ACE-Flow: Auto Color Encoding for Enhanced Low-Light Image Restoration

Jiachen Qiu, Yushen Zuo and Kin-Man Lam
Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong
E-mail: jiachen.qiu@connect.polyu.hk, yushen.zuo@polyu.edu.hk, enkmlam@polyu.edu.hk

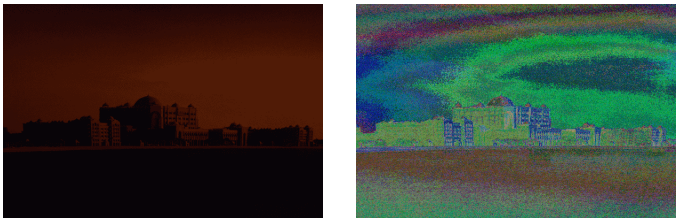
***Abstract*—Low-light image enhancement is a classic problem in low-level vision tasks, aiming to improve the quality of images captured in poor-lighting scenarios. Conventional deep enhancement models often produce distorted content (e.g., deviated lighting conditions, color bias) in extremely dark regions because they fail to capture comprehensive color information during the reconstruction process. To address these issues, we propose a novel normalizing flow-based model that incorporates an auto-color encoding method, called ACE-Flow, for low-light image enhancement. By leveraging auto-color encoding, our method can encode color information during feature extraction and effectively restore the corrupted image content in challenging regions. Furthermore, our approach can accurately learn the mapping from low-light images to high-quality ground-truth images, because the invertibility property of the normalizing flow implicitly regularizes the learning process. Experiments demonstrate that our method significantly outperforms other promising low-light enhancement models in terms of reconstruction and perceptual metrics. Additionally, the enhanced images produced by our model exhibit rich details with minimal distortion, resulting in superior visual quality.**

I. Introduction

Low-light image enhancement is an attractive topic in low-level vision tasks, with the aim of jointly improving luminance and removing undesirable noise caused by sensors and dim environments. Low-light image enhancement techniques have significant industrial values, including applications in modern imaging devices, surveillance, and autonomous driving, attracting considerable researcher attention.

In the past years, many deep learning-based low-light image enhancement models have been proposed and have achieved promising results. However, the methods in [1]–[4] often produce unacceptable artifacts when processing real-world Ultra-High-Definition images [5]. Wu *et al.* [6] and Xu *et al.* [7] proposed methods to simultaneously enhance luminance and remove noise in the spatial domain, resulting in degraded performance when the images are captured in challenging lighting conditions. Despite the promising results shown in [8], the heavy network structures highly limit its applications in real-world scenarios. All the mentioned methods largely ignore the profound color information in the reconstruction process, which inevitably leads to suboptimal solutions. In addition, after training, these methods cannot be generalized to other kinds of low-light images, such as infrared images.

Unlike previous studies, this paper focuses on more challenging image data, i.e., infrared images. When the infrared filter is removed in the imaging process, the captured images easily suffer from under-exposure in environments with inadequate lighting conditions. As a result, the main image content is unavoidably corrupted by the noise generated by sensors and other hardware devices. To handle these challenges, we propose a novel normalizing flow-based model that incorporate an auto-color encoding method, namely ACE-Flow. To restore the image content underlying the dark regions, our proposed method can effectively encode the color information of objects and image content by leveraging the proposed auto-color encoding method, providing beneficial prior information for restoration. As low-light image enhancement is intrinsically ill-posed, our method adopts the normalizing flow model to learn the mapping from low-light images to ground-truth images, because the invertible property of the normalizing flow implicitly regularizes the forward imaging process from ground-truth images to low-light images. Furthermore, due to the invertibility of normalizing flows, which allows for exact inference and reconstruction of input data, we have empirically observed that normalizing flows outperform diffusion models. Consequently, our method leverages normalizing flows to produce higher-quality images from low-light counterparts.



(a) Low light image (b) CE image

Fig. 1: A Low-Light Image (LLI) compared with its Color Encoding image (CE).

The main contributions of this paper are summarized as follows:

1) To address the challenging issues of Low-Light Infrared images (LLIR), we propose a novel normalizing flow-based model for low-light image enhancement.
2) Due to the ill-posed nature of low-light images, our proposed method leverages auto-color encoding (ACE) to capture the color information of low-light images, benefiting the restoration of image contents in challenging regions.

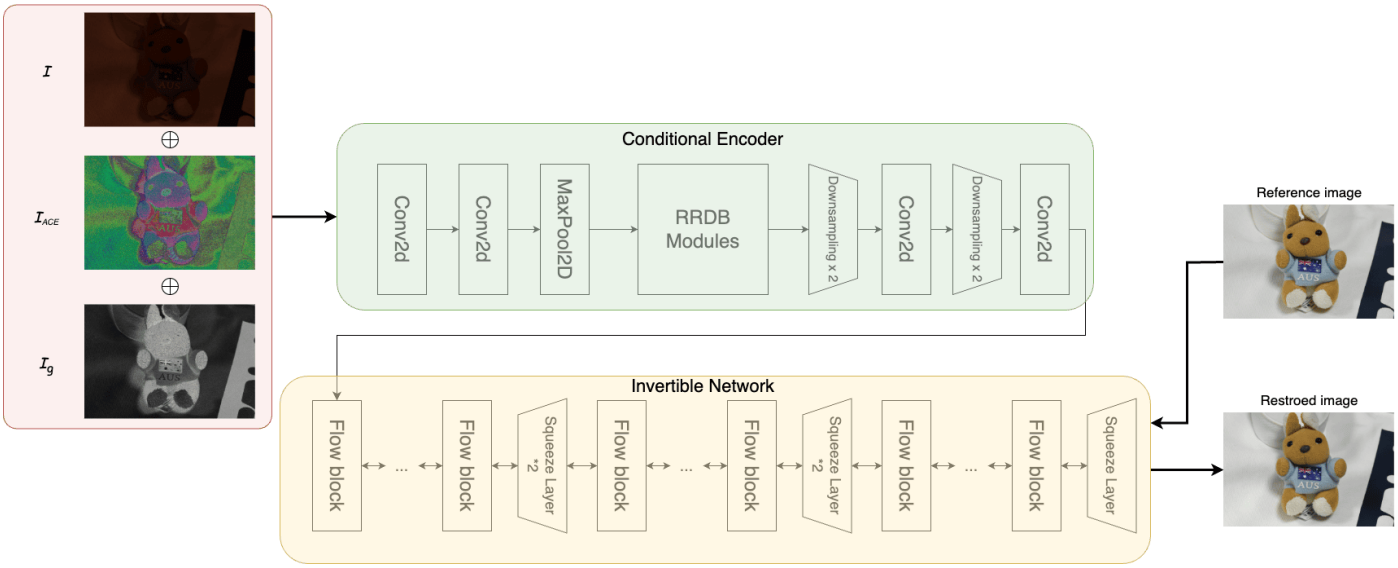


Fig. 2: Overview of our proposed framework (ACE-Flow). The inputs to the framework include the degraded input image (3 channels) I, the color encoding image (3 channels) I_{ACE}, and the inverted grayscale version of the low-light image I_g. $\oplus$ denotes concatenation. The RRDB modules inside the Conditional Encoder represent the Residual-in-Residual Dense Blocks (RRDB) [9].

3) We leverage the invertible property of the normalizing flow to effectively learn the mapping from low-light images to the corresponding ground-truth images, implicitly regularizing the forward imaging process.
4) Experiments show that our proposed method can achieve better performance than other promising deep low-light enhancement models for LLIR images, both in terms of reconstruction and perceptual metrics. Additionally, our method can effectively produce images with rich details and minimal distortion, resulting in superior visual quality.

II. Related Work

A. Low-light image enhancement

Low-Light Image Enhancement (LLIE) is a well-researched area in computer vision, with numerous methods proposed over the years to address the challenges of improving image quality under poor lighting conditions. Here, we categorize the related work into traditional methods, learning-based methods, and recent advancements in the field. LLIE is an active and attractive research area, with various models based on different architectures proposed to tackle this challenging topic [10]–[13]. Retinex-based models mainly focus on decomposing the image into illumination and reflectance components to suppress noise, improve dim lighting conditions, and remove artifacts. Deep learning-based models utilize the computational power of GPUs to train relatively big models with greater numbers of parameters, enabling them to tackle much more complex situations using larger datasets and achieve more robust image restoration [14].

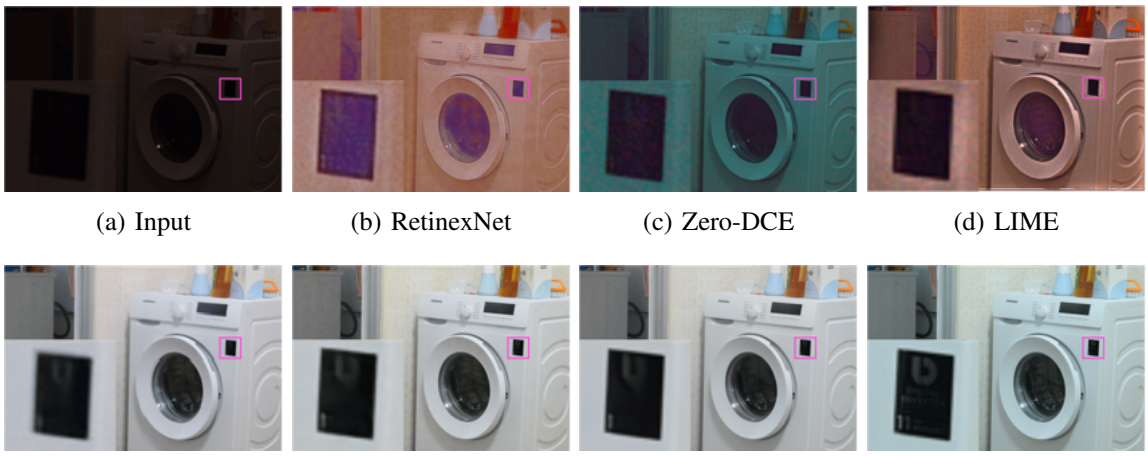
B. Normalizing flow

Normalizing flow is a powerful technique in deep learning for modeling complex data distributions. It transforms a simple probability distribution into a more complex one through a series of invertible transformations, making it highly suitable for tasks requiring detailed probabilistic modeling. Methods like Glow [15] and RealNVP [16] have demonstrated the efficacy of normalizing flow in generating high-quality images by learning the underlying data distribution. These models are particularly useful for tasks such as image synthesis and density estimation. Recently, normalizing flow has been applied to LLIE, as seen in LLFlow [17], which models the distribution of low-light images to effectively handle noise and artifacts. By leveraging the invertibility of flow-based models, LLFlow achieves superior performance in enhancing low-light images without introducing significant artifacts.

III. Methodology

A. The Overall Pipeline of Proposed Method

In this paper, we propose a novel low-light image enhancement technique based on a novel colour encoding method and using the normalizing flow model as the processing backbone. Fig. 2 shows the overall pipeline of the proposed method, consisting of three main parts: the new colour encoding block, a conditional encoder, and an invertible network. The colour encoding block draws inspiration from the powerful positional encoding technique in the Transformer models. Instead of coding positional information, we use a series of sine and cosine waves to encode the colour pixels directly. The rationale behind this design is to leverage this encoding scheme to extract powerful and discriminative representations that are



(a) Input (b) RetinexNet (c) Zero-DCE (d) LIME

(e) DiffIR (f) Restormer (g) Ours (h) Reference

Fig. 3: Perceptual quality comparison with SOTA methods in the field of low-light enhancement in the IR dataset. Our model performs better in noise reduction and restores more details with smooth surfaces. The small pink region within each image was magnified and repositioned to the lower-left corner for detailed comparison.

not easily discernible in low-light images to the human eye. We will demonstrate the effectiveness of this method later. In addition to this new colour encoding scheme, we also extract the gray map of the input low-light images, which, together with the sinusoidal wave-encoded colour map, serves as prior information in the restoration process. The conditional encoder is responsible for extracting features from the input and generating the input for the invertible network. The invertible network, which consists of flow blocks, aims to generate high-quality images based on the features of the input.

B. Preliminary

A normalizing flow model transforms a simple probability distribution into a more complex one through a series of invertible and differentiable transformations. Previous deep learning-based models primarily relied on pixel reconstruction loss. However, due to the challenge of separating undesirable artifacts from the true distribution of the reference images, these models often produce restored images with poor perceptual quality, resulting in noise and blurriness [8], [18].

Motivated by the good reconstruction performance of flow-based models [19]–[21], we realized that utilizing conditional probability distributions can address the aforementioned problem by encompassing various distributions of natural images. In particular, the state-of-the-art (SOTA) method LLFlow [17] demonstrated outstanding performance by using normalizing flow conditioned on low-light images to model the constrained distribution of the reference images. In our method, we adopt the core concept of conditional flow, along with the likelihood evaluation method proposed in LLFlow, as the backbone model of our framework. The conditional probability density function (PDF) for well-exposed images is formulated as follows:

$$f_{cnf}(x_{hq}|x) = f_z(\Theta(x_{hq};x)) \left| \det \frac{\partial \Theta}{\partial x_{hq}}(x_{hq};x) \right| \quad (1)$$

where $f_{cnf}(\cdot)$ represents the conditional PDF, and $\Theta(\cdot)$ is the bidirectional network composed of N invertible layers $\theta_1, \theta_2, \ldots, \theta_N$. The latent variable $z = \Theta(x_{hq};x)$ is derived by transforming the corrupted inputs x into normally exposed images x_{hq}. By employing maximum likelihood estimation, the model can be optimized using the negative log-likelihood loss function. The loss function is formulated as follows:

$$\begin{aligned} L_{nll}(x, x_{hq}) &= -\log f_{cnf}(x_{hq}|x) \\ &= -\log f_z(\Theta(x_{hq};x)) - \sum_{n=0}^{N-1} \log \left| \det \frac{\partial \theta_n}{\partial z_n}(z_n; g_n(x)) \right| \end{aligned} \quad (2)$$

where $g(\cdot)$ denotes the encoder that generates the conditional embedding of the layers θ_i from the bidirectional network.

C. Sinusoidal modulation color encoding

Inspired by the positional encoding scheme of the Transformer model [22], we generate three channels of colour modulated sinusoidal waves and create a color-encoded low-light input. Using a method similar to that used for sequential data in Transformer models, the color encoding for images is computed to derive a more powerful and discriminative representation. For a given input image, the color encoding is calculated as follows:

$$ACM(I_{(x,y,c)}, 2i) = \sin\left(\frac{I_{(x,y,c)}}{10000^{\frac{2i}{d}}}\right) \quad (3)$$

$$ACM(I_{(x,y,c)}, 2i+1) = \cos\left(\frac{I_{(x,y,c)}}{10000^{\frac{2i}{d}}}\right) \quad (4)$$

where $I_{(x,y,c)}$is the pixel value of the input image at spatial location (x, y) and $c \in \{0, 1, 2\}$ represents the red, green, and blue channels of the input image respectively, $i = 0, 1, ..., d/2$ is the modulation dimension index, and d

represents the modulation dimension, which corresponds to the total number of channels of the input image provided to the model. It is evident that Equations (3) and (4) map a $d-$D pixel vector at the position (x, y) to an auto colour encoding matrix $ACM(I_{(x,y,c)})$ in a specific channel $c \in \{0, 1, 2\}$. To obtain sinusoidal modulated color encoding, we follow a similar operation as in the positional encoding method of the Transformer model. Then, we sum over the modulation dimension indices, as follows:

$$ACE(x, y, c) = \frac{1}{d} \sum_{i=0}^{i=d} ACM(I_{(x,y,c)}, i) \tag{5}$$

where $ACE(x, y, c)$ is a sinusoidal modulated colour encoding at the spatial location (x, y) and channel c. Then, we concatenate $ACE(x, y, c)$, $c \in \{0, 1, 2\}$, to form a color encoding image ACE, as follows:

$$I_{ACE} = [ACE_r, ACE_g, ACE_b]. \tag{6}$$

where ACE_r, ACE_g and ACE_b denote the color encoding images of the red, green and blue channels.

D. Graymap

We also generate a graymap as an extra input image. The graymap I_g is calculated by taking the average of the red, green, and blue values at each pixel position, and then subtracting this average from 255, as follows:

$$I_g(x, y) = 255 - I_{mean}(x, y), \tag{7}$$

where I_{mean} represents the average value of the red, green and blue pixel values of the input image, i.e., the grayscale image. This operation provides an additional channel that highlights the inverse of the average intensity of the image. By emphasizing the areas of the image that have lower intensities, the graymap can guide the model to better understand the overall brightness and contrast. This additional channel can be particularly useful in enhancing the model's ability to detect and process features that might be more subtle or less obvious in the original image.

Finally, as illustrated in Equation (7), we concatenate the original image I, the sinusoidal modulated colour encoding image I_{ACE}, and the graymap I_g together to form the input I_{input} to the backbone network:

$$I_{input} = [I, I_{ACE}, I_g]. \tag{8}$$

Clearly, I_{input} has seven channels, each with the same resolution as the original input image. The first three channels are the original red, green, and blue channels, and the next three channels are the sinusoidal modulated colour encoding of the red, green, and blue pixels (see Equations (3), (4) and (5)),and the last channel is the graymap.

TABLE I: Quantitative comparison of our method with various state-of-the-art methods

Methods	PSNR(↑)	SSIM(↑)	LPIPS(↓)
RetinexNet [2]	11.14	0.628	0.586
LIME [23]	11.31	0.639	0.560
Zero-DCE [24]	11.40	0.592	0.443
DiffIR [18]	20.74	0.684	0.200
Restormer [8]	24.73	0.842	0.125
LLFlow [17]	25.42	0.866	0.118
Ours	**25.99**	**0.875**	**0.106**

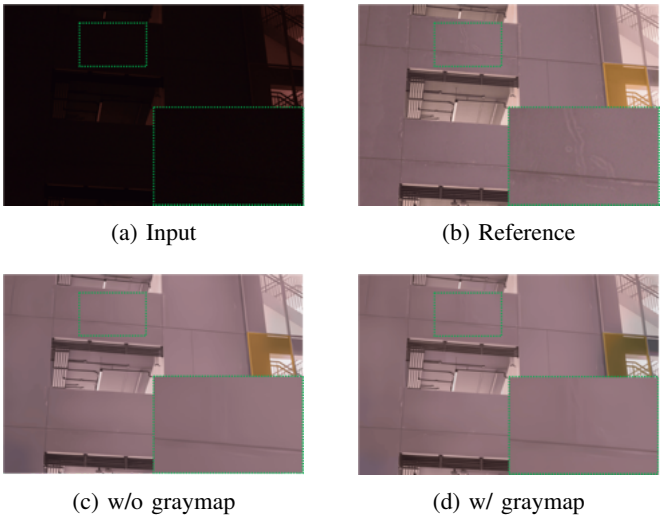
IV. Experiments

A. Experimental settings

To facilitate various experimental settings, we crop the original images of 400 × 600 resolution to 256 × 256 patches, optimizing I/O operations for better efficiency and time savings. Our training setup includes a total of 2,832 image pairs, while the evaluation set comprises 87 image pairs to assess the performance of our trained models. We evaluated our proposed model against several leading methods, including RetinexNet [2], LIME [23], Zero-DCE [24], DiffIR [18], Restormer [8], and LLFlow [17].

B. Experimental results

To evaluate the performance of different methods on the IR dataset, we retrained all the methods using the same training data, i.e., the training set of the IR dataset. For a fair comparison, we explored a wide range of hyperparameters for the compared methods and reported the best performance obtained. The experimental results are presented in Table I, and a visual comparison is shown in Fig. 3.



(a) Input (b) Reference

(c) w/o graymap (d) w/ graymap

Fig. 4: Visual comparison of different configurations of graymap. The green rectangular regions in the images are enlarged for easy comparison. (a) An input image, (b) the reference, (c) the restored image without using graymap and (d) the restored result using the graymap.

(a) Reference (b) CE of LLI (c) w/o ACE (d) w/ ACE

Fig. 5: Visual comparison of different configurations of ACE. We can see that by adding ACE, the model's performance in restoring low-light images is highly improved.

Based on our evaluation and analysis of the experimental results, we observe that RetinexNet exhibits limited generalization ability and produces unsatisfactory outputs, e.g., RetinexNet [2]. We conjecture that this limitation arises because the method assumes the existence of an invariant reflectance map across low-light inputs and ground-truth images, requiring a shared network to extract both illumination and reflectance maps, which is not feasible in our setting.

In contrast, methods based on other principles, such as Restormer [8] with a Transformer architecture and DiffIR [18] using a diffusion model, show better performance. However, our method achieves the best performance among all competitors in terms of both fidelity and perceptual quality.

C. Ablation Study

An ablation study was performed to evaluate the effectiveness of ACE and the graymap in enhancing model performance. The visual results, which highlight the impact of these components, are shown in Fig. 4 and Fig. 5.

The experimental results demonstrate that incorporating the graymap into the model significantly enhances its sensitivity to subtle image details. Specifically, in the green rectangular region of Fig. 4(c), it is evident that the model without the graymap struggles to capture edge information effectively. However, after integrating the graymap, the model is able to generate much clearer and more defined details that correspond to the edges present in the reference image, as illustrated in Fig. 4(b).

To further substantiate these findings, quantitative results are provided in Table II, where "A" and "G" represent the use of ACE and the graymap, respectively. The data in Table II clearly indicates that the combined use of ACE and the graymap yields substantial improvements in the model's performance, particularly in terms of PSNR. Notably, while the integration of the graymap alone results in a more pronounced increase in PSNR, it only brings about a slight enhancement in SSIM. On the other hand, the inclusion of ACE delivers a comparable improvement in SSIM, though the gain in PSNR is marginally lower compared to that achieved with the graymap. When both the graymap and ACE are employed together, the best performance is achieved.

TABLE II: Quantitative Results of Ablation Study. "A" means that ACE is used, while "G" means that graymap is used.

G	A	PSNR (↑)	SSIM (↑)
✗	✗	25.42	0.866
✓	✗	25.87	0.874
✗	✓	25.78	0.874
✓	✓	**25.90**	**0.876**

V. Conclusion

In this paper, we introduce a novel framework, the ACE-Flow model, designed to address the Low-Light Image Enhancement (LLIE) task using an infrared (IR) dataset. This dataset captures more photons by cutting off the infrared filter block inside the camera. By incorporating the ACE-Flow structure and a unique graymap, our model can more effectively interpret the color information and intricacies of low-light images, resulting in high-quality outputs. Extensive experiments demonstrate that our model excels in both quantitative evaluations and perceptual quality, showcasing its superior performance.

References

[1] L. Zhao, S.-P. Lu, T. Chen, Z. Yang, and A. Shamir, "Deep symmetric network for underexposed image enhancement with recurrent attentional learning," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 12 075–12 084.

[2] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[3] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net: Low-light image enhancement using deep convolutional network," *arXiv preprint arXiv:1711.02488*, 2017.

[4] S. Park, S. Yu, M. Kim, K. Park, and J. Paik, "Dual autoencoder network for retinex-based low-light image enhancement," *IEEE Access*, vol. 6, pp. 22 084–22 093, 2018.

[5] C. Li, C.-L. Guo, M. Zhou, *et al.*, "Embedding fourier for ultra-high-definition low-light image enhancement," *arXiv preprint arXiv:2302.11831*, 2023.

[6] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5901–5910.

[7] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17 714–17 724.

[8] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.

[9] X. Wang, K. Yu, S. Wu, *et al.*, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.

[10] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3291–3300.

[11] J. Xiao, W. Jia, and K.-M. Lam, "Feature redundancy mining: Deep light-weight image super-resolution model," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 1620–1624.

[12] H. Xie, Z. Huang, F. H. Leung, Y. Ju, Y.-P. Zheng, and S. H. Ling, "A structure-affinity dual attention-based network to segment spine for scoliosis assessment," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2023, pp. 1567–1574.

[13] Y. Ju, K.-M. Lam, J. Xiao, C. Zhang, C. Yang, and J. Dong, "Efficient feature fusion for learning-based photometric stereo," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2023, pp. 1–5.

[14] J. Xiao, Z. Lyu, C. Zhang, Y. Ju, C. Shui, and K.-M. Lam, "Towards progressive multi-frequency representation for image warping," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 2995–3004.

[15] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," *Advances in neural information processing systems*, vol. 31, 2018.

[16] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real nvp," *arXiv preprint arXiv:1605.08803*, 2016.

[17] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot, "Low-light image enhancement with normalizing flow," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, 2022, pp. 2604–2612.

[18] B. Xia, Y. Zhang, S. Wang, *et al.*, "Diffir: Efficient diffusion model for image restoration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 13 095–13 105.

[19] A. Lugmayr, M. Danelljan, L. Van Gool, and R. Timofte, "Srflow: Learning the super-resolution space with normalizing flow," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, Springer, 2020, pp. 715–732.

[20] V. Wolf, A. Lugmayr, M. Danelljan, L. Van Gool, and R. Timofte, "Deflow: Learning complex image degradations from unpaired data with conditional flows," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 94–103.

[21] M. Xiao, S. Zheng, C. Liu, *et al.*, "Invertible image rescaling," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, Springer, 2020, pp. 126–144.

[22] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, *Attention is all you need*, 2023. arXiv: 1706.03762 [cs.CL]. [Online]. Available: https://arxiv.org/abs/1706.03762.

[23] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.

[24] C. Guo, C. Li, J. Guo, *et al.*, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1780–1789.