

# A Lightweight CNN for Detail Enhancement and Color Correction of Low-Light Capsule Endoscopy

Shuocheng Wang<sup>1</sup>, Jiaming Liu<sup>1</sup>, Ruoxi Zhu<sup>1</sup>, Jiazheng Lian<sup>1</sup>, Chengkang Huang<sup>1</sup>, Hu Wei<sup>2</sup>, Yibo Fan<sup>1\*</sup>

<sup>1</sup>State Key Laboratory of Integrated Chips and Systems, Fudan University, Shanghai, China

<sup>2</sup>Molchip Co.,ltd, Shanghai, China

**Abstract**—Wireless capsule endoscopy (WCE) is a non-invasive medical procedure that involves swallowing a small capsule equipped with a camera to capture images of the digestive tract. However, limitations in gastrointestinal structures and equipment performance may result in deficient illumination, thereby affecting diagnostic accuracy. In recent years, deep learning (DL) has been increasingly applied in the field of low-light medical image enhancement. However, current DL methods often face challenges such as limited sensitivity to fine structures and color distortion. In this paper, we propose a novel method for addressing insufficient light in endoscopic images based on a residual neural network. We introduce a transform kernel (TFK) convolution to extract details from feature maps, making the recovered image more similar to real ones. Additionally, we propose a learnable image-guided enhancement block (IGEB), enhancing the input image in luminance and chroma distinctively at different stages of the network to improve the quality of output. The proposed approach significantly outperforms previous low-light image enhancement (LLIE) algorithms both qualitatively and quantitatively, even with much fewer parameters. Furthermore, the restored images achieve better performance in some high-level tasks such as image segmentation.

**Index Terms**—wireless capsule endoscopy, low-light image enhancement, deep learning

## I. INTRODUCTION

Wireless capsule endoscopy (WCE) has emerged as a revolutionary technology in the field of gastrointestinal diagnostics, offering significant advantages over traditional endoscopic procedures [1]. WCE involves the ingestion of a small capsule equipped with a camera. This allows for non-invasive imaging of the entire gastrointestinal tract, particularly the small intestine, which was previously challenging to visualize. However, the complexity of gastrointestinal structures, coupled with environmental low-light conditions and limitations in camera performance, may result in underexposed images. Inadequate exposure in captured images brings challenges for both manual and automated diagnosis, as early lesions and abnormalities may be difficult to detect under low-light conditions [2]. Therefore, there is a growing need for advancements in low-light medical image enhancement.

Numerous conventional techniques have been developed for the enhancement of low-light images such as histogram equalization [4] and Retinex theory [5]. For WCE, Long *et al.*

This work was supported in part by the National Key R&D Program of China (2023YFB4502802), in part by the National Natural Science Foundation of China (62031009), in part by the “Ling Yan” Program for Tackling Key Problems in Zhejiang Province (No.2022C01098), in part by Alibaba Innovative Research (AIR) Program, in part by Alibaba Research Fellow (ARF) Program, in part by the Fudan-ZTE Joint Lab, in part by CCF-Alibaba Innovative Research Fund For Young Scholars.

\*Yibo Fan is the corresponding author.

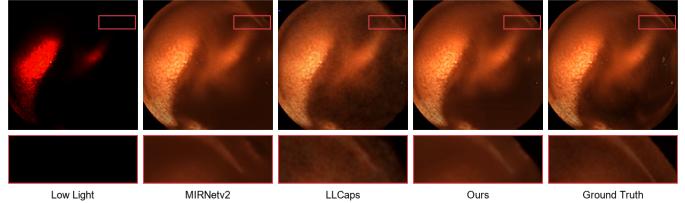


Fig. 1. Comparison with two state-of-the-art approaches on the RLE [3] dataset, which reveals that our model has successfully restored color fidelity and visually appealing content.

[6] proposed a method called guide image based enhancement to enhance low quality images by using the information of a high quality image of a similar scene. [7] proposed an adaptive fraction gamma transformation to enhance the quality of WCE images. However, traditional methods usually rely on specific prior knowledge or lack generality across different scenarios. In recent years, deep learning has been widely applied in the field of low-light image enhancement (LLIE) [8]–[10]. MIRNetv2 [10] introduced an architecture aimed at preserving high-resolution, spatially-accurate representations throughout the network while also integrating contextual information from low-resolution inputs. Several deep learning LLIE strategies have been proposed for medical endoscopy applications [11], [12]. Yue *et al.* [11] proposed an unsupervised deep learning scheme based on the Cycle Generative Adversarial Network. Long *et al.* [12] presented LLCaps, a framework that leverages the multi-scale convolutional neural network (CNN) and reverse diffusion process.

However, many deep learning methods encounter challenges such as insensitivity to fine structures and details in images, as well as color distortion, which can hinder their effectiveness in low-light enhancement (shown in Fig. 1). In this paper, we propose an efficient CNN model (shown in Fig. 2) that delivers state-of-the-art (SOTA) performance on low-light WCE images. Our main contributions of this paper can be summarized as follows:

- We propose a low-light enhancement network tailored for WCE images. We introduce a trainable image-guided enhancement block (IGEB) designed to enhance the input image separately in luminance and chroma, effectively extracting features from the input and improve the quality of the output images.
- We present a transform convolution residual block (TCRB) in the network, leveraging a transform kernel

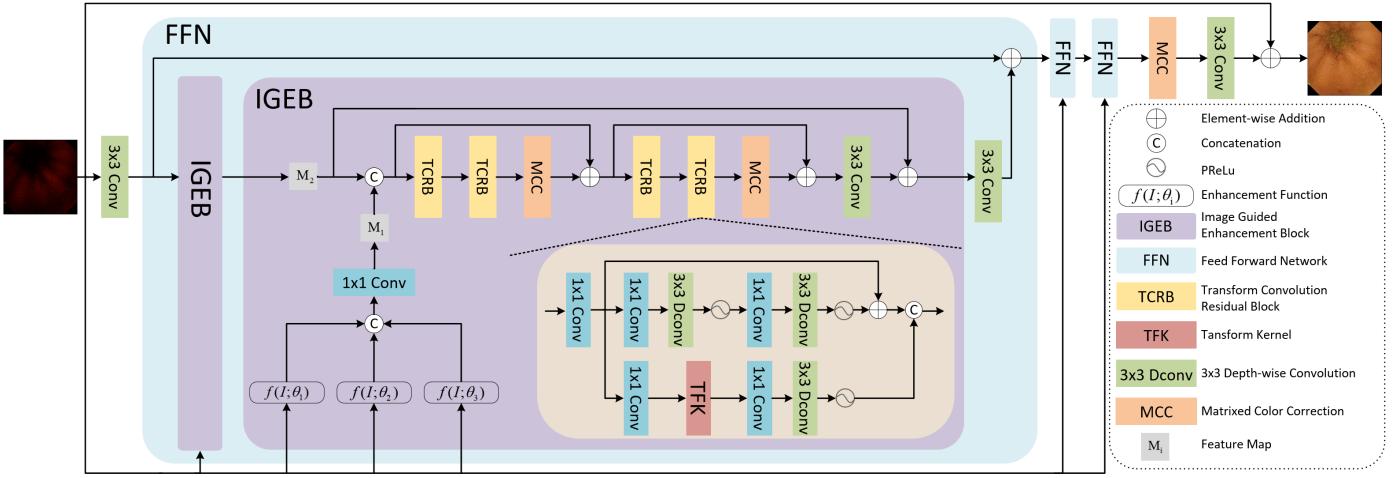


Fig. 2. Overview of our proposed model. Our model consists primarily of three feed forward networks (FFNs), each FFN containing two image guided enhancement blocks (IGEBs). Each IGEB contains three independent image guided enhancement functions (IGEFs) and four transform convolution residual blocks (TCRBs).

(TFK) convolution to better extract detailed features from the feature map.

- Extensive experiments on two public datasets demonstrate that our method achieves SOTA image qualities both objectively and perceptually with fewer parameters compared to existing algorithms. Moreover, our approach demonstrates improved performance in certain downstream tasks such as image segmentation.

## II. PROPOSED METHOD

Since many deep learning approaches encounter difficulties such as insensitivity to details and color distortion, our overall network design takes these factors into account and incorporates image guided enhancement blocks (IGEBs) and transform convolution residual blocks (TCRBs) to address them. The framework of our algorithm is illustrated in Fig. 2. Our model mainly consists of three feed forward networks (FFNs), each containing two IGEBs. And each IGEB comprises three independent image guided enhancement functions (IGEFs) and four TCRBs. What's more, we integrate the MCC module [9] to rectify color distortion.

### A. Image Guided Enhancement Block

As illustrated in Fig. 2, the original image is fed to each layer of the network as guidance to preserve the rich detail information in the original image. However, low-light images often face the problem of underexposure and color distortion, which poses a challenge to feature extraction. Therefore, we propose to apply image enhancement functions before extracting the features of the input. This ensures that the network retains important features and characteristics from the input image throughout the processing stages. Using only one enhancement function may limit the network to learn a set of globally optimal function parameters. Since we need to consider the local details of the input image, we incorporate three independent enhancement functions within each IGEB

and fuse these enhanced images to achieve the best visual result.

As depicted in IGEB in Fig. 2, the input image passes through three independent enhancement functions and the output images are concatenated together. Then we employ a  $1 \times 1$  convolution and concatenate the result ( $M_1$  in Fig. 2) with the output of the previous block ( $M_2$  in Fig. 2) as the final input of IGEB. This operation enables the IGEB to combine the feature extracted by the previous block and the abundant information of the input image enhanced by the enhancement functions. The whole process can be described as the following equation:

$$M_1 = O_C(Concat(f(I; \theta_1), f(I; \theta_2), f(I; \theta_3))) \quad (1)$$

$$IGEB_{input} = Concat(M_2, M_1) \quad (2)$$

where  $f$  denotes the image guided enhancement function (IGEF).  $\theta_i (i = 1, 2, 3)$ , including  $\lambda_{ij} (j = 1, 2, 3)$  and  $\epsilon$  in Eq. (3), denotes the parameters in  $f$ .  $I$  denotes the input image and  $O_C$  means the convolution operation.

Considering the different characteristics of luma and chroma components, we design distinct enhancement algorithm for each component. For the luminance component, we employ the logarithmic curve [22] to address the issue of darkness. By incorporating learnable parameters, the logarithmic curve can effectively amplify contrast in low-luminance regions while preserving detail in high-luminance areas attributed to its non-linear contrast enhancement capabilities. Regarding the chroma component, we introduce a learnable parameter to mitigate the deficiencies in chroma. The whole algorithm can be written in the following form:

$$I_e = f(I; \theta_i) = \begin{cases} Y_e = \log(\lambda_{i1} \cdot \frac{Y}{\max(Y)} + \epsilon) / \log(\lambda_{i1} + \epsilon) \\ U_e = U \cdot \lambda_{i2} \cdot 2 \\ V_e = V \cdot \lambda_{i3} \cdot 2 \end{cases} \quad (3)$$

TABLE I  
QUANTITATIVE COMPARISON OF DIFFERENT METHODS ON KVASIR-CAPSULE [13] AND RLE [3] DATASET.

Models	Kvasir-Capsule				RLE				Parameter	RLE Segmentation	
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	CIEDE2000 $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	CIEDE2000 $\downarrow$		mIoU $\uparrow$	Dice $\uparrow$
DUAL [14]	11.61	0.2901	0.4532	20.667	14.64	0.1611	0.4903	17.245	-	61.89	78.15
Zero-DCE [15]	14.03	0.4631	0.4917	13.749	14.86	0.3418	0.4519	12.973	<b>0.0794</b> M	54.77	71.46
EnlightenGAN [16]	27.15	0.8503	0.1769	5.202	23.65	0.8051	0.1864	6.079	54.42 M	61.97	74.15
LLFlow [17]	29.69	0.9257	0.0774	4.641	25.93	0.8519	0.1340	5.091	-	61.06	<b>78.55</b>
HWMNet [18]	27.62	0.9209	0.1507	4.231	21.81	0.7611	0.3624	7.641	66.56 M	56.48	74.17
MIRNet [19]	31.23	0.9577	0.0436	3.446	25.77	0.8694	0.1519	5.169	31.79 M	59.84	78.32
StillGAN [20]	28.28	0.9130	0.1302	4.565	26.38	0.8333	0.1860	6.716	78.69 M	58.32	71.56
SNR-Aware [21]	30.32	0.9492	0.0521	-	27.73	0.8844	0.1094	-	40.08 M	58.95	70.26
MIRNetv2 [10]	31.67	0.9522	0.0486	2.150	32.85	0.9269	0.0781	<b>2.514</b>	5.86 M	63.14	75.07
LLCaps [12]	<b>35.24</b>	<b>0.9634</b>	<b>0.0374</b>	<b>1.254</b>	<b>33.18</b>	<b>0.9334</b>	<b>0.0721</b>	2.598	119.72 M	<b>66.47</b>	78.47
<b>Ours</b>	<b>39.17</b>	<b>0.9792</b>	<b>0.0217</b>	<b>1.118</b>	<b>33.32</b>	<b>0.9304</b>	<b>0.0729</b>	<b>2.281</b>	<b>0.87</b> M	<b>65.50</b>	<b>78.87</b>

The best and the second best results are highlighted in bold and blue respectively.

where  $I_e$  denotes the result of the enhancement function.  $Y_e$ ,  $U_e$ ,  $V_e$  and  $Y$ ,  $U$ ,  $V$  denote the YUV components of  $I_e$  and  $I$  respectively.  $\lambda_{ij}$  ( $j = 1, 2, 3$ ) and  $\epsilon$  are learnable parameters in IGEF during training.

### B. Transform Convolution Residual Block

Medical images should retain rich image details for diagnostic accuracy. Thus we design a transform kernel (TFK) in transform convolution residual block (TCRB) to extract high frequency details of the feature map. The TCRB's architecture is shown in Fig. 2 and is separated into two parts: a spatial convolution (SPC) branch on the top and a TFK branch on the bottom.

Firstly we send the input feature into a convolution layer,

$$X_c = O_{CT}(X) \quad (4)$$

where  $X$  is the input of the block and  $X_c$  is the output of convolution. Then we feed  $X_c$  into two different branches.  $X_s$  is the outcome of SPC branch and  $X_{tf}$  is the outcome of TFK branch, which is proposed to contain high frequency information of  $X_c$ ,

$$X_s = O_{spatial}(X_c) \quad (5)$$

$$X_{tf} = O_{transform}(X_c) \quad (6)$$

where  $O_{spatial}$  represents the SPC branch and  $O_{transform}$  is the TFK branch. The  $X_s$  can also be written as

$$X_s = O_{CDP}(O_{CDP}(X_c)) + X_c \quad (7)$$

where  $O_{CDP}$  denotes an operation which contains a  $1 \times 1$  convolution, a  $3 \times 3$  depth-wise convolution and a  $PReLU$  non-linear activation.

In the TFK branch, we choose Laplacian kernel as the transform kernel, since Laplacian kernel possesses a strong proficiency in extracting the detailed texture from the feature map. This capability aids the TCRB module in acquiring a deeper understanding of the feature map's high-frequency characteristics. Leveraging such high-frequency information enables our network to capture finer details and make output

image more distinct and vivid. The original form of Laplacian kernel

$$K_{Laplacian} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (8)$$

is a 2D kernel and we replicate the kernel parameter in each channel to acquire the shape  $K_{Transform} \in \mathbb{R}^{C \times W \times H}$ . The  $X_{tf}$  can also be written as

$$X_{tf} = O_{CDP}(O_{CT}(X_c)) \quad (9)$$

where  $O_{CT}$  means an operation with a  $1 \times 1$  convolution and a transform kernel convolution. Finally, we concatenate  $X_s$  and  $X_{tf}$  together as the output of TCRB.

$$TCRB_{output} = Concat(X_s, X_{tf}) \quad (10)$$

## III. EXPERIMENTS

### A. Implementation Details

We conduct abundant experiments on two public WCE datasets using in LLCaps [12]: the Kvasir-Capsule dataset (KCL) [13] and the Red Lesion Endoscopy (RLE) dataset [3]. The KCL dataset contains 2000 training pairs and 400 evaluation pairs. The RLE dataset has 946 training pairs and 334 evaluation pairs. We combine the two datasets together to train the model. Our model is trained using Adam optimizer for 150 epochs with a batch size of 6 and a learning rate of  $5 \times 10^{-3}$ . We adopt Charbonnier loss as loss function:

$$L_{Charbonnier} = \sqrt{\|I_{output} - I_{gt}\|^2 + \varepsilon^2} \quad (11)$$

where  $I_{output}$  and  $I_{gt}$  represent the output of the model and the ground truth image respectively, and  $\varepsilon$  is set as  $1 \times 10^{-3}$ .

For quantitative assessment, we utilize three widely employed metrics for image quality evaluation: PSNR, SSIM [23], and Learned Perceptual Image Patch Similarity (LPIPS) [24]. We additionally employ the metric CIEDE2000 [25] to validate the accuracy of color correction. To assess the effectiveness of the LLIE methods in downstream medical applications, we perform segmentation of red lesions on the

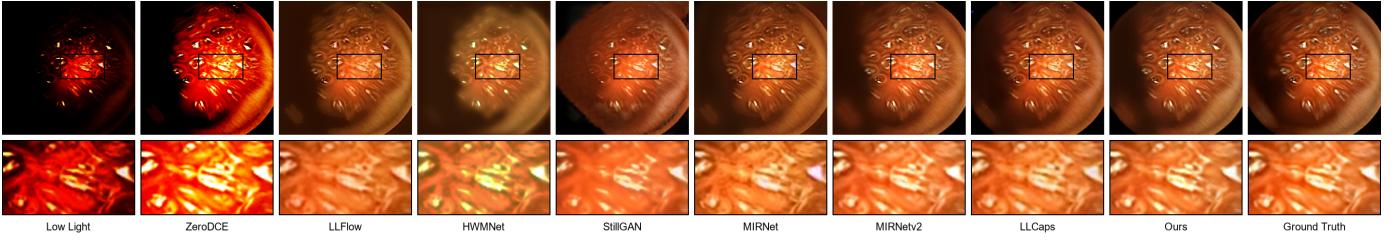


Fig. 3. Visual comparisons on RLE dataset.

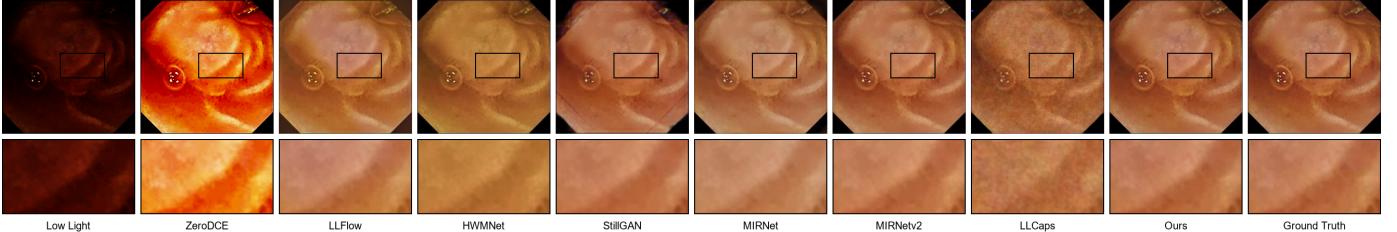


Fig. 4. Visual comparisons on Kvasir-Capsule dataset.

RLE test dataset and assess performance using mean Intersection over Union (mIoU) and Dice similarity coefficient (Dice). The segmentation network is a UNet [26] pre-trained on the RLE dataset using the Adam optimizer for 20 epochs, with a batch size of 4 and a learning rate set to  $1 \times 10^{-4}$ .

#### B. Performance Comparisons

From Table I, it is evident that our approach yields favorable outcomes compared to other methodologies. Specifically, our method surpasses LLCaps [12] by 3.93 dB and 0.14 dB in terms of PSNR on KCL and RLE dataset respectively. Additionally, our method outperforms LLCaps [12] by 0.136 and 0.317 on these two datasets in CIEDE2000, indicating significant advancements in color correction. Furthermore, the parameter count of our model is merely 0.87M, considerably fewer than numerous existing SOTA models. The qualitative outcomes are presented in Fig. 3 and Fig. 4. It is obvious that the outcomes generated by Zero-DCE [15], LLFlow [17], HWMNet [18] and StillGAN [20] encounter significant degradation in both color and structure. From Fig. 3, we can observe that the images restored by MIRNet [19] and MIRNetv2 [10] exhibit errors in the corners. LLCaps [12] may lead to some loss of details and noise in the final output (obviously in Fig. 4) compared to ours.

Moreover, we performed a downstream task focusing on red lesion segmentation to evaluate the practical utility of our model in clinical settings. As depicted in Table I, our model achieves superior performance compared to all SOTA methods in Dice, indicating that the images generated by our model exhibit more intricate structures and sharper edges.

#### C. Ablation Study

To evaluate the effectiveness of our proposed contributions respectively, we train four versions of the model: (i) without both image guided enhancement function (IGEF) and transform kernel convolution (TFK), (ii) only with TFK, (iii)

TABLE II  
RESULTS OF ABLATION STUDIES ON KVASIR-CAPSULE DATASET [13].

TFK	IGEF	Kvasir-Capsule			
		PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	CIEDE2000 $\downarrow$
✗	✗	38.42	0.9541	0.0343	1.249
✓	✗	38.88	0.9682	0.0277	1.178
✗	✓	38.56	0.9680	0.0280	1.209
✓	✓	<b>39.17</b>	<b>0.9792</b>	<b>0.0217</b>	<b>1.118</b>

only with IGEF, and (iv) with both. We compute the PSNR, SSIM, LPIPS and CIEDE2000 on the Kvasir-Capsule dataset and compare the results quantitatively. Experimental results are shown in Table II, from which we can clearly see that without TFK, the network fails to adequately extract detailed information from the feature maps, resulting in restored images lacking the richness of original details and the loss of the four metrics. The absence of IGEF at various stages of the network results in the loss of input image information in deeper layers, compromising the restoration efficacy of low-light images.

## IV. CONCLUSION

In this paper, we present an end-to-end low-light WCE image enhancement network. A transform convolution residual block is proposed to extract high frequency components from the feature maps. We also design a learnable image enhancement function for the input and employ the enhanced images to help the network learn more details and color information. Experimental results demonstrate that our method achieves SOTA performance with much fewer parameters compared to existing deep learning models, signifying the potential of applying in real WCE applications.

## REFERENCES

- [1] Levin J Slier and Gastone Ciuti, "Flexible and capsule endoscopy for screening, diagnosis and treatment," *Expert review of medical devices*, vol. 11, no. 6, pp. 649–666, 2014.
- [2] Mingzhu Long, Zhuo Li, Xiang Xie, Guolin Li, and Zhihua Wang, "Adaptive image enhancement based on guide image and fraction-power transformation for wireless capsule endoscopy," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 5, pp. 993–1003, 2018.
- [3] Paulo Coelho, Ana Pereira, Argentina Leite, Marta Salgado, and Antônio Cunha, "A deep learning approach for red lesions detection in video capsule endoscopies," in *Image Analysis and Recognition*, Aurélio Campilho, Fakhri Karray, and Bart ter Haar Romeny, Eds., Cham, 2018, pp. 553–561, Springer International Publishing.
- [4] Yun-Fu Liu, Jing-Ming Guo, and Jie-Cyun Yu, "Contrast enhancement using stratified parametric-oriented histogram equalization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 6, pp. 1171–1181, 2017.
- [5] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.
- [6] Mingzhu Long, Zhuo Li, Yuchi Zhang, Xiang Xie, Guolin Li, Shigang Yue, and Zhihua Wang, "Guide image based enhancement method for wireless capsule endoscopy," in *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2017, pp. 1–4.
- [7] Mingzhu Long, Zehua Lan, Xiang Xie, Guolin Li, and Zhihua Wang, "Image enhancement method based on adaptive fraction gamma transformation and color restoration for wireless capsule endoscopy," in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2018, pp. 1–4.
- [8] Kai-Fu Yang, Cheng Cheng, Shi-Xuan Zhao, Hong-Mei Yan, Xian-Shi Zhang, and Yong-Jie Li, "Learning to adapt to light," *International Journal of Computer Vision*, vol. 131, no. 4, pp. 1022–1041, Apr 2023.
- [9] Xin Jin, Ling-Hao Han, Zhen Li, Chun-Le Guo, Zhi Chai, and Chongyi Li, "Dnf: Decouple and feedback network for seeing in the dark," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 18135–18144.
- [10] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, "Learning enriched features for fast image restoration and enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1934–1948, 2023.
- [11] Guanghui Yue, Jie Gao, Lvyin Duan, Jingfeng Du, Weiqing Yan, Shuigen Wang, and Tianfu Wang, "Colorectal endoscopic image enhancement via unsupervised deep learning," *Multimedia Tools and Applications*, Jul 2023.
- [12] Long Bai, Tong Chen, Yanan Wu, An Wang, Mobarakol Islam, and Hongliang Ren, "Llcaps: Learning to illuminate low-light capsule endoscopy with curved wavelet attention and reverse diffusion," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, Hayit Greenspan, Anant Madabhushi, Parvin Mousavi, Septimiu Salcudean, James Duncan, Tanveer Syeda-Mahmood, and Russell Taylor, Eds., Cham, 2023, pp. 34–44, Springer Nature Switzerland.
- [13] Pia H. Smedsrød, Vajira Thambawita, Steven A. Hicks, Henrik Gjestang, Oda Olsen Nedrejord, Espen Næss, Hanna Borgli, Debesh Jha, Tor Jan Derek Berstad, Sigrun L. Eskeland, Mathias Lux, Håvard Espeland, Andreas Petlund, Duc Tien Dang Nguyen, Enrique Garcia-Ceja, Dag Johansen, Peter T. Schmidt, Ervin Toth, Hugo L. Hammer, Thomas de Lange, Michael A. Riegler, and Pål Halvorsen, "Kvasir-capsule, a video capsule endoscopy dataset," *Scientific Data*, vol. 8, no. 1, pp. 142, May 2021.
- [14] Qing Zhang, Yongwei Nie, and Weishi Zheng, "Dual illumination estimation for robust exposure correction," *Computer Graphics Forum*, vol. 38, 2019.
- [15] Chunlu Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1777–1786.
- [16] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.
- [17] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex C Kot, "Low-light image enhancement with normalizing flow," *arXiv preprint arXiv:2109.05923*, 2021.
- [18] Chi-Mao Fan, Tsung-Jung Liu, and Kuan-Hsien Liu, "Half wavelet attention on m-net+ for low-light image enhancement," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 3878–3882.
- [19] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, "Learning enriched features for real image restoration and enhancement," in *Computer Vision – ECCV 2020*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, Eds., Cham, 2020, pp. 492–511, Springer International Publishing.
- [20] Yuhui Ma, Jiang Liu, Yonghui Liu, Huazhu Fu, Yan Hu, Jun Cheng, Hong Qi, Yufei Wu, Jiong Zhang, and Yitian Zhao, "Structure and illumination constrained gan for medical image enhancement," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3955–3967, 2021.
- [21] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia, "Snr-aware low-light image enhancement," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 17693–17703.
- [22] Yael Vinker, Inbar Huberman-Spiegelglas, and Raanan Fattal, "Unpaired learning for high dynamic range image tone mapping," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 14637–14646.
- [23] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [24] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595.
- [25] Gaurav Sharma, Wencheng Wu, and Edul Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research Application*, vol. 30, pp. 21 – 30, 02 2005.
- [26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, Eds., Cham, 2015, pp. 234–241, Springer International Publishing.