Colorization of Characters Based on the Generative Adversarial Network

Changtong Liu, Lin Cao, and Kangning Du(✉)

Beijing Information Science and Technology University, No. 35, North Fourth Ring Road, Beijing, China
kangningdu@outlook.com

Abstract. With the development of economy, global demand for electricity is increasing, and the requirements for the stability of the power grid are correspondingly improved. The intelligence of the power grid is an inevitable choice for the research and development of power systems. Aiming at the security of the smart grid operating environment, this paper proposes a gray-scale image coloring method based on generating anti-network, which is used for intelligent monitoring of network equipment at night, and realizes efficient monitoring of people and environment in different scenarios. Based on the original Generative Adversarial Network, the method uses the Residual Net improved network to improve the integrity of the generated image information, and adds the least squares loss to the generative network to narrow the distance between the sample and the decision boundary. Through the comparison experiments in the self-built CASIA-Plus-Colors high-quality character dataset, it is verified that the proposed method has better performance in colorization of different background images.

Keywords: Image colorization · Intelligent electric grid · Generative Adversarial Network

1 Introduction

The construction of the global smart grid is developing rapidly, and its protection devices are distributed in different areas and play an extremely important role in the safe and stable operation of the power grid [1–5]. However, the protection devices of the current power grid lack intelligent monitoring of the operating environment of the equipment, especially the images taken by optical, infrared and other monitoring devices at night are mostly gray-scale images. Aiming at the above problems, this paper proposes a gray image coloring method based on generating anti-network, which achieves accurate coloring of characters and background in the image and improves the use value of the image. The application model of the colorization system is shown in Fig. 1. Especially at night, the gray-scale image of the device environment is collected in real time by the image sensor and transmitted to the smart grid control center. The collected gray-scale image is judged by the control center, and the gray-scale image with the suspicious problem is selected. Then it is uploaded to the gray-scale image colorization subsystem

X. Zhang et al. (Eds.): CloudComp 2019/SmartGift 2019, LNICST 322, pp. 636–649, 2020.
https://doi.org/10.1007/978-3-030-48513-9_51

to colorize the image with suspicious problems and transmit the colorized image back to the smart grid control center. Relative to the gray-scale image, the color image contains more valuable information, and the control center makes an accurate judgment for the color image. In recent years, in the field of image, the application of digital processing technology and deep learning methods have been gradually deepened [6, 7].

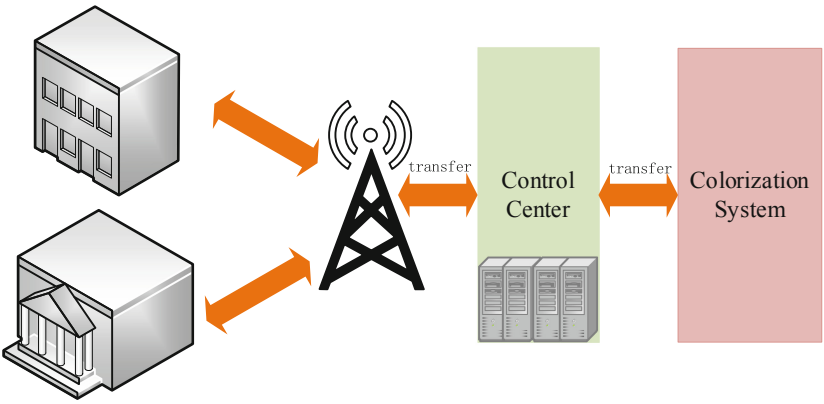


Fig. 1. Color system application model

Traditional image coloring methods mainly include local color expansion of methods [8, 9] and color transfer of methods [10, 11]. The traditional image colorization method requires a large number of artificial work or reference images with similar scenes, so it is greatly influenced by artificial factors and the coloring complexity is high.

In order to reduce the influence of artificial factors in the traditional image coloring method, the method based on deep learning [12–15] has been gradually proposed. Iizuka et al. [12] used a two-channel network to combine the local features of the image with global prior information to achieve automatic coloration of gray-scale images. Larsson et al. [13] used the VGG network [16] to extract image features and predict the color distribution of each pixel. Zhang et al. [14, 15] proposed a gray image coloring method for classifying pixel points and based on user guidance.

In recent years, the Generative Adversarial Network (GAN) [17] has achieved great success in the field of image generation. Compared with the traditional neural network [18], GAN generates higher image quality. However, GAN's training is unstable and prone to mode collapse. Therefore, Zhu et al. [19] proposed a Cycle Generative Adversarial Network (Cycle-GAN) based on Isola [20] to improve the stability of the training network.

In summary, for the security problem of power grid equipment, this paper proposes a method of portrait intelligent coloring, which is mainly used for colorization of grayscale images in nighttime environment, which can enable observers to obtain more information from color images and improve the ability to monitor the environment. The generation network is switched to the U-Net network [21] to improve the image generation details and to add the least squares loss as an optimization target for image coloring to improve the stability of the network. In the discriminant network, the image features are extracted

by means of multi-layer feature fusion, so that the extracted features more represent image details. Finally, the model is trained with improved consistency loss. In this paper, experiments are carried out on the self-built CASIA-Plus-Colors dataset. The results verify the effectiveness of the proposed method.

2 Approach

2.1 Color Model

In the color space, the RGB model based on color combination does not adapt to the human eye color, can only compare the visual characteristics of brightness and color temperature, and cannot directly reflect the intensity of the illumination information in the image. Therefore, this paper transforms an image from an RGB color model to a Lab color model based on the human eye's perception of color.

The coloring process is to map the component x of L channel to the chroma channels a and b through a network model. The output of the network model and the L channel are recombined into a new three-channel image, and the resulting colorized image is $\tilde{X} = (X_L, \tilde{X}_a, \tilde{X}_b)$. The ultimate goal of training the color model is to obtain an optimal mapping relationship of $X_L \rightarrow \tilde{X}_{ab}$.

2.2 Network Structure

The traditional GAN is one-way generation, using a single loss function as a global optimization target, which may map multiple samples to the same distribution, resulting in a pattern crash. CycleGAN uses a loop to generate confrontation, effectively avoiding this deficiency of traditional GAN. This paper proposes a colorization model based on CycleGAN, which improves the traditional GAN reverse-transfer optimization network. It uses the consistency loss of reconstructed data to reverse the whole network and strengthens the stability of the original network. At the same time, in order to improve the integrity of the generated image information, the residual network is used to improve the original generation network; and the multi-feature fusion method is introduced into the discriminant network, so that the extracted features more represent the details of the image.

Color Network Model. The colorization model in this paper consists of four subnetworks which are respectively the generation network G: responsible for mapping the L-channel gray-scale image to the ab-channel color component, generating the network F: responsible for mapping the ab-channel color component to the L-channel gray-scale image, discriminating the network: used to discriminate between the L-channel grayscale image and the gray-scale image generated by the generated network, and the discrimination network: used to distinguishing between the real color image and the composite image of the generated ab-channel component and L-channel gray-scale image. The goal of the network is to obtain an optimal correspondence by training the L channel component and the ab channel color component, and mapping the L channel gray-scale image to the ab channel color component. The network structure is shown in Fig. 2.

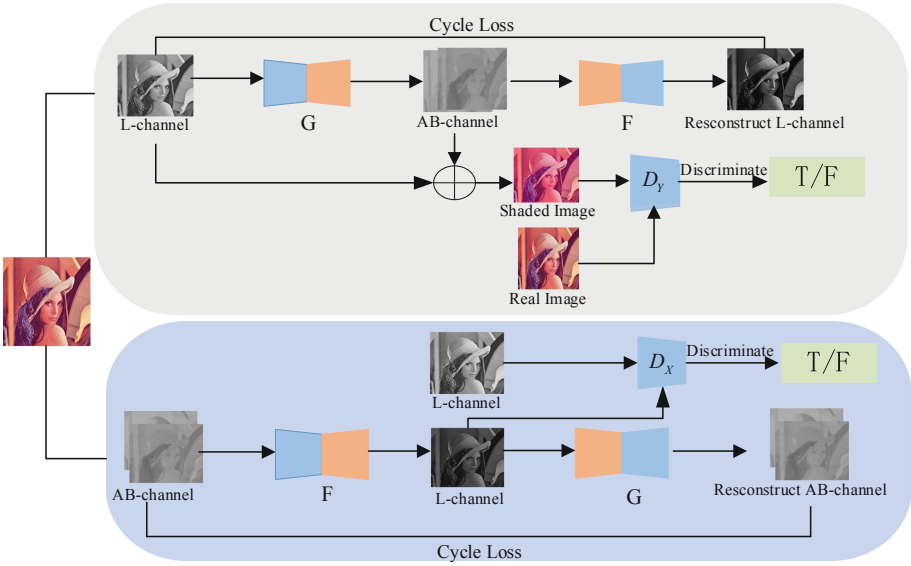


Fig. 2. Generative adversarial network structure

The above sub-networks constitute a pair of cyclic generation networks that respectively map the input samples to the intermediate domain and then reconstruct the data of the intermediate domains back into the original domain. For example, an L-channel gray-scale image is input, which is eventually mapped back to the gray-scale image, and the intermediate domain data is the generated ab-channel color component. Similarly, when the input is an ab-channel color component, it will eventually be reconstructed back into the original domain. The middle domain is the gray-scale image generated by the F network.

In the original CycleGAN, the two cyclic generative networks are independent of each other. When the two cyclic processes are completed, the consistency loss of the cyclic generative network is calculated separately. The colorization model of this paper combines the data reconstructed by two loop generation networks which recombines the output ab-channel color component and L channel gray-scale image to obtain a reconstructed color image. Then, the L1 distance between the reconstructed color image and the real color image is calculated as the cyclic consistency loss of the network, and the reverse transmission optimization of the entire network is realized together.

Generative Network Model. In the traditional GAN, the generative network consists of only a simple convolution layer and a deconvolution layer. When extracting features, it is easy to lose local information of the image and limit the coloring effect of the network. As shown in Fig. 3, in order to avoid the above problem, the generated network uses residual network to connect the features outputted by each layer in the underling layer to the corresponding upper layer through a skip connection. The purpose of adding skip connections is to pass shallow information directly to the deconvolution layer of the same height, forming thicker features and improving image generation details.

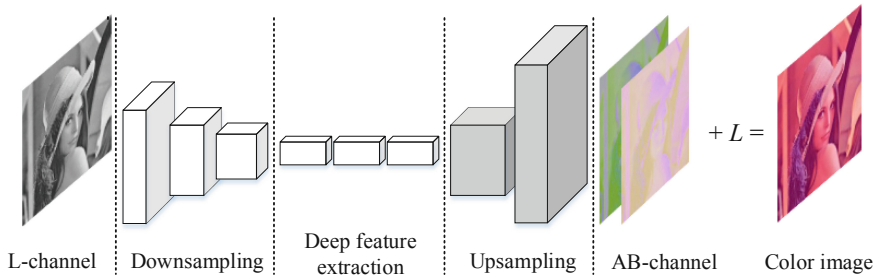


Fig. 3. Generative network structure

The generation network is composed of two parts: upper and underling layer. The underling section has 4 layers, and the number of filters is [64, 128, 256, 512]. As can be seen in Fig. 3, during the underlying process, each layer of the image features is convolved twice, and the filter size is 3×3, which of purpose is to extract basic information such as image texture structure. After the convolutional layer, the Batch Normalization (BN) layer [25] is designed to adjust the data distribution after convolution so that the output of the convolution is distributed in the near-origin field of the activation function, reducing the gradient dispersion rate and avoiding the gradient disappearing. The active layer uses a Leaky Rectified Linear Unit (LRELU) instead of the original Rectified Linear Unit (RELU) [19], which aims to reduce the computational complexity and does not lead to negative neurons in the value area are all zero. The purpose of generating the network is to map the input to the distribution of the target domain space, for example, according to the lip shape feature corresponding to the reddish process.

Discriminant Network Model. The traditional discriminant network uses a single layer feature to express the entire image, and it is easy to lose the details of an image. Therefore, this paper introduces the multi-feature fusion method of the discriminant network (see Fig. 4). The use of the fused features enhances the expression of image detail and improves the accuracy of image recognition. At the same time, in order to avoid dimensional disaster, coding network is added behind the fusion layer to reduce dimensions of feature.

The generated component $\tilde{X}_{\text{ab}}$ is combined with the L-channel component to form a colored image, and the discriminating network D_Y distinguishes between it and the real color image. Because of the correlation between the two, the discriminator can learn through the convolutional neural network to obtain more effective image features and correctly classify the two types of images. For discriminating the network, first input a three-channel 256×256 color image, and then after 6 convolutions with strides, finally output 256 feature maps which of size is 5×5. The convolution kernel size of the feature extraction is 5×5, the convolution step is 2, and the number of feature maps of each convolution layer respectively are 8, 16, 32, 64, 128, 256. After the convolution, the features are fused, and the fourth layer and the fifth layer are subjected to the average pooling layer of different size to generate 448 feature maps. Then feature map is stretched to a vector of 11264-dimensional length, and the feature's dimension is reduced to

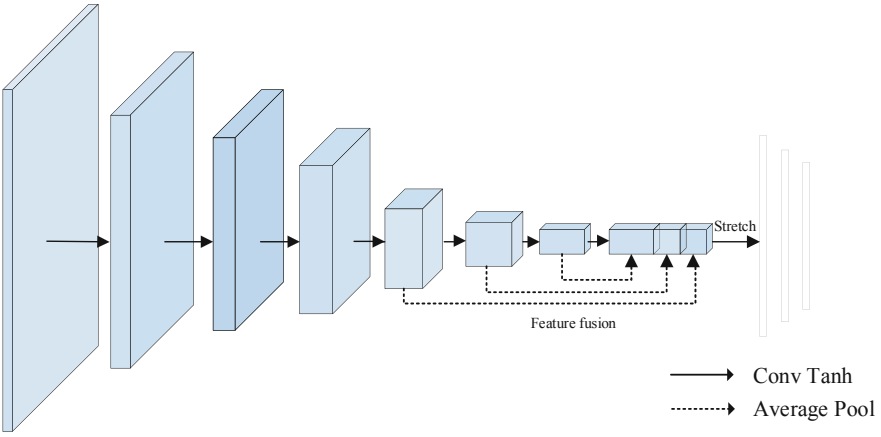


Fig. 4. Adversarial network model

1024 dimensions using a multi-layer full connection. In order to further prevent over-fitting during feature dimension reduction, the Dropout layer is added after the fully connected layer, and the probability value is set to 0.7. Finally, the feature vector is input to the Sigmoid function [26] to determine whether the generated image conforms to the distribution of the real image. For the other discrimination network, the input image is a single-channel gray-scale image, and the model structure is the same as the discrimination network D_Y.

2.3 Loss Function

The traditional GAN only uses the generative adversarial loss function, and there is redundant mapping space. This paper combines the traditional generative adversarial loss and the cyclic consistency loss to supervise the training network jointly, effectively avoiding this problem.

Generative Adversarial Network. The cyclic generation applies a process of mapping the input image to the intermediate domain image against the resistive loss in the network. The original cross entropy loss is as shown in Eq. (1), which makes it impossible for the generator to further optimize the generated image recognized by the discriminator which may result in a low quality of the network generated image. Inspired by Mao [22], the paper uses the least squares loss instead of the traditional generative adversarial loss. Comparing with the original loss function, the least squares loss will process the generated samples away from the decision boundary and the decision is true, and the generated samples far from the decision boundary will be placed near the decision boundary. According to the different distance from the decision boundary, a stable and high-quality network is constructed.

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim P_{data}(y)}[\log D_Y(y)] + E_{x \sim P_{data}(x)}\{\log\{1 - D_Y[G(x)]\}\} \tag{1}$$

Where $x \sim P_{data}(x)$ and $y \sim P_{data}(y)$ are the probability distributions of the samples X and Y respectively. $E_{x \sim Pata(x)}$ and $E_{y \sim Pata(y)}$ are the expected values for the respective sample distributions.

Therefore, for the generative network $G : X \rightarrow Y$ and its adversarial network D_Y, the least squares loss function definition herein is as shown in Eq. (2).

$$\begin{aligned} \min_{D_Y} L_{LSGAN}(D_Y) &= \frac{1}{2} E_{y \sim P_{data}(y)}[(D_Y(y) - 1)^2] + \frac{1}{2} E_{x \sim P_{data}(x)}[(D_Y(G(x)))^2] \\ \min_{G} L_{LSGAN}(G) &= \frac{1}{2} E_{x \sim P_{data}(x)}[(D_Y(G(x)) - 1)^2] \end{aligned} \tag{2}$$

The generating network G generates the X domain data to match the target of the Y domain distribution, and discriminant network D distinguishes the real Y domain data $\{y\}$ and the generated sample $\{G(x)\}$.

Training the discriminator, the loss function goal is to make the discriminator distinguish between the real sample and the generated sample, maximizing $D_Y(y)$ and minimizing $D_Y(G(x))$. Training generator, the goal of the loss function is to make the generated data close to the real data, maximizing $D_Y(G(x))$. The target of the loss is shown in Eq. (3).

$$G^* = \arg \min_{G, D_Y} L_{GAN}(G, D_Y, X, Y) \tag{3}$$

For the generation network $F : Y \rightarrow X$ and the corresponding discriminant network, the same generative adversarial loss is also introduced, and the loss target is as shown in Eq. (4).

$$F^* = \arg \min_{F, D_X} L_{GAN}(F, D_X, Y, X) \tag{4}$$

Cyclic Consistency Loss. The traditional GAN only uses the adversarial loss training network to learn the relationship of the input image mapping to the target image, but it cannot solve the redundant mapping problem existing in the generated network. The cyclic generation network uses the loop consistency loss to better ensure the stability of the generated data and reduce other redundant mapping relationships.

Based on this idea, this paper proposes to recombine the reconstructed data and calculate its L1 loss with the input color image as the network's cyclic consistency loss.

In Fig. 2, the samples of the L-channel gray-scale image and the ab-channel color component are x_{ab} and x_L.

$x_{ab} \rightarrow G(x_{ab}) \rightarrow F(G(x_{ab})) \approx \hat{x}_{ab}$ and $x_L \rightarrow F(x_L) \rightarrow G(F(x_L)) \approx \hat{x}_L$ in the network are respectively two reconstruction processes. This paper separately calculates the consistency loss of reconstruction data. The equation is as follows:

$$L_{cyc}(G, F) = E_{x \sim P_{data}(x)}[||F(G(x) - \mathrm{x}||_1] + E_{y \sim P_{data}(y)}[||G(F(y)) - y||_1] \tag{5}$$

The complete objective function includes adversarial loss and cyclic consistency loss, as shown in Eq. (6):

$$L(G, F, D_X, D_Y) = L_{LSGAN}(G, D_Y, X, Y) + L_{LSGAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \tag{6}$$

Among them, the x parameter is used to adjust the weight of the least square loss and the cycle consistency loss.

We aim to solve:

$$G^*, F^* = \arg\min_{G,F,D_X,D_Y} L(G, F, D_X, D_Y) \tag{7}$$

3 Experimental Data

There are many publicly available Face Databases, such as Public Figures Face Database and Celebrity Faces Attributes Database, which are mainly used in face recognition and other fields. The portrait images of these characters are mostly concentrated in the face area of the person, and the image quality is different. The effect of training with these pictures is not good. In order to solve the problem of the database, based on the CASIA-FaceV5 database, the data set is expanded by crawling technology. The final database contains a total of 9,500 color images of various poses and backgrounds, referred to as CASIA-PC (CASIA-Plus-Color).

CASIA-FaceV5 is a database published by the Chinese Academy of Sciences. The database is an open Asian portrait dataset containing 2,500 high-quality color images from 500 people. Through observation, it is found that most of the characters in the database are frontal photos in a monochrome background, lacking scenes in the actual environment.

(1) (2) (3) (4)

Fig. 5. This is an example of the self-built data set in this article. (1) From CASIA (2)–(4) From WEB

In order to solve the problem of lack of the real scene portrait in CASIA-FaceV5 database, based on the database, this paper uses the crawler technology to complete the task of automatically and modularly collecting 7000 colored portraits.

In this paper, the CASIA-PC dataset was used. The size of all the images was adjusted to 225 × 225 pixels, and the database was divided into training set and test set. The training set consisted of 8600 randomly selected images; the remaining images were used as test sets. Examples of the database are shown in Fig. 5, wherein the first column is picked from the CASIA database and portraits of the second to the fourth columns are picked from collected portraits. As can be seen in Fig. 5, the scenes of the self-built dataset are rich and the color is colorful, which increases the difficulty of coloring.

In order to objectively evaluate the quality of the generated image, Structural Similarity Index (SSIM) [24] and Peak Signal to Noise Ratio (PSNR) are used to evaluate the image.

4 Experiment and Analysis

4.1 Experimental Procedure

Experimental Pretreatment. In the pre-processing stage of the experiment, the color model of each image is transferred from RGB to Lab model, and the L-channel and the ab-channel component of the colorful portraits are separated as the input of the network.

Setting Parameters. In the data training process, both the generator and the discriminator use the Adam optimizer with an initial learning rate of 0.0002 and a momentum of 0.5 to update the parameters of the network, and the linear attenuation method is used to reduce the learning rate gradually. After continuous iterative training until the model converges, storing the parameters of the entire network.

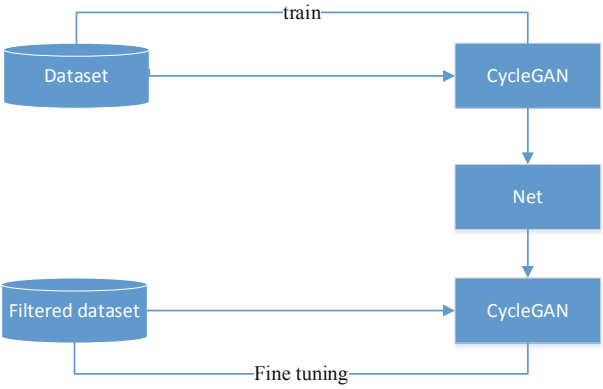


Fig. 6. Model training process

Experiment Procedure. The experimental process is shown in Fig. 6 and can be divided into two phases. In the first stage, 8600 training samples are used to train the entire network to obtain a colorization model. In order to avoid over-fitting of the network, when the network is trained with a large data set, due to the different quality of images,

some will have color distortion and image blurring, which will affect the coloring effect of the model. Therefore, in the second stage of the experiment, some training samples with relatively high quality were selected in the larger original data set, and the parameters of the network were fine-tuned. The first stage of the experiment is to make the generative network G learn the correspondence from the gray image mapping to the ab-channel color component, and the second stage is to improve the accuracy of the model coloring.

4.2 Experimental Results and Analysis

The coloring results of different methods are shown in Fig. 7. The first column is the gray image of the L-channel; the second column is the origin color image; the third column is the coloring result of the original CycleGAN; the fourth column is based on the third column method, which of generative structure is changed to the residual network; the last column is the coloring result of the method in this article.

Fig. 7. Comparison of different methods of coloring effects. (1) Gray Scale (2) Original image (3) CycleGAN (5) CycleGAN + Residual (5) Method of this paper

According to the results of different colorization model, when using the original CycleGAN model for coloring, the effect is rough, the color saturation and coloring accuracy are low, and problems such as coloring wrong or overflow may occur. For example, in the third line of Fig. 7, the leaves in the original image are mistakenly colored into other colors. And in the fifth line, the color belonging to the face area is beyond its own range and spreads to areas such as trees and sky. The method of generative

network using the residual network, the coloring result of the model, as shown in the second line, greatly improves the coloring accuracy of a single image. Although under complex background images, the effect of coloring has been improved. There is still the problem of coloring wrong, which is more obvious in its third line. In contrast, the coloring results of the method of this paper are more accurate and natural, and even in images with complex backgrounds, the true colors of portraits and backgrounds can be more accurately given. And it can correctly distinguish different targets in the image and reduce the phenomenon of color overflow, as shown in lines 3 and 4 of the paper's method. In addition, the coloring result of the first line is worth noting. The change in color of the clothing happened. This is because the database lacks the same style costume, or the similar clothing is mostly gray-black. This shows that the training set has a great influence on the coloring result.

This paper compares the PSNR and SSIM average indicators of the three models in a single color and a complex background. As shown in Tables 1 and 2, under the objective index evaluation, with the enrichment of the network structure of three models, the coloring effect in the monochrome background and the complex background is improved in turn. In addition, the colorization of the monochrome background image is relatively easy, because its structure and texture are relatively simple, and the average index of the image in the monochrome background of the same model is significantly higher than that of the image in the complex background.

Table 1. Comparison of average SSIM and PSNR indicators of different methods in complex background

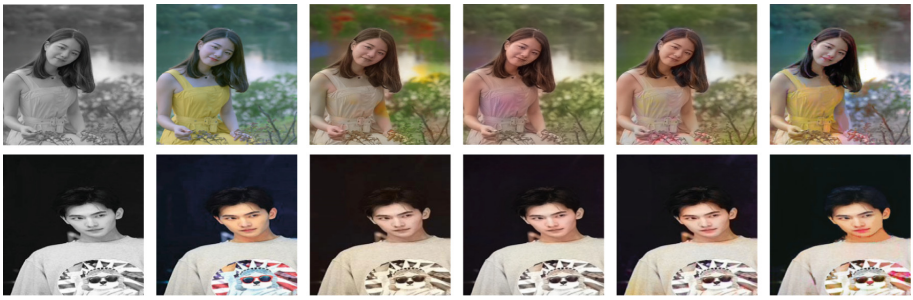
Network model	SSIM/%	PSNR/dB
CycleGAN	95.4718	30.0658
CycleGAN + Residual	96.7481	34.1508
Method of this paper	97.2621	38.4907

Table 2. Comparison of average SSIM and PSNR indicators of different methods in monochrome background

Network model	SSIM/%	PSNR/dB
CycleGAN	96.5804	35.6529
CycleGAN + Residual	97.3104	38.0507
Method of this paper	98.2316	38.6084

In addition, this paper is compared with other coloring models, and the results are shown in Fig. 8. Iizuka [12] uses a two-channel convolution network, and the colorization results are brighter in color, but the coloring accuracy is low. Larsson [13] uses the VGG

network to extract image features. The coloring model solves the problem in coloring wrong, but the partial area of the portrait becomes blurred. Zhang [14] classifies each pixel in the image which has high coloring accuracy and clear characters, but low color saturation. When using the method of the paper, the accuracy of colorization is high, the discrimination of different targets is higher, and the color is more natural. However, when coloring the same target, there are problems such as uneven color distribution and low saturation. This paper compares the SSIM and PSNR indexes of the proposed method in different background with other color models, as shown in Tables 3 and 4, respectively. In different scenarios, the image using the proposed method has a higher SSIM and PSNR values than other methods, indicating that the results of the method are more similar to the original image, and the distortion is smaller.



(1) (2) (3) (4) (5) (6)

Fig. 8. Comparison of colorization results of different models (1) Gray Scale (2) Original image (3) Larsson (4) Iizuka (5) Zhang (6) Method of this paper

Table 3. Comparison of average SSIM and PSNR indicators of different models in monochrome background

Network model	SSIM/%	PSNR/dB
Iizuka	95.4205	34.6785
Larsson	97.3620	34.6668
Zhang	98.8255	36.9591
Method of this paper	99.0219	38.4917

5 Conclusion

In order to achieve effective monitoring of operation of smart grid on night, this paper proposes a coloring method based on Generative Adversarial Network. The method adopts an improved consistency loss, calculates the L1 loss of the color image and generative image, and achieves the reverse transmission optimization of the entire network.

The generative network uses the Residual Network to extract image features to maintain the stability of the information; the discriminant network extracts image features by means of multi-feature fusion. The experiment proves that compared with the similar methods, the model of this paper is suitable for character coloring in complex scenes, and can complete the task of protecting grid equipment at night effectively.

References

1. Zhang, H.Y., Li, S.D.: Design of adaptive line protection under smart grid. In: 2011 International Conference on Advanced Power System Automation and Protection, vol. 1, pp. 599–603. IEEE (2011)
2. Jing, S., Huang, Q., Wu, J., et al.: A novel whole-view test approach for onsite commissioning in smart substation. IEEE Trans. Power Delivery **28**(3), 1715–1722 (2013)
3. Lin, H., Wang, C., Lei, W.: The design of overvoltage testing system for communication equipment in smart grid. In: 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), pp. 1812–1816. IEEE (2016)
4. Halim, H.A., Amirruddin, M., Noorpi, N.S., et al.: An improved protection scheme for smart distribution grid. In: 2013 1st International Conference on Artificial Intelligence, Modelling and Simulation, pp. 337–341. IEEE (2013)
5. Long, H., Xiang, W., Zhang, Y., et al.: Secrecy capacity enhancement with distributed precoding in multirelay wiretap systems. IEEE Trans. Inf. Forensics Secur. **8**(1), 229–238 (2012)
6. Xiang, W., Wang, G., Pickering, M., et al.: Big video data for light-field-based 3D telemedicine. IEEE Network **30**(3), 30–38 (2016)
7. Xiang, W., Barbulescu, S.A., Pietrobon, S.S.: Unequal error protection applied to JPEG image transmission using turbo codes. In: Proceedings 2001 IEEE Information Theory Workshop (Cat. No. 01EX494), pp. 64–66. IEEE (2001)
8. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. ACM Trans. Graph. (TOG) **23**(3), 689–694 (2004)
9. Heo, Y.S., Jung, H.Y.: Probabilistic Gaussian similarity-based local color transfer. Electron. Lett. **52**(13), 1120–1122 (2016)
10. Xiao, Y., Wan, L., Leung, C.S., et al.: Example-based color transfer for gradient meshes. IEEE Trans. Multimedia **15**(3), 549–560 (2012)
11. Qian, Y., Liao, D., Zhou, J.: Manifold alignment based color transfer for multiview image stitching. In: 2013 IEEE International Conference on Image Processing, pp. 1341–1345. IEEE (2013)
12. Iizuka, S., Simo-Serra, E., Ishikawa, H.: 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. (TOG) **35**(4), 110 (2016)
13. Larsson, G., Maire, M., Shakhnarovich, G.: Learning representations for automatic colorization. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 577–593. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_35
14. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 649–666. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_40
15. Zhang, R., Zhu, J.Y., Isola, P., et al.: Real-time user-guided image colorization with learned deep priors. arXiv preprint arXiv:1705.02999 (2017)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

17. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
18. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
19. Zhu, J.Y., Park, T., Isola, P., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
20. Isola, P., Zhu, J.Y., Zhou, T., et al.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
22. Mao, X., Li, Q., Xie, H., et al.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794–2802 (2017)
23. Glorot, X., Bordes, A,. Bengio, Y.: Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pp. 315–323 (2011)
24. Wang, Z., Bovik, A.C., Sheikh, H.R., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
25. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
26. Li, C.H., Lee, C.K.: Minimum cross entropy thresholding. Pattern Recogn. **26**(4), 617–625 (1993)