Attention Based Low Light Image Enhancement

Mohammad Muhibur Rahman*, Anushua Ahmed, Ehsanur Rahman Rhythm, Md Sabbir Hossain and Annajiat Alim Rasel
Department of Computer Science and Engineering (CSE)
School of Data and Sciences (SDS)
Brac University
66 Mohakhali, Dhaka - 1212, Bangladesh
{mohammad.muhibur.rahman*, anushua.ahmed, ehsanur.rahman.rhythm, md.sabbir.hossain1}@g.bracu.ac.bd, annajiat@gmail.com

***Abstract*—Enhancing low-light images is challenging due to brightness and noise issues. Existing methods struggle with extremely low light and end up amplifying noise. In order to modify these previous approaches and bring more favorable results, this paper discusses attention-based residual recursive model to produce high-quality enhanced low-light images. One way is to use spatial and channel attention modules to reduce color fringing and noise. Specifically, the spatial attention module works on denoising by using non-local correlations in the images. At the same time, the channel attention module helps the network to refine redundant color features. An innovative pooling layer is also added that selects valuable information from previous features. This method involves a synthetic dataset for training, two attention maps for distinguishing underexposed areas and noise. Experimental results for both these approaches demonstrate significant advancements over existing methods quantitatively and visually.**

***Index Terms*—Low-Light Image Enhancement, Deep Neural Network, Synthetic Dataset, Image Denoising, Attention Mechanism**

I. Introduction

The amount of ambient light and the camera setting affect how bright the image is. Poor visibility, low contrast, and unexpected noise are typical image degradation that occur when images are recorded in an environment with insufficient irradiances. Unfortunately, we constantly see photographs like these, especially at night or indoors. Poor visual effects make it challenging to use such images as input for other visual tasks like target detection and recognition. Despite the fact that the auto-exposure system (such as ISO, shutter, flashlight, etc.) can correctly increase image brightness, it frequently results in other unanticipated abnormalities (such as noise, blurring, over-saturation, etc.). Therefore, in practical application, recovering normally exposed, high-quality photos from low-light images is crucial.

Several approaches [1]–[3] have been put out recently for recovering low-light photos, but there is still much space for improvement. These techniques, like [2], [3] failed in extremely low-light situations because they concentrated on boosting contrast and brightness while disregarding the impact of significant noise, which causes noise amplification. Even while the networks suggested by [4], [5] may concurrently boost brightness and produce high-quality images with processing noise, color artifacts still substantially impair visual equality.

We provide a novel residual recursive model based on an attention mechanism for processing low-light images in order to address the issue of noise amplification and color artifacts in earlier efforts. We found that a broader receptive field is essential for decreasing color artifacts in low-light photos because a wider range of data can instruct the network on how to behave when faced with significant noise. We build a new block, called a dual attention block, to efficiently fuse local and global characteristics in our network as opposed to just stacking residual layers to increase the receptive field [5]. The suggested dual attention block, which consists of modules for channel attention and spatial attention, successfully mutes unwanted chromatic aberration and noise. The network is directed to improve redundant color features via the channel attention module. Utilizing the non-local correlation in the image, the spatial attention module concentrates on denoising.

Fig. 1. Original Image vs MIRNet Enhanced

Our contributions generally fall into two categories:

- To produce normally exposed, high-quality, and noise-free photos, we suggest an end-to-end residual recursive network based on a dual attention block. Spatial attention and channel attention, which can consider both local and global information, are both included in this dual attention block.
- We test our technique on the LOL dataset, and the experimental findings show that it performs at a cutting-edge level.

II. Literature Review

A. Conventional Enhancement Methods

Conventional techniques can be divided generally into two classes. The histogram equalization (HE) method serves as the foundation for the first category. The differences between various HE-based approaches come from the unique extra assumptions and constraints they each use. Specifically, [6] attempts to dynamically maintain image brightness; [7] suggests analyzing and penalizing unnatural visual effects for better visual quality; [8] introduces and uses the differential gray-level histogram; [9] makes use of interpixel contextual information; and [10] focuses on the layered difference representation of 2D histogram in an effort to increase the gray-level differences between adjacent pixels. Instead of accounting for illumination, these methods widen the dynamic range and focus on improving the overall contrast of the image. They might exacerbate the issue with over- and under-enhancement. The second category is based on the Retinex hypothesis [11], which holds that illumination and reflection make up an image. The illumination map is typically recovered and used for low-light image enhancement using methods like MSR [12] and SSR [13]. The recent weighting approach proposed by AMSR [14] is based on SSR. To prevent over-enhancing, NPE [15] strikes a balance between the enhancement level and image naturalness. The illumination map is processed using MF [16] on multiple scales to boost local contrast while preserving naturalness. A weighted vibrational model is developed by SRIE [17] for estimating illumination maps. LIME [18] proposes a structure-aware smoothing model to estimate the illumination map. [19] suggests a dual-exposure fusion method, and [20] enhances it further by using the camera response model. For the purpose of enhancing low-light photos with intense noise, [21] offers a solid Retinex model taking into account the noise map. The estimation of the illumination map, which is made by hand and depends on careful parameter adjustment, is the key to these Retinex-based approaches. In addition, the majority of these Retinex-based techniques enhance the noise rather than attempting to reduce it.

B. Learning-Based Enhancement Methods

Deep learning has recently had considerable success with low-level image processing [22], modeling and comprehension of nighttime settings [23] [24], and low-level picture understanding [25] [26]. End-to-end networks and GANs, two potent techniques, have been applied to picture improvement [27]. The multilayer perception auto-encoder is used by LLNet [28] to improve and denoise low-light images. To approach the intended picture transformation, HDRNet [29] learns to make local, global, and content-dependent judgments. Since they are not end-to-end solutions, LLCNN [30] and [31] rely on some conventional techniques to handle brightness/contrast augmentation and denoising simultaneously. By utilizing several Gaussian convolution kernels, MSRNet [2] learns an end-to-end mapping between dark and bright images. A multi-branch network architecture is used in MBLLEN [32] to learn the mapping from low-light photos to ones with normal light. Only synthetic data acquired through straightforward data simulations is used to test the methodology. Our work likewise uses a multi-branch network architecture, similar to [33]. However, our uniqueness resides in techniques with more intricate designs, such the reinforce-net and two attention guidances. This research also introduces a new and improved low-light data generation methodology. In order to estimate the illumination map and improve the low-light photos by modifying the illumination map, [34] integrates the Retinex theory with CNN. [35] creates a comparable network by including a Restoration-Net for noise cancellation. For the improvement of low-light images, [36] presents a novel hybrid network with a content stream and a salient edge stream. Lv et al. [33] suggests a compact model with quick processing. In order to improve underexposed photos, DeepUPE [37] suggests a network that estimates an image-to-illumination mapping. However, the low-light noise is not taken into account. Additionally, DPED [38] [39] suggests an end-to-end method for converting poor-quality images taken with a cell phone into DSLR-grade photos using a composite perceptual error function. To cut down on computational costs, PPCN [40] creates a small network that combines teacher-student information transfer. The WESPE [41] method suggests a weakly-supervised approach to get around the limitations of needing paired images. By enhancing two-way GANs, Chen et al. [3] suggests an unpaired learning strategy for picture enhancement. Chen et al. [4] creates a CNN-based pipeline to immediately interpret raw sensor images for settings with very little light, SID. In order to create high-quality photos, Lv et al. [42] suggests decoupling the visible and near-infrared signals from a single image. Using [43], the near-infrared signal can likewise be acquired. Many of these learning-based techniques even rely on conventional denoising techniques when they do not expressly include them. Our method, however, takes the impacts of noise into account and employs two attention maps to direct the boosting and denoising processes. Our approach thus complements previous learning-based approaches.

C. Image Denoising

There are a ton of existing works for picture denoising. BM3D [44] and DnCNN [45] are examples of filter-based and deep-learning-based approaches for Gaussian denoising, respectively. NLPCA [46] uses a modified version of Principal Component Analysis to integrate dictionary learning components with sparse patch-based representations of pictures for Poisson denoising. Azzari and Foi [47] suggest combining an iterative approach using BM3D filter [44], variance-stabilizing transformation (VST), and the former. To eliminate Poisson noise, DenoiseNet [48] directly adds the calculated negative noise components to the original noisy image using a deep convolutional network. CBDNet [1] introduces a convolutional blind denoising network for Gaussian-Poisson mixed denoising by combining asymmetric learning. By practicing on both artificial and actual photos, it can be applied to real noisy images. Xu et al. [49] offers a trilateral weighted sparse coding technique for real-world picture denoising. Chen et al [3] suggests a two-step framework that includes CNN-based denoising and GAN-based noise distribution estimation. Blurring will arise when these procedures are combined directly with enhancing techniques. Our approach simultaneously conducts boosting and denoising to prevent this.

III. Methodology

The improvement of low-light images captured by the camera sensor is a challenging issue. White balance, demosaicking, denoising, and other subtasks make up the traditional Image Signal Process (ISP) approach. However, this process produces images with a high noise level and less vibrant color. We present a unique Attention-based Low-light image Enhancement Module that directly converts raw image to color image to address these issues.

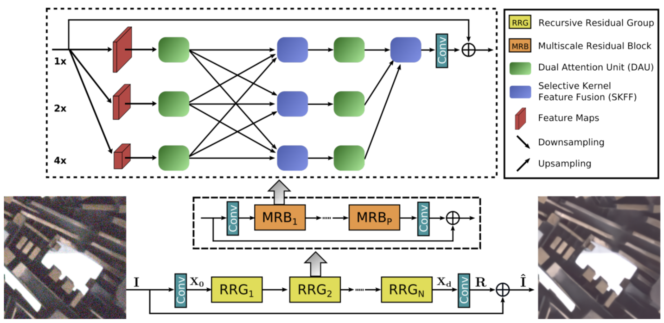


Fig. 2. MIRNet Architecture

A. MIRNet Architecture

- A feature extraction model that keeps the original high-resolution features in order to preserve fine spatial details while computing a complementary collection of features at various spatial scales.
- A frequently occurring process for information transmission where the characteristics from many branches with different resolutions are gradually combined for better representation learning.
- A fresh method for fusing features from different scales that correctly maintains the original feature information at each spatial level while dynamically combining varying receptive fields.
- A recursive residual design enables the building of very deep networks by gradually decomposing the input signal to streamline the overall learning process.

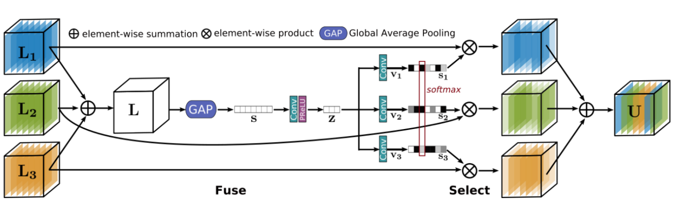


Fig. 3. Kernal Feature Fusion

B. Selective Kernel Feature Fusion

Two operations—Fuse and Select—perform the dynamic tuning of receptive fields in the Selective Kernel Feature Fusion or SKFF module. The information from multiple-resolution streams is combined by the Fuse operator to produce global feature descriptors. These descriptors are used by the Select operator to recalibrate the feature maps (of various streams) before their aggregation.

- Fuse: Three parallel convolution streams conveying various informational scales provide input to the SKFF. We first integrate these multi-scale features using an element-wise sum, and then across the spatial dimension, we use Global Average Pooling (GAP). To create a compact feature representation, we then apply a channel-downscaling convolution layer. This layer then passes through three concurrent channel-upscaling convolution layers (one for each resolution stream), giving us three feature descriptors.
- Select: This operator generates the corresponding activations for the multi-scale feature maps by applying the softmax function to the feature descriptors. The product of the relevant multi-scale feature and the feature descriptor are summed up to determine the aggregated features.

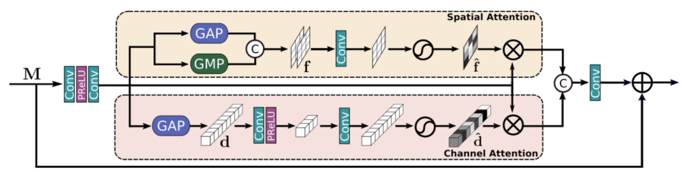


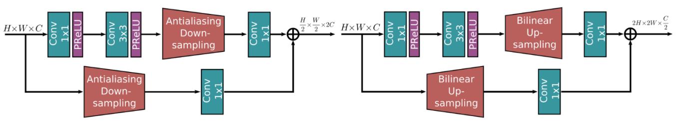
Fig. 4. Dual Attention Block

C. Dual Attention Unit

Features in the convolutional streams are extracted using the Dual Attention Unit, or DAU. The DAU block provides the means to communicate information within a feature tensor along the spatial and channel dimensions while the SKFF block fuses information across multi-resolution branches. Only more informative elements are allowed to continue after being suppressed by the DAU. Utilizing the Channel Attention and Spatial Attention methods, this feature calibration is accomplished.

- Channel Attention: By using squeeze and excitation operations, this branch takes use of the convolutional feature maps' inter-channel interactions. The squeeze operation takes as input a feature map, performs Global Average Pooling across spatial dimensions to encode global context, and produces a feature descriptor as a result. The excitation operator passes this feature descriptor through two convolutional layers followed by the sigmoid gating and generates activations. Rescaling the input feature map using the output activations results in the output of the Channel Attention branch.

- Spatial Attention: Utilizing the inter-spatial interdependence of convolutional features is the goal of the spatial attention branch. Specifically, creating a spatial attention map and using it to recalibrate the incoming features is the aim of this branch. The Spatial Attention branch first independently performs Global Average Pooling and Max Pooling operations on input features along the channel dimensions, concatenates the outputs to form a resultant feature map, and then obtains the spatial attention map by passing the resultant feature map through convolution and sigmoid activation. The input feature map is then scaled using this spatial attention map.



(a) Downsampling module (b) Upsampling module

Fig. 5. Multi-Scale Residual Architecture

D. Multi-Scale Residual Block

The Multi-Scale Residual Block may maintain high-resolution representations and produce an output that is spatially exact while also getting rich contextual data from low-resolutions. The MRB is composed of several (in this paper, three) fully-convolutional streams that are coupled in parallel. In order to aggregate high-resolution features with the aid of low-resolution features and vice versa, it enables information interchange between parallel streams. A recursive residual design (with skip connections) is used by the MIRNet to facilitate information flow during the learning process. Residual resizing modules are utilized to carry out the downsampling and upsampling operations that are employed in the Multi-scale Residual Block in order to preserve the residual nature of our design.

IV. Experiment

A. Dataset and Evaluation Metrics

We adopt the Low-Light dataset (LoL) [cite] to evaluate the performance of our method. The LoL dataset provides 500 image pairs. A low-light input image and its associated well-exposed reference image make up each pair of images in the dataset. From the LoL Dataset's available samples, we select 300 of these image pairs for training, 185 for validation, and the remaining 15 for testing. A batch size of 4 is used during the training process and PSNR is employed as a metric to assess the network's capacity for improving low-light images. The PSNR metric is an expression for the ratio of a signal's maximum achievable strength to the power of distorted noise that impairs the accuracy of its representation.

B. Training and Testing

We used TensorFlow to create our network, and the LoL dataset was used to train it over 50 epochs. We create random 128×128 size crops from the image pairings to be utilized for both training and validation. The Adam optimizer is applied during training and the starting learning rate is set to 0.0001.

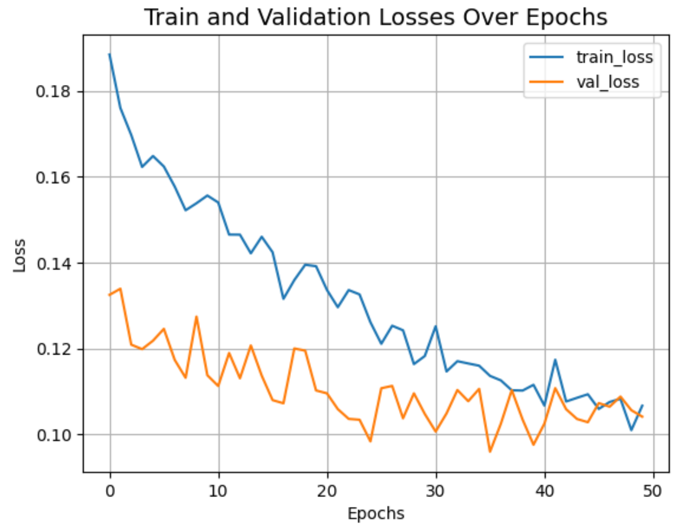


Fig. 6. Loss Curve

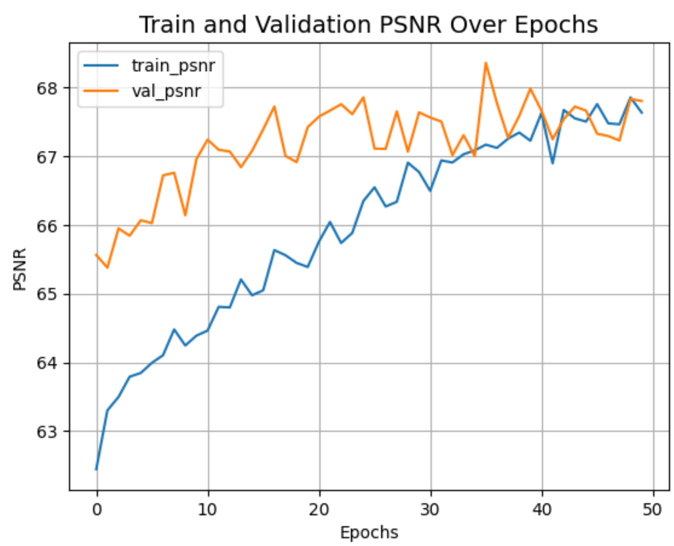


Fig. 7. Peak Signal to Noise Ratio

C. Results Analysis

When comparing the outcomes, it is clear that our network's improved images have noticeably high quality. First, even from low-light photographs with significant noise, our approach can recover a significant amount of features and textures. Second, by restoring accurate, natural color and preventing color spreading, our method can improve the realism and reality of the improved photographs. We use PSNR to compare the effectiveness of our method quantitatively.

V. Conclusion and Future Work

We suggest an attention-based architecture in the present study to improve the raw images and produce color images

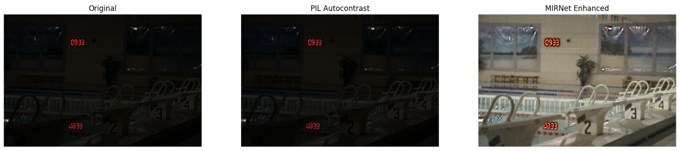


Fig. 8. Original vs Enhanced Image

with high contrast and minimal noise. In order to extract features and increase the network's performance, our strategy employs the mixed attention block, which blends spatial and channel attention. Experiments show that our approach produces superior images with reduced noise and color artifacts, scoring best on the LOL dataset. We will look into a more efficient attention module in the future to reduce processing power and enhance network generalization capability.

References

[1] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 1712–1722.

[2] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net:low-light image enhancement using deep convolutional network," 2017.

[3] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3155–3164.

[4] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," 2018.

[5] P. Maharjan, L. Li, Z. Li, N. Xu, C. Ma, and Y. Li, "Improving extreme low-light image denoising via residual learning," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, 2019, pp. 916–921.

[6] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1752–1758, 2007.

[7] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 1921–1935, 2009.

[8] K. Nakai, Y. Hoshi, and A. Taguchi, "Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms," in *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, 2013, pp. 445–449.

[9] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3431–3441, 2011.

[10] 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.

[11] [Online]. Available: https://www.scientificamerican.com/article/the-retinex-theory-of-color-vision/

[12] D. Jobson, Z. Rahman, and G. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.

[13] ——, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–462, 1997.

[14] C.-H. Lee, J.-L. Shih, C.-C. Lien, and C.-C. Han, "Adaptive multiscale retinex for image contrast enhancement," in *2013 International Conference on Signal-Image Technology & Internet-Based Systems*, 2013, pp. 43–50.

[15] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[16] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168416300949

[17] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2782–2790.

[18] 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, 2017.

[19] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," 2017.

[20] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new low-light image enhancement algorithm using camera response model," 10 2017.

[21] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.

[22] V. Sharma, A. Diba, D. Neven, M. S. Brown, L. Van Gool, and R. Stiefelhagen, "Classification-driven dynamic image enhancement," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4033–4041.

[23] F. Radenovic, J. L. Schönberger, D. Ji, J.-M. Frahm, O. Chum, and J. Matas, "From dusk till dawn: Modeling in the dark," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5488–5496.

[24] Y. Zheng, M. Zhang, and F. Lu, "Optical flow in the dark," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 6748–6756.

[25] D. Dai and L. V. Gool, "Dark model adaptation: Semantic image segmentation from daytime to nighttime," 2018.

[26] C. Sakaridis, D. Dai, and L. V. Gool, "Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation," 2019.

[27] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[28] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003132031630125X

[29] M. Gharbi, J. Chen, J. Barron, S. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Transactions on Graphics*, vol. 36, 07 2017.

[30] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie, "Llcnn: A convolutional neural network for low-light image enhancement," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, 2017, pp. 1–4.

[31] L. Tao, C. Zhu, J. Song, T. Lu, H. Jia, and X. Xie, "Low-light image enhancement using cnn and bright channel prior," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3215–3219.

[32] F. Lv, F. Lu, J. Wu, and C. Lim, "Mbllen: Low-light image/video enhancement using cnns." in *BMVC*, vol. 220, no. 1, 2018, p. 4.

[33] F. Lv, B. Liu, and F. Lu, "Fast enhancement for non-uniform illumination images using light-weight cnns," 2020.

[34] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," 2018.

[35] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," 2019.

[36] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, 2019.

[37] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6842–6850.

[38] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. V. Gool, "Dslr-quality photos on mobile devices with deep convolutional networks," 2017.

[39] E. de Stoutz, A. Ignatov, N. Kobyshev, R. Timofte, and L. V. Gool, "Fast perceptual image enhancement," 2018.

[40] Z. Hui, X. Wang, L. Deng, and X. Gao, "Perception-preserving convolutional networks for image enhancement on smartphones," in *Proceedings*

of the European Conference on Computer Vision (ECCV) Workshops, 2018, pp. 0–0.

[41] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. V. Gool, "Wespe: Weakly supervised photo enhancer for digital cameras," 2018.

[42] F. Lv, Y. Zheng, Y. Li, and F. Lu, "An integrated enhancement solution for 24-hour colorful imaging," 2020.

[43] F. Lv, Y. Zheng, B. Zhang, and F. Lu, "Turn a silicon camera into an ingaas camera," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5980–5988.

[44] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3d filtering," in *Image processing: algorithms and systems, neural networks, and machine learning*, vol. 6064. SPIE, 2006, pp. 354–365.

[45] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[46] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local pca," *Journal of mathematical imaging and vision*, vol. 48, pp. 279–294, 2014.

[47] L. Azzari and A. Foi, "Variance stabilization for noisy+ estimate combination in iterative poisson denoising," *IEEE signal processing letters*, vol. 23, no. 8, pp. 1086–1090, 2016.

[48] T. Remez, O. Litany, R. Giryes, and A. M. Bronstein, "Deep convolutional denoising of low-light images," 2017.

[49] J. Xu, L. Zhang, and D. Zhang, "A trilateral weighted sparse coding scheme for real-world image denoising," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 20–36.