

Received 20 November 2022, accepted 2 December 2022, date of publication 6 December 2022, date of current version 12 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3227069

RESEARCH ARTICLE

Low Light Image Enhancement Based on Multi-Scale Network Fusion

XUAN LIU¹, CHENFENG ZHANG¹, YINGZHI WANG¹, KAI DING¹, TAILIN HAN¹,
HONG LIU², YU TIAN², BO XU¹, AND MINGCHI JU¹

¹School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China

²School of Optoelectronic Engineering, Changchun University of Science and Technology, Changchun 130022, China

Corresponding author: Tailin Han (hantl@cust.edu.cn)

This work was supported in part by the funding from Jilin Province Science and Technology Development Plan Project under Grant 20210201042GX, and in part by the Innovation Center of Jilin Provincial Department of Science and Technology under Grant 20200602005ZP.

ABSTRACT At present, researchers have made great progress in the research of object detection, however, these studies mainly focus on the object detection of images under normal lighting, ignoring the target detection under low light. And images in the fields of automatic driving at night and surveillance are usually obtained in low-light environments. These images have problems such as poor brightness, low contrast, and obvious noise, which lead to a large amount of information loss in the image. And the performance of object detection in low light is reduced. In this paper, we propose a low-light image enhancement method based on multi-scale network fusion to solve the problems of images in low-light environments. Aiming at the problem that the effective information of low-light images is relatively small, we propose a preprocessing method for image nonlinear transformation and fusion, which improves the amount of available information in the light image. Then, in order to obtain a better enhancement effect, a multi-scale feature fusion method is proposed, which fuses features from different resolution levels in the network. The details of low-light areas in the image are improved, and the problem of feature loss caused by too deep network layers is solved. The experimental results show that our proposed method can achieve better enhancement effects on different datasets compared with the current mainstream methods. The average recall value of the object detection with our method is improved by 38.25%, which shows that our proposed method is effective and can promote the development of autonomous driving, monitoring, and other fields.

INDEX TERMS Image preprocessing, low illumination image enhancement, multi-branch network fusion.

I. INTRODUCTION

With the development of technology, computer vision has been widely used in various fields, such as autonomous driving, video surveillance, object detection, classification, etc. However, in our daily life, weather, lighting, etc. are the key factors that affect image quality. Images obtained in uneven illumination or weak light environments have some problems such as small overall gray value, low brightness, low contrast, and low signal-to-noise ratio, which reduce the amount of information contained in the image. And these problems will reduce the performance of image processing algorithms,

The associate editor coordinating the review of this manuscript and approving it for publication was Antonio J. R. Neves¹.

causing algorithms such as object detection and classification to fail to achieve expected results, and even cause danger to life in the field of autonomous driving. Therefore, it is necessary to study low-light image enhancement algorithms, which has aroused the interest of many researchers.

Traditional algorithms and deep learning-based algorithms have dominated low-light image processing technologies in recent years. Histogram equalization methods and methods based on the Retinex model are both traditional low-light enhancement methods, with the former being more common algorithms like Contrast Limited Adaptive Histogram Equalization [1](CLAHE) and Ibrahim's Brightness Preserving Dynamic Histogram Equalization [2](BPDHE) and so on. In the processed image, this approach results in

unnatural contrast and detail loss. Jobson proposed the Single Scale Retinex (SSR) and Multi-Scale Retinex (MSR) [3], [25], both of which are based on Retinex theory. And other sub-algorithms of Retinex, such as the method with color restoration MSRCR (Multi-Scale Retinex with Color Restoration) [4]. These algorithms estimate the illumination map in the number field, but they frequently cause noise and take a long time to run. Compared with traditional algorithms deep learning-based methods [5], [6], [23], [24], [29], [37], [38] are faster and have better accuracy and robustness, so deep learning-based algorithms are gradually becoming mainstream, such as Lore et al. proposed a deep self-encoder-based algorithm [7](LLNet) to identify low luminance image features and enhance the image by a self-supervised enhancement network. Although the image luminance is effectively improved, the enhanced image will have an overexposure problem, leading to image distortion. There are also convolutional neural networks for image enhancement based on retinex theory, such as the Retinex-Net algorithm [9], [22], [26] proposed by Wang et al, which consists of a Decom-Net module capable of decomposing the input image into light-independent reflectance, structure-aware smooth illumination, and a luminance enhancement module Enhance-Net module, and this method can be used to improve image quality. Although this approach recovers color and brightness adequately, the resulting image is noisy. In addition, the Zero-Dce++ algorithm [8] treats light enhancement as an image-specific curve estimation task, taking a low-light image as input and producing higher-order curves as output. These curves are used as adjustments to the dynamic range of the image at the pixel level and result in a final enhanced image. This curve-based method does not require any paired or unpaired data during training. They implement zero-reference learning through a set of non-reference loss functions. Semi-supervised learning has been proposed in recent years. Yang et al. proposed a semi-supervised deep recurrent band network (DRBN) [8]. DRBN first recovers linear band representations of augmented images under supervised learning, and then recombines the given bands through a learnable linear transformation based on unsupervised adversarial learning to obtain improved band representations. DRBN is extended to achieve better augmentation performance by introducing a long short-term memory (LSTM) network and an image quality assessment network pre-trained on the analysis dataset. Apart from semi-supervised in-depth learning, Jiang et al. proposed an unsupervised learning method, EnlighenGAN [33], which uses attention-guided U-Net as a generator and global-local discriminators to ensure that the enhanced results look like real normal light images. In addition to global and local antagonistic losses, global and local self-feature preservation losses are also proposed to preserve image content before and after enhancement. At present, low illumination image enhancement algorithms based on deep learning still have some problems, such as noise, artifacts, and distortion. The existing deep learning-based low-illumination picture enhancement

algorithm exhibits noise, artifacts, and distortion in the improved image quality.

This research provides a low illumination image enhancement technique using multi-level network fusion (MFIE-Net) to overcome the problems of noise, artifacts, and distortion that emerge in photos enhanced by the above algorithms. The primary contribution of this algorithm is that: a new enhancement approach is proposed. First, in order to enrich the effective information of the low-light image, the image is first preprocessed with a nonlinear transformation, and then the modified image is fused with the original image. Then use the convolutional neural network to extract features from the fusion image, use the feature extraction module (FEM) and feature enhancement module (EM) to extract and enhance the features of each layer, and fuse the feature maps of the same scale of the network [36], it can effectively alleviate the feature loss caused by the network downsampling. Finally, a loss function that combines perceptual, structural, and color loss is devised to improve the model's enhancing impact. The results of the experiments reveal that the algorithm effectively reduces image noise without causing color distortion and that the evaluation index is higher than that of other techniques.

II. MULTI-SCALE FEATURE FUSION NETWORK

A. STRUCTURAL DESIGN

This research proposes an end-to-end multi-scale network fusion approach to address issues such as artifacts and noise during image enhancement that are present in previous methods. Figure 1 depicts the network structure.

To begin, the image is logarithmically changed to expand the low gray value part of the image and compress the high gray value part to achieve the effect of enhancing the dark details, followed by channel fusion with the original image before entering the network to enrich the original features; as the network layer depth increases, the shallow features degrade, so the multi-scale network feature fusion method is used to fuse the low layer information with the high layer information. Reduce the amount of information that is lost due to feature loss. The encoder is made up of a 3×3 size convolutional kernel, while the decoder is made up of a transposed convolutional and activation layer with the same scale. The feature-extracted image is then fused with the input image, and the final enhanced image is output.

1) IMAGE LOGARITHMIC CHANGE PRE-PROCESSING

Because the brightness of most low-light photographs is too low, the features of the over-dark parts are not visible, and no useful features are retrieved when the images pass through the neural network, the images cannot be effectively upgraded. The majority of current low-light image preprocessing approaches use linear amplification of the overall image gray value, but this will lead to overexposure in some places.

As a result, this design uses logarithmic variation in nonlinear [10] variation to preprocess the image. Logarithmic

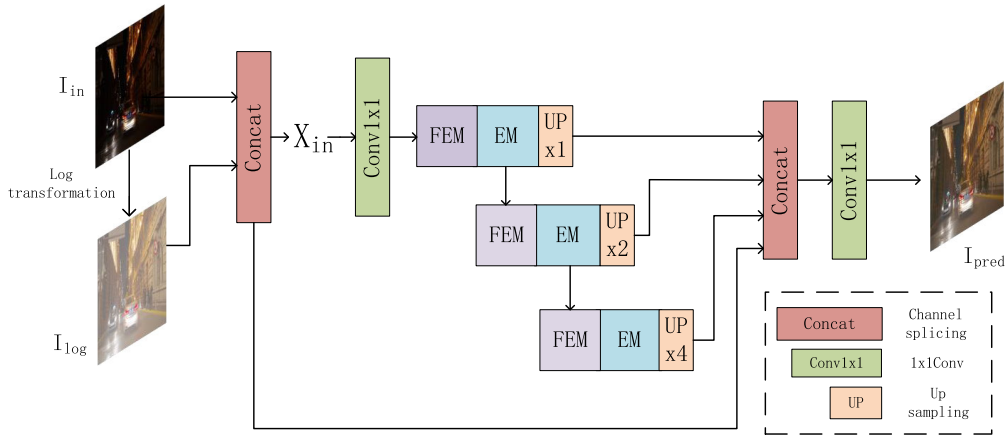


FIGURE 1. This figure shows the overall network structure of the algorithm in this paper. The enhancement process is mainly divided into three steps: logarithmic change preprocessing, multi-scale feature enhancement, and reconstruction. In the logarithmic change preprocessing step, logarithmically change the gray value of the original image to improve the low gray value area features, and fuse with the original image to reduce the feature loss of high gray value. In the multi-scale enhancement step, we perform feature enhancement and extraction at different network scales on the fused image. Finally, all the enhanced feature layers are transformed into feature maps of the same scale for fusion reconstruction.

variation is known by the function curve, whose slope is higher when the gray value is low and lower when the gray value is higher, allowing the darker pixels in the compressed high-value image to be effectively extended. This method can effectively enrich the feature information of the low gray value area, suppress the overexposure area, and improve the feature extraction effect of the network. The formula is as follows:

$$S = c \cdot \log_{v+1}(1 + v \cdot r) \quad (1)$$

Among them, c is a constant, r is the gray value after image normalization, v is a constant, and the larger v is, the more obvious the improvement of gray level is. In order to prevent the loss of detail features of the original image after preprocessing, the processed image and the original image are fused into a six-channel tensor of $RGB * 2$, which is transmitted to the neural network for training.

2) FEATURE EXTRACTION MODULE (FEM) DESIGN

The feature extraction module utilizes a layer of convolution and ReLU activation function [30], [31], where the convolution kernel size of the convolution layer is 3×3 and stride 1. Each layer of convolution is followed by a ReLU activation function to enhance the nonlinear representation of the network. After each layer of FEM, there is an enhancement module pair to enhance the image feature information, and the output of the feature enhancement module (EM) of the previous layer is used as the input of the next layer of the FEM. The network has 3 FEM modules that can extract features at different resolutions.

3) ENHANCEMENT MODULE (EM) DESIGN

The number of feature extraction modules is equal to the number of improvement modules. According to the U-net network

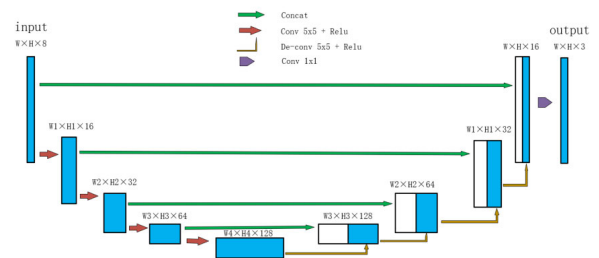


FIGURE 2. This figure shows the enhancement module.

[11], [27], [28] structure, the EM module in this paper uses an encoder-decoder structure (shown in Figure 2), in which the encoding part consists of a 5×5 size convolutional kernel and an activation function layer. The encoder as a whole exhibits a gradually shrinking structure, continuously shrinking the resolution of feature maps to capture contextual information. And it is divided into 4 stages. In each stage, convolution is used for downsampling, and the final feature map is reduced by 16 times. The decoder part consists of a deconvolutional layer and an activation function layer, And it is divided into 4 stages. In each stage, after upsampling the input feature map, it is spliced with the feature map of the corresponding scale in the encoder, and the final feature map is enlarged by 16 times. Unlike U-net, the decoder in this paper does not use a pooling layer, which can reduce the feature loss.

B. LOSS FUNCTION DESIGN

A single loss function cannot provide a decent recovery effect for the difficulties of low-light photos with low brightness and visible noise. In this paper, we consider using the combination of structural similarity loss L_{SSIM} , perception error

loss $L_{Perceptual}$, and color loss L_{colors} as the overall loss function.

1) LOSS OF STRUCTURAL SIMILARITY

In the process of low-light image recovery, which usually leads to structural distortions such as blurring and artifacts, structural similarity can be determined by calculating the contrast, brightness, and structure of the original image and the labeled image to determine the similarity of the two images, using the mean, standard deviation, and covariance to estimate the brightness, contrast and structural similarity, respectively. As shown in Equation (2):

$$S = \frac{2u_x u_y + C_1}{u_x^2 + u_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (2)$$

among u_x, u_y is the image of x and y , respectively, represents the standard deviation of the images x and y , represents the variance of the image x and y . And for constant fractions, they are set to 0.01 and 0.02, in this case, to prevent a zero denominator. The structural similarity of two identical images is equal to 1, so the expression of the structure loss function is designed here as follows:

$$L_{SSIM} = 1 - S = 1 - \frac{2u_x u_y + C_1}{u_x^2 + u_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (3)$$

2) PERCEIVED LOSS

In addition to the low-level information on structural similarity loss, we also need to use the high-level information of the image to recover the image. Referring to algorithms such as style migration, the labeled image is first convolved to obtain the feature map and compared with the features of the enhanced low-illumination image to make the high-level information of both closes. In this paper, we use VGG16 as a feature extraction network for perceptual loss, as in Equation(4):

$$L_{Perceptual}^{\phi, j} = \frac{1}{C_j H_j W_j} \|\phi_j(x) - \phi_j(y)\|_2^2 \quad (4)$$

where j represents the number of layer j of the network, C_j, H_j and W_j represent the channel number, height, and width of the characteristic graph of the layer j , $\phi_j(x)$ and $\phi_j(y)$ represents the feature maps of the enhanced image and labeled image at layer j , respectively. The perceptual loss ensures the quality of the enhanced image through high-level information.

3) COLOR LOSS

The problem of color bias can exist in low-illumination images, L_{SSIM} and $L_{Perceptual}$ functions can enhance the image through the structure and features, but there will be color difference in the enhanced image. Therefore, in this paper, loss functions with color recovery are designed. The RGB pixels are treated as three-dimensional vectors, and the color is adjusted by computing and adding the angles between each vector. The expression is as follows:

$$L_c = \sum_p \mathcal{L}(\sum_i x_i)_p, (\sum_i y_i)_p \quad (5)$$

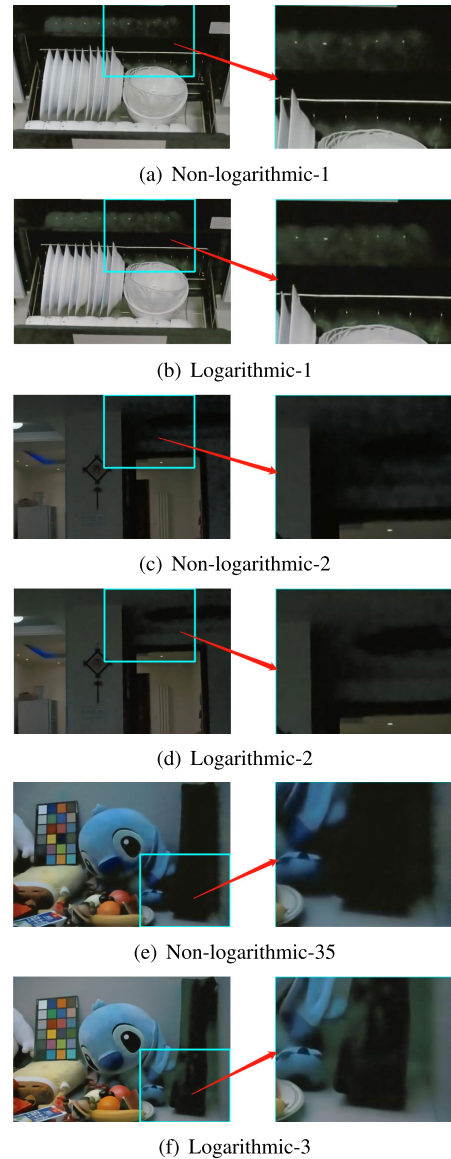


FIGURE 3. (a),(c),(e) represent the network-enhanced image without logarithmic change, and (b),(d),(f) represent the enhanced result of the network with logarithmic change.

In which $\mathcal{L}(\cdot)$ represents an operator for calculating the angle of two colors, (\cdot) stands for a pixel. L_c is the summation of the angles between the color vectors for each pixel pair.

In summary, the total loss function L designed in this paper consists of three components as in Equation (6):

$$L = \sum_p w_1 L_{SSIM} + w_2 L_{Perceptual} + w_3 L_{color} \quad (6)$$

where w_1, w_2 and w_3 represent the weight of each loss function ($w_1 = 0.45, w_2 = 0.45, w_3 = 0.1$).

III. EXPERIMENTAL SETUP

In this paper, we select 15144 images in VOC dataset [14], [17] which are processed with random gamma adjustment as the training set and the test set, with 14,000 sets of images in the training set and 1,144 sets of images in the validation set.

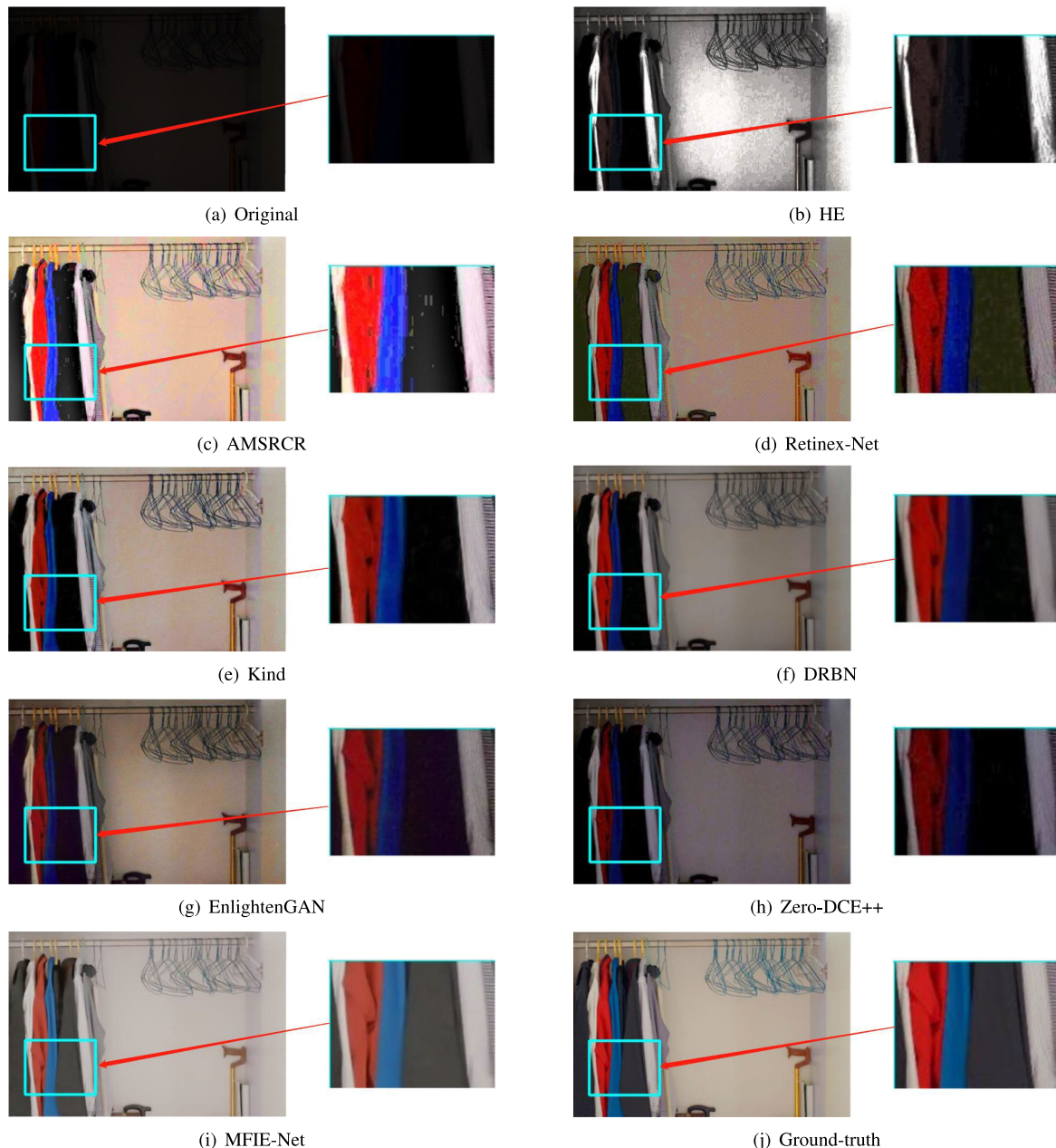


FIGURE 4. Example of low-light enhancement on a testing image. The left side of each group of images is the original-scale image, and the image pointed by the red arrow is the detailed image enlarged by the blue rectangular area in the original-scale image on the left.

To objectively represent the training results, we use the LOL dataset (Low-Light dataset) [15] as the test set. The whole network is trained on NVIDIA GTX 2080Ti GPU and Intel Xeon E5-2690 CPU using the PyTorch framework.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. ABLATION EXPERIMENT RESULT

In order to illustrate the effectiveness of each part proposed in this paper, under the same conditions as other experiments, ablation experiments are performed on the enhanced network

with and without preprocessing. The unprocessed RGB 3-channel image and the pre-processed RGB*2 6-channel image are used as the input of the network respectively. The experimental results of the two groups are compared to prove the effectiveness of fusing the preprocessed images to improve the image enhancement effect.

Through the comparison of the output of the network structure with or without logarithmic change preprocessing in Figure 3, it can be seen that the image preprocessed by logarithmic transformation can effectively improve the image features in the area with a lower gray value, and it can be more

obvious. Improve the enhancement ability of the network in low gray value regions.

B. VISUAL QUALITY COMPARISON

In order to visualize the enhancement effect, the method in this paper is compared with other low-light image enhancement methods for experimental comparison, which mainly include HE of traditional methods and AMSRCR based on Retinex theory, Retinex-Net, Kind [34], DRBN, EnlightenGAN and Zero-DCE++ based on deep learning. We performed experiments on images with different lighting conditions. The LOL dataset and Night-driving dataset provided the images, and Figure 4 displays the results. After enhancement by the HE method, the gray level is reduced, the image details are lost and the contrast is unnatural, the AMSRCR method will have problems such as over-enhancement of color and contrast, and the image will also be noisy, the Retinex-Net method has better color recovery but the recovered image has obvious noise, Kind and DBRN algorithms can enhance the color and brightness of low-light images, but the effect of enhancing local features with the low gray value of the image is not ideal. The image enhanced by EnlightenGAN algorithm which bases on the unsupervised learning has noise. Although the Zero-DCE++ algorithm can effectively enhance the image details, the overall image brightness is low. In comparison to other algorithms, the MFIE-Net algorithm proposed in this paper is more perfect in overall and local details, more balanced in color recovery, and the image details are clearly discernible and do not generate additional noise, resulting in a visual effect that is closer to that of the label image.

In order to objectively reflect the effect of the algorithm, besides subjective evaluation, PSNR, SSIM [16] and LPIPS are used to evaluate the performance of the model.

PSNR: peak signal-to-noise ratio (PSNR), commonly used as a measure of image and signal reconstruction quality. The larger the value, the smaller the image distortion.

$$PSNR = 10 \cdot \log_{10}(\frac{MAX_I^2}{MSE}) \tag{7}$$

where MAX_I^2 represents the maximum possible pixel value for the image, if each pixel is represented by 8-bit binary, then its value is 255. MSE is the mean square error between the original images and the resulting images.

SSIM: structure similarity, which measures the similarity of two images in brightness, contrast, and structure. The larger the value, the more similar the two images are.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

In which μ_x and μ_y represent the mean and variance of the image, respectively. c_1 and c_2 stand for two constants, avoiding division by zero.

LPIPS: learned perceptual image patch similarity, used to measure the difference between two images [32]. The lower

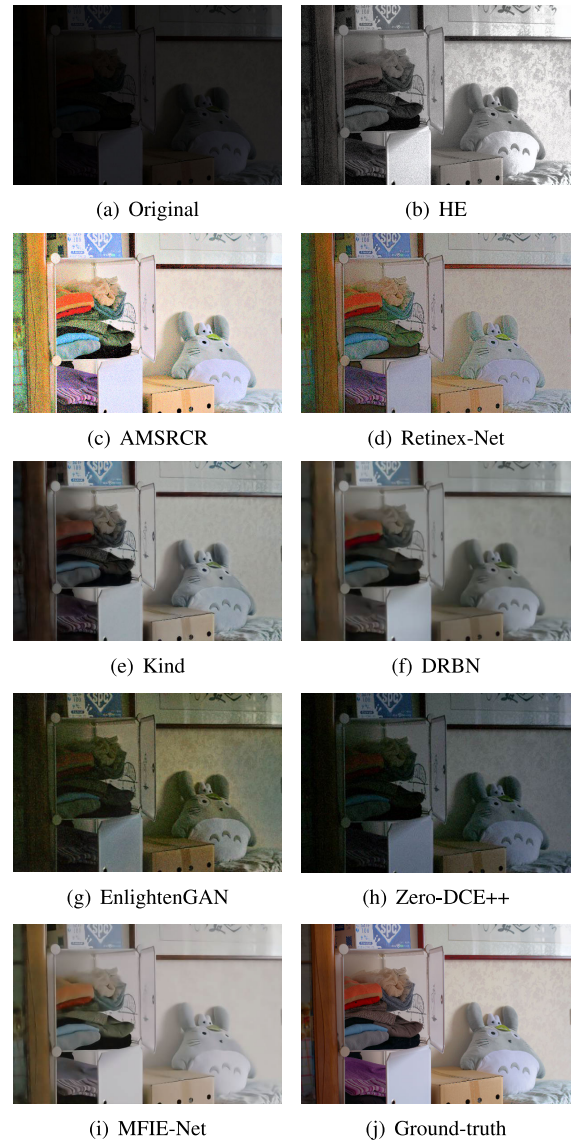


FIGURE 5. Example of low-light enhancement on other low-light images of testing dataset.

the value, the more similar the two images are.

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{y}_{hw}^l - \hat{y}_{0hw}^l)\|_2^2 \tag{9}$$

where, d is the distance between x and x_0 . Feature stacks are extracted from the L layer and unit-normalized in the channel dimension. And $\hat{y}^l, \hat{y}_0^l \in \mathbb{R}^{H_l \times W_l \times C_l}$ is for layer l . The activations are scaled channel-wise by vector $w^l \in \mathbb{R}^{C_l}$.

Table 1 shows the comparison results with other methods on the LOL dataset(testing pairs). From the data, it can be seen that the method in this paper has advantages in PSNR, SSIM and LPIPS indexes, which indicates that the algorithm in this paper is more effective in low-light image enhancement.

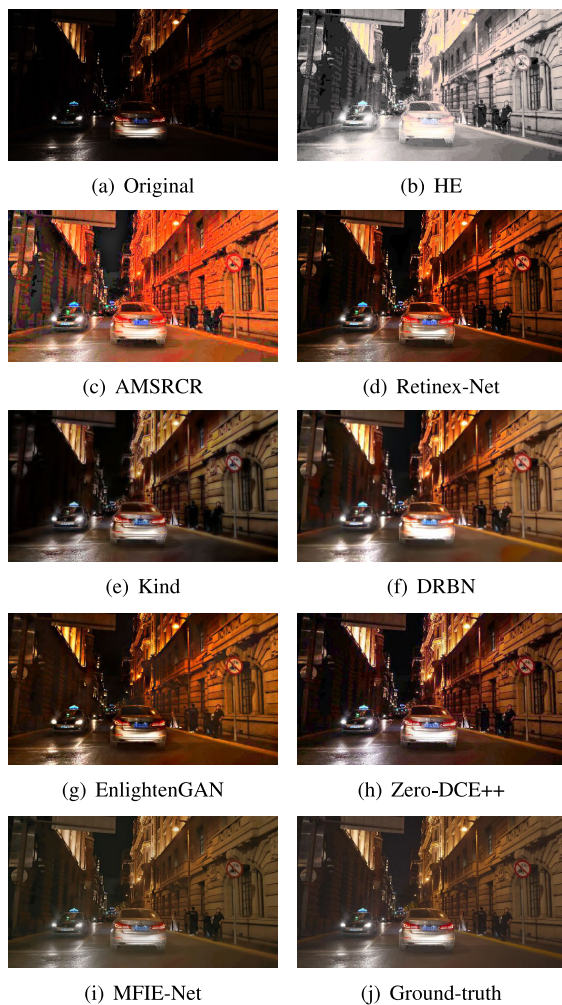
Table 2 shows the comparison results of the image processing time of different algorithms. Although the traditional

TABLE 1. Average PSNR and SSIM. The best data is marked in red.

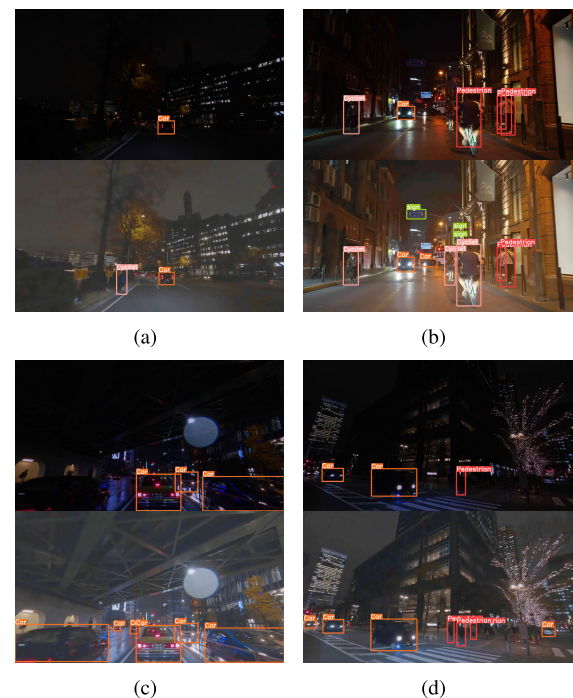
Method	HE	AMSRCR	Retinex-Net	Kind	DRBN	EnlightenGAN	Zero-DCE++	MFIE-Net
PSNR(dB)	11.62	15.16	17.18	20.69	16.71	16.67	16.47	24.91
SSIM	0.784	0.853	0.955	0.963	0.918	0.931	0.853	0.987
LPIPS	0.469	0.308	0.355	0.211	0.239	0.226	0.191	0.147

TABLE 2. Average time of processing images.

Method	HE	AMSRCR	Retinex-Net	Kind	DRBN	EnlightenGAN	Zero-DCE++	MFIE-Net
Time(s)	0.01	2.95	2.28	0.65	0.58	0.47	0.45	0.49

**FIGURE 6.** This figure shows the comparison between the algorithm in this paper and other algorithms in the actual night road environment enhancement effect.

HE algorithm has the fastest processing time, it has the worst image enhancement effect. Compared with other deep learning-based algorithms, in spite of EnlightenGAN and Zero-DCE++ are slightly faster than the proposed method, our proposed method has clear advantages in terms of vision and evaluation metrics, so our algorithm is more suitable for low-light image processing field.

**FIGURE 7.** Comparison of detection effect before and after enhancement(under image).

In order to prove that the algorithm proposed in this paper has a good enhancement effect in different scenarios, Figure 5 and Figure 6 mainly show the comparison between the algorithm in this paper and other algorithms in other datasets and actual night road pictures. As can be seen from other photos in the test set, the MFIE-Net algorithm does better in terms of brightness, contrast, and noise suppression. In the actual night road scene, the image enhanced by the algorithm in this paper is the closest to the real image in terms of brightness and color. It intuitively shows that the algorithm in this paper is excellent.

C. NIGHTTIME DETECTION

In the actual night detection process, in order to reflect the enhanced effect will be enhanced before and after the image for comparison, this paper uses YOLOv5 [18], [19], [20], [21] detection algorithm to compare the original image and the enhanced image of this paper algorithm.

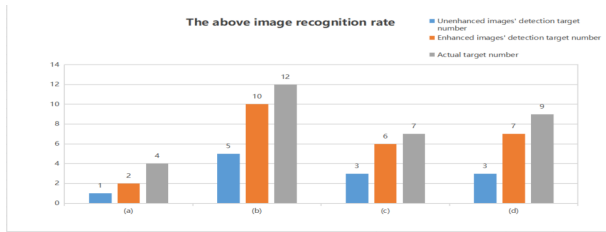


FIGURE 8. Data histogram detection effect before and after enhancement.

As shown in Figure 7 and Figure 8, in the above 4 groups of images, the first row of images is the detection result without low-light enhancement, and the second row of images is the detection result after low-light image enhancement. The recall of (a) is increased from 0.25 to 0.5, (b) is increased from 0.5 to 0.83. Especially in (a), the recognition algorithm can accurately identify roadside cyclists and pedestrians, which can effectively reduce the missed detection rate of visual recognition at night. It can be clearly seen that the detection effect after image enhancement has a significant improvement (including (c), (d)), and average recall increased by 38.25%, in keeping within the actual application requirements.

V. DISCUSSION

This section states some discussions and future works.

A. IMAGE LOGARITHMIC CHANGE PRE-PROCESSING

The logarithmic curve has a large slope where the pixel value is low, and a small slope where the pixel value is high. After the image is logarithmically transformed, the contrast of the darker area will be improved, so the features of the low gray value part of the image are enhanced. According to this feature, we logarithmically transform the image and fuse it with the original image into an RGB 6-channel image as the input of the network. Through the ablation experiments with and without logarithmic change preprocessing, it can be clearly seen that logarithmic change preprocessing can effectively enhance the feature information of the dark part of the image.

B. MULTISCALE NETWORK DESIGN

We use a combination of multiple feature extraction and feature enhancement modules to enhance the image, using a convolution kernel with a convolution kernel size of 3×3 and a ReLU activation function as the feature extraction module. The feature enhancement module is similar to U-Net, in order to prevent feature loss, we discarded the pooling layer. Through image processing by multiple feature extraction and enhancement modules, we can extract image features at different resolutions, then the feature maps at different resolutions are unified to the same scale using bilinear interpolation upsampling, and finally, we use the 1×1 convolution kernel to merge them.

C. FUTURE WORK

In the future, our work will focus on using methods such as dynamic range conversion methods [35] to study the influence of network visualization and network intermediate layers on low-light image enhancement to further improve low-light image enhancement. Secondly, we will build data sets for specific fields such as night road scenes and night surveillance videos to improve the algorithm performance of specific scenes. Finally, we intend to cascade the algorithm network of this paper and the target detection network to improve the practical application of the algorithm.

VI. CONCLUSION

In this paper, we have proposed a multi-scale network fusion enhancement method (MFIE-Net), which solves the problem of low-light image enhancement and improves the enhancement effect. We first perform nonlinear transformation on the input low-light image and fuse it with the input image to achieve information enhancement in the low-light area, so that the network can learn more features of the low-light area. Then in order to improve the enhancement effect of the low-light image, we design a multi-scale feature fusion module to fuse the features of different levels of resolution in the low-light area. It improves the detail information in low-light areas, fuses deep features with shallow features, and avoids feature loss caused by too deep network layers. The experimental results show that our method is better than other algorithms in terms of PSNR, SSIM, LPIPS, and other evaluation indicators. And the speed of our method is similar to the latest methods. In summary, our method can be applied to low-light fields such as automatic driving at night and video surveillance and helps to improve the performance of tasks such as object detection and recognition in low-light environments.

REFERENCES

- [1] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, vol. 38, no. 1, pp. 35–44, 2004.
- [2] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.
- [3] Z.-U. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Sep. 1996, pp. 1003–1006.
- [4] A. B. Petro, C. Sbert, and J.-M. Morel, "Multiscale retinex," *Image Process. Line*, vol. 4, pp. 71–88, Apr. 2014.
- [5] C. Dong, C. C. Loy, K. He, and X. Tang, "Image superresolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Apr. 2015.
- [6] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4364–4375, Sep. 2019.
- [7] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.
- [8] C. Li, C. Guo, and C. L. Chen, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4225–4238, Apr. 2022.
- [9] Z. Jiang, H. Li, L. Liu, A. Men, and H. Wang, "A switched view of retinex: Deep self-regularized low-light image enhancement," *Neurocomputing*, vol. 454, pp. 361–372, May 2021.

- [10] M. Ravikumar, B. J. Shivaprasad, and D. S. Guru, "Enhancement of MRI brain images using notch filter based on discrete wavelet transform," *Int. J. Image Graph.*, vol. 22, no. 1, Jan. 2022, Art. no. 2250010.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
- [12] S. Parthasarathy and P. Sankaran, "An automated multi scale retinex with color restoration for image enhancement," in *Proc. Nat. Conf. Commun. (NCC)*, Feb. 2012, pp. 1–5.
- [13] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1780–1789.
- [14] W. Wang, X. Wu, X. Yuan, and Z. Gao, "An experiment-based review of low-light image enhancement method," *IEEE Access*, vol. 8, pp. 884–887, 2020.
- [15] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," 2018, *arXiv:1808.04560*.
- [16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [17] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, pp. 303–338, Jul. 2010.
- [18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
- [19] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [20] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.
- [21] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [22] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.
- [23] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using CNNs," in *Proc. BMVC*, 2018, vol. 220, no. 1, p. 4.
- [24] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.
- [25] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–128, Dec. 1977.
- [26] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. BMVC*, 2018, pp. 1–15.
- [27] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++," *Remote Sens.*, vol. 11, no. 11, p. 1382, 2019.
- [28] H. Wang, P. Cao, J. Wang, and O. R. Zaiane, "UCTransNet: Rethinking the skip connections in U-Net from a channel-wise perspective with transformer," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 3, pp. 2441–2449.
- [29] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.
- [30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [31] Q. Chen, J. Xu, and V. Koltun, "Fast image processing with fully-convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2497–2506.
- [32] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.
- [33] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and C. Fang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.
- [34] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 1632–1640.
- [35] S. Mukherjee, G.-M. Su, and I. Cheng, "Adaptive dithering using curved Markov-Gaussian noise in the quantized domain for mapping SDR to HDR image," in *Proc. Int. Conf. Smart Multimedia*. Cham, Switzerland: Springer, 2018, pp. 193–203.
- [36] Y. Zhang, L. Fu, Y. Li, and Y. Zhang, "HDFNet: Hierarchical dynamic fusion network for change detection in optical aerial images," *Remote Sens.*, vol. 13, no. 8, p. 1440, Apr. 2021.
- [37] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang, "RetinexDIP: A unified deep framework for low-light image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1076–1088, Mar. 2022.
- [38] J. Liu, D. Xu, W. Yang, M. Fan, and H. Huang, "Benchmarking low-light image enhancement and beyond," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1153–1184, Apr. 2021.



XUAN LIU received the bachelor's degree in science, in 2014. He is currently pursuing the Ph.D. degree with the Changchun University of Science and Technology. His research interests include intelligent information processing and photoelectrical imaging.



CHENFENG ZHANG was born in Jilin, China, in 1995. He received the B.Eng. degree in electronic information engineering, in 2018. He is currently pursuing the M.Eng. degree in electronics and communication engineering with the Changchun University of Science and Technology, China. His research interest includes image processing. He won a National Scholarship, in 2020.



YINGZHI WANG received the B.Eng. degree in applied electronic technology, in 2002, and the M.Eng. degree in electronics and communication engineering, in 2007. He is currently working as a Professor with the School of Electronic and Information Engineering, Changchun University of Science and Technology. His current research interests include sensor signal processing and embedded systems.



KAI DING received the Bachelor of Engineering degree, in 2019. He is currently pursuing the master's degree with the Changchun University of Science and Technology. His main research interests include image processing and object detection.



TAILIN HAN received the B.Eng. degree in applied electronic technology, in 1991, and the M.Eng. and Ph.D. degrees in physical electronics, in 1999 and 2010, respectively. He is currently working as a Professor and a Doctoral Supervisor with the School of Electronic Information Engineering, Changchun University of Science and Technology. He presided over and participated in many national and provincial projects. He has published several scientific papers in national and international conference proceedings and journals, including two books, 24 papers, and 14 patents. His current research interests include special signal acquisition and processing, and digital image processing. He has received the title of the 14th Batch of Young and Middle-Aged Professionals and Technical Talents with Outstanding Contributions, Jilin, in 2014; the Third Prize of the Jilin Provincial Science and Technology Award, in 2015; and the Second Prize of Jilin Province Science and Technology Award, in 2019.



YU TIAN received the bachelor's degree, in 2020. She is currently pursuing the master's degree with the Changchun University of Science and Technology. Her research interests include photoelectric detection technology and quality control.



BO XU received the B.Eng. degree in communication engineering, in 2016. She is currently pursuing the Ph.D. degree in information and communication engineering with the Changchun University of Science and Technology, China. Her research interests include special signal processing, digital signal processing, dynamic signal detection and processing, and transient signal compressed sensing.



HONG LIU received the B.Eng. degree in applied electronic technology, in 1991, and the M.Eng. degree in physical electronics, in 2004. She is currently working as a Professor and a Master's Supervisor with the School of Optoelectronics and Information, Changchun University of Science and Technology. Her current research interests include modern testing technology, image processing, and pattern recognition.



MINGCHI JU received the B.Sc. degree in electronic information science and technology, in 2018. He is currently pursuing the Ph.D. degree in information and communication engineering with the Changchun University of Science and Technology, China. His research interests include special signal processing, digital signal processing, and transient signal compressed sensing.

...