DarkDeblur: Agricultural images deblurring in low-light condition

A report submitted in partial fulfillment of the requirements for Summer Project (BCCS-2999)

Bachelor of Technology

in

Computer Science and Engineering

by

Bhukya Madhu Naik
Roll: 2022BCS-014

Gaddala Sathvik
Roll: 2022BCS-024

Paranjay Dubey
Roll: 2022BCS-051

विश्वजीवनामृतं ज्ञानम्

ABV INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT

GWALIOR - 474015

July 2024

CANDIDATES DECLARATION

We hereby certify that the work, which is being presented in the report, entitled **Dark-Deblur: Agricultural images deblurring in low-light condition**, in partial fulfillment of the requirement for summer project (BCCS-2999) for **Bachelor of Technology in Computer Science and Engineering** and submitted to the institution is an authentic record of our own work carried out during the period *May 2024* to *July 2024* under the supervision of **Prof.Karm Veer Arya**. We also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

Date: Signature of the Candidate

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Date: Signature of the Supervisor(s)

ABSTRACT

Single-shot image deblurring in a low-light condition is known to be a profoundly challenging image translation task. This study tackles the limitations of the low-light image deblurring with a learning-based approach and proposes a novel deep network named as DarkDeblurNet. The proposed DarkDeblur- Net comprises a dense-attention block and a contextual gating mechanism in a feature pyramid structure to leverage content awareness. The model additionally incorporates a multi-term objective function to perceive a plausible perceptual image quality while performing image deblurring in the low-light settings. The practicability of the proposed model has been verified by fusing it in numerous computer vision applications. Apart from that, this study introduces a benchmark dataset collected with actual hardware to assess the low-light image deblurring methods in a real-world setup. The experimental results illustrate that the proposed method can outperform the state-of-the-art methods in both synthesized and real-world data for single-shot image deblurring, even in challenging lighting environments.

Keywords: single-shot image deblurring, low-light condition, learning-based approach, deep network, DarkDeblurNet, dense-attention block, contextual gating mechanism, feature pyramid structure, multi-term objective function, perceptual image quality, computer vision applications, real-world setup, state-of-the-art methods.

ACKNOWLEDGEMENTS

We are highly indebted to **Prof.Karm Veer Arya**, and are obliged for giving us the autonomy of functioning and experimenting with ideas. We would like to take this opportunity to express our profound gratitude to them not only for their academic guidance but also for their personal interest in our project and constant support coupled with confidence boosting and motivating sessions which proved very fruitful and were instrumental in infusing self-assurance and trust within us. The nurturing and blossoming of the present work is mainly due to their valuable guidance, suggestions, astute judgment, constructive criticism and an eye for perfection. Our mentor always answered myriad of our doubts with smiling graciousness and prodigious patience, never letting us feel that we are novices by always lending an ear to our views, appreciating and improving them and by giving us a free hand in our project. It's only because of their overwhelming interest and helpful attitude, the present work has attained the stage it has.

Finally, we are grateful to our Institution and colleagues whose constant encouragement served to renew our spirit, refocus our attention and energy and helped us in carrying out this work.

Bhukya Madhu Naik
Gaddala Sathvik
Paranjay Dubey

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND LITERATURE SURVEY

This chapter includes the details of wireless ad hoc networks, fuzzy theory, our objective, platform used to implement the project, and literature review related to work done in this field.

1.1 INTRODUCTION

In this section, we briefly describe our project in which end-to-end delay of wireless ad hoc networks is implemented by the Fuzzy Inference System.

1.1.1 Challenges in Low-Light Photography with Digital Cameras

Digital cameras have illustrated a promising performance gain over recent years. Despite evolving in both hardware and image processing aspects, image sensors still suffer from quantum inefficiency in low-light conditions [3]. Due to these shortcomings, digital cameras employ long-exposure settings in inferior lighting environments. Therefore, it is frequent to encounter undesired blur artifacts while taking image shots with a hand-held setup, especially in stochastic conditions [4]. Most notably, such blind motion blurs are inevitable and substantially deteriorate the perceptual image quality [7, 38].

Table 1.1: A brief comparison between different approaches and the proposed work.

Category	Methods	Strength	Weakness
Single-shot image deblurring (a)Conventional approaches	Use 2D/3D blur-kernel estimation and empirical algorithmic approaches to restore latent sharp image	Faster inference time with low complexity, Free from obligation of training data	Failed to generalize to large blurs, Perceptual quality of restored images is unsatisfactory
Single-shot image deblurring (b) Learning-based approaches	Use machine-learning techniques to learn from diverse data samples	Capable to handle a diverse range of complex scenes, End-to-end solution without explicit blur-kernel prediction	Lacking of content awareness, Not specialized for low-light conditions
Low-light image enhancement	Enhance perceptual quality of the images captured in low-light condition	Use content-awareness and adversarial guidance to pursue plausible images	Not specialized in deblurring
Learning-based image deblurring in low-light (proposed)	Content-aware deep FPN for image deblurring in-light conditions	Content-awareness and multi-term objective function, Specialized for low-light condition, Can handle real and synthesized data	Learns from synthesized data

1.1.2 Enhancing Perceptual Quality of Blurry Images in Low-Light Conditions

Perceptual quality enhancement from degraded blurry images refers to a deconvolution operation, where the intended image comprises a blur kernel with additive sensor noises [7]. A substantial amount of noise factors can make the restoration process considerably complicated. Notably, in low-light conditions, image sensors capture the sensor noises most [3]. Also, a single-shot image deblurring process does not incorporate reference information of the motion trajectory from neighbor frames [13]. Therefore, a single-shot latent image restoration process in a low-light environment is far more challenging than a typical well-lit image deblurring process.

1.1.3 Evaluating Learning-Based Methods

In recent years, learning-based image deblurring methods [2, 7, 18, 22, 38] have drawn significant momentum in the single-shot image deblurring domain. However, most recent studies focused on the blind deconvolution process without explicitly considering the lighting conditions. Arguably, the lighting condition has a direct impact on the blur removal process. Nevertheless, to study the feasibility of the state-of-the-art (SOTA) deblurring method in low-light conditions, an initial experiment has been conducted. It is perceptible that even the SOTA single-shot image deblurring methods illustrate deficiencies and are susceptible to producing visually disturbing artifacts while removing blurs in low-light conditions.

1.1.4 Promises and Applications

Despite the unsatisfactory performance of the SOTA methods, the single-shot image deblurring in low-light conditions has numerous promises in computer vision, robot vision, autonomous driving, surveillance solutions, etc. For instance, in recent years, night photography with single-shot hand-held cameras like the smartphone has gained noteworthy interest among end-users [24]. Regrettably, motion blurs are inclined to appear most in night shots and deliver an inadmissible photography experience. Contrarily, single-shot image deblurring can enhance the perceptual quality of such vulnerable image samples. Apart from that, low-light image deblurring has numerous real-world applications in autonomous driving and surveillance solutions. Also, single-shot image deblurring can accelerate the performance of computer vision applications (e.g., segmentation, detection, recognition, classification, etc.). The widespread applicability of such image enhancement inspired this study to tackle the challenges of single-shot low-light image deblurring.

1.1.5 Introducing DarkDeblurNet and DarkShake

This study proposes a novel content-aware learning approach to address the deficiencies of low-light image deblurring. The proposed deep model appropriates channel attention [17] in a multi-level feature pyramid structure [6, 16, 18] for global image correction. Also, the proposed model utilizes a contextual gating mechanism [21, 22] to leverage spatial enhancement in a residual manner [15, 34]. The proposed model has been optimized with a multi-term loss function combining reconstruction loss, structure loss, perceptual feature loss, and adversarial guidance to perceive visually pleasurable images. Apart from that, a blur–sharp image dataset has been developed to assess the performance of deblurring methods in low-light conditions. It is worth noting that the image pairs of the proposed dataset were collected with actual hardware rather than simulating blur artifacts on synthesized data. This study denoted the proposed deep model and real-world benchmark dataset as the “DarkDeblurNet” and “DarkShake” in the rest of the sections.

1.2 Motivation

In the world of agriculture, accurate diagnosis of plant health is crucial for ensuring sustainable crop production. However, farmers and agronomists often face challenges when capturing clear images of leaves under low-light conditions, whether it’s early morning, late evening, or in shaded areas of the field. Blurry images can obscure critical details, making it difficult to identify diseases, pests, or nutrient deficiencies.

This project was born from the need to bridge the gap between technology and real-world agricultural challenges. By developing a method to deblur leaf images taken in low-light conditions, we aim to empower farmers and agricultural experts with clearer, more accurate visual data. This, in turn, will help them make better-informed decisions, leading to healthier crops and more efficient farming practices.

1.3 Objectives

This project introduces a novel deep learning model named **DarkDeblurNet**. This model addresses the challenges of single-shot image deblurring in low-light conditions by leveraging a dense-attention block and a contextual gating mechanism. And the main objectives are as follows:

- Create a content-aware deep network specialized in single-shot image deblurring in low-light conditions.
- Combine multiple losses (reconstruction, structure, perceptual feature, and adversarial loss) to achieve visually plausible images.
- Collect a blur–sharp image dataset using actual hardware to assess the performance of deblurring methods in real-world low-light conditions.
- Perform extensive experiments to demonstrate the superiority of the proposed method over state-of-the-art methods in both synthesized and real-world data.

1.4 Python

We have used Python and a Python library named PyTorch to implement our project. Python is widely used in the field of deep learning and neural networks due to its simplicity, versatility, and the availability of powerful libraries and frameworks. PyTorch, developed by Facebook's AI Research lab (FAIR), is known for its dynamic computation graph and ease of use in model development.

1.4.1 Ease of Use and Readability

Python's syntax is clear and easy to understand, making it accessible for beginners and experts alike. This readability is crucial when working on complex deep learning models.

1.4.2 Deep Learning Approaches

Deep learning techniques have also been applied to image deblurring tasks. Frameworks like TensorFlow and PyTorch can be used to develop and train deep neural networks for learning to deblur images directly from data.

1.4.3 Mathematical and Statistical Tools

Python's extensive libraries for numerical computing (e.g., NumPy, SciPy) provide essential functions and algorithms used in image processing and deblurring tasks, such as Fourier transforms and statistical modeling of blur.

1.5 LITERATURE REVIEW

This section reviews recent single-shot image deblurring methods as well as SOTA low-light image enhancement works, which are related to the proposed works.

1.5.1 Single-shot Image Deblurring

Single-shot image deblurring is known to be a challenging task for several decades. Typically, an image blur formulates as follows:

$$I_B = K(M) * I_S + N \tag{1}$$

Here, i_B represents the blurred image, I_S represents the latent sharp image, $*$ represents the convolution operation, $K(\cdot)$ represents the blur kernel confined by the motion field M, and N represents the sensor noise. Based on Eq. (1), several novel works have been proposed in the recent past. The goal of these methods was mainly to restore the latent sharp image (I_S) convoluted by a blur kernel ($K(M)$). Depending on the solving approaches, the existing works are divided into two major categories: (1) Traditional approaches and (2) Learning-based approaches.

Traditional Approaches: In modern days, image deblurring is considered a deconvolutional operation. Fergus et al. [7] are known to be one of the early adopters of this concept. In their work, they used maximum a-posteriori estimation (MAP) on a Bayesian approach to remove unknown blur that occurred due to camera shake. Considering their work as a success, many researchers developed deconvolutional methods [11, 12, 14, 19] for blind motion deblurring in later years. For example, [11] proposed using a heavy-tailed gradient distribution on a regularized probabilistic model for blur kernel estimation. In another work, [4] suggested estimating the strong edges for faster convergence. Later, [12] exerted spatial prior to ignore the small structures for high-quality kernel estimation. In another work, [3] proposed using LightStreaks hint for blur-kernel prediction and the Richardson–Lucy optimization method to remove blurs from the image captured in low-light conditions. All of these methods considered camera shake in 2D space. Some previous works presented the camera motion blur with a 3D kernel as well, where both in-plane translation and camera rotation were considered. In a novel work, [12] estimated a motion density function of the camera motion and analyzed the scope of this approximation in 3D space. Similarly, [12] parameterized a geometric model to remove undesired blurs based on the rotational velocity. Diversely, [12] used the homography blur with a fast Fourier transform to recover the corresponding sharp image. Although traditional approaches are known to be efficient in terms of computational complexity, erroneous blur-kernel prediction performed by these methods failed to generalize the diverse range of blurs removal [22]. Subsequently, the perceptual quality of restored sharp images was unsatisfactory.

Learning-based Approaches: In recent years, learning-based methods illustrated significant dominance in the single-shot image deblurring field. These methods mostly focused on learning from diverse data samples rather than applying empirical algorithmic approaches. The first encouraging attempt of utilizing machine learning in the deblurring field was made by [2], where authors introduced the concept of a multi-scale iterative blind deconvolution method. The method comprised feature extraction, blur-kernel estimation, and sharp image estimation modules. Later, [2] proposed estimating a global blur kernel through complex Fourier coefficients of a deconvolution filter and restored the latent sharp image through a non-blind EPLL method [2]. Although these non-deep learning methods are far from reality, they made a substantial influence on the research community to drive forward with learning-based approaches. Recent studies on deblurring focused on developing deep learning-based methods with end-to-end training strategies. [22] proposed a convolutional neural network (CNN) to predict the field of non-uniform motion blur and effectively restore latent sharp images. Similarly, [18] proposed a residual learning-based deep neural network architecture for large-scale image deblurring. Their network, named as “DeblurGAN,” integrated both blur kernel estimation and latent image restoration in a single model. In contrast, [21] utilized a generative adversarial network (GAN) to develop a complex deep architecture, which showed significant improvements over the prior methods. In recent years, several works [22, 25, 31, 35] focused on multi-scale feature learning with complex architectures. For example, [35] proposed a multi-stream network with a feature pyramid structure to enhance the perceptual quality of deblurred images. [25] utilized a multi-scale pyramid network for the same purpose.

1.5.2 Recent Advances in Low-light Image Enhancement

Low-light image enhancement remains a challenging task. Recent studies have been conducted to address perceptual quality enhancement in images captured in low-light conditions. For example, [5] proposed a deep network for low-light image enhancement. In their approach, they developed a dataset based on capturing images in low-light conditions and used an encoder-decoder network to enhance perceptual quality. Similarly, [15] proposed a network for enhancing low-light images with different light intensities and backgrounds. Their method integrated an autoencoder and a generative adversarial network to address various aspects of perceptual quality enhancement.

1.6 Summary

Recent advances in learning-based methods for image deblurring are discussed, showcasing their momentum in improving single-shot image deblurring but also noting their shortcomings, especially in low-light scenarios. Despite these challenges, the potential applications in various fields, including computer vision, autonomous driving, and surveillance, underscore the importance of continued research in this area. The chapter concludes by introducing the proposed model, DarkDeblurNet. In the next chapter, we delve into detailed design of the deep model, the multi-term optimization scheme, the preparation of the dataset, and the implementation process. Each section will provide a thorough explanation of the methodologies and technologies used to develop and test the proposed approach, highlighting how this innovative method addresses the challenges outlined in this chapter.

CHAPTER 2

DESIGN DETAILS AND IMPLEMENTATION

A novel learning-based low-light image deblurring method has been proposed in this study. This section details the design of the deep model, the multi-term optimization scheme, dataset preparation, and the implementation process consecutively.

2.1 Network design

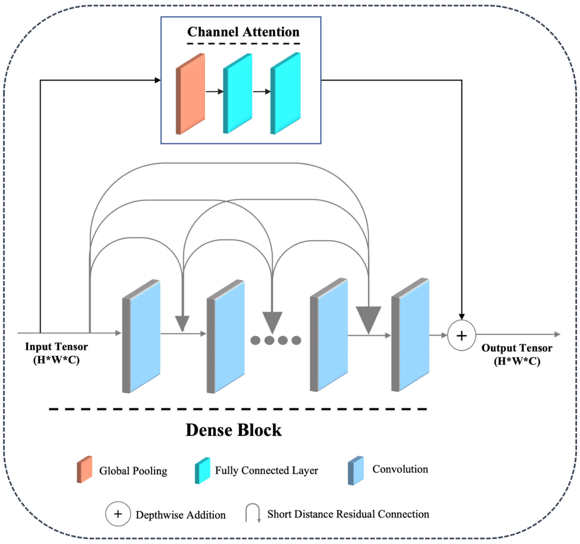
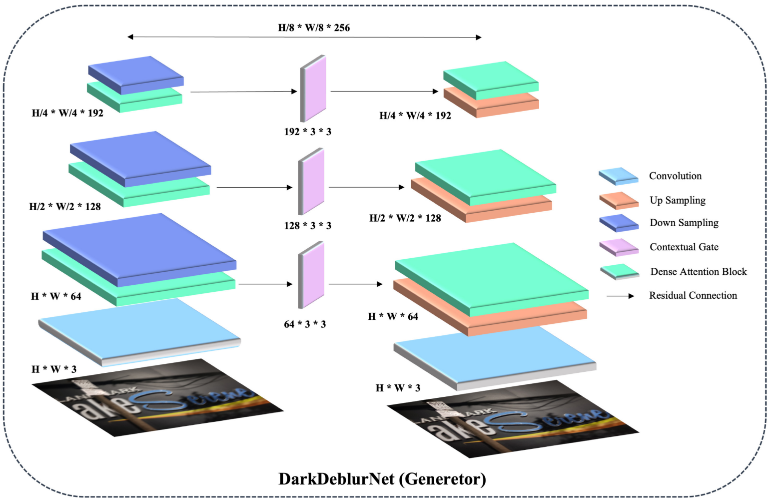
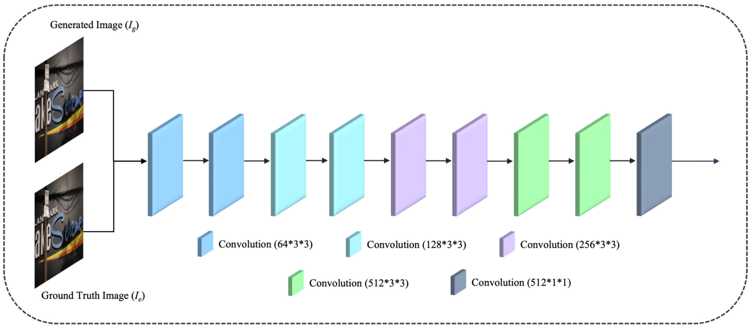


Fig. 2.1. Overview of the proposed dense-attention block. The dense-attention block combines a residual dense block and channel attention mechanism. It aims to capture the global feature interdependencies in different feature levels[71].

Typically, deblurring an image through the deconvolution process is considered a variant of image-to-image translation. Hence, the proposed DarkDeblurNet sets the aim of deblurring in low-light conditions as $F : I_B \rightarrow I_D$, where the mapping function F learns to generate a sharper image I_D from a blurry input I_B such that $I_D \in [0, 1]^{H\times W\times 3}$.



(a)



(b)

Fig. 2.2. The overview of the proposed DarkDeblurNet. The proposed network incorporates a novel dense-attention block and contextual gating mechanism in a feature pyramid structure. Also, it follows the principle of generative adversarial networks. (a) The architecture of the generator. (b) The architecture of the discriminator[71].

H and W represent the height and width of the input as well as output images.

As Fig 2.2 illustrates, the proposed DarkDeblurNet incorporates the concept of GAN. Here, the generator of the proposed DarkDeblurNet design utilizes the advantages of a feature pyramid structure with a novel dense-attention block. Additionally, the features learned at different feature levels are propagated with a contextual gating mechanism to leverage spatial awareness.

Dense-attention block: Fig. 2 depicts the overview of the proposed dense-attention block. The key idea of a dense-attention block is to go deeper and wider while learning global feature interdependencies. The proposed dense-attention block is developed by taking inspiration from the residual-dense block [19] and squeeze-excitation networks [17].

Typically, a dense block connects all the previous layers as:

$$X_l = H_l([X_0, X_1, \ldots, X_{l-1}]) \tag{2}$$

where $[X_0, X_1, \ldots, X_{l-1}]$ refers to the concatenated feature maps obtained through $l = 5$

number of convolutional layers [19]. This study sets the number of convolutional layers l as $l = 5$. All convolution layers have kernels of size 3×3, stride of 1, padding of 1, and are activated with the LeakyReLU function.

On the other hand, global feature interdependencies can be perceived by applying global average pooling [6, 17], where channel-wise squeezed descriptors $Z \in R^C$ of an input feature map can be calculated as:

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U_c(i, j) \tag{3}$$

Here, Z_c, $H \times W$, and U represent global average pooling, spatial dimensions, and the feature map. To pursue aggregated global dependencies, a gating mechanism is applied as follows:

$$W = \sigma(W_E(\delta(W_S(Z)))) \tag{4}$$

Here, σ and δ denote the sigmoid and ReLU activation functions applied after W_E and W_S convolutional operations.

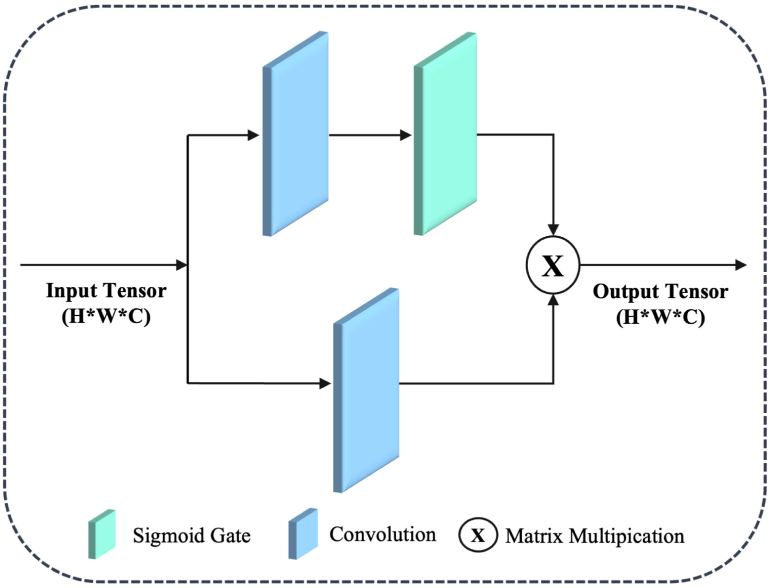


Fig.3. Overview of the contextual gate. The contextual-gate aims to propagate the spatial dependencies between different feature levels [71].

The final channel attention map is achieved by rescaling the feature map as follows:

$$\hat{S}_c = W_c \cdot S_c \tag{5}$$

Here, W_c and S_c denote the scaling factor and feature map. Typically, squeeze-and-excitation-based channel attention $\hat{S}_c$ is calculated over the output of convolutional blocks (i.e., residual block, dense block, etc.) [17]. However, this study proposes calculating the squeeze-and-excitation descriptors from the input of the convolutional block and propagating them as a residual connection to perceive long-distance channel-wise attention.

The final output of the dense-attention block is obtained as follows:

$$D_a = X_l + \hat{S}_c \tag{6}$$

Contextual gate: Fig. 3 illustrates the overview of the context gate. Here, the contextual gate aims to propagate only the important features [21]. Unlike a typical residual connection [16], the context gate does not pass trivial features from the lower level. The context gate is obtained as follows:

$$G_{m,n} = \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{W}_g \cdot I \tag{7}$$

$$F_{m,n} = \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{W}_f \cdot I \tag{8}$$

$$O_{m,n} = \phi(G_{m,n}) \odot \delta(F_{m,n}) \tag{9}$$

Here, ϕ and δ represent the LeakyReLU and sigmoid activations. W_g and W_f denote convolutional operations.

Level Transition: The proposed DarkDeblurNet traverses different feature dimensions (i.e., upscaling or downscaling) with convolutional operations. The downsampling operation is obtained as follows:

$$F_{\downarrow} = H_{\downarrow}(X_0) \tag{10}$$

Here, $H_{\downarrow}$ represents downsampling through convolutional operations with a kernel size of 3×3, stride of 2, and padding of 1.

Conversely, the upscaling is obtained as follows:

$$F_{\uparrow} = H_{\uparrow}(X_0) \tag{11}$$

Here, $H_{\uparrow}$ represents pixel shuffle convolution operations [5], activated with a PReLU function. Pixel shuffle convolution aims to avoid checkerboard artifacts [5, 20].

Conditional Discriminator: The proposed DarkDeblurNet appropriates the concept of adversarial guidance. This study adopts a well-established variant of GAN known as conditional GAN (cGAN) [11, 12]. Therefore, the goal of the discriminator is set to maximize $E_{X,Y}[\log D(X, Y)]$. The overall discriminator architecture can be considered a stacked Convolutional Neural Network (CNN), where all layers before the output layer include convolutional kernels of size 3×3, followed by batch normalization and activated with a swish function. The feature depth of the discriminator starts from 64 channels. Every $(2n-1)$-th layer increases the feature depth and reduces the spatial dimension by a factor of 2.1 Here, the spatial dimension is reduced by applying a stride of 2. The output from the final layer is obtained with a convolution operation, where the layer comprises a kernel size of 1×1 and is activated by a sigmoid function.

2.2 Objective function

The proposed DarkDeblurNet learns to deblur with a sophisticated mapping function F with parameterized weights W. Given the training set $\{I_t^B, I_t^S\}_{t=1}^P$ consisting of P image pairs, the training process aims to minimize the objective function described as follows:

$$W^* = \arg\min_W \frac{1}{P}\sum_{t=1}^{P} \mathcal{L}_T(F(I_t^B), I_t^S) \tag{12}$$

Here, L_T denotes the proposed multi-term loss for single-shot image deblurring in low-light conditions. The goal of this multi-term loss is to improve the perceptual quality (details, texture, color, structure, etc.) of a given blurry image.

Reconstruction loss: A pixel-wise loss is adopted to perceive coarse-to-refine reconstruction. Typically, an L1 or L2 distance is used as a pixel-wise loss function. However, among these, L2-loss is directly related to PSNR and tends to produce smoother images [69]. Therefore, an L1 objective function has been considered as the reconstruction loss in this study, calculated during the training phase as:

$$\mathcal{L}_R = \|I^S - I^D\|_1 \tag{13}$$

Here, I^D and I^S represent the output obtained through $F(I^B)$ and the reference sharp image.

Structure loss: One drawback of deblurring methods is to produce structural distortion. To address this limitation, this study proposes using a structural similarity loss as one of the objective functions. Previous studies have reported that SSIM loss works well with L1-loss to improve structural deficiencies [69]. Thus, SSIM loss has been utilized as the structural loss and calculated as:

$$\mathcal{L}_S = SSIM(I^S, I^D) \tag{14}$$

Here, a multi-scale variant of SSIM loss has been used.

Perceptual Feature loss: The perceptual loss is introduced based on activation maps produced by ReLU layers of a pre-trained VGG-19 network[6, 18, 20, 23]. Instead of measuring per-pixel differences between two images, this loss encourages images to have similar feature representations to achieve similar perceptual quality. It is noted that the perceptual loss works best with the L1 norm extracted from the top-layer [18,20]. Otherwise, the perceptual feature loss may produce inconsistent brightness and visual artifacts. However, previous studies on deblurring [31, 32] utilizing perceptual loss did not incorporate this basic principle. Therefore, this study defines the perceptual loss as perceptual feature loss and formulates it as:

$$\mathcal{L}_F = \frac{1}{H_j \times W_j \times C_j} \|\psi_j(I^S) - \psi_j(I^D)\|_1 \tag{15}$$

Here, ψ and j denote the pre-trained VGG network and the j-th layer.

Adversarial loss: GANs have demonstrated superior performance in producing sharper and realistic images by comparing generated and reference images [69]. They are also capable of recovering textures from smoother inputs [18,20]. The cGAN loss used in this study minimizes the cross-entropy loss function as follows:

$$\mathcal{L}_G = -\sum_t \log D(I^D, I^S) \tag{16}$$

Here, D denotes the conditional discriminator used as a global critic.

Multi-term loss: The multi-term objective $\mathcal{L}_T$ is obtained as follows:

$$\mathcal{L}_T = \mathcal{L}_R + \mathcal{L}_S + \lambda_F \cdot \mathcal{L}_F + \lambda_G \cdot \mathcal{L}_G \tag{17}$$

Here, λ_F and λ_G represent the loss regulators, set as $\lambda_F = 1e^{-2}$ and $\lambda_G = 1e^{-4}$ to stabilize the adversarial training.

2.3 Implementation details

The DarkDeblurNet has been implemented using the PyTorch framework (PyTorch, 2016). The proposed generator leverages four consecutive feature levels (e.g., 64, 128, 192, 256, etc.) in a multi-level feature pyramid structure. Additionally, the contextual gates of the proposed generator utilize filter sizes of 64, 128, and 192 to match the depth of feature levels. Network details are depicted in Fig. 1.

The Adam optimizer [18] has been utilized to optimize both the generator and discriminator of the proposed DarkDeblurNet. The hyperparameters for both networks are tuned as $\beta_1 = 0.9, \beta_2 = 0.99$, and learning rate = $1e^{-4}$. The DarkDeblurNet was trained with image patches of size $128 \times 128 \times 3$ for 100,000 steps with a constant batch size of 16. All experiments were conducted on a local machine.

2.4 Summary

This chapter introduced a novel learning-based low-light image deblurring method named DarkDeblurNet. The network design incorporates a dense-attention block, which combines a residual dense block with a channel attention mechanism to effectively capture global feature interdependencies at different feature levels. The network follows the principles of generative adversarial networks (GANs) and includes a feature pyramid structure enhanced by a contextual gating mechanism, which aids in propagating spatial dependencies between different feature levels. The training process involved extensive experimentation using image patches to optimize the network's performance. In the upcoming chapter, the effectiveness of the proposed DarkDeblurNet will be evaluated. This chapter will present the experimental results, comparing the performance of DarkDeblurNet with existing state-of-the-art methods.

CHAPTER 3

RESULTS AND DISCUSSION

In this section, we present the evaluation of the proposed DarkDeblurNet across various benchmark datasets and compare its performance with state-of-the-art (SOTA) methods. The comprehensive analysis demonstrates the effectiveness of our approach, particularly in challenging low-light conditions.

3.1 Comparison with existing methods

The proposed method is trained with synthesized and dynamic deblurring datasets. Nevertheless, the feasibility of the proposed DarkDeblurNet is verified with synthesized and real-world blurry images. The results obtained from the proposed method are compared with the SOTA single-shot learning-based image deblurring methods. The following SOTA methods were selected for the comparison: (1) DeepDeblur [46], (2) Scale-Recurrent-Network (SRN) [38], and (3) DeblurGANv2 [19]. It is worth noting that none of the existing methods are developed to handle single-shot image deblurring in low-light conditions. For a fair comparison, all target models are retrained with low-light blurry images and suggested hyper-parameters. Apart from that, we also studied the feasibility of non-learning deblurring methods with diverse data samples. The following deblurring methods have been incorporated throughout the proposed study: (1) LightStreaks [18] and (2) Dark Channel Prior [4].

3.1.1 Low-light deblurring (Synthesized data)

Table 2 depicts the performance of the evaluated methods. The results were obtained with different synthesized datasets: (1) ExDark [1], (2) Lai Dataset [46], and (3) Kohler Dataset [38]. It is worth noting that handheld image deblurring is considered a special case of dynamic deblur. However, to study further, night shot images from Lai and Kohler datasets have been included. The performance of deblurring methods has been summarized with three evaluation metrics: PSNR, SSIM, and DeltaE [10]. The DeltaE metric intends to evaluate the color and brightness consistency obtained through the deblurring methods. As Table 2 shows, the proposed DarkDeblurNet illustrates a significant improvement over existing methods while removing blurs in challenging low-light conditions. Notably, the performance gain is consistent in all evaluation metrics. Also, the learning-based methods are consistent on diverse data samples compared to their non-learning counterparts. The proposed method demonstrates its superiority over existing methods. It can produce visually pleasing images along with recovering far more details than its counterparts. In a nutshell, the proposed method can reduce artifacts and produce natural results.

3.1.2 Low-light deblurring (Real data)

Fig. 3.1 Testing Dataset Sample (a) Original. (b) Generated Re-blurred. (c) Resultant De-blurred image

Although all models were trained with synthesized data, it is salient to observe the performance of deep models in real-world data samples. Therefore, the performance of deep models has been studied on the proposed DarkShake dataset. Quantitative comparison between SOTA methods and the proposed DarkDeblurNet. A total of 100 image pairs from the DarkShake dataset were used to calculate the mean PSNR, SSIM, and DeltaE metrics. In all evaluation metrics, the DarkDeblurNet illustrated the consistency and outperformed the existing methods.

Table 3.1 illustrates the performance of deep models on the proposed DarkShake dataset. Here, the quantitative results are calculated on 25 blur–sharp image pairs. It is visible that the proposed model can outperform the existing methods in real-world data as well. It shows an improvement of 0.59 dB in PSNR metrics, 0.018 in SSIM metrics, and 0.17 in DeltaE metrics over SOTA methods. On real-world data, the pro-

Table 3.1: Quantitative comparison of SOTA methods and DarkDeblurNet on the DarkShake dataset

Method	PSNR (↑)	SSIM (↑)	DeltaE (↓)
LightStreaks [18]	22.40	0.7299	5.18
DarkChannel [4]	24.32	0.8214	3.58
DeepDeblur [46]	24.54	0.7349	3.98
SRN [38]	24.80	0.7410	3.92
DeblurGANv2 [19]	23.32	0.7727	4.55
DarkDeblurNet	**25.39**	**0.8401**	**3.75**

posed DarkDeblurNet demonstrates consistency and depicts its superiority over SOTA methods. It is noteworthy that the proposed method can recover more details without producing any visually disturbing artifacts.

3.1.3 Well-lit deblurring

The feasibility of the proposed method has been studied in well-lit conditions. Subsequently, learning-based deblurring methods have been retrained with GoPro [46] and REDS [46] datasets. Later, the performance of all deblurring methods has been evaluated with testing samples from the same datasets. Table 3.2 illustrates the performance of SOTA deblurring methods on well-lit deblurring. It can be seen that the proposed method illustrates consistency in well-lit conditions as well. In addition to the quantitative evaluation, the performance of the proposed method has been confirmed with a qualitative study. It is noticeable that the proposed method can produce sharper images without producing any visually disturbing artifacts.

Table 3.2: Quantitative comparison between SOTA methods on the GoPro and REDS datasets. The DarkDeblurNet can outperform the existing methods for well-lit image deblurring.

Method	**PSNR (GoPro)**	**SSIM (GoPro)**	**DeltaE (GoPro)**	**PSNR (REDS)**	**SSIM (REDS)**	**DeltaE (REDS)**
LightStreaks	21.25	0.7378	6.26	22.11	0.7363	5.58
DarkChannel	24.75	0.8730	3.22	23.08	0.8059	3.99
DeepDeblur	26.03	0.7936	3.29	27.17	0.7958	3.18
SRN	26.38	0.8119	3.13	27.77	0.8148	3.04
DeblurGANv2	24.33	0.8208	4.39	25.30	0.8164	4.47
DarkDeblurNet	**27.39**	**0.8281**	**2.75**	**28.72**	**0.8308**	**2.66**

3.2 Applications

Digital cameras can produce blurry images due to numerous factors such as long-exposure settings, faster-moving objects, handshakes while holding the camera, etc., explicitly, with a single-shot setup. Regrettably, any extent of motion blurs drive camera hardware to capture the target scene in an unusable form. Apart from being a significant application of computer vision, single-shot image deblurring can accelerate the performance of numerous real-world applications such as segmentation [15; 24], detection [19], recognition [31], microscopy [13], facial expression analysis [1], 3D image analysis [14], medical imaging [20], space observation [37], etc. Considering such widespread real-world applications, top computer vision societies like the Computer Vision Foundation (CVF) arrange numerous competitions and challenges in their top-tier conferences (i.e., CVPR, ECCV, ICCV, etc.) to encourage the development of deblurring solutions. However, most existing tracks and works for developing deblurring solutions are dedicated to normal lighting conditions. Contrarily, camera systems are commonly affected by motion blurs in low-light conditions due to the utilization of slower shutter speeds. This study addresses such inevitable motion blurs in challenging low-light conditions. Among the countless applications, this study illustrates the implication of single-shot low-light image deblurring in the three most widely used scenarios throughout the experiments. Here, the DarkDeblurNet is applied to a blurry image and then inferred with a SOTA segmentation method known as Mask R-CNN [15]. It is evident that the proposed method dramatically improves the performance of the segmentation method by allowing it to segment more objects than it performs with the blurry image. For object detection, a SOTA method known as YOLOv3 [19] is used. It is clear that the performance of the detection method is improved with a sharper image, which is restored through the proposed method. For text recognition, the Tesseract OCR framework [31] is applied.

3.3 Ablation study

The components proposed in this study have been verified through sophisticated experiments.

Table 3.3: Ablation study of the proposed method. The importance of the proposed component is verified with sophisticated experiments. Each component plays a significant role in improving the performance of the DarkDeblurNet on both ExDark and DarkShake datasets.

Model	**ExDark**			**DarkShake**		
	PSNR ↑	**SSIM ↑**	**DeltaE ↓**	**PSNR ↑**	**SSIM ↑**	**DeltaE ↓**
DarkDeblurNet$_{\text{Base}}$	25.37	0.7730	3.97	22.29	0.6711	5.86
DarkDeblurNet$_{\text{CA}}$	28.29	0.7905	3.30	22.47	0.6238	6.26
DarkDeblurNet$_{\text{CG}}$	31.56	0.8667	2.07	24.24	0.6904	3.96
DarkDeblurNet	**34.56**	**0.9146**	**1.78**	**25.39**	**0.8401**	**3.75**

As shown in Table 3.3, ablation experiments started with a baseline network architecture using a traditional dense block. The baseline variant replaced the multi-term objective function with a simple L1-norm. Subsequently, proposed components such as channel attention (CA) with short-distance residual connection, contextual gate (CG), and multi-term loss (ML) function were integrated into the baseline architecture in a modular manner. In Table 3.3, the baseline network with the traditional dense block is denoted as DarkDeblurNet_Base, DarkDeblurNet_Base with residual CA as DarkDeblurNet_CA, and DarkDeblurNet_CA with CG as DarkDeblurNet_CG. The network variant incorporating attention mechanisms and multi-term loss function is denoted as DarkDeblurNet.

Table 3.3 demonstrates that each proposed component significantly contributes to performance gains. Particularly, the performance gap between the proposed DarkDeblurNet and its baseline variant (DarkDeblurNet_Base) is substantial. The proposed components, such as short-distance residual connection with CA, CG, and multi-term loss, show a significant performance improvement of 9.19 dB in PSNR metrics, 0.1416 in SSIM metrics, and 2.19 in DeltaE metrics for synthesized data. Similarly, consistent performance gains are observed on real-world data, with improvements of 2.51 dB in PSNR metrics, 0.0643 in SSIM metrics, and 1.96 in DeltaE metrics.

It is evident that each of the proposed components is vital to achieving superior performance. The component-wise validation also verifies the efficacy of each element in the proposed method.

3.4 Discussion

This section discusses the key findings, limitations, and future scope of the proposed work. Network Architecture: The proposed DarkDeblurNet is presented as an FPN with a dense-attention block and contextual gates to learn feature interdependency precisely. Despite being deeper and wider, the complexity of the proposed network is calculated as 16.0 G floatingpoint operations (FLOPs). It is worth noting, the proposed network takes only 0.02 s and 0.04 s to infer on a single-shot image with a spatial dimension of $256 \times 256 \times 3$ and $512 \times 512 \times 3$, which is three times faster than the existing learning-based methods [46, 70]. Also, The proposed network is fully convolutional. Hence, it can take arbitrary image dimensions for inference. Limitations: One of the limitations of the proposed study is identified as a lack of real-world low-light training data. Notably, the synthesized low-light blurry images differ from real-world blurry images and deteriorate the performance in real-world image samples. In some extreme cases, it can also compel the proposed network to demonstrate unwanted structural distortion. Due to hardware limitations, the proposed study resized the large-dimensional DarkShake image samples into the smaller spatial dimension to fit in the GPU memory. Thus, the compression artifacts can be observed in visualizing real-world examples.

Future Scope: The limitations of the proposed study lead to an interesting future direction. The findings of this study indicate that learning from real-world blur–sharp image pairs can be helpful for image deblurring in low-light conditions. Therefore, the proposed DarkShake dataset can be extended in such a way that it can use for training purposes. In another way, low-light image deblurring can be performed in an unsupervised manner. It can also resolve the training data limitations apart. It would be a very challenging but exciting task to combine image deblurring with other low-light image enhancement tasks (i.e., lowlight to well-lit mapping). The performance of the proposed DarkDeblurNet can be studied to find the feasibility of joint image deblurring and low-light to well-lit mapping in the foreseeable future.

3.5 Summary

In this chapter, we evaluated the performance of the proposed DarkDeblurNet on various benchmark datasets and compared it against several state-of-the-art (SOTA) methods. The comparison with existing methods demonstrated the superior performance of DarkDeblurNet, particularly in low-light conditions, where traditional methods typically struggle. The proposed model outperformed the other methods across multiple evaluation metrics such as PSNR, SSIM, and DeltaE, both on synthesized and real-world data. In well-lit conditions, DarkDeblurNet continued to show consistent improvements over existing approaches. The next chapter will summarize the key findings of this study, reflecting on the strengths and limitations of the proposed DarkDeblurNet. We will also discuss potential future directions for this research.

CHAPTER 4

CONCLUSION

This study presented a learning-based single-shot image deblurring method specialized in low-light environments. The proposed DarkDeblurNet facilitated a novel dense-attention block in the feature pyramid structure for global image correction. Also, a contextual gating mechanism was used to propagate spatially enhanced features between different feature levels. The proposed architecture was optimized with a multi-term objective function. The feasibility of the proposed study was verified with sophisticated experiments. The proposed method illustrated superiority over SOTA methods in both synthesized and real-world testing. Apart from that, this study introduced a real-world blur–sharp image dataset, which can be used for low-light deblurring evaluations.

4.1 Limitations and future work

While the proposed method demonstrates significant improvements over existing techniques, there are certain limitations. The method's performance may degrade with extremely noisy images or in scenarios where the blur intensity is beyond the model's training data range. Moreover, the current model primarily focuses on single-shot deblurring and might not perform optimally for sequences of blurry images. Future work could explore the following aspects:
(1) Incorporating temporal information for video deblurring to handle sequences of blurry images effectively,
(2) Enhancing robustness against extreme noise and large blur sizes.
(3) Investigating the application of the proposed method to different domains, such as medical imaging or satellite imagery.

REFERENCES

[1] G. Ali, A. Ali, F. Ali, U. Draz, F. Majeed, and S. Yasin. Artificial neural network based ensemble approach for multicultural facial expressions analysis. *IEEE Access*, 8:134950–134963, 2020.

[2] A. Chakrabarti. A neural approach to blind motion deblurring. In *European Conference on Computer Vision*, pages 221–235. Springer, 2016.

[3] P. Chatterjee, N. Joshi, S. B. Kang, and Y. Matsushita. Noise suppression in low-light images through joint denoising and demosaicing. In *CVPR*, pages 321–328. IEEE, 2011.

[4] S. Cho and S. Lee. Fast motion deblurring. *ACM SIGGRAPH Asia 2009 papers*, pages 1–8, 2009.

[5] X. Cortés and F. Serratosa. An interactive method for the image alignment problem based on partially supervised correspondence. *Expert Systems with Applications*, 42(1):179–192, 2015.

[6] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11065–11074, 2019.

[7] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *ACM SIGGRAPH 2006 papers*, pages 787–794. ACM, 2006.

[8] S. Gai and Z. Bao. New image denoising algorithm via improved deep convolutional neural network with perceptive loss. *Expert Systems with Applications*, 138:112815, 2019.

[9] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics*, 36(4):1–12, 2017.

[10] C. Gómez-Polo, M. P. Muñoz, M. C. L. Luengo, P. Vicente, P. Galindo, and A. M. M. Casado. Comparison of the cielab and ciede2000 color difference formulas. *The Journal of Prosthetic Dentistry*, 115(1):65–70, 2016.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, and et al. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[12] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In *European Conference on Computer Vision*, pages 171–184. Springer, 2010.

[13] L. Han and Z. Yin. Refocusing phase contrast microscopy images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 65–74. Springer, 2017.

[14] M. Hanif, R. A. Naqvi, S. Abbas, M. A. Khan, and N. Iqbal. A novel and efficient 3d multiple images encryption scheme based on chaotic systems and swapping operations. *IEEE Access*, 8:123536–123555, 2020.

[15] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969, 2017.

[16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[17] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[18] Z. Hu, S. Cho, J. Wang, and M.-H. Yang. Deblurring low-light images with light streaks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3382–3389, 2014.

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[20] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3277–3285, 2017.

[21] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, and et al. Enlightengan: Deep light enhancement without paired supervision. In *arXiv preprint arXiv:1906.06972*, 2019.

[22] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997.

[23] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.

[24] K. Khan, Z. A. Khan, M. Sharif, M. Ahmed, A. Gani, S. U. Khan, and et al. Face segmentation: A journey from classical to deep learning paradigm, approaches, trends, and directions. *Computers, Materials & Continua*, 63(3):1017–1045, 2020.

[25] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.

[26] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1637–1645, 2016.

[27] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[29] N. H. Le, T. L. Le, T. D. Pham, and H. T. Nguyen. Underwater image restoration methods: A comprehensive survey. *Journal of Visual Communication and Image Representation*, 65:102668, 2019.

[30] C. Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeply-supervised nets. In *Artificial Intelligence and Statistics*, pages 562–570. PMLR, 2015.

[31] Y. Li, C. Chen, Y.-K. Lai, and K. Y. K. Wong. Interactive deep colorization with simultaneous classification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 835–851, 2018.

[32] Y. Li, S. Pirk, E. Belyaev, and H.-P. Seidel. Exploiting spatial redundancy for image denoising with convolutional neural networks. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019.

[33] D. Liu, B. Wen, Y. Fan, D.-Y. Yeung, and G. Hua. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, 2017.

[34] D. Liu, B. Wen, and T. Sim. Progressive and multiscale deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[35] G. Liu, F. Lin, J. Zhang, and J. Liu. Partial convolution based padding. *arXiv preprint arXiv:1811.11718*, 2018.

[36] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia. Semantic image segmentation via deep parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 801–810, 2016.

[37] X. Liu, C. Liu, X. Zhang, Q. Liu, X. Ye, L. Niu, and S. Zhang. Deep learning for remote sensing image classification: a survey. *Remote Sensing*, 11(18):2057, 2019.

[38] Z. Liu, L. Zhang, Y. Wei, S. Liu, X. Cao, X. Shen, and T. Huang. Multiscale gated fusion network for image deblurring. *IEEE Transactions on Image Processing*, 29:8777–8790, 2020.

[39] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[40] Z. Luo, J. Gu, Y. Guo, and Y. Zheng. Multispectral image denoising via deep convolutional neural networks. *IEEE Transactions on Image Processing*, 29:3775–3789, 2020.

[41] X. Mao, C. Shen, and Y.-B. Yang. Image restoration using convolutional autoencoders with symmetric skip connections. *arXiv preprint arXiv:1606.08921*, 2016.

[42] S. Mehta, M. Rastegari, and L. G. Shapiro. Optimizing convolutional neural networks for mobile platforms. *Journal of Signal Processing Systems*, 92(4):409–424, 2020.

[43] M. Micheletti, S. Zuffi, A. Monti, A. Just, E. Demirović, S. Zuffi, and et al. Diverse image-to-image translation via disentangled representations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2820–2829, 2019.

[44] N. Miller and M. N. Do. No-reference image quality assessment using multi-scale gradient magnitude similarity deviation. *IEEE Transactions on Image Processing*, 29:6327–6342, 2020.

[45] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1261–1264. IEEE, 2012.

[46] S. Nah, T. H. Kim, and K. M. 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.

[47] R. A. Naqvi, S. Khatri, S. Hussain, M. Saleem, and A. Rehman. A novel approach for recovering underwater image by eliminating backscatter using optical configuration and machine learning algorithm. *Ocean Engineering*, 137:142–156, 2017.

[48] Rizwan Ali Naqvi, Amir Haider, Hak Seob Kim, Daesik Jeong, and Seung-Won Lee. Transformative noise reduction: Leveraging a transformer-based deep network for medical image denoising. *To be published*, 2024. Department of AI and Robotics, Sejong University; Korea Agency of Education, Promotion and Information Service in Food, Agriculture, Forestry and Fisheries; Division of Software Convergence, Sangmyung University; School of Medicine, Sungkyunkwan University. †These authors contributed equally to this work.

[49] L. M. Nguyen, P. Gamba, T. D. Pham, Y. Li, P. M. Atkinson, M. A. Cho, and et al. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 8(2):8–36, 2020.

[50] J. Pan, Z. Hu, H. Su, and M.-H. Yang. Learning a discriminative prior for blind image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6576–6585, 2018.

[51] J. Pan, X. Wu, S. Long, C. Zhang, and Y. Qiao. Deep convolutional neural network based underwater image enhancement: A survey. *IEEE Access*, 8:149539–149553, 2020.

[52] X. Pan, J. Shi, P. Luo, and X. Tang. A2-nets: Double attention networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 261–270, 2019.

[53] S. Park, J. Jeong, J. Kim, J. Park, and H. Kim. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, 37(1):44–64, 2020.

[54] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[55] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99, 2015.

[56] Eli Schwartz, Raja Giryes, and Alex Bronstein. Deepisp: Toward learning an end-to-end image processing pipeline. *IEEE Transactions on Image Processing*, 27(10):5140–5152, 2018.

[57] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1874–1883, 2016.

[58] W. Shi, X. Li, L. Yi, X. Qi, J. Wang, and J. Jia. Robust single-image dehazing via estimation of transmission and atmospheric light. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5115–5123, 2016.

[59] J. Shin, R. Willett, and P. Marziliano. Interferometric phase estimation using deep learning. *IEEE Transactions on Image Processing*, 28(2):846–861, 2018.

[60] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[61] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8934–8943, 2018.

[62] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014.

[63] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8174–8182, Salt Lake City, UT, USA, 2018. IEEE. The Chinese University of Hong Kong, YouTu Lab, Tencent, Megvii Inc.

[64] R. Timofte, V. De Smet, and L. Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126, 2014.

[65] T. Wang, M. Yang, C. Liu, and Z. Xu. Deep single image rain removal via multi-scale convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3984–3993, 2019.

[66] X. Wang, K. Yu, C. Dong, and C. C. Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 606–615, 2018.

[67] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[68] Y. Wang, H. Liu, X. Hu, W. Liao, and X. Liu. A retrieval-based method for remote sensing image quality assessment. *Remote Sensing*, 12(15):2391, 2020.

[69] Z. Wang, Y. Guo, L. Xu, J. Zhu, and Y. Liu. Image enhancement using local contrast optimization and multi-scale processing based on singular value decomposition. *IEEE Access*, 8:180268–180279, 2020.

[70] J. Wei, C. Zhang, K. Yu, F. Wen, and X. Chang. Deep region-based image deblurring. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 204–220, 2018.

[71] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9):2108–2123, 2018.