IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 February 2023, accepted 24 February 2023, date of publication 1 March 2023, date of current version 6 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3250616

TOPICAL REVIEW

A Comprehensive Review of Deep Learning-Based Real-World Image Restoration

LUJUN ZHAI[1], (Student Member, IEEE), YONGHUI WANG[2], (Senior Member, IEEE), SUXIA CUI[1], (Senior Member, IEEE), AND YU ZHOU[3]

[1]Department of Electrical and Computer Engineering, Prairie View A&M University, Prairie View, TX 77446, USA
[2]Department of Computer Science, Prairie View A&M University, Prairie View, TX 77446, USA
[3]CSAT Solutions LP, Houston, TX 77064, USA

Corresponding author: Yonghui Wang (yowang@pvamu.edu)

This work was supported in part by the United States NSF under Grant OAC-1827243 and Grant OAC-1925764, in part by the U.S. Department of Education under Grant P120A180114, and in part by the Princeton Alliance for Collaborative Research and Innovation (PACRI) under Grant PVAMU-01.

ABSTRACT Real-world imagery does not always exhibit good visibility and clean content, but often suffers from various kinds of degradations (e.g., noise, blur, rain drops, fog, color distortion, etc.), which severely affect vision-driven tasks (e.g., image classification, target recognition, and tracking, etc.). Thus, restoring the true scene from such degraded images is of significance. In recent years, a large body of deep learning-based image processing works has been exploited due to the advances in deep neural networks. This paper aims to make a comprehensive review of real-world image restoration algorithms and beyond. More specifically, this review provides overviews of critical benchmark datasets, image quality assessment methods, and four major categories of deep learning-based image restoration methods, i.e., based on convolutional neural network (CNN), generative adversarial network (GAN), Transformer, and multi-layer perceptron (MLP). The paper highlights the latest developments and advances in each category of network architecture to provide an up-to-date overview. Moreover, the representative state-of-the-art image restoration methods are compared visually and numerically. Finally, for real-world image restoration, the current situations are objectively assessed, challenges are discussed, and future directions and trends are presented.

INDEX TERMS Image restoration, denoising, deblurring, deraining, dehazing, super-resolution, image quality assessment, benchmark datasets, review.

I. INTRODUCTION

In recent years, computer vision-based autonomous systems, such as autonomous driving, underwater robotics, video surveillance, and medical imaging, have been widely used [1]. The clarity of the images captured by cameras directly affects the performance of these autonomous systems. However, the acquired images in the real world are not always clear and may suffer from various kinds of degradations, as they are usually taken in complicated situations: bad weather conditions, underwater environments, uneven illumination, moving cameras, etc. For example, images taken by surveillance cameras and medical imaging equipment usually exhibit low-resolution [2]; images taken by moving cameras tend to have motion blur [3], [4]; underwater images have color distortions and noises [5], [6], [7]; images taken in hazy, rainy, or foggy weathers contain different levels of intensity blurs and noises [8]. Such image degradations cause a severe performance drop of visual systems in segmentation, detection, and target tracking [9], [10], [11], [12]. Therefore, it is critical to develop efficient image restoration (IR) algorithms to enhance the environmental adaptability of the visual systems. However, as a typical ill-posed inverse problem, real-world image restoration remains extremely challenging.

In general, image restoration is the process of recovering a high-quality image with good visibility and clean content from a degraded image. As presented in Fig. 1, conventionally, IR techniques in the low-level vision domain

The associate editor coordinating the review of this manuscript and approving it for publication was Yi Zhang.

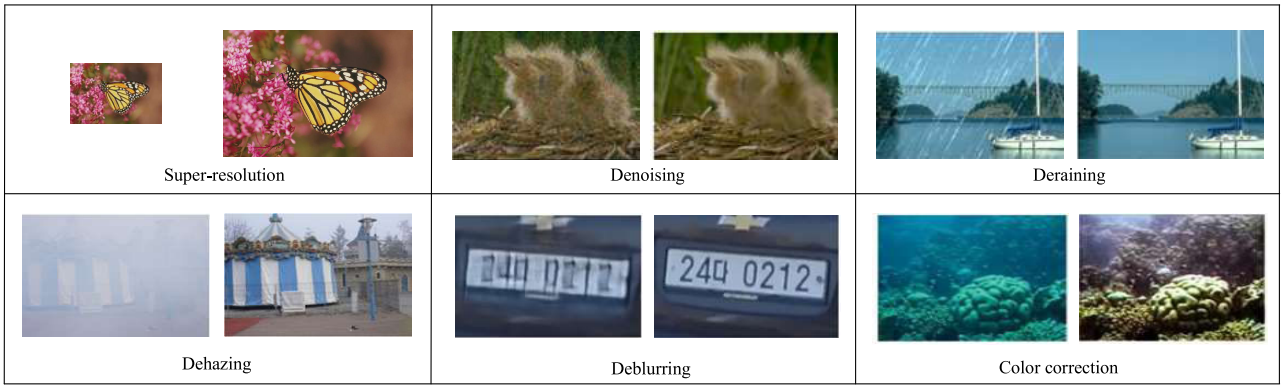


FIGURE 1. Categories of low-level image restoration.

can be grouped into six main categories according to the type of degradation involved in the image or video to be processed, i.e., super-resolution (SR), denoising, deraining, dehazing, deblurring, and color correction. Specifically, SR aims to reconstruct a high-resolution image from one or more low-resolution images [13]. Draining (dehazing, deblurring, denoising) is the task of raindrop (haze, blur, noise) removal [11], [14], [15]. Color correction deals with color distortion, especially for underwater images [16]. While typically there are many different branches or classes of IR to handle different types of degradation, many new emerging models can handle multiple tasks after being re-trained on different datasets, owing to the powerful learning prior capabilities of deep neural networks and publicly available benchmark datasets for image processing.

Considering the remarkable progress of IR has been made by deep learning models, several researchers [6], [9], [17], [18], [19] reviewed recent deep learning-based IR techniques and the primary difference among them is the categorization. In other words, each of these works examined the IR problem from a single perspective. For example, Chen et al. [17] reviewed the image restoration methods, datasets, and assessment metrics for real-world single image super-resolution (SISR). Xu et al. [9] surveyed video and image defogging algorithms and image quality assessment methods for the defogged images. Thakur et al. [20] focused on image de-noising analysis and compared different types of denoisers on benchmark datasets. Jian et al. [18] and Wang et al. [6] made a summary of underwater image processing approaches for problems with underwater turbulence, distortion, spectral absorption, and attenuation. Su et al. [19] discussed image restoration techniques for deblurring, denoising, dehazing, and super-resolution as four independent tasks. In summary, these works reviewed image restoration models from completely different perspectives or topics and ignore the generality of deep learning-based methods. However, it is worth noting that many newly emerging Transformer-based IR models, e,g., Restormer [12], SwinIR [21], and Uformer [22], have generalization ability to handle multi-tasks and as cutting-edge research works, they are rarely discussed in existing image processing reviews.

Targeting the shortcomings of the existing IR review works, a comprehensive and systematic study of the advances in IR is given in this work, which is the first attempt to make an overview of IR techniques to our best knowledge. The main contributions of this work are summarized below:

- Benchmark datasets are categorized based on their degradation types.
- Both conventional and new emerging assessment metrics for comparing recovered image qualities are summarized.
- The existing deep learning-based IR methods and their achievements are comprehensively reviewed in general, followed by detailed reviews on CNN-, GAN-, Transformer-, and MLP-based networks in four categories based on the network architectures in particular.
- The reconstruction qualities and efficiencies of the representative IR algorithms are compared on benchmark datasets.
- Potential challenges and future research directions of IR are discussed and analyzed.

The remainder of this review is organized as follows. The background of IR is briefly introduced in Section II. Benchmark datasets for each degradation type are presented in Section III. Section IV introduces image quality assessment metrics. Section V reviews state-of-the-art IR technologies and methods by categories. The comparisons among representative IR algorithms are presented in Section VI. In Section VII, current challenges and future research directions of IR are analyzed. Finally, concluding remarks are given in Section VIII.

II. PROBLEM FORMULATION

Conventional methods for restoring degraded images employ degradation modeling to solve inverse problems, mostly based on the maximum likelihood or Bayesian approaches to iteratively correct the estimated degradations [13], [19]. The general degradation model of a low-quality image I_L can be

formulated as follows:

$$I_L = I_H \otimes k + n, \tag{1}$$

where I_L and I_H denote low-quality (degraded) and high-quality (clean) images respectively, $\otimes$ represents the convolution operation, k is a degradation kernel, and n is noise [23]. By using the maximum a posteriori (MAP) estimation, the latent image I_H is formulated as follows:

$$\widehat{I_H} = \arg\max_{I_H} [\log(P(I_L|I_H)) + \log(P(I_H))], \tag{2}$$

where $P(I_L|I_H)$ denotes the likelihood of degraded observation I_L given clean image I_H, and $P(I_H)$ represents the prior distribution of clean image I_H.

It is worth noting that degradations in the real world are more complex than the assumed or predefined degradations in conventional methods because degradation from the physical world is affected by various unknown factors. The deep learning-based methods (e.g., blind image restoration) have more advantages over traditional modeling-based methods in handling complex unknown degradations due to their powerful feature learning capability, and gradually become the dominant methods for IR. Before reviewing the deep learning-based methods in detail, the benchmark datasets for training or testing deep learning models are presented in the next section, followed by a section introducing current existing image quality metrics.

III. DATASETS

Training (testing) datasets are the cornerstones of the IR algorithms, as deep learning-based models are highly dependent on datasets to learn various degradations. In this section, six categories of related benchmark datasets for six different tasks are briefly introduced.

A. DATASETS FOR SUPER-RESOLUTION

For the training and testing of SR models, the widely used datasets including DIV2K [24], BSDS500 [25], T91 [26], Set5 [27], Set14 [28], Urban100 [29], Manga109 [30] are summarized in Table 1. Datasets vary in image amount, quality, scene, diversity of contents, and resolution. Several datasets comprise paired data, while some datasets contain only HR images and the corresponding LR images usually need to be generated by bicubic downsampling with a set of degradations (e.g., combinations of different levels of Gaussian blurs and noises) [23], [31]. Table 1 summarizes the main characteristics of critical datasets and relevant information on total image count, resolution of HR image, type of dataset, and classes of images.

B. DATASETS FOR DENOISING

A variety of datasets with noisy-clean pairs for image denoising have been collected under different conditions, some of which are dedicated to specific applications (e.g., smartphone cameras [40] and fluorescence microscopy), and most are provided for real-world image noise removal. The datasets' details are listed in Table 2.

C. DATASETS FOR DEHAZING

Existing widely used dehazing datasets containing pairs of real/synthetic hazy and corresponding haze-free images are summarized in Table 3. Few hazy image datasets are collected from real-world scenes while most hazy images are synthetic or artificially generated by a haze machine.

D. DATASETS FOR DEBLURRING

Several datasets with blurry-sharp pairs for image or video deblurring have been collected covering a wide range of scenes, motions, etc. Dataset details are listed in Table 4.

E. UNDERWATER IMAGE DATASETS

Underwater image processing is a newly emerging research field in recent years. Due to the complexity of the underwater circumstances and the high workload, only a few publicly available datasets are collected. Table 5 summarizes publicly available typical databases for underwater image processing and analysis.

F. DATASETS FOR DERAINING

Image deraining aims to restore the clean vision from the degraded image taken on a rainy day. Numerous single-image deraining datasets have been recently constructed. Most datasets are synthesized in two different ways: (1) with the photorealistic rendering techniques proposed by [65] or (1) by adding simulated sharp lines slightly in a certain direction. Table 6 summarizes publicly available databases for draining tasks.

IV. IMAGE QUALITY ASSESSMENT

Image quality assessment (IQA) plays a vital role in effective model comparison in the field of image processing. The goal of IQA is to accurately predict the perceived quality by human viewers and further benefit image processing algorithms to improve the image quality to an acceptable level for the human viewers.

In general, IQA can be briefly grouped into two categories, i.e., human perception-based subjective assessment and quality metrics-based objective assessment. Overall, human evaluation is a more direct, easier way, and more in line with practical needs. It is typically referred to as the mean opinion score (MOS), which is an average rating that human raters assign to images. However, the disadvantage of subjective evaluation is two-fold: (i) the evaluation result is easily affected by personal preferences, and (ii) as a non-automated process, subjective assessment is often costly and time-consuming. While several pre-trained CNN or Transformer models [71], [72] based on a large number of human preference scores have been proposed to solve the labor-consuming problem, the predicted quality scores are not always accurate and the model training process still needs extensive human-judged score collection.

By contrast, objective evaluation is more convenient, although the results by different assessment metrics may not

TABLE 1. List of benchmark datasets used in super-resolution.

Name	Type	Image/pairs amount	Res. of HR image	Scale factors	Category
DIV 2K [24]	Paired	1,000	2048 × 1024	×2, ×3, ×4	People, handmade objects, animals, scenery, etc.
ImagePairs [32]	Paired	11,421	3474 × 2292, etc.	×2	Document, board, office, face, car, object, scenery, etc.
SupER [33]	Paired	85,050	2040 × 1080, etc.	×2, ×3, ×4	Banknote, book, coffee, doll, newspaper, loader, etc.
SR-RAW [34]	Paired	2,000	2040 × 1080, etc.	×2, ×3, ×4	Office, Car, Guidepost, bike, toy, building, book, street etc.
DRealSR [35]	Paired	2,507	5748 × 3746, etc.	×2, ×3, ×4	People, building, plant, poster, book, statue, etc.
City100 [36]	Paired	200	1218 × 870, etc.	×2.9, ×2.4	People, building, scenery, boat, bridge, street, etc.
T91 [26]	Unpaired	91	250 × 200	–	Fruits, cars, faces, etc.
BSDS500 [25]	Unpaired	500	430 × 370	–	people, animal, building, scenery, plant, etc. etc.
Set5 [27]	Unpaired	5	300 × 340	–	Butterfly, baby, bird, head, and women.
Set14 [28]	Unpaired	14	500 × 450	–	Face, animal, flower, animated character, insect, etc.
Urban100 [29]	Unpaired	1,000	800 × 1150	–	Urban buildings
CelebA [37]	Unpaired	202,599	2048 × 1024	–	Celebrities with 40 defined attributes
MS-COCO [38]	Unpaired	164,000	640 × 480	–	Labeled objects with over 80 classes
VOC2012 [39]	Unpaired	11,530	500 × 400	–	Labelled objects with over 20 classes
Manga109 [30]	Unpaired	109	300 × 340	–	Manga

TABLE 2. List of benchmark datasets used in denoising.

Name	Pairs amount	Details
SIDD [40]	30,000	SIDD is specially for smartphone cameras noise removal. Noisy images are captured from 10 scenes under different lighting conditions and ground truth (GT) images are generated through a 4-stage procedure.
SID [41]	5,094	Noisy images are taken under extremely low light with severely limited illumination and short exposure (1/30 and 1/10 sec). GT images are taken from longer exposure references (10 to 30 sec).
FMD [42]	60,000	Noisy fluorescence microscopy images with different noise levels and averaged images served as GT.
PolyU [43]	100	Designed specifically for realistic noise removal. This dataset contains 40 different scenes captured by 5 cameras. The mean of the images captured with low ISO is used as GT.
NIND [44]	126	DSLR-like images with varying levels of ISO noise which is large enough to train models for blind denoising. GT images were taken with the camera's lowest ISO setting.
DND [45]	50	They captured noisy images with different ISO values and appropriately adjusted exposure times. The low ISO images serve as GT.

be necessarily consistent with each other as well as subjective evaluation. The existing image quality metrics can be grouped into two categories, no-reference (NR), and full-reference (FR) metrics, depending on whether ground truth images are required. Table 7 reports widely used metrics for image quality assessment, including no-reference and full-reference metrics.

A. MS-SSIM

As an FR IQA, multi-scale structural similarity (MS-SSIM) [73] first performs contrast comparison, structure comparison, and luminance comparison on multi-scale images, and then combines the measurement at different scales. Further, an image synthesis approach is adopted to calibrate the parameters of cross-scale image quality models to define the relative importance between scales.

B. PSNR

A representative of common and widely used FR quality metric is the peak signal-to-noise ratio (PSNR), which focuses more on the proximity between pixels and assumes pixel-wise independence, resulting in the low consistency with perceptual quality in some cases.

C. SSIM

The Structure similarity index (SSIM) [74] is an FR image quality metric that measures structural similarity and also performs luminance and contrast comparisons. Compared to

TABLE 3. List of benchmark datasets used in dehazing.

Name	Type	Pairs amount	Scene
D-Hazy [46]	Synthetic	1,400+	Indoor
I-HAZE [47]	Artificial	35	Indoor
SOTS [48]	Synthetic	500	Indoor
O-HAZE [49]	Artificial	45	Outdoor
HazeRD [50]	Synthetic	15	Outdoor
Dense-HAZE [51]	Artificial	33	Outdoor
NH-HAZE [52]	Artificial	55	Outdoor
BeDDE [53]	Natural	200+	Outdoor
MRFID [54]	Natural	200	Outdoor
SOTS [48]	Synthetic	500	Outdoor

PSNR, SSIM reflects visual quality better than PSNR [74]. Generally, PSNR and SSIM are used jointly to evaluate the quality of the restored image.

D. IFC

The information fidelity criterion (IFC) [75] is an FR objective quality assessment criterion based on natural scene statistics. IFC considers natural images as signals with certain statistical properties and quantifies the mutual information between the test and reference images via the signal source and distortion models.

E. VSNR

Visual signal-to-noise ratio (VSNR) [76] is an FR image quality metric based on near-threshold and supra-threshold properties of human vision. First, contrast thresholds for the detection of distortions are computed via a wavelet-based model. If perceived contrast is below the threshold, it is deemed to be of perfect visual quality, otherwise, contrast and global precedence are then taken into account as an alternative measure of structural degradation.

F. FSIM

FSIM [77], as an FR IQA, is based on the assumption that HVS perceives an image mainly based on its salient low-level features. Phase congruency (PC) is used as the primary feature in FSIM, the image gradient magnitude (GM) is employed as the secondary feature to encode contrast information, and they represent complementary aspects of the image visual quality.

G. NIQE

The natural image quality evaluator (NIQE) [78] is a completely blind IQA model without any prior knowledge of distortions. The quality of the distorted image is expressed as the distance between the quality-aware natural scene statistic (NSS) feature model and the distorted image's multivariate Gaussian (MVG) model.

H. SR-SIM

Spectral residual based similarity (SR-SIM) [79], as an FR real-time IQA, is based on the spectral residual visual saliency model (SRVS) [80]. The feature map obtained from SRVS characterizes the local quality of an image and the bottom-up visual saliency model utilizing a bottom-up visual attention mechanism provides a weighting function to reflect the importance of a local region in the quality map when pooling the final quality score.

I. PIQUE

The perception-based image quality evaluator (PIQUE) [81] is a blind NR quality evaluation metric. PIQUE first divides the test image into non-overlapping blocks and block-level analysis is performed to identify distortion and grade quality. Further, the overall quality of the test image can be obtained by pooling the block-level scores.

J. CCF

Considering that most of the images taken from the underwater environment have no reference images, hence, NR metrics would be the best choice for evaluating the quality of underwater color images. Wang et al. [82] proposed CCF as a linear combination of colorfulness index, contrast index, and fog density index to predict the color loss caused by absorption, the blurring caused by forwarding scattering, and the foggy caused by backward scattering, respectively.

K. UCIQE

Underwater color image quality evaluation metric (UCIQE) [83], quantifies the non-uniform color casts, blurring, and low contrast, and then linearly combines these three components.

L. UIQM

Another NR criterion for evaluating the quality of underwater images, UIQM [84], is a linear combination of three measures: a colorfulness measure (UICM), a sharpness measure (UISM), and a contrast measure (UIConM), and each attribute metric can be used separately for a specific underwater image processing task. Specifically, UICM utilizes asymmetric alpha-trimmed mean to measure the colorfulness; UISM uses enhancement measure estimation (EME) to measure the sharpness of the grayscale edge map obtained by multiplying the original image with the edge map from the Sobel edge detector; the contrast is measured by applying the logAMEE measure [85] on the intensity image.

M. NRQM

NRQM [86] is a learned NR IQA metric for assessing super-resolved images. Statistical properties including local frequency features, global frequency features, and spatial features are modeled with three independent regression forests. The perceptual scores from linear regression are used to predict the quality of reconstructed SR images.

TABLE 4. List of benchmark datasets used in deblurring.

Name	Type	Blur Model	Content	Details
GoPro [55]	Image	Non-uniform	Outdoor	3,214 image pairs in dynamic scenes. Simulate real-world blur by frame averaging, and the central frame is used as the sharp image.
HIDE [56]	Image	Non-uniform	Pedestrians	8,422 image pairs with human motion and dynamic scenes. The blurry images are synthesized by averaging continuous frames, and the central frame is used as the sharp image.
REDS [57]	Video	Non-uniform	Objects & Scenes	300 video pairs in dynamic scenes. Blurry frames are generated by the interpolated videos.
DVD [58]	Video	Non-uniform	Objects & Scenes	71 video pairs with camera shake motion blur by averaging frames and corresponding sharp version.
Shen *et al.* [59]	Image	Uniform	Face	6,564 sharp images, 130 M training / 16K testing blurry face images with blur kernels and Gaussian noise.
Real Blur [60]	Image	Non-uniform	Scenes	4,738 pairs of images of 232 different scenes including reference pairs.

TABLE 5. List of underwater image benchmark datasets.

Name	Application	Details
UIEB [61]	Image Enhancement	UIEB (Underwater image enhancement benchmark) consists of 890 real-world underwater images, in which 890 has reference, and 60 images without references.
MUED [62]	Object Detection	MUED contains 8,600 underwater images of 430 individual groups of conspicuous objects with labeled ground-truth information.
RUIE [63]	Object Detection & Image Enhancement	RUIE contains 4,000+ images in three groups: Image Quality Subaggregate, Color Cast Subaggregate and higher-level task-driven Subaggregate.
OUC-VISION [64]	Object Detection	OUC-VISION dataset provides 4,400 underwater images and bounding box annotations.
OUC [63]	Object Detection & Image Enhancement	OUC provides underwater images, corresponding reference images generated by UIE algorithms, and bounding box annotations.

TABLE 6. List of benchmark datasets used in deraining.

Name	Image/pairs amount	Usage	Details
Rain12 [66]	100	Test	Synthesized images with only one type of rain streak using photo-realistic rendering techniques.
Rain100L [67]	12	Test	100 synthesized rainy images with only one type of rain streak.
Rain100H [68]	100	Test	Synthesized dataset with five streak directions.
Rain14000 [69]	14,000	Train & Test	A dataset containing 14,000 pairs of rainy/clean images.
Raindrop [70]	1,119	Train	1,119 pairs of rainy/clean images and raindrops attached to a glass window
Rain1800 [68]	1,800	Train	Synthesized rainy images in two types: photo-realistic rendering and simulated sharp line streaks.

N. LPIPS

The learned perceptual image patch similarity (LPIPS) [87] is a reference-based assessment metric focused on perceptual similarity. It relies on deep features from deep networks trained on large-scale, highly varied, perceptual similarity dataset to predict the perceptual similarity between two images.

O. IQT

Image quality transformer (IQT) [72] is a perceptual FR image quality assessment. It uses a CNN backbone to extract feature representations from paired clean/distorted images and the extracted feature maps are fed into a transformer to predict the reconstructed SR image quality score. IQT was

TABLE 7. Commonly used IQA.

Method	Publication	Full/No-reference	Keywords
MS-SSIM	ACSSC-2003 [73]	Full-reference	Multi-scale SSIM, image synthesis, cross-scale calibration
PSNR	-	Full-reference	Mean squared error
SSIM	TIP-2004 [74]	Full-reference	Structure similarity, luminance, contrast, structures
IFC	TIP-2005 [75]	Full-reference	Nature scene statistics, Gaussian scale mixtures
VSNR	TIP-2007 [76]	Full-reference	Two-stage, contrast threshold, global precedence
FSIM	TIP-2011 [77]	Full-reference	Low-level feature, phase congruency
NIQE	SPL-2012 [78]	No-reference	Quality-aware features, multivariate Gaussian model, natural scene statistic model
SR-SIM	ICIP-2012 [79]	Full-reference	Spectral residual visual saliency, bottom-up attention
PIQUE	NCC-2015 [81]	No-reference	Perceptually significant spatial regions, block level distortion map
CCF	ISO4-2018 [82]	No-reference	Combination of colorfulness index, contrast index and fog density index
UCIQE	TIP-2015 [83]	No-reference	Linear combination of chroma, saturation, and contrast
UIQM	ISO4-2016 [84]	No-reference	Colorfulness measure (UICM), sharpness measure (UISM), contrast measure (UIConM)
NRQM	CVIU-2017 [86]	No-reference	Statistical features, regression forests, linear regression model
LPIPS	CVPR-2018 [87]	Full-reference	Deep features, human perceptual similarity
IQT	CVPR-2021 [72]	Full-reference	CNN, Transformer, super-resolution
NeuralSBS	CVPR-2021 [88]	No-reference	CNN, human preference prediction, super-resolution

ranked 1st among 13 participants in the NTIRE 2021 perceptual IQA challenge.

P. NeuralSBS

Neural side-by-side (NeuralSBS) [88] is a NR image quality measure. It adopts a CNN model trained on a paired image dataset with labeled human evaluation scores to predict a probability of being preferable to their counterparts.

V. REVIEW OF IMAGE RESTORATION MODELS

Researchers have been studying IR methods for many decades. Significant progress has been achieved in IR thanks to the advances of deep neural networks. Thus in this review, we focus more on deep learning-based methods. Fig. 2 presents the overall taxonomy of existing IR techniques. According to the type of architecture, they are grouped into four categories: CNN-, GAN-, Transformer-, and MLP-based methods. Table 8 summarizes existing state-of-the-art (SOTA) IR methods. To have a better presentation of existing studies on IR, some necessary diagrams are provided. The following subsections present these methods in detail.

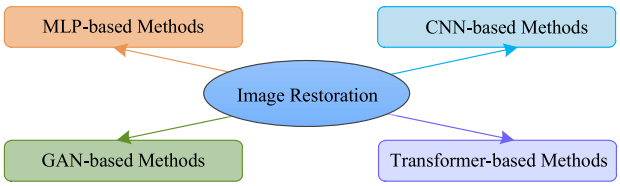


FIGURE 2. The taxonomy of existing single image restoration techniques.

A. CNN-BASED METHODS

CNN has dominated computer vision for nearly 10 years. Recently, CNN architectures [89], [93], [115], [116], [117], [118] have made significant advancements for IR and have shown vast superiority over conventional restoration approaches [119], [120], [121], [122], as they can learn generalizable priors from large-scale datasets. Driven by recent

TABLE 8. An overview of state-of-the-art works on deep learning-based image restoration.

Methods	Publication	Category	Keywords
MIRNet	ECCV-2020 [89]	CNN	Multi-scale, attention-based aggregation, selective kernel feature fusion, deraining, deblurring, denoising
Self2Self	CVPR-2020 [90]	CNN	Self-supervised learning, encoder-decoder network, denoising
Neighbor2Neighbor	CVPR-2021 [91]	CNN	Self-supervised framework, neighbor sub-sampler image pair generation, denoising
HINet	CVPRW-2021 [92]	CNN	Half instance normalization, multi-stage network, deraining, deblurring, denoising
MPRNet	CVPR-2021 [93]	CNN	Interlinked multi-stage architecture, encoder-decoder, supervised attention module, deraining, deblurring, denoising
NBNet	CVPR-2021 [94]	CNN	Subspace Projection, non-local subspace attention, UNet-based architecture, denoising
SPAIR	ICCV 2021 [95]	CNN	Distortion-guided feature extraction, spatially-guided restoration, deblurring, shadow-removal, deraining
DGUNet	CVPR-2022 [96]	CNN	Interpretable deep unfolding network, proximal gradient descent, multi-stage network, denoising, deblurring
AirNet	CVPR-2022 [97]	CNN	All-in-one solution, multi-pairs of encoder-decoder, contrastive learning, denoising, deraining, dehazing
Water-Net	TIP-2019 [61]	CNN	Gated fusion network, feature transformation unit, underwater image enhancement
FFA-Net	AAAI-2020 [98]	CNN	Channel attention with pixel attention, feature fusion structure, dehazing
DPIR	TPAMI-2021 [99]	CNN	Plug-and-play image restoration, image prior, non-blind denoising
CinCGAN	CVPRW-2018 [100]	GAN	Unsupervised SR, Cycle-in-Cycle GAN, domain translation, super-resolution
DeblurGAN	CVPR-2018 [101]	GAN	End-to-end blind motion deblurring, gradient penalty, perceptual loss, deblurring
KMSR	ICCV-2019 [102]	GAN	Blur-kernel modeling, blur kernel pool augment, super-resolution
DeblurGAN-v2	ICCV-2019 [103]	GAN	Feature pyramid deblurring, double-scale RaGAN-LS discriminator, deblurring
DSGAN	ICCVW-2019 [104]	GAN	Frequency separation, unsupervised learning, domain translation, super-resolution
LIR	CVPR-2020 [105]	GAN	Unsupervised learning, disentangle deep representation, domain-transfer
DDGAN	CVPRW-2020 [106]	GAN	Unsupervised learning, blind restoration, CinCGAN-based, double discriminators, image colorization
DBGAN	CVPR-2020 [107]	GAN	Distribution-induced bidirectional, structure-aware prior distribution estimation, Magnetic resonance imaging (MRI) reconstruction
BSRGAN	ICCV-2021 [108]	GAN	Isotropic and anisotropic Gaussian kernels, bicubic and bilinear downsampling method, Gaussian noise, JPEG compression noise, camera sensor noise, blind super-resolution
ECycleGAN	TMM-2022 [109]	GAN	Unsupervised blind IR, CinCGAN-based, content constraint loss, super-resolution
DGP	TPAMI-2021 [110]	GAN	Deep image prior, multi-stage, super-resolution, colorization, inpainting
AquaGAN	CVPRW-2022 [110]	GAN	Attenuation coefficients estimation, weighted combination of content and style loss, underwater image restoration
IPT	CVPR-2021 [111]	Transformer	Pre-trained model on ImageNet, multiple encoder-decoder pairs, denoising, deraining, super-resolution
SwinIR	ICCVW-2021 [21]	Transformer	Swin Transfortmer-based, lightweight model, super-resolution
Uformer-B	CVPR-2022 [22]	Transformer	U-Shaped hierarchical encoder-decoder, nonoverlapping window-based self-attention, denoising, motion deblurring, defocus deblurring, deraining
Restormer	CVPR-2022 [12]	Transformer	Transposed attention, high-resolution image restoration, denoising, motion deblurring, defocus deblurring, deraining
DehazeFormer	arXiv-2022 [11]	Transformer	5-stage U-Net architecture, shifted window partitioning, multiscale fusion, masked multihead self-attention, dehazing
HAT-L	arXiv-2022 [48]	Transformer	Combination of channel attention and self-attention, overlapping cross-attention, super-resolution
MLP-Mixer	NIPS-2021 [112]	MLP	Pure MLP, token-mixing, channel-mixing, milestone of MLP in vision
gMLP	NIPS-2021 [113]	MLP	Spatial Gating Unit (SGU), attention-free, comparable performance to Transformers in vision applications
MAXIM	CVPR-2022 [114]	MLP	Multi-axis gating, multi-axis self-attention, Cross Gating Block (CGB), denoising, deblurring, deraining, dehazing, enhancement

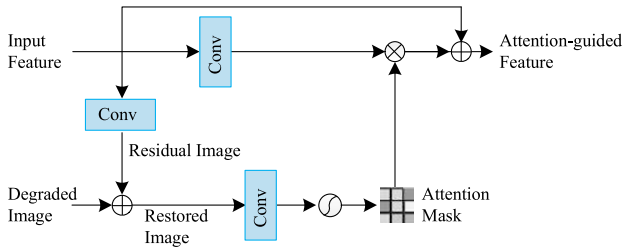


FIGURE 3. Supervised attention module.

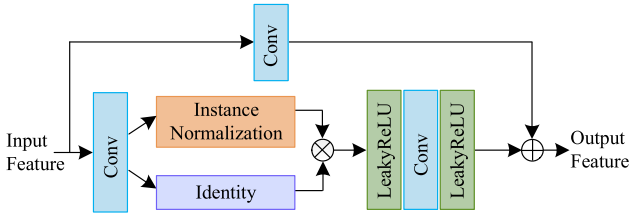


FIGURE 4. The diagram of HINet Block [92].

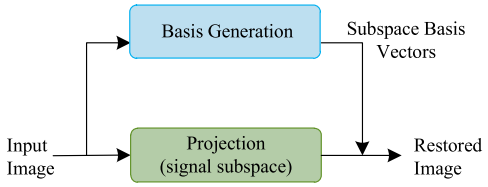


FIGURE 5. NBNet [94] denoises via subspace projection: 1) Basis generation: generates subspace basis vectors from feature representations; 2) Projection: maps feature representations into the signal subspace.

enormous efforts on building vision benchmarks, numerous CNNs have been developed and achieved promising performance on a wide variety of image restoration and enhancement tasks. These increased performance gains can be mainly attributed to novel architecture designs, invented or borrowed modules and units, including residual learning [123], [124], [125], dilated convolutions [68], [126], dense connections [125], hierarchical structures [103], [127], [128], encoder-decoder [41], [93], [129], [130], [131], [132], [133], [134], multi-stage frameworks [14], [55], [135], [136], [137], and attention mechanisms [138], [139].

Among CNN designs, encoder-decoder architectures [41], [129], [130], [131], [133], [134] have been extensively studied for IR due to their hierarchical multi-scale representation achieved by progressively mapping the input to low-resolution representations and a corresponding inverse mapping while remaining computationally efficiency.

Although the encoder-decoder architectures are effective in capturing broad context by spatial-resolution reduction, these approaches are unreliable in preserving fine spatial details. High-resolution single-scale networks [118], [124], [140], [141] produce images with spatially more accurate details, and are less effective in encoding contextual information due to their limited receptive field. To address this problem, MIRNet [89] employs parallel multi-scale residual blocks while maintaining the original high-resolution features to preserve precise spatial details. Elective kernel feature fusion and dual attention allow for feature aggregation and feature recalibration along the spatial and channel dimensions, respectively.

To provide a balanced design of spatial details and high-level contextualized information while recovering images, encoder-decoder-based MPRNet [93] employs a multi-stage approach, which progressively restores images by decomposing the challenging IR into smaller easier subtasks. Unlike previous multi-stage approaches [14], [55], [135], [136], [137] that simply cascade stages, MPRNet introduces a supervised attention module (SAM, as shown in Fig.3) between stages to provide ground-truth supervisory signals useful for the progressive IR, which facilitates achieving significant performance gain.

To enhance the transformation modeling capability of CNNs, a deformable convolutional network [142] was developed with deformable convolution and deformable RoI pooling, by augmenting the spatial sampling locations with additional offsets and learning the offsets. Inspired by the projective motion path blur (PMPB) model and deformable convolution, a novel constrained deformable convolutional network (CDCN) [143] using blur kernels estimation approach with projective motion path blur-based deblurring loss function was designed for blind deblurring.

The self-supervised learning model Self2Self [90] targeted the absence of noisy-clean image pairs (i.e., only the available noisy images) for training. It was trained with dropout on the pairs of Bernoulli-sampled instances of the input images. Bernoulli dropout scheme was adopted in both training and testing for variance reduction. Neighbor2Neighbor [91] took one step further and adopted a random neighbor sub-sampler for the generation of training image pairs.

The model-driven CNN-based IR methods, e.g., plug-and-play IR [144], [145], have shown that a denoiser can implicitly serve as the image prior for model-based methods to solve the inverse problem. The representative work DPIR plugs the deep denoiser prior as a modular part into a half quadratic splitting-based iterative algorithm to solve the IR problem.

Instance normalization (IN) is widely used in high-level computer vision tasks, but its performance degrades severely in low-level tasks. To address this problem, the half instance normalization (HIN) block (Fig. 4) was introduced as a building block in HINet [92], which significantly boost the performance of IR.

NBNet [94] achieves denoising from a new perspective of subspace projection, based on the observation that projection can naturally preserve the local structure of the input signal. NBNet learns to generate a set of reconstruction basis for the signal subspace with a subspace attention (SSA) module. By projecting the input into signal subspace, the signal can be enhanced after separation from noise (Fig. 5).

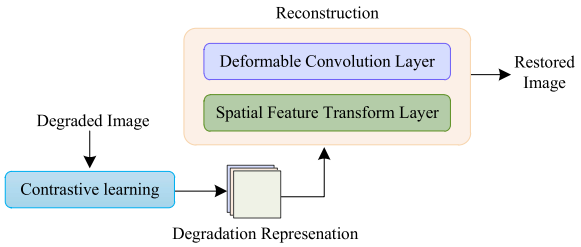


FIGURE 6. The diagram of AirNets [97].

DGUNet [96] employs a novel interpretable deep unfolding network for single image restoration (SIR), which integrates a gradient estimation into the proximal gradient descent (PGD) algorithm. Inter-stage information pathways broadcast multi-scale features in a spatial-adaptive normalization way, which rectifies the intrinsic information loss.

Degradation-specific CNNs have achieved promising results on benchmark datasets. However, these methods suffer a severe performance drop when degradation is different in practical application. Towards this challenge, SPAIR [95] exploits distortion-localization information and uses distortion guidance to perform spatially-varying modulation on degraded pixels. By employing multi-pairs of encoder-decoder to cope with each type of corruption, AirNet [97] can be used as an all-in-one solution that is capable of handling multiple types of degradations (see Fig. 6). To be specific, contrastive learning is used to extract the degradation representation from the input. Conditioned on the extracted features, the subsequent image restoration network is degradation-guided.

Overall, 'convolution' in CNNs provides local connectivity and translation equivariance. These properties bring efficiency and generalization to CNNs, but they also cause two main issues: (1) the limited receptive field of convolution makes it hard in modeling long-range pixel dependency, and (2) the convolution filters have static weights at inference, and thereby cannot flexibly adapt to the input content.

B. GAN-BASED METHODS

GAN is another class of deep generative models, which has recently gained significant attention [146]. It adopts an architecture in which two opposite networks compete with each other to generate desired data. Discriminator and generator play the two-player minimax game. The generator learns to produce new samples with the same distribution as the target domain and is trained to fool the discriminator and to capture the real data distribution. The minimax game with the value function $V(D, G)$ is formulated as:

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]. \quad (3)$$

As a popular GAN variant, Conditional GAN [147], taking the class label and the latent code as inputs, has been widely applied to image-to-image translation problems including image restoration and enhancement as special cases. To recover the finer texture details when upscaling an image, as an image SR model based on GAN, SRGAN employs a deep residual network (ResNet [148]) with skip-connection and an optimized perceptual loss calculated on feature maps of the VGG network [149]. To deal with motion blur, DeblurGAN [101], as an end-to-end blind IR model, exploits Wasserstein GAN [150] with the gradient penalty [151] and the perceptual loss [152]. DeblurGAN-v2 [103], featured with light-weight and fast, introduces a feature pyramid network (FPN) as a core building block into its generator. KMSR [102] improves the blind SR performance by integrating the blur-kernel estimation into GAN.

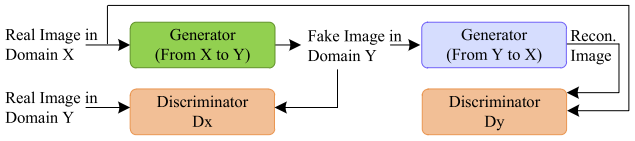


FIGURE 7. The diagram of CycleGAN [97].

CycleGAN [153], using unpaired image-to-image translation from a source domain X to a target domain Y, was proposed to overcome the absence of ground truth. Two generators are trained simultaneously to learn a pair of inverse mappings by forcing cycle consistency and adversarial losses (Fig. 7). Based on CycleGAN, several unsupervised learning models [153], [154], [155] are developed to do image SR from unpaired LR-HR training samples, which have obvious advantages in dealing with unknown types of degradation. DSGAN [104] makes a step further and separates the low and high image frequencies and treats them differently during training. Unsupervised SR model CinCGAN [100] makes remarkable progress in the SR task by using Cycle-in-Cycle GAN. DDGAN, a CinCGAN-based unsupervised learning model with double auto-encoding discriminators, was proposed to solve the color crossing problem. To improve the reconstruction ability of the cycle consistent network and preserve more fidelity of the reconstructed image, as an unsupervised blind image restoration model, ECycleGAN [109] makes two-fold explorations: (1) preserving the low-frequency content fidelity and (2) suppressing the high-frequency artifacts.

Directly applying a domain transfer approach for IR would lead to domain-shift problems in translated images due to the lack of effective supervision. Instead, LIR [105] learns invariant presentation from noisy data and reconstructs clear observations by introducing discrete disentangling representation and adversarial domain adaption into a general domain transfer framework.

DBGAN [107] is a distribution-induced bidirectional GAN, which first proposes graph representation learning and utilizes a structure-aware prior distribution estimation based on DPP [156] for latent representation by prototype learning. The diagram of DBGAN is depicted in Fig. 8.

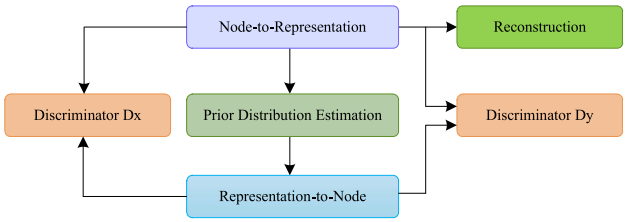


FIGURE 8. The diagram of DBGAN [107].

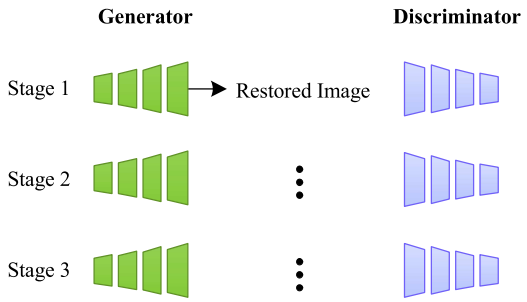


FIGURE 9. The diagram of DGP [97].

SinGAN [157] shows that a randomly-initialized GAN model is able to capture rich patch statistics being trained from a single image. DGP [110] goes one step further, exploiting deep generative prior for image restoration by being trained on large-scale natural images and adopting a progressive reconstruction strategy that fine-tunes the generator gradually (Fig. 9). DGP achieved a more precise and faithful reconstruction for real images on a range of different tasks (e.g., colorization, inpainting, SR, Image morphing, etc.). GANs have achieved amazing performance even in extremely complex application (e.g., underwater) [158], [159], [160]. A representative work, AquaGAN [161], proposes a weighted combination of content and style loss for the first time, and generates clean underwater images. It is worth noting that the attenuation coefficient in AquaGAN is very sensitive, and the recovery results corresponding to different values vary significantly.

Overall, GANs can generate data that looks similar to the original data. However, there exist major challenges in the training of GANs, i.e., mode collapse, non-convergence, and instability. The trained models may vary a lot between adjacent iterations. It is easy to trap a bad local minimum when training the model, and the generalization ability of the final trained model cannot be guaranteed.

C. TRANSFORMER-BASED METHODS

More recently, another class of neural architectures, Transformer, has achieved great success in natural language processing (NLP) and high-level vision tasks. Subsequent research explorations on Vision Transformers (e.g., ViT [162]) have exemplified their great potential as alternatives to the go-to CNN models.

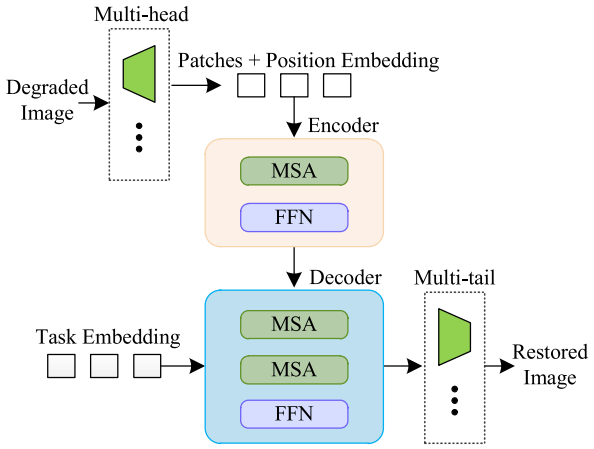


FIGURE 10. The architecture of IPT [111].

By leveraging the self-attention (SA) mechanism [163], Transformers mitigates the shortcomings of CNNs (i.e., limited receptive field and inadaptability to input content) with the capability to capture long-range dependencies between image patch sequences and adaptability to given input content. The SA mechanism plays a key role in modeling global connectivity. It calculates the response at a given pixel by a weighted sum of all other positions, which can be described as mapping a query and a set of key-value pairs to an output. The SA scoring function is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{dk}})V, \tag{4}$$

where key $K \in R^{C\times HW}$, query $Q \in R^{HW\times C}$, value $V \in R^{HW\times C}$, and $\frac{1}{\sqrt{dk}}$ is the scaling factor.

By exploiting the powerful feature representation capability of the Transformer and using the strong computing power of modern hardware, a pre-trained image processing transformer (IPT [111]) on the ImageNet dataset is developed, which can be used for super-resolution, denoising, or deraining task after fine-tuning on a task-specific dataset. The IPT model consists of multiple pairs of heads and tails for different tasks and a shared transformer body including an encoder and decoder, as shown in Fig. 10. Specifically, the multi-head takes the degraded images as input and converts them to feature maps then splits them into patches as "visual words" for subsequent processing in Transformer. The clean images are reconstructed by ensembling output patches. Notably, IPT requires the degradation information in priori and specifies the associated head.

Pioneer vision Transformer works for low-level vision [111], [164] by accepting relatively small patches (tokens) spitted from the input image, which inevitably causes patch boundary artifacts when applied to larger images. To tackle this problem, Swin Transformer [165] adopts a shifted windowing scheme that limits self-attention computation to non-overlapping local windows while also allowing for cross-window connection. SwinIR [21] takes advantage of Swin

Transformer and utilizes several Swin Transformer layers for local attention and cross-window interaction. Besides that, SwinIR uses MLP (2 layers with GELU [166]) for feature transformations.

Based on the observation that Transforms can only utilize a limited spatial range of input information through attribution analysis, to address this problem, a hybrid attention Transformer (HAT [167]) combines channel attention and self-attention schemes and makes use of their complementary advantages. To enhance the interaction between neighboring window features, an overlapping cross-attention module is employed in HAT.

While the SA mechanism in Transformer has shown its superiority over CNNs in capturing long-range pixel interactions when dealing with low-level vision tasks, its computational complexity grows quadratically with the spatial resolution, therefore it is infeasible to adapt SA to HR images. By adopting transposed attention across channels, the computational loads are significantly reduced from $\mathcal{O}(W^2H^2)$ to $\mathcal{O}(C^2)$. With this idea, Restormer [12] provides a solution for taking the entire HR image as input instead of the image patches, which effectively avoids the boundary issue when fusing the restored patches into an intact restored image. Transposed attention is defined as:

$$\text{Attention}(Q, K, V) = V\,\text{softmax}(\frac{KQ}{\alpha}), \tag{5}$$

where key $K \in R^{C\times HW}$, query $Q \in R^{HW\times C}$, value $V \in R^{HW\times C}$, and α is a learnable scaling parameter.

Uformer [22] employs U-shaped hierarchical encoder-decoder architecture and leverages local-enhanced window Transformer with two core designs: (1) non-overlapping window-based self-attention (W-MSA), and (2) locally-enhanced feed-forward network (LeFF). Given a degraded image $I \in R^{3\times HW}$ and window size of $M \times M$, W-MSA reduces the computational cost from $\mathcal{O}(W^2H^2)$ to $\mathcal{O}(M^2WH)$. LEFF incorporates convolution designs into visual Transformers to capture local information. With these designs, Uformer shows a strong generalization ability on various degradations.

Overall, SA is highly effective in capturing long-range pixel interactions, but its complexity grows quadratically with spatial resolution. How to reduce the computational complexity of SA and maintain efficiency in modeling global connectivity will be the focus of future research.

D. MLP-BASED METHODS

Most recently, deep multilayer perceptron (MLP) models have roused great interest in the vision community. MLPs are considered the "classic" form of a neural network, consisting of a series of simple fully-connected layers or perceptrons, which were first developed in 1958. MLP-Mixer [112], an architecture based exclusively on multi-layer perceptrons, presented by Google Brain in May 2021, led MLPs revival. MLP-Mixer exemplifies that, while convolutions and attention are both sufficient for good performance, neither of them is necessary, by leveraging two types of layers: one with MLPs applied independently to image patches for mixing local features, and the other with MLPs applied across patches for mixing spatial information.

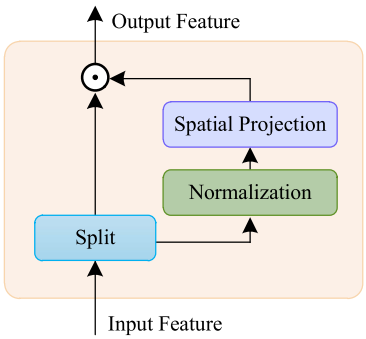


FIGURE 11. Spatial Gating Unit (SGU) in gMLP [113].

The Google Brain team further studied the necessity of self-attention modules in Transformers and proposed the gMLP [113] model, based on MLPs with gating. The spatial gating unit (SGU) is the key design element of gMLP used for cross-token interactions (Fig. 11). The gating function is defined as:

$$s(Z) = Z_1 \odot f_{W,b}(Z_2), \tag{6}$$

where Z_1 and Z_2 are two independent split parts of the input feature Z along the channel dimension and $\odot$ denotes the element-wise multiplication (linear gating).

Zhao et al. [168] proposed a multi-axis self-attention embedded in the HiT model. They first split the input image patches into two groups along the channel dimension, one of which performs regional attention operation within fixed windows and the other performs dilated attention across windows. Therefore, self-attention gets enhanced by considering local (within windows) as well as global (across windows) relations.

MAXIM [114], multi-axis MLP for image processing, takes advantage of multi-axis self-attention and gating mechanisms in gMLP. The proposed multi-axis gated MLP block (MAB) (Fig. 12) can enjoy a global receptive field, with linear complexity. The cross-gating block (CGB) as an extension of MAB built on gMLP allows global contextual features to gate the skip-connections. With multi-stage multi-scale architecture, MAXIM achieves state-of-the-art performance on more than ten benchmarks across a range of image processing tasks.

Notably, MLPs have an advantage in capturing global attention, but applying gMLP to low-level vision tasks has to overcome a missing attribute of basic CNNs.

Overall, the above four different types of IR solutions have their unique characteristics, and thus it is necessary to weigh their pros and cons according to the application requirements. More specifically, CNNs, serving as backbone networks, have achieved impressive results with dense

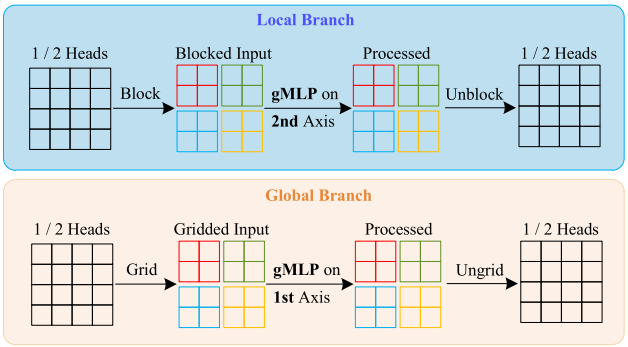


FIGURE 12. Multi-axis gated MLP [114].

connections [169] and more complex forms of convolution [142], [170], [171], but have a limitation in modeling long-range pixel interactions and processing images with large size. Confronted with the lack of well-aligned image pairs, GAN-based domain translation methods are more flexible, but the generalization issue still exists. Transformer has a significant advantage of capturing long-range dependencies at the cost of quadratic computational complexity due to self-attention. Although attention-free MLP can achieve comparable performance in IR to Transformers, it requires further exploration to make breakthroughs in reducing the complexity and capturing global and local connectivity.

VI. COMPARISONS AMONG STATE-OF-THE-ARTS

In this section, representative SOTA IR methods are visually and numerically compared on benchmark datasets. More specifically, the selected competitors include DeblurGAN-v2 [103], MPRNet [93], MIRNet [89], DGUNet [96], MAXIM [114], Restormer [12], Uformer-B [22], IPT [111], SwinIR [21], HAT-L [48], and BSRGAN [108], covering multiple kinds of approaches mentioned in Section V. Experiments are conducted on six benchmark datasets for different IR tasks, GoPro [55] for motion deblurring, SIDD [40] for denoising, Rain100L [40] for deraining, SOTS [48] for outdoor dehazing, Set14 [28] for SR, and UIEB [61] for underwater image restoration. Several different IQAs are used to measure the quality of the restored images. The numerical results are presented in Tables 9 to 14 and the visual comparisons are shown in Fig. 13.

Experiment Results Analysis: (1) Benefiting from the design of a multi-scale restoration modulator, Uformer-B [22] shows its advantage in motion blur removal tasks. The image restored by Uformer-B is more clear, compared with other methods (See Table 9 and the 1st row in Fig. 13). (2) For the denoising task, the image reproduction quality of Restormer [12] is more faithful to the ground truth than other methods (See Table 10 and the 2nd row in Fig. 13), owing to the multiscale hierarchical design incorporating gating mechanism in the feed-forward network. (3) IPT [111] shows its significant superiority over other methods in deraining (Table 11 and the 3rd row in Fig. 13), which can be attributed to its pre-training strategy, as the pre-trained model enjoys a convenience of self-generating training instances based on the original real images. (4) DehazeFormer-B [11], based on Swin Transformer [165], improved performance in dehazing by modifying the normalization layer, adapting softRELU, and aggregating spatial information. It achieved the best results of dehazing in terms of details restoration (See Table 12 and the 4th row in Fig. 13). (5) The combination of channel attention and self-attention, and the overlapping cross-attention module significantly enhanced the performance of HAT-L [48] in recovering the finer details on the SR task. HAT-L outperforms other SOTA SR models, as shown in Table 13 and the 5th row of Fig. 13. (6) Two existing SOTA underwater IR methods, Water-Net [61] and AquaGAN [110], are compared. From the last row of Fig 13, we can observe that the restored image by AquaGAN has better quality in terms of color correction and texture details restoration. Numerical results listed in Table 14 match the visual observation.

TABLE 9. Motion deblurring performance of the representative IR algorithms on GoPro [55].

Method	PSNR	SSIM	FSIM	IFC
DeblurGAN-v2 [103]	29.55	0.934	0.9431	2.6929
MPRNet [93]	32.66	0.959	0.9634	3.1568
HINet [92]	32.71	0.959	0.9772	3.4845
MAXIM [114]	32.86	0.961	0.9654	3.5642
Restormer [12]	32.92	0.961	0.9839	3.671
Uformer-B [22]	32.97	0.967	0.9847	3.6807

TABLE 10. Denoising performance of the representative IR algorithms on SIDD [40].

Method	PSNR	SSIM	FSIM	IFC
MPRNet [93]	39.71	0.958	0.9234	3.3584
MIRNet [89]	39.72	0.959	0.9472	3.4816
Uformer-B [22]	39.89	0.960	0.9675	3.4582
MAXIM [114]	39.96	0.960	0.9851	3.6423
HINet [92]	39.99	0.958	0.9751	3.6829
DGUNet [96]	39.91	0.962	0.9809	3.6713
Restormer [12]	40.02	0.965	0.9827	3.6804

VII. CHALLENGES AND FUTURE SUGGESTIONS

In this section, current challenges faced by IR are analyzed from different perspectives and directions are discussed for future research.

A. IMAGE DATASETS

Most benchmark datasets for IR are synthetic-based or simply handcrafted, while few datasets are collected in limited

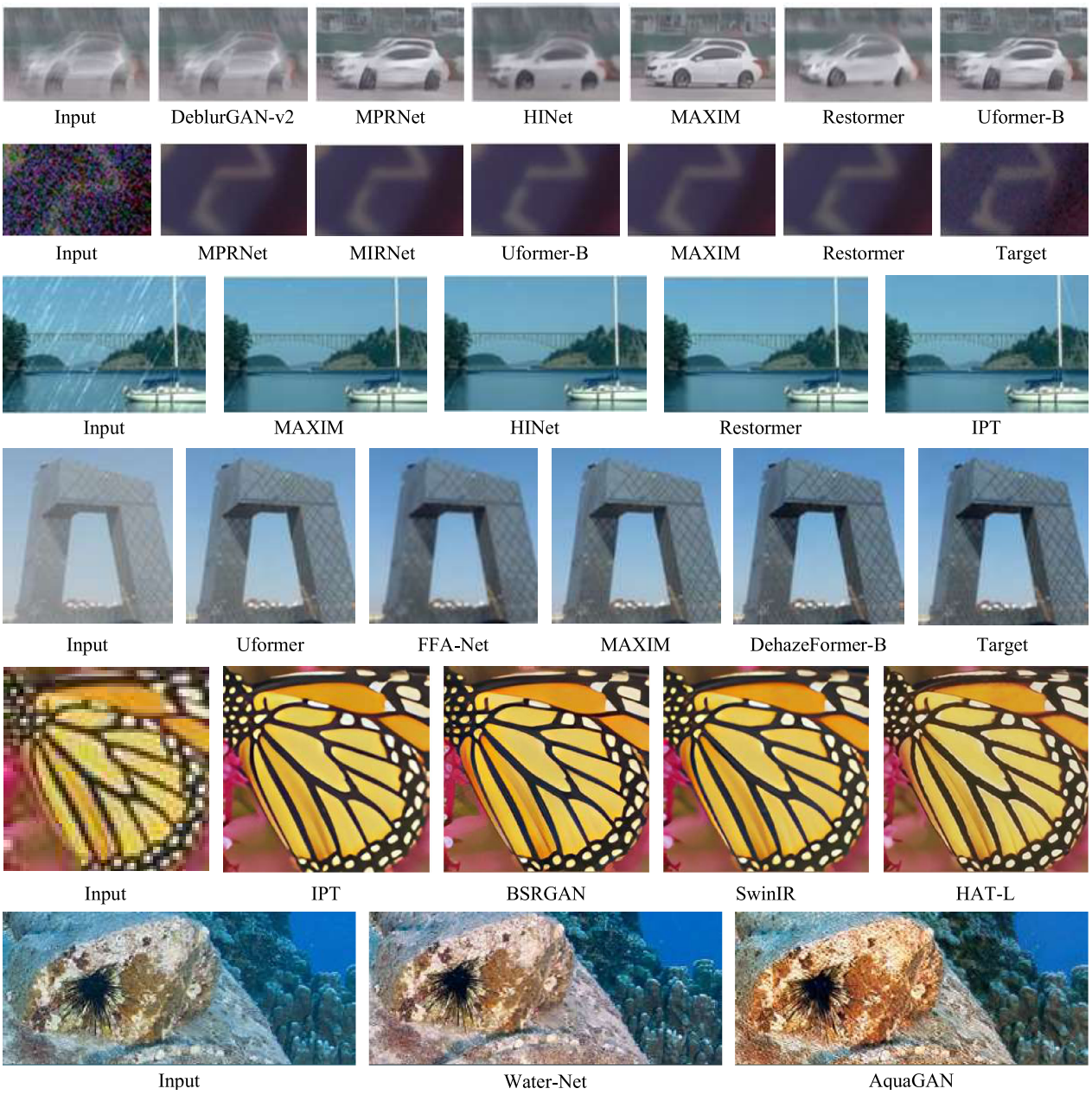


FIGURE 13. Visual comparisons of representative SOTA IR algorithms for six different tasks. From top row to bottom row: motion deblurring on GoPro [55], denoising on SIDD [40], draining on Rain100L [40], outdoor dehazing on SOTS [48], 4 × upscaling super-resolution on Set14 [28], and underwater image restoration on UIEB [61].

scenes with real degradations, due to the infeasibility of collecting the degraded/undegraded image pairs in most real scenes. However, the lack of high diversity of real degradations in learning samples fails to provide strong priors for learning-based models, which causes severe performance degradation for trained models in real applications. How to generate more realistic degradation in learning samples could be a research direction in the future.

B. IMAGE QUALITY ASSESSMENT

IQA plays a critical role in image processing tasks. Although it is easy for human beings to distinguish perceptually better images, it has been proved to be difficult for algorithms. GAN-based image processing algorithms have posed particular challenges, as they bring completely new characteristics to the output images. The growing discrepancy between the quantitative evaluation results and the perceptual quality will affect the development of image processing algorithms if the IQA methods cannot objectively compare their perceptual quality. Therefore, it is desired to develop more suitable IQA methods accordingly to adapt to the emerging image processing algorithms.

C. IMAGE RESTORATION ALGORITHMS

Although promising results have been achieved in a specific area, such as denoising, deblurring, deraining, dehazing, and

TABLE 11. Deraining performance of the representative IR algorithms on Rain100L [40].

Method	PSNR	SSIM	FSIM	IFC
HINet [92]	37.28	0.9710	0.9643	3.6013
MAXIM [114]	38.92	0.9772	0.9715	3.6815
Restormer [12]	38.99	0.978	0.9823	3.7206
IPT [111]	41.62	0.9881	0.9862	3.7542

TABLE 12. Dehazing performance of the representative IR algorithms on SOTS Outdoor [48].

Method	PSNR	SSIM	FSIM	IFC
Uformer-B [22]	31.10	0.9760	0.9562	3.5329
FFA-Net [98]	33.57	0.9804	0.9627	3.6817
MAXIM [114]	34.19	0.9702	0.9813	3.7412
DehazeFormer-B [11]	34.95	0.9840	0.9853	3.7541

TABLE 13. 4× upscaling performance of the representative SR algorithms on Set14 [28].

Method	PSNR	SSIM	FSIM	IFC
IPT [111]	29.01	0.7824	0.9325	2.5336
BSRGAN [108]	29.08	0.7923	0.9207	2.5723
SwinIR [21]	29.15	0.7958	0.9314	2.6278
HAT-L [167]	29.47	0.8015	0.9423	2.6804

TABLE 14. Performance comparison of SOTA underwater IR algorithms on UIEB [61].

Method	UCIQE	UIQM	CCF
Water-Net [61]	5.4637	0.7853	26.49
AquaGAN [110]	5.9612	0.8391	33.24

super-resolution, image restoration has encountered the following obstacles in practice. 1) The type of degradation must be known in advance to select a competing model since most existing methods can only handle one specific degradation in the inference stage, although they can handle different IR tasks after retraining. 2) In real applications, the degradation usually changes in complex environments and the images may suffer from various degradations consecutively or even simultaneously (e.g., rainy and hazy weather). Confronted with these challenges, it is necessary to build a generic model that can handle all possible various degradations in one solution.

Despite the promising performance gains achieved with novel architecture designs and newly invented modules or units, the increase in computational complexity of the SOTA frameworks limits their real-time applications. Lightweight models are desired for practical needs. Optimizing the IR models to have a better trade-off between the restoration performance and running time would be an unavoidable topic.

VIII. CONCLUSION

With the rapid development of consumer and industry cameras and smartphones, the requirements of obtaining high-quality images are constantly growing. Real-world image restoration plays a crucial role in recovering clean images and has been receiving increased attention. Due to its ill-posed nature, IR remains a challenging problem. In this paper, the commonly used datasets and assessment metrics for IR models are first summarized. Then, recent IR methods for reproducing realistic images, including CNN-, GAN-, Transformer-, and MLP-based algorithms, are comprehensively reviewed. The pros and cons of each type of architecture are presented. Finally, the challenges of IR are analyzed and the analysis shows that although some progress has been made on IR in the past few years, these unsolved problems indicate promising directions for future explorations.

REFERENCES

[1] M. O. Macaulay and M. Shafiee, "Machine learning techniques for robotic and autonomous inspection of mechanical systems and civil infrastructure," *Auto. Intell. Syst.*, vol. 2, no. 1, pp. 1–25, Dec. 2022.

[2] A. Aakerberg, A. S. Johansen, K. Nasrollahi, and T. B. Moeslund, "Semantic segmentation guided real-world super-resolution," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 449–458.

[3] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," 2022, *arXiv:2204.04676*.

[4] K. Purohit and A. N. Rajagopalan, "Motion deblurring with an adaptive network," 2019, *arXiv:1903.11394*.

[5] J. Lu, F. Yuan, W. Yang, and E. Cheng, "An imaging information estimation network for underwater image color restoration," *IEEE J. Ocean. Eng.*, vol. 46, no. 4, pp. 1228–1239, Oct. 2021.

[6] Y. Wang, W. Song, G. Fortino, L.-Z. Qi, W. Zhang, and A. Liotta, "An experimental-based review of image enhancement and image restoration methods for underwater imaging," *IEEE Access*, vol. 7, pp. 140233–140251, 2019.

[7] S. Cui, Y. Zhou, Y. Wang, and L. Zhai, "Fish detection using deep learning," *Appl. Comput. Intell. Soft Comput.*, vol. 2020, Jan. 2020, Art. no. 3738108.

[8] Y. Pei, Y. Huang, Q. Zou, X. Zhang, and S. Wang, "Effects of image degradation and degradation removal to CNN-based image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1239–1253, Apr. 2021.

[9] Y. Xu, J. Wen, L. Fei, and Z. Zhang, "Review of video and image defogging algorithms and related studies on image restoration and enhancement," *IEEE Access*, vol. 4, pp. 165–188, 2016.

[10] Z. Hao, S. You, Y. Li, K. Li, and F. Lu, "Learning from synthetic photorealistic raindrop for single image raindrop removal," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 4340–4349.

[11] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," 2022, *arXiv:2204.03883*.

[12] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5728–5739.

[13] L. Wang, Y. Wang, X. Dong, Q. Xu, J. Yang, W. An, and Y. Guo, "Unsupervised degradation representation learning for blind super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10581–10590.

[14] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, "Lightweight pyramid networks for image deraining," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 6, pp. 1794–1807, Jun. 2020.

[15] Y. Zhou, J. Jiao, H. Huang, Y. Wang, J. Wang, H. Shi, and T. Huang, "When AWGN-based denoiser meets real noises," 2019, *arXiv:1904.03485*.

[16] L. Zhai, Y. Wang, S. Cui, and Y. Zhou, "Enhancing underwater image using degradation adaptive adversarial network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 4093–4097.

[17] H. Chen, X. He, L. Qing, Y. Wu, C. Ren, E. Ray Sheriff, and C. Zhu, "Real-world single image super-resolution: A brief review," *Inf. Fusion*, vol. 79, pp. 124–145, Mar. 2022.

[18] M. Jian, X. Liu, H. Luo, X. Lu, H. Yu, and J. Dong, "Underwater image processing and analysis: A review," *Signal Process., Image Commun.*, vol. 91, Feb. 2021, Art. no. 116088.

[19] J. Su, B. Xu, and H. Yin, "A survey of deep learning approaches to image restoration," *Neurocomputing*, vol. 487, pp. 46–65, May 2022.

[20] R. S. Thakur, S. Chatterjee, R. N. Yadav, and L. Gupta, "Image de-noising with machine learning: A review," *IEEE Access*, vol. 9, pp. 93338–93363, 2021.

[21] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," 2021, *arXiv:2108.10257*.

[22] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general U-shaped transformer for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17683–17693.

[23] J. Gu, H. Lu, W. Zuo, and C. Dong, "Blind super-resolution with iterative kernel correction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1604–1613.

[24] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 126–135.

[25] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, Oct. 2010.

[26] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[27] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, p. 135.

[28] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.* Berlin, Germany: Springer, 2012, pp. 711–730.

[29] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.

[30] A. Fujimoto, T. Ogawa, K. Yamamoto, Y. Matsui, T. Yamasaki, and K. Aizawa, "Manga109 dataset and creation of metadata," in *Proc. 1st Int. Workshop coMics Anal., Process. Understand.*, Dec. 2016, pp. 1–5.

[31] H. Yuan, Y. Wang, G. Xu, and F. Wang, "Research on super-resolution reconstruction of single-frame image of infrared focal plane polarization imaging," in *Proc. 8th Symp. Novel Photoelectron. Detection Technol. Appl.*, Mar. 2022, pp. 603–609.

[32] H. R. Vaezi Joze, I. Zharkov, K. Powell, C. Ringler, L. Liang, A. Roulston, M. Lutz, and V. Pradeep, "ImagePairs: Realistic super resolution dataset via beam splitter camera rig," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Los Alamitos, CA, USA, Jun. 2020, pp. 2190–2200.

[33] T. Köhler, M. Bätz, F. Naderi, A. Kaup, A. Maier, and C. Riess, "Toward bridging the simulated-to-real gap: Benchmarking super-resolution on real data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2944–2959, Nov. 2020.

[34] X. Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to learn, learn to zoom," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3757–3765.

[35] P. Wei, Z. Xie, H. Lu, Z. Zhan, Q. Ye, W. Zuo, and L. Lin, "Component divide-and-conquer for real-world image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 101–107.

[36] C. Chen, Z. Xiong, X. Tian, Z.-J. Zha, and F. Wu, "Camera lens super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Los Alamitos, CA, USA, Jun. 2019, pp. 1652–1660.

[37] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738.

[38] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Lawrence Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision–(ECCV)*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 740–755.

[39] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, pp. 98–136, Jun. 2014.

[40] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1692–1700.

[41] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.

[42] Y. Zhang, Y. Zhu, E. Nichols, Q. Wang, S. Zhang, C. Smith, and S. Howard, "A Poisson-Gaussian denoising dataset with real fluorescence microscopy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11710–11718.

[43] J. Xu, H. Li, Z. Liang, D. Zhang, and L. Zhang, "Real-world noisy image denoising: A new benchmark," 2018, *arXiv:1804.02603*.

[44] B. Brummer and C. D. Vleeschouwer, "Natural image noise dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1777–1784.

[45] T. Plotz and S. Roth, "Benchmarking denoising algorithms with real photographs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1586–1595.

[46] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2226–2230.

[47] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images," in *Advanced Concepts for Intelligent Vision Systems*. Cham, Switzerland: Springer, 2018, pp. 620–631.

[48] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.

[49] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 754–762.

[50] Y. Zhang, L. Ding, and G. Sharma, "HazeRD: An outdoor scene dataset and benchmark for single image dehazing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3205–3209.

[51] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1014–1018.

[52] C. O. Ancuti, C. Ancuti, and R. Timofte, "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 444–445.

[53] S. Zhao, L. Zhang, S. Huang, Y. Shen, and S. Zhao, "Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines," *IEEE Trans. Image Process.*, vol. 29, pp. 6947–6962, 2020.

[54] W. Liu, F. Zhou, T. Lu, J. Duan, and G. Qiu, "Image defogging quality assessment: Real-world database and method," *IEEE Trans. Image Process.*, vol. 30, pp. 176–190, 2021.

[55] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3883–3891.

[56] Z. Shen, W. Wang, X. Lu, J. Shen, H. Ling, T. Xu, and L. Shao, "Human-aware motion deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5572–5581.

[57] S. Nah, S. Baik, S. Hong, G. Moon, S. Son, R. Timofte, and K. M. Lee, "NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1996–2005.

[58] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, and O. Wang, "Deep video deblurring for hand-held cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1279–1288.

[59] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8260–8269.

[60] J. Rim, H. Lee, J. Won, and S. Cho, "Real-world blur dataset for learning and benchmarking deblurring algorithms," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2020, pp. 184–201.

[61] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, and S. Kwong, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[62] M. Jian, Q. Qi, H. Yu, J. Dong, C. Cui, X. Nie, H. Zhang, Y. Yin, and K.-M. Lam, "The extended marine underwater environment database and baseline evaluations," *Appl. Soft Comput.*, vol. 80, pp. 425–437, Jul. 2019.

[63] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.

[64] M. Jian, Q. Qi, J. Dong, Y. Yin, W. Zhang, and K.-M. Lam, "The OUC-vision large-scale underwater image database," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 1297–1302.

[65] K. Garg and S. K. Nayar, "Photorealistic rendering of rain streaks," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 996–1002, Jul. 2006.

[66] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2736–2744.

[67] W. Yang, R. T. Tan, J. Feng, J. Liu, S. Yan, and Z. Guo, "Joint rain detection and removal from a single image with contextualized deep networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 6, pp. 1377–1393, Jun. 2020.

[68] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Deep joint rain detection and removal from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1357–1366.

[69] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3855–3863.

[70] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2482–2491.

[71] V. Khrulkov and A. Babenko, "Neural side-by-side: Predicting human preferences for no-reference super-resolution evaluation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4986–4995.

[72] M. Cheon, S.-J. Yoon, B. Kang, and J. Lee, "Perceptual image quality assessment with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 433–442.

[73] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Nov. 2003, pp. 1398–1402.

[74] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[75] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[76] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[77] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[78] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2012.

[79] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 1473–1476.

[80] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[81] N. Venkatanath, D. Praneeth, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *Proc. Twenty 1st Nat. Conf. Commun. (NCC)*, Feb. 2015, pp. 1–6.

[82] Y. Wang, N. Li, Z. Li, Z. Gu, H. Zheng, B. Zheng, and M. Sun, "An imaging-inspired no-reference underwater color image quality assessment metric," *Comput. Elect. Eng.*, vol. 70, pp. 904–913, Aug. 2018.

[83] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.

[84] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2015.

[85] K. Panetta, S. Agaian, Y. Zhou, and E. J. Wharton, "Parameterized logarithmic framework for image enhancement," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 41, no. 2, pp. 460–473, Apr. 2011.

[86] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Comput. Vis. Image Understand.*, vol. 158, pp. 1–16, May 2017.

[87] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

[88] V. Khrulkov and A. Babenko, "Neural side-by-side: Predicting human preferences for no-reference super-resolution evaluation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4988–4997.

[89] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *Proc. ECCV*, 2020, pp. 492–511.

[90] Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2Self with dropout: Learning self-supervised denoising from single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1890–1898.

[91] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, "Neighbor2Neighbor: Self-supervised denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14781–14790.

[92] L. Chen, X. Lu, J. Zhang, X. Chu, and C. Chen, "HINet: Half instance normalization network for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 182–192.

[93] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proc. CVPR*, Jun. 2021, pp. 14821–14831.

[94] S. Cheng, Y. Wang, H. Huang, D. Liu, H. Fan, and S. Liu, "NBNet: Noise basis learning for image denoising with subspace projection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4896–4906.

[95] K. Purohit, M. Suin, A. N. Rajagopalan, and V. Naresh Boddeti, "Spatially-adaptive image restoration using distortion-guided networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2309–2319.

[96] C. Mou, Q. Wang, and J. Zhang, "Deep generalized unfolding networks for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17399–17410.

[97] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, "All-in-one image restoration for unknown corruption," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 17452–17462.

[98] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 11908–11915.

[99] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. V. Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.

[100] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 701–710.

[101] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.

[102] R. Zhou and S. Susstrunk, "Kernel modeling super-resolution on real low-resolution images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2433–2443.

[103] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8878–8887.

[104] M. Fritsche, S. Gu, and R. Timofte, "Frequency separation for real-world super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3599–3608.

[105] W. Du, H. Chen, and H. Yang, "Learning invariant representation for unsupervised image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14483–14492.

[106] S. Yan, Y. Liu, J. Li, and H. Xiao, "DDGAN: Double discriminators GAN for accurate image colorization," in *Proc. 6th Int. Conf. Big Data Inf. Analytics (BigDIA)*, Dec. 2020, pp. 214–219.

[107] S. Zheng, Z. Zhu, X. Zhang, Z. Liu, J. Cheng, and Y. Zhao, "Distribution-induced bidirectional generative adversarial network for graph representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7224–7233.

[108] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4791–4800.

[109] S. Wu, C. Dong, and Y. Qiao, "Blind image restoration based on cycle-consistent network," *IEEE Trans. Multimedia*, early access, Jan. 5, 2022, doi: 10.1109/TMM.2021.3139209.

[110] X. Pan, X. Zhan, B. Dai, D. Lin, C. C. Loy, and P. Luo, "Exploiting deep generative prior for versatile image restoration and manipulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7474–7489, Nov. 2022.

[111] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, "Pre-trained image processing transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12299–12310.

[112] I. O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "MLP-mixer: An all-MLP architecture for vision," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 24261–24272.

[113] H. Liu, Z. Dai, D. So, and Q. V. Le, "Pay attention to MLPs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 9204–9215.

[114] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "MAXIM: Multi-axis MLP for image processing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5769–5780.

[115] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[116] S. Anwar and N. Barnes, "Densely residual Laplacian super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1192–1204, Mar. 2022.

[117] A. Dudhane, S. W. Zamir, S. Khan, F. S. Khan, and M.-H. Yang, "Burst image restoration and enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5759–5768.

[118] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2480–2495, Jul. 2020.

[119] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[120] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski, "Deep photo: Model-based photograph enhancement and viewing," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 1–10, 2008.

[121] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.

[122] T. Michaeli and M. Irani, "Nonparametric blind super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 945–952.

[123] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[124] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[125] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV) workshops*, 2018, pp. 63–79.

[126] S. Anwar and N. Barnes, "Real image denoising with feature attention," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3155–3164.

[127] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and C. Fang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.

[128] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.

[129] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[130] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 1632–1640.

[131] A. Abuolaim and M. S. Brown, "Defocus deblurring using dual-pixel data," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2020, pp. 111–126.

[132] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4641–4650.

[133] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1406–1415.

[134] Z. Yue, Q. Zhao, L. Zhang, and D. Meng, "Dual adversarial network: Toward real-world noise removal and noise generation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 41–58.

[135] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 254–269.

[136] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3937–3946.

[137] M. Suin, K. Purohit, and A. N. Rajagopalan, "Spatially-attentive patch-hierarchical network for adaptive motion deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3606–3615.

[138] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.

[139] S. Chaudhari, V. Mithal, G. Polatkan, and R. Ramanath, "An attentive survey of attention models," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 5, pp. 1–32, Oct. 2021.

[140] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[141] A. Ignatov, N. Kobyshev, R. Timofte, and K. Vanhoey, "DSLR-quality photos on mobile devices with deep convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3277–3285.

[142] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773.

[143] S. Tang, Y. Wu, H. Qin, X. Xie, S. Yang, and J. Wang, "A constrained deformable convolutional network for efficient single image dynamic scene blind deblurring with spatially-variant motion blur kernels estimation," 2022, *arXiv:2208.10711*.

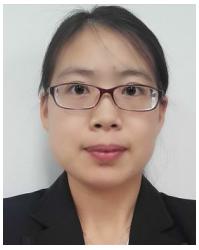
[144] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, Apr. 2012.

[145] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proc. 15th Eur. Signal Process. Conf.*, Sep. 2007, pp. 145–149.

[146] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[147] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[148] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[149] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[150] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
[151] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[152] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2016, pp. 694–711.
[153] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
[154] A. Lugmayr, M. Danelljan, and R. Timofte, "Unsupervised learning for real-world super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3408–3416.
[155] S. Maeda, "Unpaired image super-resolution using pseudo-supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 291–300.
[156] A. Kulesza and B. Taskar, "k-DPPs: Fixed-size determinantal point processes," in *Proc. 28th Int. Conf. Mach. Learn.*, Madison, WI, USA, 2011, pp. 1193–1200.
[157] T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4570–4580.
[158] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," *IEEE Trans. Knowl. Data Eng.*, early access, Nov. 23, 2022, doi: 10.1109/TKDE.2021.3130191.
[159] N. Wang, Y. Zhou, F. Han, H. Zhu, and J. Yao, "UWGAN: Underwater GAN for real-world underwater color restoration and dehazing," 2019, *arXiv:1912.10269*.
[160] X. Yu, Y. Qu, and M. Hong, "Underwater-GAN: Underwater image restoration via conditional generative adversarial network," in *Pattern Recognition and Information Forensics*. Cham, Switzerland: Springer, 2019, pp. 66–75.
[161] C. Desai, B. S. S. Reddy, R. A. Tabib, U. Patil, and U. Mudenagudi, "AquaGAN: Restoration of underwater images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 296–304.
[162] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
[163] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[164] J. Cao, Y. Li, K. Zhang, and L. Van Gool, "Video super-resolution transformer," 2021, *arXiv:2106.06847*.
[165] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted Windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.
[166] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.
[167] X. Chen, X. Wang, J. Zhou, and C. Dong, "Activating more pixels in image super-resolution transformer," 2022, *arXiv:2205.04437*.
[168] L. Zhao, Z. Zhang, T. Chen, D. Metaxas, and H. Zhang, "Improved transformer for high-resolution GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 18367–18380.
[169] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[170] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.
[171] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–16.

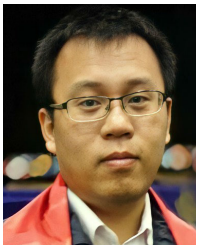


LUJUN ZHAI (Student Member, IEEE) received the B.S. degree in electrical and automation engineering from the University of Jinan, China, in 2012, and the M.S. degree from Prairie View A&M University, in 2017, where she is currently pursuing the Ph.D. degree in electrical and computer engineering. Her research interests include computer vision, image processing, deep learning, and especially focus on face recognition.

YONGHUI WANG (Senior Member, IEEE) received the B.S. degree in optoelectronics from Xidian University, Xi'an, China, in 1993, the M.S. degree in electrical engineering from the Beijing University of Technology, Beijing, China, in 1999, and the Ph.D. degree in computer engineering from Mississippi State University, Starkville, MS, USA, in 2003. From 1993 to 1996, he was a Research Engineer with the 41st Electrical Research Institute, Bengbu, China. From July 1999 to December 1999, he worked as an IT Specialist with IBM China, Beijing. From 2000 to 2003, he was a Research Assistant with the Visualization, Analysis, and Imaging Laboratory (VAIL), GeoResources Institute (GRI), Mississippi State University. He is currently with the Department of Computer Science, Prairie View A&M University, Prairie View, TX, USA. His research interests include image/video signal processing, computer vision, and artificial intelligence.

SUXIA CUI (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the Beijing University of Technology, Beijing, China, in 1999, and the Ph.D. degree in computer engineering from Mississippi State University, Starkville, MS, USA, in 2003.

She is currently an Associate Professor with the Electrical and Computer Engineering Department, Prairie View A&M University, Prairie View, TX, USA. Her research interests include cybersecurity, machine learning, and image processing. Her research is sponsored by the U.S. Department of Agriculture, the National Science Foundation, the Department of Defense, and the Department of Education.



YU ZHOU received the Ph.D. degree in electrical engineering from Prairie View A&M University, Prairie View, TX, USA, in 2019. His research interests include natural language processing, image processing, machine learning, deep learning, embedded systems, and autonomous underwater vehicle design.