2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

A Retinex Structure-based Low-light Enhancement Model Guided by Spatial Consistency

Miao Zhang[1],Yiqing Shen [2], Zhuowei Li[3], Guofeng Pan[4] and Shuai Lu[1*]

Abstract— **Images captured by robotics under low-light conditions are often plagued by several challenges, including diminished contrast, increased noise, loss of fine details, and unnatural color reproduction. These factors can significantly hinder the performance of computer vision tasks such as object detection and image segmentation. As a result, improving the quality of low-light images is of paramount importance for practical applications in the computer vision domain. To effectively address these challenges, we present a novel low-light image enhancement model, termed Spatial Consistency Retinex Network (SCRNet), which leverages the Retinex-based structure and is guided by the principle of spatial consistency. Specifically, our proposed model incorporates three levels of consistency: channel level, semantic level, and texture level, inspired by the principle of spatial consistency. These levels of consistency enable our model to adaptively enhance image features, ensuring more accurate and visually pleasing results. Extensive experimental evaluations on various low-light image datasets demonstrate that our proposed SCRNet outshines existing state-of-the-art methods, highlighting the potential of SCRNet as an effective solution for enhancing low-light images.**

I. INTRODUCTION

In the development of future human habitats, there will be a significant reliance on profound semantic perception capabilities and intelligent interaction mechanisms. In this context, robotics deploying various computer vision algorithms hold paramount significance, as it will greatly extend the depth of perception and enhance the quality of interactive experiences. However, most computer vision algorithms under low-light conditions often pose several challenges [15, 28, 33], such as a decrease in contrast, the presence of unexpected noise, loss of fine details, and unnatural color reproduction. These degradation issues can significantly increase the complexity of high-level tasks such as object detection and image segmentation, making it challenging to obtain accurate results. Therefore, enhancing low-light images is of immense practical value in the field of computer vision.

Various adjustments can be made to the camera imaging parameters to acquire high-quality images in low-light environments, such as increasing the ISO, extending the exposure time, or utilizing a flash. However, each of these methods has its limitations. For example, a high ISO setting can amplify the image sensor's sensitivity to light, but it can also intensify noise, leading to a low signal-to-noise ratio (SNR). Prolonged exposures can result in blurred outcomes, especially when capturing dynamic scenes, and using a flash can cause unbalanced lighting and unnatural colors in the photograph. As a result, developing an effective low-light image enhancement technique that can simultaneously reduce darkness and mitigate degradation issues is of utmost importance. Overcoming these challenges remains a significant obstacle in the field of computer vision.

To address the challenges mentioned above, a large number of algorithms have been proposed to enhance the subjective and objective quality of low-light images. These methods can be broadly categorized into three groups based on their design concepts: distribution mapping-based methods, Retinex-based methods, and deep learning-based methods. Distribution mapping-based methods mainly employ curve transformation, histogram equalization, and other techniques to change the pixel distribution and obtain clear and highlighted images[24, 34]. these methods are limited by their inability to obtain intrinsic distribution and semantic information between pixels, leading to color distortion and detail anomalies in the generated images [6]. The retinex-based approach shows that the image can be decomposed into reflection and illumination components [20, 17]. Reflection is an inherent property of a scene, while illumination is influenced by ambient illumination. However, due to the limitations of their design and a priori knowledge, these conventional algorithms often result in underexposure images, unsaturated colors, and significant artifacts or noise[32, 26]. With the technological explosion of deep learning, the learning-based method [29, 2, 21] has become a mainstream approach for obtaining mappings between low-light inputs and enhanced images by designing heuristic network structures.

While learning-based models have shown promising results in enhancing low-light images, they often process RGB images as a whole, ignoring the fact that each color channel carries different detail information. This can cause interference between channels and result in loss of details. To address this issue, we propose a novel enhancement and denoising model for R, G, and B channels based on the assumption of single-channel consistency.

In more detail, our model is built on the Retinex model and consists of two main parts: Decomposition and Enhancement. The Decomposition part takes the low light image as input and outputs the illumination map and reflectance map through semantic, instance, and texture paths. The En-

[1]Miao Zhang and Shuai Lu are with the Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China. Email: zhangmiao@sz.tsinghua.edu.cn, shuai.lu@sz.tsinghua.edu.cn

[2]Yiqing Shen is with the Johns Hopkins University, USA. Email: yshen92@jhu.edu

[3]Zhuowei Li is with the University of Nottingham Ningbo, Yongjiang Laboratory ,China. Email: zhuowei.li@nottingham.edu.cn

[4]Guofeng Pan is with the Shenzhen Yijiahe Technologies, Shenzhen, China. Email: nchu_pgf@163.com

*Corresponding author

hancement part utilizes the reflectance map for detail repair, denoising, and progressive color correction by following the semantic and texture consistency. The brightness and contrast of the illumination map are adjusted using the Adjustment network. Finally, the corrected illumination map and the reflectance map are combined to obtain the enhanced image. Both qualitative and quantitative empirical results demonstrate that our method produces accurate illumination maps and achieves more natural results with better details compared to state-of-the-art methods.

The main contributions of this work can be summarized as follows.

- The Decompose-net section of our paper presents two novel modules. Firstly, we introduce the Cascading Texture-Instance-Semantic Feature Fusion Module (CFM), which consists of three branches to capture semantic, instance, and texture information separately. Secondly, we propose the Channel-dependent Spatially Consistent Denoising Module (CDM), which performs bilateral filtering on each of the three RGB channels individually to achieve spatially consistent denoising.
- For the task of Detail Restoration, this paper presents an encoder-decoder structure and introduces a novel and effective Regional Consistency based Non-rigid Sampling Pyramid Module (RPM) that focuses on restoring the dark regions of the image based on the illumination distribution at different scales.
- The Progressive Binomial Color Correction Module (PCM) is introduced in this paper, which utilizes a color matrix with binomial expansion for more effective nonlinear color correction.
- Comprehensive experiments were conducted on several benchmark datasets to demonstrate the superiority of our method over the existing state-of-the-art methods. Furthermore, we conducted an ablation study to validate the effectiveness of our proposed structure.

II. The Proposed Method

In this section, we will first provide an overview of the Retinex model, which serves as the basis for our approach. We will then introduce the novel components proposed in our model and explain how we estimate the illumination and reflectance (as well as noise), including details on the network structure, loss function, and implementation.

A. Retinex Model

Our network is based on the Retinex model, which describes the luminance and color perception of human vision and has been widely used for low-illumination image enhancement. It where *r,g,b* represents the RGB three channels, * represents the pixel-level multiplication, *S* represents the observed image (i.e., weakly illuminated image), *R* represents the reflectance component, and *I* represents the illumination component. In this network, *N* represents noise. Since the

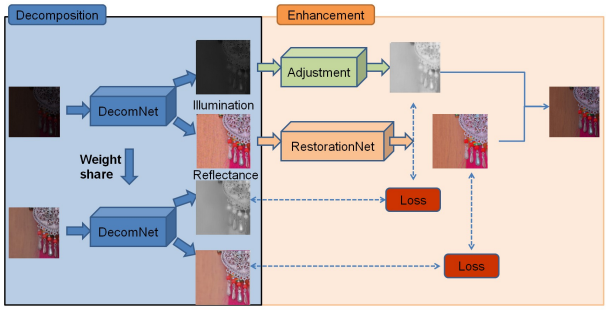

Fig. 1. The overview of our method The overview of the proposed SCRNet. That consists of the decomposition part and enhancement part

different color channels (i.e., red, green, and blue) may have different noise characteristics, we perform denoising and color correction separately on each channel. This helps to ensure that the enhancement process is performed in a more homogeneous way across all channels.

$$S = (R^{*}I + N)_{\{r,g,b\}} \tag{1}$$

B. Overall Network Architecture

For a low-light image enhancement model to be effective, it must be able to recover image details while also addressing the challenges of noise, color distortion, and degradation that often occur in low-light environments. To achieve this, we propose a new deep network architecture consisting of two main components: a Decomposition component and an Enhancement component, as shown in Figure 1. The Decomposition component includes a decomposition network that separates the low-light image into its illumination and reflectance components, while also reducing noise. The Enhancement component comprises two separate networks: a restoration network that focuses on the reflectance and improves its quality, and an adjustment network that enhances contrast and adjusts the lighting of the illumination component. In the following sections, we will describe these two networks in detail, explaining how they function and contribute to the overall effectiveness of our model.

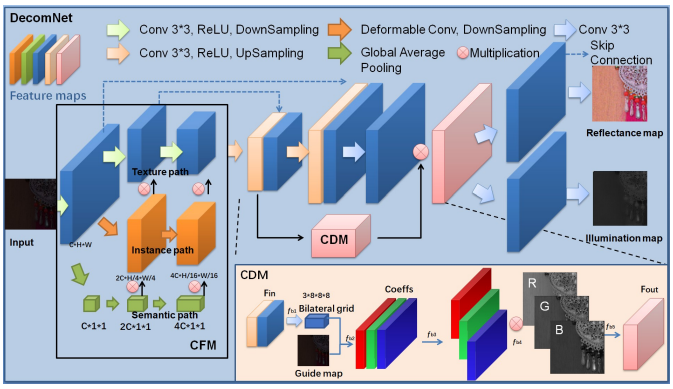

Fig. 2. The architecture of Decomposition. The encoder-decoder Net mainly includes CFM and CDM.

C. Decomposition

In Figure 2, the Decomposition component is illustrated, which takes the input image and decomposes it into two components: R and I. This is achieved by passing the input image through an encoder-decoder CNN network that extracts feature maps. The network uses a two-path approach: one path incorporates semantic information inspired by [13] to obtain high-level features, while the other path uses a Unet-like structure to obtain instance and texture features at lower levels. These features are then fused together to obtain a combined feature map. Finally, bilateral filtering is applied to denoise and correct the R-, G-, and B- channels individually.

- CFM(Cascading Texture-Instance-Semantic Feature Fusion Module)

It is important to note that enhancing under-exposed photos is a challenging task that requires adjustments to texture, instance, and semantic features. Semantic features refer to the overall properties of the image, such as its color balance, brightness, and scene category, while texture features pertain to specific regions of the image and include elements such as highlighting detail sharpness, and contrast. However, traditional image enhancement algorithms often focus either on adjusting texture features, such as methods based on local contrast enhancement[27] or on adjusting semantic features, such as histogram equalization [25]. It is difficult to balance both aspects[14]. Some deep-learning-based methods extract semantic and texture features only on a single feature map[30, 9] or only fuse them once at the end of the model[7]. Therefore, in this paper, CFM (Cascading semantic, instance, and texture feature fusion module) is introduced. CFM characterizes semantic, instance, and texture features from different scale feature maps and fuses them at each block.

For semantic features (Semantic path), a global average pooling operation is been used to compress the feature map *Fin* with a dimension of *C*H*W* into *C*1*1*. A fully connected (FC) operation is then performed for channel expansion to expand the number of channels to 2C, allowing for the representation of the combinations of different channels. Finally, the channel size is expanded to *4C*1*1* to represent the importance of each channel. This is then done by multiplying the weights by the original feature map *Fin* to produce a more robust feature representation, *Fout*.

For instance features (Instance path), a deformable CNN has been used to capture irregular instance information. The feature map with *C*H*W*, is continuously downsampled to *2C*H/4*W/4* and *4C*H/16*W/16* to extract instance features at different scales. For the texture features (Texture path), a classical encoder-decoder structure is used to extract texture features. To avoid gradient vanishing and information loss, a skip connection structure is employed where the output of each convolutional layer is connected to the input of the corresponding deconvolutional layer in the upsampling path. This enables the model to learn detailed information in the image.

- CDM (Channel-dependent Spatially Consistent Denoising Modules)

During the process of image formation, noise is inevitably introduced, such as image grain noise and color noise. Previous studies[11, 9, 16] have applied global denoising methods on all three RGB channels, assuming that the information is uniformly distributed across the various channels. Nonetheless, this assumption is frequently not valid, and this technique may lead to various types of noise being treated as identical noise types, resulting in interference among channels and diminishing the efficacy of denoising.

To avoid inter-channel interference and improve the denoising effect, this paper proposes a separate denoising approach for each of the three RGB channels, coupled with bilateral filtering. Unlike previous methods that assumed the identical distribution of information across channels, this approach considers the differences in noise type between channels. The bilateral filter, as shown in equation 2, takes into account both the spatial distance and intensity difference between pixels, preserving edge information and enhancing image details and contrast. This helps to prevent the generation of noise and artifacts in low-light scenes.

$$\hat{I}(p) = \frac{1}{W_p} \sum_{q \in S} I(q) f_r(\|p-q\|) f_s(\|I(p)-I(q)\|) \qquad (2)$$

Here, $\hat{I}(p)$ is the filtered value of the pixel p, S is the spatial domain of the filter, W_p is the normalization factor, f_r and f_s are the range kernels for distance and intensity difference, respectively. The range kernels determine the influence of the pixel q on the filtered value of the pixel p, based on the spatial distance between p and q, and the intensity difference between p and q. The larger the distance or intensity difference, the smaller the influence.

More specifically, bilateral grids [4, 5, 9] have been proposed to extend bilateral filters to the high-dimensional feature space of images, allowing for the smoothing and enhancement of image features in a similar way to conventional pixels. This is achieved by mapping image features into a grid, where the similarity between pixels can be defined as the similarity between neighboring cells in the grid, based on spatial consistency.

In this paper, Deep Bilateral is performed as follows.

1. The input is the intermediate feature layer *Fin*, and the bilateral grid is generated with a size of (3,8,8,8) by performing downsampling.

2. The bilateral grid is trilinear interpolated to obtain the coefficients (Coeffs), which have the same resolution as the guidance map with dimensions of *C*H*W*.

3. The obtained coefficients are combined with the input feature map in the backbone, resulting in a single-channel filter.

4. The combination operation is then performed for all three RGB channels to obtain the final denoised output.

D. Enhancement

The enhancement part consists of a lightweight multi-branch restoration network that performs various restoration functions such as contrast enhancement, noise suppression, and color correction. Additionally, an illumination adjustment network is included to modify the light level.

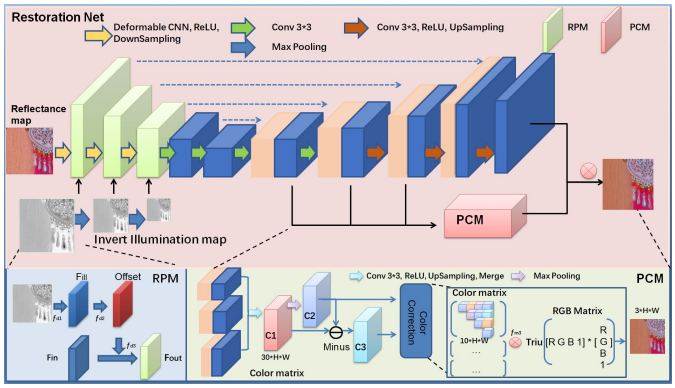

Fig. 3. The architecture of Restorationnet. The encoder-decoder Net is divided into RPM and PCM.

1) Restoration Net: The Restorationnet consists of an encoder-decoder framework and multiple branches designed for different restoration tasks, as shown in Figure 3. Specifically, it includes branches for detail restoration, denoising, and color correction.

- RPM(Regional Consistency based Non-rigid Sampling Pyramid Module)

Due to low light or poor environmental conditions, images often have low contrast and blur, so the operation of detail preservation for dark light enhancement is performed.

The illumination image is used as a guidance map to restore the details of the dark parts in a targeted manner. To make the computation more efficient during runtime, the illumination map is inverted to obtain the inverted illumination map.

To guide the detailed recovery of images at different scales, we have introduced a Laplace pyramid structure [19] at three scales. The pyramid structure helps in preserving high-frequency details by processing the information of images at different scales. This is particularly useful in avoiding local artifacts that often occur in the output of optical enhancement networks when fed with high-resolution inputs. In our encoder, the pyramid structure is implemented through sequential maximum pooling to obtain a multi-scale illumination map that extracts hierarchical features.

The model focuses on the dark parts of the image, which are key areas for improvement. Lighting is a holistic feature of the region, and its spatial consistency is present [38, 3]. However, due to varying lighting conditions, the spatial distribution of dark areas can become irregular. Traditional convolutional operations cannot effectively extract relevant information, so deformable convolution [8] is used instead. Unlike other studies [22, 23], the invert illumination map is used as the offset map instead of the input feature. This directs the model to focus more on the dark areas that contribute to noise and loss of detail.To provide the offset in the x,y direction, the offset has a dimension of 2N, where N is the area of the convolution kernel. The adaptive sampling of the input feature map *Fin* can then be performed according to the offset. To emphasize the dark areas as the key areas, an attention mechanism is introduced. This attention mechanism allows the model to focus more on these areas and achieve targeted detail recovery.

- PCM(Progressive Binomial Color Correction Module)

In low light conditions, the images may have significant color deviation due to the influence of noise in the camera sensor [3]. Therefore, color correction is needed to ensure the accuracy and stability of the image.

This section introduces a novel approach that fully utilizes feature maps of different scales. Firstly, three feature maps of different scale sizes are adjusted to the same resolution, denoted as *C1*, using convolution. Since color is spatially continuous and has a correlation between adjacent positions, maximum pooling is used to obtain local area information denoted as *C2*, which considers information from multiple surrounding points and represents overall characteristics to some extent. However, *C2*'s correction only yields a coarse result. To obtain finer correction information, fine-grained correction information *C3* is obtained based on the offset between each point's own value (*C1*) and the pooled value (*C2*). Finally, *C2* and *C3* are superimposed to obtain the final color-corrected result.

For the color correction part, previous studies[9, 22] typically used a 3*3 or 3*4 matrix for color correction, which only allows for RGB linear conversion with limited fitting effectiveness. To improve on this, we utilize a transformation method based on a 10*H*W color matrix, which is derived from the binomial expansion of the above-mentioned coefficients. *Triu (·)* is the vectorized form of the elements in the upper triangular matrix. This approach provides a better fit for nonlinear color correction.

E. Adjustment

The Adjustment network is a composite network comprising a 6-layer UNET and ResNet. By leveraging both long and short skip connections, this network establishes a mapping mechanism aimed at accurately restoring image details and preserving colors.

F. Loss Function

Decomposition The loss L_{decom} consists of two terms: the first term is reflectance similarity, the second term is reconstruction loss:

$$\mathscr{L}_{\text{decom}} = \left\| R_{\text{low}} - R_{\text{high}} \right\|_1 + \sum_{i=\text{low,high}} \left\| R_{\text{i}} \circ I_{\text{i}} - S_{\text{i}} \right\|_1 . \quad (3)$$

where R_{low} and R_{high} denote the reflectance of low image and high image.I and S denote the illumination map and rgb image.$\|\|_1$ means the ζ^1 norm.

Restoration net

The loss L_{re} consists of three terms: the first term is also reflectance similarity, the second term is reconstruction loss:

$$\mathscr{L}_{\text{re}} = \left\|\hat{\mathbf{R}} - \mathbf{R}_{high}\right\|_2^2 - \text{SSIM}\left(\hat{\mathbf{R}}, \mathbf{R}_{high}\right) + \left\|\nabla\hat{\mathbf{R}} - \nabla\mathbf{R}_{high}\right\|_2^2 \tag{4}$$

where SSIM$(\cdot,\ \cdot)$ is the structural similarity measurement, and $\hat{R}$ corresponds to the restored reflectance. ∇ stands for the first order derivative operator containing ∇x (horizontal) and ∇y (vertical) directions. $\|\|_2$ means the ζ^2 norm(MSE).

Illmination adjustment net

The loss L_{ill} consists of two terms: the first term is illumination map similarity, and the second term is the edge similarity:

$$\mathscr{L}_{\text{ill}} = \left\|\hat{\mathbf{I}} - \mathbf{I}_{high}\right\|_2^2 + \left\|\nabla\hat{\mathbf{I}} - \nabla\mathbf{I}_{high}\right\|_2^2 \tag{5}$$

where $\|\|_2$ means the ζ^2 norm(MSE).

G. Our Dataset

In our experiments, we employed two datasets for training our model: LOL [35] and MIT5K [1]. The LOL dataset contains 500 pairs of real-world low-light/normal-light images, which is the first dataset created for low-light image enhancement. We split the LOL dataset into three subsets with 400, 50, and 50 image pairs for training, validation, and testing, respectively. The MIT5K dataset contains 5,000 images, out of which we used 4,500 images for training, and the remaining 500 images were used for validation and testing.

Evaluation metrics. PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index) are commonly used metrics to evaluate the quality of the enhanced images compared to the ground truth images. Higher PSNR values indicate less distortion between the two images, while higher SSIM values indicate higher structural similarity between the two images. Both metrics are widely used in the image processing field to measure the quality of the image enhancement. In general, higher PSNR and SSIM values indicate better results and more realistic human perception of the images.

III. Experimental Results

In this section, we use a large number of experiments to evaluate our method. First, we compare our method with the current state-of-the-art enhancement methods in terms of quality and quantity. Then, we provide additional analyses to fully demonstrate the advantages of our method.

A. Implementation Details

Our proposed method was implemented using the PyTorch framework and trained on an Nvidia 2080T GPU. We trained the lightweight network for 400 iterations on our proposed dataset, with random clipping, flipping, and rotation used to augment the data and prevent overfitting. We used the Adam optimizer [18] to optimize the network, with hyperparameters $\alpha = 0.001$, $\beta 1 = 0.9$, $\beta 2 = 0.999$, and $\varepsilon = 10^{-8}$.

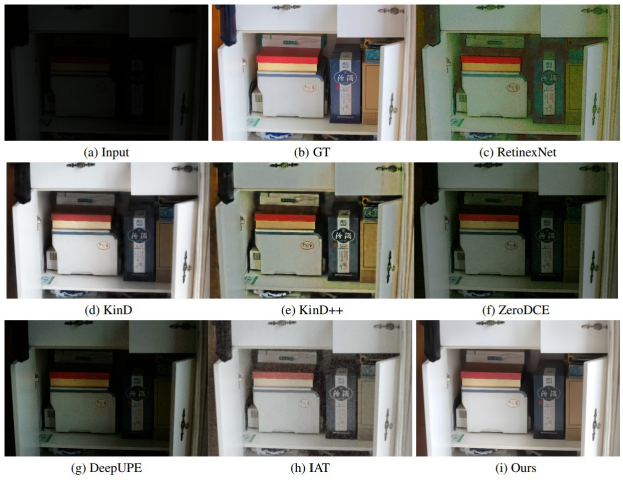

Fig. 4. Visual results on LOL dataset

TABLE I

Quantitative comparison on LOL dataset in terms of PSNR, SSIME. The best results are highlighted in bold.

Datasets	Method	PSNR	SSIM
LOL	Retinex-Net	16.674	0.490
	KinD	20.882	0.791
	KinD++	21.311	0.821
	ZeroDCE	14.96	0.573
	DeepUPE	16.798	0.519
	IAT	22.681	0.818
	ours	**23.162**	**0.835**

B. Qualitative Results

In this section, we present the results of a comprehensive set of experiments to evaluate the effectiveness of our proposed method. We start by comparing the performance of our method with the state-of-the-art enhancement methods in terms of both quality and quantity. We then provide additional analyses to demonstrate the advantages of our method.

We compare our proposed method with several existing state-of-the-art methods, including RetinexNet [35], KinD[36], KinD++[37], DeepUPE[31], IAT[7] and ZeroDCE[10], on two publicly available datasets: LOL and MIT5K. Our model is trained on the LOL dataset and evaluated on both datasets using two commonly used image quality metrics, namely PSNR and SSIM [12]. In Table I, we report the quantitative comparison results. Our proposed method achieves superior performance in terms of both metrics.

TABLE II

Quantitative comparison on MIT5K dataset in terms of PSNR, SSIME. The best results are highlighted in bold.

Datasets	Method	PSNR	SSIM
MIT5K	RetinexNet	14.80	0.720
	KinD	17.58	0.686
	DeepUPE	23.04	0.893
	ZeroDCE	16.99	0.813
	IAT	25.32	0.920
	ours	**26.16**	**0.962**

Based on Table I, our proposed method shows superior performance compared to existing state-of-the-art methods in terms of PSNR and SSIM on the LOL dataset. Our method achieves the best performance with an average PSNR of 23.162 dB and SSIM of 0.835, which is 0.481 dB higher than the second-best method (IAT) in PSNR and 0.017 in SSIM. It is also 1.851 dB higher than the third-best method (KinD++) in PSNR and 0.024 in SSIM. However, RetinexNet, DeepUPE, and ZeroDCE are not as effective. On the other hand, according to Table II, our method achieves the highest scores with PSNR of 26.16 dB and SSIM of 0.962 on the MIT5K dataset.

C. Visual Comparisons

A visual comparison of the enhancement results produced by our method and other Retinex theory-based methods (e.g. KinD, RetinexNet) is shown in Figure 4. Our method effectively enhances the contrast, improves details, and removes noise, as can be seen in the visual comparison. Previous methods, on the other hand, tend to blur the details or amplify the noise, which is well demonstrated in the experimental results.

The visual comparisons in Figure 4 demonstrate the performance of various enhancement methods on the LOL dataset. We can observe that RetinexNet produces images with significant color distortion and noise. Although KinD++ performs comparably, the results are slightly dark and lack subtle details. ZeroDCE and DeepUPE produce dimmer images with heavily hidden details. IAT improves brightness but still suffers from noise and color distortion. In contrast, our proposed method recovers true colors and obtains more texture details without noise compared to other methods.

Overall, our method is capable of enhancing dark areas without overexposure artifacts and maintaining high-contrast texture details. The pyramid structure, luminance-aware guidance, and contrast-attentive mechanism all contribute to the model's ability to predict reasonable adjustments and reconstruct high-quality images.

D. Ablation Study

To assess the effectiveness of each module in our network, we performed experiments on the LOL dataset by removing each of the four modules (CFM, CDM, RPM, PCM) individually. The results are presented in Table III. It can be seen that removing the CDM and PCM modules leads to a significant decrease in the performance of our model. For more details, please refer to Figures 5 and 6.

(a) Baseline (b) w/CFM&CDM (c) GT

Fig. 5. Illumination maps

(a) Baseline (b) w/CFM&CDM (c) w/RPM&PCM (d) GT

Fig. 6. Reflectance maps

TABLE III
ABLATION OF CFM, CDM, RPM AND PCM ON LOL DATASET IN TERMS OF PSNR.

Case	CFM	CDM	DPM	PCM	LOL
1		✓	✓	✓	22.925
2	✓		✓	✓	21.964
3	✓	✓		✓	22.647
4	✓	✓	✓		21.683
5	✓	✓	✓	✓	23.162

IV. CONCLUSION

In this paper, we propose a new end-to-end enhancement network based on Retinex theory for low-light pictures and normal-light pictures. The network consists of two parts: the decomposition network, and the enhancement network. In this paper, guided by spatial consistency, we combine semantic and texture information, bilateral filtering, and binomial color correction to edge recovery, denoise and color correction for a single channel of RGB image, so that the enhancement results obtained by our method have better visual effects. And it is verified that our scheme of low illumination image recovery is correct and feasible. Experimental results on the LOL dataset show that our method can improve the image contrast and good noise rejection, and obtain the highest PSNR and SSIM scores, which are superior to other methods.

ACKNOWLEDGMENT

The work described in this paper is partially supported by Shenzhen Fundamental Research (General Program)(WDZC20231129163533001).

REFERENCES

[1] Vladimir Bychkovsky et al. "Learning photographic global tonal adjustment with a database of input/output image pairs". In: *CVPR 2011*. IEEE. 2011, pp. 97–104.

[2] Jianrui Cai, Shuhang Gu, and Lei Zhang. "Learning a deep single image contrast enhancer from multi-exposure images". In: *IEEE Transactions on Image Processing* 27.4 (2018), pp. 2049–2062.

[3] Chen Chen et al. "Learning to see in the dark". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 3291–3300.

[4] Jiawen Chen, Sylvain Paris, and Frédo Durand. "Real-time edge-aware image processing with the bilateral grid". In: *ACM Transactions on Graphics (TOG)* 26.3 (2007), 103–es.

[5] Jiawen Chen et al. "Bilateral guided upsampling". In: *ACM Transactions on Graphics (TOG)* 35.6 (2016), pp. 1–8.

[6] Heng-Da Cheng and XJ Shi. "A simple and effective histogram equalization approach to image enhancement". In: *Digital signal processing* 14.2 (2004), pp. 158–170.

[7] Ziteng Cui et al. "Illumination Adaptive Transformer". In: *arXiv preprint arXiv:2205.14871* (2022).

[8] Jifeng Dai et al. "Deformable convolutional networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 764–773.

[9] Michaël Gharbi et al. "Deep bilateral learning for real-time image enhancement". In: *ACM Transactions on Graphics (TOG)* 36.4 (2017), pp. 1–12.

[10] Chunle 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.

[11] Jiang Hai et al. "R2rnet: Low-light image enhancement via real-low to real-normal network". In: *Journal of Visual Communication and Image Representation* 90 (2023), p. 103712.

[12] Alain Hore and Djemel Ziou. "Image quality metrics: PSNR vs. SSIM". In: *2010 20th international conference on pattern recognition*. IEEE. 2010, pp. 2366–2369.

[13] Jie Hu, Li Shen, and Gang Sun. "Squeeze-and-excitation networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.

[14] Zhenghua Huang et al. "Global–local image enhancement with contrast improvement based on weighted least squares". In: *Optik* 243 (2021), p. 167433.

[15] Haiyang Jiang and Yinqiang Zheng. "Learning to see moving objects in the dark". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 7324–7333.

[16] Yonglong Jiang et al. "DEANet: Decomposition Enhancement and Adjustment Network for Low-Light Image Enhancement". In: *Tsinghua Science and Technology* 28.4 (2023), pp. 743–753.

[17] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. "A multiscale retinex for bridging the gap between color images and the human observation of scenes". In: *IEEE Transactions on Image processing* 6.7 (1997), pp. 965–976.

[18] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[19] Wei-Sheng Lai et al. "Deep laplacian pyramid networks for fast and accurate super-resolution". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 624–632.

[20] Edwin H Land and John J McCann. "Lightness and retinex theory". In: *Josa* 61.1 (1971), pp. 1–11.

[21] Chongyi Li et al. "LightenNet: A convolutional neural network for weakly illuminated image enhancement". In: *Pattern recognition letters* 104 (2018), pp. 15–22.

[22] Jinxiu Liang et al. "Deep bilateral retinex for low-light image enhancement". In: *arXiv preprint arXiv:2007.02018* (2020).

[23] Xiaokai Liu et al. "LAE-Net: A locally-adaptive embedding network for low-light image enhancement". In: *Pattern Recognition* 133 (2023), p. 109039.

[24] Etta D Pisano et al. "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms". In: *Journal of Digital imaging* 11 (1998), pp. 193–200.

[25] Ganta Raghotham Reddy et al. "Enhancement of Images Using Optimized Gamma Correction with Weighted Distribution Via Differential Evolution Algorithm". In: *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*. IEEE. 2022, pp. 1–5.

[26] Xutong Ren et al. "LR3M: Robust low-light enhancement via low-rank regularized retinex model". In: *IEEE Transactions on Image Processing* 29 (2020), pp. 5862–5876.

[27] Abdullah Amer Mohammed Salih, Khairunnisa Hasikin, and Nor Ashidi Mat Isa. "Adaptive fuzzy exposure local contrast enhancement". In: *IEEE Access* 6 (2018), pp. 58794–58806.

[28] Yukihiro Sasagawa and Hajime Nagahara. "Yolo in the dark-domain adaptation method for merging multiple models". In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*. Springer. 2020, pp. 345–359.

[29] Liang Shen et al. "Msr-net: Low-light image enhancement using deep convolutional network". In: *arXiv preprint arXiv:1711.02488* (2017).

[30] Ruixing Wang et al. "Underexposed photo enhancement using deep illumination estimation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 6849–6857.

[31] Ruixing Wang et al. "Underexposed photo enhancement using deep illumination estimation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 6849–6857.

[32] Shuhang Wang et al. “Naturalness preserved enhancement algorithm for non-uniform illumination images”. In: *IEEE transactions on image processing* 22.9 (2013), pp. 3538–3548.

[33] Wenjing Wang, Wenhan Yang, and Jiaying Liu. “Hla-face: Joint high-low adaptation for low light face detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 16195–16204.

[34] Yu Wang, Qian Chen, and Baeomin Zhang. “Image enhancement based on equal area dualistic sub-image histogram equalization method”. In: *IEEE transactions on Consumer Electronics* 45.1 (1999), pp. 68–75.

[35] Chen Wei et al. “Deep retinex decomposition for low-light enhancement”. In: *arXiv preprint arXiv:1808.04560* (2018).

[36] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. “Kindling the darkness: A practical low-light image enhancer”. In: *Proceedings of the 27th ACM international conference on multimedia*. 2019, pp. 1632–1640.

[37] Yonghua Zhang et al. “Beyond brightening low-light images”. In: *International Journal of Computer Vision* 129 (2021), pp. 1013–1037.

[38] Xuan Zou, Josef Kittler, and Kieron Messer. “Illumination invariant face recognition: A survey”. In: *2007 first IEEE international conference on biometrics: theory, applications, and systems*. IEEE. 2007, pp. 1–8.