ISSN 2616-5767 Vol.8, Issue 1: 56-64, DOI: 10.25236/AJETS.2025.080109

Research on Low-Light Image Enhancement Algorithm Based on Multi-Dimensional Attention

Xiaoting Niu*, Guanghai Zheng

School of Software, Dalian Jiaotong University, Dalian, China
905922247@qq.com
Corresponding author

Abstract: *This study addresses the issue of noise in low-light images, which is prevalent and significantly affects image quality. A convolutional neural network (CNN)-based low-light image denoising module is proposed to tackle the problem that existing single-stage CNN models cannot effectively remove noise in dark areas. The module extracts high and low-frequency features through constructed convolutional blocks and utilizes an attention module to better focus on key features. Additionally, a residual structure is employed to preserve image detail information. In terms of the loss function, this paper adopts the smooth L1 loss, which combines the robustness of L1 loss against outliers with the numerical stability of L2 loss. Experimental results show that the pre-trained model achieves a peak signal-to-noise ratio (PSNR) of 25.133 dB and a structural similarity (SSIM) of 0.913 on the LOL dataset, and a PSNR of 19.427 dB and an SSIM of 0.827 on the GladNet-Dataset for noise image recovery. This indicates that the proposed model is significantly effective in enhancing and denoising low-light images under non-uniform lighting conditions and performs well in denoising natural low-light scene images.*

Keywords: *CNN; Multidimensional attention; Image Denoising; Low-Light Enhancement; Residual Learning*

1. Introduction

High quality image provides guarantee for the subsequent processing of image. However, due to the error of image acquisition system, special shooting environment and other factors, it is often impossible to obtain clear images directly. Noise and illumination are the two main factors affecting image quality. Noisy images will produce blur and local details are not clear, etc., while dark images contain a lot of noise in low contrast areas, which is not only difficult to distinguish but also has a great impact on subsequent image processing.

Image de-noising methods based on image filtering are widely used [1] and perform best when dealing with all-band noise, adaptive selection dependent on noise level, combined with other de-noising techniques, and specific types of noise (such as Gaussian noise) [2]. However, there are obvious limitations in dealing with complex noise and maintaining image detail. However, using learn-based methods, such as DnCNN, FFDNet, etc., denoising is realized by learning the mapping from noisy image to clean image [3-5], but there are also some challenges, especially in terms of data requirements, computing resources and model generalization ability. In the actual process of improving image quality, the enhancement of light is accompanied by the expansion of noise signal, and the removal of noise signal will often blur the features of dark light image. In recent years, with the rapid development of deep learning, the method based on convo-lutional neural network (CNN) has been widely used in image processing, and has also achieved unprecedented achievements in dark light enhancement and image denoising [6-9]. The real-time low-light enhancement algorithm (Zero-DCE) proposed by Guo C et al., which does not require reference data, uses neural networks to fit a brightness mapping curve, and then generates a brightening image based on the curve and the original image [9]. However, Zero-DCE can only achieve brightness enhancement and does not work on image deblurring. Lin L et al proposed a deep unfolding network based on Retinex (URetinex-Net), which decomposed low-light images into reflection layer and illumination layer, adaptive fitting implicit priori in a data-driven way, and realized noise suppression and detail retention of the final decomposition results [10]. However, although this deep learning method can achieve the adjustment of the overall brightness, it lacks the suppression of degraded information such as noise. Later, Wriza W et

ISSN 2616-5767 Vol.8, Issue 1: 56-64, DOI: 10.25236/AJETS.2025.080109

al used URetinex-Net and TRBA to enhance low-light image in license plate recognition on this basis, and this method had a good effect on improving the environmental accuracy of night license plate recognition system [11]. Lin S et al. proposed a synchronous multi-scale dark light enhancement network (SMNet) method, which learns feature streams of different scales in a top-down to bottom-up manner through multiple L&G modules in series to achieve image denoising. This method has good effects on dark light image enhancement and noise suppression in natural scenes [12].

In this paper, we construct an N-net network [13], and design a global and local aggregate attention module in order to better achieve the enhancement of dark images with non-uniform illumination. Specifically, input features are respectively input into the global attention module and the local attention module, and then the features are extracted respectively to adapt to different brightness [14,15]. Finally, the output attention map is residual from the input features to prevent the original features from being lost too much. In addition, a denoising module is designed to remove noise during the process of dark light enhancement. High and low frequency features are extracted by constructing convolutional blocks. Attention modules are used to better focus on key features of extracted high and low frequency features, and residual structures are used to maintain image details [16].

2. Design of De-noising Model for Dark Light Image

2.1 Model Construction

Figure 1 shows the overall framework of the N-Net model.

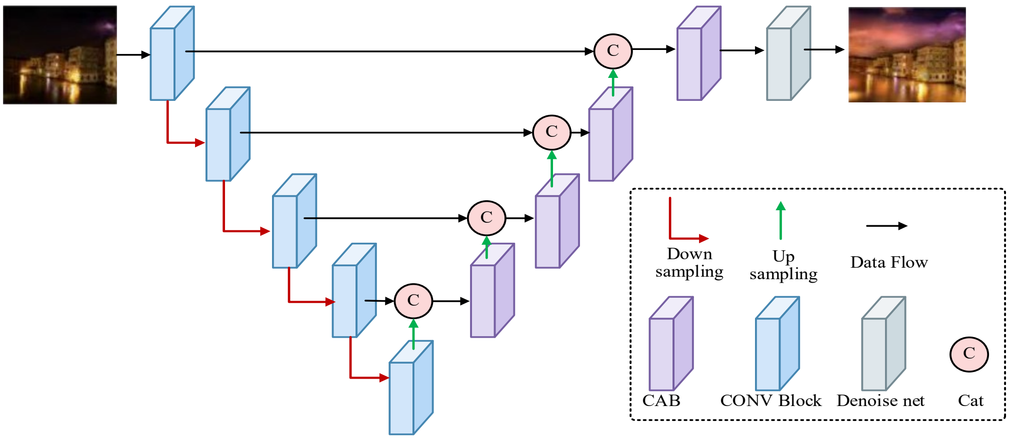

Figure 1: Overall framework of dark image denoising model

N-Net model can be divided into two modules, which are dark light enhancement and noise reduction module. The dark light enhancement module takes U-net network as the whole structure, and its encoder part is composed of 5 conv blocks (see Figure 2). Each stage extracts features with different resolutions through subsampling.

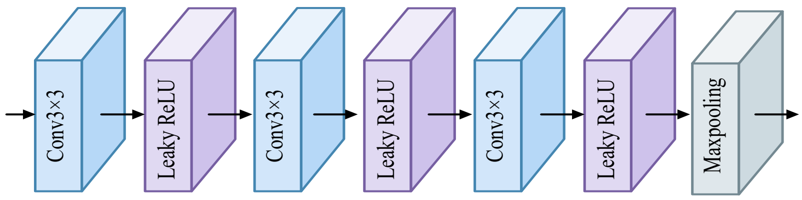

Figure 2: Conv block

The decoder part is composed of 4 conv blocks and global local aggregation module (CAB). The output features of each level of decoder are sampled up, and the output features of each level are input into the next level of decoder to achieve feature fusion to achieve the effect of dark light enhancement. The denoising module extracts different scales through convolution of different kernel sizes and carries out feature interaction, and removes noise by combining batch normalization (BN) and residual structure. The design of the model aims to improve the image quality under low light conditions by improving the image quality through dark light enhancement module and removing noise through noise removal module.

ISSN 2616-5767 Vol.8, Issue 1: 56-64, DOI: 10.25236/AJETS.2025.080109

The encoder part consists of multiple conv blocks, each with a 3x3 convolution kernel and a step size of 1 configuration. Each convolutional block internal structure contains a combination of three groups (convolution +LeakyReLU), where convolution is used for feature extraction and LeakyReLU introduces nonlinearity as an activation function.

The light distribution of low-light pictures is often uneven, for example, there will be a small bright area in a low-light picture scene. Therefore, in order to better adapt to the enhancement of non-uniform light image, a global and local aggregation attention module is designed. Figure 3 is a schematic of global and local attention aggregation, combining global attention and local attention to enhance feature representation.

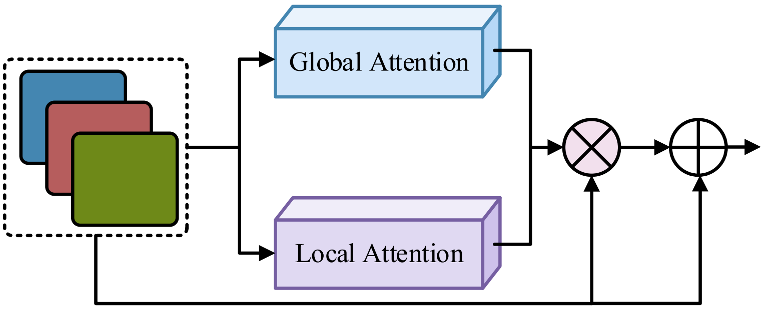

Figure 3: Global and local aggregation attention mechanisms

Figure 4 shows the structure of global attention mechanism. Global attention (GAM) enhances the interdimensional interaction by retaining both channel and space information [14].

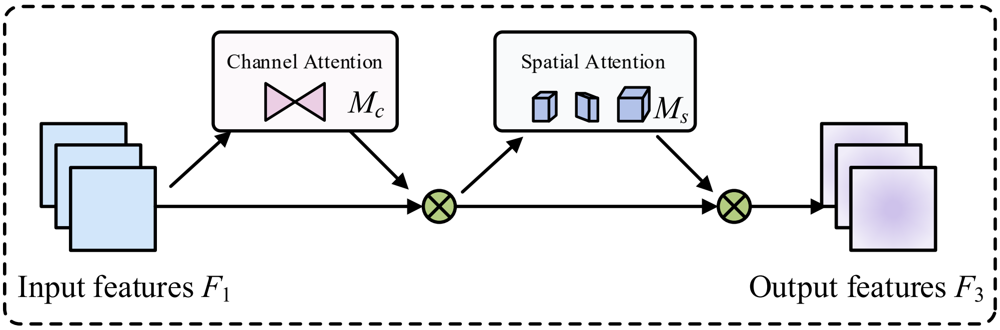

Figure 4: Structure of global attention mechanism

$$F_2 = M_C(F_1) \otimes F_1$$
$$F_3 = M_S(F_2) \otimes F_2 \qquad (1)$$

Here F_1, F_2, and F_3 represent the feature maps of the different stages, and M_C and M_S represent channel attention and spatial attention mechanisms, respectively. This approach of combining channel and spatial attention can make the network pay more attention to the features and regions that contribute to the task, thus improving the performance and generalization ability of the model. While paying attention to the brightness and position of images, GAM combines local attention to extract local features and adapt to different brightness [15].

Figure 5 shows the structure of the local attention mechanism, which helps the model to better focus on important local areas in the input feature map.

The specific work flow of the local attention mechanism is as follows:

The input feature maps are first averaged in the X and Y directions to obtain feature maps in both directions. The formula is as follows:

$$z_c^h(h) = \frac{1}{H} \sum_{0 \le i < H} x_c(h,i)$$
$$z_c^w(w) = \frac{1}{W} \sum_{0 \le j < w} x_c(j,w) \qquad (2)$$

ISSN 2616-5767 Vol.8, Issue 1: 56-64, DOI: 10.25236/AJETS.2025.080109

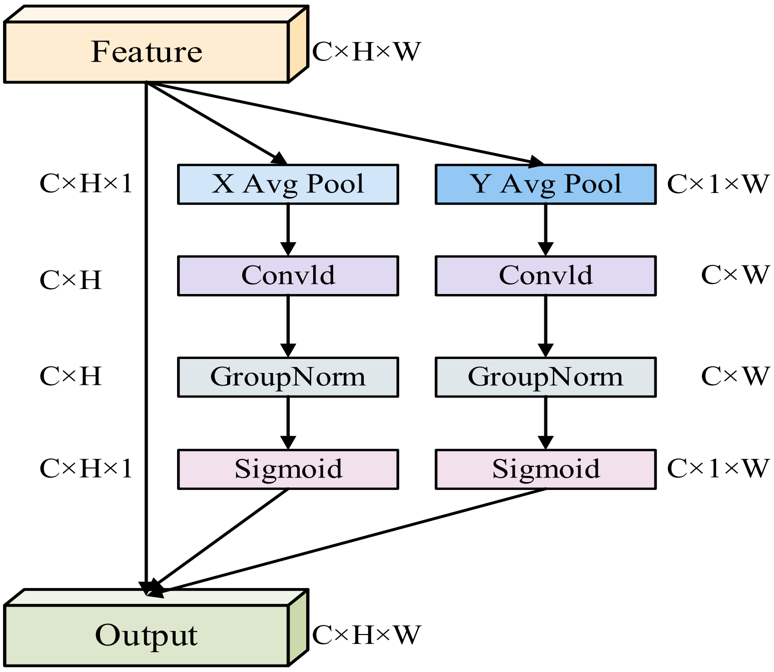

Figure 5: Structure of local attention mechanism

The two feature graphs are convolved by a one-dimensional convolution layer, and then normalized by a group normalization layer.

$$f = \theta(BN(F_1([z^h, z^w]))) \tag{3}$$

Where θ represents the Sigmoid function.

The normalized feature graph is converted to the weight graph by Sigmoid activation function.

$$g_c^h = \theta(F_h(f^h))$$
$$g_c^w = \theta(F_w(f^w)) \tag{4}$$

Finally, the two weight graphs are multiplied with the original feature graph at element level to get the final output feature graph.

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \tag{5}$$

This helps the algorithm better understand the overall structure and light distribution of the image, so that it can more fully understand and process non-uniform low-light images, and generate more natural and realistic enhancement effects. At the same time, the traditional dark light enhancement algorithm may lose some important information in the process of processing, resulting in the deterioration of the image quality after enhancement. GAM combined with high-efficiency channel attention ELA can effectively reduce information loss, retain more original image information, and make the enhanced image closer to the real light image. Finally, the output attention map is residuated from the input features to prevent the loss of the original features [16].

In addition to the noise itself when the imaging device collects images, some of the noise is hidden in the dark, and even in the process of dark light enhancement, noise is often introduced. Figure 6 is the schematic diagram of the noise removal subnetwork.

Firstly, features are extracted from conv+reLu shapes, in which 3x3 convolution and 5-way convolution can extract features of different scales through different receptive fields and perform interactive fusion. The channel attention mechanism is used to focus the network on the most relevant auxiliary information and features that are most conducive to the denoising process. Make it better aware of more relevant information and more critical features. Finally, residual learning combined with batch normalization (BN) improves the training efficiency and performance of the network, enhances the generalization ability of the model, and contributes to the stability and convergence of the network.

ISSN 2616-5767 Vol.8, Issue 1: 56-64, DOI: 10.25236/AJETS.2025.080109

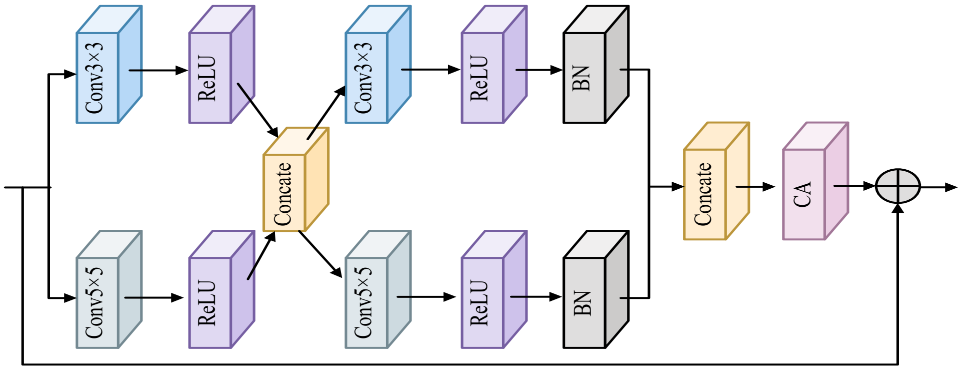

Figure 6: De-noise subnetwork

2.2 Parameter Setting

In the PyTorch framework, NVIDIA 3090 Gpus are used for deep learning training. Set the initial learning rate to 0.0001 and the decay rate to 0.1. Optimized using the Adam optimizer, batchsize equals 4. The image resolution is cropped to 512×512 pixels.

3. Model Training

3.1 Data Set Preparation

The experiment referenced the GladNet-Dataset dataset [17], LSRWhuawei dataset, and LOL dataset, the mainstream real world dataset of dark light images, which consisted of 500 pairs of low and normal light images, and the images contained the noise generated during the photographing process.

3.2 Loss Function

Smooth L1 Loss is a loss function in machine learning, especially in object detection and regression problems. This loss function also has the effect of de-noising, mainly reflected in its reasonable control of gradient, robustness to outliers, and avoidance of gradient disappearance and explosion problems, which make it excellent in regression tasks such as de-noising. It is an improvement on L1 loss and L2 loss designed to combine the advantages of both: the robustness of L1 loss to outliers and the numerical stability of L2 loss. Assuming x is the numerical difference (residual) between the predicted and real boxes, the formula is as follows (to control the transition point between L1 and L2 losses is set to 1) :

$$smoothL_1(x)=\begin{cases}0.5x^2 & if\ |x|<1\\ |x|-0.5 & \text{otherwise}\end{cases} \tag{6}$$

Derivation:

$$\frac{\mathrm{d}\,smoothL_1}{dx}=\begin{cases}x & if\ |x|<1\\ \pm 1 & \text{otherwise}\end{cases} \tag{7}$$

Smooth L1 loss Compared with L1 loss, smooth L1 loss is improved. Compared with L2 loss, when x is large, it is not sensitive to outliers like L2 and is a slowly changing loss. When the residual is small, it appears as L2 loss, which contributes to the numerical stability of the gradient. When the residuals are large, it appears as an L1 loss, which makes the model more robust to outliers. Due to its piecewise linearity, Smooth L1 Loss is differentiable over the entire domain, which makes it perform well in gradient descent optimization algorithms.

4. Experimental Results and Analysis

The proposed algorithm N-Net was tested in three datasets respectively, and the test results were compared with other algorithms. The comparison algorithms include: ZERO (Z-Score Normalization)[9], URetinex(a deep expansion network based on Retinex theory), ILL

ISSN 2616-5767 Vol.8, Issue 1: 56-64, DOI: 10.25236/AJETS.2025.080109

(Illumination-Adaptive Transformer network), SMNet (Synchronous Multi-Scale Dark Light Enhancement Network)[12].

Figure 7 shows the recovery results of different algorithms on the LOL dataset.

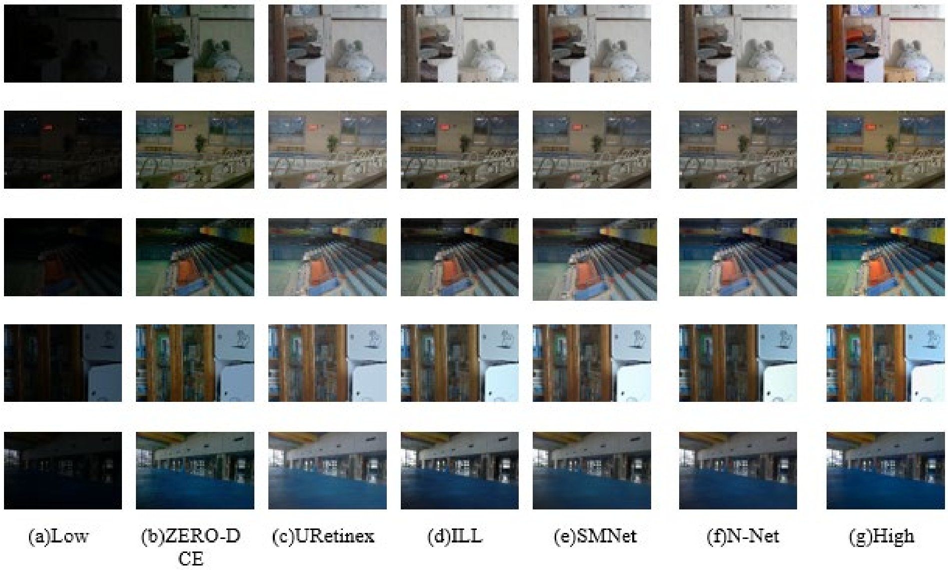

Figure 7: Restoration results of the same algorithm on the LOL dataset

Through comparison, it is found that the effect of ZERO-DCE algorithm on the shadow part is dark, and the visibility of local details is low, as shown in Figure 7(b). This may be due to the fact that the curve in the Zero-DCE method must be monotonically increasing, otherwise it may appear that the originally brighter parts of the image will become darker instead. The color contrast of images processed by URetinex algorithm is generally low, which is easy to cause color distortion, as shown in the blue ground part in Figure 7(c). The overall color temperature of the image processed by ILL algorithm is low and there are artifacts, as shown in Figure 7(d). The overall color temperature of the image processed by SMNet algorithm is low and the processing effect in the shadow part is not ideal, as shown in the playground part in Figure 7(e). Through observation and comparison, it is found that the proposed algorithm N-Net achieves good noise reduction effect in Figure 7(f), with relatively high detail retention and moderate color contrast, which is more in line with human visual perception.

Figure 8 shows the restoration results of different algorithms on the GladNet-Dataset dataset.

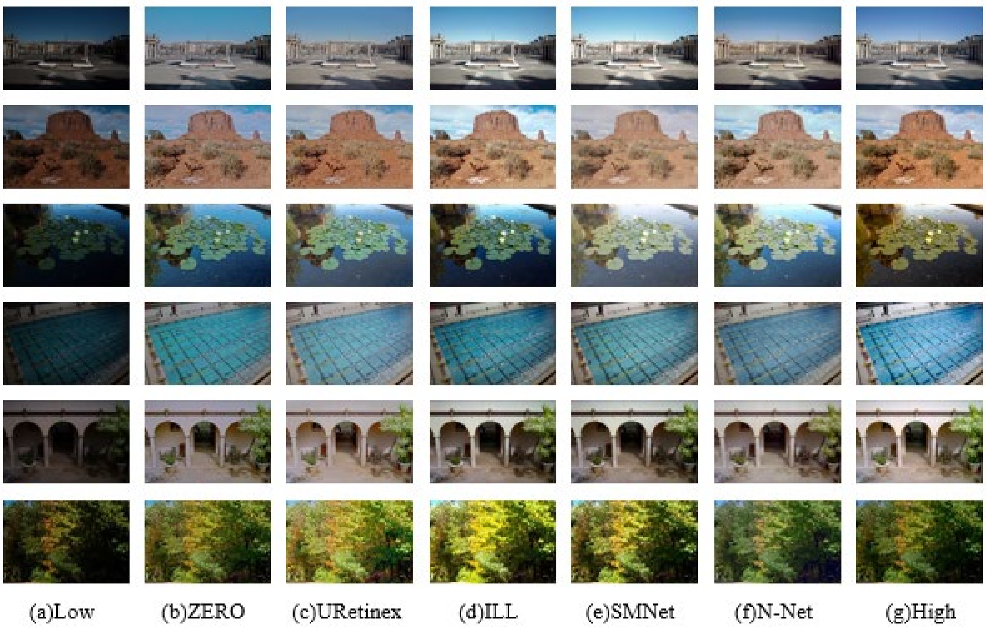

Figure 8: Restoration results of different algorithms on the GladNet-Dataset

ISSN 2616-5767 Vol.8, Issue 1: 56-64, DOI: 10.25236/AJETS.2025.080109

The restoration result is a synthetic data set generated by raw images. By comparing the outdoor natural light images, it is found that the color contrast of the images processed by the ZERO-DCE algorithm is generally high, which is easy to cause color distortion, as shown in Figure 8(b). After URetinex algorithm processing, the overall color temperature of the image is high, and the effect of fog removal on outdoor images is not good, as shown in Figure 8(c). This may be due to the large noise of the input image or the low resolution, which affects the performance of the algorithm. The overall contrast of the image processed by ILL algorithm is high and there are artifacts, such as the leaves and swimming pool in Figure 8(d). The overall color temperature of the image processed by SMNet algorithm is low and there is color distortion, as shown in the sky part in Figure 8(e). Through observation and comparison, it is found that the proposed algorithm N-Net has a reasonable color and no image artifacts after processing in Figure 8(f), and has a good retention of image details.

Figure 9 shows the restoration results of different algorithms on the LSRWhuawei dataset.

Figure 9: Comparison of restoration results of 5 different algorithms on LSRWhuawei dataset

By comparing the images, it is found that the ZERO-DCE algorithm has improved compared with the original low-light image, retaining some details, but the overall brightness is not improved enough, and the details and colors are not rich enough, as shown in the red poster text part in Figure 9(b). The URetinex algorithm balances the brightness and contrast of the image well, but in some cases it may over-enhance certain areas, resulting in loss of detail or increased noise, as can be clearly seen in the sixth figure in Figure 9(c). ILL algorithm is improved in brightness and detail by iterative method, but it is not effective in dealing with complex scenes and requires a long calculation time. The SMNet algorithm, through spatial and multi-scale processing, has better performance in detail and texture, but produces unnatural enhancement effects in some areas, such as excessive smoothness of the ground in Figure 9(e). Through observation and comparison, it is found that the neural network in the proposed algorithm N-Net can learn complex features and has a good performance in image enhancement.

In addition, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were used to quantify the accuracy of each type for image quality assessment. The higher the PSNR, the better the image quality and the less distortion. The closer the SSIM value is to 1, the better the image quality is and the closer it is to the original image. Universal Image Quality Index (UQI) is an image quality assessment method based on information theory, which takes into account the local mean, variance and correlation of images. The closer the UQI value is to 1, the better the image quality. Learning Perceptual Image Block Similarity (LPIPS) is an image quality evaluation index based on deep learning. The lower the value, the better the image quality and the closer to the original image. Among them, UQI and LPIPS are more focused on the overall quality of the image and human visual perception. Specific results are shown in Table 1, Table 2 and Table 3.

ISSN 2616-5767 Vol.8, Issue 1: 56-64, DOI: 10.25236/AJETS.2025.080109

Table 1: Comparison of objective indicators of different algorithms on the LOL data set

Model	PSNR	SSIM	UQI	LPIPS
Guo-ZERO	14.867	0.782	0.694	0.244
Wu- URetinex	18.398	0.794	0.856	0.301
Cui-ILL	23.382	0.808	0.927	0.180
Lin-SMNet	24.51	0.864	0.916	0.149
N-Net	25.133	0.913	0.937	0.098

Table 2: Comparison of objective indicators of different algorithms on the GladNet-Dataset data set

Model	PSNR	SSIM	UQI	LPIPS
Guo-ZERO	18.398	0.746	0.736	0.136
Wu- URetinex	20.136	0.755	0.871	0.223
Cui-ILL	18.143	0.784	0.223	0.153
Lin-SMNet	17.138	0.722	0.857	0.141
N-Net	19.427	0.827	0.866	0.124

Table 3: Comparison of objective indicators of different algorithms on the LSRW\huawei data set

Model	PSNR	SSIM	UQI	LPIPS
Guo-ZERO	16.404	0.776	0.718	0.257
Wu- URetinex	19.144	0.755	0.854	0.255
Cui-ILL	17.192	0.822	0.736	0.228
Lin-SMNet	17.242	0.752	0.794	0.231
N-Net	22.242	0.866	0.868	0.225

From these three tables, it can be seen that the N-Net algorithm has the highest PSNR value on three different data sets, which indicates that the pixel-level difference between the images generated by N-Net on these data sets and the original image is minimal. Overall, N-Net algorithm performs well in image quality recovery, both in pixel level error (PSNR), structure information retention (SSIM), overall image quality (UQI) and perceived similarity (LPIPS). This shows that N-Net algorithm can effectively restore image quality in image processing tasks and provide visual experience closer to the original image.

5. Conclusions

Compared with the other four methods, the PSNR and SSIM values of the proposed model in the LOL dataset can reach 25.133dB and 0.913 at the same time, achieving excellent results. At the same time, the PSNR and SSIM of the GladNet-Dataset noise image recovery reached 19.427dB and 0.827. The results show that the model proposed in this paper has a remarkable effect on the enhancement of the dark light image adapted to non-uniform illumination, and a good effect on the natural dark light scene image. However, the problem of dark light enhancement for image colour bias needs to be further improved, which may be caused by the algorithm's processing of colour information is not accurate enough or the limitation of the algorithm itself. For this problem, I will constantly adjust and optimize to achieve the best results.

References

[1] Y. Chen. Research on image denoising algorithm based on image filter optimization[D]. Southeast University, 2020.

[2] Z. F. Li, L. J. Gao, T. H. Zhang. Simulation of Multi-Feature Detail Recognition of ExposedImage Based on Filtering Algorithm[J]. Computer Simulation, 2023, 40 (12): 242-246+395.

[3] Y. C. Hou, H. Song. Remote Sensing Image Denoising Based on WaveletTransform and Improved DnCNN[J]. Computer Measurement & Control, 2022, 216-221.

[4] T. Wang, C. Zhao, Z. Z. Chen, et al. Research on denoising method of high frequency ground wave radarocean echo spectrum based on DnCNN network[C]// Chinese Institute of Electronics. Proceedings of the 18th National Annual Conference on Radio Transmission. School of Electronic Information, Wuhan University, 2023: 4.

[5] Wu Y, Xu Q, Zhang Z, et al. SAR Change Detection Algorithm Combined with FFDNet Spatial Denoising [J]. Journal of Environmental & Earth Sciences, 2023, 5 (2): 1-15.

ISSN 2616-5767 Vol.8, Issue 1: 56-64, DOI: 10.25236/AJETS.2025.080109

[6] Y. Guo, R. L. Liang, R. M. Wang. Cross-Domain Adaptive Object Detection Based on CNN Image Enhancement in Foggy Conditions[J]. Computer Engineering and Applications, 2023, 59(16): 187-195.
[7] X. Z. Yang, W. Wu, T. S. Ren, et al. Complex Environment License Plate Recognition Algorithm Basedon Improved Image EnhancementandCNN[J]. Computer Science, 2024, 51 (S1): 574-580.
[8] W. T. Jiang, Y. F. Bu. Image Denoising Network Fusing with CNN and Transformer[J]. Computer Systems & Applications, 2024, 33(07): 39-51.
[9] C. G, C. L, J. G, et al. Zero-reference deep curve estimation for low-light image enhancement[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, 1777-1786.
[10] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang and J. Jiang, "URetinex-Net: Retinex-based Deep Unfolding Network for Low-light Image Enhancement," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 5891-5900.
[11] Saputra W V, Suciati N, Fatichah C. Low Light Image Enhancement in License Plate Recognition using URetinex-Net and TRBA[J]. Procedia Computer Science, 2024, 234404-411.
[12] S. Lin, F. Tang, W. Dong, X. Pan and C. Xu, "SMNet: Synchronous Multi-Scale Low Light Enhancement Network With Local and Global Concern," in IEEE Transactions on Multimedia, vol. 25, pp. 9506-9517, 2023.
[13] Z. J. Zhu, G. Q. Zhao, Y. Gao, et al. Compressible Flowfields Machine Learning Inference Method for Airfoil Based on UNet[J]. Journal of Nanjing University of Aeronautics & Astronautic, 2024, 56(02): 253-263.
[14] Liu Y, Shao Z, Hoffmann N. Global attention mechanism: Retain information to enhance channel-spatial interactions[J]. arxiv preprint arxiv: 2112. 05561, 2021.
[15] Xu W, Wan Y. ELA: Efficient Local Attention for Deep Convolutional Neural Networks[J]. arXiv preprint arXiv: 2403. 01123, 2024.
[16] X. Liu, Y. J. Sun. Seismic Signal Denoising Based on Convolutional Neural Networkwith Residual and Attention Mechanism[J]. Journal of Jilin University(Earth Science Edition), 2023, 53(02): 609-621.
[17] W. Wang, C. Wei, W. Yang and J. Liu, "GLADNet: Low-Light Enhancement Network with Global Awareness," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 2018, pp. 751-755.