Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 15, no. 1, January 2019, pp. 270-280
DOI: 10.23940/ijpe.19.01.p27.270280

Image Colorization Algorithm based on Dense Neural Network

Na Zhang, Pinle Qin, Jianchao Zeng*, and Yulong Song

School of Data Science and Technology, North University of China, Taiyuan, 030051, China

Abstract

In most scenes, color images have richer information than grayscale images. This paper presents a method of grayscale image pseudo coloring that constructed and trained an end-to-end deep learning model based on dense neural network aims to extract all kinds of information and features (such as classification information and detail feature information). Entering a grayscale picture to the trained network can generate a full and vibrant vivid color picture. By constantly training the entire network on a wide variety of data sets, you will get the most adaptable, high-performance pseudo color network. The experiments show that the method proposed has a higher utilization of features and can obtain a satisfactory coloring effect. Compared with the current advanced pseudo color methods, it has also made remarkable improvements, and to a certain extent, the problem during the coloring processing have been improved, such as color overflow, loss of details, low contrast etc.

Keywords: image coloring; densely connected convolutional networks; grayscale image; feature utilization; information loss

(Submitted on October 21, 2018; Revised on November 23, 2018; Accepted on December 24, 2018)

© 2019 Totem Publisher, Inc. All rights reserved.

1. Introduction

Color information is an important part of the image, which can show an abundant space hierarchy by combining with the semantic and surface texture information of the scene in the image. Research shows that the human eye is highly sensitive to the color intensity and its transformation, while the sensitivity to the grayscale value transformation is far less. In terms of human psychology, color images can give observers a more pleasant and enjoyable feeling, which helps to understand the content of images and obtain more comprehensive and rich information from them to enhance the use value of images.

Grayscale image coloring (i.e., pseudo-color processing) is a technology that is generated under the above requirements by assigning a color to grayscale values [1] by a specified rule to restore, enhance, or change the color information of the image. Currently, there are mainly two kinds of methods for image coloring: ***User-guided color spreading*** and ***Data-driven image colorization***.

The algorithms based on user-guided color spreading is when the user conducts the guiding coloring of the key part or area, and then the algorithm performs color expansion according to the set rule or conversion specification. Levin et al. [2] proposed a global optimization colorization algorithm. After the user`s coloring, a color image matched with the user's scribbles was generated, and the priori definition of the image was supported. Lagodziński et al. [3] proposed a novel colorization method that uses the morphological distance transform and the image structure to propagate the colors scribbled by users in grayscale images automatically. The method mentioned above can achieve good results, but different color regions need to be clearly indicated. It usually requires intensive user interaction, and the shade of colors are not well characterized and implemented; color overflow will also easy to appear due to the inappropriate label and the gray-values which are extremely similar etc. Based on the optimized colorization method, Shah et al. [4] used three correlation thresholds to evaluate their performance in information loss. LI et al. [5] also used the colorization method based on thresholds. Such methods have a high dependence on the selected threshold and have a limited number of colors, which make the color shows relatively curt.

* Corresponding author.
E-mail address: zjc@nuc.edu.cn

The method of pseudo coloring based on data-driven include many categories, including example-based approach. Welsh et al. [6] proposed a method that compared the brightness and texture information between the grayscale image and the example color image to realize the coloring of the grayscale image. Liu et al. [7] proposed that the reference instance images for colorization related to the target grayscale image may be directly searched from the Internet. Liu, Michael et al. [8] and Morimoto et al. [9] achieved grayscale images coloring by color conversion and image analysis. This method works well when the target grayscale image is similar to the reference image but finding reference images and matching is really a time-consuming job. When the colorization target or scene is very complicated or rare, the coloring effect is more difficult to be guaranteed. Irony et al. [10] used the texture feature matching to do the image segmentation of the reference image and the grayscale image firstly. With reference to the reference image, paint the part in the grayscale image with the same color when the texture was similar, although good results can be achieved, the segmentation processing will increase the image processing burden.

The development of deep learning methods and the advent of high-performance GPUs have opened new directions for data-driven image colorization methods. This kind of method uses neural network. By building different network architecture, it can extract and analyse the content and features of the image and look for the mapping relation between the grayscale image and the color image and train the corresponding model and realize the image colorization. Cheng et al. [11] used large-scale data modelling, post-processing based on joint bilateral filtering and adaptive image clustering to integrate image global information. Deshpande et al. [12] used the LEARCH framework to train the quadratic objective function in the chromaticity diagram and achieve image colorization by minimizing the objective function. The structure of such network is relatively simple, and the colorization effect is limited. Zhang et al. [13] proposed to extract image features by VGG convolutional neural network [14] and predicted the color histogram of each pixel to color the image. Later, they put forward a new idea to extract the information by using U-Net network [15] and combined it with the user interaction [16]. Lizuka et al. [17] constructed a dual-stream structure network; they extracted the global classification information and local characteristic information of the image simultaneously and fused the two types of information to predict the pixel color. Compared with the previous methods, these three methods have made great improvement. However, due to the subsampled and up-sampling they have used during the image processing, there is a certain degree of information loss. Qin et al. [18] used the *Residual neural network* [19] to extract the detail features, and then combined with the guidance of classification information. Their method has helped to reduce the information loss in some extent, but there are still some problems such as incomplete details coloring and color overflow. In summary, the existing grayscale image colorization techniques mainly have the following problems:

The information loss on details. Natural images are often rich in details; a good grayscale image colorization algorithm should have the ability to show a variety of details as much as possible. Due to the limited efficiency of the corresponding mapping, the existing algorithms may lose some information in the processing of feature extraction, which leads to some part of the image cannot be set as appropriate color especially the small objects. The influence on image colorization is that the content that should be independent often becomes "submerged" in the main color of the whole image; this is very detrimental to the full utilization of image information.

Color overflow. Each object in color images has its own different color. However, the existing algorithms for grayscale image colorization often have the problem of “color overflow” on a certain extent. That is, there exists color leakage; the color of object A spreads to object B. This phenomenon is often due to that the mapping of the colorization algorithms is not sensitive on the subtle grayscale difference, and part of the low-level semantic information and object contour have been lost during the feature extracting.

The weak contrast. One of the main purposes of pseudo-color technology is to make full use of the high sensitivity of the human eye on different color values and intensities of the color image so as to achieve more efficient information acquisition. The existing pseudo-color algorithms show a weak contrast on a certain degree. In some cases, it tends to give the image an overall color tone. Different objects are difficult to separate by color, which is similar to adding a layer mask or a filter on the whole image, making the colored image has a lower richness and sharpness.

Discordant of colorization. Due to the above problems, the colorization result is not very satisfactory; the whole picture is not very naturally. We can carefully draw a conclusion that the main problems of the existing pseudo-color algorithms are mainly that the information utilization rate is not high when establishing the mapping relationship between the gray image and the color image, especially since the algorithm based on deep learning has obvious gradient vanishing problem when the network is deepened, and during the abstract processing on low-level features the information may have lost gradually.

This paper proposed a new method based on Densenets [20] to solve the mentioned problems. The method is based on

the idea of skip-connection and has high utilization rate of information; the loss of the low-level semantic information and object contour information is also small. In this paper, we used its characteristics to extract the texture and detail features in the image. At the same time, we obtained the classification information of the image through the VGG sub-network. The network combines the texture details and classification information for feature re-extraction and predicts the color according to the integrated features to obtain the output. We compared the output image with the original color image and calculated the mean square error. After several optimizations training, we got a final colorization model to achieve the conversion from grayscale image to color image. The network does not change the size of the feature map when extracting the detail information. The feature information that is gradually discarded or lost in the traditional network is reused through Densenets, which effectively reduces the gradient vanishing problem and enhances the transmission and utilization of features. As a result, the information utilization ratio of the original image by using the network is higher, the obtained coloring image is better, and the details are also more complete and abundant.

2. Related work

Convolution neural network. Convolution neural network has already become a research hotspot in image recognition, speech analysis, natural language processing and other fields. It has feature sharing, which can effectively reduce the complexity of the network. It can also play a very effective role in feature extraction and establishing feature mappings, especially in image analysis and processing, images can be input directly into the network, which can avoid the complexity of data reconstruction in feature extraction and classification tasks. At present, many excellent networks based on the convolutional neural network have already emerged such as AlexNet [21], VGG, GoogleNet [22], ResNet and other structures; the Top5 error rate of image classification task has been reduced below 5%.

Skip-connection. In the deep learning network, as the depth of the network deepens, the gradient vanishing problem will become more obvious. Aiming to solve this problem, many researchers have put forward the targeted solutions, such as ResNet, Highway Networks [23], Stochastic depth [24], FractalNets [25], etc. The network structure of such algorithms varies, but the core is same: creating short connection paths between early and later levels. Among them, **Residual neural network** (ResNet) has been adopted by more researchers due to its better performance and structural simplicity. It introduces a short connection between the output of the residual block and the input, rather than only stacking the network simply. The mapping relationship can be expressed as $F(x) + x$, as shown in Figure 1.

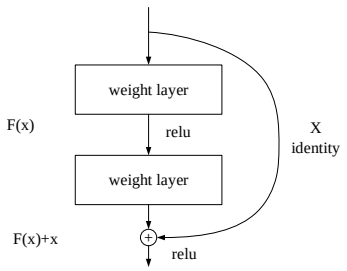

Figure 1. Residual structure

Densely Connected Convolutional Networks. Unlike ResNet that lets the network get deeper or wider, DenseNets tries to achieve better performance and potential by feature re-use to generate the compression modelling, which is easy to train and parameter-efficient. By adding connections between feature maps in different levels, the subsequent changes in the input layer will increase, and the network will be more effective. This is also the main difference between DenseNets and ResNet. The dense block structure is shown as Figure 2.

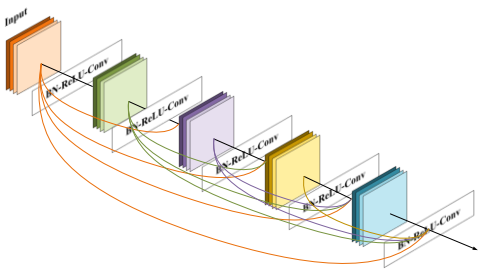

Figure 2. The dense block structure

In a convolution neural network with l levels, each layer l corresponds to a nonlinear mapping $H_l(\cdot)$. Each mapping usually contains the following items: Batch Normalization, ReLU Activation, Pooling and Convolution. When an image X_0 enters the network, after the lth layer, the output will be X_l. The traditional feedforward network only connects the output of the $(l - 1)^{th}$ layer to the lth layer directly, that is $X_l = H(X_{l-1})$. ResNet added a skip-connection, as shown in Equation (1).

$$X_l = H_l(X_{l-1}) + X_{l-1} \tag{1}$$

In DenseNets, the network connects the convolutional layer directly to all subsequent levels, i.e. l^{th} layer receives the features from all the previous convolution layers as a new input, as seen in Equation (2):

$$X_l = H([X_1, X_2, \cdots, X_{l-1}]) \tag{2}$$

Where $[X_1, X_2, \cdots, X_{l-1}]$ is processing to connect the feature maps from 0, 1, 2,⋯, (l - 1) layers. Compared with Resnet, DenseNets makes full use of the feature information in the previous layer without adding too many parameters, improving the performance of the network effectively.

3. The Image Colorization Algorithm Based on Dense Neural Network

3.1. The Network Structures

The existing grayscale image colorization networks based on deep learning mainly extract the detail texture feature of the image by constructing convolution neural network; the coloring effect is acceptable. However, if there is no proper way to learn the image global context information correctly (such as whether the scene is indoor or outdoor, etc.), the network may have obvious errors. Lizuka et al., Qin et al. incorporated the category information of the images into the network and used the information to co-train the model, which played an informational guiding role for the entire colorization network.

This paper draws on the advantages of the method of Lizuka et al. Qin et al. constructed a dual-stream structure based on the dense neural network, learning the classification information and the detail texture information of the image at the same time. In this paper, we used CIE Lab color space. When the image enters the network, it only learns and predicts the color information of the channel a and b. Then, it combines with the information in channel L from the grayscale image to achieve the coloring. The network structure is shown as Figure 3.

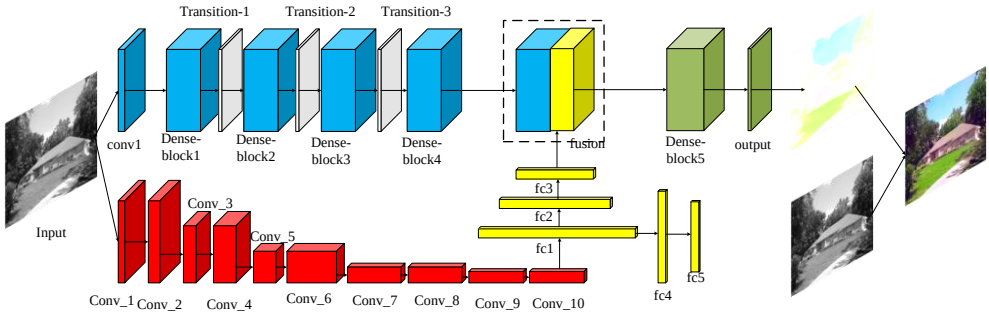

Figure 3. The network structure

The whole colorization network is composed of three parts: **the feature extraction part based on Densenets** (the blue part, including the gray transition connection layer), **the VGG-based classification guidance part** (the red and yellow parts), **the fusion and output part** (the part circled by dotted lines, green part and output). The detailed structure of the Dense block in the first part is shown in Figure 4 .

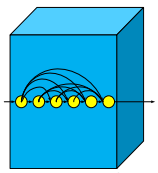

Figure 4. The structure of the dense block

Images with category labels will be converted from ***BGR color space*** to ***Lab color space*** before entering the network. The entire network model is supervised learning model. The input —channel L of the original image— entered the feature extracting part and the classification part of the network, and then generated the prediction of the ab channel and the prediction of classification correspondingly. Compared with the original ab channel information and the classification labels, we can calculate the colorization loss and the classification error. Then, all losses are feedback to the network and weights are adjusted to train the entire network. The detailed structure of the network is shown in Table 1.

Table 1. The detailed structure of the network

(a) Densely connected Network

Layers name	Structure	Outputs
conv1	7x7 conv, stride 2	16
Dense-block1	[1x1 conv, 3x3 conv] X12	160
Transition-1	1x1 conv	80
Dense-block2	[1x1 conv, 3x3 conv] X12	224
Transition-2	1x1 conv	112
Dense-block3	[1x1 conv, 3x3 conv] X12	256
Transition-3	1x1 conv	128
Dense-block4	[1x1 conv, 3x3 conv] X12	272
(c) Combine and output		
Layers Name	**Kernel**	**Outputs**
Fusion	1x1	128
Dense-block4	[1x1 conv, 3x3 conv] X12	64
Output	3x3	2

(b) Global Features Network

Layers Name	Kernel	Stride	Outputs
conv_1.	3x3	2x2	64
conv_2.	3x3	1x1	128
conv_3.	3x3	2x2	128
conv_4.	3x3	1x1	256
conv_5.	3x3	2x2	256
conv_6.	3x3	1x1	512
conv_7.	3x3	2x2	512
conv_8.	3x3	1x1	512
conv_9.	3x3	2x2	512
conv_10	3x3	1x1	512
fc1	-	-	4096
fc2	-	-	256
fc3	-	-	64
fc4	-	-	1024
fc5	-	-	205

3.1.1. Feature Extraction

The image of H X W X1 enters the feature extraction section based on dense neural network. It passes through a layer of convolution before entering four dense blocks one after another. The convolution layers in each dense block are closely connected with the subsequent convolution layers, and every layer in each block outputs 12 feature maps. In view of the denseness of Densenets, we put a 1x1 convolution before each 3x3 convolution (as the dense block part shown in Table 1 (a)). This processing decreases the number of input feature maps, reduces the dimensionality, cuts back the amount of computation required to a large extent, as well as blends the features of each channel.

Between every two Dense blocks, we added a 1x1 convolution (the Transition layer). This way, the output feature map of the previous dense block can be reduced (in this paper, the network is set to be reduced to half), so as to avoid the network being too large effectively, as well as reduce the computation burden of the next dense block.

After the image passes through the above network, a large number of features and texture information will be extracted. Because the convolutions in dense blocks are connected to all the previous layers, low-level information is used, which effectively reduced the information loss and improved the problem of gradient vanishing.

3.1.2. Classification Guidance

When the image enters the classification guidance network, the network will extract the classification information of the image. The fully connected layer fc1 reconstructs the extracted features into a 1x4096 eigenvector, and then integrates them via fc2 and fc3 to obtain a 1x64 eigenvector. This helps the entire network determine the category of the image content.

3.1.3. Fusion and Output

After the feature extraction network and the classification guiding network both complete the information extraction, two

parts of information will be reconstructed into a feature map with the same dimension and fused in the fusion layer. After the feature re-extraction by Dense-block4, the network will finally generate a H X W X2 output after a convolution operation.

3.2. Loss

As described in Section 3.1, the loss of the network will serve as an important reference for adjusting weights. Loss consists of two parts: one is the feature extraction loss (L1) from the feature extraction network and the other is the classification loss (L2) from the classification guidance network. Feature extraction loss and classification loss are fed back to the network independently and do not interact with each other.

During the training, the network read n images per batch. After obtaining the output prediction, we compared the color prediction with the original image and use the ***Mean Squared Error*** (MSE) to measure the disparity between the network output and the true value. This is shown in Equation (3) and Equation (4):

$$MSE = \frac{1}{w \times h} \sum_{i=1}^{w} \sum_{j=1}^{h} \left(Y_p(i, j) - X_p(i, j) \right)^2 \tag{3}$$

$$L_1 = \frac{1}{n} \sum_{k=1}^{n} MSE_k \tag{4}$$

Where w and h are the width and high of the sample, Y_p is the color of the ab channel of the original image, while X_p is the prediction value by the network, and n refers to the number of images contained in a training batch.

As for the classification guidance part, let the classification information of the input image to be the guidance label y^{label}, while the predicted classification of the network is y^{out}. Then, use ***Cross-Entropy*** to measure the disparity between the classification prediction of the network and the real classification, as shown in Equation (5):

$$L_2 = \frac{1}{n} \sum_{i=1}^{n} [-\sum_{i} y_i^{\text{label}} \log(y_i^{\text{out}})] \tag{5}$$

When calculating the log function value on y_i^{out}, if the value of y_i^{out} is 0, $\log(y_i^{\text{out}})$ will be positive infinity. So, when the number is smaller than $1e-10$, set it equal to $1e-10$.

3.3. The Result evaluation——Information Loss

As stated in chapter 1, the main purpose of coloring the grayscale images is to obtain more information from the colored results than the grayscale images. Whether the result is clear and the amount of information contained is sufficient can be regarded as an important index to measure the merits of the colorization algorithm.

Image information entropy is a widely used index to evaluate the information adequacy in an image. The amount of information contained in an image can be judged by calculating its information entropy. The larger the information entropy value, the higher the grayscale/chromaticity levels does the picture has; correspondingly, the amount of color information is more adequate. By using information entropy, we can judge whether the coloring results is consistent with subjective feelings of the human eyes from the objective point of view. It is calculated as Equation (6):

$$InEn = -\sum_{i=0}^{c} P(i) \log_2 \left(P(i) \right) \tag{6}$$

Where $InEn$ is the image information entropy value of the picture, $P(i)$ represents the appearance frequency of the color whose value is i in the whole image, and c represents the range of color values in Lab color space. By comparing the information entropy between the results of the colorization network and the original image and calculating the discrepancy between them, this paper adopts the evaluation criteria based on information entropy. A smaller the discrepancy means the color information loss is fewer. That is to say, the effectiveness of the colorization network is better.

The network proposed mainly estimates the value of ab channel based on the Lab color space; it does not calculate and process the value of channel L (represent the grayscale information of the image) repeatedly. Therefore, to improve the efficiency, we only consider the information contained in the ab channels when evaluating the colorization effect, and compare the information entropy of the ab channels between the output image and the original image. This is shown as Equation (7):

$$Info_loss = InEn_C - InEn_O \tag{7}$$

Where $Info_loss$, $InEn_C$, $InEn_O$ denote the degree of loss of color information, the entropy of the original image, and the information entropy of the output image in this paper. It can be seen from the above definition that the smaller the Info_loss value, the better the colorization result.

4. Experiments and Analysis

4.1. Experimental Data Set and Environment

As a supervised colorization method, the network proposed needs a large number of color images with classified labels as training data sets. Therefore, we chose the *MIT Places Database* (containing 205 scenes and more than 2.5 million images) [26] and *ImageNet* (containing 1000 scenes and more than 1.2 million images) [27] to train the network. *HDF5* is used to process the data set and generate a data file of ".h5" type, which is no longer necessary to read a large number of single pictures in sequence to facilitate the operation and maintenance. The colorization network proposed needs a lot of matrix calculation. To improve the training efficiency, we used GPU (*Graphics Processing Unit*) to do the training; the GPU type is NVIDIA Tesla M40. In the implementation of the method, we used Python programming environment, and chose TensorFlow [28] to build the network.

4.2. Coloring Result and the Comparison with Advanced Methods

4.2.1. The Coloring Effect Comparison that Have Reference of the Original Image

To verify the effectiveness of the proposed algorithm, we choose some representative images and compared them with the existing excellent algorithms such as Zhang et al., Lizuka et al., Qin et al. This is shown in Figure 5. We compared the coloring effect from the following aspect: **the information loss in details** (the numbers under each picture in Figure 5 is the corresponding Info_loss value calculated according to the formula (7)), **whether the color overflows**, **color contrast**, and **the overall harmony**. In the comparison, the coloring results of Zhang et al. and Lizuka et al. are all derived from the coloring models of their external publicity websites. The coloring effect of Qin et al. is based on their final coloring model.

The information loss in details. As group (1) shows in Figure 5, none of Zhang et al., Lizuka et al., Qin et al. have colored the plant next to the window in green properly. In group (2), Zhang et al., Lizuka et al. did not distinguish between the sparse parts of grassland when coloring the lower left corner. In group (4), neither of Zhang et al., Lizuka et al., Qin et al. did give due green color to the plants in the picture frame. In group (7), Lizuka et al., Qin et al. did not color the plant properly. Our method performed well in these details.

Whether the color overflows. In group (2), Zhang et al. spread a little bit green of the trees to the roof boundary. Qin et al. appeared abnormal blue on the right side of the roof. In group (3), Zhang et al. have some anomalous green around the image, and a color leak appears beside the skirt of the woman dancer in the result of Qin et al. In group (7), Zhang et al. have appeared unclear boundary and color leakage phenomenon beside the left side of the green leaf and at the roof of the room.

Color contrast. In the above groups, the overall coloring results of Zhang et al., Lizuka et al., Qin et al. tend to be biased toward the overall color tone, and the contrast of colors in different objects in the image is not strong, In particular, the walls and floors of the three in group (1), the ceiling of Zhang et al., Lizuka et al. in group (4). The sky and sand of Zhang et al., Lizuka et al. in group (5), The walls, ceilings and floors of Zhang et al., Lizuka et al. in group (6), (7). It can be seen that the method of Qin et al. have solved the problems of low contrast and border leakage to a certain extent, but the contrast of the overall effect is still not as significant as the method proposed in this article.

The overall harmony. Combined with the mentioned directions, our method is superior to Zhang et al., Lizuka et al., and Qin et al. In addition, it can be seen from the comparison that the subjective feeling is always better when the Info_loss

value is smaller, which also shows that the loss of information can indeed play a role in judging the coloring effect. This is consistent as the subjective observation feelings of the human eyes. Overall, the Info_loss value of our method is smaller than that of the other three methods, which also objectively verifies the superiority of our method.

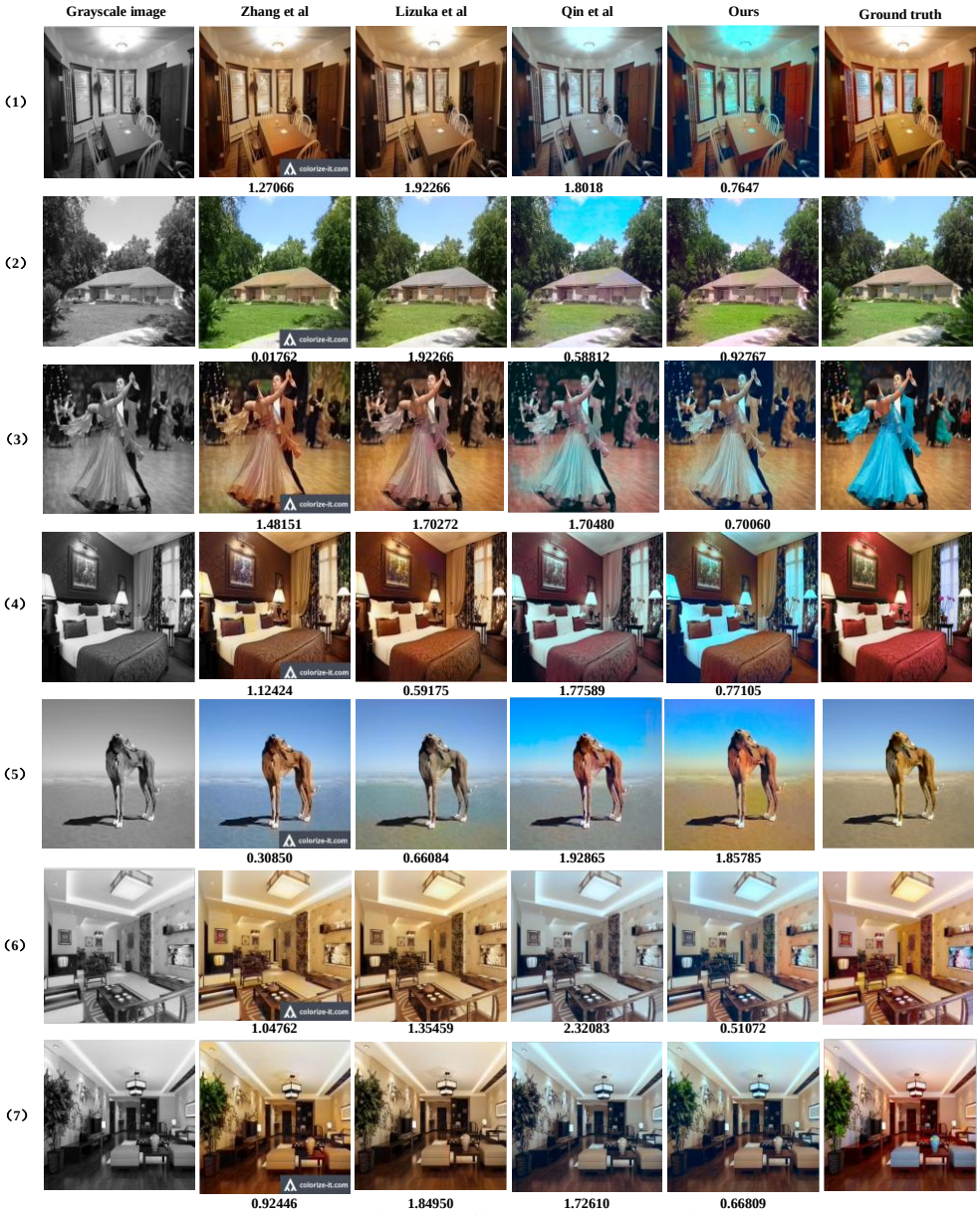

Figure 5. The comparison of coloring effect (with the guidance of original image)

4.2.2. The Coloring Effect Comparison of Old Photos and Grayscale Image

To verify the universality of our algorithm, we selected some old photos and grayscale images to compare the coloring effects, as shown in Figure 6. From the coloring effect comparison of old photos and grayscale images, we can see that our method has less color overflow, better contrast and more abundant details than the algorithm of Zhang et al., Lizuka et al., and Qin et al. (Such as in group (1) our method gave the tree trunk the proper color brown).

4.2.3. Algorithm Comparison (*with Qin et al.*)

As we can see, compared with the prior advance colorization method based on deep learning. the colorization method based on Residual neural network proposed by Qin et al. have reduced the information loss to a certain degree; the coloring effect is also improved. We compared them by calculating the Info_loss value of the two methods. We chose 5000 pictures

randomly, and then calculated the average. The mean of Qin et al.'s method is 2.31481063099, and the mean of the method proposed in this paper is 1.92628092468. Figure 7 shows the result comparison. The blue line in Figure 7 refers to the method proposed in this paper, while the red line refers to the method of Qin et al. The horizontal axis indicates the value of Info_loss and the vertical axis indicates the frequency of images.

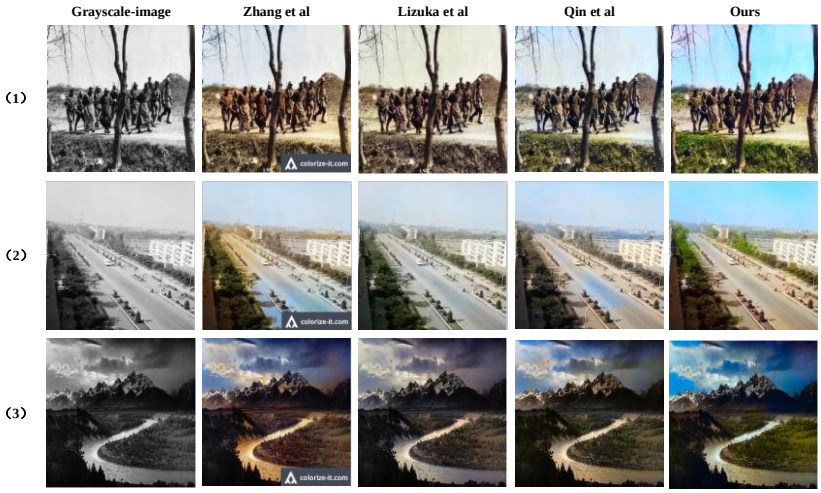

Figure 6. The comparison of coloring effect (old photos and grayscale images)

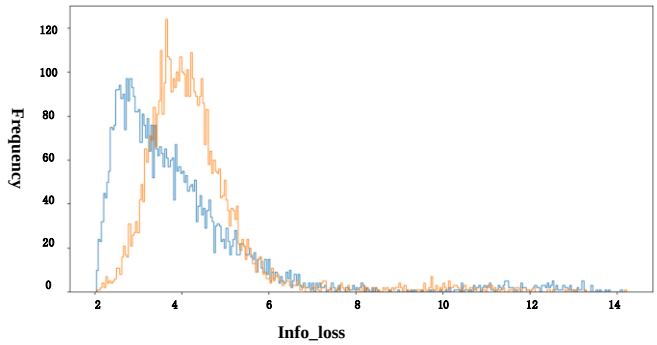

Figure 7. The comparison of the two methods

5. Conclusion and Prospection

5.1. Conclusion

In this paper, we proposed a gray-scale image colorization algorithm based on dense neural network. The algorithm includes sub-networks and VGG sub-networks, which are composed of dense blocks to extract detailed texture information and classification information respectively. Two kinds of information fused together to generate the output of the network as the prediction of color picture. The experimental results show that the proposed method is better than the existing excellent grayscale image colorization algorithm in terms of detail information richness and contrast, and the color overflow is also significantly reduced, using in old photos and grayscale images coloring can also get good performance.

5.2. Shortage and Future Research Direction

Due to the denseness of the network, the performance requirements of the running equipment are high, and the network training needs a long time. At the same time, the coloring effect of our method may not ideal for an image that has not been learned since the data set hasn't cover all the image categories. In the next phase of research, we plan to enhance the universality and utility of our approach by optimizing the whole network architecture and training more type of images.

References

1. R. Gonzalez and R. Woods, "Digital Image Processing," Publishing House of Electronics Industry, Beijing, 2011
2. A. Levin, D. Lischinski, and Y. Weiss, "Colorization using Optimization," *ACM SIGGRAPH*, Vol. 23, No. 3, pp. 689-694, 2004

3. P. Lagodziński and B. Smolka, "Medical Image Colorization," *Journal of Medical Informatics & Technologies*, Vol. 11, pp. 47-57, November 2007
4. A. A. Shah, M. Gandhi, and K. M. Shah, "Medical Image Colorization using Optimization Technique," *Acta Medica Okayama*, Vol. 62, No. 141, pp. 235-248, 2013
5. F. Li, L. Zhu, L. Zhang, Y. Liu, and A. Wang, "Pseudo-Colorization of Medical Images based on Two-Stage Transfer Model," *Chinese Journal of Stereology & Image Analysis*, Vol. 18, No. 2, pp. 135-144, February 2013
6. T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring Color to Grayscale Images," *Acm Transactions on Graphics*, Vol. 21, No. 3, pp. 277-280, July 2002
7. X. Liu, L. Wan, Y. Qu, T. T. Wong, and S. Lin, "Intrinsic Colorization," *Acm Siggraph Asia*, Vol. 27, No. 5, pp. 1-9, December 2008
8. Y. Liu, C. Michael, U. Matt, and R. Szymon, "AutoStyle: Automatic Style Transfer from Image Collections to Users' Images," *Eurographics Symposium on Rendering Eurographics Association*, Vol. 33, No. 4, pp. 21-31, July 2014
9. Y. Morimoto, Y. Taguchi, and T. Naemura, "Automatic Colorization of Grayscale Images using Multiple Images on the Web," in *Proceedings of the SIGGRAPH 2009*: *Talks ACM*, pp. 1-1, New York, USA, August 2009
10. R. Irony, D. Cohen-Or, and D. Lischinski, "Colorization by Example," *Eurographics Symposium on Rendering Techniques, Konstanz, Germany*, Vol. 5, No. 10, pp. 201-210, June 2005
11. Z. Cheng, Q. Yang, and B. Sheng, "Deep Colorization," in *Proceedings of the 2015 IEEE International Conference on Computer Vision* (*ICCV*), pp. 415-423, Santiago, Chile, December 2015
12. A. Deshpande, J. Rock, and D. Forsyth, "Learning Large-Scale Automatic Image Colorization," in *Proceedings of the 2015 IEEE International Conference on Computer Vision* (*ICCV*), pp. 567-575, Santiago, Chile, 2015
13. R. Zhang, P. Isola, and A. A. Efros, "Colorful Image Colorization," in *Proceedings of the European Conference on Computer Vision* (*ECCV*), pp. 649-666, Amsterdam, The Netherlands, October 2016
14. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv Preprint arXiv: 1409.1556, September 2014
15. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proceedings of the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234-241, Munich, Germany, October 2015
16. R. Zhang, J. Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, et al., "Real-Time User-Guided Image Colorization with Learned Deep priors," *Acm Transactions on Graphics*, Vol. 36, No. 4, pp. 119:1-119:11, July 2017
17. S. Lizuka, 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 Transactions on Graphics*, Vol. 35, No. 4, pp. 1-11, July 2016
18. P. Qin, Z. Cheng, Y. Cui, J. Zhang, and Q. Miao, "Research on Image Colorization Algorithm based on Residual Neural Network," in *Proceedings of the CCF Chinese Conference on Computer Vision*, pp. 608-621, Tianjin, China, October 2017
19. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp. 770-778, Las Vegas, USA, June 2016
20. G. Huang, Z. Liu, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp. 1-9, Honolulu, Hawaii, USA, July 2017
21. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 1097-1105, Nevada, December 2012
22. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., "Going Deeper with Convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp. 1-9, Boston, MA, USA, June 2015
23. R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway Networks," arXiv Preprint arXiv:1505.00387, May 2015
24. G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep Networks with Stochastic Depth," in *Proceedings of the 2016 European Conference on Computer Vision* (*ECCV*), pp. 646-661, Amsterdam, Netherlands, October 2016
25. G. Larsson, M. Maire, and G. Shakhnarovich, "FractalNet: Ultra-Deep Neural Networks without Residuals," arXiv Preprint arXiv:1605.07648, May 2016
26. B. Zhou, A. L. Garcia, J. Xiao. A. Torralba, and A. Oliva, "Learning Deep Features for Scene Recognition using Places Database," in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, pp. 487-495, Montréal, CANADA, December 2014
27. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F. Li, "ImageNet: A Large-Scale Hierarchical Image Database," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp. 248-255, Florida, USA, August 2009
28. M. Abadi, A. Agarwal, P. Barham. E. Brevdo, and Z. F. Chen, C. Citro, et al., "Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," arXiv Preprint arXiv:1603.04467, March 2016

Na Zhang received her undergraduate degree in School of Information and Statistics from Guangxi University of Finance and Economics, Nanning, Guangxi, P.R. China, in 2015. Currently, she is pursuing master's degree in NUC, and her areas of interest are digital image processing, medical image processing and computer vision.

Pinle Qin received the PhD degree in computer application technology from Dalian University of Technology (DLUT),

Dalian, Liaoning, P.R. China, in 2008. He is currently an associate professor with the School of Data Science and Technology, North University of China (NUC). His current research interests include computer vision, medical image processing and deep learning.

Jianchao Zeng received the PhD degree from Xi'an Jiaotong University, Xi'an, Shanxi, P.R. China, in 1990. He is currently the vice president of North University of China (NUC). His current research interests include computer vision, medical image processing and deep learning.

Yulong Song is pursing master`s degree in NUC, and his research areas are digital image processing and computer vision.