IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.DOI

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[1]

[1]Electrical and Computer Engineering Department, Prairie View A&M University, 100 University Dr, Prairie View, TX 77446
[2]Computer Science Department, Prairie View A&M University, 100 University Dr, Prairie View, TX 77446

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

This work was partially sponsored by the United States NSF grant OAC-1827243 and the US Department of Education grant P120A180114.

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]–[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]–[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 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-

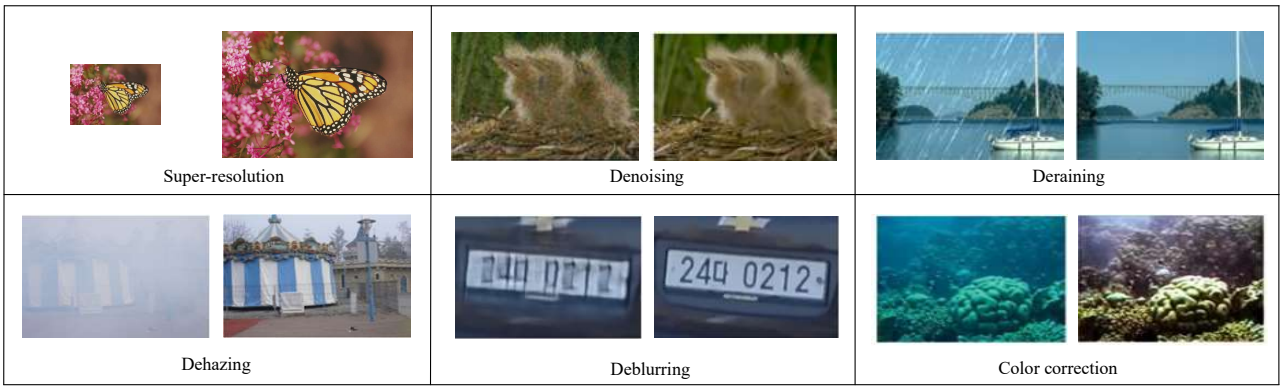

FIGURE 1. Categories of low-level image restoration.

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]–[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 denoising 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], [23]. 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 convo-

lution operation, k is a degradation kernel, and n is noise [24]. By using the maximum a posteriori (MAP) estimation, the latent image I_H is formulated as follows:

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

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 [25], BSDS500 [26], T91 [27], Set5 [28], Set14 [29], Urban100 [30], Manga109 [31] 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) [32], [33]. 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 [42] 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 [67] or (2) 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 [73], [74] 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 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

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

Name	Type	Image/pairs amount	Res. of HR image	Scale factors	Category
DIV 2K [25]	Paired	1,000	2048 × 1024	×2, ×3, ×4	People, handmade objects, animals, scenery, etc.
ImagePairs [34]	Paired	11,421	3474 × 2292, etc.	×2	Document, board, office, face, car, object, scenery, etc.
SupER [35]	Paired	85,050	2040 × 1080, etc.	×2, ×3, ×4	Banknote, book, coffee, doll, newspaper, loader, etc.
SR-RAW [36]	Paired	2,000	2040 × 1080, etc.	×2, ×3, ×4	Office, Car, Guidepost, bike, toy, building, book, street etc.
DRealSR [37]	Paired	2,507	5748 × 3746, etc.	×2, ×3, ×4	People, building, plant, poster, book, statue, etc.
City100 [38]	Paired	200	1218 × 870, etc.	×2.9, ×2.4	People, building, scenery, boat, bridge, street, etc.
T91 [27]	Unpaired	91	250 × 200	–	Fruits, cars, faces, etc.
BSDS500 [26]	Unpaired	500	430 × 370	–	people, animal, building, scenery, plant, etc. etc.
Set5 [28]	Unpaired	5	300 × 340	–	Butterfly, baby, bird, head, and women.
Set14 [29]	Unpaired	14	500 × 450	–	Face, animal, flower, animated character, insect, etc.
Urban100 [30]	Unpaired	1,000	800 × 1150	–	Urban buildings
CelebA [39]	Unpaired	202,599	2048 × 1024	–	Celebrities with 40 defined attributes
MS-COCO [40]	Unpaired	164,000	640 × 480	–	Labeled objects with over 80 classes
VOC2012 [41]	Unpaired	11,530	500 × 400	–	Labelled objects with over 20 classes
Manga109 [31]	Unpaired	109	300 × 340	–	Manga

TABLE 2. List of benchmark datasets used in denoising.

Name	Pairs amount	Details
SIDD [42]	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 [43]	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 [44]	60,000	Noisy fluorescence microscopy images with different noise levels and averaged images served as GT.
PolyU [45]	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 [46]	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 [47]	50	They captured noisy images with different ISO values and appropriately adjusted exposure times. The low ISO images serve as GT.

images are required. Table 7 reports widely used metrics for image quality assessment, including no-reference and full-reference metrics.

***MS-SSIM*:** As an FR IQA, multi-scale structural similarity (MS-SSIM) [75] 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.

***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.

***SSIM*:** The Structure similarity index (SSIM) [76] is an FR image quality metric that measures structural similarity and also performs luminance and contrast comparisons. Compared to PSNR, SSIM reflects visual quality better than PSNR [76]. Generally, PSNR and SSIM are used jointly to evaluate the quality of the restored image.

***IFC*:** The information fidelity criterion (IFC) [77] 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.

***VSNR*:** Visual signal-to-noise ratio (VSNR) [78] is an

TABLE 3. List of benchmark datasets used in dehazing.

Name	Type	Pairs amount	Scene
D-Hazy [48]	Synthetic	1,400+	Indoor
I-HAZE [49]	Artificial	35	Indoor
SOTS [50]	Synthetic	500	Indoor
O-HAZE [51]	Artificial	45	Outdoor
HazeRD [52]	Synthetic	15	Outdoor
Dense-HAZE [53]	Artificial	33	Outdoor
NH-HAZE [54]	Artificial	55	Outdoor
BeDDE [55]	Natural	200+	Outdoor
MRFID [56]	Natural	200	Outdoor
SOTS [50]	Synthetic	500	Outdoor

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.

***FSIM*:** FSIM [79], 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.

***NIQE*:** The natural image quality evaluator (NIQE) [80] 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.

***SR-SIM*:** Spectral residual based similarity (SR-SIM) [81], as an FR real-time IQA, is based on the spectral residual visual saliency model (SRVS) [82]. 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.

***PIQUE*:** The perception-based image quality evaluator (PIQUE) [83] 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.

***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.* [84] 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.

***UCIQE*:** Underwater color image quality evaluation metric (UCIQE) [85], quantifies the non-uniform color casts, blurring, and low contrast, and then linearly combines these three components.

***UIQM*:** Another NR criterion for evaluating the quality of underwater images, UIQM [86], 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 [87] on the intensity image.

***NRQM*:** NRQM [88] 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.

***LPIPS*:** The learned perceptual image patch similarity (LPIPS) [89] 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.

***IQT*:** Image quality transformer (IQT) [90] 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 ranked 1st among 13 participants in the NTIRE 2021 perceptual IQA challenge.

***NeuralSBS*:** Neural side-by-side (NeuralSBS) [91] 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

TABLE 4. List of benchmark datasets used in deblurring.

Name	Type	Blur Model	Content	Details
GoPro [57]	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 [58]	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 [59]	Video	Non-uniform	Objects & Scenes	300 video pairs in dynamic scenes. Blurry frames are generated by the interpolated videos.
DVD [60]	Video	Non-uniform	Objects & Scenes	71 video pairs with camera shake motion blur by averaging frames and corresponding sharp version.
Shen *et al.* [61]	Image	Uniform	Face	6,564 sharp images, 130 M training / 16K testing blurry face images with blur kernels and Gaussian noise.
Real Blur [62]	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 [63]	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 [64]	Object Detection	MUED contains 8,600 underwater images of 430 individual groups of conspicuous objects with labeled ground-truth information.
RUIE [65]	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 [66]	Object Detection	OUC-VISION dataset provides 4,400 underwater images and bounding box annotations.
OUC [65]	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 [68]	100	Test	Synthesized images with only one type of rain streak using photo-realistic rendering techniques.
Rain100L [69]	12	Test	100 synthesized rainy images with only one type of rain streak.
Rain100H [70]	100	Test	Synthesized dataset with five streak directions.
Rain14000 [71]	14,000	Train & Test	A dataset containing 14,000 pairs of rainy/clean images.
Raindrop [72]	1,119	Train	1,119 pairs of rainy/clean images and raindrops attached to a glass window
Rain1800 [70]	1,800	Train	Synthesized rainy images in two types: photo-realistic rendering and simulated sharp line streaks.

TABLE 7. Commonly used IQA.

Method	Publication	Full/No-reference	Keywords
MS-SSIM	ACSSC-2003 [75]	Full-reference	Multi-scale SSIM, image synthesis, cross-scale calibration
PSNR	-	Full-reference	Mean squared error
SSIM	TIP-2004 [76]	Full-reference	Structure similarity, luminance, contrast, structures
IFC	TIP-2005 [77]	Full-reference	Nature scene statistics, Gaussian scale mixtures
VSNR	TIP-2007 [78]	Full-reference	Two-stage, contrast threshold, global precedence
FSIM	TIP-2011 [79]	Full-reference	Low-level feature, phase congruency
NIQE	SPL-2012 [80]	No-reference	Quality-aware features, multivariate Gaussian model, natural scene statistic model
SR-SIM	ICIP-2012 [81]	Full-reference	Spectral residual visual saliency, bottom-up attention
PIQUE	NCC-2015 [83]	No-reference	Perceptually significant spatial regions, block level distortion map
CCF	ISO4-2018 [84]	No-reference	Combination of colorfulness index, contrast index and fog density index
UCIQE	TIP-2015 [85]	No-reference	Linear combination of chroma, saturation, and contrast
UIQM	ISO4-2016 [86]	No-reference	Colorfulness measure (UICM), sharpness measure (UISM), contrast measure (UIConM)
NRQM	CVIU-2017 [88]	No-reference	Statistical features, regression forests, linear regression model
LPIPS	CVPR-2018 [89]	Full-reference	Deep features, human perceptual similarity
IQT	CVPR-2021 [90]	Full-reference	CNN, Transformer, super-resolution
NeuralSBS	CVPR-2021 [91]	No-reference	CNN, human preference prediction, super-resolution

studies on IR, some necessary diagrams are provided. The following subsections present these methods in detail.

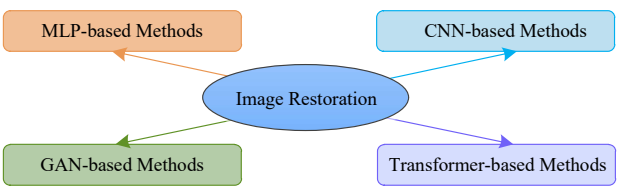

FIGURE 2. The taxonomy of existing single image restoration techniques.

A. CNN-BASED METHODS

CNN has dominated computer vision for nearly 10 years. Recently, CNN architectures [118]–[123] have made significant advancements for IR and have shown vast superiority over conventional restoration approaches [124]–[127], as they can learn generalizable priors from large-scale datasets. Driven by recent 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 [128]–[130], dilated convolutions [70], [131], dense connections [130], [130], hierarchical structures [106], [132], [133], encoder-decoder [43], [122], [134]–[139], multi-stage frameworks [14], [57], [140]–[142], and attention mechanisms [143], [144].

Among CNN designs, encoder-decoder architectures [43], [134]–[136], [138], [139] 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 [123], [129],

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

Methods	Publication	Category	Keywords
MIRNet	ECCV-2020 [92]	CNN	Multi-scale, attention-based aggregation, selective kernel feature fusion, deraining, deblurring, denoising
Self2Self	CVPR-2020 [93]	CNN	Self-supervised learning, encoder-decoder network, denoising
Neighbor2Neighbor	CVPR-2021 [94]	CNN	Self-supervised framework, neighbor sub-sampler image pair generation, denoising
HINet	CVPRW-2021 [95]	CNN	Half instance normalization, multi-stage network, deraining, deblurring, denoising
MPRNet	CVPR-2021 [96]	CNN	Interlinked multi-stage architecture, encoder-decoder, supervised attention module, deraining, deblurring, denoising
NBNet	CVPR-2021 [97]	CNN	Subspace Projection, non-local subspace attention, UNet-based architecture, denoising
SPAIR	ICCV 2021 [98]	CNN	Distortion-guided feature extraction, spatially-guided restoration, deblurring, shadow-removal, deraining
DGUNet	CVPR-2022 [99]	CNN	Interpretable deep unfolding network, proximal gradient descent, multi-stage network, denoising, deblurring
AirNet	CVPR-2022 [100]	CNN	All-in-one solution, multi-pairs of encoder-decoder, contrastive learning, denoising, deraining, dehazing
Water-Net	TIP-2019 [63]	CNN	Gated fusion network, feature transformation unit, underwater image enhancement
FFA-Net	AAAI-2020 [101]	CNN	Channel attention with pixel attention, feature fusion structure, dehazing
DPIR	TPAMI-2021 [102]	CNN	Plug-and-play image restoration, image prior, non-blind denoising
CinCGAN	CVPRW-2018 [103]	GAN	Unsupervised SR, Cycle-in-Cycle GAN, domain translation, super-resolution
DeblurGAN	CVPR-2018 [104]	GAN	End-to-end blind motion deblurring, gradient penalty, perceptual loss, deblurring
KMSR	ICCV-2019 [105]	GAN	Blur-kernel modeling, blur kernel pool augment, super-resolution
DeblurGAN-v2	ICCV-2019 [106]	GAN	Feature pyramid deblurring, double-scale RaGAN-LS discriminator, deblurring
DSGAN	ICCVW-2019 [107]	GAN	Frequency separation, unsupervised learning, domain translation, super-resolution
LIR	CVPR-2020 [108]	GAN	Unsupervised learning, disentangle deep representation, domain-transfer
DDGAN	CVPRW-2020 [109]	GAN	Unsupervised learning, blind restoration, CinCGAN-based, double discriminators, image colorization
DBGAN	CVPR-2020 [110]	GAN	Distribution-induced bidirectional, structure-aware prior distribution estimation, Magnetic resonance imaging (MRI) reconstruction
BSRGAN	ICCV-2021 [111]	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 [112]	GAN	Unsupervised blind IR, CinCGAN-based, content constraint loss, super-resolution
DGP	TPAMI-2021 [113]	GAN	Deep image prior, multi-stage, super-resolution, colorization, inpainting
AquaGAN	CVPRW-2022 [113]	GAN	Attenuation coefficients estimation, weighted combination of content and style loss, underwater image restoration
IPT	CVPR-2021 [114]	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 [50]	Transformer	Combination of channel attention and self-attention, overlapping cross-attention, super-resolution
MLP-Mixer	NIPS-2021 [115]	MLP	Pure MLP, token-mixing, channel-mixing, milestone of MLP in vision
gMLP	NIPS-2021 [116]	MLP	Spatial Gating Unit (SGU), attention-free, comparable performance to Transformers in vision applications
MAXIM	CVPR-2022 [117]	MLP	Multi-axis gating, multi-axis self-attention, Cross Gating Block (CGB), denoising, deblurring, deraining, dehazing, enhancement

[145], [146] 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 [121] 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 [96] employs a multi-stage approach, which progressively restores images by decomposing the challenging IR into smaller easier subtasks. Unlike previous multi-stage approaches [14], [57], [140]–[142] 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.

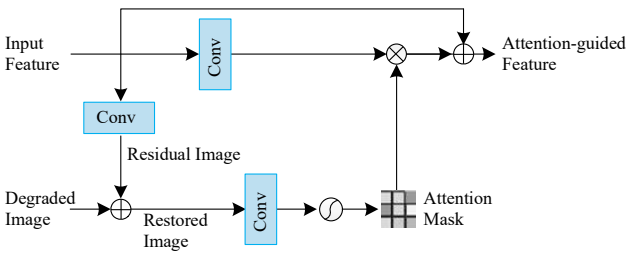

FIGURE 3. Supervised attention module.

To enhance the transformation modeling capability of CNNs, a deformable convolutional network [147] 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) [148] 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 [93] 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 [94] 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 [149], [150], 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

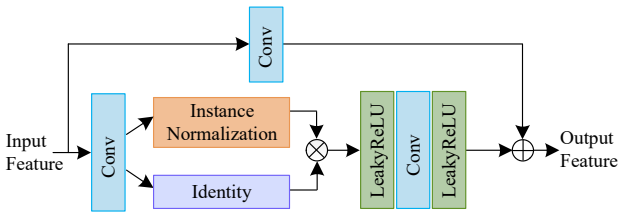

FIGURE 4. The diagram of HINet Block [95].

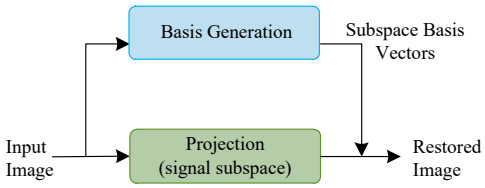

FIGURE 5. NBNet [97] denoises via subspace projection: 1) Basis generation: generates subspace basis vectors from feature representations; 2) Projection: maps feature representations into the signal subspace.

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 [95], which significantly boost the performance of IR.

NBNet [97] 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).

DGUNet [99] 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 [98] 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 [151] 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.

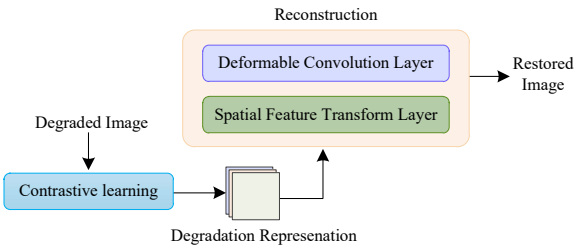

FIGURE 6. The diagram of AirNets [151].

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 [152]. 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:

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

As a popular GAN variant, Conditional GAN [153], 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 [154]) with skip-connection and an optimized perceptual loss calculated on feature maps of the VGG network [155]. To deal with motion blur, DeblurGAN [104], as an end-to-end blind IR model, exploits Wasserstein GAN [156] with the gradient penalty [157] and the perceptual loss [158]. DeblurGAN-v2 [106], featured with light-weight and fast, introduces a feature pyramid network (FPN) as a core building block into its generator. KMSR [105] improves the blind SR performance by integrating the blur-kernel estimation into GAN.

CycleGAN [159], 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 [159]–[161] are developed to do image SR from unpaired LR-HR training samples, which have obvious advantages in dealing with unknown types of degradation. DSGAN [107] makes a step further and separates the low and high image frequencies and treats them differently during training. Unsupervised SR model CinCGAN [103] 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 [112] makes two-fold explorations: (1) preserving the low-frequency content fidelity and (2) suppressing the high-frequency artifacts.

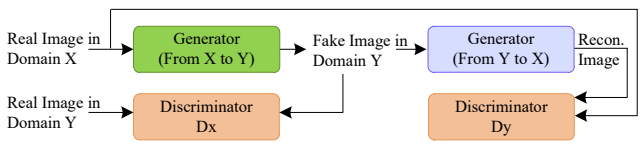

FIGURE 7. The diagram of CycleGAN [151].

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 [162] 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 [110] is a distribution-induced bidirectional GAN, which first proposes graph representation learning and utilizes a structure-aware prior distribution estimation based on DPP [163] for latent representation by prototype learning. The diagram of DBGAN is depicted in Fig. 8.

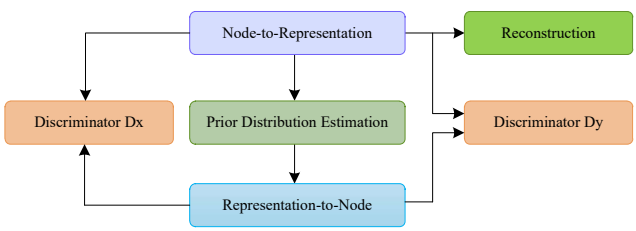

FIGURE 8. The diagram of DBGAN [110].

SinGAN [164] shows that a randomly-initialized GAN model is able to capture rich patch statistics being trained from a single image. DGP [113] 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) [165]–[167]. A representative work, AquaGAN [168], 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.

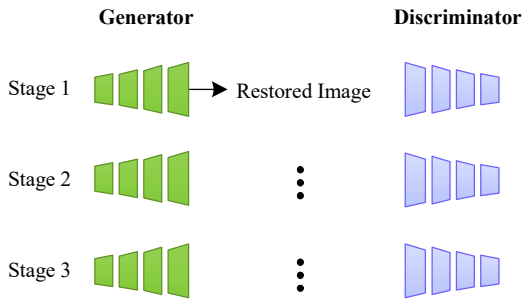

FIGURE 9. The diagram of DGP [151].

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 [169]) have exemplified their great potential as alternatives to the go-to CNN models.

By leveraging the self-attention (SA) mechanism [170], 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:

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

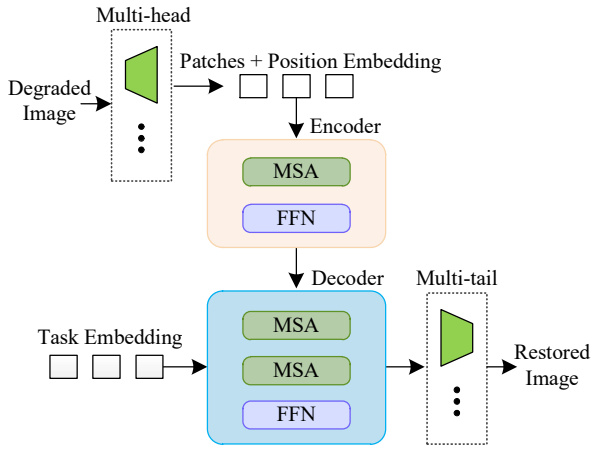

FIGURE 10. The architecture of IPT [114].

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 [114]) 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 [114], [171] 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 [172] 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 [173]) 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 [174]) 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 [115], 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.

The Google Brain team further studied the necessity of self-attention modules in Transformers and proposed the gMLP [116] 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).

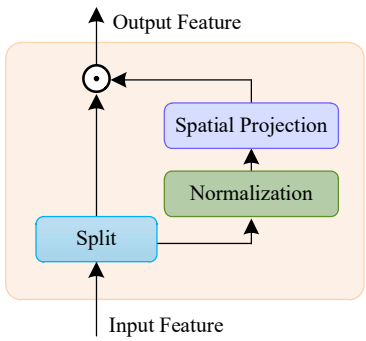

FIGURE 11. Spatial Gating Unit (SGU) in gMLP [116].

Zhao *et al.* [175] 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 [117], 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.

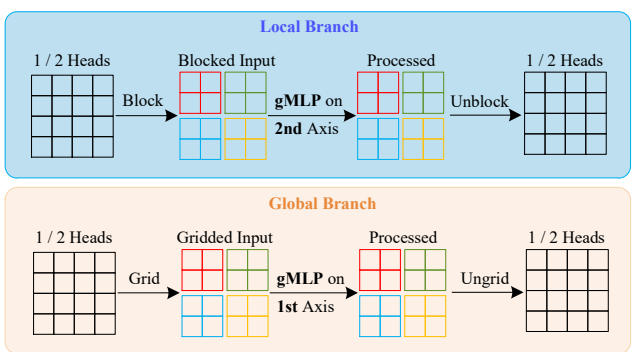

FIGURE 12. Multi-axis gated MLP [117].

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 connections [176] and more complex forms of convolution [177]–[179], 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 [106] , MPRNet [96], MIRNet [92], DGUNet [99], MAXIM [117], Restormer [12], Uformer-B [22], IPT [114], SwinIR [21], HAT-L [50], and BSRGAN [111], covering multiple kinds of approaches mentioned in Section V. Experiments are conducted on six benchmark datasets for different IR tasks, GoPro [57] for motion deblurring, SIDD [42] for denoising, Rain100L [42] for deraining, SOTS [50] for outdoor dehazing, Set14 [29] for SR, and UIEB [63] 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 [114] 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 [172], 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 [50] 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 [63] and AquaGAN [113], 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 [57].

Method	PSNR	SSIM	FSIM	IFC
DeblurGAN-v2 [106]	29.55	0.934	0.9431	2.6929
MPRNet [96]	32.66	0.959	0.9634	3.1568
HINet [95]	32.71	0.959	0.9772	3.4845
MAXIM [117]	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 [42].

Method	PSNR	SSIM	FSIM	IFC
MPRNet [96]	39.71	0.958	0.9234	3.3584
MIRNet [121]	39.72	0.959	0.9472	3.4816
Uformer-B [22]	39.89	0.960	0.9675	3.4582
MAXIM [117]	39.96	0.960	0.9851	3.6423
HINet [95]	39.99	0.958	0.9751	3.6829
DGUNet [99]	39.91	0.962	0.9809	3.6713
Restormer [12]	40.02	0.965	0.9827	3.6804

TABLE 11. Deraining performance of the representative IR algorithms on Rain100L [42].

Method	PSNR	SSIM	FSIM	IFC
HINet [95]	37.28	0.9710	0.9643	3.6013
MAXIM [117]	38.92	0.9772	0.9715	3.6815
Restormer [12]	38.99	0.978	0.9823	3.7206
IPT [114]	41.62	0.9881	0.9862	3.7542

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.

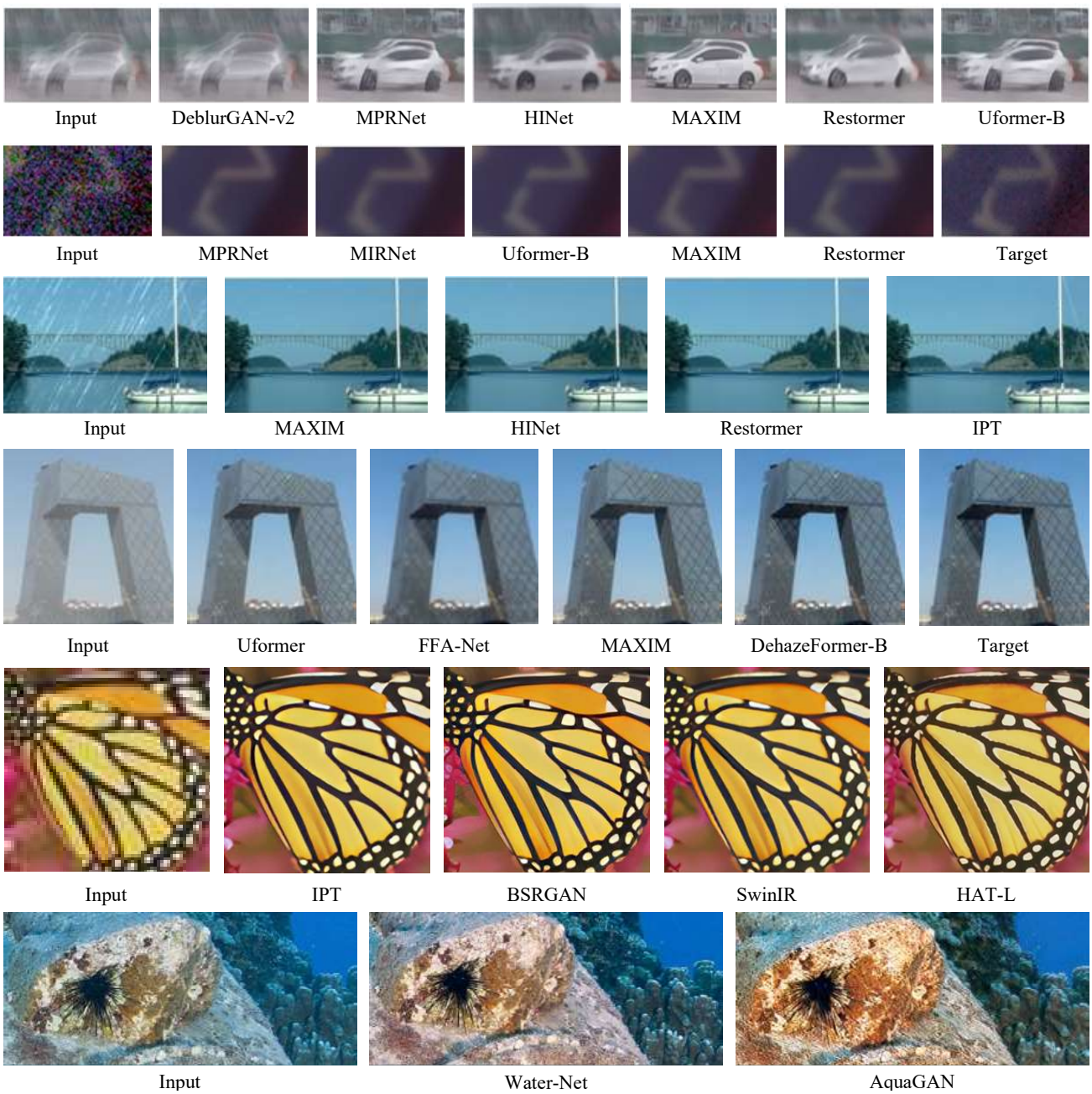

FIGURE 13. **Visual comparisons of representative SOTA IR algorithms for six different tasks. From top row to bottom row: motion deblurring on GoPro [57] , denoising on SIDD [42], draining on Rain100L [42], outdoor dehazing on SOTS [50], 4 $\times$ upscaling super-resolution on Set14 [29], and underwater image restoration on UIEB [63].**

TABLE 12. **Dehazing performance of the representative IR algorithms on SOTS Outdoor [50].**

Method	PSNR	SSIM	FSIM	IFC
Uformer-B [22]	31.10	0.9760	0.9562	3.5329
FFA-Net [101]	33.57	0.9804	0.9627	3.6817
MAXIM [117]	34.19	0.9702	0.9813	3.7412
DehazeFormer-B [11]	34.95	0.9840	0.9853	3.7541

TABLE 13. **4$\times$ upscaling performance of the representative SR algorithms on Set14 [29].**

Method	PSNR	SSIM	FSIM	IFC
IPT [114]	29.01	0.7824	0.9325	2.5336
BSRGAN [111]	29.08	0.7923	0.9207	2.5723
SwinIR [21]	29.15	0.7958	0.9314	2.6278
HAT-L [174]	29.47	0.8015	0.9423	2.6804

TABLE 14. Performance comparison of SOTA underwater IR algorithms on UIEB [63].

Method	UCIQE	UIQM	CCF
Water-Net [63]	5.4637	0.7853	26.49
AquaGAN [113]	5.9612	0.8391	33.24

A. IMAGE DATASETS

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

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] Michael O Macaulay and Mahmood Shafiee. Machine learning techniques for robotic and autonomous inspection of mechanical systems and civil infrastructure. Autonomous Intelligent Systems, 2(1):1–25, 2022.

[2] Andreas Aakerberg, Anders S Johansen, Kamal Nasrollahi, and Thomas B Moeslund. Semantic segmentation guided real-world super-resolution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 449–458, 2022.

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. arXiv preprint arXiv:2204.04676, 2022.

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

[5] Jianxiang Lu, Fei Yuan, Weidi Yang, and En Cheng. An imaging information estimation network for underwater image color restoration. IEEE Journal of Oceanic Engineering, 46(4):1228–1239, 2021.

[6] Yan Wang, Wei Song, Giancarlo Fortino, Li-Zhe Qi, Wenqiang Zhang, and Antonio Liotta. An experimental-based review of image enhancement and image restoration methods for underwater imaging. IEEE Access, 7:140233–140251, 2019.

[7] Suxia Cui, Yu Zhou, Yonghui Wang, and Lujun Zhai. Fish detection using deep learning. Applied Computational Intelligence and Soft Computing, 2020, 2020.

[8] Yanting Pei, Yaping Huang, Qi Zou, Xingyuan Zhang, and Song Wang. Effects of image degradation and degradation removal to CNN-based image classification. IEEE transactions on pattern analysis and machine intelligence, 43(4):1239–1253, 2019.

[9] Yong Xu, Jie Wen, Lunke Fei, and Zheng Zhang. Review of video and image defogging algorithms and related studies on image restoration and enhancement. IEEE Access, 4:165–188, 2016.

[10] Zhixiang Hao, Shaodi You, Yu Li, Kunming Li, and Feng Lu. Learning from synthetic photorealistic raindrop for single image raindrop removal. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pages 4340–4349, 2019.

[11] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. arXiv preprint arXiv:2204.03883, 2022.

[12] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 5728–5739, June 2022.

[13] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In CVPR, 2021.

[14] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, and John Paisley. Lightweight pyramid networks for image deraining. IEEE transactions on neural networks and learning systems, 31(6):1794–1807, 2019.

[15] Yuqian Zhou, Jianbo Jiao, Haibin Huang, Yang Wang, Jue Wang, Honghui Shi, and Thomas Huang. When AWGN-based denoiser meets real noises. arXiv preprint arXiv:1904.03485, 2019.

[16] Lujun Zhai, Yonghui Wang, Suxia Cui, and Yu Zhou. Enhancing underwater image using degradation adaptive adversarial network. In 2022 IEEE International Conference on Image Processing (ICIP), pages 4093–4097. IEEE, 2022.

[17] Honggang Chen, Xiaohai He, Linbo Qing, Yuanyuan Wu, Chao Ren, Ray E. Sheriff, and Ce Zhu. Real-world single image super-resolution: A brief review. Information Fusion, 79:124–145, 2022.

[18] Muwei Jian, Xiangyu Liu, Hanjiang Luo, Xiangwei Lu, Hui Yu, and Junyu Dong. Underwater image processing and analysis: A review. Signal Processing: Image Communication, 91:116088, 2021.

[19] Jingwen Su, Boyan Xu, and Hujun Yin. A survey of deep learning approaches to image restoration. Neurocomputing, 487:46–65, 2022.

[20] Rini Smita Thakur, Shubhojeet Chatterjee, Ram Narayan Yadav, and Lalita Gupta. Image de-noising with machine learning: A review. IEEE Access, 9:93338–93363, 2021.

[21] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. arXiv preprint arXiv:2108.10257, 2021.

[22] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general U-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 17683–17693, 2022.

[23] Jingwen Su, Boyan Xu, and Hujun Yin. A survey of deep learning approaches to image restoration. Neurocomputing, 487:46–65, 2022.

[24] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 1604–1613, 2019.

[25] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pages 126–135, 2017.

[26] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. IEEE transactions on pattern analysis and machine intelligence, 33(5):898–916, 2010.

[27] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. IEEE transactions on image processing, 19(11):2861–2873, 2010.

[28] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie line Alberi Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In Proceedings of the British Machine Vision Conference, pages 135.1–135.10. BMVA Press, 2012.

[29] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In International conference on curves and surfaces, pages 711–730. Springer, 2010.

[30] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5197–5206, 2015.

[31] Azuma Fujimoto, Toru Ogawa, Kazuyoshi Yamamoto, Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa. Manga109 dataset and creation of metadata. In Proceedings of the 1st international workshop on comics analysis, processing and understanding, pages 1–5, 2016.

[32] Hongwu Yuan, Yiqing Wang, Guoming Xu, and Feng Wang. Research on super-resolution reconstruction of single-frame image of infrared focal plane polarization imaging. In Eighth Symposium on Novel Photoelectronic Detection Technology and Applications, volume 12169, pages 603–609. SPIE, 2022.

[33] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[34] Hamid Reza Vaezi Joze, Ilya Zharkov, Karlton Powell, Carl Ringler, Luming Liang, Andy Roulston, Moshe Lutz, and Vivek Pradeep. ImagePairs: Realistic super resolution dataset via beam splitter camera rig. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 2190–2200, Los Alamitos, CA, USA, June 2020. IEEE Computer Society.

[35] Thomas Köhler, Michel Bätz, Farzad Naderi, André Kaup, Andreas Maier, and Christian Riess. Toward bridging the simulated-to-real gap: Benchmarking super-resolution on real data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(11):2944–2959, 2020.

[36] Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. Zoom to learn, learn to zoom. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 3757–3765, 2019.

[37] Pengxu Wei, Ziwei Xie, Hannan Lu, ZongYuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In Proceedings of the European Conference on Computer Vision, 2020.

[38] Chang Chen, Zhiwi Xiong, Xinmei Tian, Zheng-Jun Zha, and Feng Wu. Camera lens super-resolution. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 1652–1660, Los Alamitos, CA, USA, jun 2019. IEEE Computer Society.

[39] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In Proceedings of International Conference on Computer Vision (ICCV), December 2015.

[40] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, Computer Vision – ECCV 2014, pages 740–755, Cham, 2014. Springer International Publishing.

[41] Mark Everingham, S. M. Ali Eslami, Luc Van Gool, Christopher K. I. Williams, John M. Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, 111:98–136, 2014.

[42] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1692–1700, 2018.

[43] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3291–3300, June 2018.

[44] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A Poisson-Gaussian denoising dataset with real fluorescence microscopy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[45] Jun Xu, Hui Li, Zhetong Liang, David Zhang, and Lei Zhang. Real-world noisy image denoising: A new benchmark. arXiv preprint arXiv:1804.02603, 2018.

[46] Benoit Brummer and Christophe De Vleeschouwer. Natural image noise dataset. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, June 2019.

[47] Tobias Plötz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[48] Cosmin Ancuti, Codruta Orniana Ancuti, and Christophe De Vleeschouwer. D-HAZY: A dataset to evaluate quantitatively dehazing algorithms. 2016 IEEE International Conference on Image Processing (ICIP), pages 2226–2230, 2016.

[49] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images. In International Conference on Advanced Concepts for Intelligent Vision Systems, pages 620–631. Springer, 2018.

[50] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing, 28(1):492–505, 2019.

[51] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pages 754–762, 2018.

[52] Yanfu Zhang, Li Ding, and Gaurav Sharma. Hazerd: an outdoor scene dataset and benchmark for single image dehazing. In 2017 IEEE international conference on image processing (ICIP), pages 3205–3209. IEEE, 2017.

[53] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In 2019 IEEE international conference on image processing (ICIP), pages 1014–1018. IEEE, 2019.

[54] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 444–445, 2020.

[55] Shiyu Zhao, Lin Zhang, Shuaiyi Huang, Ying Shen, and Shengjie Zhao. Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines. IEEE Transactions on Image Processing, 29:6947–6962, 2020.

[56] Wei Liu, Fei Zhou, Tao Lu, Jiang Duan, and Guoping Qiu. Image defogging quality assessment: Real-world database and method. IEEE Transactions on Image Processing, 30:176–190, 2021.

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

[58] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 5572–5581, 2019.

[59] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2019.

[60] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[61] Ziyi Shen, Wei-Sheng Lai, Tingfa Xu, Jan Kautz, and Ming-Hsuan Yang. Deep semantic face deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[62] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In European Conference on Computer Vision, pages 184–201. Springer, 2020.

[63] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing, 29:4376–4389, January 2020.

[64] Muwei Jian, Qiang Qi, Hui Yu, Junyu Dong, Chaoran Cui, Xiushan Nie, Huaxiang Zhang, Yilong Yin, and Kin-Man Lam. The extended marine underwater environment database and baseline evaluations. Applied Soft Computing, 80:425–437, 2019.

[65] Risheng Liu, Xin Fan, Ming Zhu, Minjun Hou, and Zhongxuan Luo. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. 30(12), 2020.

[66] Muwei Jian, Qiang Qi, Junyu Dong, Yinlong Yin, Wenyin Zhang, and Kin-Man Lam. The OUC-vision large-scale underwater image database. In 2017 IEEE International Conference on Multimedia and Expo (ICME), pages 1297–1302, 2017.

[67] Kshitiz Garg and Shree K. Nayar. Photorealistic rendering of rain streaks. ACM Transactions on Graphics, 25(3):996–1002, July 2006.

[68] Yu Li, Robby T. Tan, Xiaojie Guo, Jiangbo Lu, and Michael S. Brown. Rain streak removal using layer priors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

[69] Wenhan Yang, Robby T Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. IEEE transactions on pattern analysis and machine intelligence, 42(6):1377–1393, 2019.

[70] Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1357–1366, July 2017.

[71] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3855–3863, 2017.

[72] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[73] Valentin Khrulkov and Artem Babenko. Neural side-by-side: Predicting human preferences for no-reference super-resolution evaluation. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 4986–4995, 2021.

[74] Manri Cheon, Sung-Jun Yoon, Byungyeon Kang, and Junwoo Lee. Perceptual image quality assessment with transformers. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 433–442, 2021.

[75] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, volume 2, pages 1398–1402. IEEE, 2003.

[76] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4):600–612, 2004.

[77] Hamid R Sheikh, Alan C Bovik, and Gustavo De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Transactions on image processing, 14(12):2117–2128, 2005.

[78] Damon M Chandler and Sheila S Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. IEEE transactions on image processing, 16(9):2284–2298, 2007.

[79] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. FSIM: A feature similarity index for image quality assessment. IEEE transactions on Image Processing, 20(8):2378–2386, 2011.

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

[81] Lin Zhang and Hongyu Li. SR-SIM: A fast and high-performance iqa index based on spectral residual. In 2012 19th IEEE international conference on image processing, pages 1473–1476. IEEE, 2012.

[82] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In 2007 IEEE Conference on computer vision and pattern recognition, pages 1–8. IEEE, 2007.

[83] N Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In 2015 Twenty First National Conference on Communications (NCC), pages 1–6. IEEE, 2015.

[84] Yan Wang, Na Li, Zongying Li, Zhaorui Gu, Haiyong Zheng, Bing Zheng, and Mengnan Sun. An imaging-inspired no-reference underwater color image quality assessment metric. Computers & Electrical Engineering, 70:904–913, 2018.

[85] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. IEEE Transactions on Image Processing, 24(12):6062–6071, 2015.

[86] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. IEEE Journal of Oceanic Engineering, 41(3):541–551, 2015.

[87] Karen Panetta, Sos Agaian, Yicong Zhou, and Eric J Wharton. Parameterized logarithmic framework for image enhancement. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 41(2):460–473, 2010.

[88] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. Computer Vision and Image Understanding, 158:1–16, 2017.

[89] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 586–595, 2018.

[90] Manri Cheon, Sung-Jun Yoon, Byungyeon Kang, and Junwoo Lee. Perceptual image quality assessment with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 433–442, 2021.

[91] Valentin Khrulkov and Artem Babenko. Neural side-by-side: Predicting human preferences for no-reference super-resolution evaluation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 4988–4997, June 2021.

[92] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In ECCV, 2020.

[93] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2self with dropout: Learning self-supervised denoising from single image. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 1890–1898, 2020.

[94] Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 14781–14790, June 2021.

[95] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. HINet: Half instance normalization network for image restoration. In

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 182–192, 2021.

[96] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In CVPR, 2021.

[97] Shen Cheng, Yuzhi Wang, Haibin Huang, Donghao Liu, Haoqiang Fan, and Shuaicheng Liu. NBNet: Noise basis learning for image denoising with subspace projection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 4896–4906, June 2021.

[98] Kuldeep Purohit, Maitreya Suin, AN Rajagopalan, and Vishnu Naresh Boddeti. Spatially-adaptive image restoration using distortion-guided networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 2309–2319, 2021.

[99] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 17399–17410, 2022.

[100] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-In-One Image Restoration for Unknown Corruption. In IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, June 2022.

[101] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 11908–11915, 2020.

[102] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10):6360–6376, 2022.

[103] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 701–710, 2018.

[104] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 8183–8192, 2018.

[105] Ruofan Zhou and Sabine Susstrunk. Kernel modeling super-resolution on real low-resolution images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 2433–2443, 2019.

[106] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 8878–8887, 2019.

[107] Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pages 3599–3608. IEEE, 2019.

[108] Wenchao Du, Hu Chen, and Hongyu Yang. Learning invariant representation for unsupervised image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 14483–14492, 2020.

[109] Shijie Yan, Yu Liu, Jingbei Li, and Huaxin Xiao. DDGAN: Double discriminators gan for accurate image colorization. In 2020 6th International Conference on Big Data and Information Analytics (BigDIA), pages 214–219. IEEE, 2020.

[110] Shuai Zheng, Zhenfeng Zhu, Xingxing Zhang, Zhizhe Liu, Jian Cheng, and Yao Zhao. Distribution-induced bidirectional generative adversarial network for graph representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7224–7233, 2020.

[111] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In IEEE International Conference on Computer Vision, pages 4791–4800, 2021.

[112] Shixiang Wu, Chao Dong, and Yu Qiao. Blind image restoration based on cycle-consistent network. IEEE Transactions on Multimedia, 2022.

[113] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[114] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12299–12310, 2021.

[115] Ilya O Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, et al. MLP-mixer: An all-mlp architecture for vision. Advances in Neural Information Processing Systems, 34:24261–24272, 2021.

[116] Hanxiao Liu, Zihang Dai, David So, and Quoc V Le. Pay attention to mlps. Advances in Neural Information Processing Systems, 34:9204–9215, 2021.

[117] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. MAXIM: Multi-axis mlp for image processing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5769–5780, 2022.

[118] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV), pages 286–301, 2018.

[119] Saeed Anwar and Nick Barnes. Densely residual laplacian super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

[120] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5759–5768, 2022.

[121] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In European Conference on Computer Vision, pages 492–511. Springer, 2020.

[122] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 14821–14831, 2021.

[123] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(7):2480–2495, 2020.

[124] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence, 33(12):2341–2353, 2010.

[125] Johannes Kopf, Boris Neubert, Billy Chen, Michael Cohen, Daniel Cohen-Or, Oliver Deussen, Matt Uyttendaele, and Dani Lischinski. Deep photo: Model-based photograph enhancement and viewing. ACM transactions on graphics (TOG), 27(5):1–10, 2008.

[126] Radu Timofte, Vincent De Smet, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In Proceedings of the IEEE international conference on computer vision, pages 1920–1927, 2013.

[127] Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, pages 945–952, 2013.

[128] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4681–4690, 2017.

[129] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE transactions on image processing, 26(7):3142–3155, 2017.

[130] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops, pages 0–0, 2018.

[131] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In Proceedings of the IEEE/CVF international conference on computer vision, pages 3155–3164, 2019.

[132] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. EnlightenGAN: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing, 30:2340–2349, 2021.

[133] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep Laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 624–632, 2017.

[134] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.

[135] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM international conference on multimedia, pages 1632–1640, 2019.

[136] Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In European Conference on Computer Vision, pages 111–126. Springer, 2020.

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

[138] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In Proceedings of the IEEE/CVF international conference on computer vision, pages 1406–1415, 2019.

[139] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In European Conference on Computer Vision, pages 41–58. Springer, 2020.

[140] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In Proceedings of the European conference on computer vision (ECCV), pages 254–269, 2018.

[141] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3937–3946, 2019.

[142] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3606–3615, 2020.

[143] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), September 2018.

[144] Sneha Chaudhari, Varun Mithal, Gungor Polatkan, and Rohan Ramanath. An attentive survey of attention models. ACM Transactions on Intelligent Systems and Technology, 12(5), October 2021.

[145] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2):295–307, 2015.

[146] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. DSLR-quality photos on mobile devices with deep convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, pages 3277–3285, 2017.

[147] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In Proceedings of the IEEE international conference on computer vision, pages 764–773, 2017.

[148] Shu Tang, Yang Wu, Hongxing Qin, Xianzhong Xie, Shuli Yang, and Jing Wang. A constrained deformable convolutional network for efficient single image dynamic scene blind deblurring with spatially-variant motion blur kernels estimation. arXiv preprint arXiv:2208.10711, 2022.

[149] Aram Danielyan, Vladimir Katkovnik, and Karen Egiazarian. BM3D frames and variational image deblurring. IEEE Transactions on image processing, 21(4):1715–1728, 2011.

[150] Kostadin Dabov, Alessandro Foi, and Karen Egiazarian. Video denoising by sparse 3D transform-domain collaborative filtering. In 2007 15th European Signal Processing Conference, pages 145–149. IEEE, 2007.

[151] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 17452–17462, 2022.

[152] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. Communications of the ACM, 63(11):139–144, 2020.

[153] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.

[154] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

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

[156] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In International conference on machine learning, pages 214–223. PMLR, 2017.

[157] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. Advances in neural information processing systems, 30, 2017.

[158] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision, pages 694–711. Springer, 2016.

[159] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pages 2223–2232, 2017.

[160] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pages 3408–3416. IEEE, 2019.

[161] Shunta Maeda. Unpaired image super-resolution using pseudo-supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 291–300, 2020.

[162] Wenchao Du, Hu Chen, and Hongyu Yang. Learning invariant representation for unsupervised image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.

[163] Alex Kulesza and Ben Taskar. K-DPPs: Fixed-size determinantal point processes. In Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11, page 1193–1200, Madison, WI, USA, 2011. Omnipress.

[164] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. SinGAN: Learning a generative model from a single natural image. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2019.

[165] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. IEEE Transactions on Knowledge and Data Engineering, pages 1–1, 2021.

[166] Nan Wang, Yabin Zhou, Fenglei Han, Haitao Zhu, and Jingzheng Yao. UWGAN: underwater gan for real-world underwater color restoration and dehazing. arXiv preprint arXiv:1912.10269, 2019.

[167] Xiaoli Yu, Yanyun Qu, and Ming Hong. Underwater-GAN: Underwater image restoration via conditional generative adversarial network. In Pattern Recognition and Information Forensics, pages 66–75, Cham, 2019. Springer International Publishing.

[168] Chaitra Desai, Badduri Sai Sudheer Reddy, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. AquaGAN: Restoration of underwater images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 296–304, 2022.

[169] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

[170] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.

[171] Jiezhang Cao, Yawei Li, Kai Zhang, and Luc Van Gool. Video super-resolution transformer. arXiv preprint arXiv:2106.06847, 2021.

[172] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 10012–10022, 2021.

[173] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (GELUs). arXiv preprint arXiv:1606.08415, 2016.

[174] Xiangyu Chen, Xintao Wang, Jiantao Zhou, and Chao Dong. Activating more pixels in image super-resolution transformer. arXiv preprint arXiv:2205.04437, 2022.

[175] Long Zhao, Zizhao Zhang, Ting Chen, Dimitris Metaxas, and Han Zhang. Improved transformer for high-resolution GANs. Advances in Neural Information Processing Systems, 34:18367–18380, 2021.
[176] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.
[177] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5987–5995, 2017.
[178] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Oct 2017.
[179] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable DETR: Deformable transformers for end-to-end object detection. In International Conference on Learning Representations, 2021.

LUJUN ZHAI received the B.S. degree in Electrical and Automation Engineering from University of Jinan, China, in 2012, the M.S. degree from Prairie View A & M University, in 2017. She is now a PhD student at Electrical and Computer Engineering in Prairie View A & M University. Her research interests include computer vision, image processing, deep learning, and especially focus on face recognition.

YONGHUI WANG (S'01–M'04–SM'14) received the B.S. degree in Optoelectronics from Xidian University, Xi'an, China, in 1993, the M.S. degree in electrical engineering from Beijing University of Technology, Beijing, China, in 1999, and the Ph.D. degree in computer engineering from Mississippi State University, Starkville, MS, in 2003. From 1993 to 1996, he was a Research Engineer with the 41^{st} Electrical Research Institute, Bengbu, China. From July 1999 to December 1999, he worked as an IT Specialist in IBM China, Beijing, China. From 2000 to 2003, he was a research assistant with the Visualization, Analysis, and Imaging Laboratory (VAIL), the GeoResources Institute (GRI), Mississippi State University. He is currently with the Department of Computer Science, Prairie View A&M University, Prairie View, TX. His research interests include image/video signal processing, computer vision, and artificial intelligence.

SUXIA CUI (S'00-M'04–SM'14) received her B.S. and M.S. degrees in Electrical Engineering from Beijing University of Technology, Beijing, China, in 1999. She received her Ph.D. degree in Computer Engineering from Mississippi State University, Starkville, MS, in 2003.

She is currently working as an associate professor in Electrical and Computer Engineering department at Prairie View A&M University, Prairie View, TX. Her research interests include cybersecurity, machine learning, and image processing. Her research is sponsored by US Department of Agriculture, National Science Foundation, Department of Defense, and Department of Education.

YU ZHOU received his 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.