



A novel deep neural network for noise removal from underwater image

Qin Jiang ^{*}, Yang Chen, Guoyu Wang, Tingting Ji

College of Information Science and Engineering, Ocean University of China, Qingdao, 266100, China



ARTICLE INFO

Keywords:

Underwater image
Noise removal
Generative adversarial network
Self-attention
Spectral normalization

ABSTRACT

Underwater image processing technologies have always been challenging tasks due to the complex underwater environment. Images captured under water are not only affected by the water itself, but also by the diverse suspended particles that increase the effect of absorption and scattering. Moreover, these particles themselves are usually imaged on the picture, causing the spot noise signal to interfere with the target objects. To address this issue, we propose a novel deep neural network for removing the spot noise from underwater images. Its main idea is to train a generative adversarial network (GAN) to transform the noisy image to clean image. Based on the deep encoder and decoder framework, the skip connections are introduced to combine the features of low-level and high-level to help recover the original image. Meanwhile, the self-attention mechanism is employed to the generative network to capture global dependencies in the feature maps, which can generate the image with fine details at every location. Furthermore, we apply the spectral normalization to both the generative and discriminative networks to stabilize the training process. Experiments evaluated on synthetic and real-world images show that the proposed method outperforms many recent state-of-the-art methods in terms of quantitative and visual quality. Besides, the results also demonstrate that the proposed method has the good ability to remove the spot noise from underwater images while preserving sharp edge and fine details.

1. Introduction

Underwater vision technology is essential for ocean research and ocean engineering. Since the optical imaging in underwater environment is more complex than that in the atmosphere, the underwater image processing technologies have always been more challenging tasks. As light propagates through the water, absorption and scattering effects are due to not only the water itself but also the diverse suspended particles such as sand and plankton dissolved organic matters. Consequently, the underwater images generally suffer from low contrast, color distortion and noise. To tackle this problem, numerous methods of underwater image enhancement [1–9] and underwater image restoration [10–16] have been proposed in recent years. Although these methods have greatly improved the visibility and color correction of the underwater images, they have almost neglected the problem of the spot noise caused by suspended particles themselves, which interfere with the target objects when imaged in the picture. As shown in Fig. 1, the presence of the floating particles known as marine snow that wide variable in species and density [10], causing the underwater spot noise to be randomly distributed throughout the image. Moreover, they appear more visible after image dehazing or image enhancement. Therefore, it is necessary to remove the spot noise in the pre-processing stage, which is essential for the subsequent underwater image processing tasks such as segmentation, classification, target detection and recognition.

As we known, there are several researches on removing atmospheric particles such as for the rainy and snowy images have achieved good performance [18–23]. However, the suspended particles in water are neither snowflake nor gaussian noise. Although they look like snowflake in shape, they appear smaller and sporadic. Generally, these particles have the randomness of noise but do not consist with certain specific distribution characteristics. Then we call it spot noise, which is the visible interference on underwater images. It is no doubt that removing the spot noise is of great significance for underwater image pre-processing.

In this paper, we propose a novel deep neural network for removing the spot noise from underwater image. The proposed framework is designed as the end-to-end image translation model based on a generative adversarial network (GAN) [24]. To be specific, we adopt the deep encoder-decoder framework to generate denoised images, where the skip connections are introduced to combine the low-level features and high-level features of the input image. In addition, the self-attention modules that can capture global dependencies are embedded in the convolution layers to generate more detailed images. Furthermore, we apply the spectral normalization (SN) [25] to both the generative and discriminative networks, which improves the robustness of the networks effectively. Experimental results show that the proposed method has achieved outstanding performance both quantitatively and

* Corresponding author.

E-mail address: ritajq@126.com (Q. Jiang).

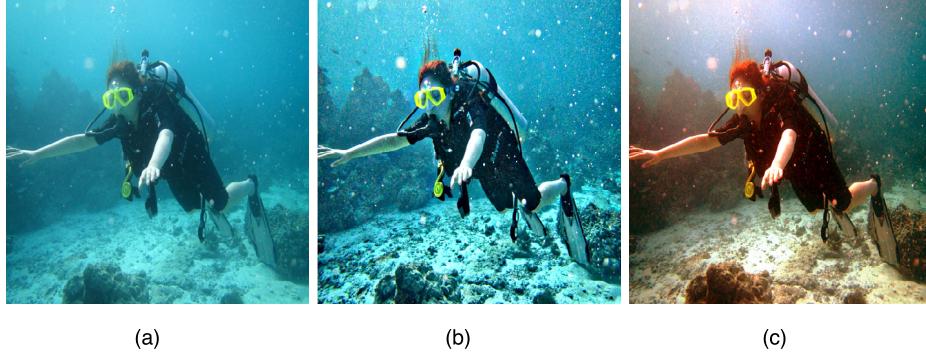


Fig. 1. The example of underwater image spot noise. (a) Raw underwater image. (b) After dehazed by He's method [17]. (c) After enhanced by Li's method [2].

qualitatively, especially has the remarkable ability to maintain the edge details while completely remove the noise from underwater images.

The rest of the paper is organized as follows. Some related work is introduced in Section 2. The framework of the proposed method is described in Section 3. Then, the experimental details and results are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. Related work

There have been many achievements in the field of image denoising in past decades. From the traditional classical method to the non-local method, especially the block matching 3D (BM3D) [26] method is popular used as the baseline. With the rapid development of deep learning methods, many of them have performed well in low-level computer vision tasks. Specifically, benefit from the advantages of large modeling capacity and tremendous advances in network training and design, the convolutional neural networks (CNNs) are widely used for image denoising task [27–30] and have achieved significant results. Jain and Seung [27] first used deep neural networks for image denoising and claimed that CNN provide comparable or even superior performance to wavelet and Markov Random Field (MRF) methods. Zhang et al. [28] proposed a deep convolutional neural network (DnCNN) for image denoising, where residual learning and batch normalization are utilized to speed up the training process as well as boost the denoising performance. J. Lehtinen et al. [31] proposed Noise2Noise method that enabled the training of CNNs from independent pairs of noisy images and achieved powerful performance in various image denoising. However, there are still some critical problems exist in the CNN-based methods because most of their objective function is the pixel wise loss function between the reconstructed and the ground truth images. Although such a method leads to high performance on PSNR metric, the reconstructed images always tend to be blurry or over-smoothing.

Recently, many works based on generative adversarial networks have shown great success in areas such as image super-resolution [32], image-to-image translation [33] and underwater image enhancement [34–36]. Meanwhile, several GAN-based methods have been explored for image denoising task. Yang et al. [37] proposed to use a generative adversarial network with the Wasserstein distance as the discrepancy measure between distributions and a perceptual loss for CT image denoising. Zhong et al. [38] proposed a Densely Connected Denoising Network (DenseNet) based on a generative adversarial network with stable Wasserstein-GAN loss function. Inspired by these advanced works, we have designed a novel deep neural network to remove the special spot noise from underwater image. Our network consists of two sub-network components: generative network and discriminative network. In terms of generator in our network, unlike existing methods that only obtain features by convolution, the attention modules are introduced into the convolutional layers. These modules help the network to capture details from every location of the image, breaking through the limitation of receptive field by convolution operation. In other

words, the attention module is a complement to the convolution layer, which combines local features with its global dependencies. Additionally, it is worth mentioning that we employ the spectral normalization to both generator and discriminator to stabilize the training process of the entire network.

3. Proposed method

Generally, one observed noisy underwater image (Y) can be decomposed into two separate images: the spot noise image (N) and the clean background image (X), represented as $Y = X + N$. In our method, instead of treating it as a decomposition problem, we aim to directly learn a mapping from noisy image to clean image by constructing a GAN-based deep network called Underwater image Denoising Network (UDnNet). The proposed network consists of two sub-networks: a generator network G and a discriminator network D . The generator network acts as a mapping function to translate an input noisy image to noise-free image such that it fools the discriminator network which is trained to discriminate the reconstructed image from the ground truth image. Through playing a minimax game between these two networks, the generator network has been trained to generate more and more realistic samples. The overall architecture of the proposed network is shown in Fig. 2. In addition, the optimization objective function can be formulated as:

$$\min_G \max_D \mathbb{E}_{y \sim p_{\text{clean}}} [\log D(y)] + \mathbb{E}_{z \sim p_{\text{noise}}} [\log(1 - D(G(z)))] \quad (1)$$

where G represents the generator network, and D represents the discriminator network. x is the input noisy image which drawn from our pool of images degraded by underwater noise, y is the ground truth image which drawn from a pool of clean images without underwater noise.

3.1. Network architecture

As the goal of underwater image denoising is to generate a noise-free image with high quality, the generator is the core of the whole framework, and it should be able to remove spot noise as much as possible without losing any details of the original image. As depicted in Fig. 2(a), the proposed generator network adopts a symmetric codec (encoder-decoder) structure, and we add skip connections following the general shape of the U-Net [39]. A lot of low-level feature information is shared between the noisy and real image, which can be transmitted directly to a higher level by using the skip connections. In this way, the feature maps passed by skip connections carry much image details, which helps deconvolution to recover a better clean image. To be specific, the generator first encodes the input image into high-dimensional representation using a stack of Convolution-InstanceNorm-LeakyReLU layers, and then, the output image can be decoded by the following Deconvolution-InstanceNorm-LeakyReLU layers. Note that the spectral normalization has been used in each convolutional layer.

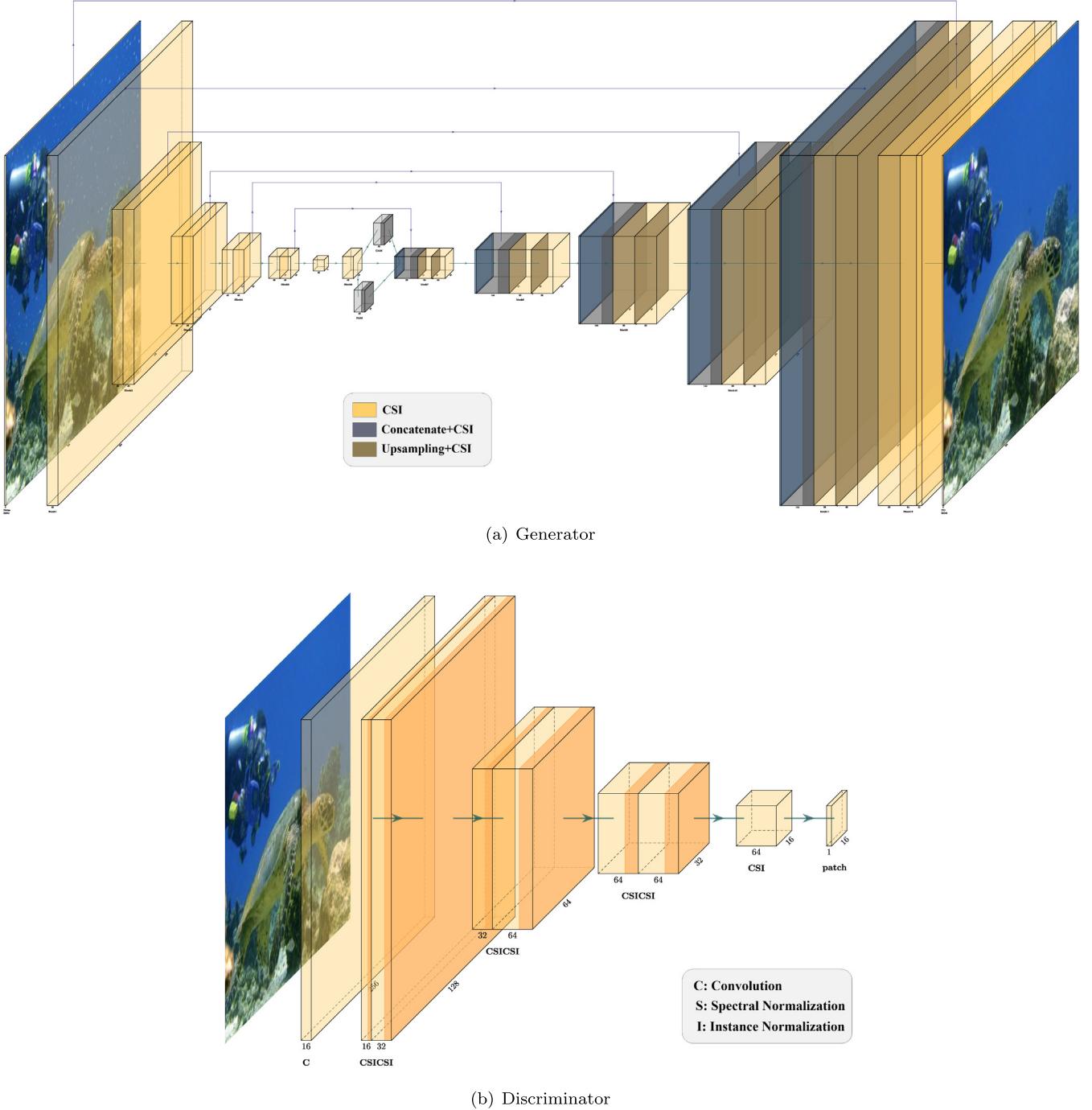


Fig. 2. The overview of the proposed method for underwater image denoising. Our network consists of two sub-networks: generator and discriminator.

In addition, we introduce the position attention module (PAM) and the channel attention module (CAM) into the bottom convolutional layer to strengthen feature representations effectively. Particularly, instead of deconvolution, we use upsampling and convolution in the decoder to avoid the checkboard artifacts in generated images.

Our discriminator (see Fig. 2(b)) is modeled as a PatchGAN [33], which tries to classify whether each patch in an image is real or fake instead of the whole image. This discriminator focuses on the high frequency details by restricting attention to the structure in local image patches. In addition, such a PatchGAN has fewer parameters, runs faster than a full-image discriminator. The proposed discriminator network consists of eight layers, where the first layer is a convolutional layer and the other layers are convolutions with spectral normalization

followed by instance normalization and LeakyReLU. Compared with the traditional discriminator that outputs a scalar value corresponding to real or fake, our patch discriminator outputs a $16 \times 16 \times 1$ matrix. We run this discriminator to convolve across the input image and average all responses to provide the final output of the discriminator.

3.1.1. Instance normalization

Adopting appropriate normalization may be a key factor in training deep network. Despite batch normalization (BN) [40] has been widely used to accelerate deep network training, it still has limitations in certain applications. Recent works [25,41] have pointed out that batch normalization will decrease the performance of image generation tasks. Instance normalization [42] has been proved to get better results than

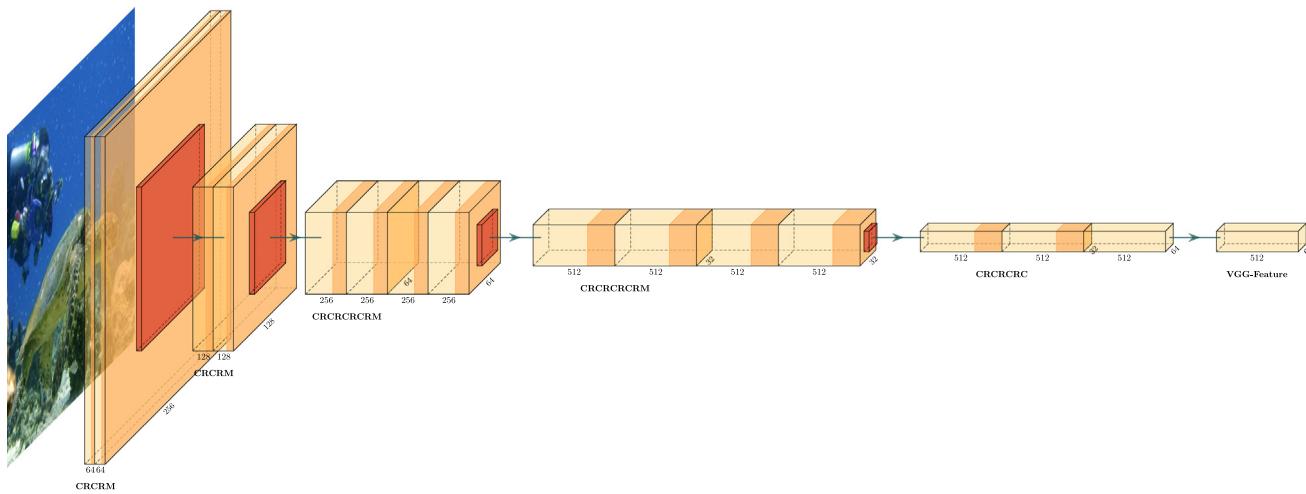


Fig. 3. Feature extractor. “C” denotes convolution layer, “R” denotes ReLU layer and “M” denotes Maxpooling.

batch normalization on the image style transform task. Consequently, we adopt instance normalization instead of batch normalization to improve the network performance. The key difference between instance normalization and batch normalization is that the former applies the normalization on a single image instead for the whole batch of images. Explained by formula, it defines $x \in \mathbb{R}^{T \times C \times W \times H}$ to be an input tensor containing a batch of T images. Let x_{tijk} denote its $tijk$ th element, where k and j are the spatial dimensions, i is the feature channel, and t is the index of the image in the batch. Then the formulation of the instance normalization is given by:

$$y_{tijk} = \frac{x_{tijk} - \mu_{ti}}{\sqrt{\sigma_{ti}^2 + \epsilon}}, \quad (2)$$

where $\mu_{ti} = \frac{1}{HW} \sum_{l=1}^L \sum_{m=1}^H x_{tilm}$ and $\sigma_{ti}^2 = \frac{1}{HW} \sum_{l=1}^L \sum_{m=1}^H (x_{tilm} - \mu_{ti})^2$. In practical, we replace batch normalization with instance normalization in every layer of the generator and discriminator.

3.1.2. Spectral normalization

Although GANs are known for their ability to generate a photorealistic image with high quality, it is difficult to stabilize training due to its own challenges such as vanishing gradients and mode collapse. In fact, our network training is unstable at first, resulting in artifacts and distortions in output images. To solve this problem, we apply the spectral normalization to both the generator and discriminator network. In this way, it restricts the Lipschitz constant of the generator and discriminator by restricting the spectral norm of each layer. To make the discriminator function satisfy 1-Lipschitz continuity, WGAN [43] and subsequent WGAN-GP [44] use weight-clipping and gradient penalty to constrain the discriminator parameters. The spectral normalization is another novel way to let the function satisfy 1-Lipschitz continuity. The 1-Lipschitz constraint can be satisfied by dividing the network parameters of each layer by the spectral norm of the parameter matrix of the layer. Lipschitz constant is the only hyperparameter to be tuned by setting the spectral norm of all weight layers to 1 consistently, which works well in practice. In our network, spectral normalization is injected into both discriminator and generator to stabilize the training of the whole network with small additional computational cost.

3.1.3. Dual attention module

Since the convolution operator has the limitation of local receptive field that prevent learning about long-term dependencies of an image, we introduce the self-attention modules into the convolutional layers. The self-attention mechanism is designed to compute the representation of each position by a weighted sum of the features at all positions, which can capture long-range relations for computer vision tasks. Many

works [45,46] have achieved good results by using the attention mechanism for image segmentation and object detection. Fu et al. [47] proposed a dual attention network that introduced a self-attention mechanism to capture features dependencies in the spatial and channel dimensions respectively. It is implemented by two types of attention modules, namely the position attention module (PAM) and the channel attention module (CAM). To be specific, the PAM is designed to enhance the expression of each feature by using the association between any two features, and the CAM is designed to enhance specific semantic responsiveness under the channel by modeling the associations between channels, then the output results of these two attention modules are fused to obtain better feature representations for pixel-level prediction. It has been proved that the dual attention modules can effectively capture long-range contextual information and provide more precise segmentation results. Therefore, we employ the dual attention modules to our generator network that can model the global dependencies between image regions to improve the details of the generated images.

3.2. Loss function

To ensure that the trained network can generate denoised pictures with good visual and quantitative scores, we proposed the loss function consisting of pixel-wise loss, perceptual loss and adversarial loss. The pixel-wise loss that minimize the pixel-level error between the reconstructed and ground truth image that can give generator some sense of ground truth, but it often produces blurring. Specifically, L1 norm is used for distance measure for it encourages less blurring. [21,33] Moreover, we adopt perceptual loss [48] that measured on the difference of high-level feature representation to enable the model to better reconstruct fine details and edges. Given an image pair $\{x, y\}$, where x is the input noisy image and y is the corresponding ground truth image, the pixel-wise loss is defined as:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y}[\|y - G(x)\|_1]. \quad (3)$$

The perceptual loss is defined as the distance between the feature representations of the reconstruct image and the corresponding ground truth image:

$$\mathcal{L}_p(G) = \mathbb{E}_{x,y}[\|\phi_i(y) - \phi_i(G(x))\|_1], \quad (4)$$

where ϕ_i is denoted as the feature map extracted by the i th convolution layer of the VGG19 network ϕ pretrained on the ImageNet dataset [49]. Specifically, we compute the perceptual loss at layer $\text{relu}5.4$ in VGG19 model [50]. In addition, the process of feature extraction is intuitively described in Fig. 3.



Fig. 4. A few samples of the synthetic dataset. The above row of images are original underwater images, the below row of images are the corresponding synthetic underwater noisy images.



Fig. 5. A few samples of real-world underwater noisy images.

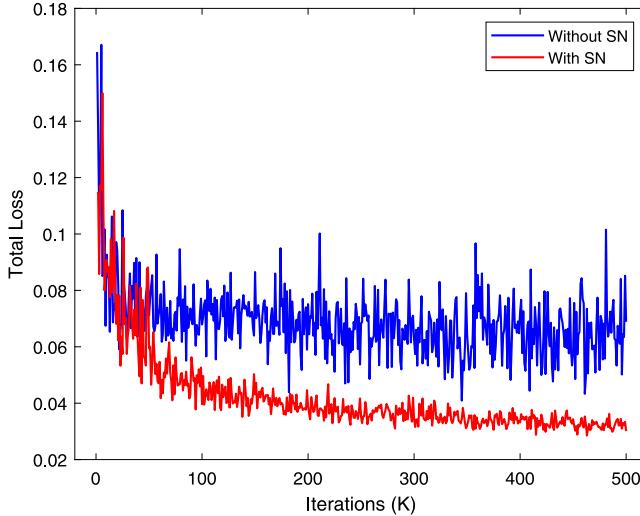


Fig. 6. Training convergence curve of the total loss.

The adversarial loss is defined as:

$$\mathcal{L}_A(G, D) = \mathbb{E}[\log D(y)] + \mathbb{E}[\log(1 - D(G(x)))] \quad (5)$$

In summary, the total loss function is the linear combination of the three losses with weights as follows:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{L1} + \lambda_2 \mathcal{L}_p + \lambda_3 \mathcal{L}_A, \quad (6)$$

where λ_1 , λ_2 and λ_3 are pre-defined weights for pixel-wise loss, perceptual loss and adversarial loss, respectively.

4. Experiments

In this section, we demonstrate the effectiveness of the proposed UDnNet method by experiments. We present the dataset and training details, and then perform an ablation study followed by comparison of the proposed method with the state-of-the-art methods on both synthetic and real-world noisy underwater images.

4.1. Dataset

Preparation of training data plays an important role in deep learning-based approaches. Due to the difficulty to obtain numerous

of clean/noisy underwater image pairs from real-world data, we synthesized a new dataset in our experiments. By selecting the underwater images without spot noise from the ImageNet dataset [49], we created our training set consisting of 5500 images and test set consisting of 1100 images. Then, we synthesized the noisy images by using Photoshop¹ to simulate a variety of suspended particles in water. Specifically, the way of adding spot noise is the same as adding snow by Photoshop. We have tried a lot of simulations to make it closer to the actual situation in water which is different from the snow in the atmosphere. Moreover, to generate the diverse dataset, we synthesize different types of spot noise for every 100 images to ensure that the simulated suspended particles with various density and orientations. All the training and test samples are resized to 256×256 . A few sample images from our dataset are shown in Fig. 4. Furthermore, in order to evaluate the effectiveness of the proposed method on real-world images, we also created a real-world test dataset contains 200 underwater noisy images downloaded from the Internet. A few real-world underwater images are shown in Fig. 5.

4.2. Training details

During the training process, we use Adam optimizer [51] with learning rate of 0.0002 and the momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$. We set the weight term in the loss function as $\lambda_1 = 1$, $\lambda_2 = 0.01$, and $\lambda_3 = 0.005$ in Eq. (6). All the parameters are set via cross-validation. A low value for λ_3 is used to ensure that the adversarial loss does not dominate the other losses. The entire network was trained on a Nvidia GTX 1080-TI GPU using the torch framework for 100 epochs. Fig. 6 illustrates the training convergence of the total loss over iterations. Compared to the network without spectral normalization, the network with spectral normalization owns a stable curve during training. It is proved that applying spectral normalization to our network indeed make the training process more stable.

4.3. Image quality assessment

To quantitatively evaluate the performance of different methods, we have carried out the full-reference evaluation and no-reference evaluation respectively. For the synthetic noisy image with ground truth, the two commonly used full-reference image metrics of PSNR and SSIM are employed to evaluate the noise removal quality regarding signal and structure similarities. Higher values of PSNR and SSIM indicate the result of denoised image is better. On the other hand, for the real-world

¹ <https://www.photoshopessentials.com/photo-effects/photoshop-weather-effects-snow/>.

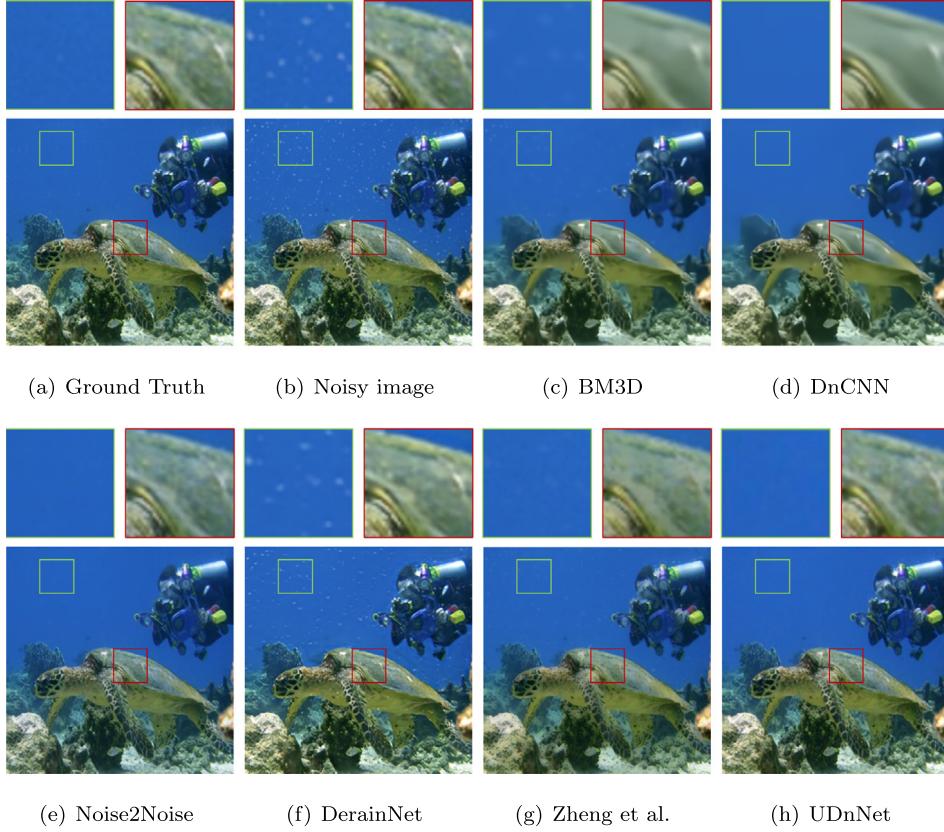


Fig. 7. The comparison of noise removal by using proposed method with other state-of-the-art methods on the synthetic noisy image. The values of PSNR and SSIM on denoised image obtained by these methods are also given followed. (a) Ground truth image. (b) Synthetic noisy image. (c) BM3D 31.82 dB/0.8992. (d) DnCNN 33.36 dB/0.9172. (e) Noise2Noise 32.50 dB/0.9658. (f) DerainNet 32.86 dB/0.9357. (g) Zheng et al. 32.53 dB/0.9285. (h) UDnNet 34.69 dB/0.9668.

Table 1
Evaluation of different components of the proposed method.

| Method | PSNR | SSIM |
|--------|--------------|---------------|
| -SC | 31.23 | 0.9175 |
| -DA | 31.46 | 0.9389 |
| -SN | 30.08 | 0.9211 |
| -IN | 31.57 | 0.9366 |
| UDnNet | 32.09 | 0.9423 |

dataset without ground truth, we adopt four no-reference image quality metrics for comprehensive evaluation. They are blind/referenceless image spatial quality evaluator (BRISQUE) [52], natural image quality evaluator (NIQE) [53], underwater color image quality evaluation (UCIQE) [54] and underwater image quality measure (UIQM) [55]. Both BRISQUE and NIQE are based on spatial feature extraction and measure the possible losses in the naturalness of an image. They are suitable for general blind image quality assessment problems and the lower scores indicate better image quality. The UCIQE combines three components of chroma, saturation, and contrast while UIQM comprises three underwater image attribute measure, such as colorfulness measure, sharpness measure and contrast measure. Higher scores of UCIQE and UIQM indicate better image quality.

4.4. Ablation study

There are several important components in the proposed method, namely the skip connection, dual attention modules, spectral normalization and instance normalization. To investigate the effect of each component on the denoising results, an ablation study has been performed involving the following experiments: (i) UDnNet without skip connection (UDnNet-SC), (ii) UDnNet without dual attention modules

Table 2
Comparison of the combination of different components.

| Method | PSNR | SSIM |
|-------------|--------------|---------------|
| SC+IN | 29.98 | 0.9206 |
| SC+IN+SN | 31.46 | 0.9389 |
| SC+IN+SN+DA | 32.09 | 0.9423 |

(UDnNet-DA), (iii) UDnNet without spectral normalization (UDnNet-SN) and (iv) UDnNet without instance normalization (UDnNet-IN). The quantitative evaluation is performed on the synthetic test set. The average results in terms of PSNR and SSIM are shown in Table 1, which illustrate the influences of different components in the proposed UDnNet. The best results of PSNR and SSIM metrics are marked in bold. Notably, each component can improve the performance of the proposed method for noise removal.

In addition, we also illustrate the different performance of three typical progressive network structures that combine different components, as shown in Table 2, which further demonstrates that our network has achieved the best results.

4.5. Comparisons with the state-of-the-art methods

To evaluate the performance of the proposed method, we compare it with the state-of-the art denoising methods including BM3D [20], DnCNN [22] and Noise2Noise [24], along with the atmospheric particle removal approaches DerainNet [13] and Zheng et al.'s method [15] that can remove both rain and snow. The quantitative results are evaluated on the 1100 images from test set by calculating the metrics of PSNR and SSIM. The average score of each method is shown in Table 3. As can be seen, the proposed UDnNet achieves the highest score, indicating that it outperforms the compared state-of-the-art methods.

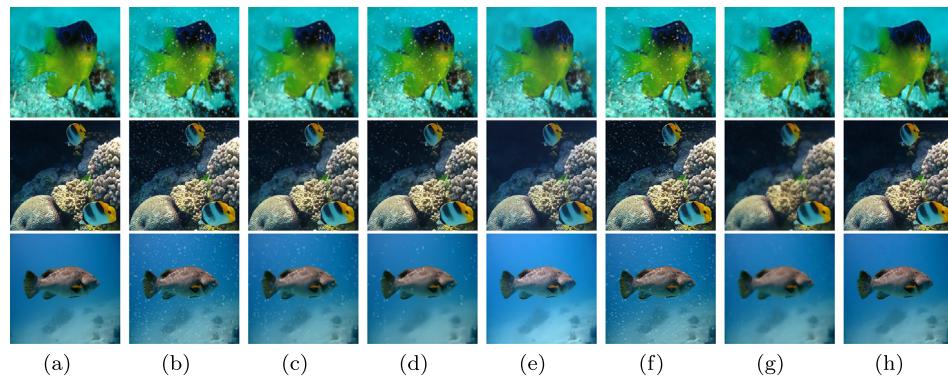


Fig. 8. Qualitative comparison of noise removal on sample images from test set. (a) Ground truth images. (b) Synthetic noisy images. (c) Results of BM3D. (d) Results of DnCNN. (e) Results of Noise2Noise. (f) Results of DerainNet. (g) Results of Zheng et al. (h) Our results.

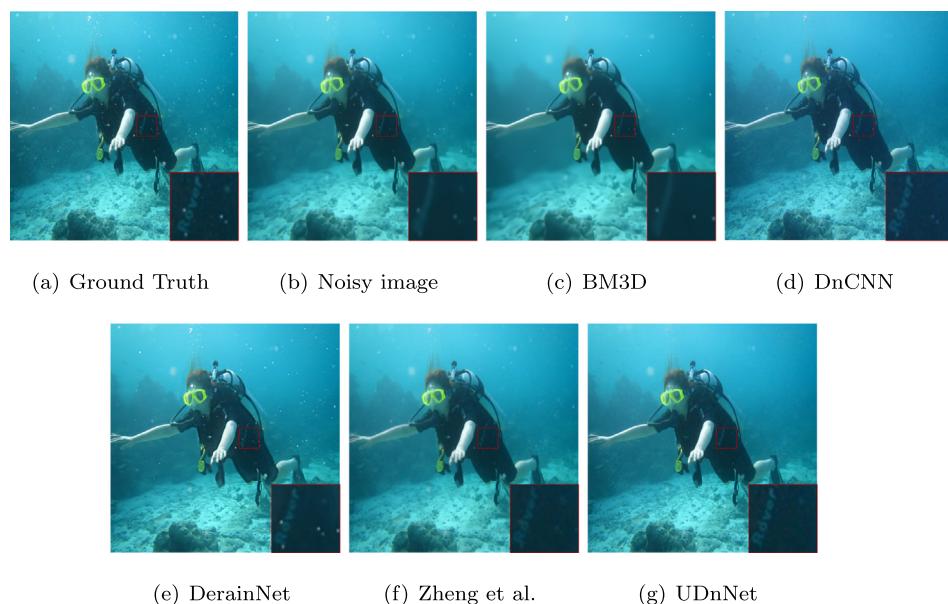


Fig. 9. Qualitative comparison of noise removal on a real-world underwater image of the “diver”.

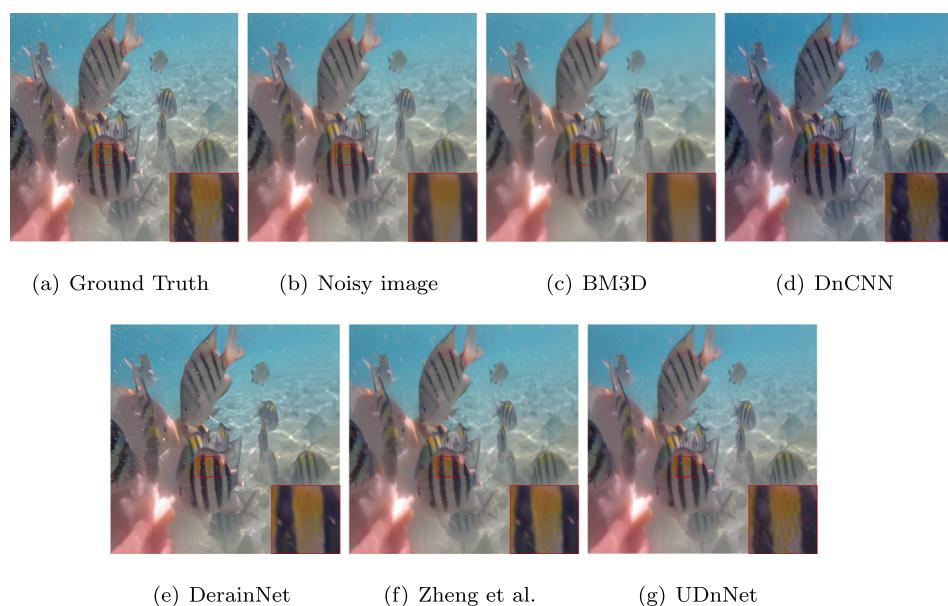


Fig. 10. Qualitative comparison of noise removal on a real-world underwater image of the “fish”.



Fig. 11. Several more results on real-world underwater noisy images. (a) Noisy images. (b) Results of BM3D. (c) Results of DnCNN. (d) Results of Noise2Noise. (e) Results of DerainNet. (f) Results of Zheng et al. (g) Our results.

Furthermore, the visual comparisons on a synthetic image from test set are demonstrated in Fig. 7. For better visual comparison, we show zoomed versions of the specific regions above the corresponding images. By looking at these enlarged regions, we can clearly observe the effects of noise removal and image details. Among these methods, BM3D still leaves a little noise and smooths the denoised image to some extent. While DnCNN removes the noise cleanly, it also over-smooths the denoised image. DerainNet preserves the details well, but hardly removes the noise. Noise2Noise is effective at removing noise, however, it introduces color distortion in the denoised images. Zheng et al.'s method effectively removes noise, whereas its corresponding results suffer from significant blurry artifacts and fail to preserve image details. UDnNet removes the noise completely while preserving sharp edges and fine details. Obviously, the output denoised by our method achieves the best visual effect that close to the ground truth image. Additional comparisons are provided in Fig. 8. It is demonstrated that the proposed method successfully removes noise and yields better results than the other state-of-the-art methods.

4.6. Comparisons with real-world images

To demonstrate the practical significance of the proposed method, we further evaluated the performance on real-world underwater noisy images. Figs. 9–11 show the qualitative comparisons on real-world underwater noisy images. As can be seen, the images denoised by BM3D and DnCNN still leave a little noise and the details are over-smoothed. DerainNet removes almost nothing from images with no effect on spot noise. While Zheng et al.'s method removes most of the noise, it tends to produce blurred results and artifacts which are not visually appealing. Noise2Noise and our method obtain impressive

results, whereas Noise2Noise produces color distortions that can be seen from the background. In comparison, our method can remove the noise completely while preserving the sharp edges and fine details, especially visible when the observed region is magnified. To sum up, the results generated by our method show the best visual performance consistently.

Furthermore, we evaluate the quantitative comparison of different methods on the 200 real-world underwater noisy images. The no-reference metrics of BRISQUE, NIQE, UCIQE and UIQM are used to measure the visual quality of denoised image. The average scores are shown in Table 4 and the best results of all metrics are marked in bold. It is observed that our method has the lowest scores of BRISQUE and NIQE, indicating that our results obtain the better image quality. In terms of UCIQE, there is little difference among all the methods, however, the score of Noise2Noise slightly higher due to its color distortion which can be seen from the visual comparisons. On the other hand, the UIQM score of our method is higher than the other methods, which indicates that the result is more consistent with the human visual perception. In summary, these quantitative evaluations provide additional evidence that our method outputs an image with greater improvement. It can be found that although UDnNet is trained on synthetic data, the learned network is also very effective for real-world images. Additionally, we exhibit the comparative effect of image enhancement before and after denoising. As shown in Fig. 12, it is obvious that the denoised images are free from the interfering spot noise, which is of great significance for the underwater image pre-processing and is helpful for subsequent processing.

5. Conclusion

We have presented a novel deep neural network called UDnNet for removing spot noise from underwater images. Based on the architecture



Fig. 12. The comparison of underwater image enhancement effect before and after denoising. The above row of images are the original underwater images enhancement, the below row of images are the corresponding denoised underwater images enhancement. All images are enhanced by Li's method [2].

Table 3

Average PSNR (dB)/SSIM results of different methods on synthetic test images.

| Method | PSNR | SSIM |
|--------------|--------------|---------------|
| BM3D | 29.67 | 0.8912 |
| DnCNN | 30.58 | 0.9025 |
| Noise2Noise | 30.22 | 0.9386 |
| DerainNet | 31.06 | 0.9207 |
| Zheng et al. | 30.72 | 0.9139 |
| UDnNet | 32.09 | 0.9423 |

Table 4

Quantitative measurement results evaluated on real-world images.

| Method | BRISQUE | NIQE | UCIQE | UIQM |
|--------------|--------------|-------------|---------------|---------------|
| BM3D | 40.77 | 5.35 | 0.4103 | 1.3665 |
| DnCNN | 46.56 | 5.70 | 0.4086 | 1.1722 |
| Noise2Noise | 34.12 | 4.07 | 0.4161 | 1.6853 |
| DerainNet | 29.28 | 4.58 | 0.4058 | 1.9706 |
| Zheng et al. | 36.43 | 4.86 | 0.4095 | 1.9195 |
| UDnNet | 27.72 | 3.39 | 0.4112 | 2.0056 |

of generative and adversarial network, we employ the skip connections and self-attention mechanism to generate images with fine details. Notably, the application of spectral normalization to both generator and discriminator has stabilize the training process. Moreover, an ablation study is carried out to illustrate improvements obtained owe to each component in the proposed method. Extensive experiments evaluated on both synthetic and real-world underwater images demonstrate that the proposed method is superior to the recent state-of-the-art image denoising methods and atmospheric particle removal approaches in terms of quantitative and visual quality. Furthermore, the proposed framework can be applied to other paired image translation tasks to achieve the outstanding effect of preserving edges and details.

CRediT authorship contribution statement

Qin Jiang: Conceptualization, Methodology, Software, Investigation, Writing - original draft, Writing - review & editing. **Yang Chen:** Software, Investigation. **Guoyu Wang:** Resources, Funding acquisition. **Tingting Ji:** Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61571407). Many thanks to the anonymous reviewers for their constructive comments and valuable suggestions.

References

- [1] C. Ancuti, C.O. Ancuti, T. Haber, P. Bekaert, Enhancing underwater images and videos by fusion, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 81–88, <http://dx.doi.org/10.1109/CVPR.2012.6247661>.
- [2] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, B. Wang, Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior, IEEE Trans. Image Process. 25 (12) (2016) 5664–5677, <http://dx.doi.org/10.1109/TIP.2016.2612882>.
- [3] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P.S. Zhang, X. Ding, A retinex-based enhancing approach for single underwater image, in: 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 4572–4576, <http://dx.doi.org/10.1109/ICIP.2014.7025927>.
- [4] C. Li, J. Guo, C. Guo, R. Cong, J. Gong, A hybrid method for underwater image correction, Pattern Recognit. Lett. 94 (2017) 62–67, <http://dx.doi.org/10.1016/j.patrec.2017.05.023>.
- [5] S. Anwar, C. Li, Diving deeper into underwater image enhancement: A survey, 2019, arXiv preprint <arXiv:1907.07863>.
- [6] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Trans. Image Process. 29 (2019) 4376–4389, <http://dx.doi.org/10.1109/TIP.2019.2955241>.
- [7] W. Zhang, L. Dong, X. Pan, P. Zou, L. Qin, W. hai Xu, A survey of restoration and enhancement for underwater images, IEEE Access 7 (2019) 182259–182279, <http://dx.doi.org/10.1109/ACCESS.2019.2959560>.
- [8] C. Li, S. Anwar, F.M. Porikli, Underwater scene prior inspired deep underwater image and video enhancement, Pattern Recognit. 98 (2020) <http://dx.doi.org/10.1016/j.patcog.2019.107038>.
- [9] R. Liu, X. Fan, M. Zhu, M. Hou, Z. Luo, Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light, IEEE Trans. Circuits Syst. Video Technol. (2020) <http://dx.doi.org/10.1109/TCSVT.2019.2963772>.
- [10] R. Schettini, S. Corchs, Underwater image processing: State of the art of restoration and image enhancement methods, EURASIP J. Adv. Signal Process. 2010 (2010) 1–14, <http://dx.doi.org/10.1155/2010/746052>.
- [11] G. Wang, B. Zheng, F.F. Sun, Estimation-based approach for underwater image restoration, Opt. Lett. 36 (13) (2011) 2384–2386, <http://dx.doi.org/10.1364/OL.36.002384>.
- [12] A. Galdran, D. Pardo, A. Picón, A. Alvarez-Gila, Automatic red-channel underwater image restoration, J. Vis. Commun. Image Represent. 26 (2015) 132–145, <http://dx.doi.org/10.1016/j.jvcir.2014.11.006>.
- [13] H. Liu, L.-P. Chau, Underwater image restoration based on contrast enhancement, in: 2016 IEEE International Conference on Digital Signal Processing (DSP), 2016, pp. 584–588, <http://dx.doi.org/10.1109/ICDSP.2016.7868625>.
- [14] C. Li, J. Quo, Y. Pang, S. Chen, J.J. Wang, Single underwater image restoration by blue-green channels dehazing and red channel correction, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 1731–1735, <http://dx.doi.org/10.1109/ICASSP.2016.7471973>.
- [15] H. Liu, L.-P. Chau, Deepset video descattering, Multimedia Tools Appl. 78 (20) (2019) 28919–28929, <http://dx.doi.org/10.1007/s11042-017-5474-3>.
- [16] Y. Wang, H. Liu, L.-P. Chau, Single underwater image restoration using adaptive attenuation-curve prior, IEEE Trans. Circuits Syst. I. Regul. Pap. 65 (2018) 992–1002, <http://dx.doi.org/10.1109/TCSI.2017.2751671>.
- [17] K. He, S. Jian, Fellow, IEEE, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2011) 2341–2353, <http://dx.doi.org/10.1109/TPAMI.2010.168>.
- [18] J. Chen, C.-H. Tan, J. Hou, L.-P. Chau, H. Li, Robust video content alignment and compensation for rain removal in a cnn framework, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6286–6295, <http://dx.doi.org/10.1109/CVPR.2018.00658>.

- [19] H. Zhang, V. Sindagi, V.M. Patel, Image de-raining using a conditional generative adversarial network, *IEEE Trans. Circuits Syst. Video Technol.* (2019) <http://dx.doi.org/10.1109/TCSVT.2019.2920407>.
- [20] X. Fu, J. Huang, X. Ding, Y. Liao, J. Paisley, Clearing the skies: a deep network architecture for single-image rain removal, *IEEE Trans. Image Process.* 26 (6) (2017) <http://dx.doi.org/10.1109/TIP.2017.2691802>.
- [21] P. Xiang, L. Wang, F. Wu, J. Cheng, M. Zhou, Single-image de-raining with feature-supervised generative adversarial network, *IEEE Signal Process. Lett.* 26 (2019) 650–654, <http://dx.doi.org/10.1109/LSP.2019.2903874>.
- [22] X. Zheng, Y. Liao, W. Guo, X. Fu, X. Ding, Single-image-based rain and snow removal using multi-guided filter, in: International Conference on Neural Information Processing, 2013, pp. 258–265, http://dx.doi.org/10.1007/978-3-642-42051-1_33.
- [23] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, J.-N. Hwang, Desnownet: Context-aware deep network for snow removal, *IEEE Trans. Image Process.* 27 (2018) 3064–3073, <http://dx.doi.org/10.1109/TIP.2018.2806202>.
- [24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.
- [25] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, 2018, arXiv preprint [arXiv:1802.05957](http://arxiv.org/abs/1802.05957).
- [26] K. Dabov, A. Foi, V. Katkovnik, K.O. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, *IEEE Trans. Image Process.* 16 (2007) 2080–2095, <http://dx.doi.org/10.1109/TIP.2007.901238>.
- [27] V. Jain, H.S. Seung, Natural image denoising with convolutional networks, in: *NIPS*, 2008.
- [28] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising, *IEEE Trans. Image Process.* 26 (2017) 3142–3155, <http://dx.doi.org/10.1109/TIP.2017.2662206>.
- [29] S. Guo, Z. Yan, K. Zhang, W. Zuo, L. Zhang, Toward convolutional blind denoising of real photographs, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 1712–1722, <http://dx.doi.org/10.1109/CVPR.2019.00181>.
- [30] J. Chen, J. Hou, L.-P. Chau, Light field denoising via anisotropic parallax analysis in a cnn framework, *IEEE Signal Process. Lett.* 25 (9) (2018) 1403–1407, <http://dx.doi.org/10.1109/LSP.2018.2861212>.
- [31] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, T. Aila, Noise2noise: Learning image restoration without clean data, 2018, arXiv preprint [arXiv:1803.04189](http://arxiv.org/abs/1803.04189).
- [32] C. Ledig, L. Theis, F. Huszár, J.A. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 105–114, <http://dx.doi.org/10.1109/CVPR.2017.19>.
- [33] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 5967–5976, <http://dx.doi.org/10.1109/CVPR.2017.632>.
- [34] C. Fabbri, M.J. Islam, J. Sattar, Enhancing underwater imagery using generative adversarial networks, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 7159–7165, <http://dx.doi.org/10.1109/ICRA.2018.8460552>.
- [35] C. Li, J. Guo, C. Guo, Emerging from water: Underwater image color correction based on weakly supervised color transfer, *IEEE Signal Process. Lett.* 25 (2018) 323–327, <http://dx.doi.org/10.1109/LSP.2018.2792050>.
- [36] Y. Guo, H. Li, P. Zhuang, Underwater image enhancement using a multiscale dense generative adversarial network, *IEEE J. Ocean. Eng.* (2019) <http://dx.doi.org/10.1109/JOE.2019.2911447>.
- [37] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M.K. Kalra, Y. Zhang, L. Sun, G. Wang, Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss, *IEEE Trans. Med. Imaging* 37 (2018) 1348–1357, <http://dx.doi.org/10.1109/TMI.2018.2827462>.
- [38] Y. Zhong, L. Liu, D. Zhao, H. Li, A generative adversarial network for image denoising, *Multimedia Tools Appl.* (2019) 1–13, <http://dx.doi.org/10.1007/s11042-019-7556-x>.
- [39] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, 2015, http://dx.doi.org/10.1007/978-3-319-24574-4_28, ArXiv, abs/1505.04597.
- [40] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015, arXiv preprint [arXiv:1502.03167](http://arxiv.org/abs/1502.03167).
- [41] S. Nah, T.H. Kim, K.M. Lee, Deep multi-scale convolutional neural network for dynamic scene deblurring, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 257–265, <http://dx.doi.org/10.1109/CVPR.2017.35>.
- [42] D. Ulyanov, A. Vedaldi, V. Lempitsky, Instance normalization: The missing ingredient for fast stylization, 2016, arXiv preprint [arXiv:1607.08022](http://arxiv.org/abs/1607.08022).
- [43] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan, 2017, arXiv preprint [arXiv:1701.07875](http://arxiv.org/abs/1701.07875).
- [44] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of wasserstein gans, in: *Advances in Neural Information Processing Systems*, 2017, pp. 5767–5777.
- [45] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, Self-attention generative adversarial networks, 2018, arXiv preprint [arXiv:1805.08318](http://arxiv.org/abs/1805.08318).
- [46] C. Li, R. Cong, S. Kwong, J. Hou, H. Fu, G. Zhu, D. Zhang, Q. Huang, ASIP-Net: Attention steered interweave fusion network for RGB-d salient object detection, *IEEE Trans. Cybern.* (2020) <http://dx.doi.org/10.1109/TCYB.2020.2969255>.
- [47] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154, <http://dx.doi.org/10.1109/CVPR.2019.00326>.
- [48] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: European Conference on Computer Vision, 2016, pp. 694–711, http://dx.doi.org/10.1007/978-3-319-46475-6_43.
- [49] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, F.-F. Li, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255, <http://dx.doi.org/10.1109/CVPRW.2009.5206848>.
- [50] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint [arXiv:1409.1556](http://arxiv.org/abs/1409.1556).
- [51] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint [arXiv:1412.6980](http://arxiv.org/abs/1412.6980).
- [52] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, *IEEE Trans. Image Process.* 21 (12) (2012) 4695–4708, <http://dx.doi.org/10.1109/TIP.2012.2214050>.
- [53] A. Mittal, Fellow, IEEE, R. Soundararajan, A.C. Bovik, Making a 'completely blind' image quality analyzer, *IEEE Signal Process. Lett.* 20 (3) (2013) 209–212, <http://dx.doi.org/10.1109/LSP.2012.2227726>.
- [54] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, *IEEE Trans. Image Process.* 24 (12) (2015) 6062–6071, <http://dx.doi.org/10.1109/TIP.2015.2491020>.
- [55] K. Panetta, C. Gao, S. Agaian, Human-visual-system-inspired underwater image quality measures, *IEEE J. Ocean. Eng.* 41 (3) (2016) <http://dx.doi.org/10.1109/JOE.2015.2469915>.