Citation information: DOI 10.1109/ICME57554.2024.10688351, 2024 IEEE International Conference on Multimedia and Expo (ICME)

HSVFormer: Robust and Unsupervised HSV-based Transformer Framework for Low-Light Image Enhancement

1st Xiaogang Du
Shaanxi Joint Laboratory of Artificial Intelligence and School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology
Xi'an, China
duxiaogang@sust.edu.cn

2nd Meng Yang
Shaanxi Joint Laboratory of Artificial Intelligence and School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology
Xi'an, China
211612079@sust.edu.cn

3rd Tao Lei ✉
Shaanxi Joint Laboratory of Artificial Intelligence and School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology
Xi'an, China
leitao@sust.edu.cn

4th Xuejun Zhang
School of Electronic and Information Engineering, Lanzhou Jiaotong University
Lanzhou, China
xuejunzhang@mail.lzjtu.cn

5th Yingbo Wang
Shaanxi Joint Laboratory of Artificial Intelligence and School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology
Xi'an, China
wangyingbo@sust.edu.cn

6th Asoke K. Nandi
Department of Electronic and Computer Engineering, Brunel University London
London, United Kingdom
asoke.nandi@brunel.ac.uk

***Abstract*—The following three factors restrict the application of existing low-light image enhancement methods: corruptions induced by the light-up process, color distortion, and a restricted generalization capacity due to limited paired training data. To address these limitations, we first combine HSV theory and Transformer, proposing a robust unsupervised low-light image enhancement framework, named HSVFormer. Secondly, we introduce brightness disturbance and design an unsupervised value enhancement network, which estimates brightness information and restores degraded brightness information to obtain enhanced reflectance. Finally, we utilize the V-subspace and devise a value-guided multi-head channel self-attention to capture brightness representations of regions with different brightness conditions and guide the modeling of non-local interactions. Experiment results on publicly available datasets demonstrate that HSVFormer can achieve superior performance compared with state-of-the-art approaches. The code is available at https://github.com/m0fig/HSVFormer.**

***Index Terms*—Unsupervised learning, Low-light image enhancement, Transformer, Retinex**

I. Introduction

Low-light images are captured in environments with low-light or poor lighting conditions, and these images tend to have quality issues such as low contrast, color distortion, and missing details. Low-light image enhancement is the process of transforming a low-light image into a normal-light image with complete structure and detail, and a natural visual effect.

In recent years, with the continuous progress in the field of computer vision, scholars have proposed many low-light image enhancement methods. Compared with traditional methods, deep learning-based methods have received extensive attention due to good accuracy and efficiency. Depending on whether labeling is required or not, these methods are mainly divided into two categories: supervised learning-based methods and unsupervised learning-based methods.

Supervised learning-based methods are used for image enhancement by training from a large number of samples (pairs of low-light and normal-light images) to obtain prior knowledge. For example, Wei et al. [1] proposed RetinexNet, which includes a decomposition module and a light adjustment module and utilizes BM3D denoising for the reflectance. However, the images enhanced by RetinexNet exhibit global color bias. To address this issue, Zhang et al. proposed KinD [2] and KinD++ [3], respectively. KinD [2] has a similar structure to RetinexNet and can effectively correct color bias. While KinD++ [3] designs a multiscale illumination attention module to improve the visual defects such as overexposure and halo that appear in the enhancement results of KinD. However, the above methods rely on handcraft priors and are time-consuming for optimization. To address these issues, Fu et al. [4] proposed regularizer-free Retinex network (RFR),

✉ Corresponding author: Tao Lei.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ICME57554.2024.10688351, 2024 IEEE International Conference on Multimedia and Expo (ICME)

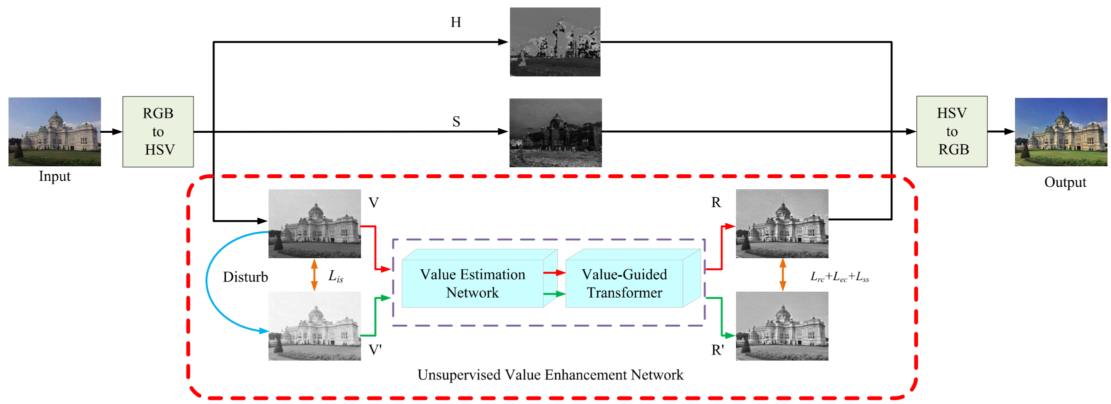

Fig. 1. The overall framework of HSVFormer.

which implicitly extracts reflectance and illuminance features and employs contrastive learning and a progressive learning strategy for self-knowledge distillation to impose constraints on Retinex decomposition. To summarized, the above methods require a large amount of paired data for training, However, real paired datasets are difficult to obtain, and collecting paired low-light images is tedious and labor-intensive. In addition, if synthetic datasets are used for training, this can result in limited model generalization capacity.

Unsupervised learning-based methods have quickly become a current research hotspot because they do not require paired labeled data and exhibit good generalization capacity. For example, Jiang et al. [5] proposed EnlightenGAN, which improves the overall and local brightness of an image through global and local discriminators. However, EnlightenGAN has strict requirements for training data and the generalization capacity is still not strong enough. To further enhance the generalization capacity, Guo et al. proposed Zero DCE [6] and Zero DCE++ [7], respectively. Zero DCE [6] is a deep curve estimation method that uses non-reference loss functions for regularization, and its accelerated and lightweight version, called Zero DCE++ [7]. However, these two methods suffer from color bias. In response to challenges related to color bias, Yang et al. [8] proposed NeRCo, which introduces semantic-oriented supervision with priors from the pretrained vision-language model to ensure the realism of the generated images at the color and detail level.

In addition to the issues mentioned above, deep learning-based low-light image enhancement methods mostly rely on convolutional neural networks, which have limitations in capturing long distance dependencies and modeling non-local self-similarity. To overcome these deficiencies, scholars have turned to Transformer-based methods for low-light image enhancement, such as Retinexformer [9], IAT [10], SNR-Net [11] and LLFormer [12] etc. However, images enhanced by these methods suffer from uneven brightness enhancement, checkerboard effect and color bias. Besides, the above methods rely typically on paired images for training.

To cope with the above problems, we propose a robust unsupervised low-light image enhancement framework, namely HSVFormer. In this work, our contributions can be summarized as follows:

- To our best knowledge, we first combine HSV theory with Transformer to propose a robust unsupervised low-light image enhancement framework. This framework is superior to popular methods since it can effectively removes hidden noise in dark areas while maintaining color information and enhancing the brightness of dark areas in an image.
- We introduce brightness disturbance into V-subspace and design an Unsupervised Value Enhancement Network (UVENet) to simulate the corruptions of the V-subspace, avoiding artifacts caused by a single target brightness value, enhancing the interpretability of the model and making it more suitable for real low-light scenes.
- We combine the V-subspace with the multi-head channel self-attention and design a Value-Guided Multi-head Channel Self-Attention (VG-MCSA). VG-MCSA uses brightness encoding information from different regions in the V-subspace to guide long-distance dependency modeling, which is helpful for the noise removal and the restoration of image details and brightness in dark areas.

II. Method

A. The Overall Framework of HSVFormer

The overall framework of HSVFormer is shown in Fig. 1. Firstly, HSVFormer converts the input image from RGB space to HSV space, preserving the color space (Hue, Saturation) and enhancing the V-subspace (Value), so as to maintain color information and avoid color bias in the enhanced image. Secondly, for the V-subspace, we design the UVENet, which is consisted of two core sub-networks: Value Estimation Network (VEN) and Value-Guided Transformer (VGT). VEN simulates corruptions in real low-light images and estimates brightness information, while VGT enhances the estimated brightness information, removes noise and restores brightness information in dark areas. Thirdly, we design the VG-MCSA to further denoise and recover image details in dark areas. Additionally,

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI 10.1109/ICME57554.2024.10688351, 2024 IEEE International Conference on Multimedia and Expo (ICME)

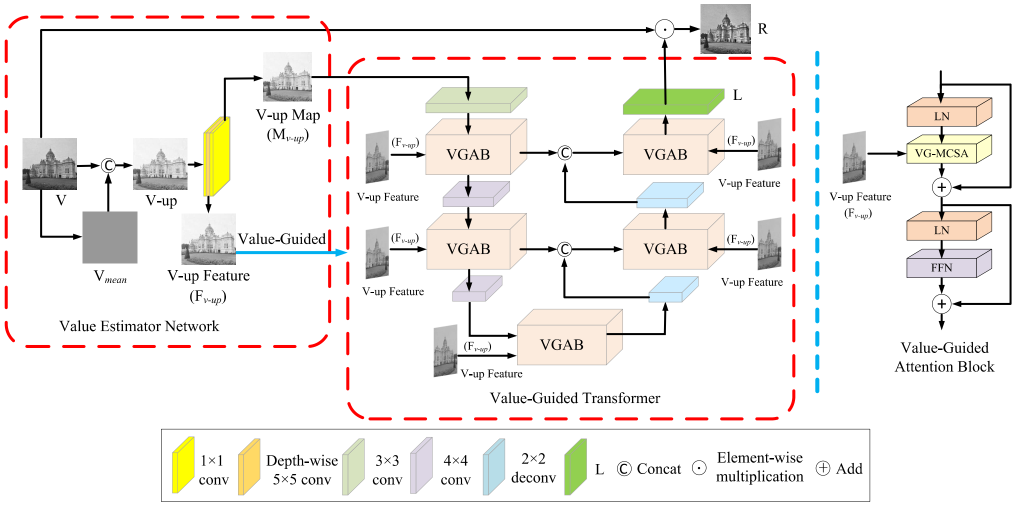

Fig. 2. Unsupervised value enhancement network.

in the absence of paired images, we train HSVFormer in an unsupervised manner. We introduce a random disturbed form of the brightness (V'), and loss functions are consisted of illumination smoothness loss (L_{is}), reflectance consistency loss (L_{rc}), exposure control loss (L_{ec}), and spatial structure loss (L_{ss}) [13]. Finally, the H and S subspaces are recombined with the enhanced V-subspace to generate the enhanced image.

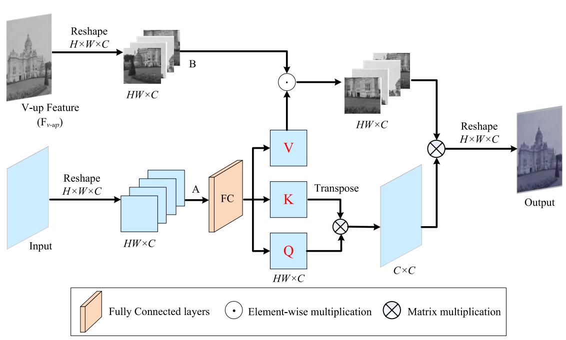

Fig. 3. Value-guided multi-head channel self-attention.

B. UVENet

Different from those low-light image enhancement methods based on deep learning and Retinex [1] [2], HSVFormer transforms the decomposition process into a generative process, and treats reflectance as enhanced brightness. Therefore, HSVFormer can adapt to various lighting conditions and has strong robustness.

Inspired by Jiang et al. [13], we combine the V-subspace with Retinex theory [14] to avoid color bias. The V-subspace can be denoted as:

$$\mathrm{V} = \mathrm{R} \odot \mathrm{I}, \tag{1}$$

where $\mathrm{R} \in \mathbb{R}^{H \times W \times 1}$ and $\mathrm{I} \in \mathbb{R}^{H \times W \times 1}$ are the reflectance and illuminance, respectively, and $\odot$ denotes element-wise multiplication. With $\mathrm{V} \in \mathbb{R}^{H \times W \times 1}$ as the input and R as the output, R can be rewritten as [13]:

$$\mathrm{R} = \mathrm{V} \odot \mathrm{L}, \tag{2}$$

where $\mathrm{L} \in \mathbb{R}^{H \times W \times 1}$ is the inverse of I. However, this variant model is designed under ideal state and is inconsistent with real low-light image scenes. To better emulate low-light scenes in the real world and avoid artifacts caused by a single target brightness value, disturbance terms for V and L are introduced to simulate corruptions in real low-light images. Therefore, R can be denoted as:

$$\begin{aligned} \mathrm{R} &= (\mathrm{V} + \tilde{\mathrm{V}}) \odot (\mathrm{L} + \tilde{\mathrm{L}}) \\ &= \mathrm{V} \odot \mathrm{L} + \mathrm{V} \odot \tilde{\mathrm{L}} + \tilde{\mathrm{V}} \odot (\mathrm{L} + \tilde{\mathrm{L}}), \end{aligned} \tag{3}$$

where $\tilde{\mathrm{V}} \in \mathbb{R}^{H \times W \times 1}$ and $\tilde{\mathrm{L}} \in \mathbb{R}^{H \times W \times 1}$ represent disturbance terms, $\mathrm{V} \odot \tilde{\mathrm{L}}$ represents underexposure or overexposure, and $\tilde{\mathrm{V}} \odot (\mathrm{L} + \tilde{\mathrm{L}})$ represents the noise and artifacts adhering to the dark corners of the V-subspace. Subsequently, we design the UVENet, which can be represented as:

$$(\mathrm{M}_{v-up}, \mathrm{F}_{v-up}) = \mathcal{D}(\mathrm{V}, \mathrm{V}_{mean}), \tag{4}$$

$$\mathrm{R} = \mathrm{V} \odot \mathcal{G}(\mathrm{M}_{v-up}, \mathrm{F}_{v-up}). \tag{5}$$

Where $\mathcal{D}$ and $\mathcal{G}$ denote VEN and VGT, respectively.

UVENet is illustrated in Fig. 2. Firstly, the disturbance map $\mathrm{V}-\mathrm{up} \in \mathbb{R}^{H \times W \times 2}$ is initially constructed by concatenating $\mathrm{V} \in \mathbb{R}^{H \times W \times 1}$ and the average pixel value of the V-subspace ($\mathrm{V}_{mean} \in \mathbb{R}^{H \times W \times 1}$), which is aimed at simulating the corruptions of the V-subspace and avoiding artifacts caused by a single target brightness value. Secondly, the brightness encoding information of different regions in the brightness feature map $\mathrm{F}_{v-up} \in \mathbb{R}^{H \times W \times C}$ is estimated by a 5×5 depth-wise separable convolution. Then, the brightness map $\mathrm{M}_{v-up} \in \mathbb{R}^{H \times W \times 3}$ is obtained through a 1×1 convolution. Finally, the brightness map $\mathrm{M}_{v-up} \in \mathbb{R}^{H \times W \times 3}$ and brightness feature map $\mathrm{F}_{v-up} \in \mathbb{R}^{H \times W \times C}$ are passed into VGT.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI 10.1109/ICME57554.2024.10688351, 2024 IEEE International Conference on Multimedia and Expo (ICME)

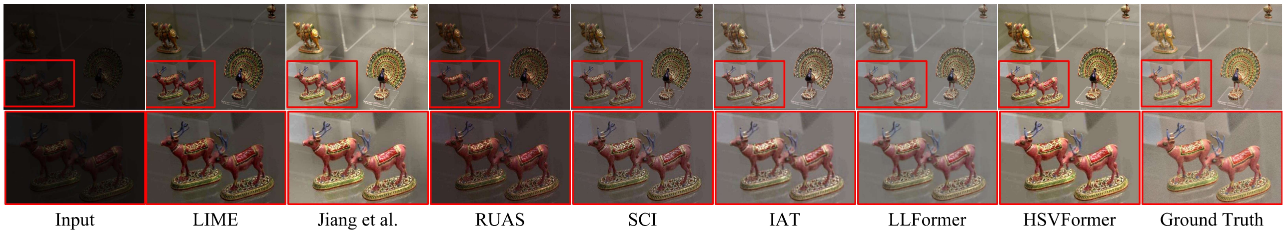

Fig. 4. Visualization results generated by different methods on the SICE dataset.

VGT is a three-scale encoder-decoder Transformer structure. The Value-Guided Attention Block (VGAB) is an important unit of VGT. Firstly, the brightness map $\mathrm{M}_{v-up} \in \mathbb{R}^{H\times W\times 3}$ is used as input, and then undergoes two downsampling operations for scale reduction, followed by upsampling, which is symmetrically structured with the downsampling operations. Additionally, the estimated brightness feature map $\mathrm{F}_{v-up} \in \mathbb{R}^{H\times W\times C}$ is input into each VGAB, contributing to denoising in dark areas, as well as the restoration of detailed information and brightness. Finally, the inverse L of the illuminance is output, and it is multiplied pixel-wise with the V-subspace in a stepwise manner to obtain the final enhancement result R.

C. VG-MCSA

The structure of the VG-MCSA is illustrated in Fig. 3. Firstly, the input feature map $\mathrm{Input} \in \mathbb{R}^{H\times W\times C}$ is reshaped into tokens $\mathrm{A} \in \mathbb{R}^{HW\times C}$, and then A is divided into k heads along the feature channel dimension ($\mathrm{A} = [\mathrm{A}_1, \mathrm{A}_2, ..., \mathrm{A}_k]$, where $\mathrm{A}_i \in \mathbb{R}^{HW\times d_k}$, $d_k = \frac{C}{k}$, $i = 1, 2..., k$). Next, for each head, we use Fully Connected layers (FC) to obtain linear projections for query element Q, key element K, and value element V of each head A_i [15]. Considering that different regions in the V-subspace may contain different brightness information, inspired by Retinexformer [9], we utilize the brightness information encoded in the brightness feature map to guide the self-attention calculation for dark areas. To ensure consistency with the shape of A, the size of the brightness feature map $\mathrm{F}_{v-up} \in \mathbb{R}^{H\times W\times C}$ is reshaped into tokens $\mathrm{B} \in \mathbb{R}^{HW\times C}$ and divided into k heads.

Retinexformer obtains brightness feature map from the input RGB image, which is a three-channel tensor. Unlike Retinexformer, HSVFormer retains color space after HSV transformation and solely encodes the brightness information of the V-subspace, which is a single-channel tensor. Additionally, Retinexformer stacks multiple Transformer blocks at different scales, respectively, with each block containing 8 attention heads, and the input dimension of each head is 64. In contrast, HSVFormer stacks single Transformer block at different scales, respectively, with each block containing 4 attention heads, and the input dimension of each head is 24. In terms of multi-head channel self-attention, HSVFormer consumes fewer computational resources and reduces parameters for self-attention computation compared to Retinexformer.

III. Experiments

A. Datasets

The experiments are conducted on the SICE dataset [16], comprising a total of 4413 images. The training dataset contains 2661 images, and the validation dataset contains 720 images. Besides, the test dataset contains 755 images. In addition, to evaluate the generalization and robustness of HSVFormer across datasets, we test 257 low-light images from five popular datasets: DarkFace [17], DICM [18], LIME [19], MEF [20], and VV [21]. Notably, these datasets lack reference images with normal lighting conditions.

TABLE I
Full reference image quality evaluation scores on the SICE dataset. Red is the best and blue is the second.

Methods	PSNR↑	SSIM↑	LPIPS↓
LIME [19]	14.666	0.503	0.245
Jiang et al. [13]	16.741	0.527	0.252
RUAS [22]	11.780	0.462	0.379
SCI [23]	14.740	0.505	0.293
IAT [10]	13.577	0.499	0.463
LLFormer [12]	14.785	0.528	0.332
HSVFormer	17.245	0.535	0.248

B. Experimental Setup

HSVFormer is implemented using the PyTorch framework and conducted training on a single NVIDIA Tesla A30 GPU. During the training process of HSVFormer, the batch size is set to 1 and the initial learning rate is 0.0001. HSVFormer employs the Adam optimizer for training. The maximum number of training epochs is set to 100, with evaluations conducted every 10 epochs.

C. Comparative Experiments

HSVFormer is compared with six popular methods, including LIME [19], Jiang et al. [13], RUAS [22], SCI [23], IAT [10] and LLFormer [12].

For the test dataset of SICE, we use Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM) [24], and Learned Perceptual Image Patch Similarity (LPIPS) [25] as evaluation metrics for full-reference image quality assessment. The quantitative results are presented in Table 1. The PSNR and SSIM score of HSVFormer are 17.245 and 0.535, respectively. In terms of suppressing noise and restoring image details, HSVFormer is superior to other popular models on the SICE dataset. Fig. 4 presents the visualization results generated

TABLE II
BRISQUE AND PIQE EVALUATION SCORES FOR SEVEN METHODS ON FIVE DATASETS. RED IS THE BEST AND BLUE IS THE SECOND.

Methods	DarkFace [16]		DICM [17]		LIME [18]		MEF [19]		VV [20]		Mean	
	BRISQUE↓	PIQE↓	BRISQUE↓	PIQE↓	BRISQUE↓	PIQE↓	BRISQUE↓	PIQE↓	BRISQUE↓	PIQE↓	BRISQUE↓	PIQE↓
LIME [19]	0.493	10.142	0.554	9.386	0.523	10.859	0.523	9.967	0.494	10.389	0.517	10.149
Jiang et al. [13]	0.490	9.421	0.551	8.176	0.457	11.050	0.515	9.261	0.598	9.292	0.522	9.440
RUAS [22]	0.491	9.887	0.492	14.336	0.494	11.928	0.495	11.869	0.492	11.869	0.493	11.978
SCI [23]	0.495	9.323	0.535	8.417	0.501	11.550	0.514	12.145	0.485	7.181	0.506	9.723
IAT [10]	0.504	13.758	0.499	16.813	0.541	16.153	0.510	19.332	0.553	12.828	0.521	15.777
LLFormer [12]	0.507	10.813	0.507	4.350	0.512	8.941	0.522	14.611	0.500	8.751	0.511	9.493
HSVFormer	0.486	9.218	0.490	7.696	0.458	10.981	0.495	9.046	0.472	9.834	0.480	9.355

Fig. 5. Visualization results generated by different methods on the DarkFace dataset.

by seven methods on the SICE dataset. Fig. 4 reveals that HSVFormer can achieve better visualization effects, especially brightness and colors than other popular methods.

For the remaining five reference-free image test datasets, we employ Blind/Reference-Free Image Spatial Quality Evaluator (BRISQUE) [26] and Perception-based Image Quality Evaluator (PIQE) [27] as evaluation metrics for no-reference image quality assessment. The results on five no-reference datasets are reported in Table 2. It is evident that HSVFormer achieves preferably score on the DarkFace, DICM, MEF, and VV datasets. Notably, average scores across the five datasets, HSVFormer attains the first in terms of BRISQUE and PIQE, demonstrating the robust generalization capacity of HSVFormer across diverse datasets. Fig. 5 presents the visualization results generated by seven methods on the DarkFace dataset. Fig. 5 reveals that HSVFormer is good at restoring smooth illumination. Furthermore, HSVFormer demonstrates superior capabilities in denoising and recovering intricate details in dark areas.

D. Ablation Study

We perform an ablation study on the SICE dataset to quantitatively evaluate the effectiveness of our contributions. Jiang et al. [13] is chosen as the baseline, and the performance is evaluated through PSNR, SSIM and LPIPS. The results are summarized in Table 4. In Table 4, it can be seen that our contributions can effectively remove noise in dark areas and help restore detailed information.

IV. CONCLUSION

In this work, we have investigated an unsupervised deep learning method for low-light image enhancement. Firstly, we propose HSVFormer, which decouples the image into two subspaces, preserving color information and adaptively enhancing the brightness. Secondly, we design the UVENet to estimate and recover brightness. Thirdly, we devise the VG-MCSA to further remove the noise and restore the brightness and detail information of the dark areas. Finally, the color and brightness are recombined to generate an enhanced image. Experimental results demonstrate that HSVFormer can effectively remove noise and artifacts, retain details and color information, and achieve the better visual effects than other popular methods. In the future, we will focus on reducing the computational complexity of HSVFormer while preserving generalization capacity.

TABLE III
RESULTS OF THE ABLATION STUDY. RED IS THE BEST AND BLUE IS THE SECOND.

Methods	PSNR↑	SSIM↑	LPIPS↓
Jiang et al. (CNN based) [13]	10.665	0.288	0.431
Jiang et al. (Transformer based)	16.164	0.514	0.266
+V-up	16.589	0.526	0.250
+V-up Map	16.736	0.527	0.249
+V-up Feature (HSVFormer)	17.245	0.535	0.248

ACKNOWLEDGMENT

This work is partly supported by National Natural Science Foundation of China (Nos. 61861024, 62271296, and 62201334), Scientific Research Program Funded by Education Department of Shaanxi Provincial Government (Nos. 23JP022, and 23JP014), and Key Research and Development Program of Shaanxi (No. 2021ZDLGY08-07).

References

[1] C. Wei, W. Wang, and W. Yang, et al., “Deep retinex decomposition for low-light enhancement,” arXiv preprint arXiv: 1808.04560, 2018.

[2] Y. Zhang, J. Zhang, and X. Guo, “Kindling the darkness: A practical low-light image enhancer,” in Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp.1632-1640.

[3] Y. Zhang, X. Guo, and J. Ma, et al., “Beyond brightening low-light images,” International Journal of Computer Vision, vol. 129, no. 4, pp. 1013-1037, 2021.

[4] H. Fu, W. Zheng, and X. Meng, et al., “You Do Not Need Additional Priors or Regularizers in Retinex-Based Low-Light Image Enhancement,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 18125-18134.

[5] Y. Jiang, X. Gong, and D. Liu, et al., “EnlightenGAN: Deep light enhancement without paired supervision,” IEEE Transactions on Image Processing, vol. 30, pp. 2340-2349, 2021.

[6] C. Guo, C. Li, and J. Guo, et al., “Zero-reference deep curve estimation for low-light image enhancement,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1780-1789.

[7] C. Li, C. Guo, and C. Loy, “Learning to enhance low-light image via zero-reference deep curve estimation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 8, pp. 4225-4238, 2021.

[8] S. Yang, M. Ding, and Y. Wu, et al., “Implicit neural representation for cooperative low-light image enhancement,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 12918-12927.

[9] Y. Cai, H. Bian, and J. Lin, et al., “Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement,” arXiv preprint arXiv: 2303.06705, 2023.

[10] Z. Cui, K. Li, and L. Gu, et al., “You Only Need 90K Parameters to Adapt Light: a Light Weight Transformer for Image Enhancement and Exposure Correction,” arXiv preprint arXiv: 2205.14871, 2022.

[11] X. Xu, R. Wang, and C. W. Fu , et al., “SNR-aware low-light image enhancement,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 17714-17724.

[12] T. Wang, K. Zhang, and T. Shen, et al., “Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2023, pp. 2654-2662.

[13] Z. Jiang, H. Li, and L. Liu, et al., “A switched view of Retinex: Deep self-regularized low-light image enhancement,” Neurocomputing, vol. 454, pp. 361-372, 2021.

[14] E. Land “The Retinex theory of color vision,” Scientific American, vol. 237, no. 6, pp. 108-129, 1977.

[15] A. Vaswani, N. Shazeer, N.Parmar, et al., “Attention is all you need”, arXiv preprint arXiv: 1706.03762, 2017.

[16] J. Cai, S. Gu, and L. Zhang, “Learning a deep single image contrast enhancer from multi-exposure images,” IEEE Transactions on Image Processing, vol. 27, no. 4, pp. 2049-2062, 2018.

[17] W. Yang, Y. Yuan, and W. Ren, et al., “Advancing image understanding in poor visibility environments: A collective benchmark study,” IEEE Transactions on Image Processing, vol. 29, pp. 5737-5752, 2020.

[18] C. Lee, C. Lee, and C. S. Kim, “Contrast enhancement based on layered difference representation of 2D histograms,” IEEE Transactions on Image Processing, vol. 22, no. 12, pp. 5372-5384, 2013.

[19] X. Guo, Y. Li, and H. Ling, “LIME: Low-light image enhancement via illumination map estimation,” IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 982-993, 2016.

[20] K. Ma, L. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3345-3356, 2015.

[21] V. Vonikakis, R. Kouskouridas, and A. Gasteratos, “On the evaluation of illumination compensation algorithms,” Multimedia Tools and Applications, vol.77, pp. 9211-9231, 2018.

[22] R. Liu, L. Ma, and J. Zhang, et al., “Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10561-10570.

[23] L. Ma, T. Ma, and R. Liu, et al., “Toward fast, flexible, and robust low-light image enhancement,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5637-5646.

[24] Z. Wang, A. Bovik, and H. Sheikh, et al., “Image quality assessment: From error visibility to structural similarity,” IEEE Transactions on Image Processing, vol.13, no. 4, pp. 600-612, 2004.

[25] R. Zhang, P. Isola, and A. Efros, et al., “The unreasonable effectiveness of deep features as a perceptual metric,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 586-595.

[26] A. Mittal, A. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” IEEE Transactions on Image Processing, vol. 21, no. 12, pp. 4695-4708, 2012.

[27] N. Venkatanath, D. Praneeth, and M. Bh, et al., “Blind image quality evaluation using perception based features,” in Proceedings of the 21th National Conference on Communications, 2015, pp. 1-6.