DarkIR: Robust Low-Light Image Restoration

Daniel Feijoo [1†], Juan C. Benito [1], Alvaro Garcia [1], Marcos V. Conde [1,2†]
[1] Cidaut AI, Valladolid, Spain
[2] Computer Vision Lab, University of Würzburg, Germany
https://github.com/cidautai/DarkIR † Corresponding Authors {danfei,marcos.conde@cidaut.es}

Low-light and Blurry (↓Noise) | RetinexFormer [3] fails with Blur | **DarkIR** (Ours)

Low-light and Noisy (↓Blur) | LEDNet [65] fails with Noise | **DarkIR** (Ours)

Figure 1. Previous Low-light Image Enhancement (LLIE) and restoration methods are not robust to blur and illumination changes. Our multi-task model is able to restore real low-light images under varying illumination, noise and blur conditions. Zoom-in to see details.

Abstract

Photography during night or in dark conditions typically suffers from noise, low-light and blurring issues due to the dim environment and the common use of long exposure. Although Deblurring and Low-light Image Enhancement (LLIE) are related under these conditions, most approaches to image restoration solve these tasks separately. In this paper, we present an efficient and robust neural network for multi-task low-light image restoration. Instead of following the current tendency of Transformer-based models, we propose new attention mechanisms to enhance the receptive field of efficient CNNs. Our method reduces the computational costs in terms of parameters and MAC operations compared to previous methods. Our model, DarkIR, achieves new state-of-the-art results on the popular LOL-Blur, LOLv2 and Real-LOLBlur datasets, being able to generalize on real-world night and dark images.

1. Introduction

During night or low-light conditions, we can define the image formation process as follows

$$\mathbf{y} = \gamma\left(\mathbf{x} \otimes \mathbf{k}\right) + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is the observed dim image, $\mathbf{x}$ the unperturbed captured scene, $\mathbf{k}$ represents the lens (point-spread function) PSF blurring kernel, $\mathbf{n}$ is the additive sensor noise, and γ is a function to control the dynamic range and pixel saturation. We use $\otimes$ to represent the convolution operator.

In comparison with daytime conditions, at low-light, the noise (shot and read) is substantially higher. During night photography, cameras usually use long exposure (slow shutter speed) to allow more available light to illuminate the image. However, long exposure could lead to ghosting and blurring artifacts. These image degradations are more notable on smartphones, since these have a fixed aperture and limited optics. For these reasons, joint low-light enhancement and deblurring is paramount for mobile computational photography [12, 65] — see Figure 1 samples.

Under any circumstances, the captured image will present certain levels of noise and blur [13]. Using a tripod to capture steady images, we can notably reduce the blur. Moreover, if we use proper illumination, noise can also be reduced [1, 20].

However, nowadays most photographs are captured using (handheld) smartphones. Due to the limited sensor size and optics, these employ more complex Image Signal Processors (ISPs) [10, 12] in comparison to DSLM (Digital Single-Lens Mirrorless) cameras. Smartphone photography is still far from DSLM quality standards, however, recent research in low-level computer vision and computational photography allows to close the gap. Low-light image enhancement (LLIE) [3, 14, 30, 31, 45, 64], image deblurring [21, 25, 38, 40] and night photography enhancement [42, 65] are popular tasks.

For instance, avoiding the presence of blur due to hand tremor (hand shaking) has been well-studied, even in low-light scenarios [29, 62]. However, in most cases, these tasks are solved individually, thus, the state-of-the-art methods for image deblurring do not generalize to nighttime images, and the best methods for low-light enhancement cannot reduce the notable blur. This presents a clear limitation, since multiple task-specific models need to be fine-tuned, stored and applied in sequence, which limits their applications in real-world cases.

To the best of our knowledge, very few works aim to solve these tasks (denoising, deblurring and LLIE) in a jointly end-to-end manner [6, 33, 63, 65], being LEDNet [65] the most notable work. We focus on this research direction since exploiting the correlation between the degradations allows us to achieve the best performance in terms of image reconstruction, usability and efficiency.

Our contribution We propose a convolutional neural network (CNN) that operates in both the spatial and frequency domains. In the spatial domain, we focus on solving the noise $\mathbf{n}$ and the non-uniform blur $\mathbf{k}$, we achieve this by using large receptive field spatial attention. On the other hand, in the Fourier domain, we are able to enhance the low-light conditions easily [28, 44], due to the global nature of the task. We can summarize our contributions as:

1. We design a lightweight neural network with frequency attention, and large receptive field attention, combining spatial and frequency information.
2. Our model, DarkIR, achieves state-of-the-art results on the popular LOLBlur and Real-LOLBlur datasets [65], improving **+1dB** in PSNR over LEDNet [65], while having less computational cost.
3. DarkIR represents a new baseline for multi-task night/dark image enhancement.

2. Related Work

Image Deblurring We can see decades of research on reconstructing sharp scenes. Reducing the blur in an image is divided into blind and non-blind methods. While the non-blind methods consider the blurring kernel $\mathbf{k}$ (or PSF) to process the image, the blind methods do not have any prior knowledge about the blur degradation.

In recent years, multiple deep learning-based approaches have been proposed for blind and non-blind deblurring [25, 38], surpassing traditional methods in both scenarios. The non-blind approaches offer a great solution, considering that only blurry-sharp image pairs are required for training such models, and we do not require PSF estimation or any information about the sensor. Most of these approaches are sensor-agnostic *i.e.,* they can enhance sRGB images captured from different cameras.

Nowadays, a big part of these methods are based on convolutional neural networks (CNNs) [4, 25, 37, 57]. In DeblurGAN [25], the authors use Generative Adversarial Networks (GANs) to solve this problem. More recently, the authors of [24] implemented an efficient frequency domain based transformer for deblurring. We also find iterative methods and diffusion models [33, 49].

Low-Light Image Enhancement (LLIE) The first methods used to consider image statistics or prior information [2, 18], being most of them based on the well-known Retinex Theory [26].

Following the deep learning tendency, nowadays LLIE methods are based on Convolutional Neural Networks (CNNs) such as RetinexNet [48] (and the corresponding LOL dataset), ZeroDCE [16] and SCI [35]. Recent methods explore the power of transformers in this task, such as in the case of RetinexFormer [3], or use Fourier frequency information to enhance the amplitude of the image, like in FourLLIE [44].

Low-Light Blur Enhancement. Even though image deblurring and low-light enhancement are tasks that capture great attention, solving both tasks at the same time is a challenging task, and very few works in the literature tackle it [6, 33, 63, 65]. NBDN [6] proposes a non-blind network to enhance night saturated images. When deconvolving the image to its sharp version, the presence of noise or saturated regions needs to be kept in mind by the algorithm. With this work, the authors proved that previous methods had issues in solving this specific task.

On the other hand, with LEDNet [65], the authors try to solve the problem of low-light enhancement considering that the images are also blurry. This is a realistic assumption, since smartphones need long exposure times for low-light environments. They developed an encoder-decoder network to solve this problem and introduced the popular LOLBlur and Real-LOLBlur datasets.

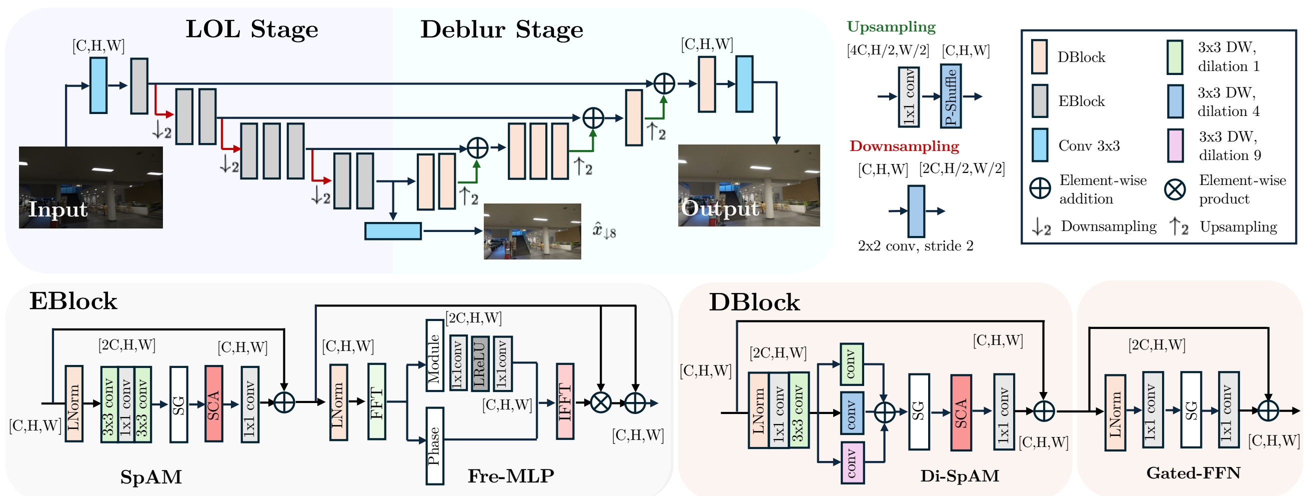

Figure 2. General diagram of **DarkIR**. The neural network follows an encoder-decoder architecture. We use different **blocks** for encoding and decoding that follow the Metaformer structure [56]. The encoder focuses on the low-light illumination issues using Fourier information. Thus, the encoder produces a low-resolution reconstructed image $\hat{x}_{\downarrow 8}$ with corrected illumination. The decoder focuses on upscaling and reducing the blur using the prior illumination-enhanced encoded features. To achieve this, the decoder uses large receptive field spatial attention. This design allows our lightweight model to have less parameters and FLOPs than previous methods.

3. Method

We follow Metaformer [56] to design our neural network. This design simplifies transformer-based architectures into simple blocks with 2 components: global attention (*e.g.,* token mixer), and a feed-forward network (FFN, MLP). The typical formulation for these blocks is:

$$z_1 = \text{Attention}(\text{LayerNorm}(z)) + z \tag{2}$$

$$z_2 = \text{FFN}(\text{LayerNorm}(z_1)) + z_1 \tag{3}$$

where z are the input features and z_2 the output features of the block. Most popular image restoration models, such as NAFNet [7], use the same structure and adapt the Attention module to different tasks.

We improve the metaformer structure by developing LLIE and deblurring specific blocks. Furthermore, we use simple gates (SG) and simplified channel attention (SCA) instead of activations as [7].

Low-light Enhancement can be solved efficiently in the frequency domain.

Many works [28, 44] proved that the low-light conditions are highly correlated with the amplitude of the image in the Fourier domain. Thus, by enhancing just the amplitude of an image (without touching the phase), we can substantially correct the illumination of the image. Moreover, this property stands at different resolutions [28]. Therefore, we can estimate an illumination-enhanced image at low-resolution and upscale it.

Sharpening and Reducing Blur usually require large receptive fields, this could be achieved by extracting deep features while downsampling the image – NAFNet's approach [7]. An alternative would be to use large kernels [32], however, this could lead to more computational complexity and memory requirements.

DarkIR Model In Figure 2, we illustrate our model. Unlike most previous methods, we use two different blocks for the encoder and decoder. The idea behind this asymmetry is to perform low-light enhancement at low-resolution in the encoder, and reduce the blur in the decoder, following a similar strategy as LEDNet [65]. The decoder will use illumination-enhanced features from the encoder, and shall focus on upsampling and improving the sharpness of the already enhanced low-resolution reconstruction $\hat{x}_{\downarrow 8}$.

For the encoder block, and to restore the low-light conditions, we work in the Fourier domain [28, 44]. The decoding block focuses on the spatial domain by incorporating dilated convolutions with large receptive field. By using task-specific blocks, we are able to use fewer blocks. This design allows to reduce notably the number of parameters and computational cost in terms of MACs (Multiply-Accumulate Operations) and FLOPs (Floating-point Operations per second).

How can we make sure the Encoder is performing low-light enhancement? As shown in Figure 2, the encoded features are linearly combined using a convolutional layer to produce an intermediate image representation $\hat{x}_{\downarrow 8}$. We will use this to regularize our model using an additional loss function. By producing a good low-resolution representation, we can ensure that the amplitude in the Fourier domain has been properly enhanced.

3.1. Low-light Enhancement Encoder

We design the encoder blocks (**EBlock**) to enhance the low-light conditions of the image using Fourier information, and following the Metaformer [56] (and NAFBlock [7]) structure. The block has two components: the spatial attention module (SpAM) and a feed-forward network in the frequency domain (FreMLP).

The spatial attention module resembles the NAF-Block [7] with an inverted residual block followed by simplified channel attention (SCA). Instead of using activations, we use a simple gating mechanism, allowing our model to extract meaningful spatial information for enhancement in the frequency domain.

As suggested by other works [33, 44] the information related to the light conditions of the image depends mainly on the amplitude in the frequency domain. To enhance this, we apply the Fast Fourier Transform (FFT) and operate only over its amplitude. After this operation, we transform again to the space domain with the Inverse Fast Fourier Transform (IFFT). The Fre-MLP serves as an additional attention mechanism. In this context, the MLP operating in the amplitude has better benefits than operating in the spatial domain (*e.g.,* channel MLP [56]).

The encoder uses strided convolutions to downsample the features. After each level, the features have half of their original spatial resolution, which implies that more encoder blocks can be used in the deep levels without significantly increasing the number of operations.

Finally, the encoder will provide illumination-enhanced features to the decoder, and already a low-resolution estimation of the clean image $\mathbf{x}$. This low-resolution image is $\hat{x}_{\downarrow 8}$ estimated as a combination of the encoded deep features, it has a resolution $8\times$ lower than the original one. Although it is a small estimation, the illumination (and amplitude) are consistent across scales [28].

3.2. Deblurring Decoder

The decoder block (**DBlock**) focuses on spatial transformations. The input of the decoder is a deep representation of $\hat{x}_{\downarrow 8}$, thus we can assume: (1) the decoder should focus on upsampling such an initial estimation, (2) the decoder should focus on reducing the blur and improving sharpness, since the illumination has been corrected by the encoder. In this block, we also maintain the metaformer structure [56]:

$$z_1 = \text{Di-SpAM}(\text{LayerNorm}(z)) + z \tag{4}$$

$$z_2 = \text{GatedFFN}(\text{LayerNorm}(z_1)) + z_1 \tag{5}$$

Inspired by Large Kernel Attention (LKA) [17], we created the Dilated-Spatial Attention Module (Di-SpAM). Unlike LKA [17], we use features at 3 different levels, using three dilated depth-wise convolutions with expand (dilation) factors 1,4,9. The attributes from the three branches are combined together, then we apply simplified channel attention to further enhance the features. Finally, we use an MLP with simple gates instead of activations [7].

3.3. Loss Function

Besides the new block designs, the loss function helped to maximize the potential of our approach. We use a combination of distortion losses and perceptual losses to optimize our model f. First, to ensure high-fidelity (low distortion), we use L_{pixel} defined as: $L_{pixel} = \|x - \hat{x}\|_1$, where $f(y) = \hat{x}$ and x are, respectively, the enhanced and the ground-truth (clean) images. Thus, L_{pixel} is the $\mathcal{L}_1$ loss.

To ensure high-fidelity, we use $l1$ norm loss and for perceptual similarity we incorporate loss L_{percep}. For the last one, we use LPIPS [60] based on VGG19 [43] to calculate the distance between features of our images:

$$L_{percep} = \text{LPIPS}\,(x, \hat{x}). \tag{6}$$

Using this loss, we make sure that the network will produce a pleasant image, close to the clean reference. Following [41], we also incorporate the gradient (∇) *edge loss*:

$$L_{edge} = \|\nabla x - \nabla \hat{x}\|_2^2\,, \tag{7}$$

that enforces consistency and accuracy in the reconstruction of the edges (high-frequencies).

Finally, similar to LEDNet [65], we included L_{lol} an architecture guiding loss to assert that the encoder focuses on the low-light enhancing. This loss works over the upsampled output of the encoder $\hat{x}_{\uparrow 8}$,

$$L_{lol} = \|x - \hat{x}_{\uparrow 8}\|_1 + \text{LPIPS}(x, \hat{x}_{\uparrow 8}) \tag{8}$$

comparing it with the reference x. Note that $\uparrow 8$ indicates an $8\times$ resolution upsampling of the encoder output, obtained by bilinear interpolation.

The complete loss function is then:

$$\mathcal{L} = \lambda_p \cdot L_{l1} + \lambda_{pe} \cdot L_{percep} + \lambda_{ed} \cdot L_{edge} + L_{lol} \tag{9}$$

The constants λ_p, λ_{pe}, and λ_{ed} are loss weights empirically set to 1, $1e^{-2}$, and 50 respectively.

4. Experimental Results

4.1. Datasets

To train our model and evaluate its ability to reconstruct low-light blurred images, we use the LOLBlur dataset [65]. Although there are other datasets for this task such as NBDN [6], LOLBlur offers a large-scale synthetic dataset produced using a sophisticated pipeline.

Table 1. **Quantitative evaluation on the LOLBlur dataset.** DarkIR achieves new state-of-the-art results in distortion and perceptual metrics. Moreover, we have 55% less parameters than LEDNet[65] and 88% less than Restormer [57], which is key in memory-constrained devices. This table –specially numbers for previous methods retrained in LOLBlur– recovers results on previous analysis [65] and adds new retrained ones. Best and second best results are bolded and underlined, respectively.

	KinD++[61]	DRBN[54]	DeblurGAN-v2[25]	MIMO[8]	NAFNet[9]	LEDNet[65]	RetinexFormer[3]	Restormer[57]	DarkIR-m (Ours)	DarkIR-l (Ours)
PSNR (dB) ↑	21.26	21.78	22.30	22.41	25.36	25.74	26.02	26.72	27.00	**27.30**
SSIM ↑	0.753	0.768	0.745	0.835	0.882	0.850	0.887	**0.902**	0.883	0.898
LPIPS ↓	0.359	0.325	0.356	0.262	0.158	0.224	0.181	**0.133**	0.162	0.137
Params (M) ↓	1.2	**0.6**	60.9	6.8	12.05	7.4	1.61	26.13	3.31	12.96
MACs (G) ↓	34.99	48.61	-	67.25	12.3	38.65	15.57	144.25	**7.25**	27.19

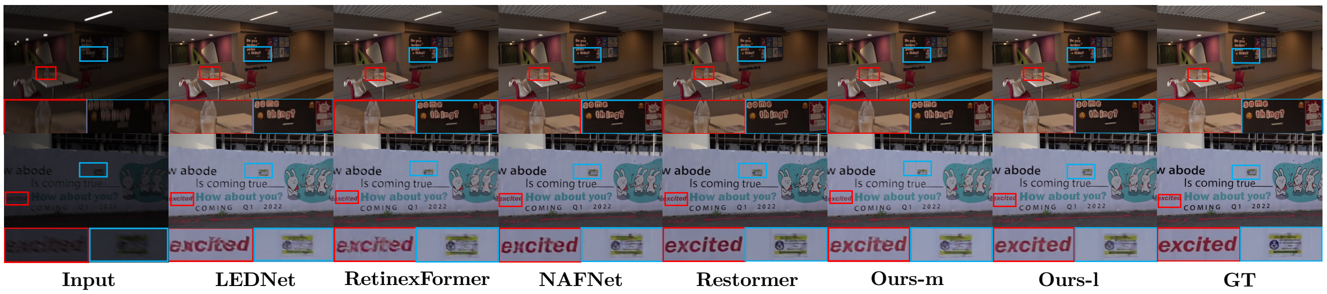

Figure 3. Qualitative comparisons on the **LOLBlur** dataset (synthetic samples from the testset).

LOLBlur has 10200 training pairs, and 1800 testing pairs. Note that this is a *synthetic dataset*, although generated in a realistic manner [65]: the data is generated by averaging frames to synthesize blur and darkening the normal-light images with EC-Zero-DCE (a variant of Zero-DCE [16]). We use the dataset variant that includes real sensor noise, which makes our model more effective in challenging conditions and real scenarios.

Real-LOLBlur is a *real-world test* dataset, that contains 482 real-world night blurry images selected from RealBlur [40] to verify the generalization of the proposed method. Note that these images do not have a ground-truth since these were captured in the wild.

LOLv2 (real) is a real-world dataset that includes 689 low/high paired images for training and 100 low/high paired images for testing. Note that LOLv2-Real [55] is the extended version of LOL[47], thus we use the v2 version directly. We also use *LOLv2-Synthetic* that includes 900 pairs of low/high images for training and 100 validation ones.

LSRW includes images from a DSLM Nikon camera and a Huawei smartphone. The LSRW-Nikon dataset is composed of 3150 training image pairs and 20 testing image pairs. The LSRW-Huawei dataset contains 2450 pairs of images and 30 pairs for training and validation, respectively.

4.2. Results

We quantitatively and qualitatively evaluate the proposed DarkIR on the LOLBlur Dataset [65]. Implementations details can be found in the supplementary. We retrained some general purpose state-of-the-art methods in LOLBlur and compare them with LEDNet [65] and our proposed network.We follow the baseline methods analyzed in previous works: (1,2) zero-shot methods trained for real-world cases, and (3) fine-tuned methods for this particular task.

1. LLIE → Deblurring. We consider Zero-DCE [16], RUAS [30] and RetinexFormer [3] as LLIE models, followed by a deblurring network like MIMO-UNet [8] or NAFNet [7].

2. Deblurring → LLIE. For deblurring, we include popular baselines such as DeblurGAN-v2 [25] trained on the RealBlur [40] dataset, and MIMO-UNet [8] or NAFNet [7] trained on the GoPro [38] dataset. We employ Zero-DCE [16] and RetinexFormer [3] for light enhancement.

3. End-to-end training on the LOLBlur dataset. We consider the following LLIE models retrained on the LOLBlur dataset: KinD++ [61], DRBN [54] and RetinexFormer[3]. In addition, we consider four deblurring networks: DeblurGAN-v2 [25], NAFNet[9], MIMO-UNet [8] and Restormer[57].

Evaluation Metrics. We employ traditional quality (distortion) metrics PSNR and SSIM for evaluation on the synthetic LOLBlur dataset. To evaluate the perceptual quality of the restored images, we use the perceptual metric LPIPS [60] between the reference and reconstructed image.

Quantitative and Qualitative Results In Table 1 we compare with fine-tuned methods for this task. We improve

Table 2. Additional quantitative evaluation on LOLBlur dataset with enhancement pipelines (Deblurring + LLIE).

	1. LLIE → Deblurring			2. Deblurring → LLIE				
	Zero-DCE [16] → MIMO [8]	RUAS [30] → MIMO [8]	RetinexFormer [3] → NAFNet [7]	Chen [5] → Zero-DCE [16]	DeblurGAN-v2 [25] → Zero-DCE [16]	MIMO [8] → Zero-DCE [16]	NAFNet [7] → RetinexFormer [3]	**DarkIR-m (Ours)** End-to-End
PSNR (dB) ↑	17.68	17.81	17.16	17.02	18.33	17.52	14.66	**27.00**
SSIM ↑	0.542	0.569	0.673	0.502	0.589	0.57	0.500	**0.883**
LPIPS ↓	0.510	0.523	0.392	0.516	0.476	0.498	0.465	**0.162**

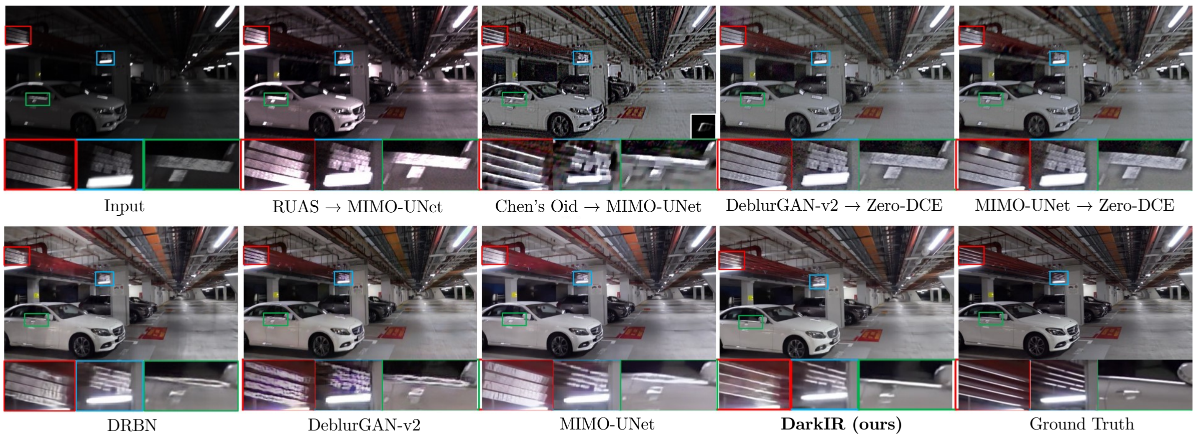

Figure 4. Additional visual comparisons on the **LOLBlur** [65] dataset with 2-step pipelines. DarkIR generates much sharper images with visually pleasing results. (Zoom in for best view).

the previous state-of-the-art LEDNet [65] by +1db in terms of PSNR, and reduce LPIPS by half. Table 1 also showcases that DarkIR performs better than most of the other general purpose methods, while reducing parameters and computing cost. Figure 3 qualitatively supports these results.

In Table 2 we compare with 2-step (zero-shot) pipelines, with a clear dominance of DarkIR. Figure 4 compares DarkIR with 2-step pipelines, where it shows more detail and sharpness. We provide more results in the appendix.

Multi-Task Results. We also trained our model for practical low-light restoration purposes, combining LOLBlur, LOLv2-Real, LOLv2-Synthetic and LSRW datasets (as all-in-one restoration methods [11, 27, 39, 59]). Results can be seen in Table 3 and Table 4. As a multi-task method, **DarkIR-mt** outperforms previous LLIE methods, while being also robust to blur (previous methods are only robust to illumination and noise). However, the performance in LOLBlur is 26.62 dB, suffering a slight -0.4dB loss in PSNR. Figures 5 and 6 showcase these results in LSRW and LOLv2-Real datasets, respectively. We provide additional details in the supplementary.

4.3. Evaluation on Real Data

We use the **Real-LOLBlur** [65] to evaluate the robustness of our method in real-world cases. Since there is no ground-truth for the test images, we use well-known blind quality assessment metrics. We present results in other *real-world (unpaired) LLIE datasets* in the appendix.

Table 3. Metrics on the LSRW dataset (50 test images from Huawei and Nikon) [19]. All the values are adopted from [35, 53].

	RetinexNet [47]	FIDE [51]	DRBN [54]	KinD [61]	STAR [50]
PSNR ↑	15.906	<u>17.669</u>	16.149	16.472	14.608
SSIM ↑	0.3725	<u>0.5485</u>	0.5422	0.4929	0.5039
	EnGAN [22]	ZDCE [16]	RUAS [30]	SCI [35]	**DarkIR-mt** (Ours)
PSNR ↑	16.311	15.834	14.437	15.017	**18.93**
SSIM ↑	0.4697	0.4664	0.4276	0.4846	**0.583**

Evaluation Metrics. We employ recent image quality assessment methods: MUSIQ [23], NRQM [34] and NIQE [36] as our perceptual metrics. Following previous works, we choose the MUSIQ model trained on the KonIQ-10k dataset, which focuses more on color contrast and sharpness assessment – quite suitable for our task. We use the pyiqa [1] implementation of these metrics.

Quantitative Evaluations. As shown in Table 5, the proposed DarkIR achieves competitive perceptual quality scores in terms of the three perceptual metrics, indicating that our method performs in tune with human perception.

[1] https://pypi.org/project/pyiqa/

Input KinD DRBN EnGAN **DarkIR-m** Ground Truth

Figure 5. Qualitative results on the real-world dataset **LSRW-Huawei** [19]. We provide more samples in the supplementary.

Input RetinexFormer [3] **DarkIR-m** Ground Truth

Figure 6. Qualitative results compared with state of the art method RetinexFormer [3] on **LOLv2-Real**.

Table 4. Results on LOLv2-Real [55] and LOLv2-Synthetic [55]. Our *multi-task model (DarkIR-mt)* obtains new SOTA results by leveraging all-in-one training, and including low-light deblurring, which proves the efficacy of the architecture. Table based on [3, 44]. MACs were calculated on $256 \times 256 \times 3$ inputs.

Methods	Complexity		LOLv2-Real		LOLv2-Syn	
	MACs (G)↓	Params (M)↓	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
UFormer [46]	12.00	5.29	18.82	0.771	19.66	0.871
RetinexNet [47]	587.47	0.84	15.47	0.567	17.13	0.798
EnGAN [22]	61.01	114.35	18.23	0.617	16.57	0.734
RUAS [30]	**0.83**	**0.003**	18.37	0.723	16.55	0.652
FIDE [51]	28.51	8.62	16.85	0.678	15.20	0.612
DRBN [54]	48.61	5.27	20.29	0.831	23.22	0.927
KinD [61]	34.99	8.02	14.74	0.641	13.29	0.578
Restormer [57]	144.25	26.13	19.94	0.827	21.41	0.830
MIRNet [58]	785	31.76	20.02	0.820	21.94	0.876
SNR-Net [52]	26.35	4.01	21.48	<u>0.849</u>	24.14	0.928
FourLLIE [44]	<u>5.8</u>	<u>0.120</u>	21.60	0.847	24.17	0.917
Retinexformer [3]	15.57	1.61	<u>22.80</u>	0.840	**25.67**	<u>0.930</u>
DarkIR-mt (Ours)	7.25	3.31	**23.87**	**0.880**	<u>25.54</u>	**0.934**

Table 5. Perceptual quality metrics on Real-LOLBlur [65].

	RUAS → MIMO	MIMO → Zero-DCE	RetinexFormer	NAFNet	LEDNet	Restormer	**DarkIR-m**	**DarkIR-l**
MUSIQ↑	34.39	28.36	45.30	**50.22**	39.11	46.6	48.36	<u>48.79</u>
NRQM↑	3.322	3.697	5.281	4.940	**5.643**	4.627	<u>4.983</u>	4.917
NIQE↓	6.812	6.892	4.576	<u>5.123</u>	4.764	**5.268**	4.998	5.051

Qualitative Evaluations. Figure 7 presents visual comparisons on real-world night blurry image from Real-LOLBlur [65]. These samples showcase the robustness of our approach in real cases with handheld motion blur, sensor noise, saturated pixels and low illumination.

4.4. Ablation Study

In addition to our results, we include three ablation studies on model design and training.

In Table 6 we compare results obtained using different block configurations, such as NAFBlock, EBlock or DBlock. The description states which changes are applied to the network, and everything else is kept exactly as in the proposed model. We can clearly see how our model achieves the best results. Unlike the final proposed model, all the models in this study were trained using crops of 256px instead of 384px, which explains the lower results in general. We also studied the decoder block spatial attention in Table 7, where the well-known Large Kernel Attention (LKA) mechanism is compared with our Dilated-spatial Attention Module (Di-SpAM). Our approach performs better and requires less parameters and operations.

In Table 8 we also explore the scalability of our model by changing its channel embedding. As expected, by increasing the embedding size, the performance raises. In ascending order of parameters, we have channel embeddings of 16, 32, and 64. We find the best efficiency/performance balance in the 32 channels embedding (**DarkIR-m**).

In the supplementary material, we present additional studies on the development of the architecture.

Input LEDNet Retinexformer Nafnet Restormer Ours-m Ours-l

Figure 7. Qualitative comparison on real night scenes from the **RealBlurLOL** dataset. (Zoom in for best view).

Table 6. Network blocks ablation study. Combination of EBlocks and DBlocks achieves the best performance.

	Params↓ (M)	MACs↓ (G)	PSNR↑	SSIM↑	LPIPS↓
EBlock is NAFBlock	3.12	7.69	26.51	0.8764	0.169
EBlock has also Phase Transform	4.08	8.38	26.30	0.871	0.179
All DBlock	3.2	8.19	26.68	0.876	0.170
All EBlock	3.44	6.46	26.17	0.863	0.187
All NAFBlock [7]	3.04	7.29	26.24	0.868	0.178
DBlock is NAFBlock	3.24	6.84	26.23	0.864	0.185
DBlock w/o Extra Depthwise	3.27	6.99	26.63	0.875	0.174
DarkIR-m	3.31	7.25	26.90	0.874	0.175

Table 7. Spatial Attention ablation study. We compare LKA (large kernel attention) [17] with our proposed Di-SpAM for the decoder blocks. MACs were calculated considering an input of 256px.

	Params↓ (M)	MACs↓ (G)	PSNR↑	SSIM↑	LPIPS↓
LKA [17]	4.06	9.14	26.45	0.876	0.172
Di-SpAM (**DarkIR-m**)	3.31	7.25	27.00	0.883	0.162

Table 8. Ablation study scaling the channel depth dimensions. We can appreciate how our model scales properly, which allows adaptation depending on memory or runtime requirements.

	Params↓ (M)	MACs↓ (G)	PSNR↑	SSIM↑	LPIPS↓
DarkIR-s (16)	0.872	2.04	26.15	0.857	0.206
DarkIR-m (32)	3.31	7.25	27.00	0.883	0.162
DarkIR-l (64)	12.96	27.19	27.30	0.898	0.137

4.5. Efficiency Discussion

Using EBlock and DBlock we are able to greatly reduce the number of parameters of the network, getting an outstanding 55% **less parameters** than LEDNet [65] (previous state-of-the-art) and 88% less than Restormer [57] (second best method). This reduction is also accompanied by a reduction in the number of operations needed to enhance the input image. Considering an image of 256px -as previous works [7]-, LEDNet uses 33.74 GMACs and Restormer 141.24 GMACs, while DarkIR only uses **7.25 GMACs** – note that 1 MAC is roughly 2 FLOPs. This means a reduction of $4\times$ the number of operations with the previous state-of-the-art method and almost $20\times$ to the second best one.

Therefore, DarkIR, while being state-of-the-art in low-light deblurring, is also lighter in all aspects, representing an advancement towards deploying this kind of models on devices with low computational power.

5. Limitations

Although we are able to reduce the computational requirements of the target device for running our model, we have done this by using depth-wise convolutions, which are not necessarily optimal in certain GPU architectures, as they lack of arithmetic intensity [15]. Due to this, the model's inference times are not drastically and proportionally reduced with the reduction in operations. In future work, we will propose new methods that can combine low computational requirements with notably faster inference times.

6. Conclusion

We propose a robust model for multi-task low-light enhancement and restoration. Our model, DarkIR, is an efficient and robust neural network that performs denoising, deblurring and low-light enhancement on dark and night scenes. DarkIR achieves new state-of-the-art results on the popular LOLBlur, LOLv2 and Real-LOLBlur datasets, being able to generalize on real-world night blurry images while being more efficient than previous methods.

Acknowledgements This work was supported by Spanish funds through Regional Funding Agency Institute for Business Competitiveness of Castile Leon (MACS.2 project "Investigación en tecnologías del ámbito de la movilidad autónoma, conectada, segura y sostenible"). The authors thank Supercomputing of Castile and Leon (SCAYLE. Leon, Spain) for assistance with the model training and GPU resources.

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, pages 1692–1700, 2018. 2

[2] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *IEEE transactions on consumer electronics*, 53(2):593–600, 2007. 2

[3] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023. 1, 2, 5, 6, 7

[4] Ayan Chakrabarti. A neural approach to blind motion deblurring. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 221–235. Springer, 2016. 2

[5] Liang Chen, Jiawei Zhang, Songnan Lin, Faming Fang, and Jimmy S Ren. Blind deblurring for saturated images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6308–6316, 2021. 6

[6] Liang Chen, Jiawei Zhang, Jinshan Pan, Songnan Lin, Faming Fang, and Jimmy S Ren. Learning a non-blind deblurring network for night blurry images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10542–10550, 2021. 2, 4

[7] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 3, 4, 5, 6, 8

[8] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4641–4650, 2021. 5, 6

[9] Xiaojie Chu, Liangyu Chen, and Wenqing Yu. Nafssr: Stereo image super-resolution using nafnet. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1239–1248, 2022. 5

[10] Marcos V. Conde, Steven McDonagh, Matteo Maggioni, Ales Leonardis, and Eduardo Pérez-Pellitero. Model-based image signal processors via learnable dictionaries. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(1): 481–489, 2022. 2

[11] Marcos V Conde, Gregor Geigle, and Radu Timofte. High-quality image restoration following human instructions. *arXiv preprint arXiv:2401.16468*, 2024. 6

[12] Mauricio Delbracio, Damien Kelly, Michael S Brown, and Peyman Milanfar. Mobile computational photography: A tour. *arXiv preprint arXiv:2102.09000*, 2021. 2

[13] Michael Elad and Arie Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6(12):1646–1658, 1997. 2

[14] Yixu Feng, Cheng Zhang, Pei Wang, Peng Wu, Qingsen Yan, and Yanning Zhang. You only need one color space: An efficient network for low-light image enhancement. *arXiv preprint arXiv:2402.05809*, 2024. 2

[15] Amir Gholami, Kiseok Kwon, Bichen Wu, Zizheng Tai, Xiangyu Yue, Peter Jin, Sicheng Zhao, and Kurt Keutzer. Squeezenext: Hardware-aware neural network design. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1638–1647, 2018. 8

[16] Chunle Guo 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 *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1780–1789, 2020. 2, 5, 6

[17] Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. Visual attention network. *Computational Visual Media*, 9(4):733–752, 2023. 4, 8

[18] 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. 2

[19] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90: 103712, 2023. 6, 7

[20] Samuel W Hasinoff. Photon, poisson noise. *Computer Vision, A Reference Guide*, 4, 2014. 2

[21] Zhe Hu, Sunghyun Cho, Jue Wang, and Ming-Hsuan Yang. Deblurring low-light images with light streaks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3382–3389, 2014. 2

[22] 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. *TIP*, 2021. 6, 7

[23] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 6

[24] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023. 2

[25] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8878–8887, 2019. 2, 5, 6

[26] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 2

[27] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *CVPR*, pages 17452–17462, 2022. 6

[28] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv preprint arXiv:2302.11831*, 2023. 2, 3, 4

[29] Orly Liba, Kiran Murthy, Yun-Ta Tsai, Tim Brooks, Tianfan Xue, Nikhil Karnad, Qiurui He, Jonathan T Barron, Dillon Sharlet, Ryan Geiss, et al. Handheld mobile photography in very low light. *ACM Trans. Graph.*, 38(6):164–1, 2019. 2

[30] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10561–10570, 2021. 2, 5, 6, 7

[31] Xiaoning Liu, Zongwei Wu, Ao Li, Florin-Alexandru Vasluianu, Yulun Zhang, Shuhang Gu, Le Zhang, Ce Zhu, Radu Timofte, Zhi Jin, et al. Ntire 2024 challenge on low light image enhancement: Methods and results. *arXiv preprint arXiv:2404.14248*, 2024. 2

[32] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. *Advances in neural information processing systems*, 29, 2016. 3

[33] Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, and Liqiang Nie. Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25378–25388, 2024. 2, 4

[34] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-rolution. *Computer Vision and Image Understanding*, pages 1–16, 2017. 6

[35] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5637–5646, 2022. 2, 6

[36] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 6

[37] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 2

[38] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 2, 5

[39] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one blind image restoration. *arXiv preprint arXiv:2306.13090*, 2023. 6

[40] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 2, 5

[41] George Seif and Dimitrios Androutsos. Edge-based loss function for single image super-resolution. In *2018 IEEE International conference on acoustics, speech and signal processing (ICASSP)*, pages 1468–1472. IEEE, 2018. 4

[42] Alina Shutova, Egor Ershov, Georgy Perevozchikov, Ivan Ermakov, Nikola Banić, Radu Timofte, Richard Collins, Maria Efimova, Arseniy Terekhin, Simone Zini, et al. Ntire 2023 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1982–1993, 2023. 2

[43] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4

[44] Chenxi Wang, Hongjun Wu, and Zhi Jin. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7459–7469, 2023. 2, 3, 4, 7

[45] Yufei Wang, Yi Yu, Wenhan Yang, Lanqing Guo, Lap-Pui Chau, Alex C Kot, and Bihan Wen. Exposurediffusion: Learning to expose for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12438–12448, 2023. 2

[46] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, 2022. 7

[47] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, 2018. 5, 6, 7

[48] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 2

[49] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G. Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16293–16303, 2022. 2

[50] Jun Xu, Yingkun Hou, Dongwei Ren, Li Liu, Fan Zhu, Mengyang Yu, Haoqian Wang, and Ling Shao. Star: A structure and texture aware retinex model. *IEEE Transactions on Image Processing*, 29:5022–5037, 2020. 6

[51] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2281–2290, 2020. 6, 7

[52] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *CVPR*, 2022. 7

[53] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12918–12927, 2023. 6

[54] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3063–3072, 2020. 5, 6, 7

[55] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE*

Transactions on Image Processing, 30:2072–2086, 2021. 5, 7

[56] Weihao Yu, Mi Luo, Pan Zhou, Chenyang Si, Yichen Zhou, Xinchao Wang, Jiashi Feng, and Shuicheng Yan. Metaformer is actually what you need for vision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10819–10829, 2022. 3, 4

[57] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 2, 5, 7, 8

[58] 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. *TPAMI*, 2022. 7

[59] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5825–5835, 2023. 6

[60] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 4, 5

[61] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021. 5, 6, 7

[62] Jiacai Zhao, Jie Ma, Bin Fang, Siwen Quan, and Fangyu Hu. Fast motion deblurring using gyroscopes and strong edge prediction. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 739–744. IEEE, 2016. 2

[63] Yuzhi Zhao, Yongzhe Xu, Qiong Yan, Dingdong Yang, Xuehui Wang, and Lai-Man Po. D2hnet: Joint denoising and deblurring with hierarchical network for robust night image restoration. In *European Conference on Computer Vision*, pages 91–110. Springer, 2022. 2

[64] Han Zhou, Wei Dong, Xiaohong Liu, Shuaicheng Liu, Xiongkuo Min, Guangtao Zhai, and Jun Chen. Glare: Low light image enhancement via generative latent feature based codebook retrieval. In *European Conference on Computer Vision*, pages 36–54. Springer, 2025. 2

[65] Shangchen Zhou, Chongyi Li, and Chen Change Loy. Lednet: Joint low-light enhancement and deblurring in the dark. In *ECCV*, 2022. 1, 2, 3, 4, 5, 6, 7, 8