International Journal of Micro Air Vehicles

ORIGINAL RESEARCH ARTICLE

Open Access

https://doi.org/10.48130/mav-0026-0006

International Journal of Micro Air Vehicles **2026**, 18: e006

FA-UNet: a fast and accurate UNet for low-light image enhancement in unmanned aerial vehicle vision navigation

Yu Cheng[1], Xixiang Liu[1*], Shuai Chen[2] and Chuan Xu[2]

[1] *School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China*
[2] *School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China*
* Correspondence: 101010902@seu.edu.cn (Liu X)

Abstract

Visual navigation is a crucial method for the high-precision autonomous navigation of unmanned aerial vehicles (UAVs). However, in low-light environments, aerial images suffer severe loss of textural detail, rendering visual navigation ineffective. In this paper, we propose a fast and accurate UNet (FA-UNet) for low-light image enhancement (LLIE) that is suitable for visual navigation. First, we introduce a dynamic switching mechanism based on illumination perception, which can adaptively select enhancement operations and effectively reduce computational redundancy. Second, inspired by zero-reference deep curve estimation, we use the FA-UNet to get the optimal light enhancment fitting curve parameters for low-light images. We add skip connection mechanism and depthwise separable convolution for lightweight implementation and introduce a convolutional attention module to help the network identify important features across channels and image spatial regions. Finally, we conduct experimental verification through public datasets and data collected from actual UAV flights. Experiments show that FA-UNet has good computational efficiency and image enhancement performance.

Keywords: Visual navigation, Low-light image enhancement, Visual scene matching navigation, U-Net, DSC, CBAM

Citation: Cheng Y, Liu X, Chen S, Xu C. 2026. FA-UNet: a fast and accurate UNet for low-light image enhancement in unmanned aerial vehicle vision navigation. *International Journal of Micro Air Vehicles* 18: e006 https://doi.org/10.48130/mav-0026-0006

Introduction

Scene matching navigation systems (SMNSs) use visual sensors to obtain real-time images of the flight area and match them with prestored satellite images to obtain high-precision information on aircraft position[1–3]. Because of its advantages such as navi-gational accuracy, reliability, and autonomy, it has attracted widespread attention from scholars[4]. However, because of the changes in light intensity, aerial images cannot sustain the scene matching task, which makes an unmanned aerial vehicle (UAV) unable to accurately locate itself. Therefore, it is particularly important to explore a low-light image enhancement (LLIE) method that is suitable for visual scene matching in UAVs.

At present, LLIE methods are mainly based on traditional prior knowledge and deep neural networks[5,6]. The methods based on traditional prior knowledge mainly include histogram equilization (HE)[7] and Retinex theory[8]. Cheng et al.[9] combined multipeak HE (MPHE) with local information to improve global HE. The combination of different local information was effective for LLIE. Liu et al.[10] used the stratified parametric-oriented histogram equalization (SPOHE) method to ensure overall contrast and avoid local distortion, and has a low computational cost. Retinex theory[8] was first proposed by Land et al. in 1977. Its essence is to determine the inherent properties of an object by removing the influence of illumination in the image. Hao et al.[11] proposed a Gaussian total variation model to constrain the illumination component and the reflectance component, and decomposed the two components in a semidecoupled manner. Ren et al.[12] first used low-rank regularized Retinex model to minimize the rank of the matrix composed of similar image blocks in the reflection component, thereby reducing the occurrence of artifacts and reducing noise interference. In summary, although LLIE algorithms based on Retinex can obtain both the illumination and reflection components, the enhancement results of such algorithms usually show unclear illumination improvement, low contrast, and unnatural phenomena.

In recent years, with the widespread application of deep learning methods in the field of computer vision, LLIE methods based on deep neural networks have attracted widespread attention from scholars[13,14]. LLIE methods based on deep learning are divided into supervised, unsupervised, and enhancement algorithms without reference images.

Supervised LLIE algorithms

Zhang et al.[15] decomposed the image into a reflection map and an illumination map, and used pairs of images captured under different illumination conditions for training. Zhou et al.[16] proposed LEDNet, which is the first joint low-light enhancement and deblurring method. LEDNet learns the residual between low-light images and normal-light images through an iterative process of enhancing and dimming brightness. Ren et al.[17] trained a fusion network which estimated the global content of the image using an encoder–decoder network and obtained the structure of the image using a dimensional varying recurrent neural network (RNN). Yao et al.[18] proposed a novel LLIE network that restores images' brightness and detail through a two-stage transformation in the spatial and frequency domains. However, these methods must use image pairs of low-light and high-light conditions, which is not practical.

Unsupervised LLIE algorithms

Unsupervised LLIE algorithms refer to networks trained using unpaired images. The earliest unsupervised LLIE algorithm was EnlightenGAN[19]. This algorithm uses an attention-guided UNet as a generator and uses global and local discriminators to make the

https://doi.org/10.48130/mav-0026-0006

enhancement results naturally. Fu et al.[20] used P-Net to remove noise and unreasonable features from the original image, L-Net to estimate the illumination component, and R-Net to estimate the reflection component. Ni et al.[21] used a single deep generative adversarial network (GAN) to obtain richer global and local features with an attention mechanism and restore the corresponding normal-light image from unpaired data. Yang et al.[22] used a hidden neural representation method for synergistic low light enhancement. Liu et al.[23] combined the advantages of Visual State Space Module (VSSM) and Local Feature Module (LFM) to ensure low computational complexity while balancing global and local feature extraction. These unsupervised methods have been shown to have good enhancement performance and generalization capabilities, and do not need paired image datasets. However, this type of algorithm has the limitation of relying on high-quality normal-light images.

LLLE algorithms without reference images

In recent years, obtaining enhancement results through neural networks when only low-light images are available is a direction worthy of in-depth research and exploration. Zhang et al.[24] used ExCNet to directly estimate the best-fitting S-curve for a given dark image. Zhu et al.[25] proposed a robust Retinex decomposition network (RRD-Net), which uses a three-branch convolutional neural network to decompose the robust Retinex model. Guo et al.[26] proposed zero-reference deep curve estimation (Zero-DCE). This algorithm treats the low-light image enhancement task as estimating an image-to-mapping curve. It outputs a series of parameters for pixel-level brightness adjustment of the low-light image and iteratively updates the mapping curve to obtain a brightness-enhanced image. Zero-reference LLIE algorithms are more realistic because they use only low-light images for network training.

Given the difficulties of deploying these LLIE methods on mobile devices and their suboptimal performance, we propose a lightweight image enhancement curve estimation network (FA-UNet) based on Zero-DCE. This method significantly improves the input image quality while effectively enhancing visual localization accuracy and computational efficiency, achieving a good balance between real-time performance and enhanced performance.

Our contributions are as follows.

(1) Considering that low-light environments do not always exist in reality, low-light enhancement of each frame would consume a large amount of computing resources. Therefore, we designed a dynamic switching mechanism for LLIE based on environmental perception. We use the average grayscale value of the image frame to measure the illumination. This mechanism effectively reduces computional redundancy.

(2) We design a fast and accurate network (FA-UNet) to obtain the parameters with the best fit between the input image and the low-light enhanced image. The network uses the encoder–decoder architecture and adds a jump connection mechanism to effectively suppress the gradient's disappearance.

(3) In FA-UNet, we use depthwise separable convolution to replace traditional convolution operations for lightweight implementation, and introduce a convolutional block attention module (CBAM) to help the network identify important features across channels and spatial regions of the image, reducing the model's computational complexity while ensuring model's stable performance and facilitating subsequent mobile deployment.

(4) To achieve low-light enhancement based on zero-reference images, we designed a set of special loss functions, which can evaluate the quality of images effectively.

(5) We conduct quantitative and qualitative analyses of multiple benchmark datasets to compare this method with other mainstream methods. The results show that this method outperforms the mainstream methods in terms of visual effects, and also surpasses mainstream methods in terms of the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) values. We designed visual positioning experiments and computational efficiency experiments, and the results show that FA-UNet significantly improves the performance of visual positioning in low-light environments and is suitable for mobile platforms with limited computing resources.

Proposed methods

Light enhancement curve

Our method's architecture is shown in Fig. 1. The light enhancement (LE) curve can automatically match low-light images to

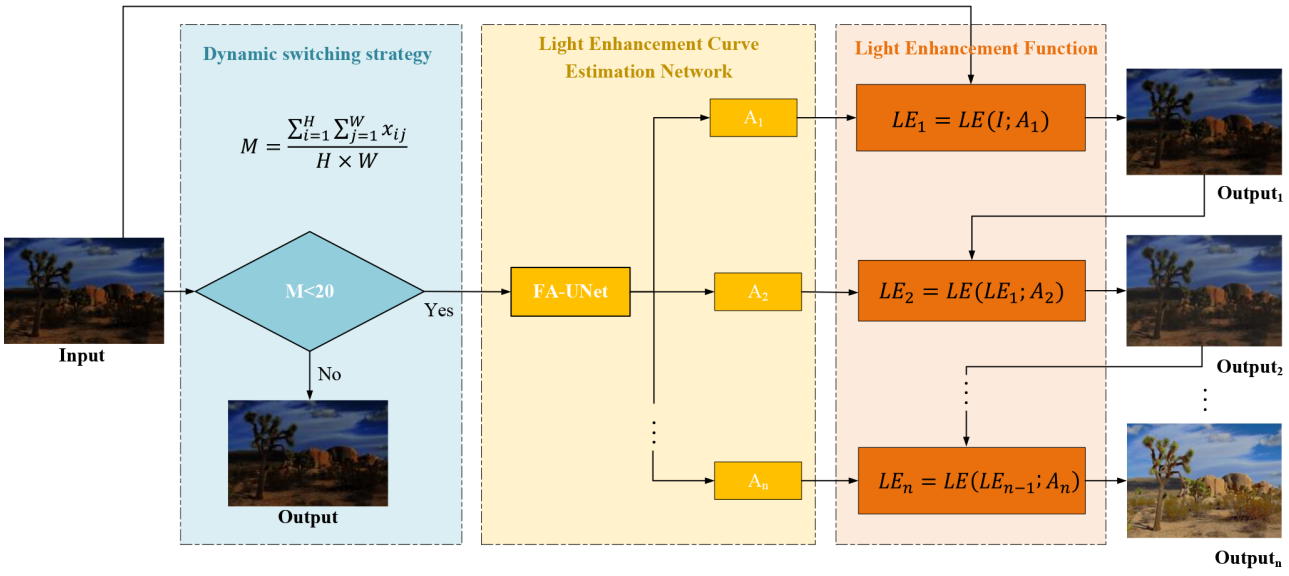

Fig. 1 The LLIE framework.

enhanced ones. The design of this curve has three goals: (1) Every pixel value of the enhanced image must normalize in [0, 1] to avoid information loss caused by overflow truncation; (2) the curve should be monotonic to maintain the difference; (3) during gradient back-propagation, the curve should be possible and differentiable. Therefore, the quadratic curve equation to fulfill these goals is as shown in Formula (1).

$$LE(Input(x);\alpha) = Input(x) + \alpha Input(x)(1 - Input(x)) \quad (1)$$

where x represents the pixel coordinates of a point in the input image, $Input(x)$ is the input image, $LE(Input(x);\alpha)$ is the corrected image, and α is the trainable parameter of the LE curve. The LE curve meets the three objectives mentioned above. To further enhance the generalizability of the curve, the curve is adjusted to a high-order curve through iteration to increase the dynamic adjustment range.

$$LE_n(x) = LE_{n-1}(x) + \alpha_n LE_{n-1}(x)(1 - LE_{n-1}(x)) \quad (2)$$

where n is the number of the LE curve's iterations, which determines the curvature of the curve. When *n* equals 1, Formula (2) degenerates into Formula (1). The generalizability of high-order curves enables them to adjust images more flexibly. However, this is still a global adjustment, which may lead to excessive or insufficient enhancement of local areas. Therefore, α is adjusted to a pixel-by-pixel parameter, and the corresponding enhancement parameter is output according to the illumination value of the pixel. Therefore, Formula (2) can be restated as Formula (3).

$$LE_n(x) = LE_{n-1}(x) + A_n LE_{n-1}(x)(1 - LE_{n-1}(x)) \quad (3)$$

where A_n is the LE parameter sets for the same pixels with $Input(x)$. $LE_n(x)$ meets the three objectives of the light enhancement curve, and each pixel value of the enhanced image is still in the range of [0, 1]. Pixel-by-pixel enhancement means that when the input image is nonuniformly illuminated, not only can low-illuminance areas be enhanced, but overexposed areas can also be suppressed.

The LE curve estimation network structure of FA-UNet

We propose a LE curve estimation network based on the improved UNet to learn the best fitting curve parameter sets of the input low-light image as shown in Fig. 2. The overall framework (Fig. 2) include an encoder–decoder architecture based on UNet. The encoder is the left half. The encoder part applies depthwise separable convolution (blue arrow) and maximum pooling (red arrow) to double the number of image feature maps and halve the image size, respectively. Inspired by characteristics of human vision, we adopt the CBAM attention mechanism (yellow arrow) to make the network focus on useful information. Every CBAM is placed after each depthwise separable convolution (DSC) to capture important features at the corresponding image scale. The decoder is the right half of Fig. 2. The bilinear upsampling operation (green arrow) is used to double the size of the feature maps. We cascade with the upsampled feature maps and the output of the encoder part with a skip connection. Finally, the number of features is halved through a DSC, and we use Tanh to get the final result.

Unlike the traditional UNet network structure, we replace the traditional convolution with DSC to reduce the network calculation effort, although in the CBAM, the traditional convolution is still applied. In addition, we introduce the CBAM into the encoder part, which helps the network identify important features across channels and image spatial regions.

Depthwise separable convolution

The depthwise separable convolution module is introduced to replace traditional convolution for lightweight implementation. The basic idea of the DSC module is to split the convolution process into two independent sub-operations: Depthwise convolution and pointwise convolution, so as to reduce the model's calculation complexity and reduce the number of parameters while ensuring that the model's performance does not deteriorate seriously. The principle is shown in Fig. 3.

The standard convolutional layer performs feature extraction and channel fusion at the same time, and the parameter q_1 and the amount of calculation p_1 are

$$q_1 = D_k \times D_k \times C \times N \quad (4)$$

$$p_1 = D_k \times D_k \times C \times N \times H \times W \quad (5)$$

where D_k is the convolution kernel size, C is the number of input channels, N is the number of output channels, and $H \times W$ is the size of the output feature map.

DSC decomposes the standard convolution process into feature extraction and channel fusion. Each channel is independently convolved in depth during feature extraction. In the channel fusion stage, a 1 × 1 convolution kernel is used to perform point-by-point convolution on all channels. The parameter q_2 and the calculation amount p_2 are

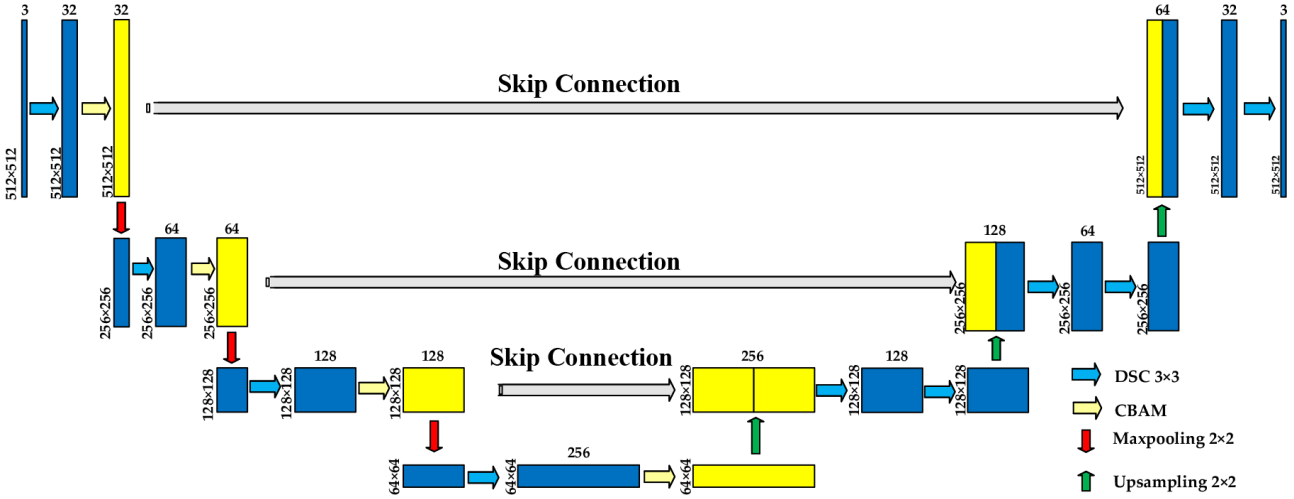

Fig. 2 FA-UNet.

https://doi.org/10.48130/mav-0026-0006

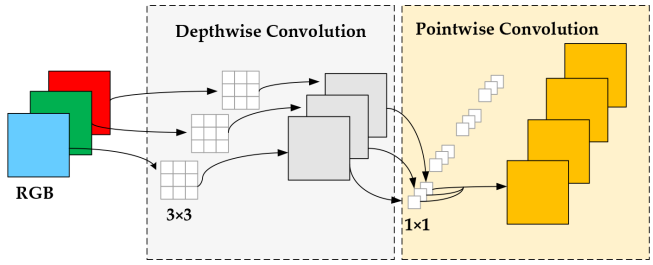

Fig. 3 Depthwise separable convolution network.

$$q_2 = D_k \times D_k \times C + 1 \times 1 \times C \times N \tag{6}$$

$$p_2 = D_k \times D_k \times C \times H \times W + 1 \times 1 \times C \times N \times H \times W \tag{7}$$

Therefore, the comparison between the two in terms of parameter quantity and calculation amount is

$$\frac{q_1}{q_2} = \frac{D_k \times D_k \times C \times N}{D_k \times D_k \times C + 1 \times 1 \times C \times N} \tag{8}$$

$$\frac{p_1}{p_2} = \frac{D_k \times D_k \times C \times N \times H \times W}{D_k \times D_k \times C \times H \times W + 1 \times 1 \times C \times N \times H \times W} \tag{9}$$

We can see that the amount of computation needed for DSC is much less than that of standard convolution. The introduction of the DSC module helps to improve the computational efficiency and achieve a lightweight model. Because the DSC uses fewer parameters for feature extraction and reduces the information overlap of convolution operations, it can effectively prevent the overfitting of the model.

Convolutional block attention module

The CBAM (Fig. 4) consists of two submodules: Channel attention and spatial attention. The design goal of the CBAM is to improve the feature expression ability of convolutional neural networks (CNNs) by explicitly modeling the channel and spatial attention. The channel attention module (as shown in Fig. 4) is used to model the dependency between channels and generate a channel attention map.

The channel attention module (CAM) is mainly divided into three steps. Step 1 is global information aggregation: Through global average pooling (GAP) and global maximum pooling (GMP) operations, the spatial dimension of the input feature is compressed to 1 to generate two channel descriptors. Step 2 is feature transformation, where the two channel descriptors are transformed through a shared Multi-Layer Perceptron (MLP) to obtain channel attention maps. Step 3 is activation, in which the values of the channel attention maps are normalized to [0, 1] through a sigmoid function. The spatial attention module (SAM, as shown in Fig. 3) is used to model the dependency between spatial positions and generate the spatial attention maps. In channel compression, we use maximum pooling and average pooling to the processed feature maps X to get single-channel image maps. In feature concatenation and convolution, we concatenate the two feature maps along channels and we use a convolutional layer to change the number of channels to 1. Finally, the weight coefficient of spatial position is obtained via a sigmoid function.

Given an input $P \in R^{C \times M \times W}$, $M_c \in R^{C \times 1 \times 1}$ (Formula 12) is obtained by the CAM. In the CAM, F^c_{avg} and F^c_{max} are obtained, respectively,

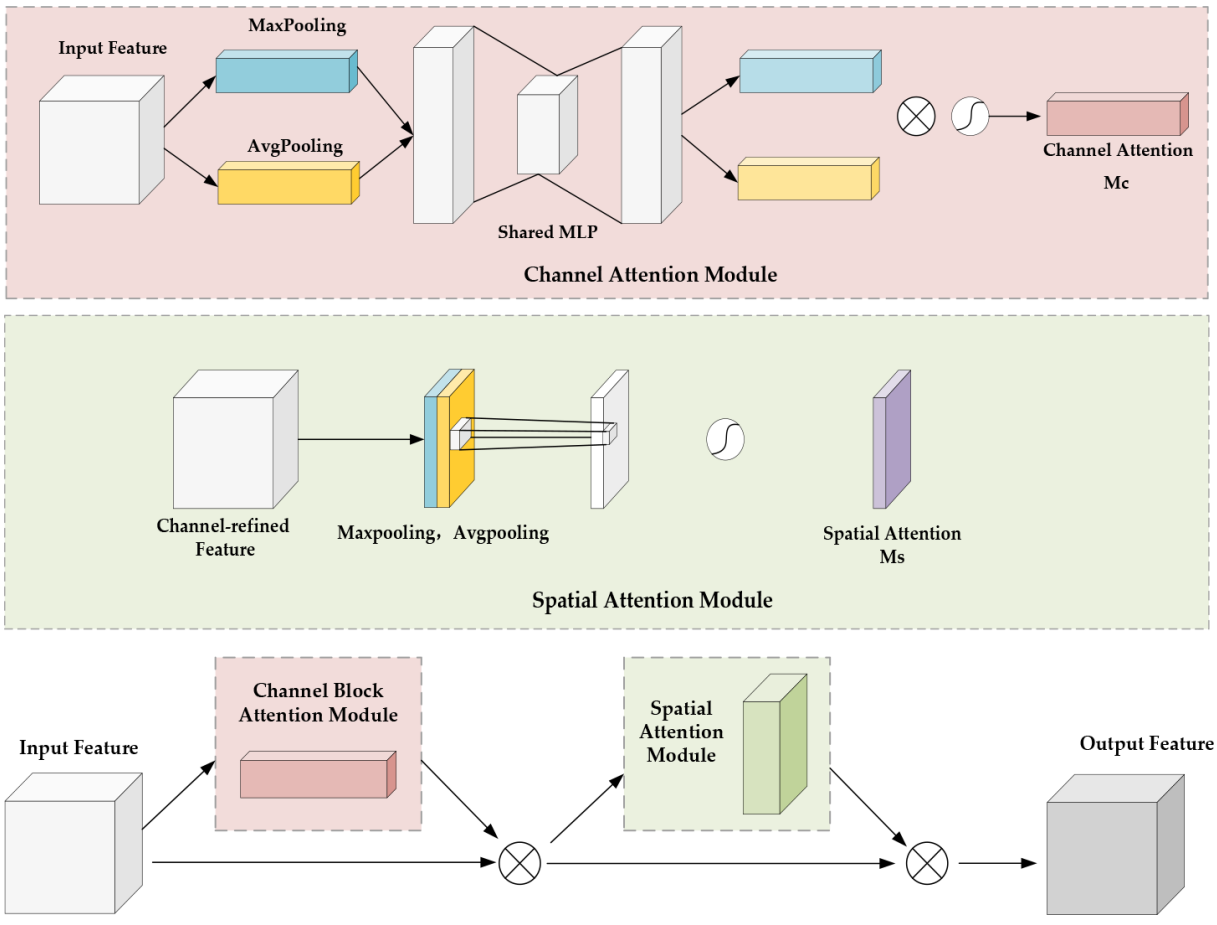

Fig. 4 Convolutional block attention module.

through average pooling and maximum pooling, as shown in Formulas 10 and 11, where σ is the sigmoid function, the the MLP structure is: Convolutional Layer, Activation Function is ReLU, Convolutional Layer, and $W_0 \in R^{C/r \times C}$ and $W_1 \in R^{C \times C/r}$ are the weights of MLP.

$$F^c_{avg} = AvgPool(F) \tag{10}$$

$$F^c_{max} = MaxPool(F) \tag{11}$$

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F))) + MLP(MaxPool(F)) \\ &= \sigma(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1(W_0(F^c_{max}))) \end{aligned} \tag{12}$$

$M_s \in R^{1 \times M \times W}$ (Formula 15) is obtained by the SAM. In the SAM, $F^s_{avg} \in R^{1 \times H \times W}$ and $F^s_{max} \in R^{1 \times M \times W}$ (Formulas 13 and 14) are obtained, respectively, via average pooling and maximum pooling, where σ is the sigmoid function, and $f^{7 \times 7}$ indicates a convolution operation with a filter of 7×7.

$$F^s_{avg} = AvgPool(F) \tag{13}$$

$$F^s_{avg} = AvgPool(F) \tag{14}$$

$$M_s(F) = \sigma\left(f^{7 \times 7}(Concat[AvgPool(F), MaxPool(F)])\right) \tag{15}$$

We multiply $M_c(F)$ obtained by the CAM by the input's original feature map F to get F', and then multiply the spatial attention weight $M_s(F')$ obtained by the SAM by F' to get F'', where $\otimes$ is element multiplication, as shown in Formulas (16) and (17).

$$F' = M_c(F) \otimes F \tag{16}$$

$$F'' = M_s(F') \otimes F' \tag{17}$$

A dynamic switching strategy based on environmental perception

We use a dynamic switching strategy based on environmental perception to dynamically control the image enhancement module. It is unnecessary and redundant to use a neural network to enhance each input frame. A potential solution is enhancement on demand, which automatically turns on the network when the system is in a low-light environment and disables low-light enhancement when the system is in a well-lit environment. We first evaluate the illumination of the input frame using an illumination metric and then design a switching strategy. We use the average grayscale value of the frame to measure the illumination of the image. The calculation formula is as follows:

$$M = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} x_{ij}}{H \times W} \tag{18}$$

where x_{ij} represents the grayscale value of the i-th row and j-th column, H is the height, W is the width, and M is the average grayscale of the image. When the average grayscale value of the image frame is less than 20, it is considered to be a dark (low-light) scene[27]. Therefore, we set the threshold to 20. If the illumination index is less than the given threshold, the input image should be enhanced by the image enhancement module and then input into the image matching algorithm for feature extraction and matching. If it is greater than the threshold, the image is directly subjected to feature extraction and matching.

Loss function

(1) Spatial consistency loss. We select this loss to keep the difference between the local regions of the input low-light and enhanced images, as defined in Formula (19).

$$L_{spa} = \frac{1}{M}\sum_{i=1}^{M}\sum\nolimits_{j \in \Omega(i)}\left(\left\|\left(Y_i - Y_j\right)\right\| - \left\|\left(I_i - I_j\right)\right\|\right)^2 M = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} x_{ij}}{H \times W} \tag{19}$$

where M is the number of local regions, and Ω represents the four regions (above, below, left, and right) of the central region i. Let Y and I represent the average intensity values of the local regions of the enhanced image and the low-light image, respectively.

(2) Illumination smoothness loss. Given the premise that illumination changes are usually smooth and slow in most practical scenes, illumination smoothness loss is used to keep the monotonic relationship in adjacent pixels in the parameter maps, as defined in Formula (20)

$$L_{ill} = \frac{1}{N}\sum_{n=1}^{N}\left(\|\nabla_x A_n\| + \left\|\nabla_y A_n\right\|\right)^2 \tag{20}$$

where ∇_x and ∇_y represent the horizontal and vertical gradients, N is the number of iterations, and A_n is the parameter maps matrix after n iterations.

(3) Exposure control loss. The illumination is adjusted by minimizing the average light intensity of the window with the preset exposure value E, as follows (Formula 21):

$$L_{exp} = \frac{1}{M}\sum_{k=1}^{M}\|Y_k - E\| \tag{21}$$

where M is the window size, E is the preset exposure value, and Y_k is the average light intensity in the window.

(4) Overall loss. The final loss function is defined as follows:

$$L = w_1 L_{spa} + w_2 L_{ill} + w_3 L_{exp} \tag{22}$$

where w_1, w_2, and w_3 are the weights of each part of the loss.

Experiment and analysis

Image enhancement performance experiment

We built the neural network based on the PyTorch framework and trained it on an NVIDIA 3050Ti GPU. We used the LOL-V1 dataset as the training set, which contains 500 pairs of low-light and normal-light images, with 485 pairs used for training and 15 pairs for testing. We set the training image size to 512×512, the batch size to 16, the optimizer to Adam, and the learning rate to 0.0001. The weights w_1, w_2, and w_3 of the loss function were set to 1, 20, and 1, respectively.

First, FA-UNet is compared with other methods in visualization quality, and some results are shown in Fig. 5. Each method has different levels of performance in the task of image enhancement. The following is a comparison from three perspectives: Image brightness enhancement, detail restoration, and color restoration.

RetinexNet performs well in color restoration and can restore the original color of the low-light image more accurately. Because of overexposure in the highlighted area of its enhancement algorithm, the image has obvious noise and artifacts, which affects the presentation of image details and makes the subsequent feature matching and positioning process difficult. Especially in complex environments, overexposure and noise problems will interfere with feature point extraction and matching stability. The HE method performs well in improving image illumination and can effectively improve image brightness, but because of its simple global HE, it can make the image overexposed and underexposed during the enhancement process. This phenomenon causes the details of some image

https://doi.org/10.48130/mav-0026-0006

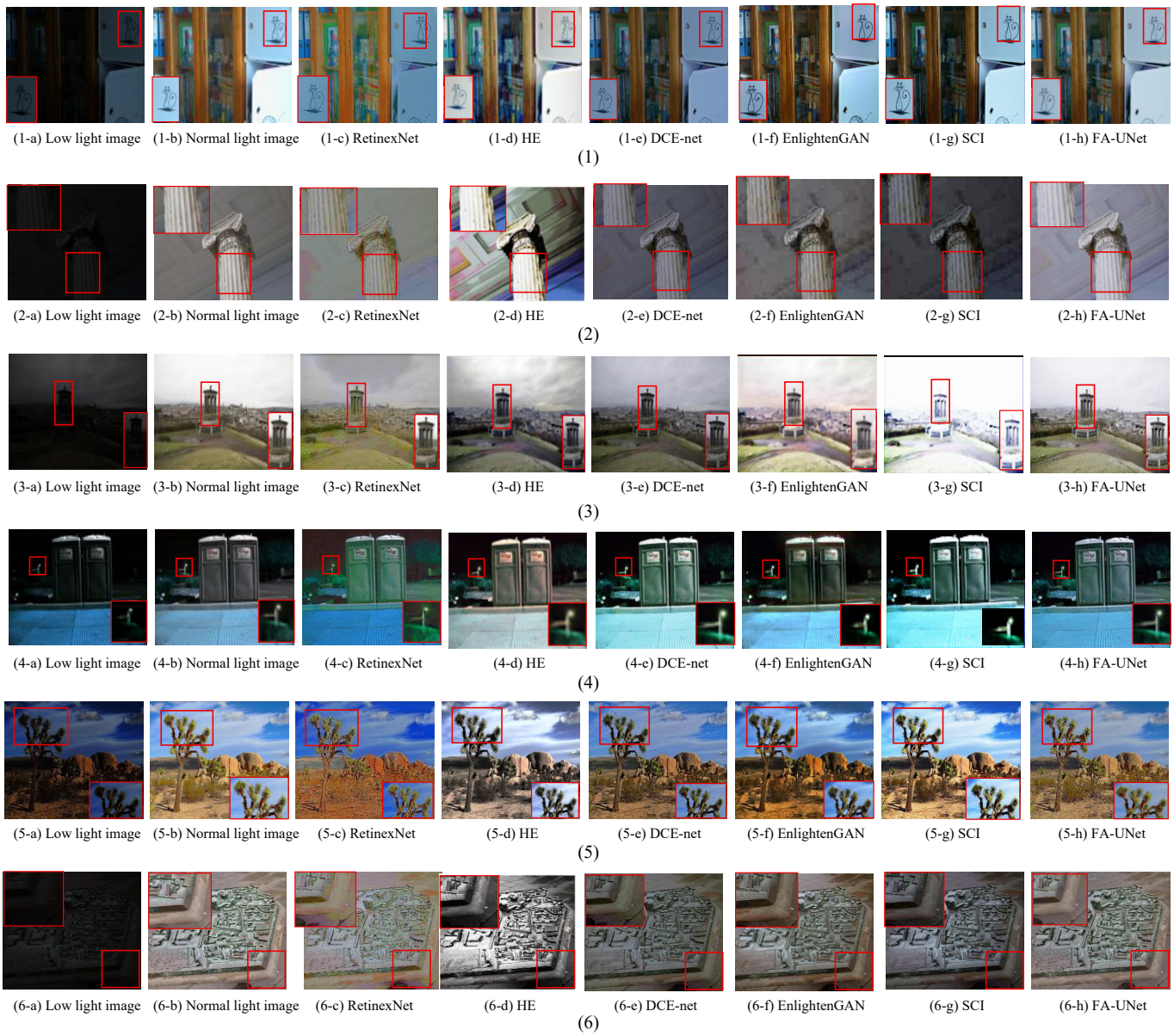

Fig. 5 Visual comparison between dark light and enhanced image.

areas to be lost, and the color reproduction is poor, affecting the overall quality of the image. DCE-net has a strong ability to restore details and can retain more image details alongside better color restoration. However, DCE-net has a general effect in terms of illumination enhancement, which does not significantly improve the brightness of the image in low light or shadowed areas, resulting in a dark image overall, affecting the effect of feature point extraction, especially in low-light environments, which may lead to reduced positioning accuracy. EnlightenGAN[19] performs well in terms of color saturation, closely resembling the images under normal light, but it also exhibits overexposure. SCI[28], on the other hand, fails to adequately restore brightness, resulting in significantly darker images after enhancement. In contrast, FA-UNet demonstrates its comprehensive advantages in many aspects. Through the UNet decoding and encoding structure and the jump connection mechanism, FA-UNet can effectively aggregate the high-level and low-level features of the image, estimate the pixel-by-pixel image enhancement curve, and achieve local enhancement of the brightness of each pixel in the image. This enhancement method not only effectively improves the image's illumination and avoids overexposure, thereby reducing the generation of noise and artifacts, but also retains more image details, making the color restoration accurate and the details clear. When enhancing the image, FA-UNet can balance brightness and detail restoration, and reduce the color distortion and noise problems commonly seen in other methods, thus demonstrating its superior performance in image enhancement tasks.

We collected real-world images through drone flight tests and used these to compare and analyze RetinexNet, HE, DCE-net, EnlightenGAN, SCI, and FA-UNet. The flight time was approximately 353 s, the flight altitude was 250 m, and the image sampling frequency was 5 fps. The results are shown in Fig. 6. RetinexNet performed well in color reproduction but exhibited noticeable noise and artifacts. The HE algorithm performed well in improving image illumination but had poor color reproduction, resulting in poor image quality overall. FA-UNet outperformed RetinexNet, HE, DCE-net, EnlightenGAN, and SCI in color reproduction and restoring the image's texture.

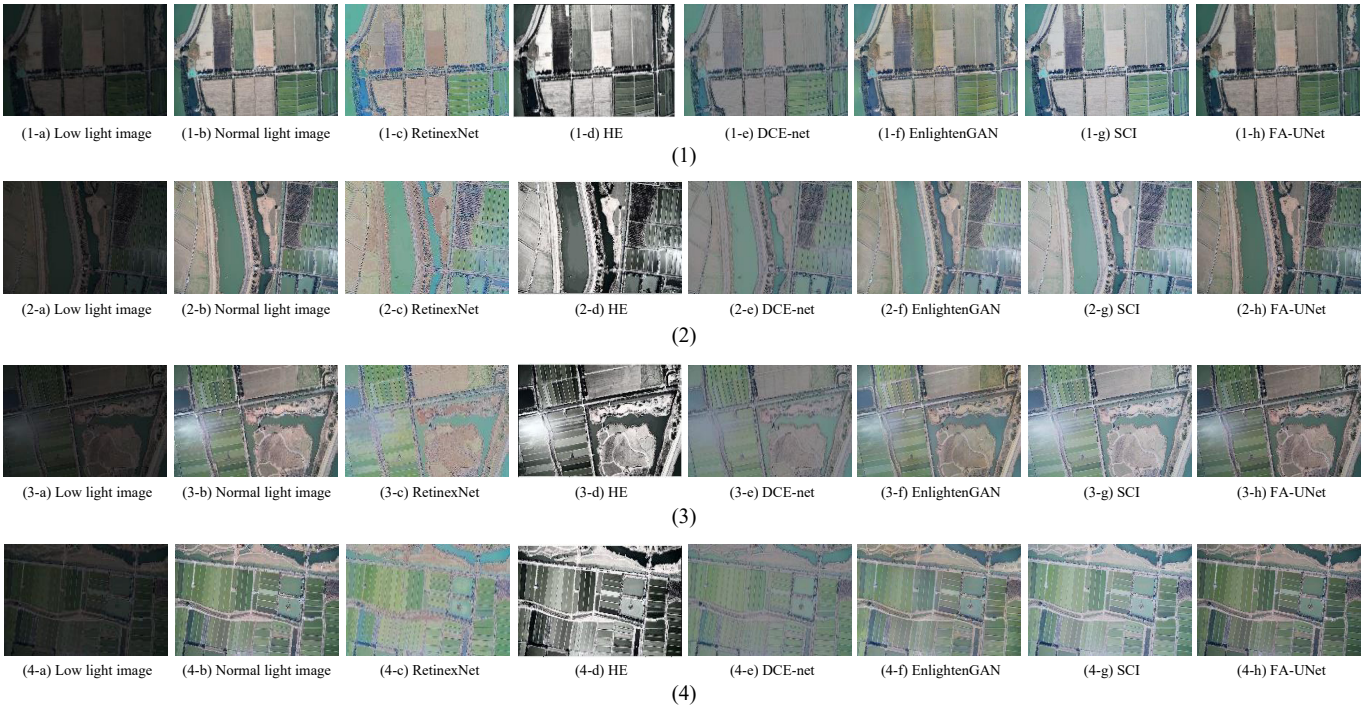

Fig. 6 Visual comparison on real-world image.

Table 1. Image quality evaluation table.

	Methods	PSNR	SSIM	NIQE	LPIPS
LOL-v1	RetinexNet	16.77	0.43	8.734	0.381
	DCE-net	15.14	0.70	7.755	0.330
	EnlightenGAN	**18.63**	**0.812**	6.869	0.412
	SCI	14.77	0.679	6.646	0.335
	FA-UNet	17.22	0.71	**5.879**	**0.318**
SICE	RetinexNet	15.84	0.73	4.368	0.407
	DCE-net	16.12	0.87	3.987	0.362
	EnlightenGAN	13.94	0.605	**2.683**	**0.253**
	SCI	13.15	0.523	3.360	0.317
	FA-UNet	**19.88**	**0.89**	3.175	0.308

We use PSNR, SSIM, the natural image quality evaluator, and learned perceptual image patch similarity (LPIPS) as quantitative indicators to objectively evaluate the proposed methods. The comparison algorithms selected were RetinexNet, DCE-net, EnlightenGAN, and SCI. The statistical results of the image quality evaluation indicators are shown in Table 1. On the LOL-V1 dataset, the LPIPS value of FA-UNet is 0.318, which is 3.6% lower than that of DCE-Net (0.33), indicating that the perceptual difference between the enhanced image and the real image of FA-UNet is the smallest. The NIQE value of FA-UNet is 5.879, which is 11.5% lower than that of SCI (6.646), indicating that the enhanced image is close to the natural distribution of the real image. FA-UNet's PSNR is 17.22, 7.6% lower than that of EnlightenGAN but 16.6% higher than that of SCI, indicating rich enhanced image details and low noise. FA-UNet's SSIM is 0.71, 12.6% lower than that of EnlightenGAN but 4.6% higher than that of SCI, indicating that the enhanced image has a similar structure to the normal-light image and a good low-light enhancement effect. On the SICE dataset, FA-UNet's PSNR is 19.88, a 23.3% improvement over DCE-Net and a 25.5% improvement over RetinexNet. FA-UNet's SSIM is 0.89, a 2.3% improvement over DCE-Net and a 21.9% improvement over RetinexNet. In terms of the NIQE metrics, EnlightenGAN performs best with 2.683, followed by FA-UNet with 3.175. On the LPIPS metric, EnlightenGAN performed best with 0.253, followed by FA-UNet with 0.308. Considering the PSNR, SSIM, NIQE, and LPIPS metrics, FA-UNet showed better overall performance, preserving the images' details and texture features well, without artifacts or distortion.

In summary, FA-UNet has obvious advantages over RetinexNet, HE, and DCE-net in improving image quality, preserving details, and avoiding overexposure. Its innovative image enhancement mechanism enables images to maintain a high degree of detail restoration under complex lighting conditions, providing more stable image input for subsequent visual tasks (such as feature extraction, matching, and positioning).

Ablation study

To analyze the contribute of various components of the loss function on the network's performance, this section analyzes the enhancement results obtained using different combinations of loss functions. Table 2 shows the training results using various loss combinations. When l_{ill} is not used, PSNR decreases by 8.3% compared with L_{Total}, and SSIM decreases by 16.9%. When l_{exp} is not used, PSNR decreases by 5.28% compared with L_{Total}, and SSIM decreases by 8.45%. When l_{spa} is not used, PSNR decreases by 1.05% compared with L_{Total}, and SSIM decreases by 4.23%. A visual comparison of the results using different loss functions is shown in Fig. 7. The result without l_{spa} has relatively low contrast. Removing l_{ill} hinders correlation between adjacent regions, resulting in

Table 2. PSNR/SSIM of different combinations of loss functions

Loss function combination	PSNR	SSIM
w/oL_{spa}	17.04	0.68
$w/o\ L_{ill}$	15.79	0.59
$w/o\ L_{exp}$	16.31	0.65
L_{Total}	17.22	0.71

https://doi.org/10.48130/mav-0026-0006

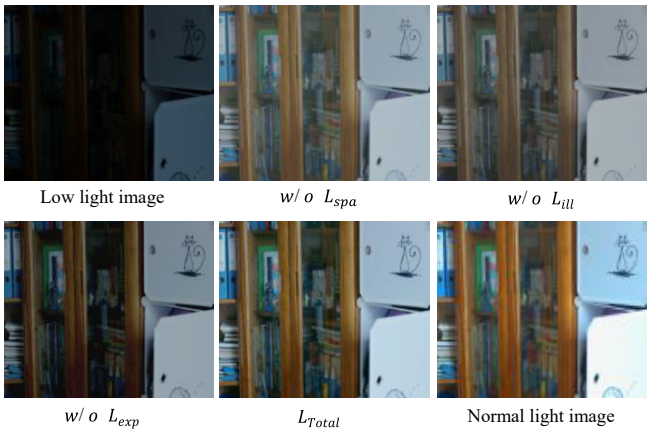

Fig. 7 Visual comparison of different loss functions

noticeable artifacts. When is removed, restoration of the low-light region is poor. The experimental results show that using L_{Total} has a better performance than other choices.

Visual positioning experiment

We used the Euroc and KITTI datasets to verify the effectiveness of LLIE algorithms for visual localization. We synthesized image sequences from the Euroc and KITTI datasets using backlighting to generate low-light images. The methods used for the comparison were RetinexNet, HE, and DCE-Net. First, we performed LLIE on the image sequences. Then we used DyPL-VO tracking to solve the low-light image sequences and the four enhanced image sequences. We calculated the root mean square error of the absolute trajectory errors, as shown in Table 3. Some enhanced images from the four algorithms are shown in Fig. 8. The HE method, though enhancing image brightness, suffers from overexposure. RetinexNet can also enhance image brightness, but textural information is severely lost. DCE-Net has limited ability to restore brightness in low-light environments, and the enhanced images are still not bright enough, which is not conducive to subsequent image matching and localization. FA-UNet avoids overexposure, achieving balanced image brightness while retaining more detail. Second, based on the multi-scale feature aggregation mechanism of UNet, FA-UNet enhances illumination while maintaining the integrity of the local textures. The experimental results show that FA-UNet outperformed existing methods in detail restoration and illumination balance for these images, and provided more stable feature information for subsequent visual scene matching.

Table 3. Comparison results of the root mean square error of the absolute trajectory error (m).

Image sequence	Length	Low-light	RetinexNet	HE	DCE-net	FA-UNet
MH_04_difficult	91.747	Fail	0.216	0.114	0.125	**0.070**
MH_05_difficult	97.593	Fail	0.198	0.829	0.100	**0.042**
V1_03_difficult	78.982	Fail	0.098	0.100	0.122	**0.087**
Syn_KITTI_Seq00	3,724.187	16.012	4.403	5.135	4.809	**3.009**
Syn_KITTI_Seq02	5,067.223	Fail	16.109	8.885	9.136	**6.891**
Syn_KITTI_Seq03	560.888	0.668	0.462	0.272	0.357	**0.195**
Syn_KITTI_Seq09	1,705.051	Fail	5.207	4.902	3.667	**3.017**

Table 3 compares the root mean square error results of the absolute trajectory error. DyPL-VO tracked successfully only with two groups of low-light sequences, but achieved successful trajectory tracking in all enhanced image sequences, indicating that image enhancement effectively improved the visual positioning performance under low-light conditions. The average root mean square errors of the four methods were 3.813, 2.891, 2.588, and 1.94 m. Among them, the accuracy of FA-UNet was improved by 49.12%, 32.90%, and 25.04% compared with the other methods. FA-UNet can effectively improve the image illumination, retain the image details, and extract more feature points for matching and positioning. Therefore, FA-UNet significantly improves the performance of visual positioning in low-light environments, improves the quality of image enhancement, and optimizes the feature extraction and matching of subsequent positioning algorithms, thereby improving positioning accuracy.

Computational efficiency evaluation

This experiment compared the performance of FA-UNet based on two metrics: Floating-point operations per second (FLOPs) and test time. According to the results in Table 4, on the Euroc dataset, FA-UNet's FLOPs were 52.17% lower than those of DCE-Net, and its test time was 58.35% lower. On the KITTI dataset, FA-UNet's FLOPs were 42.37% lower than those of DCE-Net, and its test time was 55% lower. This demonstrates that FA-UNet outperforms RetinexNet and DCE-Net in computational efficiency, making it suitable for use on mobile platforms with limited computing power.

Discussion

In this paper, we propose a fast and accurate light enhancement curve estimation network, FA-UNet, based on Zero-DCE theory. First, we use DSC instead of traditional convolution to achieve a lightweight network. In addition, we add a CBAM to help the network identify important features across channels and spatial

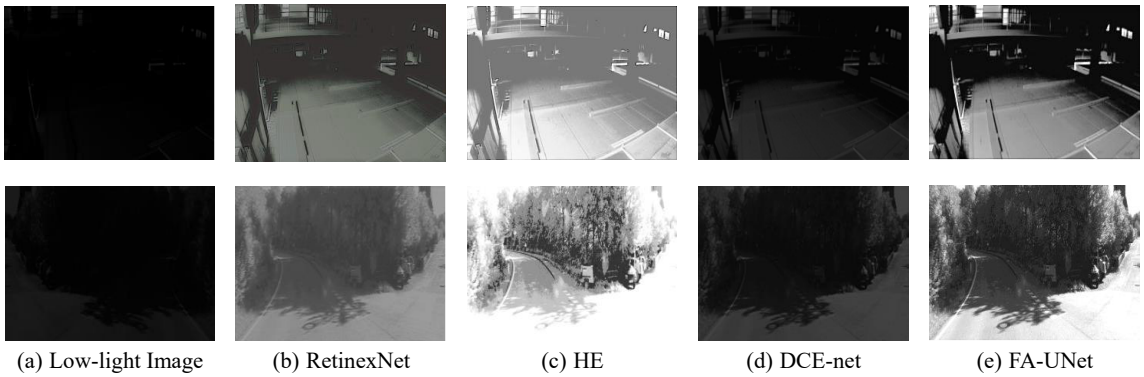

Fig. 8 Visual effects of LLIE methods.

Table 4. Model efficiency evaluation indices

Methods	RetinexNet		DCE-net		FA-UNet	
	KITTI	Euroc	KITTI	Euroc	KITTI	Euroc
FLOPs (10^9)	13.721	10.623	0.059	0.046	**0.034**	**0.022**
Test time (ms)	72.42	56.05	35.24	30.54	**15.86**	**12.72**

regions in the image. In addition, we design a dynamic switching mechanism based on environmental perception to adaptively select enhancement operations according to the illumination of the image, effectively reducing computational redundancy. By designing image enhancement performance experiments, visual positioning experiments, and computational efficiency evaluation experiments, it can be proved that the proposed UAV LLIE method based on FA-UNet not only effectively improves the images' illumination and avoids overexposure, but also reduces the generation of noise and artifacts, retains more image details, and makes the color restoration accurate and the details clear.

Conclusions

In terms of visual positioning, our method maintains the integrity of local texture while enhancing the illumination and improves the accuracy of feature point matching, and thus improves the positioning accuracy and robustness in the visual odometry task. In terms of computational efficiency, our method can achieve a good balance between lightness and high efficiency, and is suitable for scenarios with limited resources or high real-time requirements. However, in practical applications on edge devices/drones, the proposed method struggles to guarantee the sustainability of enhanced accuracy because of the high maneuverability of the vehicle and the complexity of the operating environment. Furthermore, there is still room for improvement in its real-time performance.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Cheng Y, Liu X; data collection: Cheng Y, Xu C; analysis and interpretation of results: Cheng Y, Chen S; draft manuscript preparation: Cheng Y. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The datasets generated during and/or analyzed during the current study are not publicly available because of classified information but are available from the corresponding author on reasonable request.

Acknowledgments

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 7 January 2026; Revised 3 March 2026; Accepted 19 April 2026; Published online 6 August 2026

References

[1] Tong P, Yang X, Yang Y, Liu W, Wu P. 2023. Multi-UAV collaborative absolute vision positioning and navigation: a survey and discussion. *Drones* 7(4):261

[2] Li Z, Li S, Anderson J, Shan J. 2024. Urban visual localization of block-wise monocular images with google street views. *Remote Sensing* 16(5):801

[3] Jiang S, Luo H, Liu Y. 2024. Suitable-matching areas' selection method based on multi-level saliency. *Remote Sensing* 16(1):161

[4] He J, Wu Q. 2025. A localization method for UAV aerial images based on semantic topological feature matching. *Remote Sensing* 17(10):1671

[5] Liu J, Xu D, Yang W, Fan M, Huang H. 2021. Benchmark lowlight image enhancement and beyond. *International Journal of Computer Vision* 129(4):1153–1184

[6] Ding B, Sun B, Sun X. 2025. Low-light remote sensing image enhancement *via* priors guided end-to-end latent residual diffusion. *Remote Sensing* 17(18):3193

[7] Pizer SM, Johnston RE, Ericksen JP, Yankaskas BC, Muller KE. 1990. Contrast-limited adaptive histogram equalization: speed and effectiveness. *Proceedings of the First Conference on Visualization in Biomedical Computing, May 22–25, 1990, Atlanta, GA, USA*. USA: IEEE. pp. 337–338 doi: 10.1109/VBC.1990.109340

[8] Land EH. 1977. The retinex theory of color vision. *Scientific American* 237(6):108–129

[9] Cheng HD, Shi XJ. 2004. A simple and effective histogram equalization approach to image enhancement. *Digital Signal Processing* 14(2):158–170

[10] Liu YF, Guo JM, Yu JC. 2017. Contrast enhancement using stratified parametric-oriented histogram equalization. *IEEE Transactions on Circuits and Systems for Video Technology* 27(6):1171–1181

[11] Hao S, Han X, Guo Y, Xu X, Wang M. 2020. Low-light image enhancement with semi-decoupled decomposition. *IEEE Transactions on Multimedia* 22(12):3025–3038

[12] Ren X, Yang W, Cheng WH, Liu J. 2020. LR3M: robust low-light enhancement *via* low-rank regularized retinex model. *IEEE Transactions on Image Processing* 29:5862–5876

[13] Wu J, Ai H, Zhou P, Wang H, Zhang H, et al. 2025. Low-light image dehazing and enhancement *via* multi-feature domain fusion. *Remote Sensing* 17(17):2944

[14] Sun Z, Shen Z, Chen N, Pang S, Liu H, et al. 2025. MEFormer: enhancing low-light images while preserving image authenticity in mining environments. *Remote Sensing* 17(7):1165

[15] Zhang Y, Guo X, Ma J, Liu W, Zhang J. 2021. Beyond brightening low-light images. *International Journal of Computer Vision* 129(4):1013–1037

[16] Zhou S, Li C, Change Loy C. 2022. LEDNet: joint low-light enhancement and deblurring in the dark. In *Computer Vision – ECCV 2022*. Cham, Switzerland: Springer Nature. pp, 573–589 doi: 10.1007/978-3-031-20068-7_33

[17] Ren W, Liu S, Ma L, Xu Q, Xu X, et al. 2019. Low-light image enhancement *via* a deep hybrid network. *IEEE Transactions on Image Processing* 28(9):4364–4375

[18] Yao Z, Fan G, Fan J, Gan M, Philip Chen CL. 2024. Spatial–frequency dual-domain feature fusion network for low-light remote sensing image enhancement. *IEEE Transactions on Geoscience and Remote Sensing* 62:1–16

[19] Jiang Y, Gong X, Liu D, Cheng Y, Fang C, et al. 2021. EnlightenGAN: deep light enhancement without paired supervision. *IEEE Transactions on Image Processing* 30:2340–2349

[20] Fu Z, Yang Y, Tu X, Huang Y, Ding X, et al. 2023. Learning a simple low-light image enhancer from paired low-light instances. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17–24, 2023, Vancouver, BC, Canada*. USA: IEEE. pp. 22252–22261 doi: 10.1109/cvpr52729.2023.02131

[21] Ni Z, Yang W, Wang S, Ma L, Kwong S. 2020. Towards unsupervised deep image enhancement with generative adversarial network. *IEEE Transactions on Image Processing* 29:9140–9151

https://doi.org/10.48130/mav-0026-0006

[22] Yang S, Ding M, Wu Y, Li Z, Zhang J. 2023. Implicit neural representation for cooperative low-light image enhancement. *2023 IEEE/CVF International Conference on Computer Vision (ICCV). October 1–6, 2023. Paris, France*. USA: IEEE. pp. 12872–12881 doi: 10.1109/iccv51070.2023.01187

[23] Liu M, Cui Y, Ren W, Zhou J, Knoll AC. 2025. LIEDNet: a lightweight network for low-light enhancement and deblurring. *IEEE Transactions on Circuits and Systems for Video Technology* 35(7):6602–6615

[24] Zhang L, Zhang L, Liu X, Shen Y, Zhang S, et al. 2019. Zero-shot restoration of back-lit images using deep internal learning. *Proceedings of the 27th ACM International Conference on Multimedia. October 21–25, 2019, Nice, France*. New York, USA: ACM. pp. 1623–1631 doi: 10.1145/3343031.3351069

[25] Zhu A, Zhang L, Shen Y, Ma Y, Zhao S, et al. 2020. Zero-shot restoration of underexposed images *via* robust retinex decomposition. *2020 IEEE International Conference on Multimedia and Expo (ICME). July 6–10, 2020, London, UK*. USA: IEEE. pp. 1–6 doi: 10.1109/icme46284.2020.9102962

[26] Guo C, Li C, Guo J, Loy CC, Hou J, et al. 2020. Zero-reference deep curve estimation for low-light image enhancement. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13–19, 2020. Seattle, WA, USA*. USA: IEEE. pp. 1777–1786 doi: 10.1109/cvpr42600.2020.00185

[27] Wang J, Wang R, Wu A. 2019. Improved gamma correction for visual SLAM in low-light scenes. *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). 11–13 October 2019, Chongqing, China*. USA: IEEE. doi: 10.1109/IMCEC46724.2019.8983904

[28] Ma L, Ma T, Liu R, Fan X, Luo Z. 2022. Toward fast, flexible, and robust low-light image enhancement. *arXiv* Preprint