Deep Learning of Color Constancy Based on Object Recognition

Hong-yu Zhang
School of Information Science and Engineering
Dalian Polytechnic University
Dalian, China
Universezhang0427@163.com

Yuan Fang*
School of Engineering Practice and Innovation-Entrepreneurship Education
Dalian Polytechnic University
fangy@dlpu.edu.cn

Jia-huan Wu
School of Information Science and Engineering
Dalian Polytechnic University
Dalian, China
fantianzhe0424@163.com

Wei-zhen Wang
Human Factors and Intelligent Design Research Center
Dalian Polytechnic University
Dalian, China
wz-wang@foxmail.com

Nian-yu Zou
School of Information Science and Engineering
Dalian Polytechnic University
Dalian, China
n_y_zou@dlpu.edu.cn

2023 15th International Conference on Computer Research and Development (ICCRD) | 978-1-6654-8750-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCRD56364.2023.10080343

***Abstract*—Color constancy is the ability of human beings to recognize the colors of objects independently of the characteristics of the light source. Computational color constancy aims to estimate the illuminant and subsequently use this information to correct the image and display how it would appear under a canonical illuminant. The deep learning method is among the most successful illumination estimation methods to date and typically relies on a training set of images labeled with the respective scene illuminant. Although the human visual system is often compared to a machine learning algorithm, during evolution it was never presented with ground truth illuminants. Instead, it is hypothesized that the ability of color constancy arose because it helped other crucial tasks, such as recognizing fruits, objects, and animals independently of the scene illuminant. With the development of science and the improvement of people's quality of life, the field of artificial intelligence has developed rapidly, and the progress in image recognition has been even more rapid in recent years. This paper studies object detection in low lighting environments and uses deep learning algorithms to detect and analyze images. In low lighting environments, the detected objects are compared with the big data to obtain the objects with the closest similarity for identification and confirmation and compare the accuracy of the learning model for object recognition in complex lighting environments.**

Keywords—computational color constancy, object recognition, deep learning

I. INTRODUCTION

Color constancy is the perception that the color of an object's surface remains the same when the color light illuminating the surface changes [1]. Color constancy is an important component of machine vision [2]. Color constancy is a type of perceptual constancy. It is a perceptual property in which an individual's color perception of a familiar object tends to remain consistent when the color of the object changes due to changing light conditions [3]. Without computed color constancy, the color would be an unreliable feature and inconsistent for target recognition, detection, and tracking. Hence, for the study of color constancy, also known as light estimation [4].

Computer image processing and recognition is an important aspect of computer application technology and occupies an extremely important position in many fields such as the electronics industry, artificial intelligence, automation, and medical engineering [5], including the now-familiar face recognition [6][7]. In recent years, the study of computer image processing and recognition has received extensive attention [8][9]. Image recognition is mainly the study of using computers to automatically process large amounts of physical information instead of people, directly helping people to identify information [10]. Conventional feature design is performed manually and therefore takes time and effort to properly design the function compared with automatic feature extraction using deep learning [11]. In 1987, Alexander Waibel [12] proposed a time-delay network, which was the first convolutional neural network (CNN), and in 1988, Wei Zhang [13] proposed the first two-dimensional convolutional neural network, a translation-invariant artificial neural network. CNN has been widely cited in various industries, with the most widely used area being image processing. It is a model that is used to classify images, group them by similarity and perform object recognition within scenes [14].

Although VGGNet [15] is slightly weaker than GoogLeNet [16] (ILSVRC-2014 winner) in terms of classification power, its network structure is not as complex as GoogLeNet. VGGNet is one of the most popular choices for deep learning and computer vision tasks. Color constancy can solve the problem of the low recognition rate of images in a dark environment. In this paper, we adjust the color constancy of images through the PCA method in AlexNet and do image training using three models to test the recognition rate of images after enhancing color constancy.

II. PROPOSED METHOD

Simonyan and Zisserman proposed the VGG neural network model in the Department of Science Engineering at Oxford Street University, in which the structure is configured differently depending on the size of the convolutional kernels and the number of convolutional layers. The model is divided into five blocks (blocks), and all convolutional layers are made with 3 × 3 sized convolutional kernels, with the number of kernels in each block doubling as the model gets deeper. The image is passed through a down-sampling layer (pooling layer, maxpool) after each block, and finally through three fully-connected layers to reach the Softmax layer for classification output. Compared to other deep convolutional neural networks, the structure of VGGNet has better processing power for training datasets with smaller structures and data volumes. It is also easier to implement and has better recognition rates.

The convolutional and fully connected layers have weight coefficients and are therefore also called weight layers, with a total number of 13 + 3 = 16, which is the source of 16 in VGG16 [17]. The VGG network model is a cropped segmentation of the images, for different sizes of images cropped to an image size of 224 × 224 × 3. After the convolutional layer the image data is subjected to feature extraction. After the feature extraction in the convolutional layer, the output feature map is passed to the pooling layer for feature selection and information filtering. The number of channels is doubled, increasing sequentially from 64 to 128, then to 256, until 512 remains constant and is not doubled again, while the height and width are gradually halved [18].

VGG-16 uses the feature extraction part to perform feature extraction on the input image. Good features make it easier to identify targets. Therefore, the feature extraction part consists of five convolutional blocks [19]. The convolutional algorithm excels in feature extraction. The classification part will classify based on the features captured in the feature extraction part. The classification part usually consists of fully connected layers. The features captured in the feature extraction part are usually one-dimensional vectors that can be directly applied to fully connected classification [20].

However, the drawbacks of the VGG model are obvious; the VGG model is very slow to train, with the original VGG model taking up to 2-3 weeks to train 10,000 images on an Nvidia Titan GPU. The later improved VGG-16 model, although significantly improved in time, reducing the training time to around four to six hours, trains ImageNet weights of 528MB in size, taking up a large amount of disk space and bandwidth, making it inefficient [21].

A. Improvement Based on the VGG Model

In this paper, the 13 convolutional layers, 3 fully connected layers and 5 pooling layers of VGG-16 were changed, and the network model was modified by adding fully connected layers and pooling layers in combination with AlexNet to finally generate the new network model [22]. The model is shown in Figure 1.

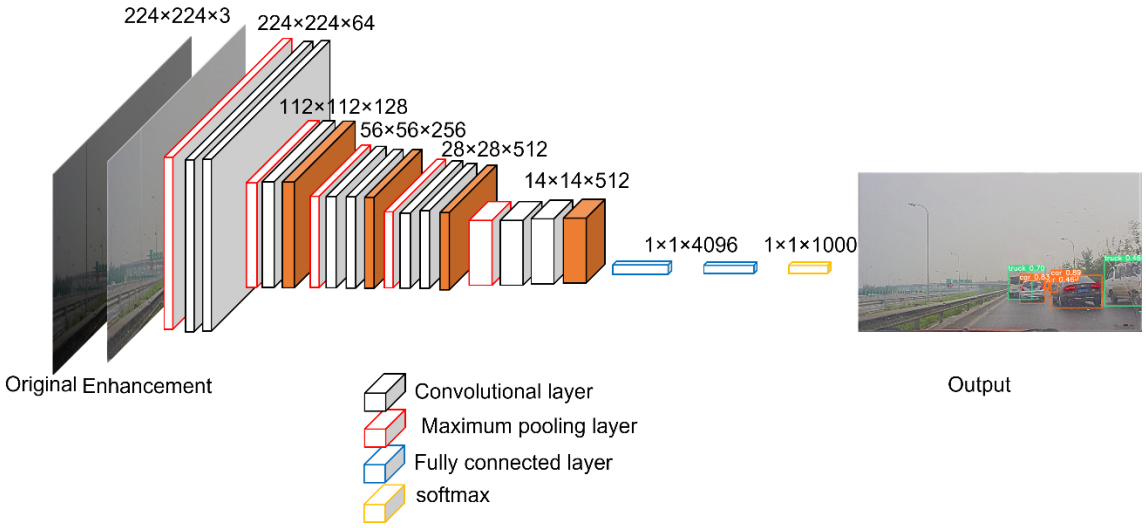


Fig. 1. Improved VGG model

The model was modified based on the previous VGG model and the best optimization algorithm, VGG-16, which is itself based on AlexNet. Compared to AlexNet, its structure is not very clear. In the convolutional layer, it is better to stack more 3×3 convolutional blocks than 5×5 convolutional blocks. The convolution kernel aspect is calculated as Eqs. (1) and (2).

$$H_{out} = \left\lfloor \frac{H_{in} + 2 \times padding[0] - dilation[0] \times (kernel_{size[0]} - 1) - 1}{stride[0]} + 1 \right\rfloor \quad (1)$$

$$W_{out} = \left\lfloor \frac{W_{in} + 2 \times padding[1] - dilation[1] \times (kernel_{size[1]} - 1) - 1}{stride[1]} + 1 \right\rfloor \quad (2)$$

H_{in} is the input image size. $\text{kernel}_{\text{size}}$ is the size of the convolution kernel. stride is the span of the convolution and the default value is 1. Padding is the padding value and the default value is 0. Dilation is the spacing between the kernel elements. The default value is 1. stride is the step size.

Finally the Soft-max function is used to calculate the classification probability, the formula is shown in equation (3).

$$Softmax(z_i) = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_c}} \tag{3}$$

z is a vector and Z_i and Z_c are an element of it. Where Z_i is the output value of the I node and c is the number of output nodes, the number of categories to be classified. The Soft-max function converts the output values of multiple categories into a probability distribution ranging from [0, 1] and 1.

B. Image Enhancement

The image enhancement module is based on AlexNet, and the data shows that AlexNet performs well in terms of color constancy. The PCA-based color enhancement method is described in AlexNet, the effect of which is that there is an overall change in the brightness (brightness) of the image and no significant change in the structure of the image or chromatic aberration occurs [23]. The rendering is shown in Figure 2.

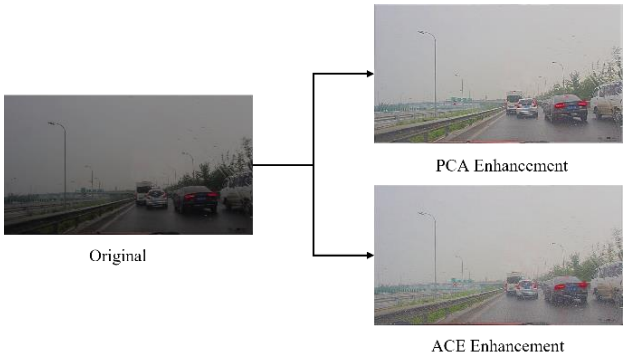


Fig. 2. Image enhancement effect

The two images in Figure 2 are each the effect after the original image has been dithered. The PCA method is compared with the ACE method in Figure 2. Automatic color balancing algorithm is proposed on the theory of Retinex algorithm. Compared with ACE enhancement, PCA obviously improves the color channel and makes the image display clearer without losing the image authenticity. It corrects the final pixel value by calculating the brightness degree and the relationship between the target pixel point and the surrounding pixel point, realizes image contrast adjustment, and generates the balance of color constancy and brightness constancy similar to that of human retina. It has good image enhancement effect [24]. The brightness of the images in Figure 2 changed significantly, the outlines of the main things in the images became very clear and the dominant colors of the things did not change; the relative color difference of the images did not change [25]. The image enhancement steps as follows:

(1) Firstly, the image is normalised according to the RGB three channels, with a mean of 0 and a variance of 1. This is done according to the RGB three channels, because we are enhancing the color, and in the RGB three channel image, the relative relationship between RGB determines the color of the image, and we cannot change the distribution of pixel values within the three channels.

(2) Flatten the image by channel into an array of size (n,3).

(3) Find the covariance matrix of the above array.

(4) Decompose the covariance matrix.

The matrix of eigenvectors is 3×3, and the array of eigenvalues multiplied by the dithering coefficients is 3×1, so the array obtained by dot multiplication is exactly 3×1 in size, and the three values are added to the R,G,B channels of the original image, which is the final enhanced image. The feature vector matrix is shown in Equation 4, where pi and λ_i are i eigenvector and eigenvalue of the 3×3 covariance matrix of RGB pixel values, respectively, and $\alpha_i$is the random variable. Each α_i is drawn only once for all the pixels of a particular training image until that image is used for training again, at which point it is re-drawn. This scheme approximately captures an important property of natural images, namely, that object identity is invariant to changes in the intensity and color of the illumination. [26] This method is a way to balance the RGB base color. This is a color distortion or color transformation method that will make your learning algorithm more robust to color changes in your photos.

$$Array = [p_1\ p_2\ p_3] \quad [\alpha_1\lambda_1\ \alpha_2\lambda_2\ \alpha_3\lambda_3]^T \tag{4}$$

III. Experiment

In this paper, we use Pytorch to construct the network structure and compare the modified VGG model with the VGG-16 model and the VGG-19 model to test the original and enhanced images to derive the model loss rates and accuracy. Experimentally, 2000 different images of driving record picture were selected for training to compare the accuracy of the three models. The images were color enhanced by PCA and the classification training was compared to the enhanced images to derive the optimal classification.

TABLE I. Improved VGG Model Experiment Setting

Model Experiment Setting	
Name	***Model***
Processor	Intel(R) Core(TM)i711800H@2.30GHz
RAM	16GB
Graphics Cards	NVIDIA GeForce RTX 3060
Primary hard drive	BC711 NVMe sj hynix 512

A. Experimental Procedure

This research builds the model and training set by VGG, constructs the pre-classification network model with Pytorch as the framework, constructs the image enhancement model of PCA based on AlexNet. The original image and the enhanced image were placed in two different files, and the improved VGG was used to recognize and judge the two types of images, and the recognition accuracy was collected and compared. Finally, the experimental results are obtained.

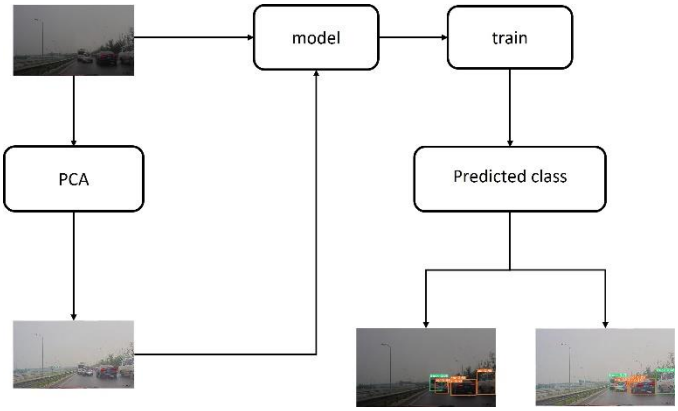


Fig. 3. Flow chart of the experiment

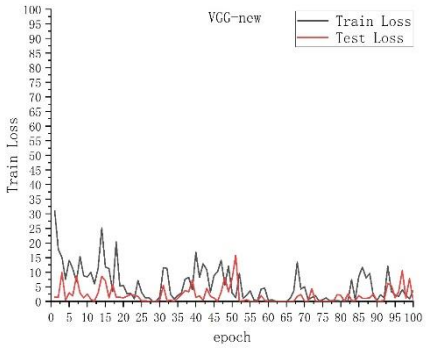


Fig. 4. Training and testing loss values for the three models

B. Result

Figure 4 shows, the training and test loss plots for VGG-19, VGG-16 and VGG-new respectively. After 100 iterations, the training loss value of VGG-19 is significantly higher than that of VGG-16, while the improved VGG-16 model can well control the training loss value within 30 and the test loss value is all below 10.

In Figure 5, the test accuracy of the different models can be seen. after 100 iterations, the accuracy of all three models was above 70%. the accuracy of the model of VGG-19 was stable at 80% after 100 iterations, on the contrary, although the accuracy of VGG-16 did not show a significant decrease after 100 iterations, it did not stabilise either. the improved VGG-16 model in this classification Accuracy reached 100% in 30 iterations but did not stabilise after 100 iterations.

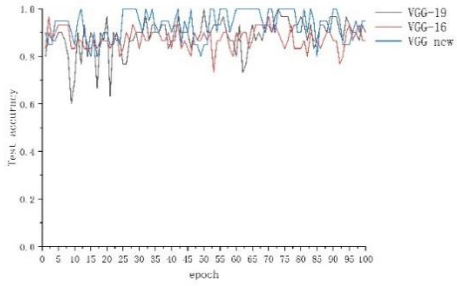
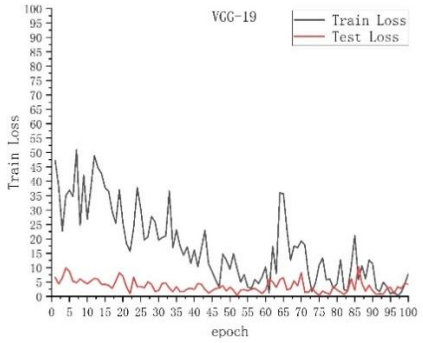


Fig. 5. Accuracy of different models tested



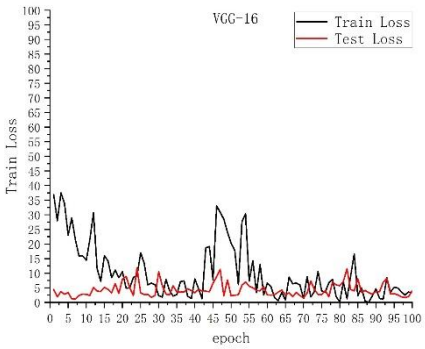


Figure 6 shows the original image and the enhanced image. In the original image, due to the influence of rainy days, the visibility presented by the picture is not clear, so the identification of vehicles is not accurate enough. After color constancy by PCA the images were clearly visible and models were able to identify cars successfully.

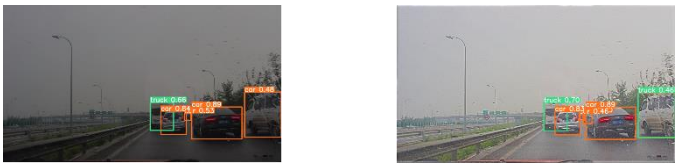


Fig. 6. Contrast the original image with the enhanced image

IV. Conclusions

In this research, the images were classified and predicted by the improved VGG model. For the images, we performed a PCA-based color enhancement method, and the color-enhanced images were imported into the three models for prediction, and the optimal model was finally compared. In this research, the improved VGG-16 algorithm was not as strong as the original algorithm, but it did not fully reflect the strength of the three models for many reasons, such as the small image repository and the small number of iterations. Although the improved VGG-16 has a better fit than VGG-19, the actual running test is slower, and after adding convolution layers, the increase in the number of convolutions will instead slow down for model training, so

the model is still not perfect. With the color enhancement of PCA, the image recognition rate of the three models was greatly improved. In today's image enhancement field, there are many other methods that can perform effective color enhancement, and there are also many denoising methods that can improve the image recognition rate. This paper is relatively single for network models, and different network models can be tested subsequently. Although PCA has a strong capability in image enhancement, there is no denoising capability, which is still not perfect for image enhancement.

ACKNOWLEDGMENT

This research was funded in part by Liaoning Natural Science Foundation, China (2022-BS-263); Liaoning Education Science Project (JG21DB054); Key Scientific Research Projects (LJKZZ20220069); Dalian Polytechnic University Reform of Education Project Foundation (JGLX2021026); Humanity and Social Science Foundation of Ministry of Education of China (21YJAZH088); University-Industry Collaborative Education Program (202102092001); Liaoning Provincial Department of Education Project (1010152); China National Textile And Apparel Council (2021BKJGLX321). We would like to express our heartfelt thanks.

REFERENCES

[1] Buzzelli M, van de Weijer J, Schettini R. Learning illuminant estimation from object recognition[C]//2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 3234-3238.

[2] Laakom F, Raitoharju J, Nikkanen J, et al. Intel-tau: A color constancy dataset[J]. IEEE Access, 2021, 9: 39560-39567.

[3] Yu H, Chen K, Wang K, et al. Cascading convolutional color constancy[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(07): 12725-12732.

[4] Afifi M, Price B, Cohen S, et al. When color constancy goes wrong: Correcting improperly white-balanced images[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 1535-1544.

[5] Yordanka Karayaneva and Diana Hintea, "Object Recognition in Python and MNIST Dataset Modification and Recognition with Five Machine Learning Classifiers," Journal of Image and Graphics, Vol. 6, No. 1, pp. 10-20 June 2018. doi: 10.18178/joig.6.1.10-20

[6] Mehmet Korkmaz and Nihat Yilmaz, "Face Recognition by Using Back Propagation Artificial Neural Network and Windowing Method," Journal of Image and Graphics, Vol. 4, No. 1, pp. 15-19, June 2016. doi: 10.18178/joig.4.1.15-19

[7] Haifeng Zhang and Shenjie Xu, "The Face Recognition Algorithms Based on Weighted LTP," Journal of Image and Graphics, Vol. 4, No. 1, pp. 11-14, June 2016. doi: 10.18178/joig.4.1.11-14

[8] P. Kasemsumran, S. Auephanwiriyakul, and N. Theera-Umpon, "Face Recognition Using String Grammar Nearest Neighbor Technique," Journal of Image and Graphics, Vol. 3, No. 1, pp. 6-10, June 2015. doi: 10.18178/joig.3.1.6-10

[9] Samuel Lukas, Aditya Rama Mitra, Ririn Ikana Desanti, and Dion Krisnadi, "Implementing Discrete Wavelet and Discrete Cosine Transform with Radial Basis Function Neural Network in Facial Image Recognition," Journal of Image and Graphics, Vol. 4, No. 1, pp. 6-10, June 2016. doi: 10.18178/joig.4.1.6-10

[10] Muhammad Hassan, Tasweer Ahmad, Nudrat Liaqat, Ali Farooq, Syed Asghar Ali, and Syed Rizwan hassan, "A Review on Human Actions Recognition Using Vision Based Techniques," Journal of Image and Graphics, Vol. 2, No. 1, pp. 28-32, June 2014. doi: 10.12720/joig.2.1.28-32

[11] Ryo Hasegawa, Yutaro Iwamoto, and Yen-Wei Chen, "Robust Japanese Road Sign Detection and Recognition in Complex Scenes Using Convolutional Neural Networks," Journal of Image and Graphics, Vol. 8, No. 3, pp. 59-66, September 2020. doi: 10.18178/joig.8.3.59-66

[12] Waibel A, Hanazawa T, Hinton G, et al. Phoneme recognition using time-delay neural networks[J]. IEEE transactions on acoustics, speech, and signal processing, 1989, 37(3): 328-339.

[13] Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition, 2018, 77: 354-377.

[14] Washington García, Cristian Mera, Leonel Santana, and Luzmila Pro, "Algorithm for the Recognition of a Silhouette of a Person from an Image," Journal of Image and Graphics, Vol. 7, No. 2, pp. 59-63, June 2019. doi: 10.18178/joig.7.2.59-63

[15] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[16] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[17] Wang J, Yuan C. Facial expression recognition with multi-scale convolution neural network[C]//Pacific Rim Conference on Multimedia. Springer, Cham, 2016: 376-385.

[18] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[19] Tammina S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images[J]. International Journal of Scientific and Research Publications (IJSRP), 2019, 9(10): 143-150.

[20] Huang S, Fan X, Sun L, et al. Research on classification method of maize seed defect based on machine vision[J]. Journal of Sensors, 2019, 2019.

[21] Ren R, Zhang S, Sun H, et al. Research on pepper external quality detection based on transfer learning integrated with convolutional neural network[J]. Sensors, 2021, 21(16): 5305.

[22] Bianco S, Buzzelli M, Schettini R. Multiscale fully convolutional network for image saliency[J]. Journal of Electronic Imaging, 2018, 27(5): 051221.

[23] Bianco S, Schettini R. Color constancy using faces[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 65-72.

[24] S. Bidon, Olivier Besson, J. Y. Tourneret. The Adaptive Coherence Estimator is the Generalized Likelihood Ratio Test for a Class of Heterogeneous Environments[J]. IEEE Signal Processing Letters, 2008, 15: 281-284.

[25] Barron J T. Convolutional color constancy[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 379-387.

[26] Cheng D, Prasad D K, Brown M S. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution[J]. JOSA A, 2014, 31(5): 1049-1058.