Unsupervised learning polarimetric underwater image recovery under nonuniform optical fields

Shilong Yang, Bowen Qu, Genshen Liu, Dingxuan Deng, Shiyuan Liu, and Xiugo Chen

State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
*Corresponding author: xiugochen@hust.edu.cn

Received 1 June 2021; revised 12 August 2021; accepted 14 August 2021; posted 17 August 2021 (Doc. ID 432994); published 10 September 2021

1. INTRODUCTION

Underwater imaging technology is widely used [1], such as in underwater equipment detection, marine resources exploration, and other aspects. However, turbid media will lead to a sharp decline in image quality [2]. In view of this problem, many studies have been proposed to enhance underwater image quality by using polarization methods. In 1995, Rowe and Pugh proposed a polarized difference algorithm [3], which introduced polarization imaging into underwater image recovery and achieved good results, but the processing method lacked physical significance. In 2005, Schechner and Karpel presented an approach for recovering clear underwater images [4], which established a simplified imaging model combined with polarization information to obtain clear images. However, some simplifications have been made in their model, such as treating the value of backscatter at an infinite distance as a constant and ignoring the polarization effect of the object, which will lead to low quality of the recovered images and even failure in nonuniform optical fields. Many studies have been made to improve the imaging model [5-7]. For example, in 2014, Liang et al. proposed to use the angle of polarization (AOP) to estimate the polarization parameters [5]. In 2018, Hu et al. proposed to use polynomial fitting for underwater image recovery [6]. In 2020, Hu et al. further proposed a polarimetric dense network based on deep learning to recover underwater images [8]. They trained the network by a large number of labeled samples and finally obtained the mapping of the polarized images to the clear recovered images.

In this paper, we focus on the combination of unsupervised learning and an improved underwater image recovery method. The value of backscatter at an infinite distance and the degree of polarization (DOP) in the nonuniform optical field are regarded as parameters changing with the spatial position. We analyze the improper simplification in nonuniform optical fields and propose the use of unsupervised learning to estimate the value of the backscatter. Compared with the method of using polynomial fitting to recover the image [6], our method is much faster in principle and has a better real-time performance. Compared with the polarimetric dense network based on supervised learning [8], our method has much less cost in preparing the training dataset. The experiments performed under different conditions have demonstrated the capability of the proposed method in recovering underwater images under different nonuniform optical fields.

2. UNDERWATER IMAGING MODEL

Our method is based on the underwater imaging model proposed in [4]. In this model, the light collected by the camera can be divided into two different parts: one is the direct transmission, and the other is the veiling light or backscatter. This can be described as follows:

\[ I(x, y) = D(x, y) + B(x, y), \] (1)
where \( I_{(x,y)} \) is the intensity of light reaching the camera, \( D_{(x,y)} \) is the light that propagates from the object and finally reaches the camera after attenuation caused by absorption and scattering in the water, and \( B_{(x,y)} \) refers to the backscatter. All of them are functions of the pixel coordinates \( x \) and \( y \). What’s more, \( D_{(x,y)} \) is given by

\[
D_{(x,y)} = L_{(x,y)} \cdot t_{(x,y)} ,
\]

where \( L_{(x,y)} \) is the ideal object radiance sensed by the camera in the absence of absorption and scattering, and \( t_{(x,y)} \) is the medium transmittance. And \( L_{(x,y)} \) is what we need to work out in order to recover the image.

The backscatter light \( B_{(x,y)} \) is given by

\[
B_{(x,y)} = A_{\infty(x,y)} \cdot (1 - t_{(x,y)}),
\]

where \( A_{\infty(x,y)} \) is the intensity of background light extending to the infinity in the water, which significantly varies with spatial location under nonuniform optical fields.

From Eqs. (1)–(3), we can obtain

\[
L_{(x,y)} = \frac{I_{(x,y)} - B_{(x,y)}}{t_{(x,y)}} = \frac{I_{(x,y)} - B_{(x,y)}}{1 - \frac{B_{(x,y)}}{A_{\infty(x,y)}}},
\]

where \( I_{(x,y)} \) is already known. Therefore, the recovery of \( L_{(x,y)} \) exclusively depends on the estimation of \( A_{\infty(x,y)} \) and \( B_{(x,y)} \). In the following section, we will discuss in detail how to utilize unsupervised learning to estimate \( A_{\infty(x,y)} \). After that, we introduce an improved method to calculate the value of \( B_{(x,y)} \). With the estimated \( A_{\infty(x,y)} \) and \( B_{(x,y)} \), we can finally acquire the recovered image \( L_{(x,y)} \).

3. UNDERWATER IMAGE RECOVERING PROCESS

The process of the proposed underwater image recovery method is shown in Fig. 1, which consists of eight steps.

Step 1: Use the polarization camera to collect polarization images at four orientations (i.e., \( I_{0(x,y)} \), \( I_{45(x,y)} \), \( I_{90(x,y)} \), and \( I_{135(x,y)} \)).

Step 2: Remove the region of the target object from the polarization images to get the incomplete background light (i.e., incomplete \( A_{\infty} \)).

Step 3: Use the unsupervised learning inpainting network as illustrated in Section 3.A to estimate the complete background light (complete \( A_{\infty} \)).

Step 4: Calculate the Stokes parameters of \( A_{\infty} \) according to Eq. (5) and filter them by the method mentioned in Section 3.A.

Step 5: Calculate the AOP of \( A_{\infty} \) according to Eq. (7), and then estimate the DOP of \( B_{(x,y)} \) from it according to Eq. (8).

Step 6: Calculate the DOP of \( I_{(x,y)} \) to obtain the DOP of \( B_{(x,y)} \), and then calculate the intensity of the polarized part of \( I_{(x,y)} \) according to Eq. (9).

Step 7: Calculate \( B_{(x,y)} \) according to Eq. (6).

Step 8: Obtain clear \( L_{(x,y)} \) by Eq. (10).

A. Estimation of \( A_{\infty(x,y)} \) based on Unsupervised Learning Inpainting Network

In 2016, Pathak et al. proposed a network to work on inpainting images via feature learning called context encoders (CE) [9], which employs a convolutional neural network (CNN) [10,11] and is trained with an adversarial loss [12].

Inspired by CE, we propose an inpainting network with unsupervised learning that can provide three additional aspects of optimization. First, we can handle arbitrary inpainting masks, which enable us to deal with incomplete \( A_{\infty(x,y)} \) in many situations. Second, we can improve the fitting degree on the edge. And third, we can obtain better visual quality overall. The network architecture, as shown in Fig. 2, is designed to realize the first and third targets. As for the second target, we use an edge weighted loss function as described in Appendix A to achieve it. For brevity, we only focus on the architecture of the inpainting network in this section. The detailed setup of the network is presented in Appendix A.
From another viewpoint, the image inpainting task can be regarded as an image generation task that implements the missing region based on the image distribution of the surrounding area. However, when the missing region is difficult to determine, the output size cannot be determined either. This problem can be solved by designing an end-to-end generative network with the same output size as the input as illustrated in the upper part of Fig. 2. We have adopted the widely used encoder-decoder architecture to form our generative network.

In the encoder part, we first stack five feature extraction blocks to extract multi-level features (each block is composed of a “same” convolution with a kernel size of $3 \times 3$, a rectified linear unit (ReLU), and a max-pooling layer). After five feature extraction blocks, a feature map with a size of $4 \times 4 \times 512$ is obtained, and then a “valid” convolution with a kernel size of $4 \times 4$ is used and activated with a ReLU to obtain a $1 \times 1 \times 4096$ long feature vector integrated global information.

For this long feature vector, we use $1 \times 1$ convolution and ReLU activation function (called “channel-wise fully connected layer”) for feature selection, calculation, and reorganization to obtain a long feature vector that is also $1 \times 1 \times 4096$. This long feature vector is considered to be the final encoding of the input image, which is then used for decoding.

In the decoder part, the decoder is symmetrical to the encoder, except that each downsampling block composed of the convolution layer, ReLU, and max-pooling layer becomes an upsampling block composed of a transposed convolution layer and ReLU. In this way, an end-to-end mapping from an input image to an output image of the same size is realized.

In order to optimize the implementation of images repaired by the generative network as illustrated in the lower part of Fig. 2, we also designed a discriminative network to determine whether the input image comes from the training dataset or the output of the generative network. (From the dataset, it will be judged as real; otherwise it will be judged as false). When training the generative network, a cheating discriminative network is part of its training target, so the output image of the generative network is confused with the image in the dataset. After that, it is difficult to distinguish the difference between them, so the output image of the generative network has a higher visual quality.

The discriminative network has the same structure as the encoder of the generative network. The difference is that after obtaining a $4 \times 4$ feature map, we use a $4 \times 4$ valid convolution layer with an output channel of 1 and the ReLU activation function to get a scalar. The scalar is used to represent the probability that the image input to the discriminative network is judged to be real.

Through the unsupervised learning inpainting network, we get four complete polarimetric images of background light in four different orientations. By definition, the background light should be smooth. So we introduce a filtering method in which we compare the gray value of one pixel with those in other pixels nearby. Here all the compared pixels are in a $7 \times 7$ square. Next, we compute the averaged gray value of the nearby pixels whose difference between the gray value of the central pixel are less than 6. Finally, we replace the gray value of the central pixel with this averaged value.

B. Improved Model to Compute $B_{(x,y)}$ and Get Optimal $L_{(x,y)}$

We now have four polarimetric images of the object ($I_0(x,y)$, $I_45(x,y)$, $I_90(x,y)$, $I_135(x,y)$) and four complete polarimetric images of background light ($A_∞0(x,y)$, $A_∞45(x,y)$, $A_∞90(x,y)$, $A_∞135(x,y)$) in four different orientations. Based on these images we can obtain the Stokes vectors of $I(x,y)$ and $A_∞(x,y)$ by

$$S_{0I(x,y)} = \frac{1}{2} \left( I_0(x,y) + I_45(x,y) + I_90(x,y) + I_135(x,y) \right) ,$$

$$S_{1I(x,y)} = I_0(x,y) - I_90(x,y) ,$$

$$S_{2I(x,y)} = I_45(x,y) - I_135(x,y) ,$$

$$S_{3I(x,y)} = \frac{1}{2} \left( A_∞0(x,y) + A_∞45(x,y) + A_∞90(x,y) + A_∞135(x,y) \right) ,$$

$$S_{1A(x,y)} = A_∞0(x,y) - A_∞90(x,y) ,$$

$$S_{2A(x,y)} = A_∞45(x,y) - A_∞135(x,y) ,$$

where $S_0$ refers to the total intensity of the incident light, $S_1$ represents the amount of horizontal or vertical polarization, and $S_2$ represents the amount of 45° or 135° linear polarization.

Assuming that backscatter is partially polarized light, then we can estimate it by

$$B_{(x,y)} = \frac{B_{p(x,y)}}{P_{\text{cat}(x,y)}}$$

where $B_{p(x,y)}$ is the light intensity of the partially polarized part of the backscatter and $P_{\text{cat}(x,y)}$ is the DOP of the backscatter.

Previous studies have shown that the DOP of the backscatter is almost the same as that of the background light $A_∞(x,y)$ [7]. Moreover, using AOP to compute DOP, we are able to reduce the influence of the reflected light coming from the object [5]. Here, AOP can be deduced from the obtained Stokes vectors by Eq. (5) as follows:

$$\theta_{A∞(x,y)} = \frac{1}{2} \arctan \left( \frac{S_{2A∞(x,y)}}{S_{1A∞(x,y)}} \right) ,$$

$$P_{A∞(x,y)} = \frac{S_{1A∞(x,y)}}{S_{0A∞(x,y)}} \cdot \frac{\cos^2 \theta_{A∞(x,y)} - \sin^2 \theta_{A∞(x,y)}}{P_{\text{cat}(x,y)}}$$

where $\theta_{A∞(x,y)}$ and $P_{A∞(x,y)}$ refer to the AOP and DOP of $A_∞(x,y)$, respectively. It can be seen from Eq. (8) that the value of $A_∞(x,y)$ will influence the estimation of $P_{A∞(x,y)}$.

Taking into consideration that most objects can strongly depolarize the incident light, the DOP of the object’s reflection light tends to be relatively small. Here we neglect the DOP of the object’s reflection light; in other words, we regard the light intensity of the polarized part of $D_{(x,y)}$ to be 0 ($\Delta D_{(x,y)} = 0$).

Therefore, from Eq. (1) we can deduce that the light intensity
of the polarized part of $B_{p(x,y)}$ is equal to that of $I_{p(x,y)}$, namely,

$$B_{p(x,y)} = I_{p(x,y)}.$$  

Similar to Eqs. (7) and (8), we can deduce the AOP of $I_{p(x,y)}$ ($\theta_{1(x,y)}$) and the DOP of $I_{p(x,y)}$ ($\rho_{1(x,y)}$), and then we can obtain the light intensity of the polarized part of $I_{p(x,y)}$ by

$$B_{p(x,y)} = S_{0}I_{p(x,y)} \times \rho_{1(x,y)}.$$

Substituting $B_{p(x,y)}$ acquired by Eq. (9) as well as $P_{\text{scat}(x,y)}$ into Eq. (6), we can then obtain $B_{s(x,y)}$.

Inspired by the lower bound of the transmission in the dark channel prior method [13,14], we modify Eq. (4) by introducing a bias factor $\varepsilon$ ($\varepsilon > 1$) as follows:

$$L_{p(x,y)} = \frac{I_{p(x,y)}}{I_{p(x,y)}} - \frac{B_{p(x,y)}}{I_{p(x,y)}} = \frac{I_{p(x,y)} - \frac{B_{p(x,y)}}{\varepsilon}}{1 - \frac{B_{p(x,y)}}{\varepsilon}}. \quad (10)$$

Before this, we have obtained $A_{\infty(x,y)}$ and $B_{s(x,y)}$. The value of $\varepsilon$ has a significant impact on the characteristics and quality of the image, and there are different optimal values in different situations. We will describe the specific impact of the bias factor $\varepsilon$ on the image in Section 6 to guide the selection of the optimal value of $\varepsilon$. Eventually, we can obtain a recovered image that is expected to be clearer than the original one.

4. EXPERIMENTAL SETUP

Figure 3 presents the experimental apparatus. We collected polarization images at four orientations ($I_{0(x,y)}$, $I_{45(x,y)}$, $I_{90(x,y)}$, $I_{135(x,y)}$) using a polarization camera (LUCID PHX050S-PC) with a resolution of 2448 $\times$ 2048. As is shown in Fig. 3, the thickness of the glass tank is 3.60 cm, and the tank is full of water. We added 100 ml milk into the water, and the volume concentration of the milk was about 2%. In order to form a nonuniform optical field, we used a concentrated light source and also made the light to incident from the top to avoid the light reflected from the surface of the glass tank. In the side view there is a coordinate system to describe the relative position of the light source. In this experiment, our underwater object is a sphere-like plastic toy bird that is rich in surface textures, making it easier to visually evaluate the effectiveness of our method.

Fig. 3. Experimental setup. The left and right images correspond to the front and left views of the experimental setup, respectively. A Cartesian coordinate system is set on the left view to describe the incident angle of the light source.

5. TRAINING OF THE UNSUPERVISED LEARNING INPAINTING NETWORK

When training the inpainting network, the dataset is composed of a series of raw and complete background light images without any manual annotation as depicted in Fig. 4(a). This is why we call our deep learning algorithm unsupervised learning. The detailed acquisition process of the dataset is introduced as follows.

First, we remove the target object from the scene in order to obtain complete images of $A_{\infty(x,y)}$. Then we change the position of the light source to get random distributions of the optical field in the water, and we use the camera to acquire the images of $A_{\infty(x,y)}$ corresponding to each position of the light source. We collected 50 groups of images altogether to train the model.

After gaining the training dataset, we do data augmentation (as described in Appendix A) and randomly erase each train sample to generate the missing region as depicted in Fig. 4(b). Additionally, the input of the network during training is an image with a missing region. Such input is the result of multiplying the corresponding position of the training sample [shown in Fig. 4(a)] and a randomly generated binary mask [shown in Fig. 4(c)]. Figure 4(d) shows the weight map used in the loss function. The details of our loss function are illustrated in Appendix A.

During training, we use the stochastic gradient descent as the optimizer. In addition, we introduce the momentum mechanism and set its value to 0.9. In order to prevent overfitting as well as possible, we set a weight decay of 0.0005 for the optimizers of both the generative network and the discriminative network. In terms of the learning rate, for the generative network, the initial learning rate is 0.01; for the discriminative network, the initial learning rate is 0.001. At the same time, both will slow down to 1/10 of the original in the middle and late stages of the training. We use the “pytorch” deep learning framework and train 100 epochs on the NVIDIA GeForce RTX 2080 Ti.

6. RESULTS AND DISCUSSION

Based on the method proposed above, we perform a series of experiments using the trained unsupervised learning network. The enhancement measure evaluation (EME) is adopted to assess the quality of each image, which is defined as

$$EME = \frac{1}{k_1 k_2} \sum_{l=1}^{k_1} \sum_{k=1}^{k_2} 20 \log \frac{i_{\text{max},l}(x, y)}{i_{\text{min},l}(x, y) + q}, \quad (11)$$

Fig. 4. (a) Raw and complete image (background light); (b) the image after randomly erasing to generate a missing region (ROI); (c) the binary mask image (the missing region is 1); (d) the weight map of the reconstructive loss [the edge weight of the ROI is 5 (white), the middle weight of the ROI is 1 (gray), and the weight of the remaining area is 0 (black)].
where \( k_1 \) and \( k_2 \) refer to the number of image regions at two orthometric dimensions, \( \max_{k,l} I \) and \( \min_{k,l} I \) are the maximum and minimum intensity in the \((k, l)\) region, and \( q \) is a small number to ensure that the denominator is not equal to zero. In the following experiments, we choose \( k_1 = k_2 = 500 \). The higher the EME value, the better the image quality.

In the first experiment, we take four polarized images with specific angles of the polarizer \( 0^\circ, 45^\circ, 90^\circ, 135^\circ \) \((I_{0(x,y)}, I_{45(x,y)}, I_{90(x,y)}, I_{135(x,y)}\) respectively, as shown in Fig. 5. Then we use the Grab Cut algorithm [16] to remove the target object in the images and get the incomplete \( A_{\infty(x,y)} \). After that, we use the trained unsupervised learning network to get the complete \( A_{\infty(x,y)} \) corresponding to the four polarization directions \((A_{\infty0(x,y)}, A_{\infty45(x,y)}, A_{\infty90(x,y)} A_{\infty135(x,y)}\) as shown in Fig. 6.

According to Fig. 6, we can observe that there is an intensity gap between the inpainted missing region and the surrounding region. On one hand, this is because of the domain difference between our training set and test set. On the other hand, as our loss function shows in Appendix A, we pay more attention to the distribution of the missing region due to the higher weight of reconstructive loss when we perform image inpainting. In order to solve this problem, we use the filtering method, which is mentioned in Section 3.A, to smooth the image. The results are shown in Fig. 7.

It can be seen from Fig. 7 that the grayscale distribution of the supplementary part is similar to that of the original part in \( A_{\infty(x,y)} \). And because the datasets used to train the model are all target-free images taken in turbid media, which can be described as the value of backscatter at an infinite distance, the complete images have their physical meaning in the grayscale distribution and are closer to the real situation, rather than a purely mathematical process.

Then we continue calculating according to the algorithm flow chart described in Fig. 1, and we try some possible values of the bias factor \( \varepsilon \), and the results are presented in Fig. 8.

As can be observed from Fig. 8, as \( \varepsilon \) continues to increase, EME shows a decreasing trend, and the noise also shows a decreasing trend. Our goal is to obtain a larger EME value while there is not too much noise, so we can comprehensively measure EME and noise to choose a suitable \( \varepsilon \). Here we choose the bias factor \( \varepsilon = 1.7 \).

At the same time, for comparison, we apply several other methods to recover the same images. The results are shown in Fig. 9.

As can be observed from the appended EME values in Fig. 9, the proposed method in this paper can achieve good results in recovering underwater images under the nonuniform optical field. But the algorithm in Ref. [5] that made the simplification may fail as indicated in Fig. 9(b). Under the nonuniform optical fields, the DOP at each position in the nonuniform optical field does not concentrate on some specific values, so it is regarded as a constant, the recovery effect will be greatly reduced. This proves that the proposed method in this paper does have a good recovery effect in the nonuniform optical field.

Fig. 5. Polarisated images: (a) \( I_{0(x,y)} \), (b) \( I_{45(x,y)} \), (c) \( I_{90(x,y)} \), and (d) \( I_{135(x,y)} \), obtained when the light source incidents at \( 90^\circ \).

Fig. 6. Complete images: (a) \( A_{\infty0(x,y)} \), (b) \( A_{\infty45(x,y)} \), (c) \( A_{\infty90(x,y)} \), and (d) \( A_{\infty135(x,y)} \).

Fig. 7. Complete images after smooth filtering: (a) \( A_{\infty0(x,y)} \), (b) \( A_{\infty45(x,y)} \), (c) \( A_{\infty90(x,y)} \), and (d) \( A_{\infty135(x,y)} \).

Fig. 8. Recovered images with different values of the bias factor \( \varepsilon \).

Fig. 9. (a) Raw intensity image. The recovered images (b) using the method in Ref. [5]; (c) using the method in Ref. [6]; (d) using the CLAHE method in Ref. [17]; (e) using Dark Chanel Prior in Ref. [13]; and (f) using our method.
As for the polynomial fitting method in Ref. [6], it is valid under nonuniform optical fields and has a good visual quality as shown in Fig. 9(c). Compared with our method, we believe that its advantage lies in the low prior preparation cost because it does not need any extra data. Relatively, our method has a lower time cost when performing (the method of Ref. [6] that we reproduced with python takes about 20 s, while our method takes less than 5 s), and our visual quality is improved to a certain extent.

We can also see that the effect of the CLAHE method is outstanding in Fig. 9(d). However, it should be noted that the polarization theory is not used in the CLAHE. Actually, CLAHE is essentially different from the method in this article, and it belongs to a postprocessing method of image data. The CLAHE can be integrated into our method as a postprocessing method, and the recovered images are expected to be further improved.

In order to further examine the performance of the proposed method, we change the incident angle and intensity of the light source. The light source incidents from the oblique tops of 45° and 135° successively. The corresponding final results are illustrated in Figs. 10 and 11, respectively (ε = 1.7).

As can be observed, the proposed method can still recover the hazy underwater images well after changing the distribution of the nonuniform optical field. This is because the datasets used to train the unsupervised learning model consist of the images of the random incident direction, so it can adapt to the nonuniform optical field incident from most angles.

Then we change the target object, which is an aluminum alloy work piece, with coating and some symbols written on it. The shape, size, and surface characteristics of this target object are greatly different from those of the target object we used before. We gradually increase the turbidity of the medium by gradually increasing the amount of milk. The proposed method is then performed to recover the hazy images. The results are shown in Fig. 12.

It can be seen that the images of target objects with large differences in surface characteristics, shape, and size can be recovered. This is because the subtraction area is randomly selected during the training of the model, so it is applicable to target objects of various shapes. It is proved that the method proposed in this paper is applicable to a wide range of target objects and does not depend on the specific surface properties of the target object.

7. CONCLUSION

In this paper, we propose a polarimetric method for underwater image recovery under nonuniform optical fields based on unsupervised learning. We successfully use unsupervised learning to estimate polarization parameters changing with the spatial position. The experimental results demonstrate that the method proposed in this paper can recover the hazy underwater images in a short time. In particular, our method can perform well in nonuniform optical fields, where the classical methods, such as the method proposed in Refs. [4,5], may become invalid. Moreover, as we do not need manual annotation, our method has much less cost when acquiring datasets compared with the supervised learning method in Ref. [8].

APPENDIX A: SETUP OF THE PROPOSED UNSUPERVISED LEARNING INPAINTING NETWORK

1. Loss Function

As for the discriminative network, we hope that it can judge the true and false well [the image coming from the training set...
is judged as true, and the image generated by the generative network is judged as false. So we use the following loss function to train the discriminative network:

\[
\text{Loss}_D = \text{BCE}(D(\text{mask} \odot \text{image}), 1) \\
+ \text{BCE}(D(\text{mask} \odot G(\text{image} \odot (1 - \text{mask})), 0)),
\]

(A1)

where \(\text{BCE}\) represents the binary cross-entropy loss function; \(D(\cdot)\) and \(G(\cdot)\) refer to the mapping of the discriminative network and the generative network, respectively; image and mask are the raw background light [as shown in Fig. 4(a)] and the randomly generated binary mask [as shown in Fig. 4(c)], respectively; and \(\odot\) denotes the pixel product of the corresponding position.

For the generative network, there are two goals. One is to fit the ROI of the training samples as much as possible, and another is to cheat the discriminative network and make it go wrong. Here, the loss function associated with the former goal is named the reconstruction loss (represented by \(\text{L}_{\text{rec}}\)), and the latter is named the adversarial loss (represented by \(\text{L}_{\text{adv}}\)) [12]. Finally, these two loss functions are combined together to obtain the joint loss of the generative network [9]:

\[
\text{Loss}_G = \lambda_{\text{rec}} \ast \text{L}_{\text{rec}} + \lambda_{\text{adv}} \ast \text{L}_{\text{adv}},
\]

(A2)

where the coefficients \(\lambda_{\text{rec}}\) and \(\lambda_{\text{adv}}\) are the weights of the two kinds of loss, respectively. Note that the reconstruction loss \(\text{L}_{\text{rec}}\) can be realized by using the mean square error loss function with the edge weighting, which can make the generated ROI fit the surrounding area more closely:

\[
\text{L}_{\text{rec}} = \text{MSE}(G(\text{image} \odot (1 - \text{mask})), \text{image} \odot \text{mask} \odot \text{weight map}),
\]

(A3)

where \(\text{MSE}\) represents the mean square error loss function. The edge weighted loss map is shown in Fig. 4(d). The adversarial loss \(\text{L}_{\text{adv}}\) can be expressed by

\[
\text{L}_{\text{adv}} = \text{BCE} \left( D(\text{mask} \odot G(\text{image} \odot (1 - \text{mask}))), 1 \right),
\]

(A4)

where \(\text{BCE}\) represents the binary cross-entropy loss function. When training, for the joint loss, here we choose \(\lambda_{\text{rec}} = 0.995\) and \(\lambda_{\text{adv}} = 0.005\), which can get a higher visual quality repair.

2. Data Augmentation

In order to improve the generalization ability of the model and make the model adapt to more application scenarios, data enhancement is very necessary.

The ROI of our binary mask is a randomly generated rectangular region. Here, the “random” includes the following three aspects: random position, random area, and random aspect ratio. At the same time, in order for the model to adapt to any brightness environment, we also perform a random grayscale transformation on the training samples.

We randomly erase the rectangular area of the image, and the gray scale of the erased region is 255, so we can get the corresponding binary mask (the erased places are 1, and the other places are 0).

The area of the erased rectangular region is 0.05 to 0.5 times the area of the training sample image, and the aspect ratio is 0.33 to 3. In the meantime, the erased rectangular region will randomly appear in any position of the training sample [as shown in Figs. 4(b) and 4(c)].

As for the grayscale transformation, first we randomly scale all the gray scale of the training samples according to the probability \(p_1\):

\[
A_{\infty,x,y} = A_{\infty,x,y} \ast \text{poly},
\]

(A5)

where the symbol “poly” refers to the coefficient of grayscale scaling, which is a randomly generated number in 0.2 to 1. And here we choose the probability \(p_1 = 0.2\). We also invert the gray level of the training samples according to the probability \(p_2\):

\[
A_{\infty,x,y} = 255 - A_{\infty,x,y}.
\]

(A6)

Here we choose the probability \(p_2 = 0.1\).

Funding. National Natural Science Foundation of China (51775217, 51727809, 52022034); National Major Science and Technology Projects of China (2017ZX02101006-004); Key Research and Development Plan of Hubei Province (2020BAA008).

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

†These authors contributed equally to this work.

REFERENCES


