Low-light image enhancement based on deep learning: a survey

Yong Wang,[a],* Wenjie Xie,[a] and Hongqi Liu[b]
[a]Jilin University, College of Communication Engineering, Changchun, China
[b]The Sixth Academy of China Aerospace Science and Industry Corporation, The 602st Institute, Hohhot, China

Abstract. Images taken under low light or dim backlight conditions usually have insufficient brightness, low contrast, and poor visual quality of the image, which leads to increased difficulty in computer vision and human recognition of images. Therefore, low illumination enhancement is very important in computer vision applications. We mainly provide an overview of existing deep learning enhancement algorithms in the low-light field. First, a brief overview of the traditional enhancement algorithms used in early low-light images is given. Then, according to the neural network structure used in deep learning and its learning algorithm, the enhancement methods are introduced. In addition, the datasets and common performance indicators used in the deep learning enhancement technology are introduced. Finally, the problems and future development of the deep learning enhancement method for low-light images are described. © 2022 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.OE.61.4.040901]

1 Introduction

In daily life, some fields require high-quality and clear images, such as biomedicine, aerospace, transportation, military, etc. But sometimes, the shooting will be affected by the environment or the equipment used, resulting in the inability to capture clear and high-quality images. For example, when in a backlight, night, or dim indoor and other special environment, even the images taken with standard equipment will appear blurry and have low brightness, loss of detail, high noise, poor visual quality, etc. Therefore, for low-light and dim images, image enhancement is very important. In the early days, low-light images were usually enhanced with traditional enhancement algorithms. However, with the wide application of deep learning in various fields, many deep learning enhancement techniques have been proposed for low-light images.

The main contributions of this article are as follows:

- The deep learning enhancement methods for low-light images are summarized, and the deep learning enhancement methods are respectively introduced according to the neural network structure and learning algorithm used in deep learning.
- A brief overview of the datasets and common evaluation indicators used in the deep learning enhancement method for low-light images.

The other parts of the article are Sec. 2 briefly outlines the traditional enhancement algorithms commonly used in low-light images. Sections 3 and 4 introduce the deep learning enhancement methods for low-light images. Sections 5 and 6 introduce the datasets and commonly used evaluation indicators in the deep learning enhancement methods. Section 6.1.4 briefly introduces other applications of deep learning in computer vision. The last section describes the problems and development of deep learning enhancement methods for low-light images.

---

*Address all correspondence to Yong Wang, Wyong82@jlu.edu.cn

2 Traditional Enhancement Algorithms for Low-Light Images

For images with insufficient illumination and low contrast, many enhancement algorithms have emerged, among which the more widely used traditional image enhancement algorithms are histogram equalization (HE) algorithm, algorithm based on physical models (atmospheric scattering model or Retinex model). In this section, the traditional enhancement algorithm is briefly introduced.

2.1 Histogram Equalization Algorithm

When enhancing images with insufficient illumination and low contrast, the simpler method is the HE algorithm,[1] which has a simple principle and a faster processing speed. The HE algorithm is a processing method of enhancing the image through the image gray histogram. When the grayscale histogram of the image is concentrated in a grayscale interval, the concentrated grayscale interval is stretched to the entire grayscale interval of the histogram through transformation, the grayscale range of the image is expanded and distributed uniformly, and the contrast is improved. However, after processing the image through HE, there will be problems such as the loss of detailed information due to gray level merging. In response to the problems in the HE algorithm, a series of improved algorithms have appeared one after another, such as adaptive HE,[2] double histogram equalization,[3] maximum brightness double histogram equalization,[4] etc.

2.2 Algorithm Based on Physical Model

The most commonly used low-light enhancement algorithms based on physical models are the atmospheric scattering model[5,6] and the Retinex model.[7-13] Dong et al.[5] found through experiments that the inversion image of low-light image is very similar to the foggy image, so they proposed an algorithm to enhance the low-illuminance image using the atmospheric scattering model. First, invert the low-light image, then apply the defogging algorithm proposed by He et al.,[14] and finally the processed image is inverted to obtain an enhanced image. Land et al.[15] proposed a Retinex theory based on the human visual system. The basic assumption of the theoretical model is that the illumination component I and the reflectivity component R work together to form the original image S seen by the human eye. The principle of the enhancement algorithm based on Retinex theory is to obtain the illumination component I by decomposing the original image S, then remove the influence of I, and finally get the enhanced result.

2.3 Retinex Model

Land et al.[16] proposed a color constancy theory based on the brightness and color perception of the human visual system in 1963, namely the Retinex theory. Retinex is a combination of the words retina and cortex. Through experiments, Land et al. found that the intensity of light irradiated on the object does not determine the color of the object but is determined by the nature of the object's reflection. Therefore, the unevenness of the illumination will not affect the color of the object seen by the human eye, which is the color constancy. After proposing the Retinex theory, a variety of Retinex algorithms have been improved and developed successively: single-scale Retinex algorithm,[17] multiscale Retinex algorithm,[18] etc. SSR and MSR are introduced below.

2.3.1 Single-scale Retinex

The light component (incident light) I and the reflected component R work together to form the image observed by the human eye. The incident light shines on the object, and then the reflected light is reflected by the object and enters the human eye to form the image seen by the human eye. The equation is

$$S(x, y) = R(x, y) \bullet I(x, y). \tag{1}$$

The light component is $I(x, y)$, the reflection property of the object is $R(x, y)$, and the image seen by the human eye is $S(x, y)$.

The principle of the enhancement algorithm of Retinex theory is to decompose the image S observed by the human eye to obtain the illumination component I, and then remove or reduce the influence of I, and the obtained reflection component R is the enhanced result. Therefore, to decompose and obtain the reflection component $R(x, y)$, it is necessary to take the logarithm of both sides of Eq. (1) and change the product relationship into an addition and subtraction relationship, that is,

$$\log(S(x, y)) = \log(R(x, y)) + \log(I(x, y)). \tag{2}$$

In the Retinex theory, the illumination component changes slowly and belongs to the low frequency component, and the reflection component belongs to the high frequency component. It is calculated by approximate estimation when solving the illumination component $I(x, y)$, as shown in Eq. (3):

$$\log(I(x, y)) = \log[F(x, y) * S(x, y)], \tag{3}$$

$$\log(R(x, y)) = \log(S(x, y)) - \log[F(x, y) * S(x, y)], \tag{4}$$

where $F(x, y)$ is the convolution kernel and c is the Gauss surround scale:

$$F(x, y) = K \exp\left[\frac{-(x^2 + y^2)}{c^2}\right]. \tag{5}$$

2.3.2 *Multiscale Retinex*

In response to the problems in SSR, Jobson et al.[18] proposed a multiscale Retinex algorithm on the basis of SSR. The equation is as shown:

$$R(x, y) = \sum_{n=1}^{N} w_n \{\log(S(x, y)) - \log[F_n(x, y) * S(x, y)]\}, \tag{6}$$

where k is the number of the convolution kernel. When the number $K = 1$, MSR is SSR.

3 Deep Learning Enhancement Method for Low-Light Images

In recent years, artificial intelligence technology has developed more and more rapidly, and its application in various fields have become more extensive, such as smart home, voice recognition, image recognition, autonomous driving, and VR. Most of the applications of artificial intelligence in these fields use machine learning technology, and deep learning is also widely used as a subcategory of machine learning. Among them, machine learning usually requires manual extraction of features, whereas deep learning mainly uses multilayer nonlinear processing units for feature extraction and conversion.

Hinton and Salakhutdinov[19] proposed a deep neural network on the basis of traditional neural networks. Krizhevsky et al.[20] proposed AlexNet, which made deep learning arouse widespread attention and was formally applied to industry. The concept of deep learning is derived from artificial neural networks, whose basic unit is artificial neurons. A deep neural network is a multilayer neural network as its name implies. The deep neural network is composed of an input layer, two or more hidden layers and an output layer. Each layer is composed of several neurons, and the input of each layer is the output of the previous layer.

As shown in Fig. 1. The hidden layer is the layer between the input layer and the output layer.

On this basis, various deep learning enhancement methods have been developed for low-light images. These methods are based on four network types: autoencoder (AE), convolutional neural network (CNN), recurrent neural network (RNN), and generative adversarial network (GAN).
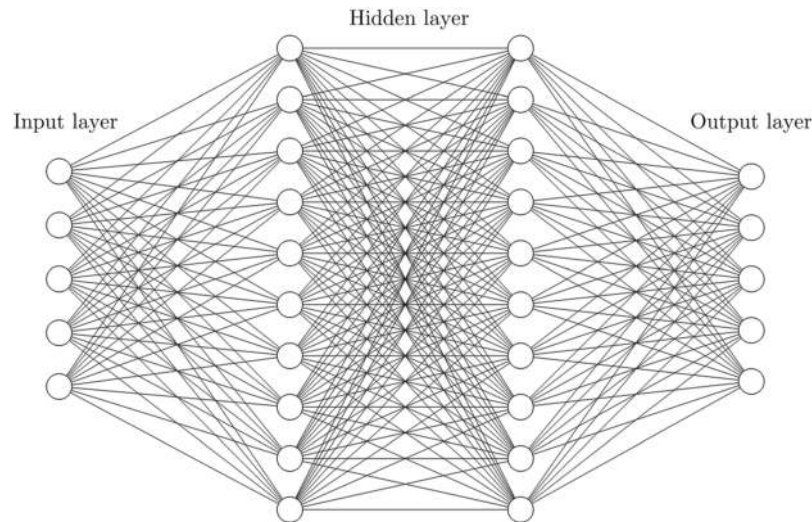
Fig. 1 Deep neural network structure.

However, there are some problems in the application of neural networks. Therefore, before formally introducing the deep learning enhancement method for low-light images, some basic concepts and existing problems of deep learning are briefly introduced.

3.1 Preliminaries on Deep Learning

3.1.1 Adversarial perturbations and adversarial examples

Adversarial perturbation is a small and imperceptible perturbation in the data sample, which is generally divided into two types: universal perturbation and image/model-dependent perturbation.[21] The universal perturbation[22] is not for a specific image, but for any image, that is, the perturbation has nothing to do with the image. The image/model dependent perturbation refers to the perturbation associated with the image/model, that is, the perturbation will be different for different images/models. Szegedy et al.[23] defined the samples subjected to small and imperceptible disturbances in these data sets as adversarial samples, and the input of these samples would lead to output errors of the model with high probability. Therefore, the vulnerability of network models to adversarial samples is one of the major challenges of deep learning. In response to this problem, a method is proposed to improve the anti-interference ability of the model using adversarial samples in the network training process, namely adversarial training. The addition of such adversarial samples can not only effectively avoid the potential security problems of deep learning in practical applications but also help improve the robustness and accuracy of the network model.

3.1.2 Stability/robustness

Robustness refers to the ability of the network to resist interference and maintain a certain performance in the presence of interference, that is, when the input information or the network is disturbed, the network model can still keep the output stable. Because of the excellent performance of deep learning, deep networks are widely used in various fields, but these networks are easily affected by some minor disturbances, which lead to unstable network output and damage the reliability and robustness of network models. In response to this problem, Giryes et al.[24] studied the performance of DNNs with random weights. Malladi and Sharapov[25] proposed an improved weight normalization method. Zheng et al.[26] proposed a training method that adds a stability term to the objective function, which makes the network more robust to small perturbations in the input and more stable output.

3.1.3 *Noisy labels*

The training data used in the network training process will inevitably be mislabeled. These mislabeled labels are called noisy labels and make the model performance suffer. Label noise is generally caused by the low quality of the obtained data labeling or mislabeling during the labeling process. When using a dataset with noisy labels for training, the network model will overfit to the noisy samples, resulting in poor generalization performance. Therefore, eliminating the adverse effects of label noise on the network is currently another research problem in deep learning. In response to this problem, Algan and Ulusoy[27] introduced the processing methods for noise labels in deep learning and divided them into noise-based models and noise-free models. Thekumparampil et al.[28] proposed two conditional GAN architectures, depending on whether the distribution of noise is known or not, making it more robust to noise labels in the training data.

3.1.4 *Interpretability/understanding*

Interpretability is understanding the structure of a system in a human-understandable term (knowledge of the relevant domain, human cognition, etc.). With the widespread application of deep learning in various fields, researchers have begun to pay more and more attention to how models accomplish tasks. However, because the network model is equivalent to a black box, the internal working mechanism of the network is not clear. Therefore, in the medical, military, financial, and other fields, allowing users to understand the decision-making process of the model in more detail will make users trust the product. Therefore, studying the interpretability of network models is of great significance for deep learning. For example, Zhang et al.[29] expounded the importance of interpretability in terms of reliability and morality and proposed a new interpretability classification method.

3.1.5 *Provability*

The training process of the neural network is a process of solving optimization. With the successful application of deep learning in various fields, the structure of neural network becomes more and more complex, and the ability to fit the model becomes more and more powerful. But this can lead to overfitting of the model, and it is also more difficult to optimize the model. Therefore, to avoid the phenomenon of neural network overfitting, methods such as adding data and adding regularization terms can be used. Vidal et al.[30] provided mathematical proofs for properties, such as global optimality of network models from three aspects: deep learning architecture, regularization techniques, and optimization algorithms. Yun et al.[31] studied global optimality in deep learning.

3.2 *Autoencoder*

Hinton and Salakhutdinov[19] proposed AE and some concepts, which made AE receive extensive attention.

AE is an unsupervised learning algorithm. The autoencoder is a data compression algorithm, in which the encoder and the decoder together constitute the autoencoder, that is, it realizes the dimensionality reduction or feature learning of the data through the encoder and the decoder. The input data are compressed by the encoder into low-dimensional variables, and then the low-dimensional variables are reconstructed by the decoder to become the original dimensions of the input, as shown in Fig. 2. The AE uses the input data as supervision through optimization methods such as back propagation algorithm and guides the network to obtain the reconstructed output. When designing an autoencoder, the input and output are not designed to be completely equal, so some constraints are usually added to the autoencoder to make the input and the reconstructed output as equal as possible. Therefore, after adding some constraints, some new encoders are obtained: denoising autoencoder,[32] sparse autoencoder,[33] stacked autoencoder, variational autoencoder,[34] and contractive autoencoder, etc.
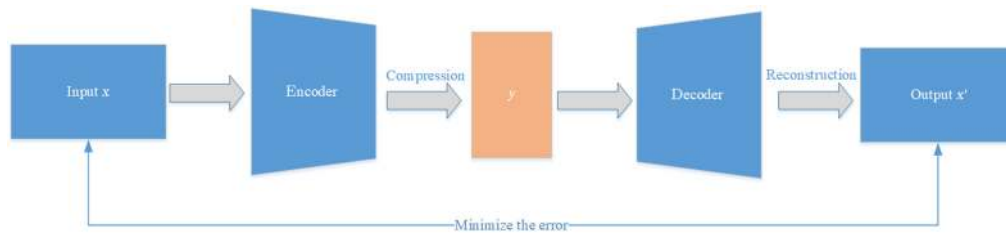
Fig. 2 AE structure.

Aiming at the problem of amplifying noise during the enhancement process, Lore et al.[35] proposed an enhancement method based on stacked sparse denoising autoencoder (SSDA),[36] called the LLNet, which is the first method to apply deep learning to low-light images. Two networks LLNet and S-LLNet are proposed in this method. LLNet is an image with low brightness and noise, and then the SSDA1 module is used for contrast enhancement and denoising, and finally an enhanced and denoised image is obtained. S-LLNet contains two independent modules: the SSDA1 module for contrast enhancement and the SSDA2 module for denoising. The enhanced and denoised images are obtained by inputting low-brightness and noisy images into SSDA1 module for contrast enhancement and SSDA2 module for denoising.

Park et al.[37] proposed a dual autoencoder network based on Retinex theory. The network consists of brightness estimation and reflection component estimation. First, the smoothed illumination component estimation is performed by the stacked autoencoder, and then the initial reflectance is obtained according to the Retinex theory. Then, the initial reflectance is denoised by the convolutional autoencoder, and finally the HSV channel is changed into RGB channel to obtain the final enhancement result.

3.3 Convolutional Neural Network

CNN is a feedforward neural network that includes convolution calculations and is one of the most commonly used networks in deep learning methods. The structure of the general CNN is shown in Fig. 3. The convolutional layer is to learn the features of the input data through the convolution kernel matrix. The activation function is a nonlinear mapping of the output of the convolutional layer. The pooling layer is mainly used for feature dimensionality reduction and data compression.

Fukushima[38] proposed a weight sharing convolutional neural layer and neocognitron based on the receptive field theory. LeCun et al.[39] proposed a CNN combining convolutional neural layers and backpropagation. CNN evolved from the multilayer perceptron. Its particularity mainly lies in the two aspects of weight sharing and local connection. Weight sharing reduces the weight parameters of CNN. Local connection means that each neuron only connects to some
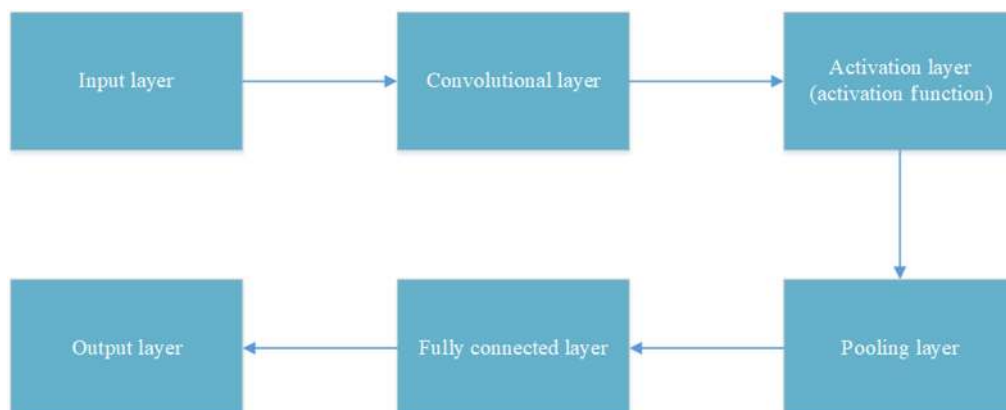

Fig. 3 CNN structure.

neurons in the previous layer and does not need to be connected to all neurons. This reduces the number of network weights and reduces model complexity.

Tao et al.[40] proposed an enhancement method based on CNN, called LLCNN. The LLCNN network structure is composed of two convolutional layers and multiple specially designed convolutional modules. This special module refers to residual learning[41] and inception[42] and proposes a dual-branch and residual learning structure. After inputting the image, the final enhancement result is obtained through these special modules and the last convolutional layer.

Shen et al.[43] proposed an enhanced network MSR-net based on MSR theory, which imitated the processing process of MSR theory. MSR-net is divided into three processing steps: multi-scale logarithmic transformation, convolution difference, and color restoration. After inputting the image, the final enhancement result is obtained through these three processing in turn.

Li et al.[44] proposed an enhancement method for illumination estimation based on Retinex theory, called LightenNet. After inputting the image, the four convolutional layers of the network are used to achieve feature enhancement, nonlinear mapping, and other operations to obtain the illumination component, and then the illumination component is enhanced through gamma correction, and finally the input image is divided by the illumination component according to Retinex theory to obtain the final enhancement result.

Cai et al.[45] proposed a CNN-based SICE method, which is divided into two stages and three networks. The first stage is to decompose the input image into low frequency information and high frequency information through weighted least squares filtering, and then the low and high frequency information are enhanced through the two networks, respectively. The second stage is to merge the two parts after enhancement, and then enhanced by a CNN network containing a BN layer (overall enhancement network), and finally the enhancement result is output.

Wei et al.[46] proposed a new enhancement method based on Retinex theory, which combines Retinex theory and deep neural network, called Retinex-Net. The network is divided into two subnetworks: decom-net and enhance-net. The decomposition network takes the low-light image and the normal image as input and then obtains the illumination component I and the reflection component R according to the Retinex theory and Decom-Net. The enhancement network enhances the decomposed illumination component and uses the BM3D[47] denoising algorithm to denoise the reflected component R and then multiply the enhanced illumination component I with the denoised R to get the final enhanced result.

Lv et al.[48] proposed a new CNN-based enhancement method, namely MELLEN, which consists of a feature extraction module (FEM), enhanced module (EM), and fusion module (FM). After inputting the image, first extract the features through the FEM, and each output of FEM is the input of the next layer and corresponding enhancement module, and then the corresponding input is enhanced by the enhancement module, and finally the output of all EM is multibranch fused through the FM to obtain the final enhanced result. Later, Lv et al.[49] proposed an attention-guided enhanced multibranch CNN based on the multibranch structure. The network consists of four modules: attention-net, noise-net, enhancement-net, and reinforce-net. Among them, attention-net is the structure of U-Net,[50] and its output ue-attention map (underexposed) guides image enhancement and Noise-Net denoising. The structure of enhancement-net is similar to MELLEN. The low-light image first passes through FEM, and then the result obtained is sent to EM together with ue-attention map and noise-map for enhancement, and then the different enhancement results are connected, and finally output the enhanced image through FM. Reinforce-Net further enhances the contrast of the Enhancement-Net output through dilated convolution and obtains the final enhancement result.

Chen et al.[51] proposed an enhancement method (SID) based on end-to-end training of FCN, which directly processes raw data. The image was taken by two cameras, the sensors of which were Bayer and X-Trans. First, the Bayer array is processed into four channels (the input is 6×6 X-Trans array are processed into nine channels), then the black level is subtracted and amplification ratio is multiplied for lighting, then input the processed data into the FCN. Finally, the sRGB spatial image of original size is obtained by upsampling, which is the final enhancement result. In addition, Chen et al.[52] extended the low-light static video enhancement method based on the SID method.

Wang et al.[53] proposed the GLADNet, which consists of global illumination estimation and detail reconstruction. First, the input image is downsampled and scaled to a certain resolution so

that the receptive field is large enough to perceive the global information, and then perform illumination prediction on the entire image, and finally the image is restored to the original input size through upsampling. Detail reconstruction is because detailed information will be lost during image scaling, so the input image is connected with the image after global illumination estimation, and the output is the final enhancement result.

Jiang et al.[54] proposed a refined network LL-RefineNet. It mainly extracts features through symmetrical convolutional layer structure and then the extracted high-resolution features are refined through four subnetworks, that is, perform multiscale feature fusion, and finally obtain high-resolution enhancement result.

Zhang et al.[55] proposed the KinD-Net. The design idea of this network is the same as Retinex-Net. It is still decomposed and enhanced, but on this basis, considering that the reflection map R is usually degraded under low light conditions, the function of illumination guided reflectance restoration and flexible mapping of arbitrary lighting operations is proposed. The network consists of a layer decomposition network, a reflectance recovery network, and a light adjustment network. The decomposition network decomposes the input image according to Retinex theory. The reflectance recovery network uses the reflectance R of the normal image as the ground truth and introduces the decomposed lighting information into the network to restore the reflectance. The illumination adjustment network: by adjusting the parameters, the illumination can be flexibly adjusted to obtain the desired enhancement result. Afterward, Zhang et al. proposed the KinD++ network[56] to solve the problems of excessive smoothing in KinD. In this network, a new module (MSIA) was proposed to alleviate the problems left in KinD.

Wang et al.[57] proposed a network (DeepUPE) that enhances underexposed images by estimating the mapping of the input image to the illumination map. The network structure is roughly the same as the HDR-Net proposed by Gharbi et al.[58] First, the input image is downsampled and local features and global features are extracted through the encoder network, and the local features and global features are combined to perform low-resolution illumination prediction, then through the bilateral grid[59] upsampling to get the full-resolution illumination map, and finally through the Retinex theory to calculate the final enhanced image.

Wang et al.[60] proposed an end-to-end enhancement network, which is composed of Retinex decomposition network (RDNet) and fusion enhancement network (FENet). After inputting the image, it is decomposed into illumination component and reflection component by RDNet, and then the decomposed illumination component is preliminarily enhanced by the camera response function.[13] Finally, the input image, the decomposed reflection component, and the preliminary enhanced illumination component are used as the input of FENet for fusion enhancement, and the final enhancement result is obtained.

Zhu et al.[61] proposed the low-light enhancement method of EEMEFN. The network framework is divided into multiexposure fusion (MEF) and edge enhancement (EE). The MEF stage first generates multiple exposure images through a given exposure ratio and then fuses the different scale information of the generated multiple exposure images through the U-net structure and the FM, and finally the initial image is generated through the 1×1 convolutional layer. EE is divided into two steps: detection and enhancement. The edge detection network proposed by Liu et al.[62] is used to extract the edge information, and then multiple exposed images, the initial image, and the obtained edge information are input into the enhancement module to obtain the final enhance result.

Fan et al.[63] combined Retinex theory with semantic information and proposed a semantic perception low-light enhancement network, which consists of three parts: information extraction, reflectivity enhancement, and illumination adjustment. Semantic information is extracted through semantic segmentation, and reflectivity is reconstructed through ReflectNet under the guidance of semantic information. Then use the restored reflectance and enhancement ratio to adjust the illumination through RelightNet and finally get the final enhancement result according to Retinex theory.

Lv et al.[64] proposed a lightweight CNN enhancement method. The network consists illumination-net, fusion-net, and restoration-net. Combining the original image, bright channel, and invert bright channel as input, first output the underexposure image and overexposure image through illumination-net, and then input the obtained underexposure image and overexposure image together with the original image into fusion-net for fusion. Multiply the weight of the

fusion-net output with the previous image and use it as the input of restoration-net to get the output of removing noise and artifacts. Finally, the output is added with the input of restoration-net to obtain the final enhancement result.

Guo et al.[65] combined multiple iterative calculations and CNN and proposed a no-reference low-light image enhancement method Zero-DCE. Inspired by the image editing software "curves adjustment," a class of mapping curve from low-light images to enhanced images is designed. First, a set of best-fit curves (LE-curve) of the input image are estimated through DCE-Net, and then the curve equation is iteratively transformed to obtain the final enhancement result.

Zhu et al.[66] proposed a new network RRDNet. The network has three branch networks, which separate the input image into illumination component, reflectivity component, and noise component, and no paired data are required during the training process. First, the decomposed illumination component is enhanced by gamma transformation, then noise is subtracted from the input image, divided by the enhanced illumination component to obtain the reflection component, and finally get the final enhancement result through Retinex theory.

Wang et al.[67] proposed a new CNN structure, namely the deep lightening network (DLN). The network is mainly composed of lighten back-project (LBP) and feature aggregation (FA) modules. Among them, LBP iteratively performs the process of brightening and darkening to learn the residual, and the FA module is to fuse the features of different scales of the image. First, perform preliminary feature extraction on the input image X, and then the residual is obtained through the LBP module and the FA module, and the final residual is multiplied by the parameter γ and added to the original image to obtain the final enhancement result Y.

Lu et al.[68] proposed a dual-branch exposure fusion enhanced network TBEFN. The network is divided into two parts: the -1E branch enhances slightly distorted images, and the -2E branch enhances the more severely distorted image with noise. Then through the FM for rough fusion and further refinement, the final enhancement result is obtained.

From the above introduction, it is found that many of the deep learning enhancement methods for low-light images are based on the Retinex theory. Therefore, summarize and show the results of several enhancement methods based on the Retinex theory on LIME dataset, as shown in Table 1 and Fig. 4.

3.4 Recurrent Neural Network

RNN developed into one of the deep learning algorithms in the early 21st century. Elman[69] proposed the first fully connected RNN. Later, due to the problems of gradient disappearance and gradient explosion, researchers made a series of improvements to RNN, among which Bi-RNN[70] and LSTM[71] are a more commonly used RNN. RNNs have memory characteristics compared with traditional networks. For traditional neural networks, its input and output are independent of each other. But for some tasks, the output is not only related to the current input but also related to the input at the previous moment, so RNN has the characteristics of memory, as shown in Fig. 5. That is, the output depends on the previous input sequence, and the RNN can also be combined with other networks to form a hybrid neural network.

Table 1 Deep learning method based on Retinex theory.

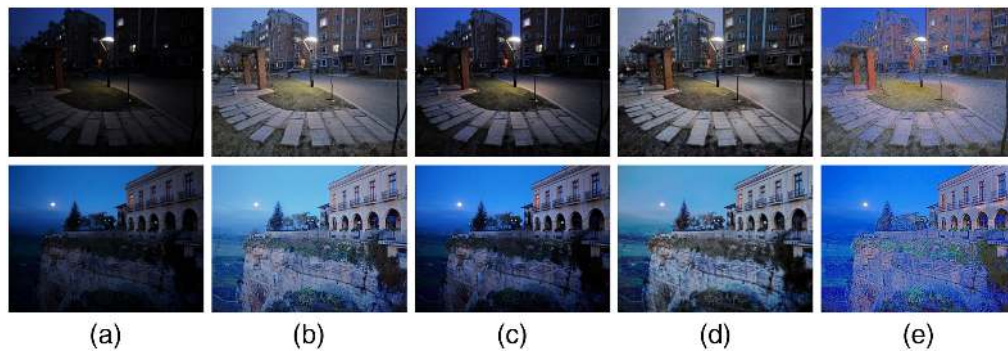| Method	Hardware devices
LightenNet[44]	Intel(R) Xeon(R) CPU E5-2660 v3 @2.60 GHz and an Nvidia Titan X GPU
Retinex-Net[46]	—
DeepUPE[57]	NVidia Titan X Pascal GPU
KinD[55]	Nvidia GTX 2080Ti GPU and Intel Core i7-8700 3.20 GHz CPU
RRDNet[66]	3.0 GHz Intel Core i7-5960X CPU and an Nvidia GeForce GTX 980Ti GPU
RDGAN[60]	Intel Xeon E5-2630 CPU and NVIDIA GTX 1080 Ti GPU

Fig. 4 The result of deep learning enhancement methods based on Retinex theory on LIME image. (a) Input, (b) RDGAN, (c) RRDNet, (d) KinD-Net, and (e) Retinex-Net.
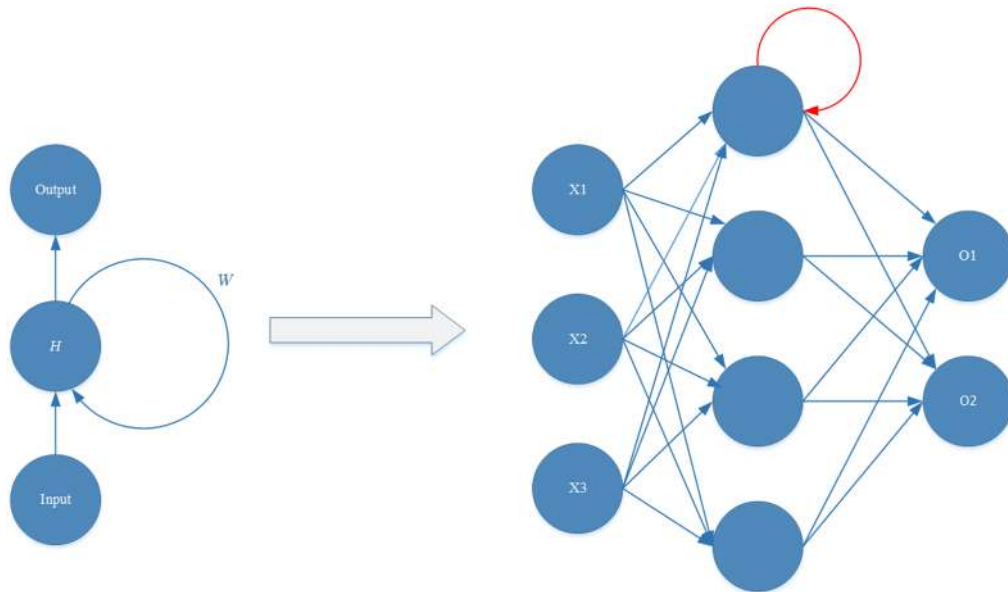

Fig. 5 RNN structure.

Ren et al.[72] proposed a hybrid neural network combining autoencoder and RNN. The network is divided into two streams. The content stream uses the autoencoder structure and skip connections to estimate the global feature information of the image, and the edge stream obtains the edge information of the image through the two weights of g and p output by CNN and the spatial variant RNN. Then, fuse global content features and edge features to get the final enhancement result.

3.5 *Generative Adversarial Network*

GAN is a commonly used network in deep learning, and it is a widely used unsupervised learning. GAN is composed of generative model G and discriminant model D. The generator receives a random noise z, and then generates an image $G(z)$ through this noise z. The discriminator is to judge whether an image is "real," input an image x, and output $D(x)$ means the probability that x is a real image. If $D(x) = 1$, it means it must be a real image; if $D(x) = 0$, it means it must not be a real image. In the training process, the task of generative model G is to generate real images as much as possible to deceive the discriminative model D; the task of discriminative model D is to distinguish the images generated by the generative model G from the real images as much as possible. The most ideal situation is that the generative model G can generate images that closely resemble real images, whereas the discriminator model D is difficult to determine whether the images generated by G are real, and $D(G(z)) = 0.5$. Its working principle is shown in Fig. 6.
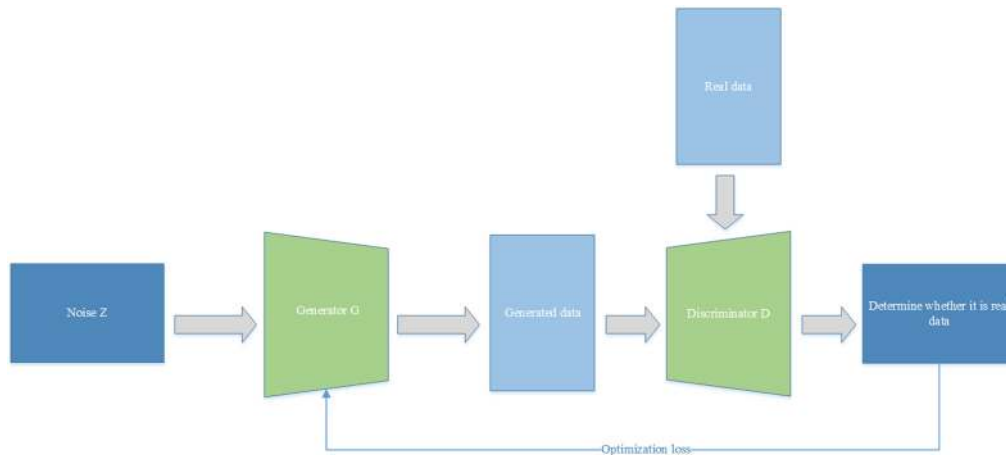
Fig. 6 GAN principles.

Jiang et al.[73] proposed an EnlightenGAN network based on GAN that does not require paired supervision. The generator in the network is an attention-guided U-Net structure, and the discriminator is a global–local dual discriminator structure. The global discriminator adopts a relative discriminator structure to improve the ability of the discriminator; the local discriminator randomly crops five image blocks from the images before and after the enhancement to distinguish, and the self feature preserving loss function is used to make the content before and after the image enhancement unchanged.

The summary of deep learning enhancement methods for low-light images is shown in Table 2, and its timeline is shown in Fig. 7.

4 Learning Method of Deep Learning Enhancement Method

In the previous section, a brief overview of deep learning enhancement methods for low-light images was given. In this section, these enhancement methods will be divided into supervised learning, unsupervised learning, and zero-shot learning (ZSL) according to the learning method of the deep learning, as shown in Table 3.

4.1 Supervised Learning

In the process of network training, the data in the training dataset have both features and labels corresponding to the features. The model is trained through these two items in the dataset, so the model can determine the corresponding label according to the features of the input data. In the enhancement methods based on deep learning, paired data are often required during training: low-light image and standard image. Most of the methods currently proposed are supervised learning, and in these methods, researchers not only propose low-light image enhancement networks but also provide some public paired datasets, such as LOL, SID, etc.

4.2 Unsupervised Learning

In low-light image enhancement, collecting a large number of paired images in the same scene is more difficult. Therefore, Jiang et al.[73] proposed the EnlightenGAN network in response to this situation and successfully introduced unpaired training in the deep learning enhancement methods.

4.3 Zero-Shot Learning

ZSL is in the process of neural network training, no training samples are needed or test samples are types that do not exist in the training samples, but the purpose can be achieved through the trained model mapping. For example, the zero-DCE network proposed by Guo et al.[65] and the RRDNet proposed by Zhu et al.[66] belong to ZSL. The zero-DCE network does not need any

Table 2 Deep learning enhancement method for low-light images.

| Method	Network type	Frame	Evaluation index	Time
LLCNN[40]	CNN	—	PSNR, SNM, LOE, SSIM	2017VCIP
MSR-Net[43]	CNN	—	SSIM, NIQE	2017ArXiv
LightenNet[44]	CNN	Caffe, MATLAB	PSNR, MAE	2018PRL
SCIE[45]	CNN	Caffe, MATLAB	PSNR, FSIM	2018TIP
LLNet[35]	AE	Theano	PSNR, SSIM	2017PR
Retinex-Net[46]	CNN	TensorFlow	—	2018BMVC
MBLLEN[48]	CNN	TensorFlow	PSNR, SSIM, AB, VIF, LOE, TMQI	2018BMVC
GLADNet[53]	CNN	TensorFlow	—	2018FG
LL-RefineNet[54]	CNN	—	PSNR, RMSE, SSIM	2018Symmetry
Park et al.[37]	AE	TensorFlow	PSNR, SSIM	2018IEEEAccess
KinD[55]	CNN	TensorFlow	PSNR, SSIM, NIQE, LOE	2019ACMMM
DeepUPE[57]	CNN	TensorFlow	PSNR, SSIM	2019 CVPR
RDGAN[60]	CNN	TensorFlow	FSIM, PSNR	2019ICME
EEMEFN[61]	CNN	PyTorch	PSNR, SSIM	2020 AAAI
Chen et al.[51]	FCN	TensorFlow	PSNR, SSIM	2018CVPR
Fan et al.[63]	CNN	—	NIQE, PSNR, SSIM	2020ACMMM
Lv et al.[64]	CNN	TensorFlow	PSNR, SSIM	2020ACMMM
Zero-DCE[65]	CNN	PyTorch	PSNR, MAE, SSIM	2020 CVPR
RRDNet[66]	CNN	PyTorch	NIQE, CPCQI	2020 ICME
Ren et al.[72]	RNN and AE	Caffe	PSNR, SSIM	2019 TIP
DLN[67]	CNN	PyTorch	PSNR, NIQE, SSIM	2020 TIP
Lv et al.[49]	CNN	TensorFlow	LPIPS, PSNR, TMQI, SSIM, VIF, LOE, AB	2021IJCV
TBEFN[68]	CNN	TensorFlow	PSNR, SSIM, NIQE	2020 TMM
EnlightenGAN[73]	GAN	PyTorch	NIQE	2021 TIP
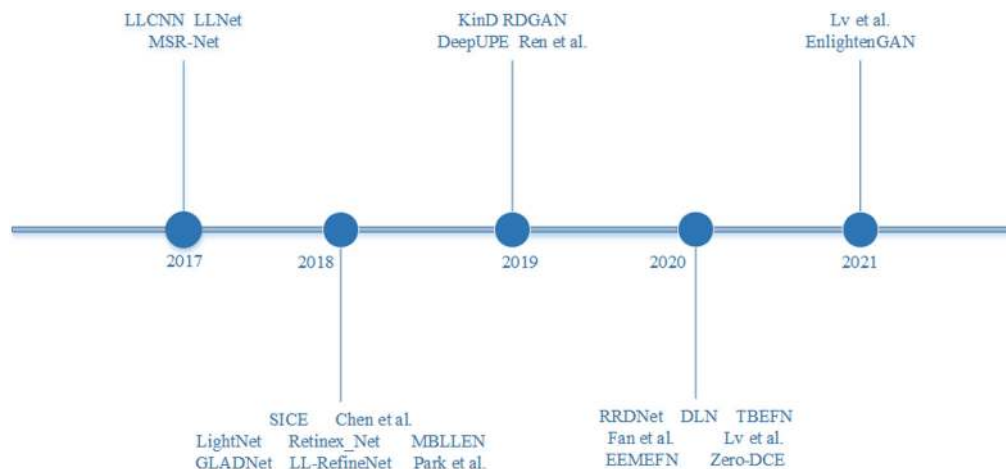
Fig. 7 Timeline of deep learning-based low-light image enhancement.

Table 3 Deep learning method based on training method classification.

Method	Hardware devices	Training style
RetinexNet[46]	—	Supervised learning
KinD[55]	Nvidia GTX 2080Ti GPU and Intel Core i7-8700 3.20 GHz CPU	Supervised learning
EnlightenGAN[73]	—	Unsupervised learning
Zero-DCE[65]	NVIDIA 2080Ti GPU	ZSL
RRDNet[66]	3.0 GHz Intel Core i7-5960X CPU and Nvidia GeForce GTX 980Ti GPU	ZSL

samples during the training process, it only needs to iterate the designed curve several times to get the final enhancement result. The RRDNet proposed by Zhu et al. does not require paired samples during the training process. It only needs to input the image to be enhanced and then iteratively minimize the loss function to enhance the input image to obtain the final enhancement result.

5 Datasets Used by Deep Learning Method

In recent years, many methods of using deep learning for enhancement have emerged for low-light images, but these methods usually require a large number of paired images during training, it is more difficult to collect images. Therefore, the existing enhancement methods for low-light images are mostly trained and evaluated on synthetic low-light image datasets. For example, Cai et al.[45] synthesized a large-scale multiexposure image dataset SICE, which contains low-contrast images of different exposures and their corresponding high-contrast reference images. The high-contrast reference image is produced by the best method among 13 most advanced MEF and HDR methods. Chen et al.[51] provided a new dataset SID dataset, which contains 5049 short-exposure images. Each short-exposure image has a corresponding long-exposure reference image. The images are divided into indoor and outdoor images. When shooting outdoors, the camera brightness is generally between 0.2 and 5 lux, and the indoor camera brightness is generally between 0.03 and 0.3 lux.

Table 4 shows the commonly used datasets of deep learning enhancement methods for low-light images.

Table 4 Datasets commonly used in deep learning methods.

Author	Dataset	Paired/Unpaired
Cai et al.[45]	SICE	Paired
Wei et al.[46]	LOL	Paired
Chen et al.[51]	SID	Paired
Wang et al.[57]	DeepUPE	Paired
Ma et al.[74]	MEF	Paired
Lee et al.[75]	DICM	Unpaired
Guo et al.[9]	LIME	Unpaired
Wang et al.[8]	NPE	Unpaired
Loh et al.[76]	ExDARK	Unpaired
Bychkovsky et al.[77]	MIT-Adobe FiveK	Paired

Table 5 Abbreviation for image quality evaluation index.

Image quality evaluation index	Abbreviation	Full/no reference
Mean square error	MSE	Full reference
Visual information fidelity	VIF[78]	Full reference
Information entropy	Entropy	No reference
Information fidelity criterion	IFC[79]	Full reference
Structural similarity	SSIM[80]	Full reference
Lightness order error	LOE[8]	No reference
Natural image quality evaluator	NIQE[81]	No reference
Feature similarity index	FSIM[82]	Full reference
Peak signal-to-noise ratio	PSNR	Full reference
Average brightness	AB[83]	No reference
Learned perceptual image patch similarity metric	LPIPS[84]	Full reference
colorfulness-based patch-based contrast quality index	CPCQI[85]	No reference
Tone mapped image quality index	TMQI[86]	Full reference

6 Commonly Used Image Quality Evaluation Indicators

After the image is enhanced, the degree of distortion deviation between the image to be evaluated (enhanced image) and the standard image is usually evaluated using evaluation indicators. Image quality evaluation is divided into subjective evaluation indicators and objective evaluation indicators according to human subjective awareness or objective standards. The objective evaluation of image quality is divided into full-reference evaluation index, reduced-reference evaluation index, and no-reference evaluation index according to whether there is a standard image as a reference.

Full reference image quality evaluation index refers to the use of a standard or ideal image as a reference image, comparing the image to be evaluated with the reference image, and obtaining the evaluation result of the image to be evaluated. Commonly used full reference evaluation indicators are: mean square error, visual information fidelity[78] (VIF), structural similarity, mean absolute error, information fidelity criterion[79] (IFC), and peak signal-to-noise ratio. Reduced-reference quality evaluation is also called partial reference. It takes partial information of the ideal image as a reference and compares it with the image to be evaluated to obtain the evaluation result. Common reduced-reference evaluation methods are based on the original image feature method, etc. Nonreference quality evaluation refers to the evaluation of images directly through several commonly used evaluation indicators without referring to any image information. The commonly used nonreference evaluation indicators are mean, standard deviation, information entropy (Entropy), natural image quality evaluation (NIQE), etc.

In this section, a brief overview of commonly used image quality evaluation indicators is given. Tables 5 and 6 give the abbreviations and mathematical equations of the evaluation indicators, respectively.

Table 6 Mathematical equation of image quality evaluation index.

Performance metrics	Formula
MSE	$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [x(i,j) - \hat{x}(i,j)]^2$
PSNR	$PSNR = 10 \bullet \log_{10}\left(\frac{MAX^2}{MSE}\right)$
SSIM	$SSIM = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}\right)\left(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}\right)$
NIQE	$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{(v_1 - v_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (v_1 - v_2)}$

6.1 *Mean Square Error*

The mean square error is one of the more commonly used indicators in image quality evaluation, which is to calculate the mean value of the square sum of the pixel value errors of the corresponding points of the image to be evaluated and the original image. As shown in Eq. (7):

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [x(i, j) - \hat{x}(i, j)]^2, \tag{7}$$

$x(i, j)$ and are expressed as the pixel value of the reference image and the image to be evaluated, respectively. The smaller the value of MSE, the better the quality of the image to be evaluated.

6.1.1 *Peak Signal-to-Noise Ratio*

MAX represents the maximum value of the pixel. If each pixel is an 8-bit table binary, then MAX = 255. The larger the PSNR, the smaller the distortion between the image to be evaluated and the reference image, and the better the quality of the image to be evaluated.

$$\text{PSNR} = 10 \bullet \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right). \tag{8}$$

6.1.2 *Structural Similarity*

SSIM[80] is an index that measures the similarity between the image to be evaluated and the reference image. It is measured from the three aspects of brightness, contrast, and structure:

$$\text{SSIM} = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \right) \left(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \right). \tag{9}$$

The larger the value of SSIM, the smaller the distortion of the image and the better the image quality. When the two images are exactly the same, SSIM = 1.

6.1.3 *Natural Image Quality Evaluator*

Mittal et al.[81] proposed a nonreference evaluation index NIQE. The NIQE score is obtained by calculating the distance between the parameters of the multivariate Gaussian model (MVG) of the image to be evaluated and the MVG parameters of the natural image:

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{(v_1 - v_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (v_1 - v_2)}. \tag{10}$$

6.1.4 *Other Applications of Deep Learning*

Due to the excellent performance of deep learning, in addition to the application in the field of low-light image enhancement, there are also many applications in other fields.[87] In this section, other applications of deep learning in the field of computer vision are briefly introduced.

1. Image segmentation: Image segmentation refers to dividing the image into several regions according to the feature information of the image for simplified analysis, which is divided into semantic segmentation, instance segmentation, and panoramic segmentation. It is an important research direction in the field of computer vision. The initial image segmentation was based on traditional segmentation methods such as thresholds. Later, with the development of deep learning technology, image segmentation technology developed rapidly, and many deep learning-based image segmentation techniques were studied,[88,89] which are widely used in various fields, such as the field of medical imaging.[90]

2. Object detection: Object detection refers to finding the target object in the image and determining the position and category of the target, which is widely used in computer

vision. With the development of deep learning and the excellent performance of CNN in image processing, some excellent target detection methods based on deep learning have emerged, such as YOLO,[91] Faster R-CNN,[92] etc. It performs well in face detection, driving vehicles, pedestrian detection,[93] etc.

3. Background subtraction: To lock the moving target from the captured video, the background subtraction method comes into being. It separates the moving object in the video from the background information without any moving object. It is often used in target detection and is one of the common methods for moving target detection. Due to the development of deep learning, neural networks are gradually applied in this field. Bouwmans et al.[94] introduced deep neural networks for background subtraction.

4. Human activity recognition: Due to the development of sensor technology and ubiquitous computing, sensor-based HAR is becoming more and more popular, so human activity recognition is currently a relatively hot research area, which is to identify the actions that are taking place from the collected sensor data. For example, motion states such as walking and running can be recorded and identified by wearing sensors or mobile phones with accelerometers and gyroscopes. Due to the advantages of deep learning to extract features, HAR based on deep learning has been gradually developed.[95]

7 Conclusion

With the development of artificial intelligence theory and technology and the emergence of the first deep learning enhancement method for low-light images, opening the door to the development of deep learning in this field, and then a large number of deep learning enhancement methods have appeared one after another. Because the deep learning method is not as complicated as traditional algorithms in adjusting parameters, which can learn appropriate parameters from the sample data through continuous training. Therefore, deep learning enhancement methods are currently widely used in various fields of life, such as medicine, transportation, and public safety. In the medical field, it is necessary to enhance the dark and blurred images under the electron microscope; and the enhancement technology can also be applied to the transportation field. When the vehicle is in backlight or at night, it can be enhanced by this technology. Therefore, deep learning enhancement methods for low-light images have been an important direction in recent years and are of great significance to the medical field and the transportation field. However, there are still some problems in the application of existing deep learning enhancement methods.

7.1 *Datasets*

In the existing deep learning enhancement methods for low-light images, most of the training datasets are synthetic data, and there is a lack of paired real-world data sets. In response to this problem, in future research on enhancement methods of deep learning, learning methods such as ZSL, self-supervision, and graph signal processing can be considered to reduce the demand for paired data.

Among them, graph signal processing is a signal that discrete signal processing extends to graphs, and graphs are structured data composed of a series of vertices and edges. Due to the existence of some non-Euclidean data, research on graph upsampling[96] and the study of graph neural network in computer vision applications,[97] GSP has attracted more and more attention in the field of computer vision. For example, inspired by the GSP method, Giraldo et al.[98,99] proposed a semisupervised method combining MOS and GPS that can achieve good results with only a small number of labeled samples. Ortega et al.[100] introduced some graph sampling strategies for the scarcity of labeled samples in semisupervised learning.

7.2 *Generalization Ability of the Method*

At present, in the existing enhancement methods, most of the datasets used in the network training process are their own dataset, so the processing effect on the images of the real scene or other datasets is not as good as the effect on the own dataset, and because of the network structure

design, the prior knowledge used by the model and other factors leads to poor generalization ability. In response to this problem, future research should pay more attention to how to improve the generalization ability of the proposed method.

Acknowledgments

References

1. L. Lu et al., "Comparative study of histogram equalization algorithms for image enhancement," *Mobile Multimedia/Image Process. Secur. Appl.* **7708**, 337–347 (2010).
2. S. M. Pizer et al., "Adaptive histogram equalization and its variations," *Comput. Vision Graphics Image Process.* **39**(3), 355–368 (1987).
3. K. T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.* **48**(1), 5548636 (1997).
4. S. D. Chen and A. R. Ramli, "Minimum mean brightness error bi-histogram equalization in contrast enhancement," *IEEE Trans. Consum. Electron.* **49**(4), 1310–1319 (2003).
5. X. Dong et al., "Fast efficient algorithm for enhancement of low lighting video," in *IEEE Int. Conf. Multimedia and Expo*, IEEE, pp. 1–6 (2011).
6. Y. F. Wang, H. M. Liu, and Z. W. Fu, "Low-light image enhancement via the absorption light scattering model," *IEEE Trans. Image Process.* **28**(11), 5679–5690 (2019).
7. Y. Gao et al., "Naturalness preserved nonuniform illumination estimation for image enhancement based on retinex," *IEEE Trans. Multimedia* **20**(2), 335–344 (2018).
8. S. Wang et al., "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.* **22**(9), 3538–3548 (2013).
9. X. Guo, Y. Li, and H. Ling, "LIME: low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.* **26**(2), 982–993 (2017).
10. X. Fu et al., "A fusion-based enhancing method for weakly illuminated images," *Signal Process.* **129**, 82–96 (2016).
11. S. Park et al., "Low-light image enhancement using variational optimization-based retinex model," *IEEE Trans. Consum. Electron.* **63**(2), 178–184 (2017).
12. X. Fu et al., "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 2782–2790 (2016).
13. Z. Ying et al., "A new low-light image enhancement algorithm using camera response model," in *Proc. IEEE Int. Conf. Comput. Vision Workshops*, pp. 3015–3022 (2017).
14. K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2341–2353 (2011).
15. E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Am.* **61**(1), 1–11 (1971).
16. E. H. Land, "The Retinex theory of color vision," *Sci. Am.* **237**(6), 108–128 (1977).
17. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.* **6**(3), 451–462 (1997).
18. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.* **6**(7), 965–976 (1997).
19. G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science* **313**(5786), 504–507 (2006).
20. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM* **60**(6), 84–90 (2017).
21. O. Poursaeed et al., "Generative adversarial perturbations," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 4422–4431 (2018).
22. S. M. Moosavi-Dezfooli et al., "Universal adversarial perturbations," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1765–1773 (2017).

23. C. Szegedy et al., "Intriguing properties of neural networks," arXiv:1312.6199 (2013).
24. R. Giryes, G. Sapiro, and A. M. Bronstein, "On the stability of deep networks," arXiv:1412.5896 (2014).
25. S. Malladi and I. Sharapov, "FastNorm: improving numerical stability of deep network training with efficient normalization" (2018).
26. S. Zheng et al., "Improving the robustness of deep neural networks via stability training," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 4480–4488 (2016).
27. G. Algan and I. Ulusoy, "Image classification with deep learning in the presence of noisy labels: a survey," *Knowl.-Based Syst.* **215**, 106771 (2021).
28. K. K. Thekumparampil et al., "Robustness of conditional GANs to noisy labels," in *Adv. Neural Inf. Process. Syst.*, Vol. 31 (2018).
29. Y. Zhang et al., "A survey on neural network interpretability," *IEEE Trans. Emerging Top. Comput. Intell.* **5**, 726–742 (2021).
30. R. Vidal et al., "Mathematics of deep learning," arXiv:1712.04741 (2017).
31. C. Yun, S. Sra, and A. Jadbabaie, "A critical view of global optimality in deep learning," arXiv:1802.03487 (2018).
32. P. Vincent et al., "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Machine Learn.*, pp. 1096–1103 (2008).
33. A. Ng, "Sparse autoencoder," *CS294A Lect. Notes* **72**, 1–19 (2011).
34. D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," arXiv:1312.6114 (2013).
35. K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: a deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.* **61**, 650–662 (2017).
36. J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Adv. Neural Inf. Process. Syst.*, pp. 341–349 (2012).
37. S. Park et al., "Dual autoencoder network for retinex-based low-light image enhancement," *IEEE Access* **6**, 22084–22093 (2018).
38. K. Fukushima, "Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.* **36**(4), 193–202 (1980).
39. Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.* **1**(4), 541–551 (1989).
40. L. Tao et al., "LLCNN: a convolutional neural network for low-light image enhancement," in *IEEE Visual Commun. and Image Process.*, IEEE, pp. 1–4 (2017).
41. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 770–778 (2016).
42. C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1–9 (2015).
43. L. Shen et al., "MSR-Net: low-light image enhancement using deep convolutional network," arXiv:1711.02488 (2017).
44. C. Li et al., "LightenNet: a convolutional neural network for weakly illuminated image enhancement," *Pattern Recognit. Lett.* **104**, 15–22 (2018).
45. J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.* **27**(4), 2049–2062 (2018).
46. C. Wei et al., "Deep retinex decomposition for low-light enhancement," arXiv:1808.04560 (2018).
47. K. Dabov et al., "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007).
48. F. Lv et al., "MBLLEN: low-light image/video enhancement using CNNs," in *BMVC*, p. 220 (2018).
49. F. Lv, Y. Li, and F. Lu, "Attention guided low-light image enhancement with a large scale low-light simulation dataset," *Int. J. Comput. Vision* **129**(7), 2175–2193 (2021).
50. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
51. C. Chen et al., "Learning to see in the dark," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 3291–3300 (2018).
52. C. Chen et al., "Seeing motion in the dark," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, pp. 3185–3194 (2019).

53. W. Wang et al., "GLADNet: low-light enhancement network with global awareness," in *13th IEEE Int. Conf. Autom. Face and Gesture Recognition*, IEEE, pp. 751–755 (2018).

54. L. Jiang et al., "Deep refinement network for natural low-light image enhancement in symmetric pathways," *Symmetry* **10**(10), 491 (2018).

55. Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: a practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. multimedia*, pp. 1632–1640 (2019).

56. Y. Zhang et al., "Beyond brightening low-light images," *Int. J. Comput. Vision* **129**(4), 1013–1037 (2021).

57. R. Wang et al., "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 6849–6857 (2019).

58. M. Gharbi et al., "Deep bilateral learning for real-time image enhancement," *ACM Trans. Graph.* **36**(4), 1–12 (2017).

59. J. Chen, S. Paris, and F. Durand, "Real-time edge-aware image processing with the bilateral grid," *ACM Trans. Graph.* **26**(3), 103-es (2007).

60. J. Wang et al., "RDGAN: retinex decomposition based adversarial learning for low-light enhancement," in *IEEE Int. Conf. Multimedia and Expo*, IEEE, pp. 1186–1191 (2019).

61. M. Zhu et al., "EEMEFN: low-light image enhancement via edge-enhanced multi-exposure fusion network," in *Proc. AAAI Conf. Artif. Intell.*, Vol. 34, No. 7, pp. 13106–13113 (2020).

62. Y. Liu et al., "Richer convolutional features for edge detection," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 3000–3009 (2017).

63. M. Fan et al., "Integrating semantic segmentation and retinex model for low-light image enhancement," in *Proc. 28th ACM Int. Conf. Multimedia*, pp. 2317–2325 (2020).

64. F. Lv, B. Liu, and F. Lu, "Fast enhancement for non-uniform illumination images using light-weight CNNs," in *Proc. 28th ACM Int. Conf. Multimedia*, pp. 1450–1458 (2020).

65. C. Guo et al., "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 1780–1789 (2020).

66. A. Zhu et al., "Zero-shot restoration of underexposed images via robust retinex decomposition," in *IEEE Int. Conf. Multimedia and Expo*, IEEE, pp. 1–6 (2020).

67. L. Wang et al., "Lightening network for low-light image enhancement," *IEEE Trans. Image Process.* **29**, 7984–7996 (2020).

68. K. Lu and L. Zhang, "TBEFN: a two-branch exposure-fusion network for low-light image enhancement," *IEEE Trans. Multimedia* **23**, 4093–4105 (2021).

69. J. L. Elman, "Finding structure in time," *Cognit. Sci.* **14**(2), 179–211 (1990).

70. M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.* **45**(11), 2673–2681 (1997).

71. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.* **9**(8), 1735–1780 (1997).

72. W. Ren et al., "Low-light image enhancement via a deep hybrid network," *IEEE Trans. Image Process.* **28**(9), 4364–4375 (2019).

73. Y. Jiang et al., "EnlightenGan: deep light enhancement without paired supervision," *IEEE Trans. Image Process.* **30**, 2340–2349 (2021).

74. K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.* **24**(11), 3345–3356 (2015).

75. C. Lee, C. Lee, and C. S. Kim, "Contrast enhancement based on layered difference representation," in *19th IEEE Int. Conf. Image Processing*, IEEE, pp. 965–968 (2012).

76. Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Comput. Vision Image Understanding* **178**, 30–42 (2019).

77. V. Bychkovsky et al., "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, IEEE, pp. 97–104 (2011).

78. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).

79. H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.* **14**(12), 2117–2128 (2005).

80. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
81. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.* **20**(3), 209–212 (2013).
82. L. Zhang et al., "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Process.* **20**(8), 2378–2386 (2011).
83. Z. Y. Chen et al., "Gray-level grouping (GLG): an automatic method for optimized image contrast enhancement-part I: the basic method," *IEEE Trans. Image Process.* **15**(8), 2290–2302 (2006).
84. R. Zhang et al., "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 586–595 (2018).
85. K. Gu et al., "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Networks Learn. Syst.* **29**(4), 1301–1313 (2018).
86. H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.* **22**(2), 657–667 (2013).
87. S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Computer Science Review* **40**, 100379 (2021).
88. S. Minaee et al., "Image segmentation using deep learning: a survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 1 (2021).
89. A. G. Garcia et al., "A survey on deep learning techniques for image and video semantic segmentation," *Appl. Soft Comput.* **70**, 41–65 (2018).
90. M. H. Hesamian et al., "Deep learning techniques for medical image segmentation: achievements and challenges," *J. Digit. Imaging* **32**(4), 582–596 (2019).
91. J. Redmon et al., "You only look once: unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 779–788 (2016).
92. S. Ren et al., "Faster R-CNN: towards real-time object detection with region proposal networks," in *Adv. Neural Inf. Process. Syst.*, Vol. 28 (2015).
93. L. Chen et al., "Deep neural network based vehicle and pedestrian detection for autonomous driving: a survey," *IEEE Trans. Intell. Transp. Syst.* **22**(6), 3234–3246 (2021).
94. T. Bouwmans et al., "Deep neural network concepts for background subtraction: a systematic review and comparative evaluation," *Neural Network* **117**, 8–66 (2019).
95. J. Wang et al., "Deep learning for sensor-based activity recognition: a survey," *Pattern Recognit. Lett.* **119**, 3–11 (2019).
96. Y. Tanaka et al., "Sampling signals on graphs: from theory to applications," *IEEE Signal Process. Mag.* **37**(6), 14–30 (2020).
97. P. Pradhyumna, G. P. Shreya, and Mohana, "Graph neural network (GNN) in image and video understanding using deep learning for computer vision applications," in Second Int. Conf. Electron. and Sustainable Commun. Syst., IEEE, pp. 1183–1189 (2021).
98. J. H. Giraldo et al., "The emerging field of graph signal processing for moving object segmentation," in *Int. Workshop Front. Comput. Vision*, Springer, Cham, pp. 31–45 (2021).
99. J. H. Giraldo, S. Javed, and T. Bouwmans, "Graph moving object segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(5), 2485–2503 (2020).
100. A. Ortega et al., "Graph signal processing: overview, challenges, and applications," *Proc. IEEE* **106**(5), 808–828 (2018).

Yong Wang received her bachelor's degree from Jilin University in 2004 and her doctorate degree from the Graduate School of the Chinese Academy of Sciences in 2010. She is currently an associate professor at Jilin University and her main research field is digital image processing.

Wenjie Xie received her bachelor's degree in engineering in 2020 and is currently studying for a master's degree at Jilin University. Her main research interests include deep learning and image enhancement.

Hongqi Liu received her master's degree from Jilin University in 2020 and is currently working at the 602th Institute of the Sixth Academy of Aerospace Science and Technology of China. Her research interests include artificial intelligence and image fusion.