ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0018

Gray-scale Image Colorization based on Conditional Deep Convolution Generation Adversarial Network

Tingting Gui, Xing Gu, Yanli Shi

Changchun University of Science and Technology National Demonstration Center for Experimental Electrical, Changchun, Jilin, 130022, China.

Abstract

Current gray-scale image colorization methods generally have problems such as border blurring, loss of detail, and boring coloring. In response to the above problems, this paper proposes an improved conditional depth convolution to generate a gray-scale image colorization method against the network. The network model fusion generates a confrontation network structure, builds a discriminant network, introduces a deep aggregation structure network into the field of image colorization, and adds long connections on the basis of the traditional network, which improves the utilization of features while alleviating the problem of gradient disappearance, thereby improving The algorithm model's ability to process image boundaries and details, and dynamically evaluate the color quality of the image, alleviate the problems of boundary blurring, loss of details and boring coloring. Experimental results show that this method has certain advantages in improving the image quality of grayscale images after colorization.

Keywords

Colorization; Generative Adversarial Network; Deep Layer Aggregation; Skip Connection.

1. Introduction

With the development of Internet technology and digital image processing technology, visual media such as images and videos have gradually replaced text information as an important source for people to obtain information. In images and videos, color is an important feature descriptor. It plays a pivotal role when expressing or transmitting information. In addition, the human visual system can distinguish thousands of tones and intensities, but can only distinguish dozens of grayscales. Compared with grayscale images, the human eye is more sensitive to color information. Colored images can better reflect the characteristics of things, enhance the expression effect of images, and give users a better visual experience. Therefore, the colorization of grayscale images is an important topic in the field of digital image processing.

Traditional colorization methods need to have two elements: reference color information and a colorization method based on these color information. The traditional colorization method not only requires manual intervention by people, but the coloring effect often has obvious color overflow, color confusion and other problems, and the results are unsatisfactory. In recent years, with the rapid development and wide application of deep learning, some scholars combine deep learning and image coloring, based on the powerful feature extraction and learning capabilities of convolutional neural networks [1], so as to realize the gray-scale image Automatic coloring. At the same time, due to the good dynamic color processing effect of the generative adversarial network, it occupies a very important position in the field of image coloring. Compared with other methods, the deep convolution condition generation method for adversarial network coloring proposed in this paper fused with the deep aggregation structure restores more image details, richer image colors, and effectively alleviates the boundary blooming and blurring caused by the colorization process. Problems such as loss of detail and boring coloring.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0018

2. Related Theoretical Basis

2.1 CIELab Color Space

Lab color space is based on people's perception of colors, and uses a digital way to express people's visual perception. Its numerical value can describe all the colors that the human eye can see. In the Lab color space, the brightness channel and the color channel are separated. L represents the brightness channel, the value range is [0, 100], which represents the brightness change from black to white. a and b respectively represent two color channels.

The purpose of image coloring is to learn the mapping function from 1 channel to 3 channel images. In this case, the CIELab color space shows its advantages. In the Lab color space, the luminance channel L can be regarded as a grayscale image, which can be directly used for neural network training. The mapping function only needs to learn to generate two channels a and b, and can use the original luminance channel to generate color images. By using the original brightness channel, most of the features of the input image can be retained in the output, making the input and output closer visually.

2.2 Deep Layer Aggregation

Yu et al. proposed a Deep Layer Aggregation (DLA) [2], which connects the low-level network with the high-level network through a jump connection, so as to fuse information of different scales and resolutions, and learn to cross different feature hierarchies. The combination of the network itself is trained to learn the importance of different depth features, to solve the effect of the size of the receptive field on the feature extraction of the object in the image feature extraction, so as to enrich the detailed features of the large and small objects in the color image. Figure 1 shows the network structure of the deep aggregation structure.

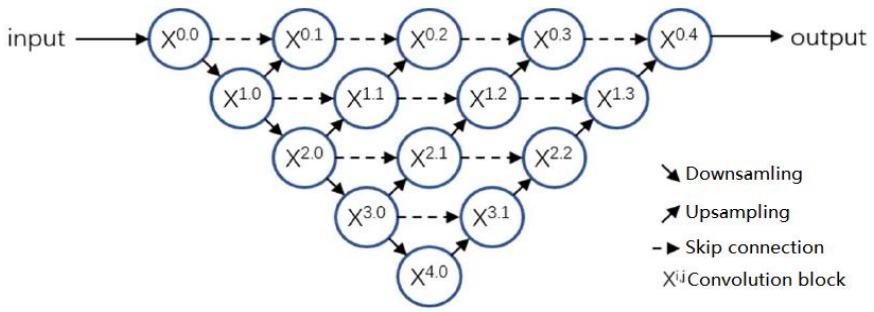


Figure 1. Network structure of deep aggregation structure

2.3 Generative Adversarial Network

Generative Adversarial Nets (GAN) [3] adopts the idea of a generator and arbiter game with each other, using the result of the decision network as the loss function of the generation network, that is, the loss function of the generation network is changed from manual designation to It is dynamically learned by the decision network, so GAN has a good performance in the field of image generation. The structure of the generated adversarial network is shown in Figure 2.

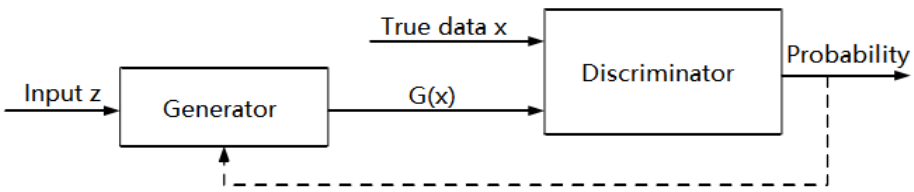


Figure 2. Generative adversarial network structure

ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0018

Radford et al. [4] combined Convolutional Neural Network (CNN) and GAN, and proposed Deep Convolutional GAN (DCGAN), which uses the advantages of CNN in image processing to improve GAN comparison. The ability to generate complex images. Mirza et al. [5] proposed Conditional Generative Adversarial Nets (C-GAN). Constraints are added to GAN, and conditional information is used to guide data generation to solve the problem of traditional GAN in the image generation process. Problems that are too free and difficult to control.

In the field of colorization of grayscale images, the generation network in the generation confrontation network can generate color images based on the input target grayscale image, and the discrimination network is used to distinguish between the generated color image and the real color image, so as to dynamically evaluate the colorization effect.

3. Algorithm in this paper

3.1 Improved Deep Aggregation Structure Network

This paper proposes an improved deep aggregation network structure, which uses a combination of long connections and short connections to complete the feature fusion and transmission of neural networks. The network structure is shown in Figure 3.

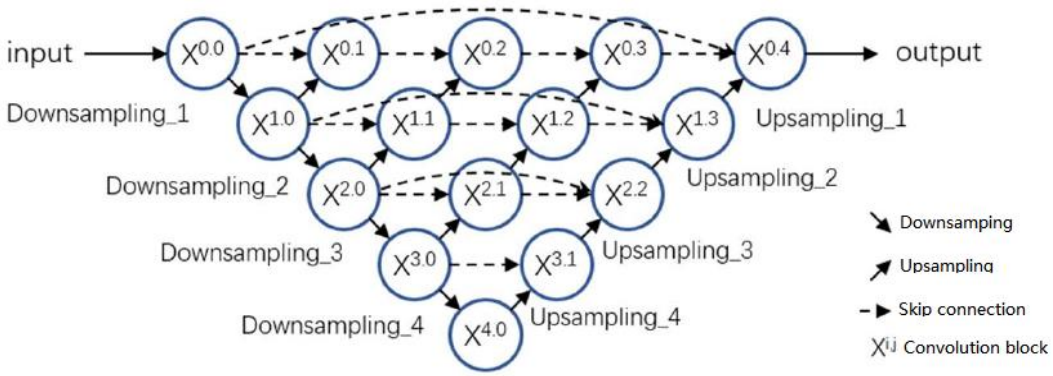


Figure 3. Improved deep aggregation network structure

This model uses a symmetrical structure of decoding and encoding. The down-sampling stage extracts the features of each scale of the grayscale image layer by layer, and the up-sampling stage reconstructs the image based on the extracted features to complete the image coloring. In the model, down-sampling extracts features of different levels, up-sampling is performed after each down-sampling, and the up-sampling result is spliced and fused with the previous feature maps of the same resolution, and the same is spliced on the last up-sampling node. The result of the first upsampling of the scale. Such a structure helps to restore the loss of detail due to downsampling and improve the problem of gradient dispersion.

Different from the traditional deep aggregation structure, this article uses step-size convolution to achieve down-sampling and step-size transposed convolution to complete up-sampling. Adopting convolution with step size and deconvolution with step size can preserve the position information of the image, which is beneficial to the colorization task.

Each convolution block in this article is composed of three convolution layers stacked, and the size of the convolution kernel of each convolution layer is 3X3, and n represents the number of convolution kernels. As the number of downsampling increases, the number of convolution kernels used doubles. This article performs four downsampling, so n is 64, 128, 256, 512, 512, respectively. At the same time, in order to prevent over-fitting problems, each convolutional layer contains batch normalization and ReLU activation functions. The structure of each convolution block in the model is shown in Figure 4.

Volume 7 Issue 9, 2021
DOI: 10.6919/ICJE.202109_7(9).0018

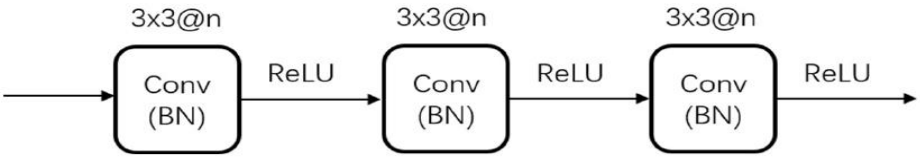


Figure 4. The structure of the convolution block

3.2 Conditional Deep Convolution Generative Adversarial Network

In this paper, the conditional depth convolutional generation confrontation network (C-DCGAN) is used to complete the colorization of grayscale images, and the convolutional neural network and conditional information are introduced into the traditional GAN at the same time. In this paper, the information of the L channel of the grayscale image to be colored is used as the conditional information input to the generation network to guide the color image generation. Both the generation network and the discrimination network use deep convolutional neural networks, which are good at mining image features and high robustness. The advantage of realizing image generation and discrimination. Among them, the generation network adopts the gray-scale image colorization model of the above-mentioned improved deep aggregation structure network. The detailed structure diagram is shown in Figure 5.

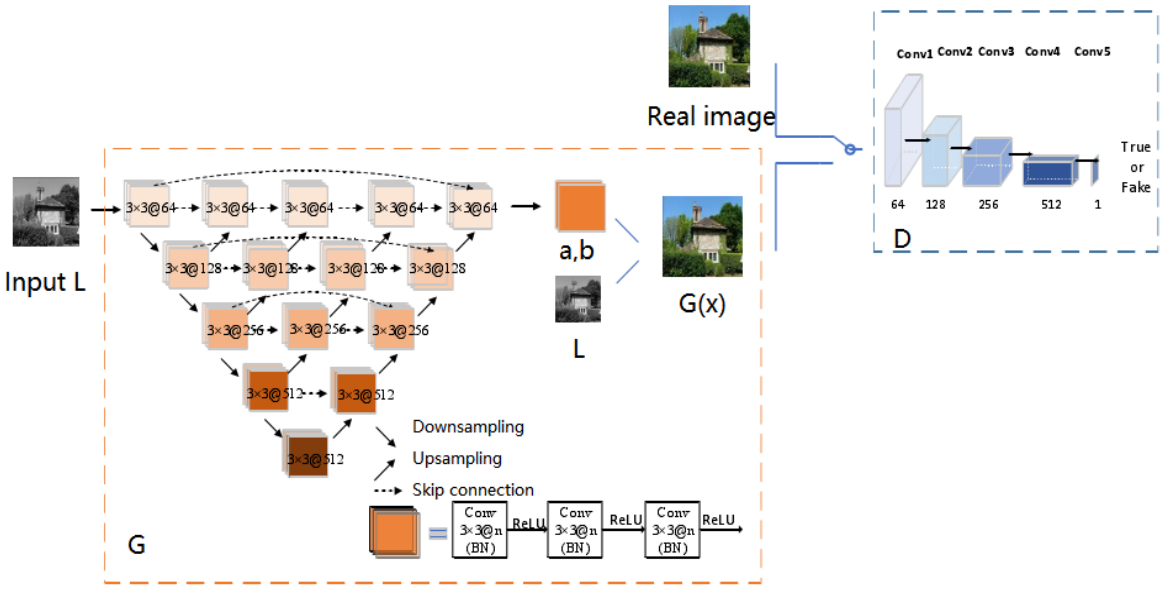


Figure 5. The overall network structure of the color model based on C-DCGAN

Regarding the design of the discriminant network structure, it uses an architecture similar to the downsampling path of the generation network, using a 4X4 convolutional layer with a stride of 2, and the number of channels is doubled after each downsampling. All convolutional layers are batch normalized, and LeakyReLU with a slope of 0.2 is used as the activation function. In the last layer, a 1x1 convolution is used to map the feature map to a one-dimensional output, and then the sigmoid function is used to return the probability value of the input being true or false.

The detailed structure of the discriminant network is shown in Table 1.

3.3 Loss Function

The loss function of the neural network is an important reference for adjusting the weight of the network model. The loss function in this article is composed of GAN loss and chroma loss.

The conditional GAN loss function is expressed as follows:

$$L_{cGAN}(G,D) = E_{x,y}[logD(x,y)] + E_x[log(1 - D(x,G(x)))] \quad (1)$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0018

Table 1. Discriminant network structure

Convolutional layer	structure	Step size	quantity
Conv1	4X4	2	64
Conv2	4X4	2	128
Conv3	4X4	2	256
Conv4	4X4	2	512
Conv5	4X4	1	1

Among them, E represents the mathematical expectation, which is the input condition, and this article is the L channel information. The optimal solution of the GAN objective function is to obtain the parameters of the generating network D and the discriminating network G when the extremely small maximum value is obtained.

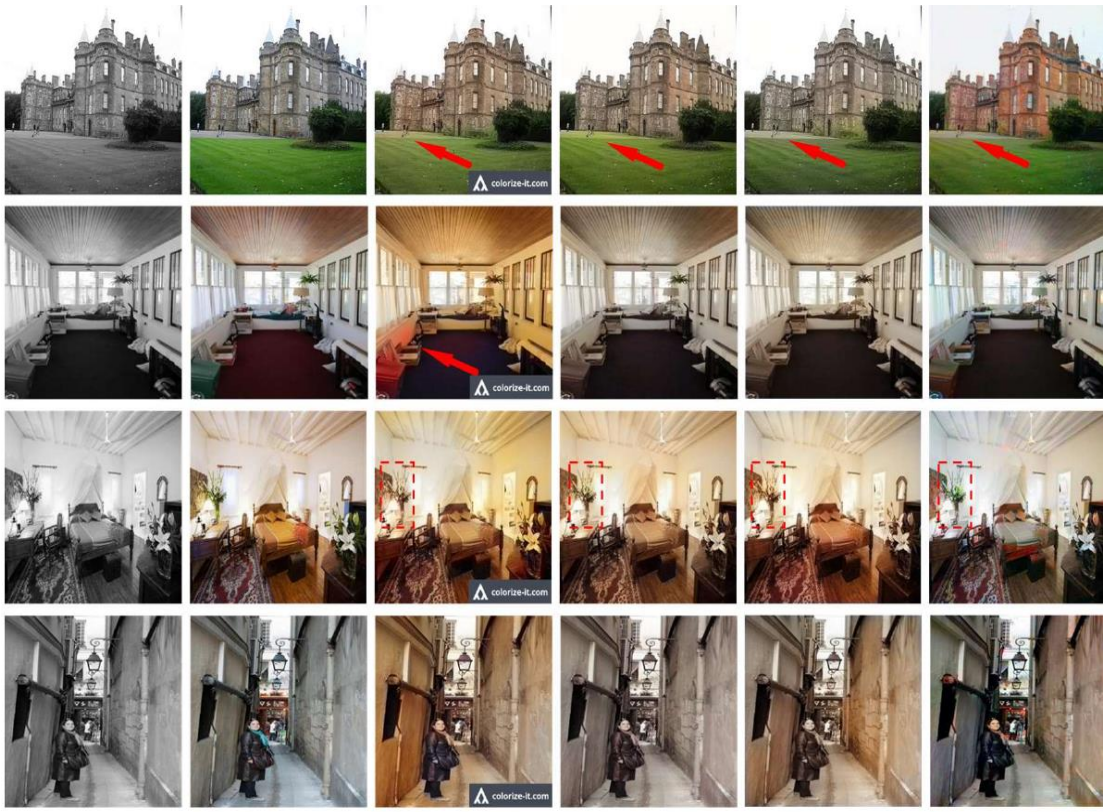
The chromaticity loss is derived from the L1 distance between the generated data and the real color data. The calculation formula is as follows:

$$L_{L1}(G) = E_{x,y}[\|y - G(x)\|_1] \tag{2}$$

Adding the L1 loss function to the GAN loss can force the generation network to generate similar results to the real image, avoiding the generation of too aggressive color deception discrimination network. At the same time, the chroma loss restricts the generator, which can speed up the convergence speed of GAN and improve the training efficiency of the overall network. Therefore, the final loss function is as follows:

$$G^* = \arg\min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G) \tag{3}$$

Among them, λ is the weight.



A B C D E F

Figure 6. Comparison of coloring effects of different colorization models. (A) Grayscale image; (B) Real image; (C) Literature [8]; (D) Literature [9]; (E) Literature [10]; (F) This paper

4. Experimental Results and Analysis

When training the model, use the training set of Places365 [6] and ILSVRC2012 [7] for training. Each batch of Places365 and ILSVRC2012 data sets has a batch size of 16, and a total of 20 iterations. Before training, first preprocess the data set, convert the image color space from RGB space to Lab space, then separate the L channel and the ab channel, and input the L channel as input to the generation network G to obtain the two-channel regression result. As the ab channel and the L channel are spliced, the data of the three channels after splicing is the colorization result generated. The optimizer in the model uses Adam, the learning rate is set to 0.0003, and the momentum is set to 0.5. The weight λ in the loss function is set to 100.

4.1 Subjective Evaluation Analysis

In order to illustrate the effectiveness of the colorization model proposed in this paper, the model in this paper is compared with other existing colorization models (reference [8], reference [9] and reference [10]). The experimental results are shown in Figure 6. From the figure, it can be seen that the algorithm in this paper reduces the boundary blurring, reduces the loss of detail, and effectively alleviates the problem of boring coloring.

4.2 Objective Evaluation Analysis

At the same time, in order to objectively illustrate the superiority of the algorithm in this paper, this paper uses two evaluation criteria of peak signal-to-noise ratio and pixel accuracy for experimental comparison. The results are shown in Table 2. It can be seen from the table that the model in this paper has a higher peak signal-to-noise ratio and pixel accuracy, indicating that the coloring effect of the model in this paper is more natural than the other three algorithms and closer to the original image.

Table 2. Comparison of objective indicators of different models

	Place365		ILSVRC2012	
Model	PSNR/dB	Acc/%	PSNR/dB	Acc/%
Literature [8]	22.59	15.7	22.43	15.2
Literature [8]	25.17	16.4	24.92	15.9
Literature [8]	25.22	16.9	24.45	15.7
This paper	25.54	17.6	25.12	17.1

5. Conclusion

In order to improve the current gray-scale image colorization methods that generally have problems such as border blurring, loss of details and boring coloring, this paper proposes an improved conditional depth convolution generation against network gray-level image colorization method. The network model fusion generates a confrontation network structure, builds a discriminant network, and introduces an improved deep aggregation structure network into the field of image colorization, which improves the feature utilization rate of the network, thereby improving the algorithm model's ability to process image boundaries and details, and at the same time dynamic Evaluate the color quality of the picture. Experimental results show that the algorithm in this paper effectively alleviates the problems of border blurring, loss of detail and boring coloring in image colorization.

References

[1] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.

[2] Yu F, Wang D, Shelhamer E, et al. Deep Layer Aggregation [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018:2403-2412.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0018

[3] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets. MIT Press, 2014: 2672-2680.

[4] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks [C]// 2015.

[5] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.

[6] Zhou B, Lapedriza A, Khosla A, et al. Places: A 10 Million Image Database for Scene Recognition.[J]. IEEE transactions on pattern analysis and machine intelligence, 2018,40(6):1452-1464.

[7] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.

[8] Zhang R, Isola P, Efros A A. Colorful Image Colorization [C]// European Conference on Computer Vision. Springer International Publishing, 2016.

[9] 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[J]. ACM Transactions on Graphics, 2016, 35(4):110.1-110.11.

[10] Isola P, Zhu J Y, Zhou T, et al. Image-to-Image Translation with Conditional Adversarial Networks [C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017:1125-1134.