American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

A Fast Near-Infrared Image Colorization Deep Learning Mode

ChunMing Tang[a], XinYi Zheng[b*], WenYan Zhu[c]

[a,b,c]*School of Electronic and Information Engineering, Tianjin Polytechnic University, Tianjin 300387, China*

[b]*Email: 664528170@qq.com*

Abstract

Near-infrared(NIR) image colorization is the main research content in the field of current near-infrared image application. It has a wide range of application value. For the problem of image colorization, such as diffuse color and even color error, and can not be automated, A fast near-infrared image colorization model consisting of a lightweight image recognition network module and an image colorization CNN module with a fusion layer, firstly using a lightweight image recognition network for image recognition of near-infrared images, and then selecting from the IamgeNet image library The image of the same class as the scene is used as the training set of the colorized network. After training with the colored CNN module with the fusion layer, the near-infrared image is input as the testing set for colorization. The experimental results show that the color is colored by the algorithm. The image details are clear, the color transfer effect is good and the running speed is fast.

Key words: Near infrared image; Colorization; Convolutional neural network; Image recognition.

1. Introduction

Because near-infrared image (NIR) can reflect the thermal radiation information of the scene target, and is not sensitive to the brightness change of the scene, many important night vision or low illumination scenes, such as mines, wild animal observation points, military bases, etc. We can use near-infrared images to achieve comprehensive monitoring.

----

* Corresponding author.

However, since the image obtained by the near-infrared image capturing device is a grayscale image, the grayscale image has the disadvantages of edge blurring and lack of detail, so it needs to be colored to increase its color and texture information to achieve the goal. Detecting efficiency, reducing judgment time,-enhancing judgment on the scene and reducing visual fatigue.

Image colorization method can be roughly divided into reference image based color migration method [1-4],Color propagation method based on line coloring [5-7]And deep learning method based on convolutional neural network [8, 9].

Reinhard [1] first proposed a color migration method based on reference images. This algorithm needs to manually select a reference image similar to the image to be painted, and migrate the color information of the reference image to the image to be colored through statistical analysis. Welsh [2] only migrates colors while preserving the brightness distribution of the original image. Liu [3] decomposes both the grayscale image and the reference image into the illumination and reflection parts, and establishes the mapping relationship between the two through the mathematical model and then performs color migration, but it requires that the gray image and the reference image have the same color and illumination conditions. Xiao[4] In order to ensure that the relevant pixels between the two images can be correctly matched, a colorization method based on gradient domain preservation is proposed to transform the color migration problem into an energy optimization problem. Although the above algorithm achieves a certain colorization effect, it is necessary to manually select a color reference image similar to the target image content.

Levin[5] first proposed a color propagation method based on line coloring. The problem of colorizing the image is solved by selecting a quadratic energy function optimal solution. Yatziv [6] proposes an efficient colorization method that determines the blending weight by calculating the distance from the colored lines,but the above algorithm will have a boundary color error.Therefore, Luan [7] proposes a colorization method for processing natural images of complex textures by introducing information such as texture features and brightness.

In recent years, with the rise of deep learning, neural networks have achieved leapfrog development in the processing of image colorization problems. The method based on convolutional neural network is to extract and analyze the color features in the image sample set by training, and complete the color migration by a certain method. This method has a better coloring effect on grayscale images and improves the disadvantages of traditional algorithms in delivering colors. Zhang. R [8] analyzes the image scene by adding a classifier, coloring according to the type, Lizuka. S [9] through fusion layer fusion by extracting global and local features of the image separately, improved the effect of colorization. However, the above method due to the large number of layers of the network, when training images, it's need Extremely high computational power requirements, it has a long time when this algorithm runs, sometimes the coloring effect is not ideal.

These three types of image colorization methods are all coloring grayscale images which in the visible light band. There is no corresponding literature for the colorization of NIR images. The biggest difference is that an NIR image edge blurring and image details are not clear. So it is more difficult to colorize NIR images. At the same time, in order to speed up the training of colorized models, reduce computing pressure. We need identify the near-infrared image first. In order to find a similar image for training, there are two main methods for image recognition. One method is based on feature matching, That is using the underlying features (such as SIFT, SURF, LoG and DoG, etc.) to extract local features from the image in fixed steps and scales, en encode the local features, then the image is evenly divided by spatial pyramid matching, and then using classifier(such as support vector machine, K-nearest neighbor, neural network and random forest) to classify the feature vector. Another method is based on deep learning models, Girshick.R proposed the target detection method R-CNN [10] in 2014 and proposed the improved model Fast-RCNN [11] in 2015. Both models are better than other network models in the same period in terms of training speed and accuracy. ShaoQing. R to further reduce network detection time, in 2016, the Regional Recommendation Network (RPN) was proposed, and a new image recognition network, Faster-RCNN [12], was built based on Fast-RCNN. Its core idea is to use CNN to directly generate suggested areas. Thus, the convolution feature of the full image can be shared with the entire detection network. It's shortened the time of target recognition Greatly.

For the problems in the above colorization algorithm, our proposes a lightweight automatic image colorization model for NIR images. Firstly, based on our improved lightweight Faster-RCNN recognition network, the scenes in the near-infrared image are identified. Then, selecting the image of the scene class similar to the NIR image from the ImageNet image library as the training set to training the coloization network, Finally, the NIR image was input as the test set input into the colorization network to finish the colorization of the near-infrared image.

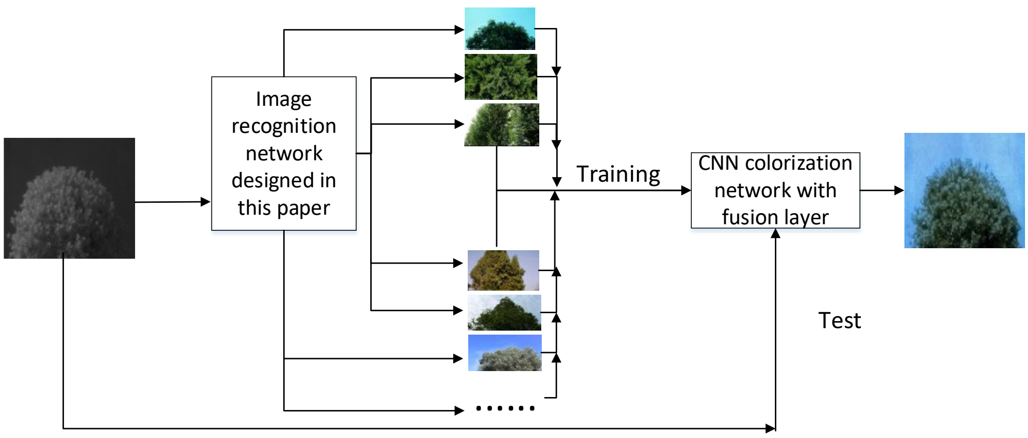


Figure 1: System framework in our paper

2. Lightweight Faster-RCNN Identification Network

Since the traditional deep learning model is used to color the image, it is necessary to select a million-level image library as the training set of the network to achieve good colorization effect. But training such a large image library often takes several weeks. Therefore, firstly we use the image recognition model to identify the targets in the near-infrared image, then select 500-1000 infrared images with similar targets, those image that from the ImageNet image library as the training set of the subsequent colorized network model. This method not only ensures the quality of the color image, but also shortens the training time of the color network.

The difference of the recognition network we designed with the Faster-RCNN is that we have deleted 3 convolution layers in our network model. The reason of the cut the convolution layer is that the deep layer network delete part of layer has little effect on the recognition accuracy, and reduce the memory requirements of the device and speed up the operation. The remaining three convolutional parameters are 32@3*3, 64@2*2, and 64@2*2. Padding equal to 1, the pooling layer uses a maximum pooling method with a size of 2*2 and a padding is 2. The number of nodes in the three fully connected layers is set to 512, 256, and 128, respectively. Finally, the Softmax layer is used as the output layer of the network. Its model structure is shown in Figure 2. The learning rate of the entire network is initialized to 0.001, using exponential decay, the attenuation factor is 0.1, the weight attenuation factor is 0. 98, when training network, the dropout parameter is set to 0.5 when test network, the dropout parameter is set to 0.5.Using Adam optimization algorithm. The specific test process is as follows：

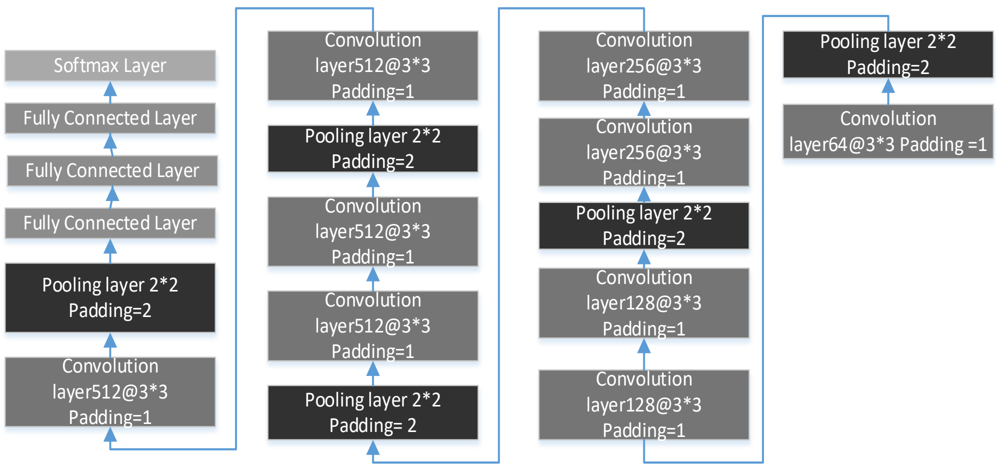


Figure 2: image identification network architecture

Step1. Input the NIR image to VGG-16;

Step2. NIR image is forwarded through CNN to the last shared convolutional layer, get the feature map for the input of the RPN network, continue to forward to the last convolutional layer to produce a higher dimensional feature map;

Step3. Feature maps get regional recommendations and regional scores via RPN network, non-maximum suppression of the regional score (threshold is 0.7), output the area of the Top-N (our article sets the highest score of 300) score to the RoI pooling layer;

step4. The high-dimensional feature map obtained in step 2 and the region output in step 3 are simultaneously input into the ROI pooling layer, and the features recommended in the corresponding region are extracted;

step5. After the region is proposed in step 4, the feature is into the fully connected layer. Finally, the target category of the region and its probability and the target bounding box of the refined target are output.

3. Near-infrared image colorization network with fusion layer

3.1 Colorized network structure

The colorization network in this paper consists of an encoder, a fusion layer,a decoder, and an Inception Resnet V2 classifier, the Inception Resnet V2 classifier is a new classifier designed by Google researchers it's combination of Resnet residual network and Inception network module. It has achieved excellent results in the ILSVRC image classification test.

The specific coloring steps of the colorization network are as follows, firstly, we first use the convolution layer to extract the local features of the near-infrared image, at the same time, the scene classification is performed by Inception Resnet V2 classifier to obtain the global characteristics of near-infrared images, then merge with local features, finally, using convolution and upsampling layers to coloring and restoring image size. We select an image with a similar scene from the NIR image from the ImageNet image library as a training set for the colorized network. The color network framework is shown in Figure 3.

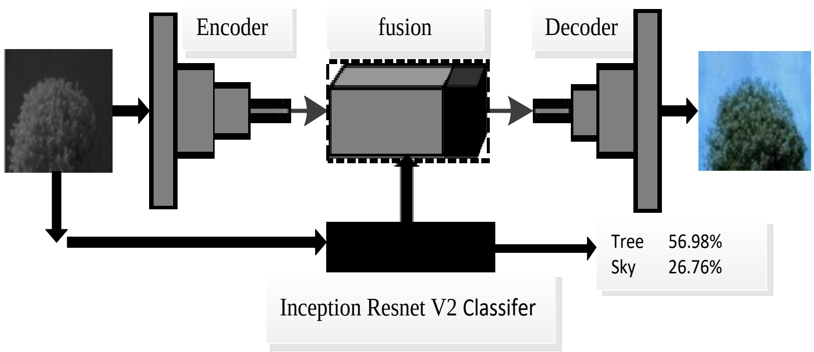


Figure 3: colorizaition network

The input of the encoder is a 256 × 256 near-infrared image, The output is a feature matrix of 8*8*1000. The Encoder consists of a normalized layer and 14 convolutional layers, the convolution layer parameters are set to 64@3×3, 128@3×3, 256@3×3 and 512@3×3. The specific architecture diagram is shown in Figure 4.

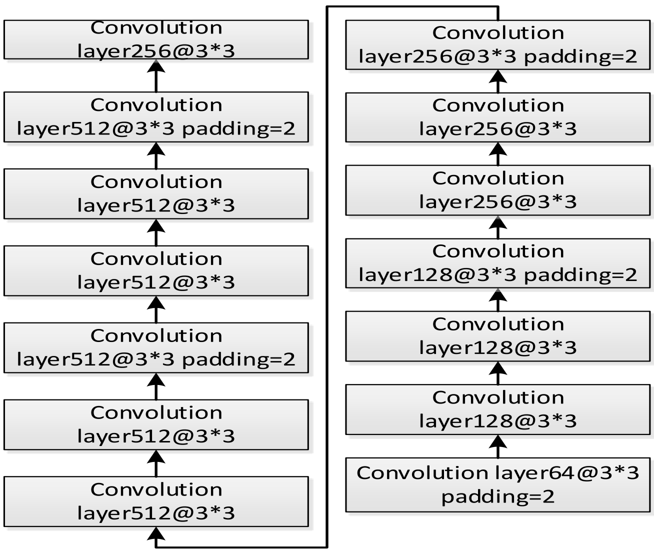


Figure 4: Encoder architecture diagram

The global features obtained by the Inception Resnet V2 classifier and the local features obtained by the Encoder are into together as the input of the fusion layer, then the fusion layer output the fusion feature. The purpose is to improve the image quality after colorization. The fusion layer is fused in the following ways:

$$y_{u,v}^{fusion} = b + w\begin{bmatrix} y^{classifer} \\ y_{u,v}^{encoder} \end{bmatrix} \tag{1}$$

Where $y_{u,v}^{encoder}$ is the feature obtained by the Encoder, $y^{classifer}$ Is the global feature of the image obtained through Inception Resnet V2, $y_{u,v}^{fusion}$ represents the fusion feature on the image coordinates *(u, v)*, *w* is the weight matrix, *b* is the offset.

The input of the decoder is 8×8×1000 feature matrix. The decoder network consists of 5 upsampling layers and 5 convolutional layers.

The setting of the upsampling layer parameters is 2×2, the convolution layer is set to 64@3×3, 32@2×2, 16@2×2, 2@2×2, 2@2×2, the entire network is trained by using MSE (Mean Square Error) guidelines, the network weight is updated by the BP algorithm.

Upsampling uses the basic nearest neighbor method, therefore the length and width of the output image are twice of the input image.

Therefore, the convolution and upsampling operations are stopped when the final result photo size is 128*128.

The convolution layer of the decoder selects Leaky ReLU as the activation function, this makes the final output the 2 numbers between the 0 and 1, these two numbers are respectively used as the values of α and β of the L αβ color space, and L is a known gray value. Finally, the feature matrix of 128 * 128 * 2 is upsampled to 256 * 256 * 2 and merged with the 256 * 256 * 1 NIR image. the colorization of the near-infrared image is completed.

3.2 Objective function and network training

Network training is divided into two phases. Forward Propagation enables feature learning from the input layer to the output layer, back propagation is based on the loss function $L(M,a)$ to calculate the output value:

$$L(M,a) = CE(M,a) = -\sum_{i=1}^{N}\sum_{j=1}^{C}\{\hat{y}_i = j\}\log p_i^j \tag{2}$$

Where $\hat{y}_i$ is the expected value of the i^{th} training sample, p_i^j is the predicted probability of the j^{th} category of the i^{th} training sample, C is the total number of categories of training samples, and N is the total number of training samples.

Throughout the training process, the loss value is calculated by forward propagation, and then back propagated by gradient descent, and the training parameters T and c of each layer are updated layer by layer. The parameter update formula is defined as:

$$T_i = T_i - \alpha\left(\frac{\partial L(T,c)}{\partial T_i} + \beta\right) \tag{3}$$

$$c_i = c_i - \alpha\left(\frac{\partial L(T,c)}{\partial c_i} + \beta\right) \tag{4}$$

a is the learning rate, β is the adjustment parameter, it's used to control the strength of back propagation of loss values.

4. Experimental results and analysis

4.1 Identification results and analysis

Figure 5 is the target recognition result of the target recognition network in our paper for NIR images expressed by position box, semantics and probability, it can be seen from Figure 5 that the recognition network can achieve good recognition accuracy for different categories of targets in the image. Table 1 lists the time required for RCNN, Fast-RCNN, Faster-RCNN, and recognition network of our article for testing images at different resolutions. By comparison, we found that the recognition network of our paper has the fastest recognition speed.

Table 1: Comparison of running time(s) of near-infrared image recognition by different network models

Network name	128*128	256*256	512*512	800*600
RCNN	78.7	164.8	321.7	478.9
fast-RCNN	7.62	15.89	33.63	46.75
faster-RCNN	0.75	1.61	3.26	4.58
Our Image recognition network	0.58	1.35	2.95	3.78

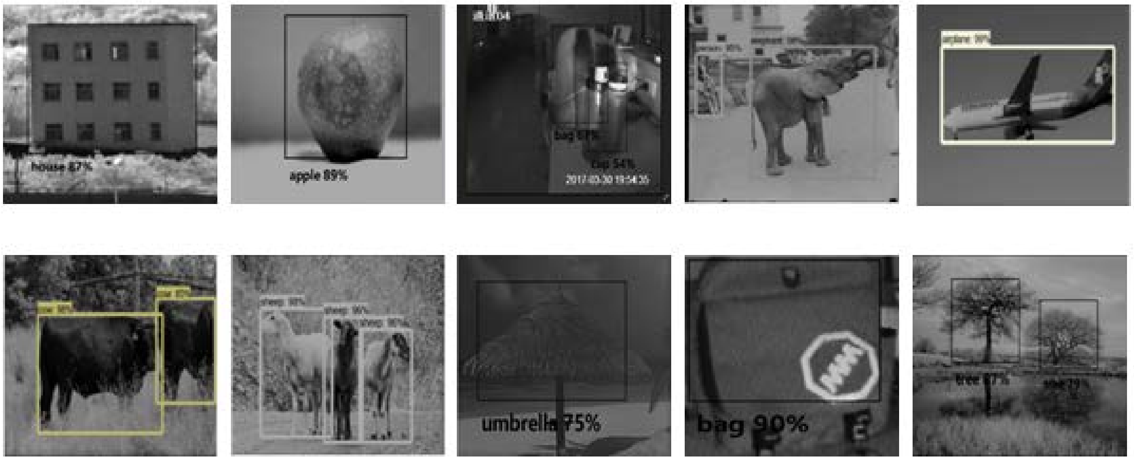


Figure 5: Recognition result of near-infrared image by using Identification network in our paper

4.2 Colorization results and analysis

4.2.1 Image acquisition device parameters

In the experiment we used the LIR X8500SC SLS LWIR to capture high definition NIR images, its spectral range is 7.5 -10.5μm and the maximum resolution is 1280*1024. using WAT-910HX to capture low-resolution NIR image, its spectral range is 8.4-9.8μm, and the maximum resolution is 768*494.

4.2.2 Colorization results and analysis of low resolution NIR images

Low-resolution NIR images have the characteristics of blurred scene edges and unclear detail textures. Figure 6 shows the contrast between the model and other colorized models. Since the model of Zhang's model does not introduce a local feature classifier, therefore, when the color is colored, the edge is diffused and the contrast is not natural enough and the color is too bright, and the training set of the Iizuka's model comes from ImageNet, therefore, it can only identify scene targets that exist in the ImageNet library, and this model was originally colored for black and white old photos, therefore, the color saturation of the image after colorization is poor. the color is not natural enough. Since the colorized model training set of this paper selects an image set similar to the scene in the NIR image, therefore, the color is more accurate and the color is more natural.

Figure 6: The results of the difference model for the colorization the low-resolution image

NIR image Zhang's network Iizuka's network our network NIR image Zhang's network Iizuka's network our network

4.2.3 Colorization Results and Analysis of High Resolution NIR Images

High definition NIR images are characterized by rich texture and sharp edges. Figure 7 shows the comparison of the colorization effect of this model with other colorization models. The Iizuka's model incorrectly identifies the grass as the sky when the second line of images is colored. The model of Zhang is mistakenly dyed into orange for the sky area in the upper right corner of the seventh line image. By contrast, the our model is more prominent in semantic recognition and color consistency.

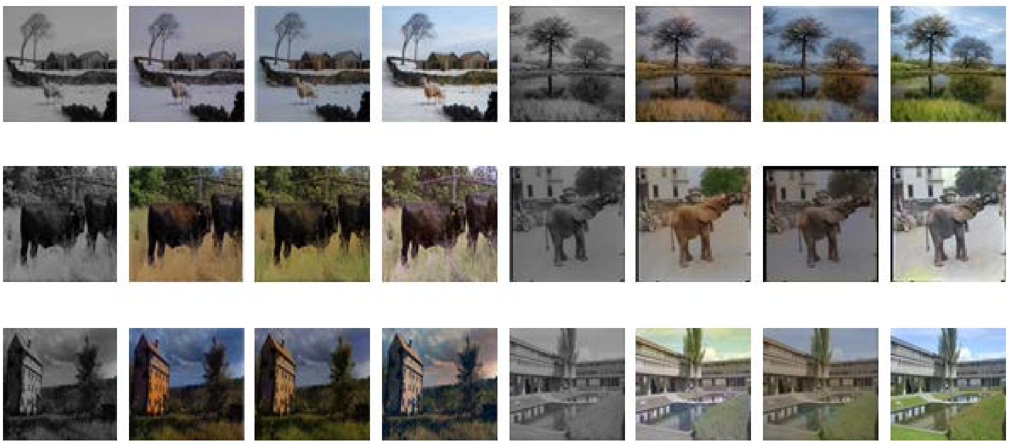


Figure 7: The results of the difference model for the colorization the high-resolution images

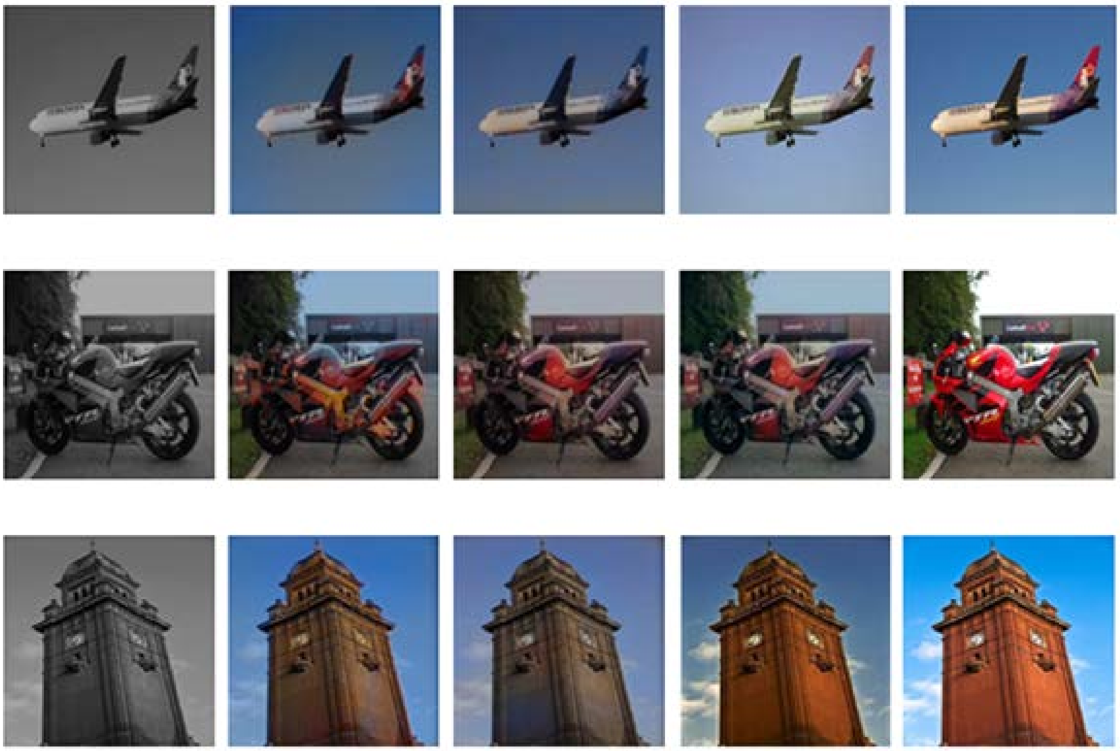
NIR image Zhang's network Iizuka's network our network NIR image Zhang's network Iizuka's network our network

4.2.4 Image quality comparison of different colorized models

(structural similarity, SSIM) is an image quality evaluation index.Table 2 gives the average SSIM (MSSIM) values for several colorization methods in Figure 8.

We took a partial image of the COCO image set as the image to be painted by the near-infrared camera. Figure 8 is a subjective evaluation of the coloring of the near-infrared image by each model. MSSIM as an objective evaluation index of image similarity is shown in Table 2.

Through the visual comparison of Figure 8, it is found that the Zhang.R model has edge overflow phenomenon when dealing with scenes in the image scene; the Iizuka model deals with low contrast of the image scene,especially the sky. The model guarantees that the scene and the edge of the object are clear in the image. Accurately enhances the contrast of the image and improves the visualization of the image.Comparing the MSSIM values with Table 2, it can be seen that the colorized image using this model is closest to the original color image.



(a1) Near infrared image (b1) Zhang.R (C1) Iizuka (E1) Ours (D1) Original color image

Figure 8: The results of the difference model for the colorization of COCO image collection

Table 2: Comparison of MSSIM Indicators in Different Methods in Figure 9

Zhang.R	Iizuka	ours
Fig.8（b1）,0.8217	Fig.8（c1）,0.8887	Fig.8（d1）,0.9002
Fig.8（b2）,0.6587	Fig.8（c2）,0.6084	Fig.8（d2）,0.6908
Fig.8（b3）,0.6634	Fig.8（c3）,0.5183	Fig.8（d3）,0.7022

5. Conclusion

This paper proposes a fast near-infrared image colorization model composed of a lightweight image recognition network module and an image colorization CNN module with a fusion layer. Our method can successfully colored images of a certain type of scene, especially natural scenery and animal images. Because the color of the near-infrared image is highly dependent on the similarity of the image training set, therefore, for images with particularly complex scenes and non-unique colors, the colorization effect is not good. As shown in Figure 9. Our follow-up work is to train a more semantic classifier model and minimize the time required, In order to better adapt to the requirements of the actual application.

Figure 9: The failures by using the colorization model in our paper

6. Job Prospects

There are still many problems that require us to analyze and study in depth, mainly in the following aspects:

（1）Parameter setting problem. In general, deep learning methods involve parameter setting issues, such as the choice of weight attenuation factors

（2）Improvements to the fusion layer. The fusion method used in the colorization network in this paper is a simple linear fusion, which will be designed in a more rigorous fusion method.

（3）Optimize the classifier model. Our method can successfully color the image of a certain scene, especially outdoor scenes, portraits and animals, but since the near-infrared image is colorless, the image quality after colorization is highly dependent on the training set image content, so it is especially special for the scene. The color rendering of complex and object color non-unique images still needs to be improved. Our follow-up work is to optimize the classifier model and train the more semantic classifier model.

References

[1] Reinhard E, Ashikhmin M, Gooch B, et al. Color Transfer between Images[J]. IEEE Computer Graphics & Applications, 2002, 21(5):34-41.

[2] Welsh T, Ashikhmin M, Mueller K. Transferring color to greyscale images[J]. Acm Trans Graphics, 2002, 21(3):277-280.

[3] Liu X P, Wan L, Qu Y G, et al. Intrinsic colorization[J]. ACM Transactions on Graphics, 2008, 27(5): Article No. 152

[4] Xiao X, Ma L. Gradient-Preserving Color Transfer[C]// Computer Graphics Forum. 2009:1879–1886.

[5]Levin A, Lischinski D, Weiss Y. Colorization using optimization[J]. ACM Transactions on Graphics, 2004, 23(3): 689-694

[6] Yatziv L, Sapiro G. Fast image and video colorization using chrominance blending[J]. IEEE Transactions on Image Processing, 2006, 15(5): 1120-1129

[7] Luan Q, Wen F, Cohen-Or D, et al. Natural image colorization[C] //Proceedings of the 18th Eurographic Conference on Rendering Techniques. Aire-la-Ville: Eurographics Association Press, 2007: 309-320

[8]Zhang R, Isola P, Efros.A A.Colorful Image Colorization[C]// European Conference on Computer Vision. Springer, Cham, 2016:649-666.

[9] Iizuka S, Simoserra 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):1-11.

[10] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//. IEEE Conference on Computer Vision and Pattern Recognition.IEEE Computer Society, 2014:580-587. 6:649-666.

[11] Girshick R. Fast R-CNN[C]// IEEE International Conference on Computer Vision.IEEE Computer Society, 2015:1440-1448.

[12] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]// International Conference on Neural Information Processing Systems. MIT Press, 2015:91-99.

[13]Bhat P, Zitnick C L, Cohen M, et al. GradientShop: A gradient-domain optimization framework for image and video filtering[J]. Acm Transactions on Graphics, 2010, 29(2):10.

[14] Sheng B, Sun H, Chen S, et al. Colorization Using the Rotation-Invariant Feature Space[J].IEEE Computer Graphics & Applications, 2011, 31(2):24-35.