IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Double-Channel Guided Generative Adversarial Network for Image Colorization

KANGNING DU[1,2], CHANGTONG LIU[1,2], LIN CAO[1,2], YANAN GUO[1,2], FAN ZHANG[1,2],AND TAO WANG[1,2]

[1]The authors are with the Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science and Technology University, Beijing, 100101, China.
[2]The authors are with the School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing, 100101, China.

Corresponding author: Lin Cao (e-mail: charlin@bistu.edu.cn).

This work was supported by the National Natural Science Foundation of China (61671069, 62001033,62001034), the Qin Xin Talents Cultivation Program of Beijing Information Science and Technology University (QXTCP A201902), and by General Foundation of Beijing Municipal Commission of Education (KM202011232021).

ABSTRACT

Image colorization has a widespread application in video and image restoration in the past few years. Recently, automatic colorization methods based on deep learning have shown impressive performance. However, these methods map grayscale image input into multi-channel output directly. In the process, it usually loses detailed information during feature extraction, resulting in abnormal colors in local areas of the colorization image. To overcome abnormal colors and improve colorization quality, we propose a novel Double-Channel Guided Generative Adversarial Network (DCGGAN). It includes two modules: a reference component matching module and a double-channel guided colorization module. The reference component matching module is introduced to select suitable reference color components as auxiliary information of the input. The double-channel guided colorization module is designed to learn the mapping relationship from the grayscale to each color channel with the assistance of reference color components. Experimental results show that the proposed DCGGAN outperforms existing methods on different quality metrics and achieves state-of-the-art performance.

INDEX TERMS grayscale image colorization; generative adversarial network; double-channel guided colorization; reference component.

I. INTRODUCTION

COMPARED with grayscale images, color images can provide richer information. Image colorization is a technique that converts grayscale images into color images. It is widely used in video processing [1], film and television production [2], and photo restoration [3]. However, predicting missing color channels from a given single-channel grayscale image is an ill-posed problem [4]. Therefore, image colorization remains a challenging research problem.

Traditional colorization methods rely on the user's manual intervention to obtain satisfactory results, such as local color expansion [5]–[10] or color transfer [11]–[17]. The local color expansion method needs to specify a particular area of the grayscale image and diffuse the local color to the entire image. It requires a lot of manual work, such as color labeling, and the quality of colorization largely depends on manual colorization techniques.

Unlike the local color expansion methods that utilize user-supplied colors, color transfer methods exploit the colors of a reference image that are similar to the input image. Although these methods eliminate the influence of some human interventions on the image colorization, it still needs to select the reference image, and the colorization quality of the colored image depends on the selection of the reference image.

To overcome manual intervention limitations in traditional methods, automatic colorization methods [18]–[26] based on deep learning gradually appeared. Using large-scale data sets such as Google Landmarks [27] or MegaFace [28] dataset, the deep learning-based methods learn the end-to-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

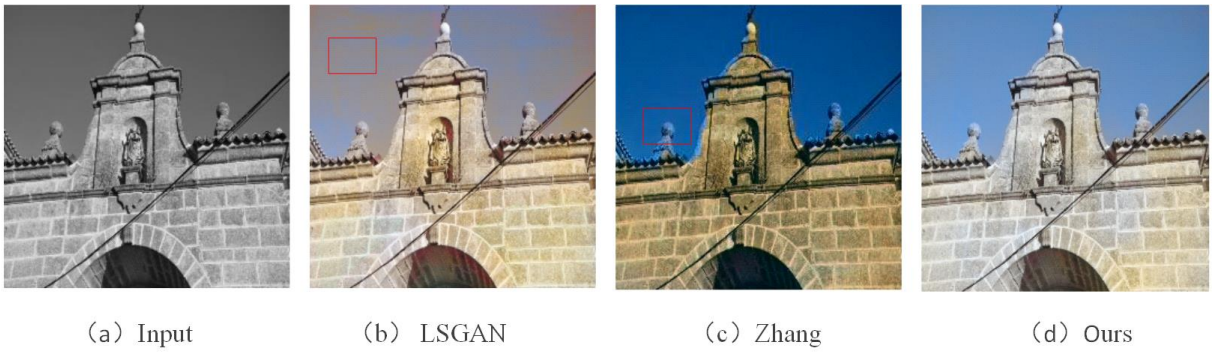

FIGURE 1: The limitations of existing deep learning methods.

end mapping relationship from grayscale to color image. Based on the deep learning method, automatic colorization can be realized by identifying different targets' semantic information in the grayscale image. However, most of the current colorization methods based on deep learning still have the problem of abnormal colors, even in the latest research [24]–[26]. Therefore, how to solve the abnormal colors in image colorization remains a challenging research problem awaiting exploration.

Figure 1 shows the common problems of the colorization results of the latest proposed methods. The result of the Least Squares Generative Adversarial Network(LSGAN) [29] method occurs lots of abnormal colors in the red box of Figure 1 (b). Compared with Figure 1(b), the image quality of Figure 1(c) has been improved, but still exists some abnormalities in some zone. For instance, in the red box of Figure 1 (c), the Zhang [21] method cannot distinguish the building's decoration from its background. The main reason is that the existing models usually lose detailed information during feature extraction and cannot effectively learn meaningful object-level semantics.

To solve the above problem, we propose a Double-Channel Guided Generative Adversarial Network (DCGGAN). The main innovation is to introduce two-color channel references as auxiliary information separately to guide the network to perform colorization. The proposed method performs a colorization task in the CIE Lab color space. In Lab color space, the L represents the luminance, the a and b represent the respective color dimension channels.

In the proposed method, a reference component matching module is designed to select the grayscale image's most suitable color components. Since the traditional deep learning methods are prone to the abnormal color problem, a double-channel guided colorization module is designed to provide color guidance for the image to be colored with reasonable reference information. In the double-channel guided colorization module, the corresponding reference components guide similar target colorization of the grayscale image input by constructing cascaded double colorization channel networks.

To summarize, this paper makes the following contributions:

(1) We propose a novel end-to-end DCGGAN framework for grayscale image colorization, which can generate rich and high-quality color images. In particular, our method can be applied to different image types (e.g., landscape and character).

(2) Use a-channel and b-channel reference components separately to guide double colorization networks to learn the corresponding color component's mapping relationship.

(3) A reference component matching module is proposed to automatically select the color reference component for different colorization channel networks.

(4) Qualitative and quantitative comparative experiments on two datasets and original images have proved the proposed method's superiority. Our method outperforms existing methods in addressing the problems of abnormal colors.

The rest of this paper is organized as follows: an overview of related works is provided in Section II; the proposed method is described in Section III; the experimental results and analysis are reported in Section IV; the conclusion is given in Section V.

II. RELATED WORK

A. LOCAL COLOR EXPANSION

Due to the image colorization's uncertainty, early attempts rely on user input to guide the image colorization process. In general, this kind of colorization methods spread the color of the user-specified area to the entire image. For instance, Levin et al. [5] encouraged assigning a similar color to adjacent pixels with similar luminance and formalized this premise using a quadratic cost function and obtain an optimization problem that can be solved using standard techniques. Huang et al. [6] integrated the colorization and the edge detection algorithm to reduce color bleeding. Luan et al. [7] and Qu et al. [8] improved the color propagation efficiency by employing a smoothness map to guide the incorporation of texture-similarity. Yatziv et al. [9] proposed the colorization method based on the luminance-weighted chrominance blending and fast intrinsic distance computations. In recent years, Heo et al. [10] proposed a color transfer method based on the Gaussian mixture model, which realized

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

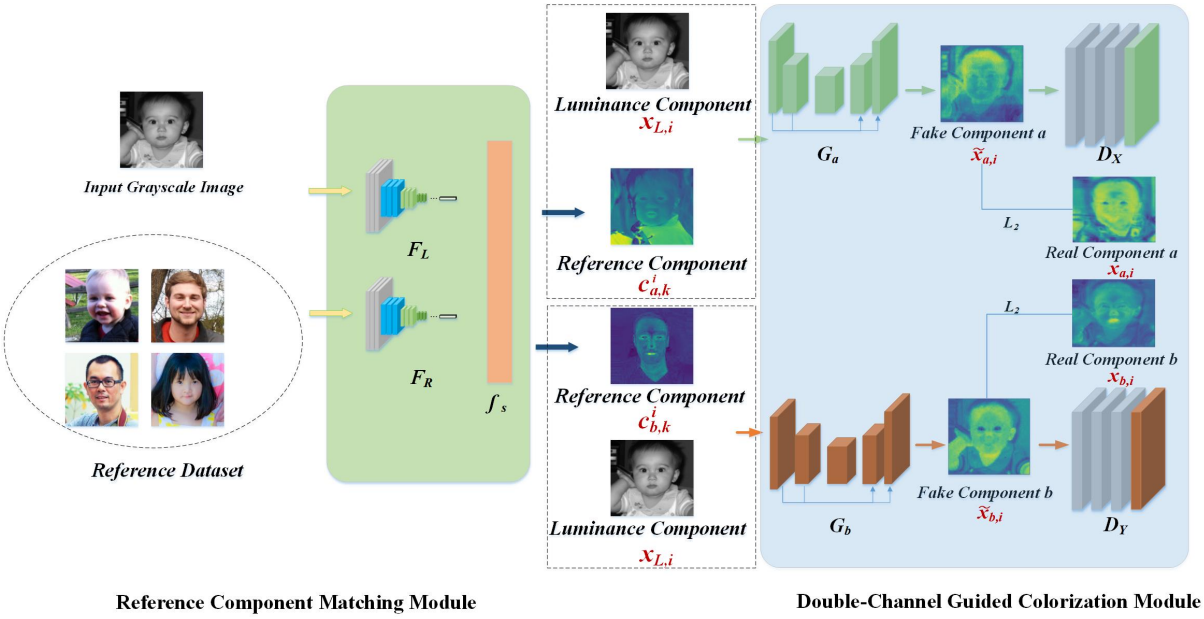

FIGURE 2: Overview of our model for automatic colorization of grayscale images.

the extension of image local color. This kind of method can realize grayscale images' colorization through the user's guidance but requires much manual operation.

B. COLOR TRANSFER

To reduce the user's workload, a colorization method based on the reference image specified by the user appeared. This colorization method required the user to specify one or more reference images with similar scenes and moved the colors to the grayscale image. For instance, Fu et al. [11] proposed a color transfer method based on a rarefaction curve that encodes the low-frequency colors, and use the Laplacian of residual colors to represent the high-frequency details to realize color transfer between images. Jin et al. [12] proposed a variational fusion model to fix up these incoherent colors obtained by a simple color transfer method with DFT and variance features. Later, these methods calculated the correspondences between the input and reference image based on some low-level similarity metrics measured at the pixel level [13], [14], or super-pixel level [15], [16]. Welsh et al. [14] introduced a general technique for "colorizing" grayscale images by transferring color between a source, color image, and a destination, grayscale image. Liu et al. [13] transferred color from the color reflectance image to the grayscale reflectance image, and obtain the final result by relighting with the illumination component of the target image. Gupta et al. [15] extracted features from the images at the resolution of superpixels, and exploit these features to guide the colorization process. Ironi et al. [16] presented a method for colorizing grayscale images by transferring color from a segmented example image. Each pixel is assigned a color from the appropriate region using a neighborhood matching metric, combined with spatial filtering for improved spatial coherence.

This kind of method highly depends on the similarity between the reference image and the input image, most of which still need to manually select the appropriate reference image. Even with an online image search system [17], it is still difficult to obtain suitable reference images.

C. LEARNING-BASED COLORIZATION

To solve the problem of user guidance dependency, the automatic colorization method based on deep learning has attracted more and more people's attention. These approaches need to learn the mapping relationship between grayscale input and the color image in a large-scale dataset. Iizuka et al. [18] and Zhao et al. [19] presented a two-branch architecture that jointly learns and fuses local image features and global priors (e.g., semantic labels). Larsson et al. [20] used the VGG neural network [30] to extract image features to predict each pixel's color. Zhang et al. [21] transformed the colorization problem into a classification problem for each pixel and built a convolutional neural network to achieve the effect of automatic grayscale image colorization. Later, they introduced sparse artificial annotation [22] to improve the naturalness of colorization. Messaoud et al. [23] proposed a deep learning framework to predict per-pixel color distributions. Hong et al. [24] employed the generative model in the joint intensity-gradient domain to improve the performance of automatic colorization and the visual perception of generated images. Zhao et al. [25] proposed a pixelated semantic color embedding and generator to generate realistic colored images.

With the development of Generative Adversarial Network (GAN) [31], a series of image style transfer methods based on GAN is proposed. The methods based on GAN can be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

applied to the field of image colorization. Mirza et al. [32] proposed a Conditional GAN (CGAN), which adds conditional information to the generator's input and the discriminator so that the generated result is not entirely free and unsupervised. Mao et al. [29] proposed the LSGAN, which replaces the cross-entropy loss in GAN with the least-squares loss, and used the distance metric to construct a stable and fast-convergent adversarial network. Arjovsky et al. [33] proposed Wasserstein GAN (WGAN), which solves the problem of unstable GAN network training by adjusting the calculation method of discriminator network activation function and loss function. Zhao et al. [26] presented a hierarchical Saliency Map-guided Colorization with GAN (SCGAN) to combine the low-level and high-level semantic information to help the system minimize semantic confusion and color bleeding. Compared with the traditional convolutional neural network, the GAN network improves the generated image's quality through confrontative learning between the generator and discriminator.

Although colorization technology has achieved remarkable results, the existing end-to-end models do not consider the partial abnormal color problem. For this reason, we propose the end-to-end DCGGAN model to study the automatic grayscale image colorization of the grayscale image under the guidance of reference components.

III. PROPOSED METHOD

A. NOTATION

To solve the problem of abnormal colors, the proposed DCGGAN comprises a reference component matching module and a double-channel guided colorization module, as shown in Figure 2. The DCGGAN introduces auxiliary information from the reference dataset to guide the process of colorization.

Specifically, we first design the reference component matching module to automatically select two reference components for the input luminance component(grayscale image). The reference components c_a and c_b are selected from the domain C_a and C_b, where the C_a and C_b respectively are a-channel and b-channel reference components domain in the reference dataset. Through the module, two pairs of luminance and reference color component c_a (c_b) are obtained.

Then, the double-channel guided colorization module is designed to learn the two mapping relationships from the luminance domain to generated color component domains $\tilde{X}_a$ and $\tilde{X}_b$, the learning process is shown in Eq. (1) and (2):

$$\widetilde{x}_a = f_a\left(x_L, c_a\right) \tag{1}$$

$$\widetilde{x}_b = f_b\left(x_L, c_b\right) \tag{2}$$

Where x_L represents the input luminance component. The $\tilde{x}_a$ and $\tilde{x}_b$ respectively represent generated a-channel and b-channel color components. The f_a and f_b represent the mapping relationships from x_L and c_a to $\tilde{x}_a$ and from x_L and c_b to $\tilde{x}_b$. In this module, we construct two cascaded colorization channel networks G_a and G_b to respectively learn the mapping relationships f_a and f_b, in which each channel network generate one color component. Finally, the input luminance and the generated color components are combined to obtain a color image.

B. REFERENCE COMPONENT MATCHING MODULE

As shown in Figure 3, the reference component matching module is proposed to select suitable reference components from the reference dataset for the input grayscale image. The reference data $C_{data} = \{(c_{L,k}, c_{a,k}, c_{b,k}) \,|c_{L,k} \in C_L, c_{a,k} \in C_a, c_{b,k} \in C_b, k = 1, \ldots, m\}$, where m is the number of the images in reference dataset, $c_{a,k}$ and $c_{b,k}$ respectively represent the k-th a-channel and b-channel color component in reference data.

The module consists of three parts: two feature extractor networks F_L and F_R and a feature similarity function f_s. Specifically, the feature extractor network F_L and F_R are used to obtain luminance and color components feature vectors respectively. To avoid an excessive increment of overall complexity, we adopt pre-trained ResNet18 [34] as the feature extractor network. The f_s is a feature similarity function. The detailed process is as follow:

In the first branch network F_L, the luminance component $x_{L,i}$ is input to obtain the feature vector $V_{L,i}$, where $x_{L,i}$ represents the i-th luminance component in the training dataset. Its input and output are as follow:

$$V_{L,i} = F_L(x_{L,i}) = \left[\begin{array}{cccc} l_{1,i} & l_{2,i} & l_{3,i} & \ldots & l_{n,i} \end{array}\right] \tag{3}$$

where n represents the size of the vector.

In the second branch network F_R, a-channel color components C_a and b-channel color components C_b in reference dataset are input to obtain the feature matrix M_a and M_b.

$$M_a = F_R\left(C_a\right) = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \cdots & a_{k,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \tag{4}$$

$$M_b = F_R\left(C_b\right) = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k,1} & b_{k,2} & \cdots & b_{k,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{bmatrix} \tag{5}$$

Where $\left[\begin{array}{cccc} a_{k,1} & a_{k,2} & \ldots & a_{k,n} \end{array}\right]$ and $\left[\begin{array}{cccc} b_{k,1} & b_{k,2} & \ldots & b_{k,n} \end{array}\right]$ are the k-th a-channel and b-channel color component feature representation of reference dataset in latent space. The m is the number of images in the reference dataset.

Then, through the feature similarity function, two feature similarity vectors are obtained as shown in Eq. (6) and (7). Among them, the $V_{a,i}$ is the feature similarity vector between the i-th luminance component and all a-channel reference

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

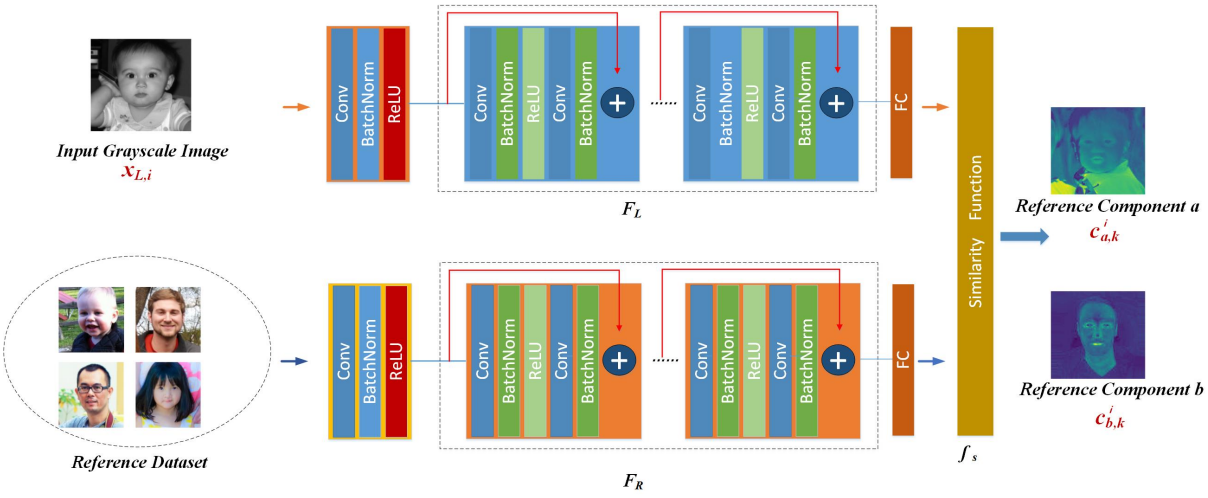

FIGURE 3: Reference Component Matching Module.

components. The $V_{b,i}$ is the feature similarity vector between the i-th luminance component and all b-channel reference components.

$$V_{a,i} = f_s\left(M_a, V_{L,i}\right) = \begin{bmatrix} \frac{\sum_{i=1}^{n} a_{1,i} \times l_i}{\sqrt{\sum_{i=1}^{n} (a_{1,i})^2} \times \sqrt{\sum_{i=1}^{n} (l_i)^2}} \\ \vdots \\ \vdots \\ \frac{\sum_{i=1}^{n} a_{m,i} \times l_i}{\sqrt{\sum_{i=1}^{n} (a_{m,i})^2} \times \sqrt{\sum_{i=1}^{n} (l_i)^2}} \end{bmatrix} = \begin{bmatrix} v^a_{1,i} & v^a_{2,i} & \cdots & v^a_{k,i} & \cdots & v^a_{m,i} \end{bmatrix}^T \tag{6}$$

$$V_{b,i} = f_s\left(M_b, V_{L,i}\right) = \begin{bmatrix} \frac{\sum_{i=1}^{n} b_{1,i} \times l_i}{\sqrt{\sum_{i=1}^{n} (b_{1,i})^2} \times \sqrt{\sum_{i=1}^{n} (l_i)^2}} \\ \vdots \\ \vdots \\ \frac{\sum_{i=1}^{n} b_{m,i} \times l_i}{\sqrt{\sum_{i=1}^{n} (b_{m,i})^2} \times \sqrt{\sum_{i=1}^{n} (l_i)^2}} \end{bmatrix} = \begin{bmatrix} v^b_{1,i} & v^b_{2,i} & \cdots & v^b_{k,i} & \cdots & v^b_{m,i} \end{bmatrix}^T \tag{7}$$

Where $v^a_{k,i}$ represents the similarity value between the k-th a-channel reference component and the i-th luminance component. The $v^b_{k,i}$ represents the similarity value between the k-th b-channel reference component and the i-th luminance component. The m is also equal to the number of images in the reference dataset.

Finally, as shown in Eq. (8) and (9), the color components corresponding to maximum similarity value in $V_{a,i}$ and $V_{b,i}$ are selected as reference components $c^i_{a,k}$ and $c^i_{b,k}$ of the i-th input luminance component.

$$c^i_{a,k} = \arg\max_k V_{a,i} = \begin{bmatrix} v^a_{1,i} & v^a_{2,i} & \cdots & v^a_{k,i} & \cdots & v^a_{m,i} \end{bmatrix}^T \tag{8}$$

$$c^i_{b,k} = \arg\max_k V_{b,i} = \begin{bmatrix} v^b_{1,i} & v^b_{2,i} & \cdots & v^b_{k,i} & \cdots & v^b_{m,i} \end{bmatrix}^T \tag{9}$$

As it is not easy to directly perceive the similarity between different color components and the luminance component through visual observation, the proposed reference component matching module avoids the manual selection problem and achieves the automatic selection of reference components. Besides, we do not select the reference color components by matching the luminance component $c_{L,k}$ of the reference dataset. Because it only considers texture information but does not consider color information. Our approach only matches one color component in each channel as the reference component to avoid misleading color.

C. DOUBLE-CHANNEL GUIDED COLORIZATION MODULE

In the CIE Lab color space, the traditional colorization methods based on deep learning need to simultaneously predict two color components (a-channel and b-channel) from a given grayscale image. To reduce the prediction uncertainty, we design a double-channel guided colorization module.

First, we construct two colorization channel networks to transform the traditional method's one-to-two (one luminance component to a-channel and b-channel color components) mapping problem into the one-to-one (one luminance component to one a-channel or b-channel color component) mapping problem in each colorization channel network. Compared with the traditional one-to-two mapping problem, for a single colorization channel, the one-to-one mapping problem can reduce the complexity of the latent mapping

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

space and decrease prediction difficulty to improve colorization quality.

Then, through reference components, we transform the one-to-one mapping problem into the two-to-one mapping problem. Compared with the one-to-one mapping problem, the additional auxiliary information is added to reduce the colorization uncertainty. The reference components as auxiliary information help improve the colorization's quality.

As shown in Figure 1, the double-channel guided colorization model is divided into two colorization channel networks G_a and G_b. In the module, the network learns from the training set.

$$T_{data} = \{(x_{L,i}, x_{a,i}, x_{b,i}) \,|x_{L,i} \in X_L, x_{a,i} \in X_a, x_{b,i} \in X_b, \\ i = 1, 2, \ldots, N\} \tag{10}$$

where N is the number of the images in the training dataset, $x_{a,i}$ and $x_{b,i}$ respectively represent the i-th a-channel and b-channel color component in training data, and $x_{L,i}$ represents the i-th L-channel color.

Specifically, for colorization network G_a, under the guidance of a-channel reference component $c^i_{a,k}$, the input luminance component $x_{L,i}$ maps to the a-channel color component $\tilde{x}_{a,i}$. For colorization network G_b, under the guidance of b-channel reference component $c^i_{b,k}$, the input luminance component $x_{L,i}$ maps to the b-channel color component $\tilde{x}_{b,i}$. These input and output are as follow:

$$\widetilde{x}_{a,i} = G_a\left(x_{L,i}, c^i_{a,k}\right) \tag{11}$$

$$\widetilde{x}_{b,i} = G_b\left(x_{L,i}, c^i_{b,k}\right) \tag{12}$$

Besides, the a-channel and b-channel color components represent two different ranges of color distributions. To decrease relevance between the generated a and b color components, we construct the cascaded two colorization channels without shared parameters.

Inspired by the idea of the UNet++ network [35], the generator network combines the features among different layers to further enhance the information representation of the feature map. The network structure is shown in Figure 4.

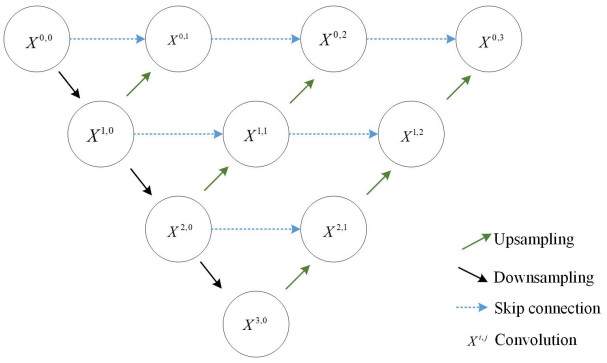

FIGURE 4: Generator network.

The generator network combines the deep-level information and the shallow-level information. Among them, the shallow-level information represents the contour information in the image, and the deep-level information means the semantic information of the image.

For instance, in the layer $x^{1,j}$, through the unite $x^{2,0}$, the shallow-level information output of $x^{1,0}$ and the deep-level information output of $x^{2,0}$ are fused, and then connected to the unit $x^{1,2}$. Combining and superimposing the feature of different layers makes the generator network better understand the input's texture and semantic information.

The discriminator networks D_X and D_Y are introduced to make the generated color components close to the actual color component. The D_X aims to distinguish the actual color component $x_{a,i}$ and the generated color component $\tilde{x}_{a,i}$. The D_Y aims to distinguish the color component $x_{b,i}$ and the generated color component $\tilde{x}_{b,i}$.

Inspired by the idea of the PatchGAN [36], the discriminator network predicts the authenticity of each pixel value of the generated component as shown in Figure 5.

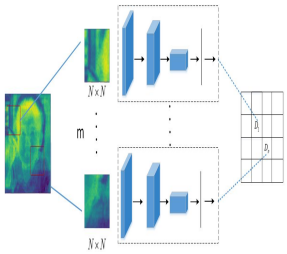

FIGURE 5: Discriminator network.

The purpose of the discriminator network is to perform a weighted summation on the entire image's authenticity. Its input is as follow:

$$\tilde{x}_{a,i} = \begin{bmatrix} \tilde{a}_{1,1} & \tilde{a}_{1,2} & \ldots & \tilde{a}_{1,j} & \ldots & \tilde{a}_{1,w} \\ \vdots & \vdots & \ddots & \vdots & \iddots & \vdots \\ \tilde{a}_{k,1} & \tilde{a}_{k,2} & \ldots & \tilde{a}_{k,j} & \ldots & \tilde{a}_{k,w} \\ \vdots & \vdots & \iddots & \vdots & \ddots & \vdots \\ \tilde{a}_{h,1} & \tilde{a}_{h,2} & \ldots & \tilde{a}_{h,j} & \ldots & \tilde{a}_{h,w} \end{bmatrix} \tag{13}$$

$$\tilde{x}_{b,i} = \begin{bmatrix} \tilde{b}_{1,1} & \tilde{b}_{1,2} & \ldots & \tilde{b}_{1,j} & \ldots & \tilde{b}_{1,w} \\ \vdots & \vdots & \ddots & \vdots & \iddots & \vdots \\ \tilde{b}_{k,1} & \tilde{b}_{k,2} & \ldots & \tilde{b}_{k,j} & \ldots & \tilde{b}_{k,w} \\ \vdots & \vdots & \iddots & \vdots & \ddots & \vdots \\ \tilde{b}_{h,1} & \tilde{b}_{h,2} & \ldots & \tilde{b}_{h,j} & \ldots & \tilde{b}_{h,w} \end{bmatrix} \tag{14}$$

where $a_{k,j}$ and $b_{k,j}$ represent the generated a-channel and b-channel color component pixel value at position k, j.

Through the discriminator network, the estimation matrices $P_{a,i}$ and $P_{b,i}$ are obtained as shown in Eq. (15) and (16). Among them, the $P_{a,i}$ and $P_{b,i}$ represent the i-th generated a-channel and b-channel color component estimation matrix.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

$$P_{a,i} = D_X\left(\tilde{x}_{a,i}\right) = \begin{bmatrix} p^a_{1,1} & p^a_{1,2} & \cdots & p^a_{1,j} & \cdots & p^a_{1,w} \\ \vdots & \vdots & \ddots & \vdots & \iddots & \vdots \\ p^a_{k,1} & p^a_{k,2} & \cdots & p^a_{k,j} & \cdots & p^a_{k,w} \\ \vdots & \vdots & \iddots & \vdots & \ddots & \vdots \\ p^a_{h,1} & p^a_{h,2} & \cdots & p^a_{h,j} & \cdots & p^a_{h,w} \end{bmatrix} \tag{15}$$

$$P_{b,i} = D_Y\left(\tilde{x}_{b,i}\right) = \begin{bmatrix} p^b_{1,1} & p^b_{1,2} & \cdots & p^b_{1,j} & \cdots & p^b_{1,w} \\ \vdots & \vdots & \ddots & \vdots & \iddots & \vdots \\ p^b_{k,1} & p^b_{k,2} & \cdots & p^b_{k,j} & \cdots & p^b_{k,w} \\ \vdots & \vdots & \iddots & \vdots & \ddots & \vdots \\ p^b_{h,1} & p^b_{h,2} & \cdots & p^b_{h,j} & \cdots & p^b_{h,w} \end{bmatrix} \tag{16}$$

Where $p^a_{k,j}$ and $p^b_{k,j}$ represent the prediction value of the generated pixel's authenticity at position k, j respectively.

D. LOSS FUNCTION

We adopt the generative adversarial loss and component constraint loss. The generative adversarial loss is responsible for supervising the image's quality to conform to the real sample distribution. The component constraint loss is responsible for constraining the colorization network's mapping space, reducing the redundant correspondence in the network's hidden layer so that the generation network can effectively learn the feature's color information. Specifically, the generative adversarial loss uses the Mean Square Error (MSE) function, which can increase the penalty for the generated images distributed far from the target domain so that the network can faster converge.

As shown in Eq. (17) and (18), L^a_{adv} and L^b_{adv} respectively supervise the authenticity of the generated color component $\widetilde{x}_a$ and $\widetilde{x}_b$.

$$\begin{aligned} L^a_{adv} = & E_{z\sim P_z}\left[\log\left(D_X\left(G_a\left(x_{L,i}, c^i_{a,k}\right)\right) - 1\right)\right] \\ & + E_{x_a\sim P_a}\left[\log D_X\left(x_{a,i}\right)\right] \end{aligned} \tag{17}$$

$$\begin{aligned} L^b_{adv} = & E_{z\sim P_z}\left[\log\left(D_Y\left(G_b\left(x_{L,i}, c^i_{b,k}\right)\right) - 1\right)\right] \\ & + E_{x_b\sim P_b}\left[\log D_Y\left(x_{b,i}\right)\right] \end{aligned} \tag{18}$$

where $E\left(*\right)$ is the expected value of the distribution, and P is the data distribution.

If the network is trained using only the generative adversarial loss, the hidden space's redundant mapping relationship cannot be solved. Therefore, we adopt the component constraint loss to reduce the distance between the generated and the target color component. As shown below, we use the L_2 distance to measure the color component constraint loss L^a_{dis} and L^b_{dis}.

$$L^a_{dis} = \left\|\tilde{x}_{a,i} - x_{a,i}\right\|_2 \tag{19}$$

$$L^b_{dis} = \left\|\tilde{x}_{b,i} - x_{b,i}\right\|_2 \tag{20}$$

where $\|\cdot\|$ represents the Euclidean distance.

Therefore, the total loss is shown in Eq. (21), which includes the generative adversarial loss L_{adv} and the component constraint loss L_{dis}. The L^a_{adv} is used to train the generator G_a and discriminator D_X. The L^b_{adv} is used to train the generator G_b and discriminator D_Y. The L_{dis} only updates the parameters of the generator network. The F_L and F_R employ the off-the-shelf pre-trained network, ResNet18, as our feature extractor.

$$L_{total} = L^a_{adv} + L^b_{adv} + \lambda L^a_{dis} + \rho L^b_{dis} \tag{21}$$

where λ and ρ are the trade-off hyperparameters between the component constraint loss and the generative adversarial loss.

The final training goal, as shown below, is to obtain the parameters of the colorization model when the total loss value achieves the minimum.

$$G^*_a, G^*_b = arg \min_{G_a, G_b} L_{total}(G_a, G_b, D_X, D_Y) \tag{22}$$

IV. EXPERIMENT

A. EXPERIMENTAL SETTING

In order to objectively evaluate the quality of the generated image. We adopt the Structural Similarity (SSIM) [37], the Peak Signal to Noise Ratio (PSNR) [38], the Spatial Frequency (SF) [39], and the Normalized Cross-Correlation (NCC) [40] as the performance evaluation indicator. Among them, the SSIM is used to measure the structural similarity between the original and the generated images. The PSNR is used to measure the average error between the original and the generated images. The SF reflects the rate of change of image grayscale. The NCC can calculate the consistency between the two targets. The larger the SSIM, the PSNR, the SF, and the NCC value, the better the quality of the generated image.

In our experiment, the Google Landmarks dataset [27] and MegaFace dataset [28] are used to evaluate our method. These two datasets contain common image types in reality and could meet different colorization demands. The MegaFace dataset is publicly released by the University of Washington Computer Science and Engineering Laboratory. It contains thousands of different postures, various light intensities of natural shots, and no restrictions on the character's clothes, background, and posture. The Google Landmarks dataset is currently the world's largest landmark landscape dataset, released by Google. It records tens of thousands of unique landmarks in every corner of the world, including different natural scenery and human-made buildings. In each dataset, 1000 images are divided into the training set and 150 images are divided into the testing set. The size of all images is cropped to 256×256 pixels. Some examples from the two databases are shown in Figure 6.

Since the large-scale reference dataset would consume massive GPU computing resources and reduce model colorization efficiency, we randomly selected 600 different images. Furthermore, by comparing the SSIM indicators between them, we only selected 45 images with varying representative styles to make up our reference dataset.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

FIGURE 6: Selected images are from the MegaFace and Google Landmarks datasets. The first and second rows are the corresponding images from the MegaFace and Google Landmarks datasets, respectively.

We adopt a three-step training process as follows. First, the reference component matching module uses a pre-trained network model to match the training image's reference component. Then, two matching pairs of the luminance component and reference component train the double-channel colorization network. Finally, the structure and network parameters of the model are saved.

The testing process is similar to the training process. First, the model of the colorization network is loaded. Then, the a-channel and b-channel reference components of the test image are matched. Finally, the generated a-channel and b-channel color components combine with the input luminance component to form the full colorization image.

The proposed framework is implemented by using the PyTorch deep learning library. The batch size is set to 8 in the stage of training. The model is trained using the Adam [41] optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The initial learning rate is set to 0.0001. The trade-off hyperparameter λ an ρ between the generative adversarial loss and the component constraint loss are set to 0.01.

B. ABLATION EXPERIMENT

In this section, we perform sufficient ablation study on MegaFace and Google Landmarks datasets to evaluate each component's contribution to our network. The settings of the ablation experiment are as follows:

(a) Baseline: The RCM module and the double-channel guided colorization module are removed simultaneously. Only one single generator network is used to generate color components. It directly maps the input luminance component into two color components $\widetilde{x}_a$ and $\widetilde{x}_b$.

(b) DCGGAN without RCM (w/o RCM): The RCM module is removed. To verify the effectiveness of the double-channel colorization network, the DCGGAN only trains the double-channel colorization module to generate color components without the assistance of reference components. In the colorization network of each channel, it respectively maps the input luminance component into the single respective color component.

(c) w/o RCM + S_{ab}: The RCM module is removed, and the reference components are randomly selected based on color components in the reference dataset. The $c_{a,k}$ and $c_{b,k}$ of random selection are reference components $c^i_{a,k}$ and $c^i_{b,k}$. The S_{ab} represents the method of directly matching color reference components $c^i_{a,k}$ and $c^i_{b,k}$. We set the ablation experiment to prove that the proposed RCM module is superior to the S_{ab} random selection method.

(d) w/o RCM + S_L: The RCM module is removed, and the reference components are randomly selcted based on the luminance component $c_{L,k}$. The color components $c_{a,k}$ and $c_{b,k}$ in the reference dataset corresponding to the $c_{L,k}$ are the reference components $c^i_{a,k}$ and $c^i_{b,k}$. The S_L represents the method of indirectly matching reference color components $c^i_{a,k}$ and $c^i_{b,k}$ through selecting $c_{L,k}$. The ablation experiment is set to prove that the proposed RCM module is superior to the S_L random selection method.

(e) DCGGAN + S_L: Through the RCM module, it matches luminance component $c_{L,k}$ in the reference dataset. The color components $c_{a,k}$ and $c_{b,k}$ corresponding to the matching luminance component $c_{L,k}$ are the reference components $c^i_{a,k}$ and $c^i_{b,k}$. We set the ablation experiment to prove the effectiveness of matching color components through the RCM module directly.

(f) DCGGAN + S_m: Through the RCM module, the multiple color components $c^i_{a,k1}$, $c^i_{a,k2}$, …, $c^i_{a,kn}$ and $c^i_{b,k1}$, $c^i_{b,k2}$, …, $c^i_{b,kn}$ are matched with the input luminance component, where n represents the number of reference components in each colorization channel. We set the ablation experiment to prove that the multiple different reference components cannot effectively guide image colorization.

The colorization results of ablation experiments are shown in Figure 7. With the comparison of the first two ablation experiments, the effectiveness of the double-channel guided colorization module is proved. The results of the baseline occur abnormal color in a larger area. For instance, in the distant mountain of the first column in Figure 7 (a), an unreasonable yellow appears. We can find the abnormal color of w/o RCM can be significantly reduced, but the results also contain a few wrong colors. For instance, in the last column of Figure 7 (b), the interference color (green) appears in the woman's hair. In addition, the colorization image's overall brightness is low, as shown in the third column of Figure 7 (b), the sky's primary color appears to be gray. Because the single colorization network is difficult to accurately predict the missing multiple color channel.

With the comparison of the rest of the ablation experiments and our approach, the reference component matching module's effectiveness can be proved. Specifically, by observation, the results of w/o RCM + S_{ab} and w/o RCM + S_L are the worst, indicating that the matched module has apparent advantages over random selection. The randomly chosen color components as reference components might mislead the colorization network channel. For instance, in the fourth column of Figure 7 (c), the boy's clothing has large areas of unreasonable colors. Because through the RCM module, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

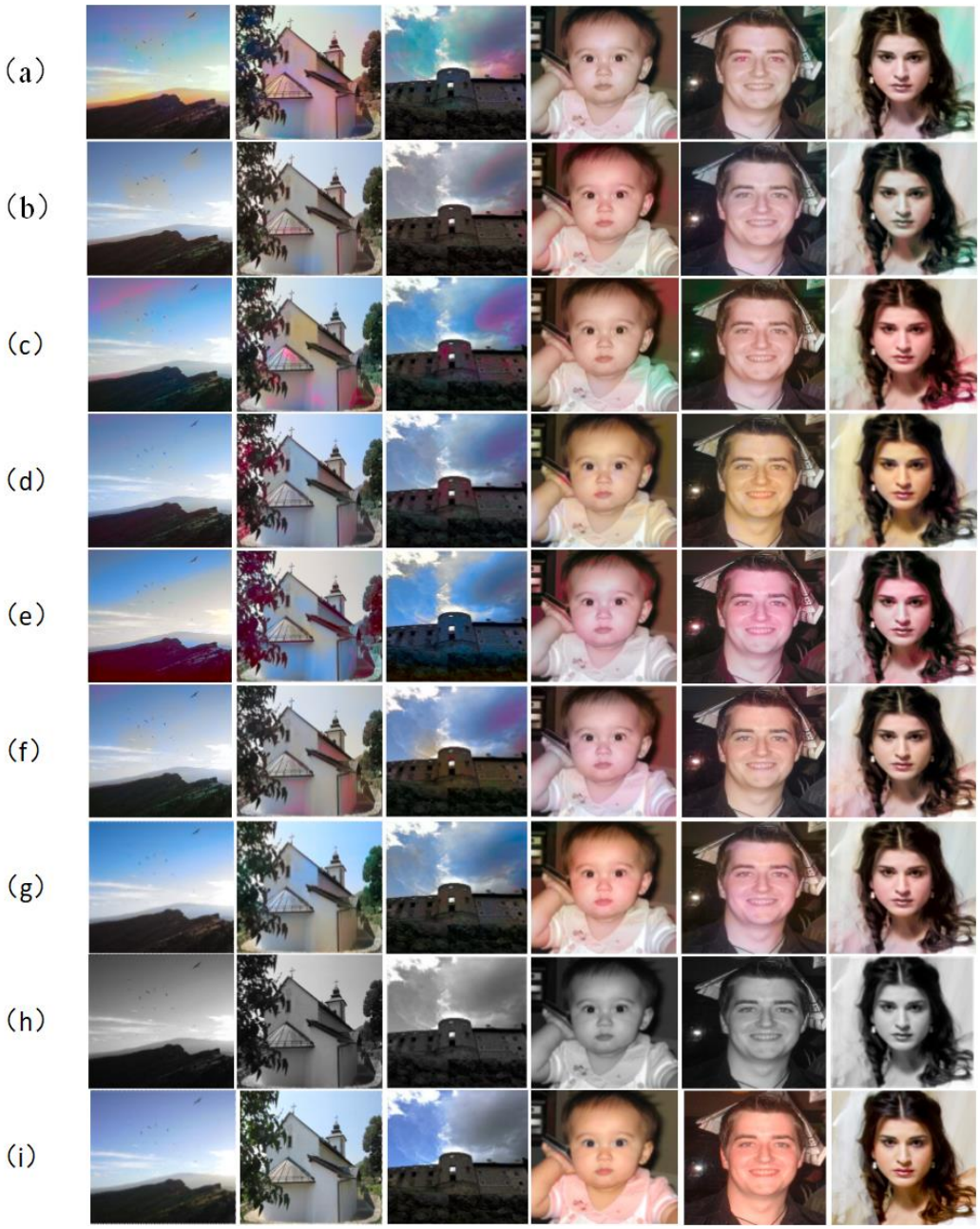

FIGURE 7: Colorization results of different ablation experiments on Google Landmarks and MegaFace dataset. (a)Baseline. (b)DCGGAN w/o RCM (Reference Component Matching Module). (c)DCGGAN w/o RCM + S_{ab}. (d)DCGGAN w/o RCM + S_L. (e)DCGGAN + S_L. (f)DCGGAN + S_m.(g)Ours. (h)Grayscale images.(i)Original images.

suitable reference components color can be selected.

We further analyze the DCGGAN + S_m. Take adding the first two closest reference color components in each colorization channel network as an example. As shown in Figure 9, in the same colorization channel, we find that the two reference color components are not from the same reference image. It indicates that the RCM module considers the similarity of the appearance structural features and refers to the rationality of the color information.

As shown in Figure 10, we compare our approach with the DCGGAN + S_m. We find that the results of our approach perform better in colorization accuracy and overall lightness. The results of the DCGGAN + S_m indicate that the more reference components are added, the worse the effect. We further compare our generated color components with the generated color components of the DCGGAN + S_m experiment. Figure 10 (b) shows that the color components appear chaotic and blurred, such as the sky's central area. Therefore, we conclude that multiple reference components mislead the colorization network. The double-channel guided

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

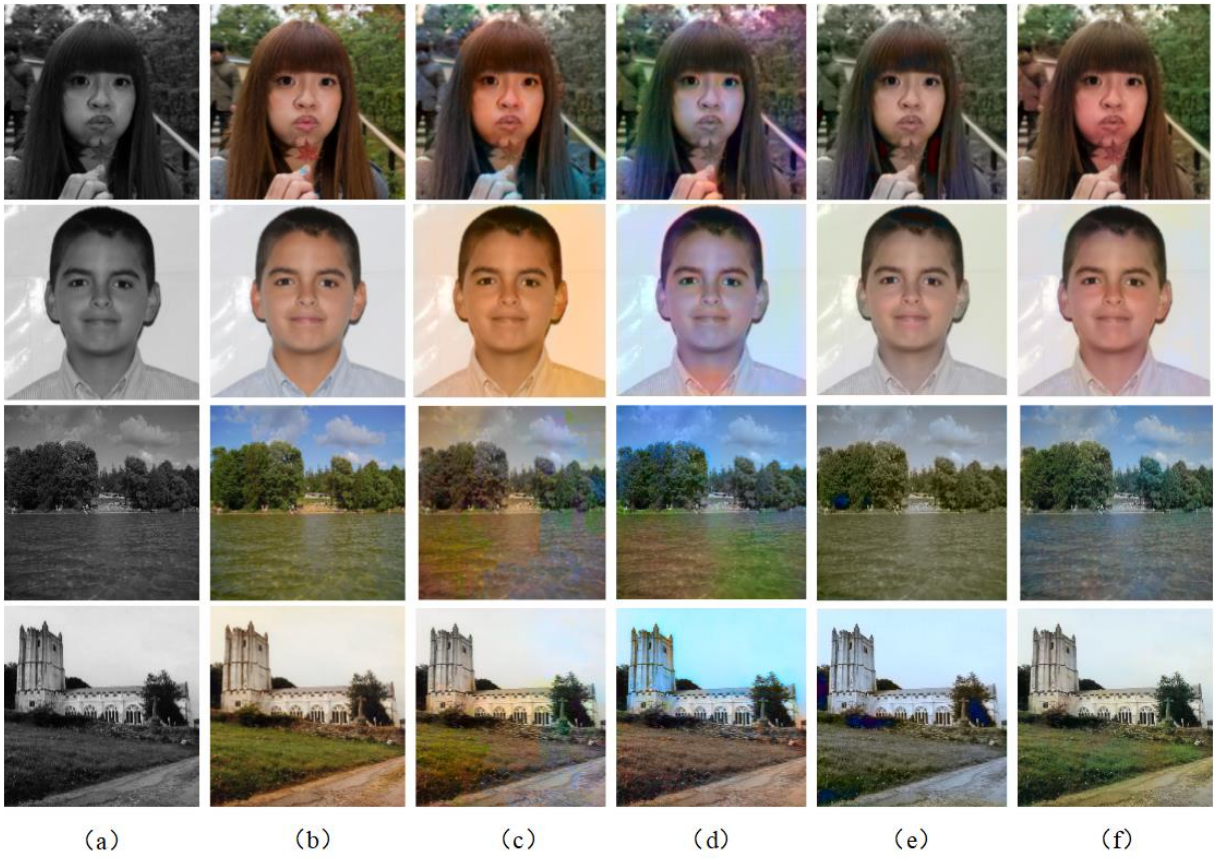

FIGURE 8: Comparison of different GAN methods. (a) L-channel image; (b) Original image; (c) CGAN [32]; (d) LSGAN [29]; (e) WGAN [33]; (f) Ours.

TABLE 1: Comparison of different indicators in the ablation experiment.

Ablation	Google Landmarks				MegaFace			
	SSIM/%	PSNR/dB	SF	NCC	SSIM/%	PSNR/dB	SF	NCC
Baseline	93.61	31.65	26.12	0.52	94.37	32.02	26.77	0.55
w/o RCM	93.72	31.76	26.29	0.56	94.52	32.16	26.85	0.57
w/o RCM + S_{ab}	93.97	32.34	26.96	0.58	94.83	32.59	27.38	0.62
w/o RCM + S_L	94.56	32.62	27.35	0.63	95.34	32.85	27.50	0.66
DCGGAN + S_L	94.87	32.85	27.62	0.66	95.59	33.13	28.09	0.69
DCGGAN + S_m	95.32	33.34	28.45	0.72	95.91	33.38	28.49	0.73
Ours	**95.45**	**33.49**	**28.58**	**0.75**	**96.02**	**33.68**	**29.24**	**0.78**

TABLE 2: One image average time consumption of different methods.

Methods	Prediction time(s)
baseline	1.02
w/o RCM(Signal layer)	0.98
w/o RCM + S_{ab}	1.14
w/o RCM + S_L	1.13
MCGGAN + S_L	1.25
DCGGAN + S_m	1.41
Proposed	**1.33**

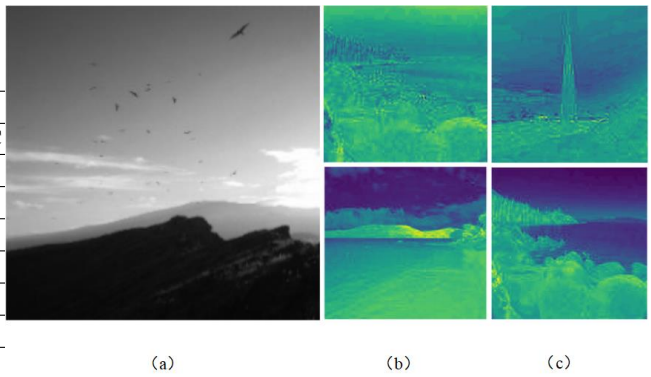

FIGURE 9: Multiple reference component matching results. (a) L-channel luminance image; (b) the first reference component of a and b channels; (c) the second reference component of a and b channels.

colorization module cannot effectively predict the luminance component's missing color information under multiple reference components.

We use four data indicators to evaluate the colorization results of different ablation experiments. As shown in Table 1, the values of different indicators are calculated on the Google

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

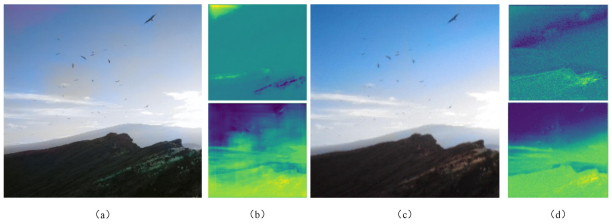

FIGURE 10: Comparison of a single reference component and multiple reference components. (a) The colorization results of multiple reference components; (b) The a-channel and b-channel of the colorization results of multiple reference components; (c) The colorization results of a single reference component; (d) The a-channel and b-channel of the colorization result of a single reference component.

Landmarks and MegaFace testing datasets, respectively. Our approach is also superior to other ablation experiments.

Finally, in different colorization models, we input multiple 256×256 grayscale images and calculate the average time it takes for each model to colorize one grayscale image, as shown in Table 2. Among them, the baseline is the one-to-two mapping relationship, and the ablation experiment w/o RCM is the one-to-one mapping relationship. The remaining ablation experiments are the two-to-one mapping relationship. In Table 2, we can conclude that the complexity of the two-to-one mapping problem is the highest. It is mainly because the reference component needs to be selected. Nevertheless, the colorization effect of the two-to-one mapping relationship is the best. Comparing with the one-to-one mapping methods, the time consumption increment of the proposed method is not significant.

C. COMPARISON OF GAN BASED METHOD

The DCGGAN proposed in this paper is a colorization model based on GAN. Therefore, we compare DCGGAN with other mainstream GAN. As shown in figure 8, the first two rows of testing data are from the MegaFace dataset, and the second two rows are from the Google Landmark testing dataset.

We found that the colorization accuracy of the CGAN is relatively high, and the color is also more natural under intuitive experience. However, the color consistency of the result is low, and the color difference is massive. For instance, in the third row of figure 8 (c), two different colors appear on the lake surface: green and blue. The colorization image of the LSGAN is more vivid than the CGAN, but there are more interference colors in the target. For example, in the fourth row of figure 8 (d), a large area of interference color appears in the sky. The WGAN removes the last layer of the discriminant network's Sigmoid function and solves the generated network's unstable training. The colorization accuracy of WGAN is higher than that of the CGAN and LSGAN, but the colorization image saturation is low, and the brightness is dark. After observation, it can be found that the colorization method proposed in this paper is significantly

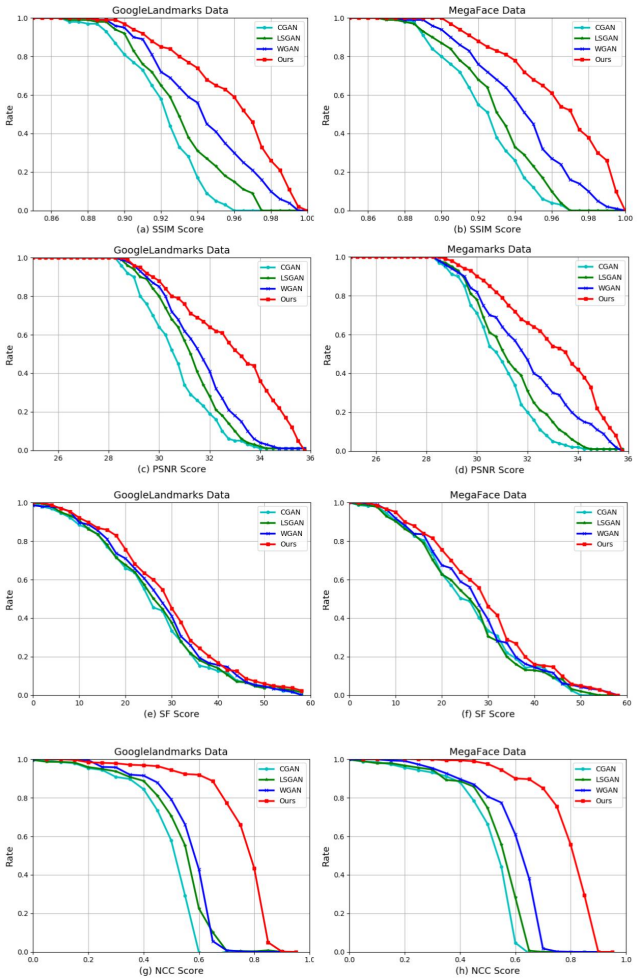

FIGURE 11: The performance of different GAN methods. (a) SSIM indicator score on Google Landmarks dataset; (b) SSIM indicator score on MegaFace dataset; (c) PSNR indicator score on Google Landmarks dataset; (d) PSNR score on MegaFace dataset; (e) SF indicator score on Google Landmarks dataset; (f) SF indicator score on MegaFace dataset; (g) NCC indicator score on Google Landmarks dataset; (h) NCC indicator score on MegaFace dataset.

better than other contrasting colorization methods in terms of color naturalness, image brightness, and colorization accuracy.

To quantitatively analyze GAN's effectiveness in different data sets, we further compare the four indicators of the mentioned GANs. As shown in Table 3, the average values of different evaluating indicators of different GAN methods in Google Landmarks and MegaFace data sets are reported. It can be seen from the table that in different data sets, the SSIM, the PSNR, the SF and the NCC indicator values of our approach are the highest. It shows that our approach's colorization image has the highest similarity with the original image, and image quality is the best. As shown in Figure 11, we count the test data distribution in different index

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

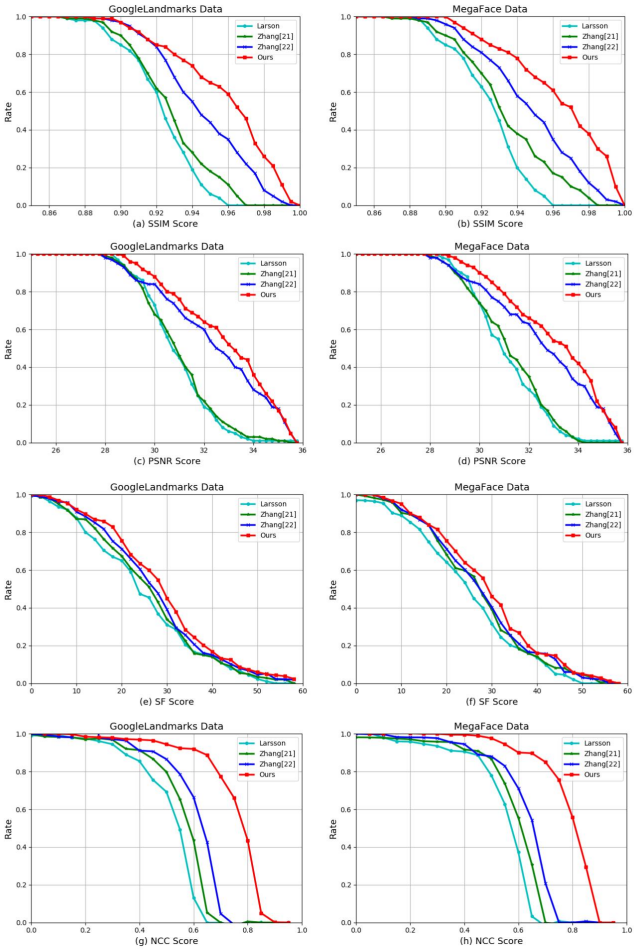

FIGURE 12: The performance of different colorization methods. (a) SSIM indicator score on Google Landmarks dataset; (b) SSIM indicator score on MegaFace dataset; (c) PSNR indicator score on Google Landmarks dataset; (d) PSNR score on MegaFace dataset; (e) SF indicator score on Google Landmarks dataset; (f) SF indicator score on MegaFace dataset; (g) NCC indicator score on Google Landmarks dataset; (h) NCC indicator score on MegaFace dataset.

values to visually represent different GAN methods' index evaluation values. The abscissa represents the index value, and the ordinate represents the proportion of test data that meets the standard. Obviously, the curve of our approach is significantly higher than other GANs in the two datasets.

Besides, it can be noticed that the index values of the same methods on the MegaFace data set are significantly higher than those on the Google Landmarks data set because landscape colorization is more complicated caused by the various landscape changes. Moreover, in different seasons, the same place also has different landscapes, making the colorization network challenging to learn.

Finally, in different GANs, we input multiple 256×256 grayscale images and calculate the average time it takes to colorize one grayscale image, as shown in Table 4. Compared

TABLE 3: Comparison of different indicators in the GAN methods.

GAN	Google Landmarks				MegaFace			
	SSIM/%	PSNR/dB	SF	NCC	SSIM/%	PSNR/dB	SF	NCC
CGAN [32]	92.86	30.23	25.85	0.48	93.02	30.93	26.43	0.50
LSGAN [29]	93.24	31.12	26.07	0.52	93.76	31.42	26.65	0.53
WGAN [33]	93.92	31.84	26.94	0.55	94.27	32.02	27.05	0.59
Ours	**95.45**	**33.49**	**28.58**	**0.75**	**96.02**	**33.68**	**29.24**	**0.78**

TABLE 4: One image average time consumption of different GANs.

Methods	Prediction time(s)
CGAN	1.16
LSGAN	1.28
WGAN	1.05
Proposed	**1.33**

with the traditional GANs, we have added the RCM module to select reference components, so the colorization process takes longer time. However, the colorization effect of the proposed method is the best. The complexity increment of the proposed method is not significant.

D. COMPARISON OF COLORIZATION METHODS

In this section, to evaluate our method's performance, we compare our colorization results with several state-of-the-art colorization methods on the MegaFace and Google Landmarks datasets. In this section, we compare our approach with other colorization methods.

As shown in figure 14, Larsson et al. [20] use the VGG network to extract image features, which of the colorization image's overall brightness is lower. The colorization results of Zhang et al. [21] are brighter than Larson, but the color accuracy is low. Compared to the first two methods, the colorization results of Zhang et al. [22] are better in colorization accuracy and overall brightness. However, there are

TABLE 5: Comparison of different indicators in the colorization methods.

Method	Google Landmarks				MegaFace			
	SSIM/%	PSNR/dB	SF	NCC	SSIM/%	PSNR/dB	SF	NCC
Larsson [20]	93.19	30.73	25.90	0.51	93.87	31.24	26.59	0.54
Zhang [21]	93.21	30.86	26.14	0.55	93.95	31.32	26.71	0.58
Zhang [22]	94.24	32.79	27.46	0.60	95.54	33.42	28.32	0.62
Ours	**95.45**	**33.49**	**28.58**	**0.75**	**96.02**	**33.68**	**29.24**	**0.78**

TABLE 6: One image average time consumption of different colorization methods.

Methods	Prediction time(s)
Larsson [20]	3.28
Zhang [21]	2.96
Zhang [22]	2.13
Proposed	**1.33**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

FIGURE 13: Comparision of complexity in the ablation experiment.

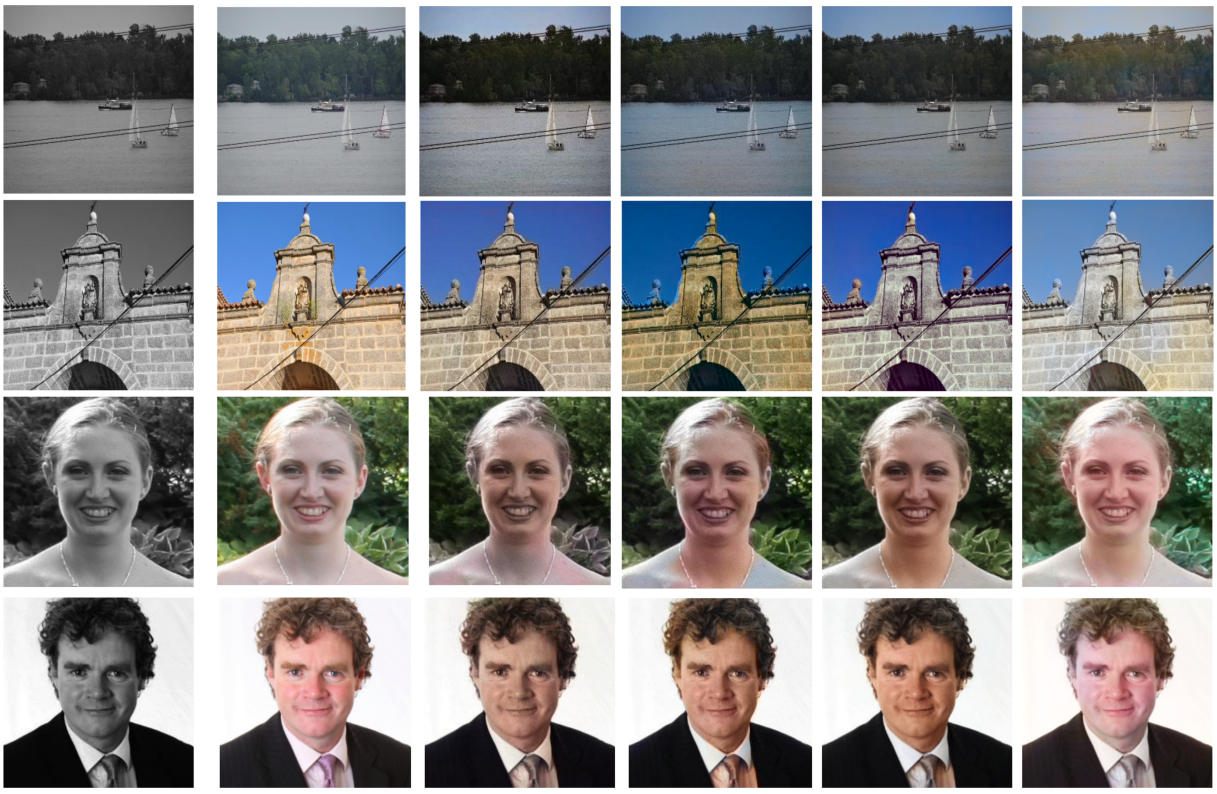

FIGURE 14: Comparison of different colorization methods. (a) Input image; (b) Original image; (c) Larsson [20]; (d) Zhang [21]; (e) Zhang [22]; (f) Ours.

interference colors in local areas.

Our approach's colorization results are natural, and the same target's interference color is the least. However, there is still the problem of uneven color distribution in the proposed method. In some areas of the colorization image, the ideal saturation is still not reached, as shown in the second row of Figure 14 (f).

We report the quantitative comparisons of other colorization methods on two datasets in Table 5. As shown in the table, the indicator values of our approach are the highest. As shown in Figure 12, we also calculate the test data distribution in different index values to visually represent other colorization methods' index evaluation values. The abscissa represents the index value, and the ordinate represents the proportion of test data that meets the standard. It can be seen from the curves that the DCGGAN achieves higher performance than other methods on SSIM, PSNR, SF and NCC scores.

Finally, in different colorization models, we input multiple 256×256 grayscale images and calculate the average time it takes to colorize one grayscale image, as shown in Table 6. With the comparison of the different colorization methods, our method is superior to other methods both in colorization effect and efficiency.

V. CONCLUSION

To address the abnormal color problems, we propose a novel DCGGAN colorization method. It consists of a reference component matching module and a double-channel guided colorization module. The reference component matching module selects suitable reference components as auxiliary information to guide colorize the grayscale image. The double-channel guided colorization module is proposed to map the grayscale image into plausible color output with the assistance of reference color components. Through experiments on the different datasets, it illustrates that our approach generates the image with reasonable color. Moreover, with the comparison of the state-of-the-art methods, abnormal color is significantly reduced, and the colorization accuracy has been dramatically improved.

REFERENCES

[1] C. Lei and Q. Chen, "Fully automatic video colorization with self-regularization and diversity," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, pp. 3753–3761, 2019.

[2] M. R. Lavvafi, S. A. Monadjemi, and P. Moallem, "Film colorization,

using artificial neural networks and laws filters," J. Comput., vol. 5, no. 7, pp. 1094–1099, 2010.

[3] F. Mameli, M. Bertini, L. Galteri, and A. Del Bimbo, "Image and video restoration and compression artefact removal using a nogan approach," in Proceedings of the 28th ACM International Conference on Multimedia, pp. 4539–4541, 2020.

[4] J. Su, H. Chu, and J. Huang, "Instance-aware image colorization," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020, pp. 7965–7974, 2020.

[5] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," ACM Trans. Graph., vol. 23, no. 3, pp. 689–694, 2004.

[6] Y. Huang, Y. Tung, J. Chen, S. Wang, and J. Wu, "An adaptive edge detection based colorization algorithm and its applications," in Proceedings of the 13th ACM International Conference on Multimedia, Singapore, November 6-11, 2005, pp. 351–354, 2005.

[7] Q. Luan, F. Wen, D. Cohen-Or, L. Liang, Y. Xu, and H. Shum, "Natural image colorization," in Proceedings of the Eurographics Symposium on Rendering Techniques, Grenoble, France, 2007, pp. 309–320, 2007.

[8] Y. Qu, T. Wong, and P. Heng, "Manga colorization," ACM Trans. Graph., vol. 25, no. 3, pp. 1214–1220, 2006.

[9] L. Yatziv and G. Sapiro, "Fast image and video colorization using chrominance blending," IEEE Trans. Image Process., vol. 15, no. 5, pp. 1120–1129, 2006.

[10] Y. Heo and H. Jung, "Probabilistic gaussian similarity-based local colour transfer," Electronics Letters, vol. 52, no. 13, pp. 1120–1122, 2016.

[11] Q. Fu, Y. He, F. Hou, J. Zhang, A. Zeng, and Y. Liu, "Vectorization based color transfer for portrait images," Comput. Aided Des., vol. 115, pp. 111–121, 2019.

[12] Z. Jin, L. Min, M. K. Ng, and M. Zheng, "Image colorization by fusion of color transfers based on DFT and variance features," Comput. Math. Appl., vol. 77, no. 9, pp. 2553–2567, 2019.

[13] X. Liu, L. Wan, Y. Qu, T. Wong, S. Lin, C. Leung, and P. Heng, "Intrinsic colorization," ACM Trans. Graph., vol. 27, no. 5, p. 152, 2008.

[14] T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring color to greyscale images," ACM Trans. Graph., vol. 21, no. 3, pp. 277–280, 2002.

[15] R. K. Gupta, A. Y. S. Chia, D. Rajan, E. S. Ng, and Z. Huang, "Image colorization using similar images," in Proceedings of the 20th ACM Multimedia Conference, MM '12, Nara, Japan, October 29 - November 02, 2012, pp. 369–378, 2012.

[16] R. Ironi, D. Cohen-Or, and D. Lischinski, "Colorization by example," in Proceedings of the Eurographics Symposium on Rendering Techniques, Konstanz, Germany, June 29 - July 1, 2005, pp. 201–210, 2005.

[17] A. Y. S. Chia, S. Zhuo, R. K. Gupta, Y. Tai, S. Cho, P. Tan, and S. Lin, "Semantic colorization with internet images," ACM Trans. Graph., vol. 30, no. 6, p. 156, 2011.

[18] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," ACM Trans. Graph., vol. 35, no. 4, pp. 110:1–110:11, 2016.

[19] J. Zhao, L. Liu, C. Snoek, J. Han, and L. Shao, "Pixel-level semantics guided image colorization," in British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018, p. 156, 2018.

[20] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization," in Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV, pp. 577–593, 2016.

[21] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III, pp. 649–666, 2016.

[22] R. Zhang, J. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros, "Real-time user-guided image colorization with learned deep priors," ACM Trans. Graph., vol. 36, no. 4, pp. 119:1–119:11, 2017.

[23] S. Messaoud, D. A. Forsyth, and A. G. Schwing, "Structural consistency and controllability for diverse colorization," in Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VI, pp. 603–619, 2018.

[24] K. Hong, J. Li, W. Li, C. Yang, M. Zhang, Y. Wang, and Q. Liu, "Joint intensity-gradient guided generative modeling for colorization," arXiv preprint arXiv:2012.14130, 2020.

[25] J. Zhao, J. Han, L. Shao, and C. G. M. Snoek, "Pixelated semantic colorization," Int. J. Comput. Vis., vol. 128, no. 4, pp. 818–834, 2020.

[26] Y. Zhao, L. Po, K. Cheung, W. Y. Yu, and Y. A. U. Rehman, "SCGAN: saliency map-guided colorization with generative adversarial network," CoRR, vol. abs/2011.11377, 2020.

[27] T. Weyand, A. Araujo, B. Cao, and J. Sim, "Google landmarks dataset v2 - A large-scale benchmark for instance-level recognition and retrieval," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020, pp. 2572–2581, 2020.

[28] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, pp. 4873–4882, 2016.

[29] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pp. 2813–2821, 2017.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

[31] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, pp. 2672–2680, 2014.

[32] M. Mirza and S. Osindero, "Conditional generative adversarial nets," CoRR, vol. abs/1411.1784, 2014.

[33] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," CoRR, vol. abs/1701.07875, 2017.

[34] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA, pp. 4278–4284, 2017.

[35] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in Deep Learning in Medical Image Analysis - and - Multimodal Learning for Clinical Decision Support - 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings, pp. 3–11, 2018.

[36] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pp. 5967–5976, 2017.

[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, 2004.

[38] V. Pullano, A. Vanelli-Coralli, and G. E. Corazza, "PSNR evaluation and alignment recovery for mobile satellite video broadcasting," in 6th Advanced Satellite Multimedia Systems Conference and 12th Signal Processing for Space Communications Workshop, ASMS/SPSC 2012, Vigo, Spain, September 5-7, 2012, pp. 176–181, 2012.

[39] J. A. Bhutto, L. Tian, Q. Du, T. A. Soomro, Y. Lubin, and M. F. Tahir, "An enhanced image fusion algorithm by combined histogram equalization and fast gray level grouping using multi-scale decomposition and gray-pca," IEEE Access, vol. 8, pp. 157005–157021, 2020.

[40] T. Lianfang, J. Ahmed, D. Qiliang, B. Shankar, and S. Adnan, "Multi focus image fusion using combined median and average filter based hybrid stationary wavelet transform and principal component analysis," International Journal of Advanced Computer Science and Applications, vol. 9, no. 6, pp. 34–41, 2018.

[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3055575, IEEE Access

KANGNING DU received his BSc degree in Telecommunication Engineering from Beijing Information Science and Technology University in 2011. He received PhD in Communication and Information System from Institute of Electronics, Chinese Academy of Sciences in 2016. Currently, he is a teacher of Electronic Engineering at Beijing Information Science and Technology University. His research interests include radar signal processing and image understanding and recognition.

CHANGTONG LIU received the BSc degree from Liaocheng University in 2017, and is currently working toward the MSc degree at Beijing Information Science and Technology University. His research interests include machine learning and computer vision.

LIN CAO received a B.Eng. degree in Telecommunication Engineering from Northeastern University China in 1999. He received Ph.D. in Signal and Information Processing from Institute of Electronics, Chinese Academy of Sciences in 2005. Currently, he is a professor at the Department of Electronic Engineering at Beijing Information Science and Technology University(BISTU). He teaches courses on digital signal processing, digital image processing, and soft design fundamentals. He is the dean of the School of Information and Communication Engineering and the deputy director of the Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument at BISTU. He is a member of China Education Society of Electronics. He has published over forty papers on image processing and pattern recognition. His research interests include radar signal processing and image understanding and recognition. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

YANAN GUO received the BSc degree from Hubei Polytechnic University in 2014, and the PhD degree from Yunnan University in 2019. Currently, she is a teacher of Electronic Engineering at Beijing Information Science and Technology University. Her research interests include machine learning and computer vision.

FAN ZHANG received the Ph.D. degree from Beijing University of Posts and Telecommunications in 2019. Currently, she is a teacher of Electronic Engineering at Beijing Information Science and Technology University. Her research interests include machine learning and computer vision.

TAO WANG received the Ph.D. degree in the School of Electronic and Information Engineering from Beihang University in 2019. Currently, he is a teacher of Electronic Engineering at Beijing Information Science and Technology University. His research interests include radar signal processing, data fusion, and target localization and tracking in intelligent transportation system or vehicle intelligent assistance system.