus process is useful to capture a robust and stable result. We examined the proposed method on two widely used real-world underwater image enhancement datasets. Experimental results demonstrate that our approach enables sampling possible enhancement predictions. Meanwhile, the consensus estimate yields competitive performance compared with state-of-the-art UIE methods. Code available at https://github.com/zhenqifu/PUIE-Net.

Keywords: underwater image enhancement, deep learning, probabilistic network, adaptive instance normalization, conditional variational autoencoder

1 Introduction

Underwater images suffer from degradation due to the poor and complex lighting conditions in the water. The degradation of underwater images is mainly rooted in the wavelength-dependent light scattering and absorption, which reduces visibility, decreases contrast, and introduces unpleasant color casts. It is important and necessary to develop Underwater Image Enhancement (UIE) methods to adjust the degraded underwater signal so that the results are more suitable for

---

* Corresponding author.

display or further analysis. In the past few years, deep learning-based UIE approaches have enabled tremendous progress. Commonly, this kind of method adopts corresponding pairs of clean and distorted images to learn a mapping between two quality levels [36, 59, 39, 14].

UIE is an important low-level vision task. For underwater scenes with adverse visual conditions, it is difficult and impractical to capture the clean image to train a deep neural network directly. This is because the degradation of underwater images is non-reversible and very complex. To solve this problem, previous methods propose to generate approximate supervisors to train a deep neural network. For example, in [36] the authors applied several state-of-the-art algorithms to generate a set of potential reference images and manually select the best one as the ground truth. Benefiting from the large-scale dataset constructed in previous works, deep learning-based methods have made profound progress in learning the mapping from a degraded underwater image to the corresponding high-quality reference image [34]. Nonetheless, considering the progress of UIE under this pipeline, we would like to argue that this kind of method fails to capture the uncertainties in labeling the ground truth.

Although the reference image achieves high visual quality in existing UIE datasets, it is generated through an approximate approach (e.g., using exiting UIE algorithms [36]) and may be influenced by various factors, such as human-specific preference during subjective selection and different algorithm parameters. In this case, UIE suffers from uncertainty issues. For ambiguous labels, directly learning a mapping between the degraded underwater image and corresponding reference is inappropriate. There are many potential solutions for the same degraded underwater image because we cannot confidently know what a true clean image looks like. Nonetheless, most of the existing deep learning-based methods treat UIE as a point estimation problem. As a result, they have to make a compromise between possible solutions because they are following a deterministic learning pipeline.

In this paper, we propose the first probabilistic network for UIE, termed PUIE-Net. Instead of directly generating a single prediction (i.e., the point estimation), we are interested in how the network produces multiple results (i.e., the distribution estimation) so that the network can handle the uncertainty issue in UIE. Furthermore, once the distribution is estimated, we can perform a consensus process to capture a deterministic result based on a set of estimations. In this paper, we introduce two consensus processes to predict the final result named: Monte Carlo likelihood estimation (MC) [51] and Maximum Probability estimation (MP). Specifically, MC is calculated by taking the average of the likelihoods. While the sample with maximum probability is regarded as the final result in MP.

The proposed network structure is based on probabilistic adaptive instance normalization (PAdaIN), which combines conditional variational autoencoder (CVAE) [48] with adaptive instance normalization (AdaIN) [23] to construct the enhancement distribution. PAdaIN is motivated by AdaIN that is originally designed for style transfer. We extend AdaIN into PAdaIN via drawing style

inputs from two posterior distributions constructed by conditional variational autoencoders. PAdaIN aims to transform the global enhancement statistics of input features. Therefore, diverse predictions can be achieved by sampling different enhancement attributes from the latent space. The whole pipeline of our method is trained following the standard training procedure of the CVAE. We conduct extensive experiments on two real-world UIE datasets to validate the effectiveness of our method.

Our main contributions are summarized as follows:

– We resolve UIE into distribution estimation and consensus process to handle the uncertainty issue in labeling the ground truth.
– We propose the first probabilistic network for UIE, which learns to approximate the posterior over meaningful appearance. Specifically, the enhancement distribution is constructed based on conditional variational autoencoder and adaptive instance normalization.
– We show that our method can generate diverse potential solutions. Besides, by inferring the consensus prediction based on a set of samples, our method achieves promising performance compared with state-of-the-art methods on two UIE datasets.

2   Related Work

2.1   Underwater Image Enhancement

According to the means of the modeling imaging process, the existing UIE methods can be roughly categorized into the following three types.

The first category is model-free methods, which enhance underwater images without considering the degradation process. Traditional contrast limited adaptive histogram equalization (CLAHE) [44], white balance (WB) [56], and Retinex [47] belong to this category. In [7], the authors proposed a fusion-based method for UIE, where the inputs and weight measures are derived only from the degraded image. An improvement version of [7] is presented in [5], which adopts a white balancing technique and a novel multi-scale fusion strategy to further promote the enhancement performance. Fu et al. [16] proposed a retinex-based UIE approach to enhance a single underwater image. Gao et al. [18] presented a teleost fish retina-guided underwater image enhancement approach to deal with the problems of nonuniform color shift and content blurring. Other relevant works can be found in [6, 19].

The second category is prior-based methods, which enhance underwater images using physical imaging models and focus on accurately estimating the parameters of the defined physical model. Chiang et al. [12] proposed to enhance underwater images via a dehazing algorithm. Galdran et al. [17] proposed a variant of the Dark Channel Prior (DCP) [21] that uses red channel information to estimate the depth map of underwater images. Li et al. [33] proposed an underwater image dehazing algorithm and a contrast enhancement method based on a minimum information loss and histogram distribution prior. Berman et al. [11]

took into account multiple spectral profiles of different water types, in which the authors additionally estimated two global parameters, i.e., the attenuation ratios of the blue-red and blue-green color channels. Akkaynak et al. [3] developed a UIE method named Sea-thru based on a revised physical imaging model. Sea-thru takes RGBD images as the input and it first estimates backscatter using the darkest pixels and their known range map. Then it calculates the attenuation coefficient based on an estimation of the spatially varying illuminant. Other relevant works of prior-based UIE methods can be found in [35, 2, 43, 42, 10].

The third category is deep learning-based methods that automatically extract representations and learn an enhancement mapping based on numerous paired/unpaired training data. Li et al. [38] first proposed a generative adversarial network to generate synthetic underwater images in an unsupervised pipeline. Then the authors trained an enhancement network using these synthetic data. Li et al. [37] proposed a weakly supervised underwater image enhancement method that relaxes the need for paired data. Guo et al. [20] enhanced degraded underwater images using a multi-scale dense generative adversarial network. Li et al. [35] introduced a lightweight UIE model based on the underwater scene prior. Li et al. [36] constructed a large scale real-world UIE dataset. The reference image is generated by 12 existing UIE methods. Besides, based on this dataset, the authors proposed a gated fusion network for enhancing underwater images. Jamadandi et al. [25] proposed to enhance underwater images by augmenting the network with wavelet corrected transformations. To deal with the challenge of underwater image degradation diversity, Uplavikar et al. [53] trained a new deep neural network to learn the domain agnostic features for a given degraded underwater image, where the domain is the Jerlov water type of the image. Li et al. [34] presented an underwater image enhancement network called Ucolor by medium transmission-guided multi-color space embedding. Kar et al. [29] presented a zero-shot underwater and hazing image restoration method by leveraging a theoretically deduced property of degradation through the physical model. More relevant works of learning-based UIE methods can be found in [58, 15, 27, 55, 41, 24, 46, 26].

2.2   VAE based Deep Probabilistic Model

Variational autoencoder (VAE) and its conditional counterpart (CVAE) [51, 48] have been widely used in various computer vision tasks. Rather than building an encoder that outputs a single value to describe each latent state attribute, VAE formulates the encoder to describe a probability distribution for each latent attribute. To train a VAE, a regularizer and a reconstruction loss are needed to penalize the disagreement of the posterior and prior distributions of the latent representation.

Despite not being explored to model ambiguities for underwater image enhancement, VAEs and CVAEs are utilized to sample diverse results from constructed posteriors. For example, in [32], the authors use VAEs to model the image background for salient object detection. In [57], the authors apply VAEs for learning motion sequence generation. In [30], a probabilistic U-Net based on

VAEs is proposed to learn a conditional distribution of medical image segmentation. [9] and [31] improve the diversity of samples in [30] via adopting a hierarchy of latent variables. In [1], a contrastive VAE is introduced which combines the benefits of contrastive learning with the power of VAEs to identify and enhance salient latent features. In [45], a VAE based denoising approach is developed by predicting a whole distribution of denoised images. In [8], the authors employ VAEs to predict multiple deprojected instances for images/videos collapsed along a dimension. In [60], the authors employ VAEs for RGB-D saliency detection by learning from the data labeling process. Other relevant works about VAEs for diverse solution sampling can be found in [4, 13, 52]

3   Method

In this section, we will introduce PUIE-Net in detail. PUIE-Net is based on PAdaIN that learns meaningful distributions of UIE. It is the first time that employs a probabilistic network to solve the UIE problem.

3.1   Motivation

Ambiguities for UIE The main idea of PUIE-Net is to introduce ambiguities for UIE. This is because the true clean image is unavailable and there is a degree of uncertainties in recording approximate labels. Existing deterministic learning-guided methods fail to capture such uncertainty and have to make a compromise between possible results. We consider that as we cannot confidently know which of the possible clean image has given rise to the distorted underwater image at hand, estimating the distribution of possible interpretations may be an advisable solution. We use an implicit variable $\mathbf{z}$ to express the uncertainty. Here, $\mathbf{z}$ can be interpreted as human subjective preferences or camera/algorithm parameters in capturing the ground truth. Let $\mathbf{x}$ and $\mathbf{y}$ refer to the corrupted observation and the clean image, respectively. UIE under a probabilistic framework can be formulated as:

$$p\left(y\left|\mathbf{x}\right.\right) \approx \frac{1}{S}\sum_{s=1}^{S} p\left(\mathbf{y}\left|\mathbf{z}^{(s)},\mathbf{x}\right.\right), \mathbf{z}^{(s)} \sim p\left(\mathbf{z}\left|\mathbf{x}\right.\right) \tag{1}$$

where $p\left(\mathbf{z}\left|\mathbf{x}\right.\right)$ denotes the distribution of uncertainty. S represents the number of samples. Equation (1) not only allows us to generate multiple enhancement predictions but also gives a straightforward way to calculate a deterministic result, i.e., the MC estimation [51] that draws samples using the prior network and takes the average of the likelihoods. Apart from MC, the Maximum Probability estimation (MP) is also considered in this paper. MP takes the enhancement sample with the maximum probability as the final result. Mathematically, MP can be expressed as:

$$p\left(y\left|\mathbf{x}\right.\right) \approx p\left(\mathbf{y}\left|\mathbf{z}_{\max},\mathbf{x}\right.\right), \mathbf{z}_{\max} \sim p\left(\mathbf{z}\left|\mathbf{x}\right.\right) \tag{2}$$

where $\mathbf{z}_{\max}$ denotes the sample with the maximum probability.

PAdaIN for Learning Appearance Distribution Since the goal of UIE is to adjust the image appearance such as colors and contrasts, rather than the content, it is important to capture such information during the enhancement. Here, we adopt a modified AdaIN [23] to capture such features. AdaIN is originally developed for style transfer, which can be expressed as:

$$\text{AdaIN}\,(\mathbf{x}, \mathbf{y}) = \boldsymbol{\sigma}(\mathbf{y}) \left(\frac{\mathbf{x} - \boldsymbol{\mu}\,(\mathbf{x})}{\boldsymbol{\sigma}\,(\mathbf{x})} \right) + \boldsymbol{\mu}(\mathbf{y}) \tag{3}$$

where $\mathbf{x}$ denotes the features of the content image and $\mathbf{y}$ denotes the features of the style image. $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ refer to the mean and standard deviation operations, respectively. AdaIN changes the appearance by adjusting the mean and standard deviation of features. We observe that UIE falls into the AdaIN paradigm. However, AdaIN relies on the known content and style images, it cannot be directly leveraged for UIE. To handle this problem, we propose PAdaIN, which can be formulated as:

$$\text{PAdaIN}\,(\mathbf{x}) = \boldsymbol{b} \left(\frac{\mathbf{x} - \boldsymbol{\mu}\,(\mathbf{x})}{\boldsymbol{\sigma}\,(\mathbf{x})} \right) + \boldsymbol{a} \tag{4}$$

where $\boldsymbol{b}$ and $\boldsymbol{a}$ are two random samples from the posterior distributions of the mean and standard deviation, respectively. Specifically, the posterior distributions can be learned with CVAEs, which will be detailed in the next subsection. Note that the proposed PAdaIN can be regarded as a generalized AdaIN.

3.2   Network Structure

The whole pipeline of PUIE-Net is illustrated in Fig. 1. The proposed network structure contains two branches. Both branches include a U-Net [49] based feature extractor. Specifically, the top branch aims to estimate the prior distribution of a single raw underwater image. While the goal of the bottom branch is to construct posterior distributions of UIE, it takes the raw underwater image and corresponding reference image as the input. In PUIE-Net, we simply modify the U-Net by adding SE-ResNet blocks [22]. This is useful to improve the network capacity of enhancement representations.

The core component of PUIE-Net is the PAdaIN after the feature extractor to encode the ambiguities. The prior/posterior block (i.e., Pr/Po block in Fig. 1) is designed to build the distribution of enhancement. We note that both Pr and Po need to construct a mean and a standard deviation distributions. After that, embeddings sampled form Pr/Po block are input to the AdaIN to transform the feature statistics. Let $\boldsymbol{f} \in \mathbb{R}^{B \times C \times H \times W}$ refers to the data matrix of Pr/Po block's input, where B, C, H, W indicate the batch size, number of channels, the height, and the width, respectively. First, we calculating the mean and standard deviation of each channel of $\boldsymbol{f}$. Then, we adopt 1×1 convolutions to obtain $\boldsymbol{\mu} \in \mathbb{R}^{B \times N \times 1 \times 1}$ and $\boldsymbol{\sigma} \in \mathbb{R}^{B \times N \times 1 \times 1}$ from the mean vector. Similarly, we adopt 1×1 convolutions to obtain $\boldsymbol{m} \in \mathbb{R}^{B \times N \times 1 \times 1}$ and $\boldsymbol{v} \in \mathbb{R}^{B \times N \times 1 \times 1}$ from the standard deviation vector. Finally, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are applied to build the N-dimensional Gaussian distribution of the mean ($\mathcal{N}_{\mathrm{m}}$). $\boldsymbol{m}$ and $\boldsymbol{v}$ are applied to
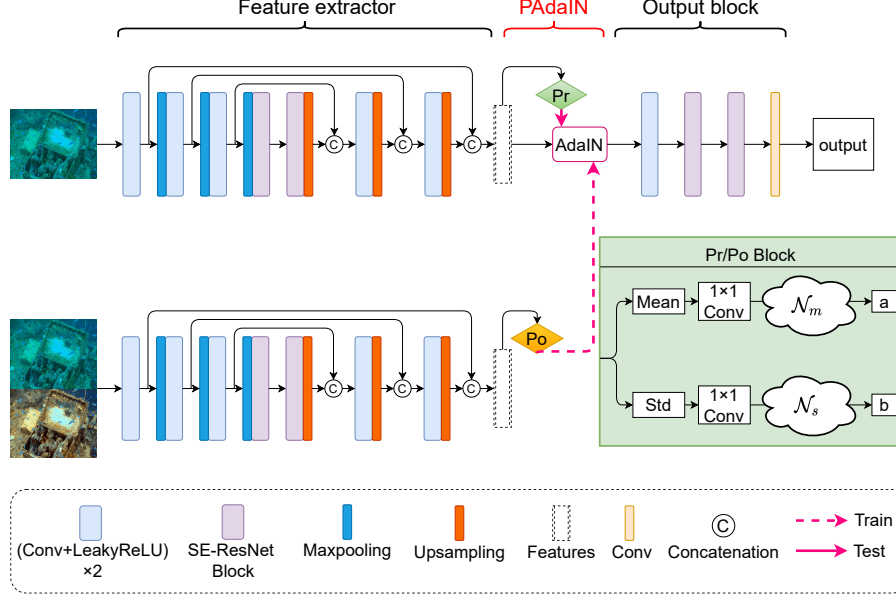
Fig. 1. The network architecture of PUIE-Net. The feature extractor is based on U-Net [49] that maps the input to representations. After the feature extractor is the PAdaIN module, which transforms the enhancement statistics of received deterministic features. **In the training phase,** features from the bottom branch are used to calculate the posterior distribution. Random samples from the posterior distributions are injected into AdaIN to transform the enhancement representation. **In the testing phase,** taking a single degraded image as the input, random samples from the Pr block are employed to generate the predictions.

build the N-dimensional Gaussian distribution of the standard deviation $(\mathcal{N}_{\mathrm{s}})$. As the two distributions are constructed, we extract random samples from them, which can be expressed as:

$$\boldsymbol{a} \sim \mathcal{N}_{\mathrm{m}}\left(\boldsymbol{\mu}\left(\mathbf{x}\right), \boldsymbol{\sigma}^2\left(\mathbf{x}\right)\right) \tag{5}$$

$$\boldsymbol{b} \sim \mathcal{N}_{\mathrm{s}}\left(\boldsymbol{m}\left(\mathbf{x}\right), \boldsymbol{v}^2\left(\mathbf{x}\right)\right) \tag{6}$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are two random samples from the mean and standard deviation distributions, respectively. Note that, in the testing phase, the latent codes ($\boldsymbol{a}$ and $\boldsymbol{b}$) are only dependent on the input image $\mathbf{x}$. While in the training phase we leverage input image $\mathbf{x}$ and corresponding reference image $\mathbf{y}$ to learn the posterior distributions, which will be described in the next subsection.

Random samples $\boldsymbol{a}$ and $\boldsymbol{b}$ are further injected into the AdaIN module to transform the statistics of received features. Typical AdaIN receives a content input and a style input, and simply aligns the mean and standard deviation of the content input to match those of the style input across channels. For UIE, the style input is unavailable in advance. Therefore, we propose to align the

mean and standard deviation of received features based on random activations extracted from the learned distributions.

3.3   Training and Testing

PUIE-Net is trained following the standard training procedure of CVAE, i.e., by minimizing the variational lower bound (Eq. 12). The main difference with respect to training a deterministic enhancement model is that our training process additionally needs to find a meaningful embedding of the enhancement statistics in the latent space. This is addressed by introducing a posterior network (i.e., the bottom branch in Fig. 1), that learns to recognize the posterior features, and map those to two posterior distributions of the mean and standard deviation. Samples from the posterior distributions can be formalized as:

$$a \sim \mathcal{N}_{\mathrm{m}}\left(\boldsymbol{\mu}\left(\boldsymbol{y}, \mathbf{x}\right), \boldsymbol{\sigma}^2\left(\boldsymbol{y}, \mathbf{x}\right)\right) \tag{7}$$

$$b \sim \mathcal{N}_{\mathrm{s}}\left(\boldsymbol{m}\left(\boldsymbol{y}, \mathbf{x}\right), \boldsymbol{v}^2\left(\boldsymbol{y}, \mathbf{x}\right)\right) \tag{8}$$

where a and b are two random samples from the mean and standard deviation posterior distributions, respectively. During the training, random samples a and b are fed into the AdaIN module to predict the enhanced image. The enhancement loss L_{e} penalizes the differences between the output of PUIE-Net and the reference. L_{e} is formulated as:

$$L_{\mathrm{e}} = L_{\mathrm{mse}} + \lambda L_{\mathrm{vgg16}} \tag{9}$$

where L_{mse} denotes the mean square error loss and L_{vgg16} denotes the perceptual loss explored by [28], λ refers to the weight.

Apart from minimizing the enhancement loss, Kullback-Leibler (KL) divergences are employed to assimilate the posterior distributions and the prior distributions:

$$L_m = D_{\mathrm{KL}}\left(\mathcal{N}_{\mathrm{m}}\left(\mathbf{x}\right) \| \mathcal{N}_{\mathrm{m}}\left(\boldsymbol{y}, \mathbf{x}\right)\right) \tag{10}$$

$$L_s = D_{\mathrm{KL}}\left(\mathcal{N}_{\mathrm{s}}\left(\mathbf{x}\right) \| \mathcal{N}_{\mathrm{s}}\left(\boldsymbol{y}, \mathbf{x}\right)\right) \tag{11}$$

where D_{KL} refers to the KL divergence between two distributions. Finally, the total loss function for training PUIE-Net is the weighted sum of above losses:

$$L = L_e + \beta(L_m + L_s) \tag{12}$$

where β is the weight. In the testing phase, we apply the network n times to the same input image to predict n enhancement variants. Note that only PAdaIN and the output block need to be re-evaluated. Diverse enhancement solutions provide users with multiple alternative results for display or analysis. More importantly, a set of samples provide sufficient inferring data for the consensus process. In this paper, the default consensus processes are the Monte Carlo likelihood estimation (MC) and Maximum Probability estimation (MP). MC predicts a final result via averaging a group of possible samples. MP takes the enhancement sample with the maximum probability as the final estimation. Equations (1) and (2) in section 3.1 describe the formulation of MC and MP, receptively.
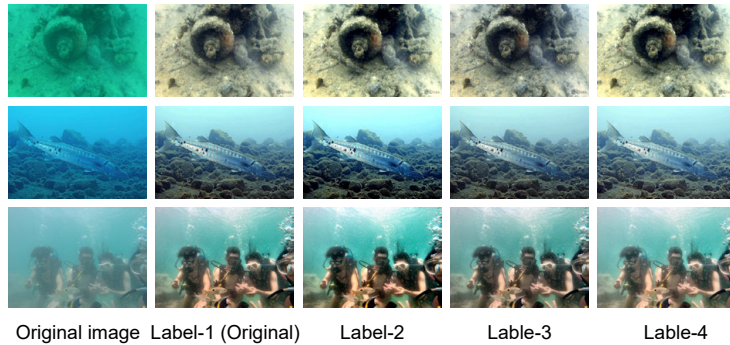
| Original image | Label-1 (Original) | Label-2 | Lable-3 | Lable-4 |

Fig. 2. Examples of our new UIE dataset. Label 2-4 denote the new labels generated by contrast adjustment, saturation adjustment, and gamma correction, respectively.

3.4 Training Data Generation

One obstacle before training our probabilistic network is that the existing UIE dataset generally only provides a single reference map for each degraded underwater image. To apply the probabilistic network, we re-build the existing UIE dataset by generating multiple reference images. Our new dataset is based on UIEBD [36], a real-world UIE dataset that contains 890 underwater images and corresponding reference maps. In the original UIEBD, the authors utilize 12 state-of-the-art enhancement algorithms to generate the potential ground truth. With raw underwater images and the 12 enhanced results, the authors invite volunteers to perform pairwise comparisons and subjectively select the best one among twelve candidates as the final reference image. Based on UIEBD, we create ambiguities by performing contrast and saturation adjustment, and gamma correction. We adopt these methods because the distortions of underwater images are mainly reflected in contrast, saturation, brightness, and colors. Note that we aim to generate ambiguous labels rather than significantly alter the original label. The contrast and saturation adjustment are performed via a simple linear transformation $y = (x - m) \times \alpha + x$, where x and y refer to the input and output, m denotes the mean of each channel. α is the adjustment coefficient. For contrast adjustment, α is the same for all pixels. For saturation adjustment, α is determined by each pixel itself. To generate a more reliable reference map, we first create two adjusted versions (i.e., over and under adjustment) per method and then choose the better one as the potential label. As a result, we obtain four reference maps (including the original label) for each raw underwater image, which can reflect the uncertainty during the ground truth recording. We show an example of the new dataset in Fig. 2.

4   Experiments

In this section, we will first show the detailed experimental settings including training and test datasets, performance criteria, compared methods, and the

implementation details. Then, we quantitatively and qualitatively evaluate our method against several state-of-the-art algorithms on two UIE datasets.

4.1   Experiment Settings

Training and testing datasets: Two UIE datasets are used for performance verification. The first one is the new UIEBD dataset that is built in this paper. As described in section 3.4, we create multiple labels for each raw underwater image. We use the first 700 original images and corresponding reference images for training, and the rest for testing. The second dataset is RUIE [39], which only contains raw underwater images. RUIE is a large-scale underwater dataset that contains three subsets, including an underwater image quality subset (3630 images), an underwater color cast subset (300 images), and an underwater higher-level task-driven subset (300 images). In this paper, we use the underwater image quality subset for testing since it contains different levels of image quality and various underwater scenes. Note that our model is trained on UIEBD and tested on both UIEBD and RUIE.

Performance criteria: In the case of applying a probabilistic network for UIE, we not only want to compare a deterministic estimation with a unique reference image, but also we are interested in the distributions of enhancement. To analyze the learned distributions, we perform subjective comparisons by visualizing the latent space. To evaluate the enhancement performance, we adopt SSIM [54], PSNR, DeltaE (CIE2000 standard) [50], and NIQE [40] to measure the image quality objectively. SSIM, PSNR, and DeltaE are full-reference metrics calculated based on the original label in UIEBD for a fair comparison with existing methods. Note that DeltaE is used for color difference evaluation. NIQE is a no-reference metric and it does not need reference images. Additionally, we conduct subjective tests to understand how users prefer the results generated by each UIE method. We use Mean Opinion Score (MOS) to quantify the subjective evaluation. 20 participants (10 male and 10 female) are invited to join the subjective test. Raw and enhanced underwater images are simultaneously displayed on a screen. The subjective score of each image is rated on a five-level scale: 5 (excellent), 4 (good), 3 (fair), 2 (poor), and 1 (bad), according to the following measures: color distortion, contrast enhancement, naturalness preservation, brightness improvement, and artifacts. We note that since RUIE only contains raw underwater images, we adopt NIQE and MOS to measure the model performance.

Compared methods: We compare PUIE-Net with 9 UIE methods, including three model-free methods (GC, Retinex [16] and Fusion [7]), three prior-based methods (IBLA [43], Histogram-Prior [33] and Haze-line [10]), and three deep learning-based approaches (Water-Net [36], Ucolor [34], and LC-Net [26]). We record the results of all competitors by conducting the same experiments using the original implementations provided by the authors for comparison fairness.

Implementation details: PUIE-Net is implemented in the Pytorch framework and trained on an NVIDIA RTX 2080Ti GPU with ADAM optimizer. The learning rate is 1×10^{-4}, the batch size is 4, and the patch size is 256×256.

(a) Original underwater images

(b) Sampling

u-3σ   u-2σ   u-1σ   u   u+1σ   u+2σ   u+3σ
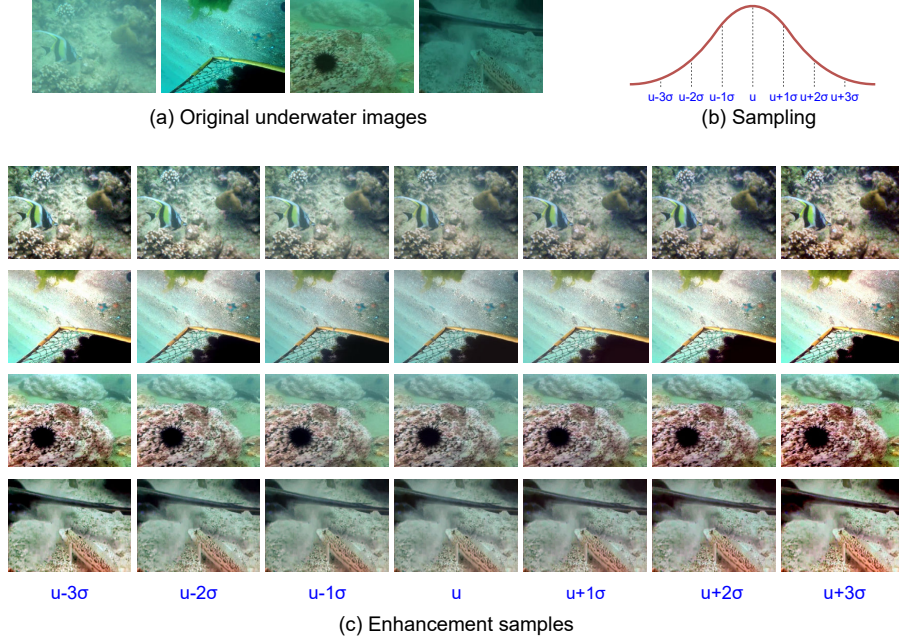
(c) Enhancement samples

Fig. 3. Visualization of the enhancement distribution of UIE. We show the original image in (a). The sampling scheme used for visualization is presented in (b) and the enhancement samples are shown in (c). μ and σ in (b) denote the mean and standard deviation of the distribution, respectively.

We augment the training data with rotation, flipping horizontally and vertically to promote network generalization. The dimension N of the latent space is 20. We adopt 1×1 convolutions to broadcast the samples to the desired number of channels before input to AdaIN. The parameter λ and β are empirically set as 1. The default kernel size of convolution layers is 3×3 and the number of channels is 64. The default sampling times for calculating MC estimations are 20.

4.2   Analysis of Enhancement Distribution

Although a lot of UIE algorithms have been developed, PUIE-Net is the first method that learns the distribution of enhancement and explicitly takes multi-solution into account. Fig. 3 shows the original image and PUIE-Net samples. We manually control the sampling interval for better visualization. From Fig. 3, we can make the following observations: 1) Each PUIE-Net sample has a different but reasonable appearance. Enhanced samples with higher sampling probability show relatively mild enhancement. Enhanced samples with lower sampling probability show strong contrast and color adjustment. This demonstrates that our method can learn the meaningful distribution of UIE and can generate diverse enhancement predictions. 2) A set of enhanced samples not only provide multiple

Table 1. Quantitative comparison on UIEBD and RUIE datasets. The best result is highlighted in red and the second best one is in blue.

Method	UIEBD					RUIE	
	PSNR ↑	SSIM ↑	DeltaE ↓	NIQE ↓	MOS ↑	NIQE ↓	MOS ↑
GC	16.14	0.761	16.11	3.789	2.6	4.656	2.4
Retinex [16]	17.53	0.773	14.82	4.074	2.7	4.593	2.5
Fusion [7]	21.18	0.822	9.079	3.747	3.4	4.488	2.9
IBLA [43]	18.51	0.762	18.64	4.290	2.2	4.767	1.9
Histogram-Prior [33]	14.39	0.573	15.69	3.780	2.4	4.486	2.5
Water-Net [36]	19.31	0.830	10.14	3.879	3.3	4.491	3.3
Haze-line [10]	14.97	0.669	17.23	3.830	2.4	4.620	1.8
Ucolor [34]	21.65	0.840	8.646	3.786	3.7	4.755	3.0
LC-Net [26]	18.54	0.812	14.25	3.800	3.3	4.721	2.9
PUIE-Net (MC)	21.86	0.870	9.556	3.626	4.2	4.512	3.7
PUIE-Net (MP)	21.05	0.854	10.26	3.668	4.1	4.555	3.6

alternatives for display or analysis but also offer effective data for the consensus process to further obtain a robust and stable result. Compared with deterministic learning-based methods which have to make a compromise between possible results, our method is more flexible and can reduce the influence of biased labels in the existing UIE dataset.

4.3   Performance Comparison

Tab. 1 presents the quantitative comparisons on UIEBD and RUIE datasets. From Tab. 1, we can perceive that PUIE-Net achieves favorable performance and outperforms other methods. Specifically, prior-based approaches obtain relatively poor results since this kind of method is highly dependent on the used prior knowledge and the predefined imaging model. We found that the performance of MC is significantly better than the others. This is reasonable since averaging a set of samples can reduce enhancement bias in the testing phase. We further present visual comparisons in Fig. 4. As can be seen, although most of the methods can somehow enhance the contrast, severe visual defects caused by the unsatisfactory adjustment of colors or artifacts remain. For example, GC and Retinex show unnatural colors and saturation, smearing image details. Prior-based methods can improve the contrasts, however, the colors are seriously degraded in these cases. Water-Net and Ucolor tend to generate over/under-enhanced results. Our method works well on all these cases and the result looks more clean and natural with fine-grained textures.

4.4   Impact of Sampling Times

Since we apply the consensus process to obtain a deterministic result, it is necessary to analyze the influence of sampling times. We calculate the standard deviation of PSNR and SSIM under different sampling times on the UIEBD
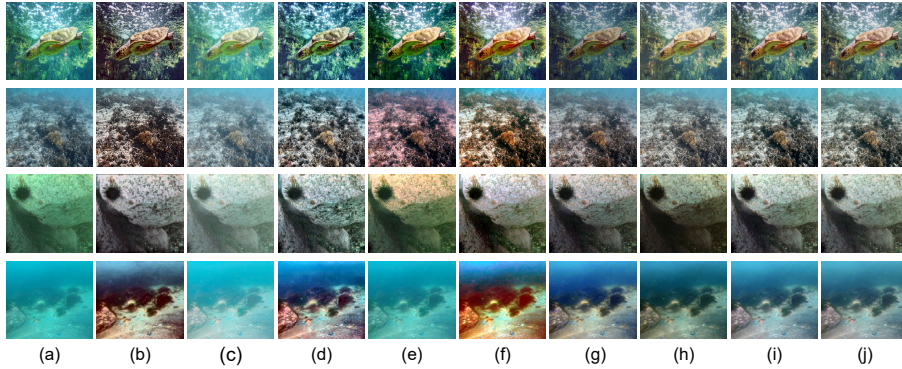
Fig. 4. Comparisons of visual results on UIEBD and RUIE datasets. (a) Original image. (b) Retinex. (c) GC. (d) Fusion. (e) IBLA. (f) Histogram-Prior. (g) Water-Net. (h) Ucolor. (i) PUIE-Net (MC). (j) PUIE-Net (MP).

Table 2. Impact of sampling times.

Sampling times	1	2	4	6	8	10	20	50
Mean std-PSNR	0.726	0.512	0.380	0.326	0.292	0.253	0.175	0.166
Mean std-SSIM	0.008	0.004	0.003	0.003	0.003	0.002	0.002	0.001

dataset. Note that we run the model 10 times to calculate the standard deviation of PSNR and SSIM (i.e., std-PSNR and std-SSIM) for each raw underwater image. The mean std-PSNR and std-SSIM of all test images are listed in Tab. 2. As can be observed, as the number of samples increases, the standard deviation of PSNR and SSIM first reduces and then becomes stable after 20 sampling times. This demonstrates that PUIE-Net can capture the diversity of enhanced images, and the consensus process can estimate a stable result based on multiple predictions. Note that increasing sampling times can further improve the stability of final predictions. In this paper, we set the default sampling times of MC estimation as 20 to better balance the stability and running time.

4.5   Discussion

Consensus Process The key idea of PUIE-Net is to learn the enhancement distribution and employ a consensus process to get the final prediction. Therefore, the effectiveness of the consensus process has a significant impact on the final results. In this paper, we have designed two consensus processes i.e., MC and MP. The former estimates an enhanced image by averaging a group of samples while the latter takes the image with the highest sampling probability as the final result. We consider that many other approaches can be applied to get the final result. For example, one can perform an image quality assessment method to select the best result according to the visual quality. In this case, both subjective and objective metrics can be used and the enhancement performance is

highly dependent on the used evaluation method. Quality based selection can be expressed as:

$$p\left(y\left|\mathbf{x}\right.\right) \approx \max_{s} Q\left(p\left(\mathbf{y}\left|\mathbf{z}^{(s)}, \mathbf{x}\right.\right), \mathbf{z}^{(s)} \sim p\left(\mathbf{z}\left|\mathbf{x}\right.\right)\right) \tag{13}$$

where Q denotes the subjective/objective image quality assessment function.

Relationship with Unsupervised Method The degradation of underwater images is non-reversible and complex, it is impracticable to record the ground truth for training a supervised UIE model directly. Therefore, we consider that unsupervised methods are more suitable for addressing the UIE problem. In this paper, instead of concentrating on developing an unsupervised method, we aim at making full use of the biased labels in existing UIE datasets. Biased reference maps of real-world underwater images are possible and cheap to obtain [36, 34]. With such reference maps, we proposed to leverage an AadIN module and a conditional variational autoencoder to learn the enhancement distribution. Therefore, the goal of both PUIE-Net and unsupervised UIE methods is to tackle the challenge of learning a UIE model without the ground truth. But our framework provides a new means to address the UIE problem i.e., resolving UIE into distribution estimation and consensus process based on biased reference maps, which is different from the unsupervised methods.

5   Conclusion

In this paper, we introduce PUIE-Net, a novel probabilistic network for underwater image enhancement. Specifically, we propose to learn the distribution of enhanced images rather than directly estimate a single result with certainty. This allows us to handle the label ambiguity issue of underwater image enhancement. The main idea is to leverage random style attributes from two posterior distributions constructed by conditional variational autoencoders, to transform the global enhancement statistics of input features. Moreover, with the consensus process, final enhancement results can be inferred via integrating a set of predictions. We demonstrate that PUIE-Net can effectively produce a set of reasonable results and the visual quality of the consensus estimation is highly competitive on two real-world UIE datasets. In the future, we plan to extend our method to other image enhancement tasks such as low-light image enhancement, dehazing, and denoising, where the ground truth image is difficult to obtain.

Acknowledgements

References

1. Abid, A., Zou, J.: Contrastive variational autoencoder enhances salient features. arXiv preprint arXiv:1902.04601 (2019)
2. Akkaynak, D., Treibitz, T.: A revised underwater image formation model. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6723–6732 (2018)
3. Akkaynak, D., Treibitz, T.: Sea-thru: A method for removing water from underwater images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1682–1691 (2019)
4. Aliakbarian, M.S., Saleh, F.S., Salzmann, M., Petersson, L., Gould, S., Habibian, A.: Learning variations in human motion via mix-and-match perturbation. arXiv preprint arXiv:1908.00733 (2019)
5. Ancuti, C.O., Ancuti, C., De Vleeschouwer, C., Bekaert, P.: Color balance and fusion for underwater image enhancement. IEEE Transactions on image processing **27**(1), 379–393 (2017)
6. Ancuti, C.O., Ancuti, C., De Vleeschouwer, C., Sbert, M.: Color channel compensation (3c): A fundamental pre-processing step for image enhancement. IEEE Transactions on Image Processing **29**, 2653–2665 (2019)
7. Ancuti, C., Ancuti, C.O., Haber, T., Bekaert, P.: Enhancing underwater images and videos by fusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 81–88 (2012)
8. Balakrishnan, G., Dalca, A.V., Zhao, A., Guttag, J.V., Durand, F., Freeman, W.T.: Visual deprojection: Probabilistic recovery of collapsed dimensions. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 171–180 (2019)
9. Baumgartner, C.F., Tezcan, K.C., Chaitanya, K., Hötker, A.M., Muehlematter, U.J., Schawkat, K., Becker, A.S., Donati, O., Konukoglu, E.: Phiseg: Capturing uncertainty in medical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 119–127 (2019)
10. Berman, D., Levy, D., Avidan, S., Treibitz, T.: Underwater single image color restoration using haze-lines and a new quantitative dataset. IEEE transactions on pattern analysis and machine intelligence **43**(8), 2822–2837 (2020)
11. Berman, D., Treibitz, T., Avidan, S.: Diving into haze-lines: Color restoration of underwater images. In: Proc. British Machine Vision Conference (BMVC). vol. 1 (2017)
12. Chiang, J.Y., Chen, Y.C.: Underwater image enhancement by wavelength compensation and dehazing. IEEE transactions on image processing **21**(4), 1756–1769 (2011)
13. Esser, P., Sutter, E., Ommer, B.: A variational u-net for conditional appearance and shape generation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8857–8866 (2018)
14. Fabbri, C., Islam, M.J., Sattar, J.: Enhancing underwater imagery using generative adversarial networks. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). pp. 7159–7165 (2018)
15. Fu, X., Cao, X.: Underwater image enhancement with global-local networks and compressed-histogram equalization. Signal Processing: Image Communication p. 115892 (2020)

16. Fu, X., Zhuang, P., Huang, Y., Liao, Y., Zhang, X.P., Ding, X.: A retinex-based enhancing approach for single underwater image. In: 2014 IEEE International Conference on Image Processing (ICIP). pp. 4572–4576. IEEE (2014)
17. Galdran, A., Pardo, D., Picón, A., Alvarez-Gila, A.: Automatic red-channel underwater image restoration. Journal of Visual Communication and Image Representation **26**, 132–145 (2015)
18. Gao, S.B., Zhang, M., Zhao, Q., Zhang, X.S., Li, Y.J.: Underwater image enhancement using adaptive retinal mechanisms. IEEE Transactions on Image Processing **28**(11), 5580–5595 (2019)
19. Ghani, A.S.A., Isa, N.A.M.: Underwater image quality enhancement through integrated color model with rayleigh distribution. Applied soft computing **27**, 219–230 (2015)
20. Guo, Y., Li, H., Zhuang, P.: Underwater image enhancement using a multiscale dense generative adversarial network. IEEE Journal of Oceanic Engineering **45**(3), 862–870 (2019)
21. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(12), 2341–2353 (2010)
22. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7132–7141 (2018)
23. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 1501–1510 (2017)
24. Huo, F., Li, B., Zhu, X.: Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 1944–1952 (2021)
25. Jamadandi, A., Mudenagudi, U.: Exemplar-based underwater image enhancement augmented by wavelet corrected transforms. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 11–17 (2019)
26. Jiang, N., Chen, W., Lin, Y., Zhao, T., Lin, C.W.: Underwater image enhancement with lightweight cascaded network. IEEE Transactions on Multimedia (2021)
27. Jiang, Q., Zhang, Y., Bao, F., Zhao, X., Zhang, C., Liu, P.: Two-step domain adaptation for underwater image enhancement. Pattern Recognition **122**, 108324 (2022)
28. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 694–711 (2016)
29. Kar, A., Dhara, S.K., Sen, D., Biswas, P.K.: Zero-shot single image restoration through controlled perturbation of koschmieder's model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16205–16215 (2021)
30. Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K., Eslami, S.A., Rezende, D.J., Ronneberger, O.: A probabilistic u-net for segmentation of ambiguous images. In: Advances in Neural Information Processing Systems (NeurIPS). pp. 6965–6975 (2018)
31. Kohl, S.A., Romera-Paredes, B., Maier-Hein, K.H., Rezende, D.J., Eslami, S., Kohli, P., Zisserman, A., Ronneberger, O.: A hierarchical probabilistic u-net for modeling multi-scale ambiguities. arXiv preprint arXiv:1905.13077 (2019)

32. Li, B., Sun, Z., Guo, Y.: Supervae: Superpixelwise variational autoencoder for salient object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). vol. 33, pp. 8569–8576 (2019)

33. Li, C.Y., Guo, J.C., Cong, R.M., Pang, Y.W., Wang, B.: Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. IEEE Transactions on Image Processing **25**(12), 5664–5677 (2016)

34. Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., Ren, W.: Underwater image enhancement via medium transmission-guided multi-color space embedding. IEEE Transactions on Image Processing **30**, 4985–5000 (2021)

35. Li, C., Anwar, S., Porikli, F.: Underwater scene prior inspired deep underwater image and video enhancement. Pattern Recognition **98**, 107038 (2020)

36. Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D.: An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing **29**, 4376–4389 (2019)

37. Li, C., Guo, J., Guo, C.: Emerging from water: Underwater image color correction based on weakly supervised color transfer. IEEE Signal processing letters **25**(3), 323–327 (2018)

38. Li, J., Skinner, K.A., Eustice, R.M., Johnson-Roberson, M.: Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. IEEE Robotics and Automation letters **3**(1), 387–394 (2017)

39. Liu, R., Fan, X., Zhu, M., Hou, M., Luo, Z.: Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. IEEE Transactions on Circuits and Systems for Video Technology **30**(12), 4861–4875 (2020)

40. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters **20**(3), 209–212 (2012)

41. Panetta, K., Kezebou, L., Oludare, V., Agaian, S.: Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with gan. IEEE Journal of Oceanic Engineering (2021)

42. Peng, Y.T., Cao, K., Cosman, P.C.: Generalization of the dark channel prior for single image restoration. IEEE Transactions on Image Processing **27**(6), 2856–2868 (2018)

43. Peng, Y.T., Cosman, P.C.: Underwater image restoration based on image blurriness and light absorption. IEEE transactions on image processing **26**(4), 1579–1594 (2017)

44. Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C., Muller, K.E.: Contrast-limited adaptive histogram equalization: speed and effectiveness. In: Proceedings of the First Conference on Visualization in Biomedical Computing. pp. 337–345 (1990)

45. Prakash, M., Krull, A., Jug, F.: Divnoising: Diversity denoising with fully convolutional variational autoencoders. arXiv preprint arXiv:2006.06072 (2020)

46. Qi, Q., Zhang, Y., Tian, F., Wu, Q.J., Li, K., Luan, X., Song, D.: Underwater image co-enhancement with correlation feature matching and joint learning. IEEE Transactions on Circuits and Systems for Video Technology (2021)

47. Rahman, Z.u., Jobson, D.J., Woodell, G.A.: Multi-scale retinex for color image enhancement. In: Proceedings of 3rd IEEE international conference on image processing. vol. 3, pp. 1003–1006. IEEE (1996)

48. Rezende, D.J., Mohamed, S., Wierstra, D.: Stochastic backpropagation and approximate inference in deep generative models. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 1278–1286 (2014)

49. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 234–241 (2015)
50. Sharma, G., Wu, W., Dalal, E.N.: The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. Color Research and Application **30**(1), 21–30 (2005)
51. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. Advances in Neural Information Processing Systems (NeurIPS) **28**, 3483–3491 (2015)
52. Tan, Q., Gao, L., Lai, Y.K., Xia, S.: Variational autoencoders for deforming 3d mesh models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5841–5850 (2018)
53. Uplavikar, P.M., Wu, Z., Wang, Z.: All-in-one underwater image enhancement using domain-adversarial learning. In: CVPR workshops. pp. 1–8 (2019)
54. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
55. Xue, X., Hao, Z., Ma, L., Wang, Y., Liu, R.: Joint luminance and chrominance learning for underwater image enhancement. IEEE Signal Processing Letters **28**, 818–822 (2021)
56. Y. C. Liu, W. H. Chan, Y. Q. Chen: Automatic white balance for digital still camera. IEEE Transactions on Consumer Electronics **41**(3), 460–466 (1995)
57. Yan, X., Rastogi, A., Villegas, R., Sunkavalli, K., Shechtman, E., Hadap, S., Yumer, E., Lee, H.: Mt-vae: Learning motion transformations to generate multimodal human dynamics. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 265–281 (2018)
58. Yang, H.H., Huang, K.C., Chen, W.T.: Laffnet: A lightweight adaptive feature fusion network for underwater image enhancement. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 685–692. IEEE (2021)
59. Yang, M., Hu, J., Li, C., Rohde, G., Du, Y., Hu, K.: An in-depth survey of underwater image enhancement and restoration. IEEE Access **7**, 123638–123657 (2019)
60. Zhang, J., Fan, D.P., Dai, Y., Anwar, S., Saleh, F.S., Zhang, T., Barnes, N.: Uc-net: Uncertainty inspired rgb-d saliency detection via conditional variational autoencoders. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8582–8591 (2020)