

A survey of deep learning approaches to image restoration

Jingwen Su, Boyan Xu, Hujun Yin *

Department of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, UK



ARTICLE INFO

Article history:

Received 19 March 2021

Revised 14 January 2022

Accepted 19 February 2022

Available online 25 February 2022

Keywords:

Deep learning

Convolutional neural networks

Image restoration

Image deblurring

Image super-resolution

ABSTRACT

In this paper, we present an extensive review on deep learning methods for image restoration tasks. Deep learning techniques, led by convolutional neural networks, have received a great deal of attention in almost all areas of image processing, especially in image classification. However, image restoration is a fundamental and challenging topic and plays significant roles in image processing, understanding and representation. It typically addresses image deblurring, denoising, dehazing and super-resolution. There are substantial differences in the approaches and mechanisms in deep learning methods for image restoration. Discriminative learning based methods are able to deal with issues of learning a restoration mapping function effectively, while optimisation models based methods can further enhance the performance with certain learning constraints. In this paper, we offer a comparative study of deep learning techniques in image denoising, deblurring, dehazing, and super-resolution, and summarise the principles involved in these tasks from various supervised deep network architectures, residual or skip connection and receptive field to unsupervised autoencoder mechanisms. Image quality criteria are also reviewed and their roles in image restoration are assessed. Based on our analysis, we further present an efficient network for deblurring and a couple of multi-objective training functions for super-resolution restoration tasks. The proposed methods are compared extensively with the state-of-the-art methods with both quantitative and qualitative analyses. Finally, we point out potential challenges and directions for future research.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Image restoration has been a long-standing research topic of digital image processing since last century [1–5] and remains an active topic in recent years. Image restoration aims to recover clean latent images from degraded observations, and is a typical inverse problem. Infinite possible mappings between multidimensional degraded observations and restored images determine the ill-posed nature of such inverse problems. For the cases where mappings are known and invertible, corresponding solutions are easy to obtain but such mappings are unique and lack of generality. In practice, the inverse mappings are unknown, thus solution space is infinite and requires regularisation techniques to apply in order to derive feasible and optimal solutions. Therefore most researches in image restoration are devoted to resorting effective analytical models and learning schemes, such that approximations of exact mappings can be found to restore degraded images.

Conventional methods for image restoration employ advanced mathematics and probabilistic models to solve inverse problems,

mostly based on the maximum likelihood or Bayesian approaches in iterative algorithms [6–8]. General formulation of a degraded image Y is assumed to be the result from convolving a clean image X with a blur kernel B and further added with noise N as follows,

$$Y = X * B + N \quad (1)$$

where $*$ denotes convolution operation. By using the maximum-a-posterior estimate, the latent image \hat{X} is formulated as follows,

$$\hat{X} = \operatorname{argmax}_X \log(P(Y|X)) + \log(P(X)) \quad (2)$$

where $P(Y|X)$ denotes the likelihood of degraded observation Y given clean image X while $P(X)$ represents prior distribution of clean image X . It can also become a constrained maximum likelihood estimation,

$$\hat{X} = \operatorname{argmin}_X \|Y - B * X\|^2 + \lambda \Phi(X) \quad (3)$$

where fidelity term $\|Y - B * X\|^2$ approximates the likelihood $P(Y|X)$, while regularisation term $\Phi(X)$ represents priors of latent image X or constraints on the solution. Priors can be adapted depending on various tasks, such as deconvolution [9–11], super-resolution [12–14], inpainting [15–17], in the field of astronomy [18–20],

* Corresponding author.

E-mail address: hujun.yin@manchester.ac.uk (H. Yin).

medicine [21–23], microscopy [24–26], etc. There are also growing interests in multiframe and video restoration that makes use of relationships among successive image frames to reconstruct high quality clean images and videos [27–32].

In the past decade, the rapid rise of deep learning (DL) techniques has greatly impacted various computer vision tasks, from recognition and classification [38–41], to regression and generation [42–45]. Convolutional Neural Networks (CNNs) firstly boosted the performance of classification and detection [46], with numerous network architectures proposed to tackle benchmark research tasks. VGGNet [47] points out that deep network architecture is beneficial, while previous studies mostly focused on shallow network [48]. ResNet [39] provides the baseline structure of image restoration, and becomes the basic structure of several following methods like EDSR [49] (for super-resolution), DeepDeblur [50] (for image deblurring), DnCNN [35] (for image denoising). DenseNet [51] further improves network performance by developing residual links to densely connected convolutional layers.

Deep learning approaches bring many benefits to image restoration such as,

- Learning based methods can boost the performance. On most benchmark datasets, deep learning based methods often outperform the traditional methods significantly.
- Deep learning makes applications more realistic. One can restore a video from degradation by taking sequential frames into consideration or fill in some missing content while it is impossible for the degradation process to be modelled mathematically (e.g., inpainting).
- By using parallel processing units such as graphic processing units (GPUs), deep learning algorithms naturally fit with computer hardware, leading to high efficiency compared to using CPUs.

However, there are still many challenges:

- From the perspective of computational complexity, deep learning based methods have considerable computational costs, making them difficult to deploy in real-time processing. In addition, matrix processing leads to higher requirement for computer hardware, in terms of GPUs and memory, which cannot be satisfied by the embedded systems used commonly in industry, e.g., microcontroller unit (MCU).
- From the perspective of performance, existing algorithms still have a lot of room for improvement.
- From the perspective of training, deep learning CNNs require large datasets, which are not easy to obtain, annotate and may not match well with the actual situations. For instance, many deblurring or super-resolution applications pay more attention on human faces, but most existing training datasets contain relatively few human face samples, while many other samples such as cars or buildings may not help for specific applications.

There are also tasks that are highly related to image restoration, such as 3D reconstruction [52] and image inpainting [53]. The ideas and novel approaches in image restoration can benefit these aforementioned tasks, and vice versa.

This survey is intended as a timely update and overview of deep learning approaches to image restoration and is organised as follows. Section 2 reviews existing deep neural networks for image restoration in general, followed by detailed reviews on models for deblurring, denoising, and super-resolution tasks in particular. Various image quality assessment criteria are also reviewed and discussed. Typical network architectures and learning strategies are reviewed and analysed in Section 3. The most recent models

are briefly considered. A couple of our proposed networks are then presented for deblurring and super-resolution tasks, along with extensive experiments and comparisons with the state-of-the-art models. Final section discusses these networks, performances and results, and remaining challenges and concludes the work. Future work and research directions are also suggested.

2. Deep Networks for Image Restoration

2.1. Image Restoration

There are several ways to apply deep learning in digital image restoration. Learning image priors or kernels by deep learning neural networks [54–56,17] is one of the popular methods. Compared to sophisticated hand-crafted image priors and considerable amount of work conducted to derive such priors, it is more efficient to learn priors through deep learning neural networks, as effective regularisation terms for the ill-posed problems. Learnt priors are integrated in the next phase of optimisation to restore degraded images by variable splitting techniques (e.g. ADMM (Alternating Direction Method of Multipliers) [57] and HQS (Half Quadratic Splitting) [58]), and contribute to achieving more superior performances than those priors based on analytical models. Besides, deep learning approaches employ various architectures [59,60,56,61] and learning strategies [62,63], in order to derive better solutions by exploiting powerful learning capability to extract important information from massive training data. A great amount of research studies have been conducted to apply prevalent DL techniques on solving image restoration tasks. Recently, generative adversarial network (GAN) based approaches have become dominant and surpass general CNN-based methods, elevating the state-of-art performance [64–66]. Burden of dedicating network design for specific applications is alleviated by great compatibility and capacity of the GAN models, at the expense of larger and deeper networks and training issues [67–69]. Moreover these advanced networks have made substantial advances in various applications including underwater imaging [63,70], light-field imaging [60], fluorescence image reconstruction [71], and computerised tomography super resolution [72].

2.2. Image Deblurring

Blurry images are common in practice, and restoration of these degraded images are intractable due to a wide range of factors, such as inevitable motions during long exposure time, physical limitations of imaging devices and imperfect system, unknown degradation process, etc. Researchers have paid many efforts and endeavoured to develop efficient and novel methods to solve these challenging problems.

Dynamic scene blurs are ubiquitous in real life image capturing. Blurs can be caused by a mixture of camera motion, object motion, and scene depth variation. Camera motion has six degrees of freedom in two categories, translational and rotational motions. Translational motion relates to depth variation [73,74], while rotational camera motion and object motion are independent factors that also lead to non-uniform blurs in the image. Since these motion blurs are spatially variant, it is not a trivial task to model imaging and degradation process, especially when there is only a single blurry image that is available. Numerous attempts contribute to building models that approximate real blur kernels by using prior knowledge and additional observations on images.

There are studies that review representative works and compare performances for single image deblurring. A review of conventional methods for image deblurring was provided by Wang et al. [75], which defined blurs occurred in common imaging and

categorised methods to five main frameworks according to respective characteristics. Since learning-based methods were not well developed back then, neural networks were only considered as a promising topic for further research. Lai et al. [76] evaluated and compared 13 single image deblurring algorithms by using their own real-world blurry images and human subject study (Amazon Turk). Recent NTIRE (New Trends in Image Restoration and Enhancement) 2020 challenge on image and video deblurring introduced the state-of-the-art methods and provided fair rankings and comparisons of their performances [77]. A most recent survey provided by Koh et al. [78] reviewed the development of deep learning based non-blind and blind deblurring techniques since 2013. In the paper, a comparative study illustrated the artifacts caused by perceptual loss, the superiority of explicit image priors and the potential of unsupervised learning.

Multi-scale deblurring networks use the "coarse-to-fine" structure to restore an image with several steps. It was firstly proposed in [50], which applied the multi-scale structure developed by Eigen et al. [79]. Tao et al. [80] and Gao et al. [81] developed multiscale deblur networks, and Zhang et al. [82] made a fundamental subsequent change to the structure and mechanism when adopting it, leading to marked differences compared to the other three methods.

DeblurGAN was proposed by Kupyn et al. [33] as the first exploitation of conditional GAN for deblurring problems. This method uses the Residual network block [39] as the main component of the generator. DeblurGAN-v2[64] is an updated version of DeblurGAN, with Feature Pyramid Network (FPN), originally proposed for object detection [83,84], being used as the generator. The authors of [85] presented an end-to-end deblurring network by a unsupervised CNN. Previously, supervised deep-learning networks extensively rely on large sets of paired data, which is highly demanding and challenging to obtain, while unsupervised training scheme can achieve comparable performances with unpaired data. Lu et al. [86] proposed another unsupervised network for domain-specific, single-image deblurring based on disentangled representations.

Authors of [87] proposed a novel network called Dr-Net. They used the Douglas-Rachford iterations to solve deblurring problem because of its more applicable optimisation procedure than the proximal gradient descent algorithm. Purohit et al. [88] reported that restoration of images affected by severe blur necessitated a network design with a large receptive field and presented a new architecture composed of region adaptive dense deformable modules that could implicitly discover the spatially varying shifts responsible for non-uniform blur in the input image and learn to modulate the filters.

Comparisons of various different methods are given in Table 1, Table 2, and Table 3.

2.3. Image Denoising

Image denoising is another important task in image restoration and is of extraordinary value to low-level vision from many aspects. First, it is often an essential pre-processing step to remove noise in various computer vision tasks. Second, image denoising is an ideal test bed for evaluating image prior models and optimisa-

Table 1

Comparison of multiscale based image deblurring methods.

Method	Year	Param. sharing	Loss function	Method description
DeepDeblur[50]	2017	No	adv + MSE	Deep Multi-scale Convolutional Neural Network
SRN-Deblur[80]	2018	Yes	MSE	Multi-scale Network with encoder-decoder structure and LSTM
PSS-NSC[81]	2019	Yes	MSE	Multi-scale Network with DenseBlocks
DSHMN[82]	2019	No	MSE	Stacked Hierarchical Multi-patch Network

Table 2

Comparison of GAN-based image deblurring methods.

Method	Year	Supervision	Loss function
DeblurGAN[33]	2018	Supervised	GAN + perceptual
DeblurGANv2[64]	2019	Supervised	GAN + MSE + perceptual
Unsupervised Deblur[85]	2018	Unsupervised	GAN + reblur + gradient
Disentangled Representation[86]	2019	Unsupervised	GAN + KL + Cycle + perceptual

Table 3

Comparison of image deblurring methods that use special convolution modules from blur mechanism.

Method	Year	Loss function	Method description
SVRNN[89]	2018	MSE	Added RNNs with weights learned by CNNs
Dual Residual Network [90]	2019	SSIM + L1	A novel style of residual connections dubbed with paired operations
Douglas-Rachford [87]	2019	GAN + MSE	Using the Douglas-Rachford iterations to solve deblurring
RADN-Deblur [88]	2019	MSE	Dense deformable module and self-attentive module

tion methods from the Bayesian perspective[94]. Traditionally, BM3D [91] is a main stream method and has enhanced the sparsity by grouping similar 2-D image fragments (e.g., blocks) into 3-D data arrays. Learning-based denoising focuses not only on deep learning, but also other machine learning methods. Such variance is due to the noising mechanism is widely suitable for many signal processing approaches. In this paper, we focus on DL based denoising and their commonalities with other image processing tasks like dehazing and denoising. For an overview in learning based image denoising, see [109]. Mathematically, noised image Y can be expressed as

$$Y = X + N \quad (4)$$

where X denotes true image and N additive noise corrupted with X . Noise can also be multiplicative in nature. Deep CNNs began to apply to image denoising in 2015 [110,111]. The first important work is [112], which first applied a very deep CNN with skip connections. In [93], Gharbi et al. developed a Monte Carlo denoising method with a kernel-splatting architecture.

Based on the type of noise, image denoising can be divided into four categories: additive white noise image (AWNI) denoising, real noisy image denoising, blind denoising, and hybrid image denoising. Among these categories, AWNI attracts most attention. However, the popularity of AWNI does not reflect real noisy images. As a result, although AWNI denoising includes Gaussian, Poisson, salt, pepper and multiplicative noise, there are still gaps with actual application scenarios.

A recent overview is given in [109], where you can find relevant. In this subsection, we aim to focus on comparisons of learning based denoising methods, along with other image restoration tasks. Many ideas and tricks developed for denoising are also use-

ful for other image inverse problems, and vice versa many important denoising networks were inspired by existing work on low-level vision. For example, DnCNN [35] first proposed residual learning in image restoration. The residual learning learning here is different from ResNet proposed by He et al. [39]. It employs a single residual unit to predict the residual image. Generally speaking, DnCNN uses a long residual link to connect the input image to the output directly, so that the networks just need to learn residual image and do not have to focus on the content of images. Residual learning approach has a great impact on image restoration, most deblurring and super-resolution networks use residual links since DnCNN. A comparison of comprehensive image denoising methods is given in Table 4.

2.4. Image Dehazing

The atmospheric scattering model is the classical description for the hazy image generation:

$$Y = X \times t + A(1 - t) \quad (5)$$

where Y is observed hazy image, X is the haze-free scene radiance to be recovered. There are two critical parameters: A denotes the global atmospheric light, and t is the transmission matrix defined as:

$$t = e^{-\beta d} \quad (6)$$

where β is the scattering coefficient of the atmosphere, and d is the distance between the object and the camera.

Caused by turbid medium such as mist and dust, the existence of haze leads to poor visibility of photos and adds data-dependent, complicated and nonlinear noise to the images, making haze removal an ill-posed and highly challenging restoration problem. Many computer vision algorithms can only work well on the scene radiance, which is haze-free. In [113], authors reformulated an image formation model that accounts for surface shading in addition to the transmission function. In [114], He et al. proposed dark channel prior (DCP) to remove haze from single image based on the assumption that image patches of outdoor haze free image often have low-intensity values. [115] is an early learning based approach. Li et al. [116] proposed an all-in-one method to directly generate clean images through a light-weight CNN, based on a reformulated atmospheric scattering model. Qu et al. [117] introduced a GAN for dehazing by using the discriminator to guide the generator to create a pseudo realistic image on a coarse scale,

while the enhancer following the generator was required to produce a realistic dehazing image on the fine scale. In [118], Chen et al. adopted smoothed dilation technique and leveraged on a gated sub-network to fuse the features from different levels. MSRL-DehazeNet [119] relies on multi-scale residual learning and image decomposition. RYF-Net [120] uses a transmission map fusion network to integrate two transmission maps and estimate robust accurate scene transmission map for haze image. [121] contains a supervised learning branch and an unsupervised learning branch. DCP-Loss [122] uses dark channel prior as loss function. Authors of [123] proposed a heterogeneous GAN based method composed of a CycleGAN for producing haze-clear images and a conditional GAN for preserving textural details. Similar work can be seen in Cycle-dehaze [124]. FAMED-Net [125] comprises encoders at three scales and a fusion module to efficiently and directly learn haze-free image. Shao et al. [126] proposed a domain adaptation paradigm, consisting of an image translation module and two image dehazing modules. Authors of [127] adopted a novel fusion-based strategy to derive three inputs from an original hazy image by applying White Balance (WB), Contrast Enhancing (CE), and Gamma Correction (GC). Similar with many image deblurring networks, DCPDN [128] adopts a densely connected structure. A comparison of learning based image dehazing methods is given in Table 5. (See Fig. 1).

2.5. Image Super-resolution

Super-resolution (SR) is a technique to reconstruct high resolution images that effectively overcomes inherent limitations of imaging systems [134]. It has drawn a great deal of attentions due to its practical value in a wide range of applications. At the early stage of development for super-resolution, availability of multiple low resolution (LR) images is considered as a basic premise, restoration and interpolation techniques as prerequisites, which collaboratively contribute to deriving a high resolution (HR) image. When there is only one LR image available, the problem becomes more challenging and is known as single image super-resolution (SISR). Different from other restoration tasks, additional process of upsampling is required in SR after deblurring to increase image dimension and obtain a HR image. On the basis of Eq. 1, degradation applies down-sampling operator D after blurring, given in Eq. 7. Observation model is depicted in Fig. 2.

Table 4
Comparison of image denoising methods.

Method	Year	Approach	Method description
BM3D [91]	2007	Traditional method	Using sparse 3-D transform-domain collaborative filter
WNNM [92]	2009	Traditional method	Based on Markov random field
DnCNN [35]	2017	CNN	Using residual learning in deep convolutional network
Monte Carlo [93]	2019	CNN	CNN that can learn to denoise Monte Carlo renderings directly from samples
FFDNet [94]	2018	CNN	CNN with a tunable noise level map as input
TNRD [95]	2016	Learning method	Using a dynamic nonlinear reaction diffusion model with time-dependent parameters
NLRN [96]	2018	CNN	A non-local recurrent network
MWCNN [97]	2018	CNN	A multi-level Wavelet-CNN
GCDN [98]	2020	Graph network	Trainable neural network employing layers based on graph convolution operations
RIDNet [99]	2019	CNN	Using residual on the residual structure to ease the flow of low-frequency information
CBDNet [100]	2019	CNN	Training network with realistic noise and real-world noisy-clean image pairs
VDN [101]	2019	CNN	Integrating noise estimation and image denoising into a Bayesian framework
SADNet [102]	2020	CNN	A spatial-adaptive denoising network
CycleISP [103]	2020	CNN	Modelling camera imaging pipeline in forward and reverse directions
MPRNet [104]	2021	CNN	A multi-stage architecture with per-pixel adaptive design
MIRNet [105]	2020	CNN	Maintaining spatially-precise high-resolution representations
Uformer-B [106]	2021	Transformer	A hierarchical encoder-decoder network using the Transformer block
HINet [107]	2021	CNN	Exploring role of instance normalization in low-level vision tasks
Restormer [108]	2021	Transformer	An efficient Transformer model with locally-enhanced window and learnable modulator

Table 5

Comparison of deep learning based image dehazing methods.

Method	Year	Approach	Method description
IHF [115]	2014	Random Forest	Using random forest to investigate different features
DehazeNet [129]	2016	CNN	CNN with layers specially designed to embody the established assumptions/priors
AOD-Net [116]	2017	CNN	A light-weight CNN based on a reformulated atmospheric scattering model
Cycle-dehaze [124]	2018	GAN	Training network by feeding clean and hazy images in an unpaired manner
GridDehazeNet [130]	2019	CNN	Attention-based multi-scale network for image dehazing
EPDN [117]	2019	GAN	Enhanced Pix2pix dehazing network
MSRL-DehazeNet [119]	2019	CNN	Multi-scale residual learning to restore the image base component
YRF-Net [120]	2019	CNN	Using CNN to estimate the accurate scene transmission map
SSID [121]	2019	Unsupervised learning	A supervised learning branch and an unsupervised learning branch
Deep DCP [122]	2019	Unsupervised learning	Using dark channel prior as loss function
FFA-Net [37]	2020	CNN	An end-to-end feature fusion attention network
MSBDN [131]	2020	CNN	A multi-Scale boosted dehazing network with dense feature fusion based on U-Net architecture
KDDN [132]	2020	CNN	Distill image dehazing with the heterogeneous task imitation
DAID [126]	2020	GAN	A domain adaptation paradigm with an image translation module
GCANet [118]	2019	CNN	Smoothed dilation technique and a gated sub-network
AECR-Net [133]	2021	CNN	An adaptive mixup operation and a dynamic feature enhancement module
FAMED-Net [125]	2019	CNN	Comprising encoders at three scales and a fusion module

$$Y = D(X * B) \downarrow_S + N \quad (7)$$

Traditional methods decompose super-resolution techniques according to the observation model, and implement upsampling by interpolation and deblurring by iterative optimisation techniques. Both require sophisticated image priors and necessary assumptions on scaling factors and blur kernels. Yang et al. [135] conducted a comparison work of ten representative SR algorithms [136,12,137–139,13,140–142], which adopted sparse representation and various image priors in iterative optimisation frameworks. Most of these methods are time-consuming and inefficient to derive a HR image based on exploiting expert knowledge. While on the other hand, DL based methods are prosperously developed. Deep neural networks learn solutions from large training data, rather than sophisticated probability models and statistical priors employed in conventional methods. Besides, advanced computational hardware and learning algorithms also contribute to the rapid development. Authors of [143] reviewed over a hundred deep learning based SR algorithms, and selected 12 DL-based models [144–150,49,151,152,36] as the most popular ones. Frequently used architectures and optimisation objectives were thoroughly discussed in the paper.

2.6. Image Quality Assessments

Image quality assessment (IQA) is critical to determine image quality, image processing algorithms and imaging systems. Only when unified quality measures are provided, fair comparisons can be conducted to reflect characteristics and properties of algorithms and systems with convincing and reliable evidences. Initially, image quality measures (IQM) were applied mostly for evaluating image compression and acquisition techniques, then generalised to other image processing tasks and image communication networks [153]. Since the ultimate receiver of images is human, the most reliable evaluation of image visual quality is through subjective human studies that collect ratings of a large number of test examples. But it is time-consuming and often too expensive to conduct such studies to provide quality evaluation for every case in practice. Therefore an objective IQA that aims to predict perceptual quality efficiently while correlating to human visual system (HVS) response is highly desirable.

The most common categorisation of objective quality measures is by the availability of the reference image, i.e. the full-reference (FR), reduced-reference (RR), and no-reference (NR) quality measures. FR measures compute similarities between distorted and

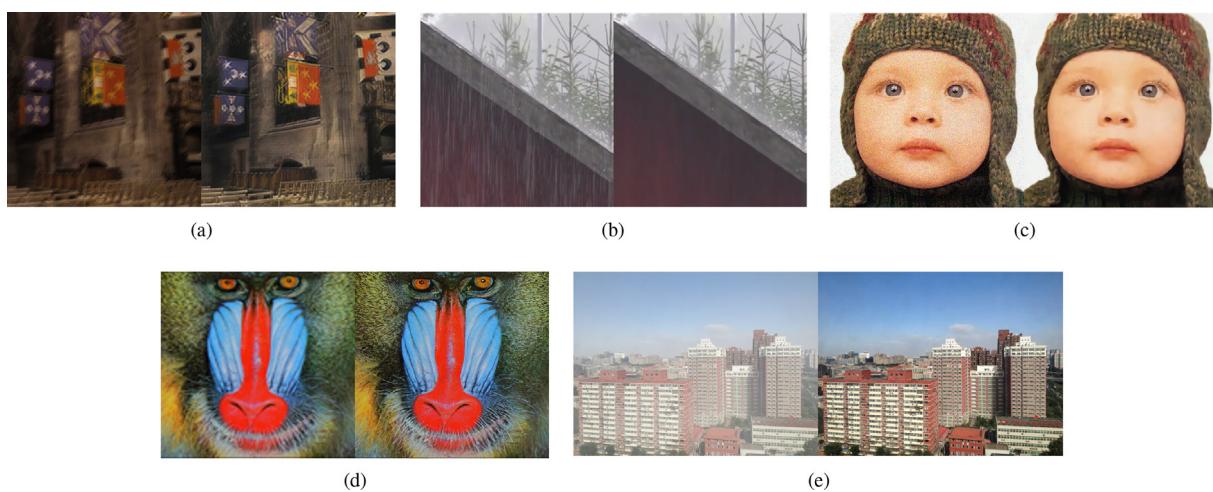


Fig. 1. Examples of image restoration tasks: (a) deblurring [33], (b) deraining [34], (c) denoising [35], (d) super-resolution [36], and (e) dehazing [37].

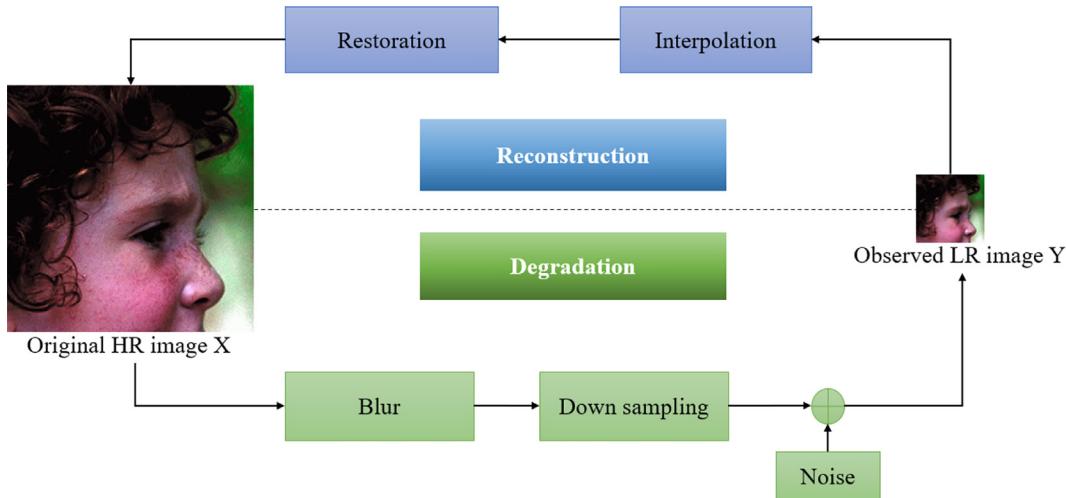


Fig. 2. Image super-resolution degradation and reconstruction model.

reference images, RR measures are applied when partial information of reference image is available, while NR measures utilise image statistics to evaluate image qualities since information of reference image is completely unavailable.

The simplest objective FR measure is the peak signal-to-noise ratio (PSNR), based on mean-squared error (MSE) between reference and degraded images. Although widely adopted, it is known that image fidelity measures like PSNR is incapable to relate well to visual quality [154,155]. Structural similarity index measure (SSIM) was introduced by Wang et al. [156] to further approximate quality evaluation of HVS, exploiting its sensitivity in variations of structural information. There are variants of SSIM, such as multi-scale SSIM [157], three-component SSIM [158], and four-component SSIM [159], proposed to further develop for generalisation. Besides, information theory can be introduced to derive image quality assessment, for example information fidelity criterion (IFC) proposed by Sheikh et al. in [160], followed by extension work of visual information fidelity measure (VIF) [161]. Furthermore, measures like feature similarity index measure (FSIM) [162], DCTune [163], wavelet-based distortion measures [164], Haar wavelet-based perceptual similarity index (HaarPSI) [165], utilise image features from other domains that approximate responses of HVS. Numerous studies provide valuable reviews on FR IQA [166–168, 160,169–171]. RR IQA measures are applied when partial information of reference image or degradation process is present, and can be considered as an intermediate case, inspired from both FR and NR IQA measures [172–175]. Representative FR and RR methods with equations are presented in Table 6.

When pristine reference image is unavailable for quality evaluation, NR IQA measures are useful. A common feature adopted in most NR IQA measures is the natural scene statistics (NSS) [177,178], which have invariant properties to various degradations and image contents, such measures include the blind/referenceless image spatial quality evaluator (BRISQUE) [179], the distortion identification-based image verity and integrity evaluator (DIIVINE) [180], the natural image quality evaluator (NIQE) [181]. Fig. 3 provides pipelines of representative NR methods: BRISQUE, BLIINDS-II [182], DIIVINE, and NIQE. There are other features also employed for NR IQA, e.g. NSS in DCT domain [183,182], NSS in multivariate Gaussian model [184], gradient magnitude [185,186], etc. Perception index (PI) proposed in [187] combines two NR methods ([181,188]) for perceptual evaluation of generated images.

In the development of IQA measures, the conflicts between distortion-based measures and perceptual quality measures are

reported in many research studies. As a consequence, the tradeoff between perception and distortion is systematically stated in [189]. Relevant studies have been conducted to analyse this trade-off [190,191], and discussions have led to a conclusion that image quality can be improved in terms of either fidelity or perceptual quality, at the expense of each other.

Recently, deep learning is developed as an alternative paradigm for IQA that learns to map images to numerical scores from training set of distorted images with human subjective quality assessment (i.e. mean opinion score (MOS) or differential mean opinion score (DMOS)) [192–200]. End-to-end training enables deep neural networks to achieve better prediction accuracy than previous hand-crafted methods. But performance optimisation and generalisation capability is subject to the limited training set, considering the difficulty of collecting ground truth MOS/DMOS values. Model complexity and fine-tuning network hyperparameters is also a non-trivial and critical task for generalising deep learning based methods.

3. Network Architectures and Learning Strategies

3.1. Baseline Models

Multilayer perceptron (MLP) [201] was one of the earliest artificial neural networks adopted for image restoration [202–206]. As restoration requires the same dimension of input and output images, MLP framework (Fig. 4a) follows the structure of fully connected networks and is enabled to learn high dimensional mappings between degraded input images and sharp latent images. However, MLP is inefficient due to redundancy in a huge amount of parameters that burdens computational resources and storage. Besides, spatial information across channel and content of multidimensional images is overlooked in MLP, which is another obstacle for further development.

Considering the structural property of images and deficiency of MLP, convolutional neural networks (Fig. 4b) are adopted (e.g. [207]) and provide more suitable solutions for image restoration. Convolutional neural networks (CNNs) have advantages of shared weights, architecture sparsity, training stability and hierarchical feature extraction, hence extraordinary performances are achieved and become the new state-of-the-art approaches. Although it is observed that increasing network depth benefits the model performance of CNNs by the means of large receptive fields and meaningful hierarchical features, training stability and computational

Table 6
Representative FR and RR IQA methods.

Method	Category	Feature	Formula	Annotation
Mean Squared Error (MSE)	FR	Euclidean distance	$MSE(X, \hat{X}) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (X(i,j) - \hat{X}(i,j))^2$	
Peak Signal-to-Noise Ratio (PSNR)	FR	Euclidean distance	$PSNR(X, \hat{X}) = 10 \log(MAX^2 / MSE(X, \hat{X}))$	M, N denote image dimension, X reference image, \hat{X} distorted or estimated images. MAX is the maximum signal value.
Similarity Structural Index (SSIM) [156]	FR	Structural information variation	$SSIM(x, \hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}$	μ is the mean, σ is the standard deviation, c_1 and c_2 are constants to ensure stability.
Visual Information Fidelity (VIF) [161]	FR	Image information extracted by HVS	$VIF = \sum_p \frac{I(\bar{C}^{p,p}, \bar{X}^{p,p} S^{p,p} = s^{p,p})}{I(\bar{C}^{p,p}, \bar{X}^{p,p} S^{p,p} = s^{p,p})}$	$p \in \text{subbands}$, $\bar{C}^{p,p}$, $\bar{X}^{p,p}$ and $S^{p,p}$ represent vectors of P blocks in subband p of wavelet decomposition of natural image signal, distorted output of HVS and reference output of HVS. $s^{p,p}$ denotes the maximum likelihood estimate of $S^{p,p}$, which is a positive scalar. At each location x , similarity $S_L(x)$ equals to the product of similarity of phase congruency and gradient magnitude features. $PC_m(x) = \max(PC_1(x), PC_2(x))$
Feature Similarity Index (FSIM) [162]	FR	Phase congruency and gradient magnitude	$FSIM = \frac{\sum_{x \in \Omega} S_L(x) \cdot PC_m(x)}{\sum_{x \in \Omega} PC_m(x)}$	f_x^p and $f_{\hat{x}}^p$ denote probability distribution function in subband p of reference and distorted image respectively, \hat{d}^p is estimation of Kullback–Leibler divergence between f_x^p and $f_{\hat{x}}^p$, D_0 is a constant that controls scale of distortion measure.
Wavelet coefficients based [174]	RR	Natural scene statistics in wavelet transform domain	$D = \log_2 \left(1 + \frac{1}{D_0} \sum_p \hat{d}^p(f_x^p f_{\hat{x}}^p) \right)$	$g_{x,p}^r$ and $g_{x,p}^d$ denote scaled entropies of reference and distorted image in subband p respectively, L_p is the size of subband p , Λ_p the number of subsampled blocks, λ index the block.
Reduced reference entropy difference index (RRED) [175]	RR	Wavelet coefficient entropy difference	$RRED_p^{\Lambda_p} = \frac{1}{L_p} \sum_{\lambda=1}^{\Lambda_p} g_{x,p}^r - g_{x,p}^d $	
DNT domain based [176]	RR	Divisive normalisation transformation (DNT) for image representation	$D_{band} = \alpha \log(d(f_x f_{\hat{x}})) + \beta \log(d_s) + \gamma \log(d_k) + \theta \log(d_s)$	$\alpha, \beta, \gamma, \theta$ are weighting parameters, d_s, d_k, d_s representative absolute difference of standard deviation, kurtosis, skewness of the DNT coefficients computed from original and distorted image respectively.

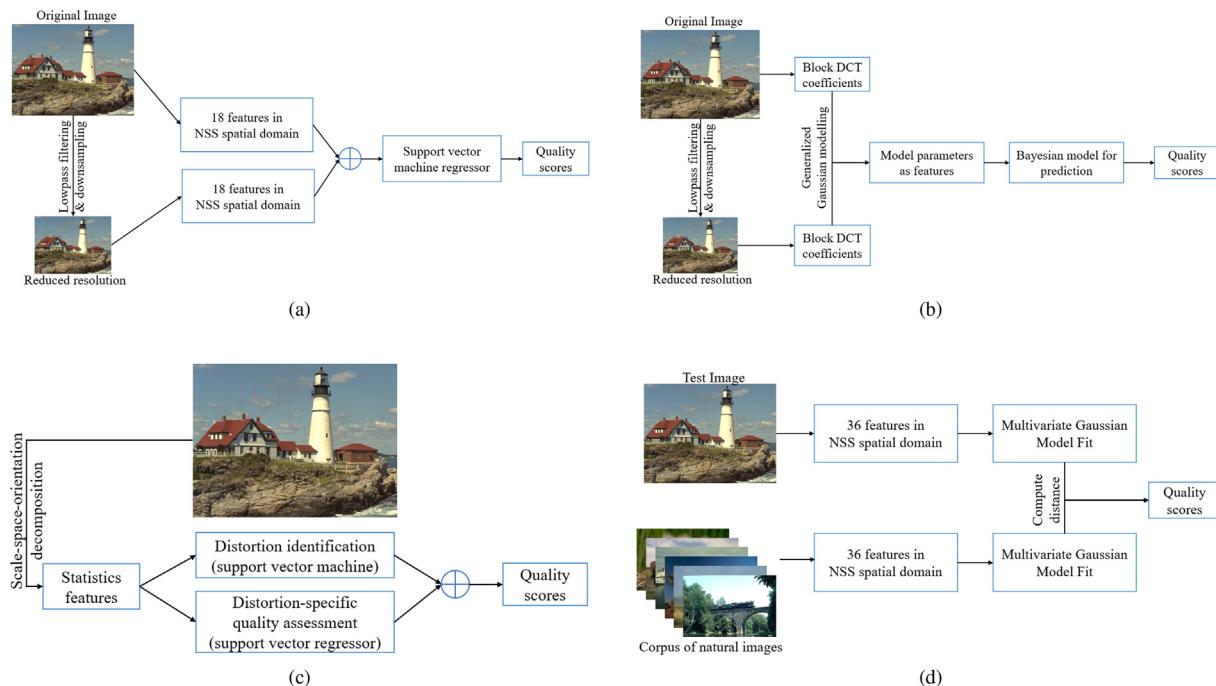


Fig. 3. Representative NR IQA method pipelines, (a) BRISQUE [179], (b) BLIINDS-II [182], (c) DIIVINE [180], and (d) NIQE [181].

resource become tricky issues. Therefore many advanced techniques to deal with these issues. Residual learning and skip connections were invented to stabilise training [39]. Residual block

(Fig. 4c) is effective for improving performance and becomes new building block of many deep residual networks (Fig. 4d). Other network paradigms include encoder-decoder, autocoder, and varia-

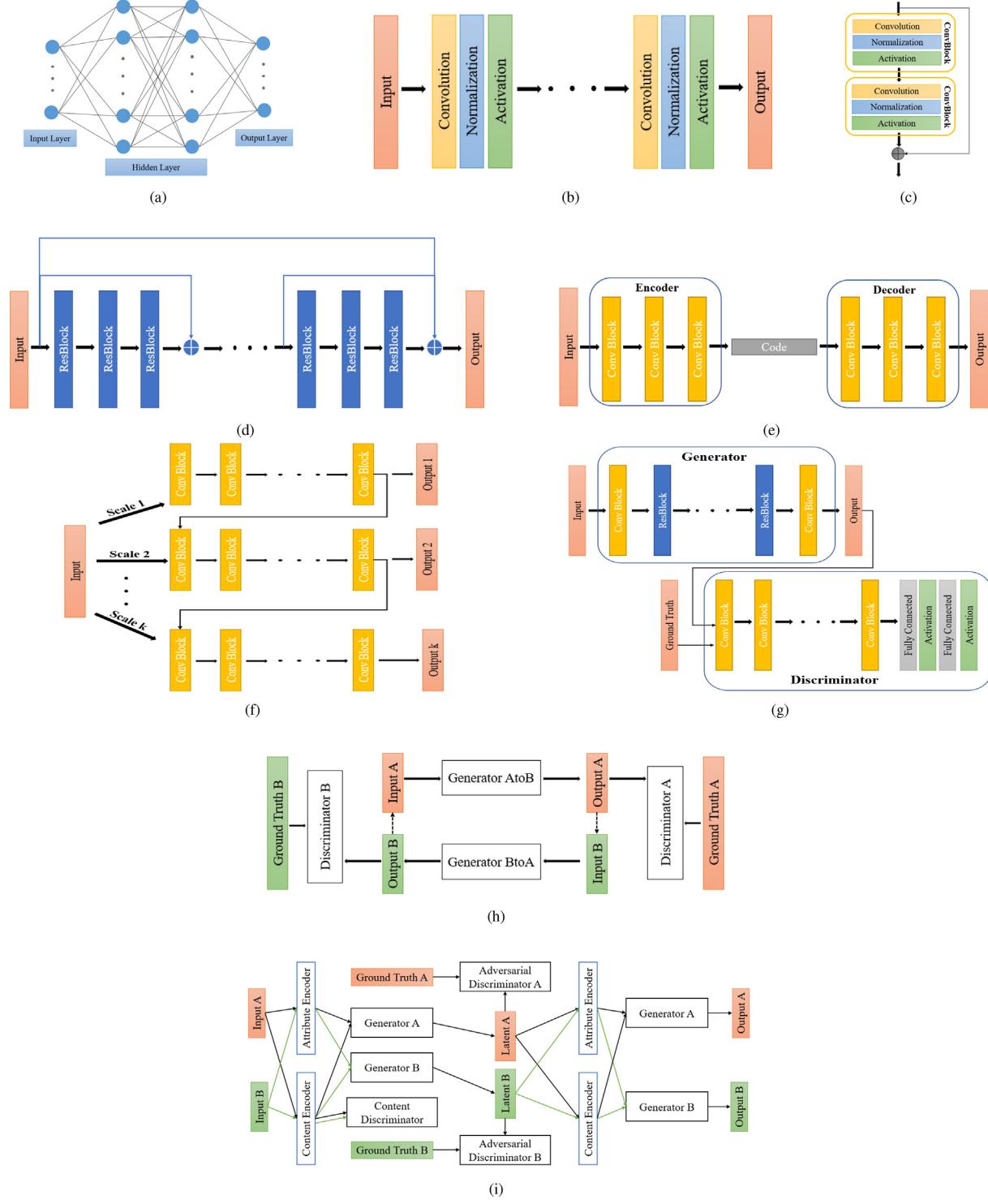


Fig. 4. Representative DNN architectures for image restoration, (a) Multilayer perceptron, (b) Convolutional neural network, (c) ResBlock [39], (d) ResNet [39], (e) auto-encoder [211], (f) Multi-scale CNN [50], (g) Generative adversarial network [208], (h) CycleGAN [209], and (i) DRIT [210].

tional auto-encoder under unsupervised learning scheme, which aims to learn high-level sparse representations from training data (Fig. 4e). Multi-scale networks (Fig. 4f) are specialised in dealing with degradation across various scales. Generative adversarial networks (GANs), as shown in Fig. 4g) [208], combine advantages of generative modelling and adversarial learning to produce plausible textures in generated images. Since paired training images

required in GAN-based models are difficult to obtain, unpaired training has been proposed, e.g. cycleGAN [209], which devised cycle consistency loss as a regularisation technique to generate high quality images (Fig. 4h). To prevent mode collapse issues, disentanglement representation was adopted in [210] (Fig. 4i) i) to provide a novel alternative to generate diverse output images without aligned training data.

3.2. Learning Strategies

3.2.1. Supervised, semi-supervised and unsupervised learning

It is common and straightforward to adopt supervised learning for training neural networks, as long as labelled data are available. Minimisation of cost function and back propagation through network layers enable powerful capability of learning under effective supervision. It encourages networks to converge towards target distribution and generate desired outputs. Typical applications are classification and regression where the purpose is for prediction or inference. However for training deep neural networks, supervised learning is prone to overfitting and/or poor generalisation due to complex underlying mapping function and limited training data. To mitigate such problems, techniques such as early stopping [212], dropout [213], and weight sharing [214] are used to regularise model complexity, and these techniques become necessary in designing and training deep neural networks nowadays. Moreover, it is time consuming to collect matched image pairs for training deep networks in image restoration.

Unsupervised learning can discover underlying structures and patterns in data and hence provides potential insights of the mapping between input and output. Therefore it is useful to learn representative features first, and learnt features can then be used in other tasks under supervised learning or in generative models [215–217,210,218,209]. Reconstruction loss between original input and reconstructed output is important for unsupervised learning to harness representational power of deep networks. Through dimension reduction and reconstruction, autoencoder employs an encoder-decoder structure to learn sparse representation of images [112,219]. For domain transfer tasks like image-to-image translation, unsupervised mechanism is indispensable [220–222]. In practice, the amount of labelled or paired training data is always scarce. In order to exploit large amount of unlabelled data and small amount of labelled data, semi-supervised learning [223] utilises intrinsic advantages of supervised and unsupervised learning. Under supervision, deep networks is empowered to generate desired outputs from training data, but performance is also limited. Unlabelled data are inexpensive and easy to obtain, unsupervised and semi-supervised learning use them to improve network performance in terms of accuracy and generalisation capability. And it is shown that unsupervised learning is able to outperform supervised learning for certain classes of problems under prespecified assumptions [224–230]. Authors of [121] adopted semi-supervised learning in a CNN that contains a supervised branch and an unsupervised branch for single image dehazing. Authors in [231] used semi-supervised learning to train a deep CNN for single image rain removal and achieved superior performance than the state-of-art methods.

3.2.2. Autoencoder and adversarial networks

Auto-encoder: Autoencoder is a type of neural networks used to learn an efficient data coding or representation in an unsupervised or self-supervised manner. The aim of an autoencoder is to learn a representation of a set of data in a reduced dimensional space. Along with dimension reduction, reconstruction part of autoencoder tries to generate from the reduced encoding a representation as close as possible to its original input. Autoencoder exists many variants, aiming to force the learned representations to assumed useful properties. For instance, regularised autoencoders (sparse, denoising and contractive), which are effective in learning representations for subsequent classification tasks. Autoencoders and variational auto-encoders can be used as an integral part of generative models. Autoencoder is wildly used in image denoising [232,233] and super-resolution [234–236]. Deblurring network like [237] is also related to autoencoder. Specifically, the authors used GAN to generate a blur image as a representation of the given

input sharp image, and the reconstruction part of the autoencoder was used as deblurring network.

Adversarial networks: The ingenious idea of generative adversarial networks, introduced by Goodfellow et al. [208], is to define a game between two competing networks: a discriminator and a generator. The generator receives information from an input and generates a sample. The discriminator learns from real and generated samples and tries to distinguish between them. The goal of the generator is to fool the discriminator by generating perceptually convincing samples that cannot be distinguished from the real samples. The game between the generator G and discriminator D has the following minimax objective:

$$\min_{G} \max_{D} \mathbb{E}_{a \sim \mathbb{P}_r} [\log D(a)] + \mathbb{E}_{\tilde{a} \sim \mathbb{P}_g} [\log (1 - D(\tilde{a}))]. \quad (8)$$

where \mathbb{P}_r is the data or real sample distribution and \mathbb{P}_g is the generator model distribution. GANs are known for its ability to generate samples of good perceptual quality in vision tasks; however, training of the vanilla version of GAN often suffers from many problems such as mode collapse and vanishing gradients, as described in [67]. Minimising the value function in GAN is equal to minimising the Jensen-Shannon (JS) divergence between the data and model distributions on a . Arjovsky et al. [238] discussed the difficulties in GAN training caused by JS-divergence approximation and proposed to use the Earth-Mover (also called Wasserstein-1) distance $W(q, p)$. The value function for Wasserstein-GAN is constructed using the Kantorovich-Rubinstein duality [239]:

$$\min_{G} \max_{D \in \mathcal{D}} \mathbb{E}_{a \sim \mathbb{P}_r} [D(a)] + \mathbb{E}_{\tilde{a} \sim \mathbb{P}_g} [D(\tilde{a})]. \quad (9)$$

where \mathcal{D} is the set of 1-Lipschitz functions and \mathbb{P}_g is once again the model distribution.

The idea here is that critic value approximates $K \times W(P_r, P_g)$, where K is a Lipschitz constant and $W(P_r, P_g)$ is a Wasserstein distance. In this setting, the discriminator network is called critic and it approximates the distance between the samples. To enforce Lipschitz constraint in WGAN, Arjovsky et al. [238] added weight clipping to $[-c, c]$. Gulrajani et al. [240] proposed to add a gradient penalty term to the value function as an alternative way to enforce the Lipschitz constraint instead:

$$\lambda \mathbb{E}_{\tilde{a} \sim \mathbb{P}_g} \left[\left(\|\nabla_{\tilde{a}} D(\tilde{a})\|_2 - 1 \right)^2 \right]. \quad (10)$$

This approach is robust to the choice of generator architecture and requires almost no hyper-parameter tuning. This is crucial for image deblurring as it allows the use of lightweight architectures in contrast to standard Deep ResNet architectures [39], previously used for image deblurring [50]. GAN based methods are also popular in denoising [241–245] and super-resolution [36,246–250].

3.3. State-of-the-Art models

Learning based single image restoration is still an active topic. Apart from motion deblur, defocus deblur attracted increasing attention. For instance, [251] exploited data available on dual-pixel (DP) sensors found on most modern cameras, and [252] proposed to perform an explicit deconvolution process in a feature space by integrating a classical Wiener deconvolution framework with learned deep features.

An important direction for further development in image restoration is combining image processing and deep learning/machine learning methods. For image super-resolution, researchers start to focus on different scenes. For example, using internal data can train a super-resolution network just as to use a single image, and this is called zero-shot super-resolution (ZSSR) [253], and MZSR [254] further extended it by adding a meta-training stage to accelerate training speed. Graph Neural Networks also

began to make way in super-resolution [255]. For image denoising, [256] introduced a self-supervised denoising framework called Nose2Siame, where a new self-supervised loss was proposed by deriving a self-supervised upper bound of the typical supervised loss. Nose2Siame requires neither J-invariance (which may lead to worse denoising models) nor extra information about the noise model and thus can be used in a wider range of applications.

4. Proposed Networks

4.1. Super-resolution

As multiple loss components are involved in the training objective function, simultaneous minimisation for various losses is required. Linear combination is the most straightforward means, but the combined loss after weighting and addition may not be convex and difficult to derive the optimum solution through gradient descent. We assume the multi-dimensional loss space naturally formed by multiple loss components is Euclidean, and each individual loss represents a dimension in the space that is independent to each other. We propose that the training objective to be minimised can be defined as the Euclidean distance between the loss location and the origin, or the hypervolume bounded by loss and loss bound (Fig. 5). Thus, the complicated multi-objective optimisation problem is transformed to single objective optimisation. Mathematical formulations of proposed schemes are described in Table 7.

A common characteristic shared by both of the proposed *Ed* and *Hypervol* formulation proposed in [257] is the gradient weighting learnt during training progress, and automatic importance assignment to each individual loss. Compared to manually fine-tuning weighting parameters adopted by most existing methods, the proposed *Ed* formulation of training objective function provides an alternative to optimise model performance for a given model structure.

These two methods differ in the gradient weighting factor. Euclidean distance based scheme utilises the projection of individual loss on Euclidean distance between the origin and the loss location. While gradient weighting factor of *Hypervol* formulation computes the inverse of distance between loss and corresponding loss bound. As we can see from Table 7, the equation of *Ed* formulation is more concise and has no additional hyperparameter to be predefined while *Hypervol* formulation requires loss bound μ_k to be determined before implementation.

We applied proposed methods on the baseline model SRGAN [36], and adopted the same implementation details as in the

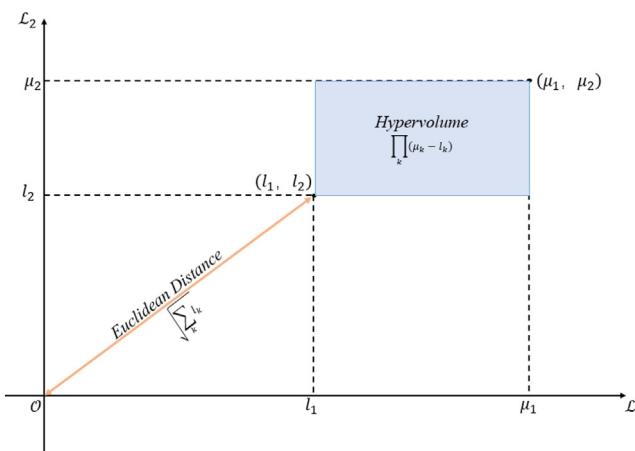


Fig. 5. Illustration of 2-dimensional loss space and proposed training objectives.

SRGAN paper. On the basis of adversarial loss \mathcal{L}_{adv} and perceptual loss \mathcal{L}_X given in the original paper [36], we also include MSE loss \mathcal{L}_{MSE} and SSIM loss \mathcal{L}_{SSIM} as additional constraints to form a multi-dimensional loss space. Loss functions are given as follows,

$$\mathcal{L}_{adv} = -\mathbf{E}[\log(D(G(I_{LR})))] \quad (11)$$

$$\mathcal{L}_X = \frac{1}{W_{ij}H_{ij}} \sum_{x=1}^{W_{ij}} \sum_{y=1}^{H_{ij}} \|\phi_{ij}(I_{HR})_{x,y} - \phi_{ij}(G(I_{LR}))_{x,y}\|_1 \quad (12)$$

$$\mathcal{L}_{MSE} = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \|G(I_{LR}) - I_{HR}\|_1 \quad (13)$$

$$\mathcal{L}_{SSIM} = 1 - SSIM(G(I_{LR}), I_{HR}) \quad (14)$$

where I_{LR} denotes input low-resolution image, I_{HR} denotes high resolution image, ϕ_{ij} is the feature map obtained after activation of the j -th convolution before i -th maxpooling layer in the VGG19 network. W_{ij} and H_{ij} are the dimensions of the feature map.

For original formulation of training objective function, on the basis of the equation given in [36], $\mathcal{L} = 10^{-3} \mathcal{L}_{adv} + \mathcal{L}_X$, MSE loss \mathcal{L}_{MSE} or SSIM loss \mathcal{L}_{SSIM} weighted by 10^{-2} is added to constitute multi-objective training function as follows,

$$\mathcal{L} = 10^{-3} \mathcal{L}_{adv} + \mathcal{L}_X + 10^{-2} \mathcal{L}_* \quad (15)$$

where \mathcal{L}_* represents \mathcal{L}_{MSE} or \mathcal{L}_{SSIM} .

We tested our methods on four SR datasets, Set5 [258], BSDS100 [259], DIV2K validation set [260] and RealSR [261]. Quantitative results are given in Table 8. For visual comparison, example test images and patches are shown in Fig. 6. We used four quality evaluation measures: distortion-based PSNR, SSIM, and perception-based VIF, PI. For PSNR, SSIM and VIF metrics, the higher the measure, the better the quality of the test image, while the lower PI (Perception Index) the better the quality. Descriptions and calculations of PSNR, SSIM and VIF metrics can be found in Table 6, PI is computed using Ma's score [188] and NIQE [181] as follows,

$$PI = \frac{1}{2}((10 - Ma) + NIQE) \quad (16)$$

From Table 8, we can find that the proposed *Ed* formulation of training objective function is effective in boosting model performances. Compared to the original formulation defined with fixed loss weights, the proposed method and *Hypervol* formulation make good use of adaptive weights and achieve the better performances. Moreover, it is shown that involvement of other losses as additional constraints for regularisation is helpful and necessary. Among them, SSIM loss is the most meaningful loss for training GAN models to generate high quality images. As reflected in quality evaluation of images generated by models trained using different training objective functions, models using training objective function $f(\mathcal{L}_{adv}, \mathcal{L}_X, \mathcal{L}_{SSIM})$ generally produce images that has higher scores than those using training objective function $f(\mathcal{L}_{adv}, \mathcal{L}_X, \mathcal{L}_{MSE})$. From experiments with training objective function $f(\mathcal{L}_{adv}, \mathcal{L}_X, \mathcal{L}_{MSE}, \mathcal{L}_{SSIM})$, we can find that compared with linear combination of multiple loss components, the proposed *Ed* formulation provides effective alternatives to enhance quality of generated images.

For visual evaluation, Fig. 6 contains example results with extracted patches for texture detail observation (finer details can be found when zoom in), as overall visual differences are too trivial to distinguish. We can observe that *Ed* formulation introduces blurring artifacts while *Hypervol* formulation is capable of recovering finer details. Moreover, although SSIM loss is useful to improve image quality quantitatively, visual image quality is not benefited

Table 7

Proposed training objective functions.

	Formulation	Gradient	Annotation
Euclidean distance (abbr. Ed)	$\mathcal{L}_E = \ \mathcal{L}\ _2 = \sqrt{\sum_k \mathcal{L}_k^2}$	$\frac{\partial \mathcal{L}_E}{\partial \theta} = \sum_k \frac{\mathcal{L}_k}{\sqrt{\sum_k \mathcal{L}_k^2}} \frac{\partial \mathcal{L}_k}{\partial \theta} = \sum_k w_k^E \Delta \mathcal{L}_k$	\mathcal{L}_E is L2 normalisation of all loss components, gradient weighting factor $w_k^E = \frac{\mathcal{L}_k}{\sqrt{\sum_k \mathcal{L}_k^2}}$
Hypervolume [257] (abbr. Hypervol)	$\mathcal{L}_H = \text{Hypervol}(\mathcal{L}_k) = -\sum_k \log(\mu_k - \mathcal{L}_k)$	$\frac{\partial \mathcal{L}_H}{\partial \theta} = \sum_k \frac{1}{\mu_k - \mathcal{L}_k} \frac{\partial \mathcal{L}_k}{\partial \theta} = \sum_k w_k^H \Delta \mathcal{L}_k$	μ_k denotes adaptive loss bound as defined in [257], gradient weighting factor $w_k^H = \frac{1}{\mu_k - \mathcal{L}_k}$

Table 8

Quantitative results on four datasets: Set5, BSDS100, DIV2K and RealSR. Bold values indicate the best performances while underlined values indicate the second best performances.

	$f(\mathcal{L}_{adv}, \mathcal{L}_X)$			$f(\mathcal{L}_{adv}, \mathcal{L}_X, \mathcal{L}_{MSE})$			$f(\mathcal{L}_{adv}, \mathcal{L}_X, \mathcal{L}_{SSIM})$			$f(\mathcal{L}_{adv}, \mathcal{L}_X, \mathcal{L}_{MSE}, \mathcal{L}_{SSIM})$		
	Original	Ed	Hypervol	Original	Ed	Hypervol	Original	Ed	Hypervol	Original	Ed	Hypervol
Set5												
PSNR(dB)	27.68	28.20	28.09	28.08	27.99	29.40	28.80	28.89	29.03	28.00	<u>29.03</u>	28.54
SSIM	0.8054	0.8113	0.8121	0.8091	0.8123	<u>0.8396</u>	0.8289	0.8383	0.8366	0.8085	0.8457	0.8296
VIF	0.4447	0.4525	0.4600	0.4547	0.4585	0.5088	0.4827	0.4956	0.4911	0.4548	<u>0.5076</u>	0.4871
PI	3.416	4.231	3.543	<u>3.476</u>	4.745	4.447	4.163	4.493	4.294	3.536	4.536	3.517
BSDS100												
PSNR(dB)	24.00	24.27	24.37	24.32	24.43	25.42	24.92	25.25	25.08	24.23	<u>25.27</u>	24.77
SSIM	0.6192	0.6290	0.6316	0.6283	0.6432	0.6693	0.6590	<u>0.6742</u>	0.6721	0.6270	0.6791	0.6553
VIF	0.2883	0.3001	0.3067	0.3042	0.3069	<u>0.3592</u>	0.3344	0.3537	0.3452	0.3018	0.3597	0.3346
PI	2.382	2.450	<u>2.366</u>	2.408	2.437	2.947	2.585	3.024	2.853	2.345	2.840	2.465
DIV2K												
PSNR(dB)	26.40	26.65	26.78	26.78	26.86	28.07	27.48	27.86	27.68	26.70	<u>27.91</u>	27.35
SSIM	0.7422	0.7429	0.7502	0.7490	0.7562	0.7854	0.7449	<u>0.7889</u>	0.7871	0.7477	0.7938	0.7751
VIF	0.3502	0.3547	0.3677	0.3650	0.3641	<u>0.4235</u>	0.3979	0.4194	0.4119	0.3637	0.4302	0.4041
PI	3.228	3.333	<u>3.188</u>	3.223	4.044	3.641	3.409	3.712	3.574	3.181	3.662	3.317
RealSR												
PSNR(dB)	36.84	37.27	37.11	37.30	37.86	38.28	38.32	<u>38.52</u>	38.02	37.38	38.71	38.21
SSIM	0.9397	0.9411	0.9418	0.9418	0.9497	0.9499	0.9528	<u>0.9568</u>	0.9519	0.9441	0.9604	0.9544
VIF	0.6397	0.6399	0.6455	0.6457	0.6482	0.6744	0.6835	<u>0.6936</u>	0.6743	0.6557	0.7120	0.6751
PI	5.633	6.071	<u>5.854</u>	5.859	6.973	6.181	6.091	6.212	6.149	5.886	6.258	6.044

from the involvement, as bizarre artifacts are explicitly present in small patches. While MSE loss is more compatible to be adopted in various formulations and generates high quality images.

4.2. Deblurring

In this paper, we also propose a deblurring network named MixNet. We adopt densely connected encoder-decoder structure to pursue strong deblurring performance. We removed all the parameter sharing used in [81]. This does not mean that parameter sharing is not useful, but the combination of parameter sharing in intra-scale, cross-scale, and multiscale structures can lead to more containment to the optimisation of shared parameters, and hence may lead to worse performance. In addition, we removed the multiscale structure to further simplify the network.

Network Architecture: As shown in Fig. 7, the network consists of convolutional layers, DenseBlocks and Inception-A blocks, where I_b and I_s denote input blur image and output image, respectively. The backbone of the network adopts DenseBlocks with ResBlock as its elements for enhanced receptive field. By default, each nonlinear DenseBlock module has four processing units. Structures of the ResBlocks used are indicated in Fig. 7, comprising two convolutional layers. For the framework, we remove both multiscale structure and parameter sharing mechanism, leading to a simplified network. The encoder-decoder structure is based on 12 DenseBlocks with independent parameters. We add four Inception-A blocks in the middle of the network, to widen the feature maps. Inception-A is one of the components of Inception-v4 and has suitable input size and feature map width for image restoration, con-

taining 382 channels. For Inception-v4, there are five types of blocks: Inception-A, Reduction-A, Inception-B, Reduction-B and Inception-C. But these blocks, except for Inception-A, have over 1000 features and hence are not suitable for image reconstruction. By adopting inception-A, the proposed network has mixed feature extraction mechanism from DenseNet and Inception-A network and therefore is termed as MixNet. Different from using kernel size of 5×5 [50,80], we use kernel size of 3×3 to control the model size, since 2 layers with 3×3 kernels can cover the same receptive fields as one layer with 5×5 kernels but saving around 25 % of parameters.

Loss function is another significant element of image deblurring networks. Based on the review in Section 2, the MSE loss is referred to as the most important loss for image deblurring. It has direct relationship with PSNR, one of the most important measures in performance evaluation, given as,

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right). \quad (17)$$

Therefore in this work, we adopt the MSE loss as the loss function. From our experience, adding other auxiliary loss such as SSIM loss or adversarial loss may not always have significant effect in deblurring.

Implementation: We implemented the proposed MixNet by TensorFlow [264] on a NVIDIA Tesla P100 GPU. During training, a 256×256 region from a blurred and its ground truth image at the same location were randomly cropped out as the training input. The batch size was set to 16 during training. All weights



Fig. 6. Resulting images generated by GANs trained using various training objective functions (Example image: 0865.PNG from DIV2K validation dataset, Quality measures: PSNR/SSIM/VIF/PI).

were initialised using the Xavier method [265], and biases were initialized to zero. The network was optimised using the Adam method [266] with default setting $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The learning rate was initially set to 0.0001 and exponentially decayed to 0 using power of 0.3. According to the experiments, 2000 epochs were sufficient for the network to converge.

Results and Comparisons: We conducted experiments on the proposed MixNet and compared with the state-of-the-art methods for dynamic scene deblurring and non-uniform deblurring on the GoPro dataset. The methods compared include DeepDeblur [50], Scale Recurrent Network (SRN-Deblur) [80], DSHMN [82], DeblurGAN [33], DeblurGANv2 [64], Unsupervised deblur [85], Domain specific [86], SVRNN [89], Dual Residual [90], Douglas-Rachford network [87], Region Adaptive [88], and Blur2Flow [262]. The results were generated by the models trained on the default GoPro

training dataset, and then tested on the GoPro testing dataset. For unsupervised learning methods, Unsupervised deblur and Domain specific, we used blurry images from the training dataset, and sharp images from the New GoPro dataset, which has higher resolution. For kernel based methods including Blur2Flow and optimisation based methods, we tested them on their released model codes. The quantitative results and evaluations are presented in Table 9.

In general, unsupervised learning methods lead to low PSNR and SSIM, as expected due to no supervision (ground truth) involved and limited size of the dataset for unsupervised learning. In addition, these networks are mainly developed for exploring such training mechanism, while the network structures are largely under-developed. Although Dr-Net had the best SSIM performance, the proposed MixNet achieved the-state-of-the-art performances

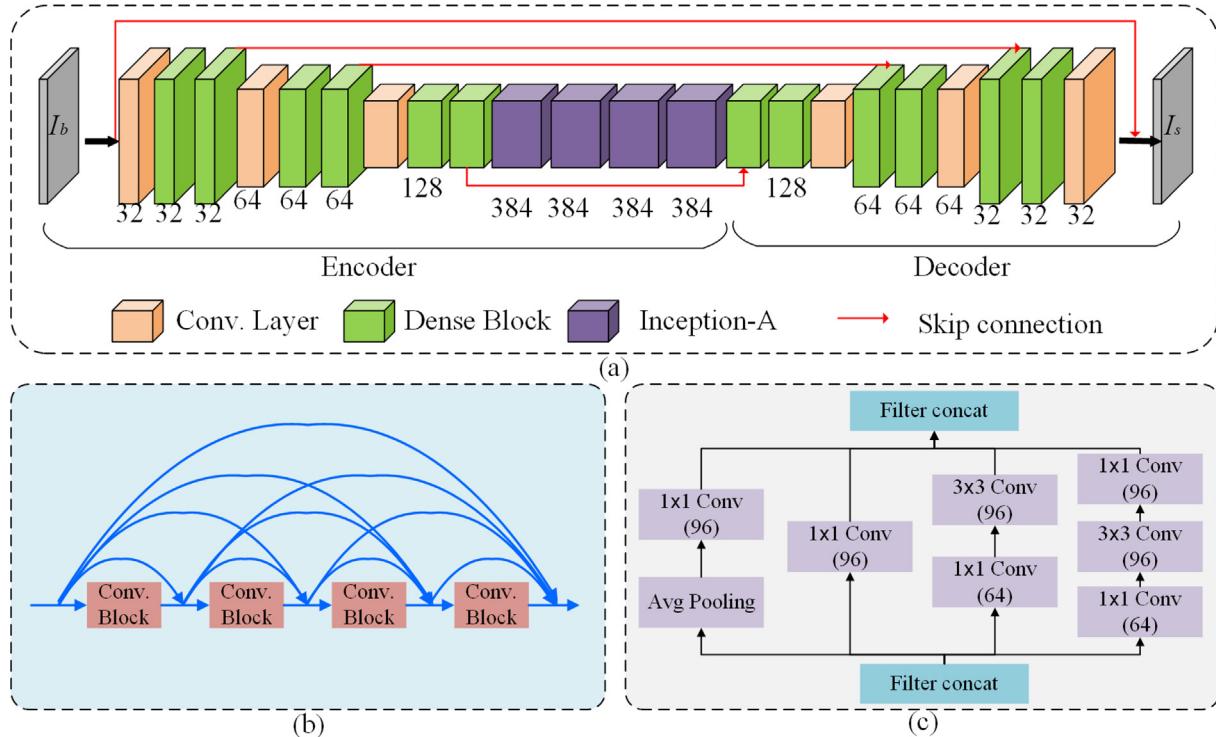


Fig. 7. Proposed MixNet network for image deblurring: (a) overall architecture, (b) DenseBlock structure, and (c) Inception-A block.

Table 9

Testing results of proposed MixNet deblurring network on GoPro dataset.

Algorithm	PSNR	SSIM	FSIM	IFC	VIF	RunTime	Size
DeepDeblur[50]	29.08	0.9135	0.9633	3.2916	0.9955	3.09s	432 MB
SRN-Deblur[80]	30.26	0.9432	0.9653	3.0750	0.9878	0.86s	27.5 MB
PSS-NSC[81]	30.92	0.9421	0.9756	3.4136	0.9914	0.68s	46.5 MB
DSHMN[82]	31.50	0.9483	0.9742	3.2697	0.9915	0.552s	86.8 MB
DeblurGAN[33]	28.70	0.9580	0.9688	3.1124	0.9732	0.85s	33.8 MB
DeblurGANv2[64]	29.55	0.9340	0.9527	2.6929	0.9697	0.35s	244.5 MB
Unsupervised deblur[85]	18.64	0.6219	0.8220	1.0233	0.6401	1.23s	165.3 MB
Disentangled Representation[86]	20.15	0.6623	0.8317	1.1790	0.6788	0.96s	636 MB
SVRNN[89]	29.19	0.9306	0.9446	2.8934	0.9822	1.40s	37.1 MB
Dual Residual[90]	29.90	0.9100	0.9436	2.6387	0.9948	1.14s	51.4 MB
Dr-Net[87]	30.35	0.961	0.9812	3.5540	0.9964	1.20s	91.8 MB
RADIN-Deblur[88]	31.76	0.9530	0.9804	3.4298	0.9975	0.038s	64.5 MB
Blur2Flow[262]	26.99	0.9233	0.9562	3.1432	0.9842	1.70s	41.2 MB
MixNet (ours)	32.21	0.9534	0.9827	3.6804	0.9987	0.54s	97.0 MB

on all other evaluation criteria. Furthermore, our MixNet has a good trade-off on run-time and performance, while Dr-Net's run-time was twice as much as MixNet. A visual comparison on GoPro evaluation dataset is shown in Fig. 8. As illustrated, the proposed model generally produces better results than other methods. Domain specific network is a representative of unsupervised deblurring, it is obviously that it has certain distortion in colour.

We also evaluated and compared our method on the HIDE dataset and quantitative results are shown in Table 10. These results were generated by the models trained on the default HIDE training dataset. As illustrated, the proposed MixNet either outperformed or matched these state-of-the-art methods in all these evaluation criteria.

4.3. Contributions

Here we summarise the key contributions of this paper.

- Comprehensive review

A comprehensive literature review was conducted on image restoration, image deblurring, image denoising, image dehaz-

ing, super-resolution and image quality assessments. All corresponding baseline deep models were also reviewed.

- New methods for super-resolution and deblurring
A novel formulation of training objective functions of GANs was proposed for super-resolution, as an extension of the *Hypervol* formulation [257]. A novel balanced image deblurring method named *MixNet* was proposed.
- Experimental verification Extensive experiments of applying *Ed* and *Hypervol* formulations to various training objective functions of SRGAN for super-resolution were conducted and improvements were obtained. Experiments were performed to compare the proposed *MixNet* with the state-of-the-art deblurring method, and the results show that *MixNet* has better performance compared with mainstream image deblurring networks.

5. Conclusions and Discussion

Image restoration is a challenging image processing task due to its ill-posed nature. Conventional methods rely on handcrafted

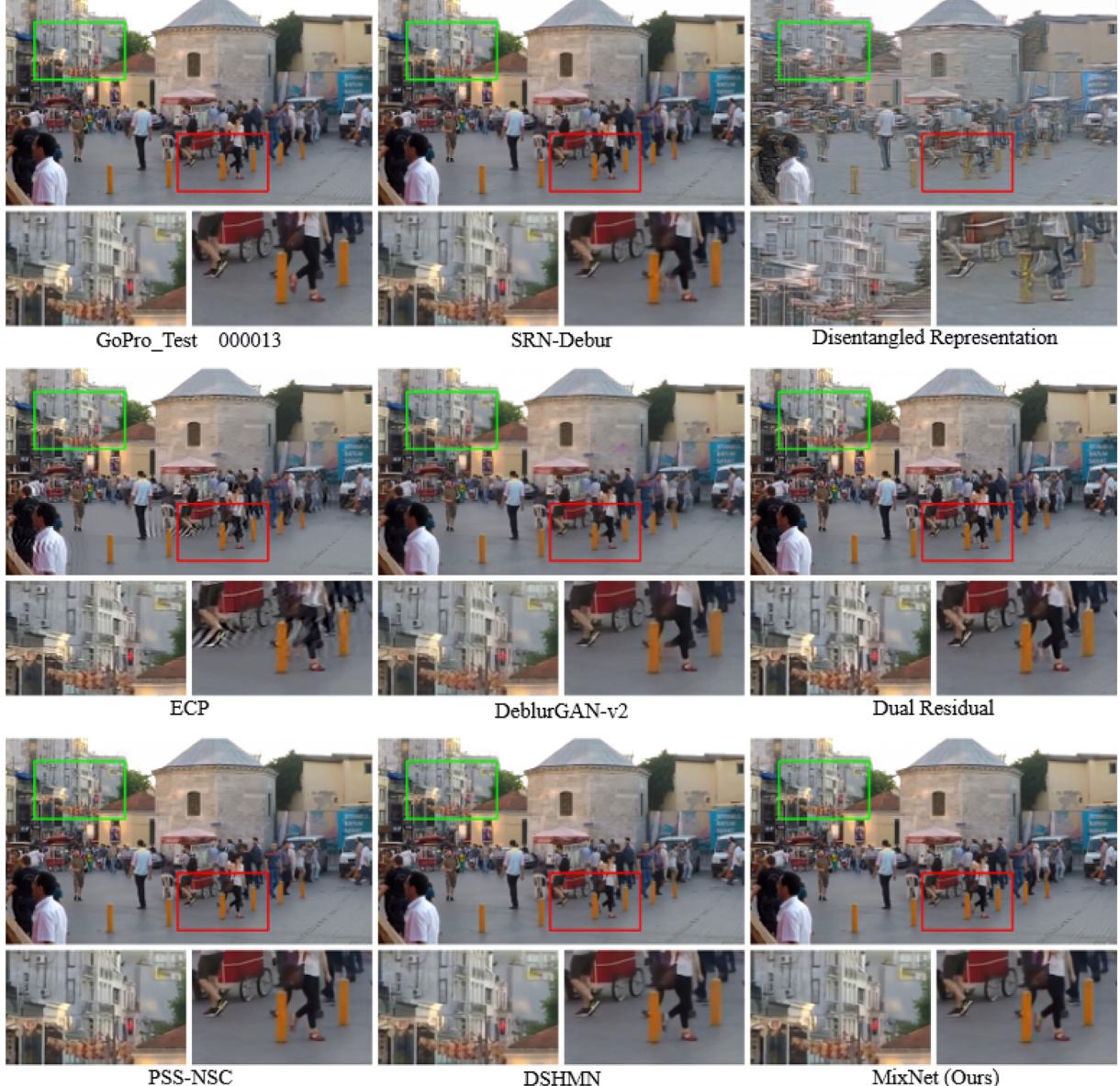


Fig. 8. Visual comparisons of deblurring results on images from GoPro test [50]. The tested methods include SRN-Deblur [80], Disentangled Representation [86], ECP [263], DeblurGAN-v2 [64], Dual Residual [90], PSS-NSC [81], DSHMN [82], and our MixNet.

Table 10
Testing results of MixNet deblurring network on HIDE datasets.

Algorithm	HIDE I (long-short)					HIDE II (close-ups)				
	PSNR	SSIM	FSIM	IFC	VIF	PSNR	SSIM	FSIM	IFC	VIF
DeepDeblur[50]	27.43	0.9020	0.9622	3.1032	0.9877	26.18	0.8780	0.9339	2.1277	0.9699
SRN-Deblur[80]	29.41	0.9137	0.9681	3.3863	0.9939	27.54	0.9070	0.9393	2.2656	0.9710
PSS-NSC[81]	29.98	0.9234	0.9728	3.5607	0.9955	28.14	0.9021	0.9470	2.3680	0.9752
DSHMN[82]	29.79	0.9247	0.9736	3.5627	0.9966	28.33	0.9099	0.9501	2.4560	0.9789
DeblurGAN[33]	26.44	0.8900	0.9440	3.0332	0.9697	25.37	0.8670	0.9244	2.0787	0.9654
DeblurGANv2[64]	28.29	0.8960	0.9591	3.1759	0.9746	26.64	0.8722	0.9248	1.9547	0.9310
Unsupervised deblur[85]	15.64	0.4988	0.7366	1.2894	0.5493	14.09	0.4625	0.7233	0.9140	0.4581
DR-Net[86]	17.23	0.5136	0.7791	1.3280	0.5785	16.41	0.4942	0.7470	0.9267	0.4925
SVRNN[89]	28.69	0.9038	0.9614	3.2097	0.9985	26.68	0.8702	0.9231	1.7689	0.9724
Dual Residual[90]	28.75	0.9022	0.9620	3.1859	0.9914	26.05	0.8605	0.9128	1.7870	0.9767
Dr-Net[87]	29.63	0.9135	0.9654	3.6233	0.9728	28.29	0.9043	0.9488	2.2420	0.9762
RADN-Deblur[88]	28.97	0.9044	0.9637	3.1517	0.9927	26.51	0.8698	0.9130	1.7519	0.9752
Blur2Flow[262]	28.25	0.9011	0.9668	3.0964	0.9923	26.87	0.8610	0.9155	1.5406	0.9722
MixNet (ours)	30.72	0.9277	0.9745	3.5747	0.9978	28.72	0.9137	0.9517	2.5837	0.9838

models to construct the degradation mechanisms and noise models. In practice degradation mechanisms and noise models are rarely simplistic and uniform. Therefore, learning based approaches become more applicable and often outperform conventional methods significantly. Deep learning networks have been particularly prevalent and there have been a large body of research studies from deblurring, denoising and dehazing to super-resolution. We have comprehensively reviewed approaches for these tasks and summarised typical and useful mechanisms and their benefits for various restoration tasks. We also proposed certain training objectives and a novel formulation for super-resolution, as well as an efficient deep networks for blind single image deblurring. Experimental results demonstrate their benefits and improved performances over the state-of-the-art methods.

Image denoising has extraordinary value for low-level vision and signal processing and can serve as an ideal test bed for evaluating image priors and optimisation methods from Bayesian perspective. Traditionally, image denoising methods are inspired from existing signal denoising approaches. Learning based methods can significantly boost the performance and recover sharp images with little prior knowledge. However, recent research is mainly based on synthesised noisy images rather than real-world noisy images, for which its distribution is unknown and rarely Gaussian as often assumed. Real-world image denoising will remain as a challenge.

Fine-detail reconstruction is the ultimate purpose in image deblurring, using either expert knowledge such as natural scene statistics or features learnt by deep neural networks. These image priors are exploited as regularisation techniques for solving image deblurring problems, added into the cost function to be minimised during optimisation. Because dynamic scene blurs are intrinsically variant across the entire image, it is difficult to adopt hand-crafted method to estimate kernels for restoration. Instead, training end-to-end deep neural networks is equipped with powerful learning capability to approximate the mapping between degraded input image and clean output image. However, careful calibration of network architectures and parameters is a necessary and requires dedicated efforts in order to optimise model performance.

Reconstruction of a high resolution image from its degraded low resolution version involves not only dimension increase but also deconvolution. On the basis of a given image, pixels to be interpolated are estimated either by classical bicubic function or learnt by deep neural networks. Deep learning neural networks have been shown to yield superior performances. Unampling layers in deep neural networks interpolate sub-pixels and increase image dimension, such that image can be rescaled and quality further enhanced with finer details and textures. Besides, because a training objective function often employs various losses as constraints for the solution space, it also determines the direction of optimisation and hence the performance. Image quality measure based loss is shown to be effective for improving model performance and is straightforward and convenient to adopt.

Although most deep learning methods are based on supervised learning, unsupervised or semi-supervised learning can also benefit image restoration in terms of better data representation and therefore has been increasingly combined with the supervised mechanisms. Further exploitation in this direction could make restoration tasks more efficient and effective. Graph and self-organising structures could be integrated into the supervised learning schemes in the deep networks to make the ill-posed inverse problems more tenable, more efficacy, and less reliant on massive number of paired training samples.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] S. Geman, D. Geman, Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* 6 (1984) 721–741.
- [2] W.H. Richardson, Bayesian-based iterative method of image restoration, *J. Opt. Soc. Amer.* 62 (1) (1972) 55–59.
- [3] J. Besag, J. York, A. Mollié, Bayesian image restoration, with two applications in spatial statistics, *Ann. Inst.Stat. Math.* 43 (1) (1991) 1–20.
- [4] A. Rosenfeld, A.C. Kak, *Digital Picture Processing*, 1976.
- [5] A. Jain, Advances in mathematical models for image processing, *Proc. IEEE* 69 (5) (1981) 502–528.
- [6] G. Chantas, N.P. Galatsanos, R. Molina, A.K. Katsaggelos, Variational bayesian image restoration with a product of spatially weighted total variation image priors, *IEEE Trans. Image Process.* 19 (2) (2009) 351–362.
- [7] R. Molina, J. Núñez, F.J. Cortijo, J. Mateos, Image restoration in astronomy: a bayesian perspective, *IEEE Signal Process. Mag.* 18 (2) (2001) 11–29.
- [8] G.K. Chantas, N.P. Galatsanos, A.C. Likas, Bayesian restoration using a new nonstationary edge-preserving image prior, *IEEE Trans. Image Process.* 15 (10) (2006) 2987–2997.
- [9] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, W.T. Freeman, Removing camera shake from a single photograph, *Int. Conf. Comput. Graphics Interactive Techniques* 25 (3) (2006) 787–794.
- [10] S.J. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, *Multiscale Modeling Simul.* 4 (2) (2005) 460–489.
- [11] A. Beck, M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, *IEEE Trans. Image Process.* 18 (11) (2009) 2419–2434.
- [12] K.I. Kim, Y. Kwon, Single-image super-resolution using sparse regression and natural image prior, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (6) (2010) 1127–1133.
- [13] J. Sun, Z. Xu, H.-Y. Shum, Image super-resolution using gradient profile prior, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, IEEE, 2008, pp. 1–8.
- [14] H. Zhang, Y. Zhang, H. Li, T.S. Huang, Generative bayesian image super resolution with natural image prior, *IEEE Trans. Image Process.* 21 (9) (2012) 4054–4067.
- [15] T.F. Chan, J. Shen, Variational image inpainting, *Commun. Pure Appl. Math.: J. Issued Courant Institute Math. Sci.* 58 (5) (2005) 579–619.
- [16] J. Shen, Inpainting and the fundamental problem of image processing, *SIAM news* 36 (5) (2003) 1–4.
- [17] Z. Xu, J. Sun, Image inpainting by patch propagation using patch sparsity, *IEEE Trans. Image Process.* 19 (5) (2010) 1153–1165.
- [18] R. Narayan, R. Nityananda, Maximum entropy image restoration in astronomy, *Ann. Rev. Astron. Astrophys.* 24 (1) (1986) 127–170.
- [19] Y. Bobichon, A. Bijaoui, A regularized image restoration algorithm for lossy compression in astronomy, *Exp. Astron.* 7 (3) (1997) 239–255.
- [20] J.-L. Starck, E. Pantin, F. Murtagh, Deconvolution in astronomy: A review, *Publications of Astronomical Soc. Pacific* 114 (800) (2002) 1051.
- [21] M.G. Sanchez, M.G. Sánchez, V. Vidal, G. Verdu, G. Verdú, P. Mayo, F. Rodenas, Medical image restoration with different types of noise, in: *Proc. Int. Conf. IEEE Engineering Medicine Biol. Soc.*, IEEE, 2012, pp. 4382–4385.
- [22] J. Jan, *Medical image processing, reconstruction and restoration: concepts and methods*, Crc press, 2005.
- [23] J. Ma, J. Huang, Q. Feng, H. Zhang, H. Lu, Z. Liang, W. Chen, Low-dose computed tomography image restoration using previous normal-dose scan, *Med. Phys.* 38 (10) (2011) 5713–5731.
- [24] L. Schaefer, D. Schuster, H. Herz, Generalized approach for accelerated maximum likelihood based image restoration applied to three-dimensional fluorescence microscopy, *J. Microscopy* 204 (2) (2001) 99–107.
- [25] P.A. Penczek, Image restoration in cryo-electron microscopy, *Methods Enzymol.* 482 (2010) 35–72.
- [26] J.B. De Monvel, S. Le Calvez, M. Ulfendahl, Image restoration for confocal microscopy: improving the limits of deconvolution, with application to the visualization of the mammalian hearing organ, *Biophysical J.* 80 (5) (2001) 2455–2470.
- [27] X. Li, Y. Hu, X. Gao, D. Tao, B. Ning, A multi-frame image super-resolution method, *Signal Process.* 90 (2) (2010) 405–414, <https://doi.org/10.1016/j.sigpro.2009.05.028>.
- [28] M. Van Noort, L.R. Van Der Voort, M.G. Löfdahl, Solar image restoration by use of multi-frame blind de-convolution with multiple objects and phase diversity, *Sol. Phys.* 228 (1) (2005) 191–215.
- [29] M. Sonogashira, T. Funatomi, M. Iiyama, M. Minoh, Variational bayesian approach to multiframe image restoration, *IEEE Trans. Image Process.* 26 (5) (2017) 2163–2178.
- [30] J. Mairal, G. Sapiro, M. Elad, Learning multiscale sparse representations for image and video restoration, *Multiscale Modeling Simul.* 7 (1) (2008) 214–241.
- [31] S.H. Chan, R. Khoshabeh, K.B. Gibson, P.E. Gill, T.Q. Nguyen, An augmented lagrangian method for total variation video restoration, *IEEE Trans. Image Process.* 20 (11) (2011) 3097–3111.
- [32] H. Ji, S. Huang, Z. Shen, Y. Xu, Robust video restoration by joint sparse and low rank matrix approximation, *SIAM J. Imaging Sci.* 4 (4) (2011) 1122–1142.

- [33] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, J. Matas, Deblurgan: Blind motion deblurring using conditional adversarial networks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 8183–8192.
- [34] Q. Guo, J. Sun, F. Juefei-Xu, L. Ma, X. Xie, W. Feng, Y. Liu, Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining, in: AAAI, 2021.
- [35] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, *IEEE Trans. Image Process.* 26 (7) (2017) 3142–3155.
- [36] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (2017) 105–114.
- [37] X. Qin, Z. Wang, Y. Bai, X. Xie, H. Jia, Ffa-net: Feature fusion attention network for single image dehazing, in: Proc. AAAI Conf. Artif. Intell., 2020, pp. 11908–11915.
- [38] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [39] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognt., 2016, pp. 770–778.
- [40] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma, Pcanet: A simple deep learning baseline for image classification?, *IEEE Trans Image Process.* 24 (12) (2015) 5017–5032.
- [41] S.-H. Zhong, Y. Liu, Y. Liu, Bilinear deep learning for image classification, in: Proc. 19th ACM Int. Conf. Multimedia, 2011, pp. 343–352.
- [42] V. Santhanam, V.I. Morariu, L.S. Davis, Generalized deep image to image regression, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 5609–5619.
- [43] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, R. Horau, A comprehensive analysis of deep regression, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (9) (2019) 2065–2081.
- [44] X. Yan, J. Yang, K. Sohn, H. Lee, Attribute2image: Conditional image generation from visual attributes, in: Eur. Conf. Comput. Vis., Springer, 2016, pp. 776–791.
- [45] K. Gregor, I. Danihelka, A. Graves, D. Rezende, D. Wierstra, Draw: A recurrent neural network for image generation, in: Int. Conf. Mach. Learn., PMLR, 2015, pp. 1462–1471.
- [46] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2012, pp. 1097–1105.
- [47] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).
- [48] C.J. Schuler, H. Christopher Burger, S. Harmeling, B. Scholkopf, A machine learning approach for non-blind image deconvolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2013, pp. 1067–1074.
- [49] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced deep residual networks for single image super-resolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2017, pp. 1132–1140.
- [50] S. Nah, T. Hyun Kim, K. Mu Lee, Deep multi-scale convolutional neural network for dynamic scene deblurring, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 3883–3891.
- [51] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 4700–4708.
- [52] H. Fan, H. Su, L.J. Guibas, A point set generation network for 3d object reconstruction from a single image, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognt., 2017, pp. 605–613.
- [53] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T.S. Huang, Generative image inpainting with contextual attention, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognt., 2018, pp. 5505–5514.
- [54] K. Zhang, W. Zuo, S. Gu, L. Zhang, Learning deep cnn denoiser prior for image restoration, in: IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2808–2817.
- [55] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep networks for image super-resolution with sparse prior, in: IEEE Int. Conf. Comput. Vis. (ICCV), 2015, pp. 370–378.
- [56] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A.V.D. Hengel, Q. Shi, From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur, *IEEE Conf.*, in: 2017 Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 3806–3815.
- [57] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein (2011), <https://doi.org/10.1561/2200000016>.
- [58] D. Geman, Chengda Yang, Nonlinear image recovery with half-quadratic regularization, *IEEE Trans. Image Process.* 4 (7) (1995) 932–946, <https://doi.org/10.1109/83.392335>.
- [59] S. Lefkimiatis, Non-local color image denoising with convolutional neural networks, in: IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 5882–5891.
- [60] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, I.S. Kweon, Learning a deep convolutional network for light-field image super-resolution, in: IEEE Int. Conf. Comput. Vis. Workshop (ICCVW), 2015, pp. 57–65.
- [61] T.M. Nimisha, A.K. Singh, A.N. Rajagopalan, Blur-invariant deep learning for blind-deblurring, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 4762–4770.
- [62] V. Jain, J. Murray, F. Roth, S. Turaga, V. Zhigulin, K. Briggman, M. Helmstaedter, W. Denk, H. Seung, Supervised learning of image restoration with convolutional networks, in: IEEE Int. Conf. Comput. Vis., 2007, pp. 1–8.
- [63] J. Li, K.A. Skinner, R.M. Eustice, M. Johnson-Roberson, Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images, *IEEE Robotics and Autom. Letters* 3 (1) (2018) 387–394.
- [64] O. Kupyn, T. Martyniuk, J. Wu, Z. Wang, Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2019, pp. 8878–8887.
- [65] R. Li, J. Pan, Z. Li, J. Tang, Single image dehazing via conditional generative adversarial network, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 8202–8211.
- [66] J. Pan, J. Dong, Y. Liu, J. Zhang, J. Ren, J. Tang, Y.W. Tai, M.-H. Yang, Physics-based generative adversarial models for image restoration and beyond, *IEEE Trans. Pattern Anal. Mach. Intell.* (2020), 1–1.
- [67] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, arXiv preprint arXiv:1606.03498 (2016).
- [68] M. Li, J. Lin, Y. Ding, Z. Liu, J.-Y. Zhu, S. Han, Gan compression: Efficient architectures for interactive conditional gans, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 5284–5294.
- [69] A. Brock, J. Donahue, K. Simonyan, Large scale gan training for high fidelity natural image synthesis, in: Proc. Int. Conf. Learn. Representations, 2018.
- [70] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, *IEEE Trans. Image Process.* 29 (2020) 4376–4389.
- [71] C. Belthangady, L.A. Royer, Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction, *Nat. Methods* 16 (12) (2019) 1215–1225.
- [72] C. You, W. Cong, M.W. Vannier, P.K. Saha, E.A. Hoffman, G. Wang, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle), *IEEE Trans. Medical Imaging* 39 (1) (2020) 188–203.
- [73] Z. Hu, L. Xu, M.-H. Yang, Joint depth estimation and camera shake removal from single blurry image, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014, pp. 2893–2900.
- [74] A. Gupta, N. Joshi, C.L. Zitnick, M. Cohen, B. Curless, Single image deblurring using motion density functions, in: Proc. IEEE Eur. Conf. Comput. Vis. (ECCV), Springer, 2010, pp. 171–184.
- [75] R. Wang, D. Tao, Recent progress in image deblurring, arXiv preprint arXiv:1409.6838 (2014).
- [76] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, M.-H. Yang, A comparative study for single image blind deblurring, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 1701–1709.
- [77] S. Nah, S. Son, R. Timofte, K. Mu Lee, Ntire 2020 challenge on image and video deblurring, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, 2020, pp. 416–417.
- [78] J. Koh, J. Lee, S. Yoon, Single-image deblurring with neural networks: A comparative survey, *Comput. Vis. Image Understanding* 203 (2021), <https://doi.org/10.1016/j.cviu.2020.103134>.
- [79] D. Eigen, C. Puhrsch, R. Fergus, Depth map prediction from a single image using a multi-scale deep network, in: Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2014, pp. 2366–2374.
- [80] X. Tao, H. Gao, X. Shen, J. Wang, J. Jia, Scale-recurrent network for deep image deblurring, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 8174–8182.
- [81] H. Gao, X. Tao, X. Shen, J. Jia, Dynamic scene deblurring with parameter selective sharing and nested skip connections, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 3848–3856.
- [82] H. Zhang, Y. Dai, H. Li, P. Koniusz, Deep stacked hierarchical multi-patch network for image deblurring, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 5978–5986.
- [83] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2117–2125.
- [84] A. Kirillov, R. Girshick, K. He, P. Dollár, Panoptic feature pyramid networks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 6399–6408.
- [85] T. Madam, Nimisha, K. Sunil, A. Rajagopalan, Unsupervised class-specific deblurring, in: Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 353–369.
- [86] B. Lu, J.-C. Chen, R. Chellappa, Unsupervised domain-specific deblurring via disentangled representations, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 10225–10234.
- [87] R. Aljadaany, D.K. Pal, M. Savvides, Douglas-rachford networks: Learning both the image prior and data fidelity terms for blind image deconvolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 10235–10244.
- [88] K. Purohit, A. Rajagopalan, Region-adaptive dense network for efficient motion deblurring, arxiv eprints, page, arXiv preprint arXiv:1903.11394 (2019).
- [89] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R.W. Lau, M.-H. Yang, Dynamic scene deblurring using spatially variant recurrent neural networks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 2521–2529.
- [90] X. Liu, M. Suganuma, Z. Sun, T. Okatanai, Dual residual networks leveraging the potential of paired operations for image restoration, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 7007–7016.

- [91] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, *IEEE Trans. Image Process.* 16 (8) (2007) 2080–2095.
- [92] S.Z. Li, Markov random field modeling in image analysis, Springer Science & Business Media, 2009.
- [93] M. Gharbi, T.-M. Li, M. Aittala, J. Lehtinen, F. Durand, Sample-based monte carlo denoising using a kernel-splatting network, *ACM Trans. Graphics (TOG)* 38 (4) (2019) 1–12.
- [94] K. Zhang, W. Zuo, L. Zhang, Ffdnet: Toward a fast and flexible solution for cnn-based image denoising, *IEEE Trans. Image Process.* 27 (9) (2018) 4608–4622.
- [95] Y. Chen, T. Pock, Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6) (2016) 1256–1272.
- [96] D. Liu, B. Wen, Y. Fan, C.C. Loy, T.S. Huang, Non-local recurrent network for image restoration, arXiv preprint arXiv:1806.02919 (2018).
- [97] P. Liu, H. Zhang, K. Zhang, L. Lin, W. Zuo, Multi-level wavelet-cnn for image restoration, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018, pp. 773–782.
- [98] D. Valsesia, G. Fracastoro, E. Magli, Deep graph-convolutional image denoising, *IEEE Trans. Image Process.* 29 (2020) 8226–8237.
- [99] S. Anwar, N. Barnes, Real image denoising with feature attention, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3155–3164.
- [100] S. Guo, Z. Yan, K. Zhang, W. Zuo, L. Zhang, Toward convolutional blind denoising of real photographs, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1712–1722.
- [101] Z. Yue, H. Yong, Q. Zhao, L. Zhang, D. Meng, Variational denoising network: Toward blind noise modeling and removal, arXiv preprint arXiv:1908.11314 (2019).
- [102] M. Chang, Q. Li, H. Feng, Z. Xu, Spatial-adaptive network for single image denoising, in: European Conference on Computer Vision, Springer, 2020, pp. 171–187.
- [103] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, L. Shao, Cycleisp: Real image restoration via improved data synthesis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2696–2705.
- [104] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, L. Shao, Multi-stage progressive image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14821–14831.
- [105] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, L. Shao, Learning enriched features for real image restoration and enhancement, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16, Springer, 2020, pp. 492–511.
- [106] Z. Wang, X. Cun, J. Bao, J. Liu, Uformer: A general u-shaped transformer for image restoration, arXiv preprint arXiv:2106.03106 (2021).
- [107] L. Chen, X. Lu, J. Zhang, X. Chu, C. Chen, Hinet: Half instance normalization network for image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 182–192.
- [108] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, Restormer: Efficient transformer for high-resolution image restoration, arXiv preprint arXiv:2111.09881 (2021).
- [109] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, C.-W. Lin, Deep learning image denoising: An overview, *Neural Netw.* (2020).
- [110] J. Liang, R. Liu, Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network, in: Proc. Int. Congress Image Signal Process. (CISP), IEEE, 2015, pp. 697–701.
- [111] Q. Xu, C. Zhang, L. Zhang, Denoising convolutional neural network, in: Proc. IEEE Int. Conf. Inf. Autom., IEEE, 2015, pp. 1184–1187.
- [112] X.-J. Mao, C. Shen, Y.-B. Yang, Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections, in: Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS), 2016, pp. 2810–2818.
- [113] R. Fattal, Single image dehazing, *ACM Trans. graphics (TOG)* 27 (3) (2008) 1–9.
- [114] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (12) (2010) 2341–2353.
- [115] K. Tang, J. Yang, J. Wang, Investigating haze-relevant features in a learning framework for image dehazing, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014, pp. 2995–3000.
- [116] B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, An all-in-one network for dehazing and beyond, arXiv preprint arXiv:1707.06543 (2017).
- [117] Y. Qu, Y. Chen, J. Huang, Y. Xie, Enhanced pix2pix dehazing network, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 8160–8168.
- [118] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, G. Hua, Gated context aggregation network for image dehazing and deraining, in: Proc. IEEE Winter Conf. Applications of Comput. Vis. (WACV), IEEE, 2019, pp. 1375–1383.
- [119] C.-H. Yeh, C.-H. Huang, L.-W. Kang, Multi-scale deep residual learning-based single image haze removal via image decomposition, *IEEE Trans. Image Process.* 29 (2019) 3153–3167.
- [120] A. Duhdane, S. Murala, Ryf-net: Deep fusion network for single image haze removal, *IEEE Trans. Image Process.* 29 (2019) 628–640.
- [121] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, N. Sang, M.-H. Yang, Semi-supervised image dehazing, *IEEE Trans. Image Process.* 29 (2019) 2766–2779.
- [122] A. Golts, D. Freedman, M. Elad, Unsupervised single image dehazing using dark channel prior loss, *IEEE Trans. Image Process.* 29 (2019) 2692–2701.
- [123] J. Park, D.K. Han, H. Ko, Fusion of heterogeneous adversarial networks for single image dehazing, *IEEE Trans. Image Process.* 29 (2020) 4721–4732.
- [124] D. Engin, A. Genç, H. Kemal Ekenel, Cycle-dehaze: Enhanced cylegan for single image dehazing, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, 2018, pp. 825–833.
- [125] J. Zhang, D. Tao, Famed-net: A fast and accurate multi-scale end-to-end dehazing network, *IEEE Trans. Image Process.* 29 (2019) 72–84.
- [126] Y. Shao, L. Li, W. Ren, C. Gao, N. Sang, Domain adaptation for image dehazing, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 2808–2817.
- [127] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, M.-H. Yang, Gated fusion network for single image dehazing, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3253–3261.
- [128] H. Zhang, V.M. Patel, Densely connected pyramid dehazing network, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogit., 2018, pp. 3194–3203.
- [129] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: An end-to-end system for single image haze removal, *IEEE Trans. Image Process.* 25 (11) (2016) 5187–5198.
- [130] X. Liu, Y. Ma, Z. Shi, J. Chen, Griddehazenet: Attention-based multi-scale network for image dehazing, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7314–7323.
- [131] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, M.-H. Yang, Multi-scale boosted dehazing network with dense feature fusion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2157–2167.
- [132] M. Hong, Y. Xie, C. Li, Y. Qu, Distilling image dehazing with heterogeneous task imitation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3462–3471.
- [133] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, L. Ma, Contrastive learning for compact single image dehazing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10551–10560.
- [134] S.C. Park, M.K. Park, M.G. Kang, Super-resolution image reconstruction: a technical overview, *IEEE Signal Process. Mag.* 20 (3) (2003) 21–36.
- [135] C.-Y. Yang, C. Ma, M.-H. Yang, Single-image super-resolution: A benchmark, in: Proc. Eur. Conf. Comput. Vis., 2014, pp. 372–386.
- [136] Q. Shan, Z. Li, J. Jia, C.-K. Tang, Fast image/video upsampling, *ACM Trans. Graphics (TOG)* 27 (5) (2008) 1–7.
- [137] C.-Y. Yang, M.-H. Yang, Fast direct super-resolution by simple functions, in: Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 561–568.
- [138] M. Irani, S. Peleg, Improving resolution by image registration, *CVGIP: Graphical Models Image Process.* 53 (3) (1991) 231–239.
- [139] G. Freedman, R. Fattal, Image and video upscaling from local self-examples, *ACM Trans. Graphics* 30 (2) (2011) 12.
- [140] W. Dong, L. Zhang, G. Shi, X. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, *IEEE Trans. Image Process.* 20 (7) (2011) 1838–1857.
- [141] D. Glasner, S. Bagon, M. Irani, Super-resolution from a single image, in: Proc. IEEE Int. Conf. Comput. Vis., 2009, pp. 349–356.
- [142] R. Timofte, V. De, L.V. Gool, Anchored neighborhood regression for fast example-based super-resolution, in: Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 1920–1927.
- [143] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, Q. Liao, Deep learning for single image super-resolution: A brief review, *IEEE Trans. Multimedia* 21 (12) (2019) 3106–3121.
- [144] M. Haris, G. Shakhnarovich, N. Ukita, Deep back-projection networks for super-resolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogit., 2018, pp. 1664–1673.
- [145] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Deep laplacian pyramid networks for fast and accurate super-resolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 5835–5843.
- [146] Y. Tai, J. Yang, X. Liu, C. Xu, Memnet: A persistent memory network for image restoration, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 4549–4557.
- [147] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2790–2798.
- [148] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 2472–2481.
- [149] J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network for image super-resolution, in: Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR), 2016, pp. 1637–1645.
- [150] W. Shi, J. Caballero, F. Huszar, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 1874–1883.
- [151] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 1646–1654.
- [152] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: Proc. Eur. Conf. Comput. Vis., 2014, pp. 184–199.
- [153] N. Joshi, A. Kapoor, H. Tang, Image quality assessment (2010).
- [154] Z. Wang, A.C. Bovik, L. Lu, Why is image quality assessment so difficult?, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process., Vol. 4, IEEE, 2002, pp. IV–3313.
- [155] B. Girod, What's wrong with mean-squared error?, *Digital images human Vis* (1993) 207–220.

- [156] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612.
- [157] Z. Wang, E. Simoncelli, A. Bovik, Multiscale structural similarity for image quality assessment, in: Proc. Asilomar Conf. Signals Syst. Comput., Vol. 2, 2003, pp. 1398–1402.
- [158] C. Li, A.C. Bovik, Three-component weighted structural similarity index, in: Proc. of SPIE, the Int. Soc. Opt. Engineering, Vol. 7242, 2009.
- [159] C. Li, A.C. Bovik, Content-partitioned structural similarity index for image quality assessment, *Signal Process. Image Communication* 25 (7) (2010) 517–526.
- [160] H. Sheikh, M. Sabir, A. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Trans. Image Process.* 15 (11) (2006) 3440–3451.
- [161] H. Sheikh, A. Bovik, Image information and visual quality, *IEEE Trans. Image Process.* 15 (2) (2006) 430–444.
- [162] L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: A feature similarity index for image quality assessment, *IEEE Trans. Image Process.* 20 (8) (2011) 2378–2386.
- [163] A.B. Watson, Dctune: A technique for visual optimization of dct quantization matrices for individual images, (1993).
- [164] A. Beghdadi, B. Pesquet-Popescu, A new image distortion measure based wavelet decomposition, in: Proc. Int. Symposium Signal Process. Its Applications, Vol. 1, 2003, pp. 485–488.
- [165] R. Reisenhofer, S. Bosse, G. Kutyniok, T. Wiegand, A haar wavelet-based perceptual similarity index for image quality assessment, *Signal Process. Image Commun.* 61 (2018) 33–43.
- [166] E. Girshtel, V. Slobodyan, J.S. Weissman, A.M. Eskicioglu, Comparison of three full-reference color image quality measures, in: Soc. of Photo-Opt. Instrumentation Engineers (SPIE) Conf. Series, Vol. 6059, 2006, p. 605908.
- [167] K. Seshadrinathan, A. Bovik, Unifying analysis of full reference image quality assessment, in: Proc. IEEE Int. Conf. Image Process., 2008, pp. 1200–1203.
- [168] L. Zhang, L. Zhang, X. Mou, D. Zhang, A comprehensive evaluation of full reference image quality assessment algorithms, in: Proc. IEEE Int. Conf. Image Process., 2012, pp. 1477–1480.
- [169] I. Avcibas, B. Sankur, K. Sayood, Statistical evaluation of image quality measures, *J. Electronic Imaging* 11 (2) (2002) 206–223.
- [170] K.-H. Thung, P. Raveendran, A survey of image quality measures, in: Proc. Int. Conf. Tech. Postgraduates (TECHPOS), 2009, pp. 1–4.
- [171] W. Lin, C.C.J. Kuo, Perceptual visual quality metrics: A survey, *J. Visual Communication Image Representation* 22 (4) (2011) 297–312.
- [172] L. Ma, S. Li, F. Zhang, K.N. Ngan, Reduced-reference image quality assessment using reorganized dct-based image representation, *IEEE Trans. Multimedia* 13 (4) (2011) 824–829.
- [173] Z. Wang, A.C. Bovik, Reduced- and no-reference image quality assessment, *IEEE Signal Process. Mag.* 28 (6) (2011) 29–40.
- [174] Z. Wang, E.P. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, in: Human Vision and Electronic Imaging X, Vol. 5666, International Society for Optics and Photonics, 2005, pp. 149–159.
- [175] R. Soundararajan, A.C. Bovik, Rred indices: Reduced reference entropic differencing for image quality assessment, *IEEE Trans. Image Process.* 21 (2) (2012) 517–526.
- [176] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, *IEEE J. Sel. Top. Signal Process.* 3 (2) (2009) 202–211.
- [177] E.P. Simoncelli, B.A. Olshausen, Natural image statistics and neural representation, *Annu. Rev. Neurosci.* 24 (1) (2001) 1193–1216.
- [178] W.S. Geisler, Visual perception and the statistical properties of natural scenes, *Annu. Rev. Psychol.* 59 (1) (2008) 167–192.
- [179] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, *IEEE Trans. Image Process.* 21 (12) (2012) 4695–4708.
- [180] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: From natural scene statistics to perceptual quality, *IEEE Trans. Image Process.* 20 (12) (2011) 3350–3364.
- [181] A. Mittal, R. Soundararajan, A.C. Bovik, Making a 'completely blind' image quality analyzer, *IEEE Signal Process. Letters* 20 (3) (2013) 209–212.
- [182] M.A. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the dct domain, *IEEE Trans. Image Process.* 21 (8) (2012) 3339–3352.
- [183] M.A. Saad, A.C. Bovik, C. Charrier, A dct statistics-based blind image quality index, *IEEE Signal Process. Lett.* 17 (6) (2010) 583–586.
- [184] L. Zhang, L. Zhang, A.C. Bovik, A feature-enriched completely blind image quality evaluator, *IEEE Trans. Image Process.* 24 (8) (2015) 2579–2591.
- [185] W. Xue, X. Mou, L. Zhang, A.C. Bovik, X. Feng, Blind image quality assessment using joint statistics of gradient magnitude and laplacian features, *IEEE Trans. Image Process.* 23 (11) (2014) 4850–4862.
- [186] K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment, *IEEE Trans. Multimedia* 17 (1) (2015) 50–63.
- [187] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, L. Zelnik-Manor, The 2018 pirm challenge on perceptual image super-resolution, in: Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops, 2018, pp. 334–355.
- [188] C. Ma, C.-Y. Yang, X. Yang, M.-H. Yang, Learning a no-reference quality metric for single-image super-resolution, *Comput. Vis. Image Understanding* 158 (2017) 1–16.
- [189] Y. Blau, T. Michaeli, The perception-distortion tradeoff, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 6228–6237.
- [190] X. Deng, R. Yang, M. Xu, P.L. Dragotti, Wavelet domain style transfer for an effective perception-distortion tradeoff in single image super-resolution, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2019, pp. 3076–3085.
- [191] S. Vasu, N.T. Madam, A.N. Rajagopalan, Analyzing perception-distortion tradeoff using enhanced perceptual super-resolution network, in: Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops, 2018, pp. 114–131.
- [192] C. Li, A.C. Bovik, X. Wu, Blind image quality assessment using a general regression neural network, *IEEE Trans. Neural Netw.* 22 (5) (2011) 793–799.
- [193] W. Hou, X. Gao, D. Tao, X. Li, Blind image quality assessment via deep learning, *IEEE Trans. Neural Netw.* 26 (6) (2015) 1275–1286.
- [194] P. Ye, J. Kumar, L. Kang, D. Doermann, Real-time no-reference image quality assessment based on filter learning, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2013, pp. 987–994.
- [195] S. Suresh, R.V. Babu, H.J. Kim, No-reference image quality assessment using modified extreme learning machine classifier, *Applied Soft Computing* 9 (2) (2009) 541–552.
- [196] X. Lu, Z. Lin, X. Shen, R. Mech, J.Z. Wang, Deep multi-patch aggregation network for image style, aesthetics, and quality estimation, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2015, pp. 990–998.
- [197] S. Dodge, L. Karan, Understanding how image quality affects deep neural networks, in: Proc. Int. Conf. Quality of Multimedia Experience (QoMEX), 2016, pp. 1–6.
- [198] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, *IEEE Trans. Image Process.* 27 (1) (2018) 206–219.
- [199] S. Bianco, L. Celona, P. Napoletano, R. Schettini, On the use of deep learning for blind image quality assessment, *Signal Image Video Process.* 12 (2) (2018) 355–362.
- [200] X. Liu, J. van de Weijer, A.D. Bagdanov, Rankqa: Learning from rankings for no-reference image quality assessment, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 1040–1049.
- [201] M. Gardner, S. Dorling, Artificial neural networks (the multilayer perceptron)- review of applications in the atmospheric sciences, *Atmos. Environ.* 32 (14) (1998) 2627–2636, [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0).
- [202] S. Zhang, E. Salari, Image denoising using a neural network based non-linear filter in wavelet domain, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process., Vol. 2, 2005, pp. 989–992.
- [203] C.J. Schuler, H.C. Burger, S. Harmeling, B. Scholkopf, A machine learning approach for non-blind image deconvolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2013, pp. 1067–1074.
- [204] A. de Castro, I. Drummond, J. da Silva, A multiscale neural network method for image restoration, *Trends Appl. Comput. Math.* 9 (1) (2008) 41–50.
- [205] K. Sivakumar, U. Desai, Image restoration using a multilayer perceptron with a multilevel sigmoidal function, *IEEE Trans. Signal Process.* 41 (5) (1993) 2018–2022.
- [206] C. Miravet, F.B. Rodriguez, A two-step neural-network based algorithm for fast image super-resolution, *Image Vis. Computing* 25 (9) (2007) 1449–1473.
- [207] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Intelligent. Signal Process.* (2001) 306–351.
- [208] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Proc. Adv. Neural Inf. Process. Syst. (NIPS), Vol. 27, 2014, pp. 2672–2680.
- [209] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 2242–2251.
- [210] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M.K. Singh, M.-H. Yang, Diverse image-to-image translation via disentangled representations, in: Proc. Eur. Conf. Comput. Vis. (ECCV), 2020, pp. 36–52.
- [211] K. Zeng, J. Yu, R. Wang, C. Li, D. Tao, Coupled deep autoencoder for single image super-resolution, *IEEE Trans. Syst. Man Cybern.* 47 (1) (2017) 27–37.
- [212] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, 2016.
- [213] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, *J. Mach. Learn. Research* 15 (1) (2014) 1929–1958.
- [214] S.J. Nowlan, G.E. Hinton, Simplifying neural networks by soft weight-sharing, *Neural Comput.* 4 (4) (1992) 473–493.
- [215] J. Cui, K. Gong, N. Guo, C. Wu, X. Meng, K. Kim, K. Zheng, Z. Wu, L. Fu, B. Xu, Z. Zhu, J. Tian, H. Liu, Q. Li, Pet image denoising using unsupervised deep learning, *Eur. J. Nuclear Med. Mol. Imaging* 46 (13) (2019) 2780–2789.
- [216] X. Chen, S. You, K.C. Tezcan, E. Konukoglu, Unsupervised lesion detection via image restoration with a normative prior, *Medical Image Anal.* 64 (2020) 101713.
- [217] W. Du, H. Chen, H. Yang, Learning invariant representation for unsupervised image restoration, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 14483–14492.
- [218] M.-Y. Liu, T. Breuel, J. Kautz, Unsupervised image-to-image translation networks, in: Proc. Adv. Neural Inf. Process. Syst. (NIPS), Vol. 30, 2017, pp. 700–708.
- [219] Z. Gao, C. Shen, C. Xie, Stacked convolutional auto-encoders for single space target image blind deconvolution, *Neurocomputing* 313 (2018) 295–305.
- [220] Z. Yi, H. Zhang, P. Tan, M. Gong, Dualgan: Unsupervised dual learning for image-to-image translation, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 2868–2876.

- [221] X. Qu, X. Wang, Z. Wang, L. Wang, L. Zhang, Perceptual-dualgan: Perceptual losses for image to image translation with generative adversarial nets, in: Proc. Int. Joint Conf. Neural Netw. (IJCNN), 2018, pp. 1–8.
- [222] M.-Y. Liu, O. Tuzel, Coupled generative adversarial networks, in: Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS), Vol. 29, 2016, pp. 469–477.
- [223] O. Chapelle, B. Schlkopf, A. Zien, *Semi-Supervised Learning* (2010).
- [224] A. Singh, R. Nowak, X. Zhu, Unlabeled data: Now it helps, now it doesn't, in: Proc. Adv. Neural Inf. Process. Syst. (NIPS), Vol. 21, 2008, pp. 1513–1520.
- [225] J.E. van Engelen, H.H. Hoos, A survey on semi-supervised learning, *Mach. Learn.* 109 (2) (2020) 373–440.
- [226] X. Zhu, A.B. Goldberg, R. Brachman, T. Dietterich, Introduction to Semi-Supervised Learning, 2009.
- [227] A.B. Goldberg, X. Zhu, A. Singh, Z. Xu, R.D. Nowak, Multi-manifold semi-supervised learning, in: Int. Conf. Artif. Intell. Stats., Vol. 5, 2009, pp. 169–176.
- [228] Y. Wang, S. Chen, Z.-H. Zhou, New semi-supervised classification method based on modified cluster assumption, *IEEE Trans. Neural Netw.* 23 (5) (2012) 689–702.
- [229] Y.-F. Li, Z.-H. Zhou, Towards making unlabeled data never hurt, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (1) (2015) 175–188.
- [230] G. Blanchard, G. Lee, C. Scott, Semi-supervised novelty detection, *J. Mach. Learn. Res.* 11 (99) (2010) 2973–3009.
- [231] W. Wei, D. Meng, Q. Zhao, Z. Xu, Y. Wu, Semi-supervised transfer learning for image rain removal, in: Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR), 2019, pp. 3877–3886.
- [232] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, L. Bottou, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, *J. Mach. Learn. Res.* 11 (12) (2010).
- [233] L. Gondara, Medical image denoising using convolutional denoising autoencoders, in: IEEE Int. Conf. Data Mining Workshops, IEEE, 2016, pp. 241–246.
- [234] K. Zeng, J. Yu, R. Wang, C. Li, D. Tao, Coupled deep autoencoder for single image super-resolution, *IEEE Trans. Cybern.* 47 (1) (2015) 27–37.
- [235] Z. Cui, H. Chang, S. Shan, B. Zhong, X. Chen, Deep network cascade for image super-resolution, in: Proc. IEEE Eur. Conf. Comput. Vis. (ECCV), Springer, 2014, pp. 49–64.
- [236] M. Zhu, A. Alperovich, O. Johannsen, A. Sulc, B. Goldluecke, An epipolar volume autoencoder with adversarial loss for deep light field super-resolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, 2019, pp. 0–0.
- [237] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, H. Li, Deblurring by realistic blurring, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 2737–2746.
- [238] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: Proc. Int. conf. mach. learn. (ICML), 2017, pp. 214–223.
- [239] F.H.N.H. Hörmander, N.S.B. Totaro, A.V.M. Waldschmidt, Grundlehren der mathematischen wissenschaften 332 (2006).
- [240] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of wasserstein gans, in: Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2017, pp. 5767–5777.
- [241] D.-W. Kim, J. Ryun Chung, S.-W. Jung, Grdn: Grouped residual dense network for real image denoising and gan-based real-world noise modeling, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, 2019, pp. 0–0.
- [242] S. Tripathi, Z.C. Lipton, T.Q. Nguyen, Correction by projection: Denoising images with generative adversarial networks, arXiv preprint arXiv:1803.04477 (2018).
- [243] Z. Chen, Z. Zeng, H. Shen, X. Zheng, P. Dai, P. Ouyang, Dn-gan: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images, *Biomed. Signal Process. Control* 55 (2020) 101632.
- [244] J. Chen, J. Chen, H. Chao, M. Yang, Image blind denoising with generative adversarial network based noise modeling, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 3155–3164.
- [245] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M.K. Kalra, Y. Zhang, L. Sun, G. Wang, Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss, *IEEE Trans. Medical Imaging* 37 (6) (2018) 1348–1357.
- [246] C.K. Sønderby, J. Caballero, L. Theis, W. Shi, F. Huszár, Amortised map inference for image super-resolution, arXiv preprint arXiv:1610.04490 (2016).
- [247] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. Change Loy, Esrgan: Enhanced super-resolution generative adversarial networks, in: Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops, 2018.
- [248] M. Chu, Y. Xie, J. Mayer, L. Leal-Taixé, N. Thurey, Learning temporal coherence via self-supervision for gan-based video generation, *ACM Trans. graphics (TOG)* 39 (4) (2020), 75–1.
- [249] W. Zhang, Y. Liu, C. Dong, Y. Qiao, Ranksrgan: Generative adversarial networks with ranker for image super-resolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 3096–3105.
- [250] R. Zhou, S. Susstrunk, Kernel modeling super-resolution on real low-resolution images, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 2433–2443.
- [251] A. Abuolaim, M.S. Brown, Defocus deblurring using dual-pixel data, in: Proc. IEEE Eur. Conf. Comput. Vis. (ECCV), Springer, 2020, pp. 111–126.
- [252] J. Dong, S. Roth, B. Schiele, Deep wiener deconvolution: Wiener meets deep learning for image deblurring, *Proc. Adv. Neural Inf. Process. Syst. (NIPS)* 33 (2020).
- [253] A. Shocher, N. Cohen, M. Irani, 'zero-shot' super-resolution using deep internal learning, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 3118–3126.
- [254] J.W. Soh, S. Cho, N.I. Cho, Meta-transfer learning for zero-shot super-resolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 3516–3525.
- [255] S. Zhou, J. Zhang, W. Zuo, C.C. Loy, Cross-scale internal graph neural network for image super-resolution, arXiv preprint arXiv:2006.16673 (2020).
- [256] Y. Xie, Z. Wang, S. Ji, Noise2same: Optimizing a self-supervised bound for image denoising, arXiv preprint arXiv:2010.11971 (2020).
- [257] J. Su, H. Yin, Efficient multi-objective gans for image restoration, in: Proc. IEEE Int. Conf. Acoust. Speech Signal (ICASSP), 2021.
- [258] M. Bevilacqua, A. Roumy, C. Guillemot, M.L. Alberi-Morel, Low-complexity single-image super-resolution based on nonnegative neighbor embedding, in: British Mach. Vis. Conf., 2012, pp. 1–10.
- [259] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Vol. 2, 2001, pp. 416–423.
- [260] E. Agustsson, R. Timofte, Ntire 2017 challenge on single image super-resolution: Dataset and study, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, 2017, pp. 1122–1131.
- [261] X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li, F. Huang, Real-world super-resolution via kernel estimation and noise injection, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, 2020, pp. 466–467.
- [262] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. Van Den Hengel, Q. Shi, From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2319–2328.
- [263] Y. Yan, W. Ren, Y. Guo, R. Wang, X. Cao, Image deblurring via extreme channels prior, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 4003–4011.
- [264] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, arXiv preprint arXiv:1603.04647 (2016).
- [265] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proc. Int. Conf. Artif. Intell. Stat., 2010, pp. 249–256.
- [266] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

Jingwen Su received the B.Eng. degree in Electrical and Electronic Engineering from the University of Manchester, UK, in July 2017. Since Sept 2017, she has been pursuing her PhD study in deep learning and image processing in the Department of Electrical and Electronic Engineering, the University of Manchester. Her current research interests are computer vision, image processing, deblurring, super-resolution, and deep learning.



Boyan Xu received the B.Eng. degree from the School of Automation, Wuhan University of Science and Technology, China, in 2019. Since Sept 2019 he has been a research student at the Department of Electrical and Electronic Engineering, the University of Manchester, UK. His current research focuses on image restoration, inverse problems and deep learning.





Hujun Yin received the PhD degree in neural networks from University of York, UK and the BEng degree in electronic engineering and the MSc degree in signal processing from Southeast University, China. Since 1998, he has been with the Department of Electrical and Electronic Engineering, The University of Manchester, UK, where he is currently a Professor of Artificial Intelligence. His main research interests include neural networks, self-organizing systems, deep learning, image classification and hyperspectral image processing, face recognition, data analytics, time series modelling and prediction, and bio-/neuro-informatics. He has extended and developed a number of learning models and methods and published over 200 peer-reviewed articles in a wide range of topics in the field. He is a Senior

Member of the IEEE, a Turing Fellow, and a member of the UK EPSRC College. He has been an Associate Editor for the IEEE Transactions on Cybernetics since 2015 and had been a member of the Editorial Board of the International Journal of Neural Systems (2005–2020). He had also served as an Associate Editor for the IEEE Transactions on Neural Networks during 2006–2010.