Bridge the Vision Gap From Field to Command: A Deep Learning Network Enhancing Illumination and Details

Zhuqing Jiang
Beijing University of Posts and Telecommunications
Beijing, China
jiangzhuqing@bupt.edu.cn

Chang Liu
Beijing University of Posts and Telecommunications
Beijing, China
m13261632677@163.com

Ya’nan Wang
Beijing University of Posts and Telecommunications
Beijing, China
wynn@bupt.edu.cn

Aidong Men
Beijing University of Posts and Telecommunications
Beijing, China
menad@bupt.edu.cn

Haiying Wang
Beijing University of Posts and Telecommunications
Beijing, China
why@bupt.edu.cn

Haiyong Luo
Institute of Computing Technology, Chinese Academy of Sciences
Beijing, China
yhluo@ict.ac.cn

Abstract—Low-light image enhancement aims to improve the brightness of images captured in low-light conditions, which has various applications in surveillance, remote sensing and computational photography. However, low-light images often suffer from poor visibility and blurring, and simply brightening the dark regions may amplify blurring and cause detail loss. In this paper, we propose a simple yet effective two-stream framework called NEID that can enhance the brightness and the details simultaneously without introducing much computational cost. Specifically, our method consists of three modules: Light Enhancement (LE), Detail Refinement (DR) and Feature Fusing (FF), that aggregates composite features for multiple tasks using a channel attention mechanism. Extensive experiments on several benchmark datasets show the effectiveness of our method and its superiority over state-of-the-art methods.

Keywords—low-light image enhancement, super-resolution, multi-task learning, codec framework, detail refinement, feature fusion

I. INTRODUCTION

In some fields such as remote sensing, maritime affairs and live broadcasting[3], images are often captured or transmitted under low-light conditions or narrow bandwidth, degrading their quality. However, the end users require bright and detail-rich images, as shown in Fig. 2. Therefore, a post-processing method that brightens images and refines details simultaneously is needed.

Various methods have been proposed over the past decades to improve the subjective and objective quality of low-light images. Early work[4] mainly focused on enhancing contrast, which may not be enough to recover image color and details. With the rapid development of deep neural networks, CNNs have become widely used in low-level computer vision problems, including low-light image enhancement. Recent work[5][6] adopts CNN-based approaches to simultaneously learn adjustments in color, contrast, brightness, and saturation to produce more expressive results. However, these methods still have limitations in brightening extremely dark images, resulting in blurred details. In this paper, we propose an effective framework to address the dilemma. As shown in Fig. 1, the NEID enhances a low-light image by brightening it up, restoring its inherent color and enhancing its details.

Fig. 1. (a) is the normal-light and high-resolution version of the low-light image (b) enhanced by the proposed NEID. (c) is a detailed region of (a). Our method brightens up the challenging low-light image while restoring the inherent color and enhancing the details.

Inspired by the extensive use of super-resolution in broadcasting[7] and broadband multimedia[8][9], we propose a novel framework that restores brightness and enhances details simultaneously, which aims to reconstruct a high-resolution image from a low-resolution input.

Our framework is a two-stream model, consisting of two parallel branches: Light Enhancement (LE) branch and Detail Refinement (DR) branch. The features of the two branches are fused by a Feature Fusing (FF) module. We integrate the idea of super-resolution into low-light image enhancement pipelines, using a U-net as the codec, which forms the LE branch. Then, the encoded features of LE are further enhanced by the fine-grained detail representation from DR with the FF module. These two branches share the same encoder. Finally, DR is optimized with reconstruction supervision only during training, and the post-process of the DR branch is removed from the network in the inference stage, without any computational cost.
The main contributions of this work can be summarized as follows:

1) Inspired by image super-resolution, the proposed method integrates low-light image enhancement and image super-resolution, producing normal-light images with rich details and high visual quality.

2) Based on dual codecs, we propose a two-stream network to tackle this unified task, in which the encoders share the same weights.

3) Our proposed network accepts a low-resolution low-light image as the input, which can reduce computational costs and increase inference speed while maintaining or even improving the performance of low-light image enhancement.

II. METHODOLOGY

A. Review of U-net Framework

The U-net framework is widely applied to low-light image enhancement since it was proposed. The architecture consists of a contracting path and an expansive path. The contracting path, which can be seen as an encoder, repeatedly extracts hierarchical features from input low-light images by applying two 3×3 convolutions followed by a 2×2 max pooling operation with stride 2 for down-sampling. These features include color, brightness, contrast and saturation information for reconstructing normal-light images. The expansive path, which can be seen as a decoder, restores the brightness at each step by fusing the features with up-sampling, a 2×2 convolution, concatenation and two 3×3 convolutions. U-net also has symmetric skip connections between the two paths, which directly propagate low-level features from the encoder to the decoder.

The low-light image enhancement task aims to restore the normal-light image with high visual quality from the low-light input image. High visual quality implies vivid color and precise details. However, most existing methods based on U-net architecture can only restore the brightness, while ignoring image details.
DR and LE share the same feature extractor, as Fig. 3 shows. We assume that although low-light images are barely visible, there are still small differences among pixels. Based on this assumption, we apply super-resolution on low-light images to extract detail information, and LE also brightens these details. Therefore, we supervise the training of LE and DR by sharing an encoder and comparing them to their respective ground truth images. The encoder learns to extract features for both tasks.

4) Feature Fusing module
As we have mentioned before, DR contains more detail information than LE. Therefore, we design a Feature Fusing module to aggregate the features extracted from two branches and use fine-grained features from DR to guide the learning of LE. Specifically, the Feature Fusing module adaptively assigns different weights to different channels based on fine-grained DR features. The weighted sum of these features is computed to obtain the fused features before sending them to the decoder of LE for brightening.

C. Optimization
The whole objective function consists of a Huber loss for Light Enhancement, a Mean Squared Error (MSE) loss for Detail Refinement, and a Color loss to relieve color distortion.

1) Huber Loss
The Huber loss function describes the penalty incurred by an estimator. Huber defines the loss function piecewise as follows:

\[
L_\delta(a) = \begin{cases} 
\frac{1}{2}a^2, & \text{for } |a| \leq \delta, \\
\delta|a| - \frac{1}{2}\delta^2, & \text{otherwise.} 
\end{cases}
\] (1)

The variable \(\alpha\) often refers to the residuals, that is to the difference between the ground truth and output of LE: \(a = I_{GT} - I_{LE}\), so the former can be expanded to

\[
L_{Huber} = \begin{cases} 
\frac{1}{2}(I_{GT} - I_{LE})^2, & \text{for } |I_{GT} - I_{LE}| \leq \delta, \\
\delta|I_{GT} - I_{LE}| - \frac{1}{2}\delta^2, & \text{otherwise.} 
\end{cases}
\] (2)

Two widely used loss functions are the Mean Squared Error (MSE) loss and the Mean Absolute Error (MAE) loss. The MSE loss has the disadvantage of being dominated by outliers. One substantial problem with the MAE loss is its large constant gradient, leading to missing minima at the end of training.

As defined above, the Huber loss function is convex in a uniform neighborhood of its minimum \(\alpha = 0\); at the boundary of this uniform neighborhood, the Huber loss function has a differentiable extension to an affine function at points \(\alpha = -\delta\) and \(\alpha = \delta\). These properties allow it to combine the sensitivity of MSE loss and the robustness of MAE loss.
2) **MSE Loss**
For DR, we apply a conventional MSE loss to supervise its training.

\[ L_{MSE} = \frac{1}{N} \sum_{i=1}^{n} ||I_{DR} - I_{GT}||^2 \]  

(3)

3) **Color Loss**
We have observed that existing low-light image enhancement methods suffer from color distortion. We minimize a Color loss in HSV color space, calculating the cosine distance between the predicted image and the ground truth image in the Hue and Saturation channels. The loss function is expressed as follows:

\[ L_{Color} = L_H + L_S \]  

(4)

\[ L_H = 1 - \cos(H_p, H_{GT}) \]  

(5)

\[ L_S = 1 - \cos(S_p, S_{GT}) \]  

(6)

where \( \cos(\cdot) \) is an operation to calculate the cosine similarity. \( H_p \) and \( H_{GT} \) refer to the Hue channel of the prediction and ground truth, respectively. Similarly, \( S_p \) and \( S_{GT} \) refer to the Saturation channel.

4) **Total Loss**
The total loss is the combination of all three losses above:

\[ L_{total} = L_{Huber} + \lambda_1 L_{MSE} + \lambda_2 L_{Color} \]  

(7)

where \( \lambda_i \) is the weight of corresponding loss. We minimize the total loss end-to-end.

### III. Experiments
We evaluate the results both qualitatively and quantitatively and compare them with the state-of-the-art methods in this part. First, we describe the datasets and provide the implementation details. Then, we report the results for low-light image enhancement.

| Method      | BIREF[13] | CRM[14] | Dong[15] | LIME[16] |
|-------------|-----------|---------|----------|----------|
| PSNR        | 13.88     | 17.20   | 16.72    | 16.76    |
| SSIM        | 0.58      | 0.64    | 0.58     | 0.56     |

| Method      | MF        | Retinex-Net | MSR[17] | NPE[18] |
|-------------|-----------|--------------|---------|---------|
| PSNR        | 18.79     | 16.77        | 13.17   | 16.97   |
| SSIM        | 0.64      | 0.56         | 0.48    | 0.59    |

| Method      | GLAD[19] | KinD[20] | Ours | |
|-------------|----------|---------|------|---|
| PSNR        | 19.72    | 20.87   | 22.83 | - |
| SSIM        | 0.70     | 0.80    | 0.78  | - |
A. Datasets and Evaluation Metrics

We train our algorithm on the LoL$^{22}$ dataset, which consists of 500 image pairs. Each image pair has a low-light input image and a well-exposed reference image. We also evaluate the detail restoration ability of our algorithm on the MIT-Adobe 5K dataset$^{23}$. It contains 5000 images captured with DSLR cameras. Five trained photographers (labeled as experts A to E) manually adjusted the tonal attributes of all images. Following$^{8,9,24}$, we consider the enhanced images of expert C as the ground-truth. Moreover, we use some no-reference datasets, such as LIME$^{16}$, DICM$^{25}$ and NPE$^{18}$, to assess the qualitative performance.

We use two common metrics, PSNR and SSIM, to measure the color and structural similarity between the predicted results and the ground truth images quantitatively. High PSNR and SSIM values indicate good results.

B. Implementation Details

The proposed model is end-to-end trained with the Adam optimizer for 2000 epochs. The initial learning rate is set to $10^{-4}$. It first decreases to $10^{-5}$ at 500th epoch and then to $10^{-6}$ at 1000th epoch. A batch size of 8 is applied. For data augmentation, we randomly crop patches of size 256×256 from normal-light images followed by vertical flips and rotations by multiples of 90 degrees. For the model input, we down-sample the ground truth patches to 128×128 with bicubic interpolation.

The filter weights of each layer are initialized with a standard zero mean and 0.02 standard deviation Gaussian function. Bias is initialized as a constant. Both weights of the standard zero mean and 0.02 standard deviation Gaussian function. Bias is initialized as a constant. Both weights of the convolutional layer are initialized with Xavier uniform initialization.

C. Comparison with state-of-the-art methods

This section evaluates the effectiveness of our algorithm for low-light image enhancement. Table 1 compares the PSNR/SSIM values of our method and several other techniques on the LoL$^{22}$ dataset. Our NEID achieves significant improvements over previous approaches. Notably, NEID obtains about 2 dB performance gain over KinD$^{20}$, the recent best method, on the LoL dataset.

We also evaluate the visual effect on widely-adopted datasets, such as LoL$^{22}$, LIME$^{16}$, NPE$^{18}$, and DICM$^{25}$. We compare our method with some of the above methods. Fig. 5 and Fig. 6 show visual comparisons of images with different contents. Although these methods can brighten the inputs, they cause severe visual degradations, such as blur and color distortion. Our method generates more natural and vivid images.

The above datasets are unsuitable for evaluating the proposed model’s performance in detail restoration, as they contain few images with rich details. Hence, we use MIT Adobe 5K, a high-resolution dataset with detailed images. Fig. 7 compares the results of our NEID and other methods in terms of details. NEID restores images with clear and abundant details, demonstrating the effectiveness of our method.

We first down-sample the input images with bicubic interpolation to generate images of the same size as other methods. Thus, NEID outperforms state-of-the-art methods with smaller input images. Our method reduces the computational costs and improves the inference speed, meeting the scenarios' requirements in Section 1.

D. Ablation Study

We also evaluate the effectiveness of different components in NEID. Table 2 shows that, using the U-net architecture as the baseline, the LE branch improves the performance from 19.60 dB to 20.78 dB. Adding the DR branch further improves the PSNR by about 1 dB. Combining the Feature Fusing module, the performance reaches 22.83 dB (3.23 dB higher than the baseline), indicating that transferring detail information from DR to LE is crucial. The ablation study confirms the necessity of NEID's components.

IV. CONCLUSION

We propose a practical framework, NEID, for low-light enhancement. Drawing on super-resolution, our network increases the output normal-light images’ resolution to suit field-to-command scenarios. The LE branch enhances the illumination and details of input images and upscals them. The DR branch shares the encoder with the LE branch and refines detail information. The Feature Fusing (FF) module fuses the features of both branches, guiding the LE branch’s learning with detail features from the DR branch. Our experiments show the clear advantages of our design over state-of-the-art alternatives.
ACKNOWLEDGMENT

This work was supported by Project No.A14B07C02-202308D1 , Beijing University of Posts and Telecommunications.

REFERENCES

[1] Nancy Mehta, Subrahmanym Murula. Con-Net: A Consolidated Light-Weight Image Restoration Network. IEEE Transactions on Broadcasting, 68(4):862–875, 2022.

[2] J Alex Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. IEEE Transactions on image processing, 9(5):889–896, 2000.

[3] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and YungYu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6306–6314, 2018.

[4] Sung Ju Hwang, Ashish Kapoor, and Sing Bing Kang. Context-based automatic local image enhancement. In European conference on computer vision, pages 569–582. Springer, 2012.

[5] Jongchun Park, Joon-Young Lee, Donggeun Yoo, and In So Kweon. Distort-and-recover: Color enhancement using deep reinforcement learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5928–5936, 2018.

[6] Jin Wan, Hui Yiu, Zhihao Liu, Aixin Chong and Yanting Liu. Lightweight Image Super-Resolution by Multi-Scale Aggregation. IEEE Transactions on Broadcasting, 67(2):372–382, 2021.

[7] Aoyang Zhang, Qing Li, Ying Chen, Xiaoteng Ma, Longhao Zou, Yong Jiang, Zhimin Xu and Gabriel-Miro Muntean. Video Super-Resolution and Caching—An Edge-Assisted Adaptive Video Streaming Solution. IEEE Transactions on Broadcasting, 67(4):799–812, 2021.

[8] Alireza Esmaeilzehi, M.Omair Ahmad and M.N.S.Swamy. SRNMSM: A Deep Light-Weight Image Super Resolution Network Using Multi-Scale Spatial and Morphological Feature Generating Residual Blocks. IEEE Transactions on Broadcasting, 68(1):58–68, 2022.

[9] Xian Zhang, GuangTao Fu, Lin Wang, Zhuqing Jiang and Aidong Men. Few Pains, Many Gains: Fast On-device Image Compression through Super Resolution. In Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pages 1-6, 2021.

[10] Yun Zhou, Tao Liang, Zhuqing Jiang and Aidong Men. Efficient Network Removing Feature Redundancy for Single Image Super-Resolution. In Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pages 1-6, 2022.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.

[12] Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Realtime video super-resolution with spatio-temporal networks and motion compensation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4778–4787, 2017.

[13] Zhenqiang Ying, Ge Li, and Wen Gao. A bio-inspired multieposure fusion framework for low-light image enhancement. arXiv preprint arXiv:1711.00591, 2017.

[14] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wemin Wang. A new image contrast enhancement algorithm using exposure fusion framework. In International Conference on Computer Analysis of Images and Patterns, pages 36–46. Springer, 2017.

[15] Xuan Dong, Guan Wang, Yi Pang, Weixin Li, Jiatao Wei, Meng, and Yao Lu. Fast efficient algorithm for enhancement of low lighting video. In 2011 IEEE International Conference on Multimedia and Expo, pages 1–6. IEEE, 2011.

[16] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on image processing, 26(2):982–993, 2016.

[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. IEEE Transactions on Image processing, 6(7):965–976, 1997.

[18] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing, 22(9):3538–3548, 2013.

[19] Wenjing Wang, Chen Wei, Wenhao Yang, and Jiaying Liu. Gladnet: Low-light enhancement network with global awareness. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pages 751–755. IEEE, 2018.

[20] 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, pages 1632–1640, 2019.

[21] Peter J Huber. Robust estimation of a location parameter. In Breakthroughs in statistics, pages 492–518. Springer, 1992.

[22] Chen Wei, Wenjing Wang, Wenhao Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560, 2018.

[23] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In Proceedings of the IEEE International Conference on Computer Vision, pages 3155–3164, 2019.

[24] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Juyan Jia. Undere xposed photo enhancement using deep illumination estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6849–6857, 2019.

[25] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In 2012 19th IEEE International Conference on Image Processing, pages 965–968. IEEE, 2012.