Image Denoising in the Deep Learning Era

Saeed Izadi
Simon Fraser University
Darren Sutton
Simon Fraser University
Ghassan Hamarneh (✉ hamarneh@sfu.ca)
Simon Fraser University

Research Article

Keywords: Image Denoising, Deep Learning, convolution Neural Networks, Recurrent Neural Networks

Posted Date: July 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1806416/v1

Image Denoising in the Deep Learning Era

Saeed Izadi, Darren Sutton and Ghassan Hamarneh

School of Computing Science, Simon Fraser University, Burnaby, BC, Canada.

*Corresponding author(s). E-mail(s): hamarneh@sfu.ca; Contributing authors: saeedi@sfu.ca; darrens@sfu.ca;

Abstract

Over the last decade, the number of digital images captured per day witnessed a massive explosion. Nevertheless, the visual quality of photographs is often degraded by noise during image acquisition or transmission. With the re-emergence of deep neural networks, the performance of image denoising techniques has been substantially improved in recent years. The objective of this paper is to provide a comprehensive survey of recent advances in image denoising techniques based on deep neural networks. In doing so, we commence with a thorough description of the fundamental preliminaries of the image denoising problem followed by highlighting the benchmark datasets and the widely used metrics for objective assessments. Subsequently, we study the existing deep denoisers in the supervised and unsupervised categories and review the technical specifics of some representative methods within each category. Last but not least, we conclude the analysis by remarking on trends and challenges in the development of better state-of-the-art algorithms and future research.

Keywords: Image Denoising, Deep Learning, convolution Neural Networks, Recurrent Neural Networks

1 Introduction

The role of digital cameras is to approximate an image of the real world by sampling from a discrete grid while maintaining image quality as judged by human perception. The visual quality of images collected by handheld consumer cameras [1, 2], medical imaging equipment [3], or industrial cameras,

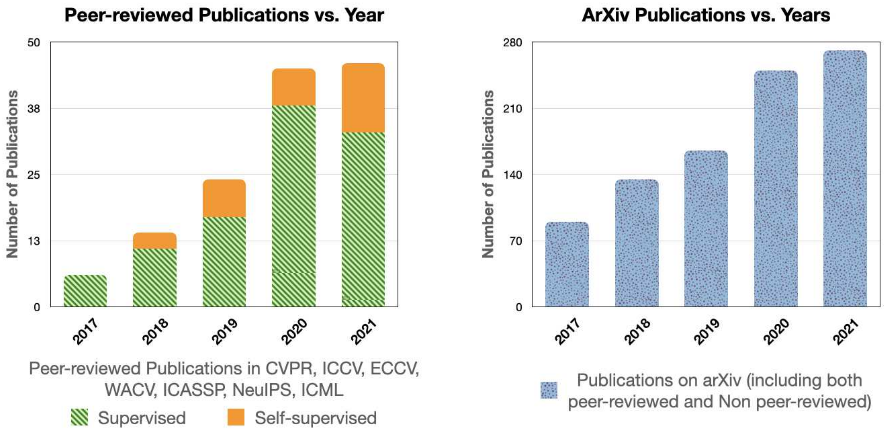

Fig. 1: Statistics of (a) peer-reviewed and (b) arXiv papers on image denoising over the past few years. (c) indicates the evolution history of image denoising algorithms in deep learning era.

may be impaired by several intrinsic or extrinsic factors related to the acquisition environment such as the pixel pitch of the sensor or the scene light level. Noise corruption caused by light interference, dark current leakage, shot noise and lens aberration can deteriorate the perceptual quality of images. Image noise can also impact subsequent higher-level computer vision tasks [4, 4], therefore it is often crucial to denoise images prior to any further higher-level image interpretations tasks.

Image denoising refers to the process of inspecting a noisy image and recovering an estimate of the underlying clean counterpart through discarding the noise artifacts. Traditionally, image denoising is framed as an optimization procedure searching for the most likely clean image and, given that more than one image-noise combinations can produce the noisy image, it falls into the family of ill-posed inverse problems [5, 6]. Traditional denoising techniques tackle this issue by imposing explicit regularization to constrain the search space. Some examples of such classic denoising methods leveraging *a priori* knowledge about the clean image include non-local self-similarity models [7, 8], sparse models [9], gradient models [10–12] and, Markov random field models [13, 14]. However, these methods typically suffer from two major deficiencies: 1) discovering clean images often involves a set of tuned hyper-parameters that do not generalize well to unseen data and, 2) all the computational steps are performed at the inference phase requiring a considerable amount of resources.

Driven by the availability of large datasets, rapid increases in computational power and, advances in algorithmic development for optimization of neural networks, deep learning has made impressive improvements in tackling many computer vision tasks [15–18]. In the context of image denoising, deep learning has attracted significant research interest and raised many new questions during the past recent years [2, 19–21] (Fig. 1). Employing neural

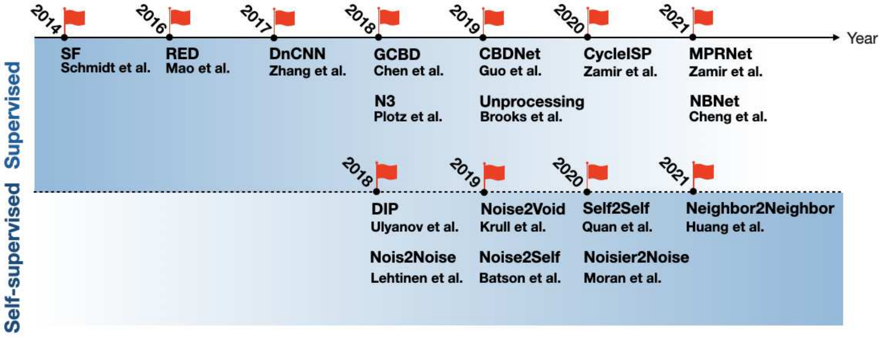

Fig. 2: Figure indicates the highly cited supervised and self-supervised works in recent years.

networks in image denoising can be traced back to the seminal works exploring the advantages of lightweight networks over classical human-engineered denoisers [22, 23]. The research question initially asked was whether neural networks could compete with engineered classical denoisers. As our understanding of neural networks has improved, deep denoising networks have become the de-facto choice for state-of-the-art denoising applications. The extensive use of neural networks for denoising has created a diverse set of approaches to choose from, ranging from convolutional networks [24] to generative adversarial frameworks [25].

The field of deep image denoising has developed rapidly but in a disparate manner. As depicted in Fig. 2, different denoising paradigms have been proposed during the past years, however, most of these methods are tailored to specific contexts and are based on benchmark datasets that are not directly comparable. Additionally, some new benchmark datasets have been proposed which are not included in the existing reviews [2, 19–21]. This motivates us to examine the recent advances in this active area to deliver an overview as well as new perspectives for interesting research directions. We provide a new taxonomy of the existing deep denoising techniques by grouping methods into two major categories: supervised and unsupervised approaches. In each category, we further organize the representative methods in accordance with their network design, adopted priors and training strategies. We present the methods in chronological order to show the advancement timeline for each training paradigm category.

To summarize, the main contributions of the survey are as follows:

1. We provide a thorough description of the preliminaries for image denoising as well as a comprehensive summary of the benchmark datasets and evaluation metrics.
2. We deliver an extensive overview of deep denoisers. We introduce a novel taxonomy of the existing methods in an effort to present a complete picture of the state of the art in deep denoising.

Table 1: Notations and abbreviations used in this report.

Notations	**Descriptions**
$\mathbb{R}$	1-dimensional Euclidean space
$a, \mathbf{a}, \mathbf{A}$	Scalar, vector, matrix
$\{\}$	Set of scalars
Φ	Degradation function
θ_η	Parameters of degradation function
Φ^{-1}	Restoration function
$\forall$	For all elements
θ_ζ	Parameters of restoration function
$\hat{\mathbf{A}}, \hat{a}$	Estimate of $\mathbf{A}$, a
a^*	Optimal value for a
$\mathcal{L}$	Generic Loss function
ρ	Regularization term
λ	Regularization coefficient
$N(\cdot)$	Normal distribution
μ	Mean
σ	Standard deviation (Noise strength)
$P(\cdot)$	Poisson distribution
α	Sensor-specific scaling factor
$\sim$	Random draw from a distribution
$\perp$	Statistical independence

3. By compiling the results of previous work, we discuss research challenges and open issues to identify new trends and future research directions for the denoising community.

This survey is organized as follows: Sec. 2 and Sec. 3 cover the problem definition and review the mainstream datasets and evaluation metrics. In Sec. 3, we investigate the representative works in the supervised denoising area. Sec. 5 delivers a summary of recent unsupervised denoising methods and, Sec. 6 provides a summary of denoising applications in other domains. We conclude this survey in Sec. 7 and Sec. 8 with a discussion on a number of open problems and future research directions. For better readability, we list the notations that will be used in this survey in Table 1.

2 Background

2.1 Problem Definition and Terminology

Formally, let $\mathbf{X} = \{x_i \in \mathbb{R}\}_{i=1}^n$ be a noisy image with n pixels that is corrupted by a degradation function Φ, and let $\mathbf{Y} = \{y_i \in \mathbb{R}\}_{i=1}^n$ be the corresponding clean counterpart. The degradation function Φ for the i-th pixel is written as:

$$x_i = \Phi(y_i;\ \theta_\eta); \quad \forall i \in \{1, 2, ..., n\} \tag{1}$$

where θ_η indicates the set of parameters associated with the degradation function and noise model. Degradation by noise is often modelled as noise addition followed by pixel-wise clipping to account for sensor saturation. Suppose that

η_i denotes the noise component for i-th pixel physically caused by light or camera. Therefore, the additive noise model can be written as:

$$x_i = \Phi(y_i, \theta_\eta) = \mathrm{clip}\left(y_i + \eta_i\right); \quad \forall i \in \{1, 2, ..., n\}, \tag{2}$$

without loss of generality assuming the pixel intensities to lie in the range $[0, 1]$, we have that $\mathrm{clip}\left(y_i\right) = \min\left(\max\left(y_i, 0\right), 1\right)$. The task of image denoising is to recover $\mathbf{Y}$ from the observed noisy data $\mathbf{X}$. Typically, the degradation function and the noise parameters are unknown. Thus, an approximation of the inverse function is learned such that:

$$\hat{y}_i = \Phi^{-1}(x_i;\ \theta_\zeta); \quad \forall i \in \{1, 2, ..., n\} \tag{3}$$

where Φ^{-1} and θ_ζ denote the denoising function and its parameters, respectively. The learning-based denoiser is implemented as a regression function that maps the noisy $\mathbf{X}$ inputs to the clean $\mathbf{Y}$ ground truth; i.e. $\Phi^{-1} : \mathbf{X} \mapsto \mathbf{Y}$. When training a neural network as a denoiser, the loss is typically composed of a fidelity term $\mathcal{L}(y_i, \hat{y}_i)$ measured between the clean estimate and the ground truth and, a regularization term $\rho(\hat{y}_i)$ to constraint the solution space adjusted with and a trade-off parameter λ. The denoising network is trained to learn an optimal parameter configuration:

$$\theta^* = \arg\min_\theta \mathcal{L}(y_i, \hat{y}_i) + \lambda\rho(\hat{y}_i); \quad \forall i \in \{1, 2, ..., n\}. \tag{4}$$

When choosing a fidelity term, the prior knowledge of the clean input may be an important consideration. A mean squared error (MSE) or $L2$ distance fidelity term tends to produce over-smoothed outputs, which may lack high-frequency details due to enforcing a Gaussian prior on the restored output. Some works have demonstrated the benefits of mean absolute error (MAE) or $L1$ distance to produce higher quality restored images with perceptually sharper edges and textures [26].

2.2 Noise Formation Model

Noise in digital images comes from many sources, such as variation in sensor sensitivity (ISO factor), thermal fluctuations, signal transmission errors, photon shot noise and, quantization noise. Noise models are approximations of the real noise created during signal conversion in the sensor and readout by an analog-to-digital converter. In this section, we elaborate on three types of most commonly studied noise models in digital imaging.

Shot Noise or Signal-dependent Noise. Photons are elementary particles travelling from the world to the camera during the exposure and arrive at the pixel sites in whole numbers, or packets. Scene irradiance is measured by the conversion of incident photons into charge at each pixel in a sensor array [27]. The packet-count varies proportionally to the square root of the count and therefore the photon count at the sensor array has an uncertainty that comes from random fluctuations in the arrival time of the photos. Such uncertainty

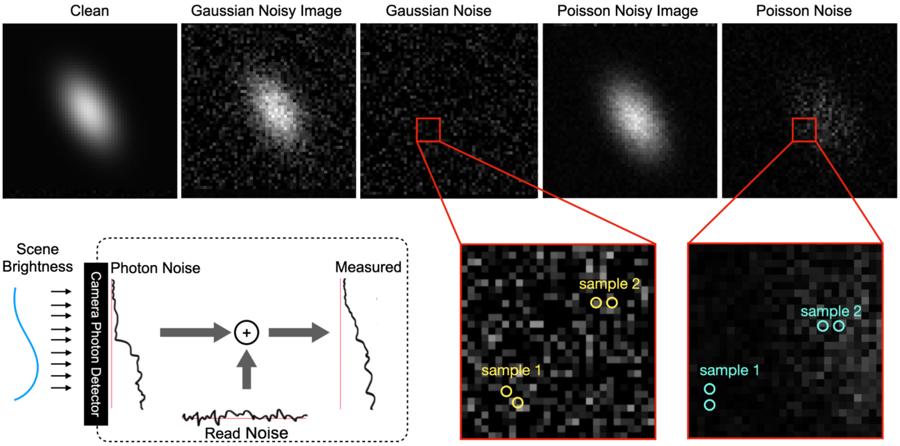

Fig. 3: Figure indicates the evolution history of image denoising algorithms in deep learning era.

is known as the shot or photon noise and is theoretically described by the Poisson distribution $\mathcal{P}(y_i)$,

$$\begin{gathered} x_i = y_i + \eta_i^p; \qquad \forall i \in \{1, 2, ..., n\} \\ \text{subject to} \quad \eta_i^p \sim \alpha \mathcal{P}(y_i) - y_i \end{gathered} \tag{5}$$

where the mean and variance of the noise at pixel location i equals the pixel intensity in the clean image y_i. The scalar coefficient α indicates the sensor-specific scaling factor of the signal.

Read Noise or Signal-independent Noise Photons accumulated at each cell during the exposure are readout as a charge or voltage that is eventually stored as a scalar pixel value. Read noise is the summation of the noise from random events during the photon to photo-electron accumulation and readout process, including lower-level noises such as thermal fluctuations, analogue-to-digital quantization noise, reset noise, and source follower noise [28, 29]. Different sensor types have different read noise characteristics. A CCD sensor typically has one read action for all pixels, thus the read noise is consistent among pixels but varies from image to image. A CMOS sensor has a read action for each pixel or column of pixels, thus there is variability in the read noise from pixel to pixel within a single image. Read noise is conservatively approximated using a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2,

$$\begin{gathered} x_i = y_i + \eta_i^g; \qquad \forall i \in \{1, 2, ..., n\} \\ \text{subject to} \quad \eta_i^g \sim \mathcal{N}(\mu, \sigma^2) \end{gathered} \tag{6}$$

with μ and σ being fixed everywhere within the spatial dimensions of the image. The read noise is often considered as white Gaussian noise when it is sampled from a zero-mean distribution. Read noise alone underestimates the actual real noise corruption occurring in real images as it only models the errors from reading the charge accumulated at each pixel and not noise from charge accumulation itself.

Poisson-Gaussian Noise. Digital imaging induces both signal-independent and signal-dependent errors and, a Gaussian or Poisson distribution alone may not be sufficient for precise noise modelling. To address such a limitation, real noise is often modelled using a combination of both Poisson and Gaussian components,

$$x_i = y_i + \eta_i^p + \eta_i^g; \quad \forall i \in \{1, 2, ..., n\}\,. \tag{7}$$

In practice, Poisson-Gaussian noise is modelled using a heteroscedastic Gaussian model. The parameters of a heteroscedastic Gaussian noise model change with respect to some quality of the signal. In image denoising, noise is often represented by a Gaussian distribution whose variance is proportional to the signal intensity, i.e.,

$$\begin{aligned} x_i &= y_i + \eta_i^{hg};\ ; \quad \forall i \in \{1, 2, ..., n\} \\ \text{subject to} \quad & \eta_i^{hg} \sim N(0, \alpha y_i + \sigma^2) \end{aligned} \tag{8}$$

The heteroscedastic Gaussian model is commonly referred to as the noise level function [30, 31].

3 Benchmarks

The recent use of deep neural networks has led to a consensus that datasets are of critical importance for a variety of computer vision and image processing applications. For image denoising, numerous publicly available datasets have emerged that greatly differ in image amounts, quality, resolution, diversity and, most importantly noise characteristics. In this section, we review some of the most widely used image denoising datasets. We group these datasets into synthetic noise and real noise categories and highlight their remarkable properties, such as image amounts, resolution, and acquisition settings. Table. 2 lists a summary of studied datasets.

3.1 Synthetic Noisy Datasets

A common strategy to train neural networks for image denoising is to consider the image datasets used for other computer vision tasks [32, 33] as a collection of clean images and simulate the noisy equivalents by imposing *i.i.d*[1] Gaussian, Poisson or Poisson-Gaussian random samples. Despite the popularity of synthetic noisy datasets, insufficient proofs are indicating that images borrowed from other datasets are genuinely clean. More importantly, the noise characteristics in the real noisy image do not always conform to those of the synthetic ones [30, 31] resulting in a significant performance discrepancy when networks

[1] independent and identically distributed

Table 2: Summary of datasets.

	Dataset	**# Images**	**Avg. Resolution**	**Content/Camera Info.**
Synthetic	BSD500 [32]	500	(481 × 321)	images of natural scenes with at least one discernible object
	DIV2K [33]	1000	(1972×1437)	covers a large diversity of contents including people, cities and natural scenes
	BSD68 [34]	68	(435×364)	part of BSD dataset largely used for image denoising
	Urban100 [35]	100	(984×797)	collected from Flicker containing urban, architectural and structured scenes
	Kodak24 [36]	24	(768×512)	uncompressed images published by Kodak Corporation including indoor, outdoor images
	Manga109 [37]	109	(826×1169)	includes images of Japanese comic books publicly available for academic research
	Set14 [38]	14	(492×446)	classic dataset introduced for low-level image processing tasks
	Set5 [39]	5	(313×336)	classic dataset introduced for low-level image processing tasks
Real	RNI15 [40]	15	(514×465)	covers variety of indoor and outdoor scenes, especially scans of some old photos
	RENOIR [41]	120	(4364×3115)	low-light noise, Canon PowerShot S90, Canon EOS Rebel T3i, and mobile Xiaomi Mi 3
	NAM [42]	11	(7360×4912)	Canon EOS-5D Mark III, Nikon D800, Nikon D600
	DND [2]	1000	(512×512)	50 scenes, Sony A7R, Olympus E-M10, Sony RX100 IV, Huawei Nexus 6P
	SIDD [1]	30,000	(4586×3035)	Apple iPhone 7, Google Pixel, Samsung Galaxy S6 Edge, Motorola Nexus 6, LG G4
	PolyU [43]	4000	(512×512)	Canon (Mark 5D, 80D, 600D), Nikon D800, Sony A7 II
	SID [44]	5094	(5078 × 3388)	Sony A7S II, Fujifilm X-T2
	NIND [45]	616	(3083 × 3864)	Canon C500D, FujifilmX-T

trained on synthetic are evaluated on real noisy images. Table. 2 summarizes the most widely synthetic datasets for training and evaluation of DL-based denoising models. Among them, we describe BSD and DIV2K in more detail below:

BSD. Berkeley Segmentation Dataset [32] is the most widely used dataset to render noisy and clean pairs via noise synthesis strategy. BSD is a collection of natural images with human-labelled segmentation ground truths consisting of 500 natural RGB images of size 481×321 with at least one discernible object. With today's standards, BSD however contains fairly low-resolution images which makes it less useful for real-world applications.

DIV2K. Recently, Luc Van Gool et al. [33] introduced a larger dataset primarily used as a benchmark for image super-resolution. In contrast to the BSD dataset, DIV2K contains images of higher resolution (2K) and larger content diversity. To fairly benchmark competing methods, 1000 images in DIV2K dataset have been partitioned into the subsets of size 800, 100 and, 100 for training, validation and test, respectively,

3.2 Real Noisy Datasets

There have been efforts to pair clean images with their real noisy equivalents to assist the denoising development in real-world applications. These pairs can be captured by constraining the extrinsic variables in the imaging environment or adjusting the intrinsic parameters of the imaging aperture. A prevalent strategy to approximate the clean ground truth is to offset the inherent noise by collecting a rapid sequence of shots from the fixed scene followed by temporal averaging. Another strategy is to consider the image taken with lower ISO factors and slower shutter as clean ground truths. In either setting, precise post-processing steps and image manipulations might be exploited to marginalize the noise in the clean ground truths even more.

RNI15. Lebrun et al [40] provided the first collection of real noisy images containing 15 images without clean counterparts. The images in RNI15 cover a variety of noise types including low-light images from smartphones, old photographs, aerial images, etc. Due to the absence of clean ground truths, RNI15 is merely used for qualitative evaluation purposes.

RENOIR. Anaya et al. [41] presented the first dataset containing both noisy and clean images. In RENOIR, images of 120 scenes are captured with low and high ISO settings. For each scene, two clean images are taken interleaved with one or two noisy ones in between. Multiple clean shots are used to secure the spatial alignment of images within the entire acquisition process. Finally, the low ISO images are averaged and paired with either or both of the noisy images. Two consumer cameras (Canon Rebel T3i, Canon S90) and a smartphone (Xiaomi T3i) are used to collect images at various ISO levels ranging from 100 to 6400. The RENOIR does not model heteroscedastic noise and, low-frequency bias is not removed.

NAM. Nam et al. [42] collected a laboratory-controlled dataset from 11 static scenes with printed pictures and few real objects. For each scene, 500 successive JPEG images were captured and used to approximate the (nearly) clean ground truth. Images are taken by three consumer cameras (Nikon D800, Nikon D600 and, Canon 5D Mark III) across three ISO factors (1600, 3200 and, 6400). A major drawback of the NAM is its use of printed pictures that deviate from the scenes in the real world. Additionally, NAM lacks modelling the heteroscedastic noise and low-frequency bias repair. Lastly, the images with misalignment or different illumination are not discarded in NAM.

DND. The Darmstadt noise dataset [2] consists of 50 scenes taken by 4 consumer cameras (Sony A7R, Olympus E-M10, Sony RX100 IV and, Huawei Nexus 6P) across different ISO ranges and shutter speeds. Image with high ISO (short exposure time) and low ISO (long exposure time) are taken as real noisy and clean images, respectively. Additional post-processing including correction of spatial misalignment and removing low-frequency bias are further adopted to derive more accurate clean ground truths for low ISO images. Moreover, the employed intensity transform is based on a heteroscedastic Tobit regression model.

SID. Chen et al. [44] introduced See-in–Dark (SID) dataset consisting of 5094 raw pairs captured with fast shutter (low exposure) and slow shutter (long exposure) using two cameras (Sony α7S II and Fujifilm X-T2). The dataset contains both indoor and outdoor images where the latter are captured at night under moonlight or street lighting.

SSID. This dataset [1] is collected from 10 scenes using five smartphones (Apple iPhone 7, Google Pixel, Samsung Galaxy S6 Edge, Motorola Nexus 6 and, LG G4) with fifteen ISO levels (50-10,000) under three illumination temperatures (3200K for tungsten or halogen, 4400K for fluorescent lamps and, 5500K for daylight) and three light brightness levels (low, normal and, high). Each scene is captured multiple times with different cameras settings and/or different lighting conditions rendering more than 30,000 images. The collected noisy images are then processed by a systematic procedure to obtain the clean ground truth. The main focus of SSID is to address the problem of noticeable noise caused by small sensor sizes in small apertures.

PolyU. Xu et al. [43] introduced a more comprehensive dataset taken from 40 versatile scenes in different lighting conditions using five cameras (Canon 5D Mark II, Canon 80D, Canon 600D, Nikon D800 and, Sony A7 II). To include more camera settings, each image is captured with 6 difference ISO factors (800, 1,600, 3,200, 6,400, 12,800 and, 25,600). Moreover, other intrinsic camera parameters such as shutter speed, aperture and, luminance are re-adjusted for each ISO to render all images normally exposed. Each scene is captured 500 $\sim$ 1000 times and the ones with spatial misalignment and luminance discrepancy are removed. Next, multiple samples of the same scene are averaged and taken as the clean ground truth. Since the image pairs are subjectively monitored, spatial misalignment is almost avoided. PolyU contains both raw s-RGB and JPEG images.

NIND. Most recently, NIND [45] was rendered from 101 scenes using two cameras (FujifilmX-T1, Canon C500D). Each scene is captured with a set of different ISO factors starting from 100 up to the highest possible value. The image with the lowest ISO is taken as the clean ground truth. Also, images with the highest ISO tend to be quite dark and therefore are correctly exposed using the software. As the ISO increases, the shutter speed decreases to match the original exposure value. On average, six images are captured for each scene rendering a dataset of a total size of 616 paired images.

3.3 Evaluation Metrics

In this section, we provide a summary of two of the well-known metrics used in evaluating the performance of denoising methods. Although the majority of the existing works use quantitative metrics for comparisons, the visual quality of denoised images is also important in deciding the best models as a human is often the end consumer of denoised images.

Peak-Signal-To-Noise Ratio. Peak-signal-to-noise ratio (PSNR), measured in decibels (dB), is the most prevailing criterion to quantify the degradation derived from losses in image transformations (e.g. compression, transmission, or reconstruction). Due to its low complexity and high simplicity, it is widely used and compared with. Given two images $\mathbf{X} = \{x_i \in \mathbb{R}\}_{i=1}^{n}$ and, $\mathbf{Y} = \{y_i \in \mathbb{R}\}_{i=1}^{n}$, PSNR is calculated as follows:

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{\mathrm{MAX}_X^2}{\mathrm{MSE}} \right) \tag{9}$$

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} \|x_i - y_i\|_2^2$$

where MAX_X is the maximum value in the dynamic range of the images. In the case of image reconstruction, higher PSNR values indicate the reconstruction of better quality, however, in some cases, it may not since it poorly correlates with the perceived quality by human eyes [46, 47].

Structural Similarity. Wang et al. [46, 48–50] proposed structural similarity (SSIM) as a more intelligent image quality assessment metric that is better linked to how human perceive the visual quality of images. SSIM measures the visual impact of changes in the image luminance, contrasts, spatial dependencies and, collectively structural information in the viewing field [49]. Given two images $\mathbf{X} = \{x_i \in \mathbb{R}\}_{i=1}^{n}$ and, $\mathbf{Y} = \{y_i \in \mathbb{R}\}_{i=1}^{n}$, SSIM is computed as follows:

$$\mathrm{SSIM} = [l_{\mathbf{X},\mathbf{Y}}]^a \, [s_{\mathbf{X},\mathbf{Y}}]^b \, [s_{\mathbf{X},\mathbf{Y}}]^c \tag{10}$$

where $a > 0, b > 0, c > 0$ control the relative significance of each terms. The luminance, contrast and, structural components are computed defined as follows:

$$l_{\mathbf{X},\mathbf{Y}} = \frac{2\mu_{\mathbf{X}}\mu_{\mathbf{Y}} + \epsilon_1}{\mu_{\mathbf{X}}^2 + \mu_{\mathbf{Y}}^2 + \epsilon_1} \tag{11}$$

$$c_{\mathbf{X},\mathbf{Y}} = \frac{2\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}} + \epsilon_2}{\sigma_{\mathbf{X}}^2 + \sigma_{\mathbf{Y}}^2 + \epsilon_2} \tag{12}$$

$$s_{\mathbf{X},\mathbf{Y}} = \frac{\sigma_{\mathbf{X},\mathbf{Y}} + \epsilon_3}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}} + \epsilon_3} \tag{13}$$

where $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$ denotes the mean, $\sigma_{\mathbf{Y}}$ and $\sigma_{\mathbf{Y}}$ represent the standard deviation and, $\sigma_{\mathbf{X},\mathbf{Y}}$ refers to the covariance of $\mathbf{X}$ and $\mathbf{Y}$. Also, ϵ_1, ϵ_2, and ϵ_3 are constants introduced to avoid instabilities when denominators are close to zero.

4 Supervised Denoising

Supervised image denoising implies using both the noisy and the clean images while training neural networks. On the other hand, optimizing the parameters in neural networks inevitably grows the need for accessing massive datasets with accurate clean ground truths for supervision. Owing to the prevalence of synthetic noisy datasets in recent years, many deep denoisers based on supervised training schemes have been dominantly presented in the literature [24, 51–53]. Apart from the synthetic datasets, the advent of large denoising datasets with real noisy and clean image pairs has contributed significantly to the success of supervised denoisers for real-world denoising problems [1, 2, 44]. In this section, we summarize the existing methods for supervised image denoising. We study the literature in two major directions, i.e. discriminative and generative approaches.

4.1 Discriminative Models

Discriminative methods have recently become increasingly prevailing for image denoising, thanks to their trade-off between denoising quality and speed at test time. In the scope of deep denoisers, the discriminative models exploit the capacity of neural networks to learn a direct mapping from noisy images to clean counterparts. In particular, these methods attempt to find the optimal parameters of a feed-forward network that maximize the conditional probability $P(y_i \mid x_i)$ directly from the training set $\mathcal{D}_{train}$. Mathematically, it can be written as:

$$\theta^* = \arg\max_{\theta} P(y_i \mid x_i); \qquad \forall i \in \{1, 2, ..., n\} \tag{14}$$

The success of discriminative deep denoisers in fast inference is attributed to fact that the learned parameters are kept fixed during the testing implying fixed computational cost for each image. However, this comes at the expense of less flexibility and the necessity of training distinct networks for different noise levels. The differences between the approaches in this line of work are mainly related to the network design, learning strategies and, modelling prior information. In the remaining parts of this section, we collect and summarize some representative works in this specific category.

4.1.1 Plain Networks

Plain feed-forward neural networks are known as the simplest DL-based models for image denoising, yet they have achieved superior performance against classic approaches such as BM3D [8] and WNNM [54]. In a nutshell, these networks are formed by assembling alternating sequences of convolutional or fully-connected layers, potentially interleaved with non-linear activations [55, 56], normalization [57, 58] and, dropout operations [59, 60]. Leveraging neural networks for image denoising arguably began gathering momentum in 2008 when Jain et al. [23] proposed to exploit the convolutional layers as a way to relax the computational expenses associated with the parameter estimation and inference popular probabilistic denoising methods. With the achievements of sparse coding models in image processing, Xie et al. [61] proposed a stacked sparse denoising auto-encoder (SSDA) framework by sequentially stacking multiple instances of the denoising auto-encoders connecting the noisy input to the network output. In addition to the reconstruction term, an auxiliary KL-divergence term was employed to ensure the sparsity of the mean of intermediate activations. Later, Agostinelli et al. agostinelli2013adaptive extended the previous SSDA framework by introducing adaptive multi-column SSDA to improve its robustness against various noise types. Coming next, Burger et al. [22] showed that a well-trained multi-layer perceptron (MLP) network over a massive collection of noisy and clean patches could outperform the well-known BM3D [8] method.

Some of the proposed methods for image denoising are based on unrolling the inference procedure in model-based techniques where the computational steps are modelled by neural layers. In this group, Schmidt et al. [62] borrowed the notion of shrinkage functions from wavelet restoration domain [63] and proposed a cascaded set of shrinkage fields (CSF) to model stage-wise predictions in an unrolled half-quadratic optimization procedure. Shrinkage functions in CSF were learned in a data-driven manner reducing the optimization procedure in each stage into a single quadratic minimization. The run-time speed was improved by leveraging convolution operation and discrete Fourier transform. Another example in this group is TNRD by Chen et al. [64, 65] which exploited advances in partial differential equations for image restoration. TRND designed a flexible denoising framework in which each stage was modelled by a convolutional layer with large trainable kernels optimized over a large dataset. Kim et al. [66] proposed to learn data-driven plain feedforward networks as the implicit regularizer in the widely adopted alternating minimization algorithms [67] for image restoration.

The major body of previous denoisers based on plain deep networks focused on noise phenomenon with spatially-fixed statistics. The work proposed by Zhang et al. [68] was among the earliest attempts to take into account the spatially-varying noise by adapting the performance of a plain deep CNN for different regions. Particularly, they proposed to augment the noisy image with a noise level map prior to feeding it to the network. Noise level maps were generated by stretching either the actual or estimated noise variance across the spatial dimensions to match the input size.

4.1.2 Residual Networks

Plain networks with deep architectures suffer from the potential risk of struggling with vanishing or exploding gradients. Therefore, assisting techniques, such as skip connections [69–71] are often utilized to facilitate unhindered information flow within the layers of the network. The image denoising literature has witnessed a significant use of residual learning in recent years [24, 51, 52,

72–79]. An early attempt to leverage the residual connections for image denoising was described in REDNet, by Mao et al. [80], where skip connections were used between corresponding layers in a mirrored encode-decoder architecture.
Residue Learning. Instead of the absolute clean images. some denoising methods leverage long skip connections in their design to learn the residue between the noisy and clean images. Notably, DnCNN by Zhang [51] introduced the earliest attempt in this avenue and exhibited superior performance with simpler architectures. The residue image produced in the output of the DnCNN is subsequently subtracted from the input noisy image to acquire the clean estimate. In other words, instead of learning a sophisticated mapping from a complete image to another, DnCNN learns the residue image that discards the noise part of the image and recovers the high-frequency details. Mathematically, residue learning based models can be written as:

$$r_i = \Phi^{-1}(x_i;\ \theta_\zeta); \qquad \forall i \in \{1, 2, ..., n\} \tag{15}$$

where r_i denotes the output residue. The clean estimate is computed as:

$$\hat{y_i} = r_i + x_i; \qquad \forall i \in \{1, 2, ..., n\} \tag{16}$$

DnCNN has been successfully employed in many model-based denoising algorithms serving as an implicit natural image prior [81, 82]. To complement DnCNN, Remez et al. [83] proposed CADN that reflected the direct impact of all intermediate layers in estimating the residue image.
Other Improvements. Some researchers have adopted more sophisticated patterns for skip connections [84] to improve the representation power of the networks. Tai et al. [24] designed MemNet by incorporating densely connected memory blocks between a low-level feature extractor and a reconstruction block. The memory blocks encompass end in 1 × 1 gating convolutions that adaptively control how much of the information received from the previous layers need to be preserved or discarded before delivering them to the subsequent module. Zhang et al [52] proposed a residual dense network intending to make full use of the hierarchical features with densely connected global memory blocks, which are themselves formed by a sequence of densely-connected convolutions. Most recently, Liu et al. [74] proposed dual residual building blocks to enhance the interaction between paired operations, e.g. down-sampling and up-sampling, occurring within the network.

4.1.3 Attention Mechanism

Attention mechanism has become an integral part of the neural networks in recent years [85, 86]. In neural networks equipped with attention mechanisms, the relationships among learned features are explicitly analyzed and exploited to help more efficient representation learning [87, 88]. In image denoising tasks, many works have tried to exploit the principal merits of attention mechanisms to achieve better denoising performance with faster training and smaller model size [77, 89–97]. Anwar et al.[89] proposed the first work that benefited from the emerging popularity of channel-wise attention mechanism. [87] for image denoising to improve the learning efficiency of the network by re-scaling the feature channels in accordance to their mutual dependencies. Later, Cheng et al. [93] proposed a novel subspace attention module in which the noisy images are projected into a learned clean subspace such that the reconstructed image can keep most of the original content and remove the noise, i.e. the irrelevant information to the generated basis vectors. Hu et al. [94] designed an efficient 3D auto-correlation that can extract vertical, horizontal and channel-wise axes simultaneously. In contrast to regular auto-correlation attention modules, this lightweight pseudo 3D module can avoid dense connection and high dimension operations.

A different class of attention mechanisms relate the features from different scales or operations. Specifically, Gu et al.[90] described a technique that connects the contextual features extracted at different resolutions to each other in a top-down processing architecture. The input image is initially down-sampled into multiple scales using the shuffling operation. Then, a hierarchical coarse-to-fine structure gradually receives and manipulates individual resolutions of the inputs. After several convolutions, the features from the coarser scale are delivered into the subsequent first-level coarser scale as a way to transfer the cross-scale contextual information within the entire multi-scale framework. Different from the previous work that aggregates multi-scale contents in a top-down manner, Zamir et al. [95] proposed a model that aggregates the contextual information among multi-scales through exchanging the information across all scales at each resolution level. Moreover, the delivered information from other resolution levels is adaptively gated and fused to the current information by a self-attention mechanism. Similar to [95], Zamir et al. [97] proposed to incorporate a supervised attention module between every two stages in a multi-stage architecture. Further, they introduced a cross-stage information exchange module to improve the feature fusion between early stages and later ones.

Most recently, Suganuma et al. [96] presented a more versatile layer architecture that embodies multiple operations such as convolutions with different kernel sizes applied on the input. In such a setting, the attention mechanism intends to produce a weight vector to determine the impact of each operation within the layer. The weight vector is then multiplied with the outputs of the operations to re-scale them in accordance with their significance.

4.1.4 Nonlinear Activation Functions

Increasing the depth of the network architecture for better learning capacity is not always doable due to the limited computational resources in many practical applications. To address this, more focus has been put on the role of activation functions in constructing efficient yet powerful networks [98–100]. Toward efficient image denoising, there have been various recent improvements that focus on ameliorating the activation functions [101, 102].

As opposed to the ubiquitous RELU [55] activation that operates per pixel, Kligvasser et al. [101] incorporated the notion of learnable activations with spatial connections into deep denoisers. RELU [55] activation can be explained as a hard-gating mechanism where irrelevant activations are discarded by binary weight map. Conversely, xUnit offers a soft-gating scheme through adopting internal convolutions and Gaussian gating modules to provide spatially-dependent continuous-valued weight maps for activations. Seemingly, this method requires more computational demands however the expanded representational power of the layers allows achieving the performance of deep networks with a smaller number of layers.

In the same vein, Gu et al. [102] crafted another learnable activation function (MTLU) that assists in boosting the learning capacity of small networks. As depicted in Eq. 17, the core methodology of MTLU has two highlights: a) dividing the activation space into several equidistant bins and, b) learning the coefficients for different linear functions per bin using the back-propagation strategy during the training.

$$f(x) = \begin{cases} a_0 x + b_0, & \text{if } x \leq c_0 \\ a_k x + b_k, & \text{if } c_{k-1} < x \leq c_k \\ \cdots & \\ a_K x + b_K, & \text{if } c_{K-1} < x \end{cases} \tag{17}$$

where the set $\{c_k\}_{k=1}^{K}$ is K hyper-parameters for MTLU and, $\{a_k\}_{k=1}^{K}$ and $\{b_k\}_{k=1}^{K}$ are the coefficient for linear functions.

4.1.5 Non-Local Similarity

Many classic image denoising methods have demonstrated the merits of self-similarity (NSS) prior on natural images for image restoration [7, 8]. Concretely, the NSS prior states that similar image patches tend to re-occur within the image in non-local regions. While the NSS has been broadly explored in the classic denoisers, a few works have attempted to incorporate this internal image property into deep networks for image denoising [53, 103–111]. Among them, we distinguish two major categories depending on the way the non-local information come to the play, i.e. non-local retrieval and implicit non-local attention.

Non-Local Retrieval. Motivated by BM3D [8] and non-local means [7], the intent of the works in this category is to explicitly find and retrieve the most similar patches to a query patch and, utilize them in subsequent stages to discard the noise component. The earliest deep denoiser exploiting non-local prior was the NLNet proposed by Lefkimmiatis et al. [103]. NLNet is a patch-based proximal gradient method unrolled into multiple stages. Each stage is efficiently modelled by a sequence of convolutions to linearly transform every patch, a block-matching to collect similar patches and, ultimately a collaborative filtering that projects all patches into a single patch representing the clean estimate. Later Xia et al. [104] proposed a patch denoising framework in which the network takes in an individual noisy patch and a set of most similar

Table 3: Summary of some representative methods in supervised denoising. The "S", "R" denotes synthetic and real noise respectively. Also, "D", "F", "LSC", and "SSC" represent effective depth, max filter size, long skip connection, and short skip connection, respectively.

Ref	Publication	Data		Architecture				Loss	keywords
		S.	R.	D.	F.	LSC	SSC		
[65]	2015, CVPR	✓	×	8	48	×	×	L2	Plain CNN
[23]	2008, NeurIPS	✓	×	4	24	×	×	L2	Plain CNN
[62]	2014, CVPR	✓	×	5	48	×	×	L2	Plain CNN
[66]	2017, CVPR	✓	×	10	64	×	×	L1	Plain CNN
[68]	2018, TIP	✓	×	15	64	×	×	L2	Plain CNN
[82]	2017, CVPR	✓	×	17	64	✓	×	L2	Residual Learning
[83]	2018, TIP	✓	✓.	20	63	✓	×	L2	Residual Learning
[80]	2016, NeurIPS	✓	×	30	64	✓	×	L2	Residual Learning
[52]	2021, PAMI	✓	×	120	64	✓	✓	L1	Dense Residual
[74]	2019, CVPR	✓	×	24	32	✓	✓	L2	Dual Residual
[51]	2017, TIP	✓	×	17	64	✓	×	L2	Residual Learning
[24]	2017, ICCV	✓	×	80	64	✓	✓	L2	Recursive Learning
[89]	2019, ICCV	✓	✓	40	64	✓	✓	L1	Attention, Residual
[90]	2019, CVPR	✓	✓	16	32	×	✓	L1	Attention
[93]	2021, CVPR	✓	✓	16	128	✓	✓	L1	Non-Local Attention
[95]	2020, ECCV	✓	✓	58	256	✓	✓	Ch.	Residual, Attention
[96]	2019, CVPR	✓	×	89	16	×	✓	L1	Multi-scale Attention
[101]	2018, CVPR	✓	×	17	64	✓	×	L2	Activation, Attention
[102]	2018, CVPR	✓	×	10	64	✓	×	L2	Activation
[104]	2020, WACV	✓	×	14	96	✓	×	L2	Non-Local
[105]	2018, NeurIPS	✓	✓	30	64	×	×	L2	Non-Local
[106]	2019, ICLR	✓	×	120	64	×	×	L2	Non-Local
[53]	2018, NeurIPS	✓	×	38	128	×	×	L2	Non-Local
[108]	2020, TIP	✓	×	30	512	✓	×	L1	Dynamic Conv
[110]	2020, ECCV	✓	×	30	512	✓	×	L1	Dynamic Conv
[113]	2020, TIP	✓	×	18	132	×	×	L2	graph-convolutional
[114]	2019, CVPR	×	✓	32	512	✓	×	L1	RAW
[115]	2019, CVPR	×	✓	16	256	✓	✓	L2, A., TV	RAW
[116]	2020, CVPR	×	✓	16	512	×	✓	L1	RAW
[117]	2020, ECCV	×	✓	28	512	✓	✓	L1	RAW
[118]	2020, CVPR	×	✓	23	256	✓	✓	L1	RAW
[119]	2021, CVPR	×	✓	32	N/A	×	✓	L1, L2	Invertible
[120]	2020, CVPR	✓	✓	110	256	✓	✓	L1, A.	Transfer Learning
[121]	2020, CVPR	×	✓	40	N/A	✓	✓	L1	RAW, Attention, Residual
[122]	2020, ECCV	×	✓	27	512	✓	✓	L1	Joint Distribution
[123]	2018, PAMI	✓	✓	35	32	✓	✓	L2	Boosting, Residual
[124]	2019, TIP	✓	×	67	64	✓	✓	L1	Boosting, Residual
[125]	2019, NeurIPS	✓	✓	17	512	✓	✓	L2	Variational
[25]	2018, CVPR	×	✓	17	64	✓	×	L2	Noise Modeling, GAN
[126]	2020, ECCV	×	✓	17	64	✓	×	L2, , Adv., FM., Tri.	Noise Modeling
[94]	2021, CVPR	✓	✓	40	64	✓	✓	L1	Attention, Residual, Non-Local
[97]	2021, CVPR	×	✓	140	80	✓	✓	L1	Multi-scale Attention
[127]	2021, ICCV	×	✓	65	256	✓	✓	L2, NLF	Non-Local, Graph Network

patches and, outputs a vector of matching scores. The denoised patch is then obtained by averaging across candidates using the matching scores. In contrast to the normal convolutions which have rigid sampling grid and kernel weights, Xu et al. [108] proposed to explicitly learn the sampling locations along with the kernel weights in a data-driven manner. Thus, the network is able to adaptively sample from the 2D input space to freely expand the respective field. Chang et al. [110] not also adopted deformable convolutions [112], but also inserted the modulated deformable convolution in their proposed network to sample the spatially relevant features for weighting.

Non-Local Attention. Most existing denoising methods suffer from having a small receptive field due to local convolutions. However, long-range similarities may be used for denoising the patches. Wang et al. [128] embedded the concept of non-local mean in the neural networks and proposed a non-local neural network leading to a considerable boost in many computer vision applications. Zhang et al.[106] adopted this work and proposed a residual trunk-and-mask [129] architecture for the task of image denoising. The trunk branch provides the intermediate features whereas the mask branch calibrates the feature based on the non-local correspondences in the spatial domain. Another novel technique is N^{3}Net [105], which proposed a continuous deterministic relaxation for the non-differentiability of *KNN* selection rule. It is then used within the internal layers of the network to concatenate every feature vector with a weighted average of the most similar feature in the 2D space of the intermediate representations. Liu et al. [53] integrated the non-local mean operation into the recurrent neural networks and, proposed to perform non-local matching in a confined region centered at query position rather than considering the entire spatial scope for matching.

Graph Neural Networks. Valsesia et al. [113, 130] proposed to exploit graph convolutions to cope with the limited receptive field in traditional convolutional layers. To be concrete, they generalized the traditional convolution layers by creating adaptive receptive fields based on nearest-neighbour graphs. During the training, distant but similar features are aggregated to leverage the non-local similarities. An additional module estimates the aggregation weights to further increase learning adaptability. Li et al. [127] designed a cross-patch graph convolutional network to explicitly cross-patch long-range contextual dependency. For every patch, their proposed network aggregates similar patches to the primary input patch, and ensembles the extracted features toward a more accurate clean patch estimate. Mou et al. [131] extended the patch-based graph convolutional networks and proposed a dynamic attentive graph in which the query patch can have a dynamic and adaptive number of neighbors.

4.1.6 Raw Denoising

Until recently, most of the deep denoisers have been leveraging pairs of simulated noisy and clean datasets for training and, this ended up with dramatic performance discrepancy once assessed on real noisy images. To narrow down this gap, recent works in image denoising attempt to perform training and validation on raw real noisy datasets in explicit [115–122, 132–136].

RAW Noise Synthesis As described in section 2.2, in-camera processing pipeline, a.k.a image signal processor (ISP) affects the nature of the noise and therefore noise can come from different sources in the real camera system. Guo et al. [115] proposed a noise model that takes into account both heteroscedastic Gaussian noise as well as demosaicing, Gamma correction and JPEG compression. The new noise model is used to simulate a large set of noisy images that resembles real-world noise. Brooks et al. [114] incorporated more ISP components in the noise modelling. Particularly, they showed that a generic clean image can be *unprocessed* into real RAW data by successively applying inverse tone mapping, gamma decompression, sRGB-to-RGB correction and, inverse white balance & digital gain. The heteroscedastic noise is then added to the RAW data to stimulate real noisy and clean pairs for Raw-to-Raw training. Similarly, Zamir et al. [121] proposed a cycle framework for learning RGB-to-Raw and Raw-to-RGB mappings through two distinct network branches.

Other Improvements Motivated by the instance normalization module [58], Kim et al. [120] adopted an adaptive instance normalization and a transfer learning scheme to reduce the domain discrepancy between the synthetic and real noise data. After training the network on synthetic noisy datasets, only the adaptive instance normalization layers are fine-tuned on the real noisy data to bridge the distribution gap. Wang et al. [117] also proposed a lightweight model for on-device denoising of real images. The core of their methodology is adopting a novel K-Sigma transform that projects noisy images captured across different ISO settings into an ISO-invariant space. This way, a single network is capable of processing images with different noise characteristics. Liu et al. [119] leveraged å invertible network for image denoising. To mitigate the issue with different distribution for input and output pairs, Liu et al. [119] proposed to transform the noisy input into a low-resolution clean image and a latent encoding for the noise. The noise component can be discarded by replacing its corresponding encoding with a sampled representation from a prior distribution.

4.1.7 Boosting

Boosting algorithms is one of the widely used techniques for improving the performance of machine learning algorithms [137, 138]. In the image denoising field, Chen et al. [123, 139] incorporated data-driven DL-based denoising networks as the base units in a cascaded boosting framework. Inspired by the Strengthen-Operate-Subtract (SOS) [138], in each boosting unit, the summation of the denoised image and the noisy input is fed into the subsequent denoising module. Next, the identical denoised image is subtracted in each step to ensure the iterability of SOS. A cascaded boosting configuration leads to a very deep architecture. To cope with this challenge, they leveraged a set of lightweight structures equipped with dense residual connections and dilated kernels in base units to reduce the overall computational burden of the network. In contrast to the cascaded boosting, Choi et al. [124] developed a convex optimization procedure to optimally aggregate the outputs of multiple denoising units (CsNet). Specifically, they solve a quadratic minimization problem to find the optimal weights for combining the complementary outcomes of different denoising networks.

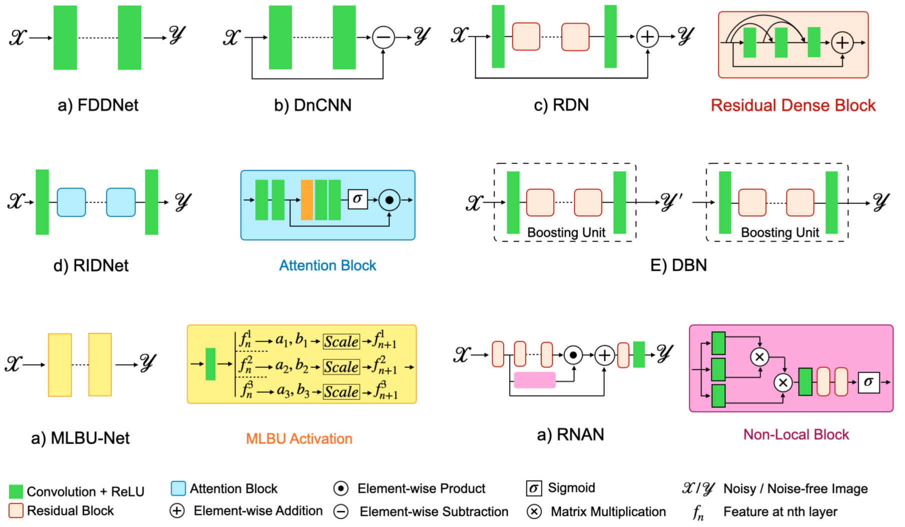

Fig. 4: A glimpse of the some methods representing different network architectures for supervised image denoising.

4.2 Generative Image Denoising

In contrast to the discriminative models that compute $P(y_i \mid z_i)$, a generative model often captures the generation process of the observed noisy example by modelling $P(z_i \mid y_i)$. In other words, the discriminative models are mostly focused on separating the underlying clean image, while the generative models try to understand the basics of noisy image formation. In the following, we will elaborate on a few representative denoisers based on the generative models. We review the existing works related to the generative models from two perspectives, methods based on variational inference and, generative adversarial networks.

4.2.1 Variational Inference

In an attempt to discern the generation process of the noisy observations, Yue et al. [125] proposed a novel variational inference framework that performs both noise estimation and noise removal in a Bayesian framework. To be concrete, their proposed framework learns an approximate posterior of the clean and noise statistics conditioned on the observed noisy image. By replacing the approximate posterior with a variational distribution, they take advantage of the independence property of variational latent variables and represent the variational distribution in form of two distinct functions. Consequently, the variational approximation of the clean image is modelled by a conjugate Gaussian prior with mean and covariance parameters. For noise estimation, the inverse Gamma distribution is taken as the conjugate prior. Two distinct deep networks were trained to learn the mapping between the noisy image and the parameters of the variational posteriors. A second important generative denoiser was proposed by Abdelhamed et al [140] (NoiseFlow) which unites the basic parametric noise models and the power of normalizing flow architectures [141], initially proposed in variational inference [142] and density estimation [143], to approximate the real noise distribution from large datasets. Starting from a simple distribution, NoiseFLow learns the transformation to a complex distribution of the real noise via a sequence of differentiable and invertible mappings. The learned noise distribution is then used to compile a set of realistic synthetic images for training deep neural networks.

4.2.2 Generative Adversarial Networks

In recent years, generative adversarial networks (GAN) [144] have received significant attention in a variety of applications in computer vision and image processing tasks, thanks to their compelling ability to generate realistic examples plausibly drawn from an existing distribution of samples. Typically, GANs consist of a generator and a discriminator network. The former is used to generate synthetic samples which are hard to be distinguished from real data and, the latter is trained to distinguish whether the sample is from real data or the generator. For image denoising, the applicability of GANs has been explored in recent years [25, 122, 126, 145–147].

Chen et al. [25] was the first to leverage GAN for real noise modelling and build a paired training dataset for training. Specifically, the generator produces synthetic noise while the discriminator is trained to distinguish between real and synthetic noise. However, this approach only takes a random vector as the input to the noise generator. Thus, the noise samples from the generator are signal-independent since it has not seen the clean intensity during the training. Kim et al. [145] improved the previous method by including more parameters such as clean image, ISO and shutter speed as additional inputs to the generator. Most recently, Chang et al. [126] proposed to decouple the noise generation and camera characteristics via two distinct networks. Particularly, a noise generative network receives and processes the clean image and an initial noise sample. In addition, a latent vector generated from a camera encoding network is adopted to transfer camera-specific characteristics of the image to the noise generative network via feature concatenation. The final synthetic noise is obtained in the output of the generative network. The noise generative and camera encoding networks are trained jointly along with a discriminator supervised by adversarial and feature matching [148] and triplet loss [149].

Inspired by TripleGan [150], Yue et al.[122] designed a dual GAN to learn the joint distribution of noisy and clean pairs. The joint distribution is approximated by its two different factorized forms. Therefore, their proposed framework consists of two networks; a) a denoiser that maps the noisy image to the clean estimate and, b) a noise generator that maps the clean image to the noisy one. Both of these networks are jointly trained along with a discriminator. After training, the learned denoiser can be directly used for noise removal.

On the other hand, the noise generator can also be utilized to build realistic noisy and clean training pairs. Marras et al. [147] proposed to constrain the residue of the noisy input. A denoising network takes the noisy input as well as encoded information about the camera and produces the residue estimate. During training, the ground-truth residue, clean image and, encoded camera information are further fed into an auto-encoder to estimate the residue estimate. Given that the decoders in both denoising and auto-encoder are shared across networks, the denoiser is explicitly constrained to generate residual estimates that are consistent with the noise manifold.

5 Unsupervised & Self-supervised Denoising

The DL-based image denoising research has flourished with hundreds of works seeking to learn the mapping between noisy and clean pairs. However, collecting clean images in some domains is very expensive, or sometimes infeasible. Accordingly, some interests have been recently put into unsupervised learning schemes for denoising. Among them, leveraging image priors [151–153, 166, 167] and/or noise statistics [155–160, 168–171] has been a prominent approach. Another line of work is to design advanced self-supervised loss functions to train networks in absence of clean images [161, 162, 172, 173]. It is noteworthy that researchers often use the terms *unsupervised denoising*, *self-supervised denoising* and, *blind denoising* interchangeably in the literature.

5.1 Unbiased MSE Estimators

Mean-squared error is recognized as an indispensable element of the deep denoisers that necessitates the availability of clean ground truths during the training. In the past, the applicability of Stein's unbiased risk estimator (SURE) [172] has been explored for unsupervised denoising in traditional frameworks [174, 175]. Given its success, SURE has attracted considerable attention in DL-based image denoising over the past few years [161, 162, 173]. Soltanayev et al. [161] proposed the first work investigating the benefits of SURE in DL-based denoisers in lieu of the MSE loss. The SURE function can be written as:

$$\mathcal{L}_{SURE} = \frac{1}{n}\mathbf{X} - \mathcal{F}(\mathbf{X};\ \theta)^2 - \sigma^2 + \frac{2\sigma^2}{n}\sum_{i=1}^{n}\frac{\partial \mathcal{F}(x_i;\ \theta)}{\partial x_i} \tag{18}$$

However, the divergence term in Eq. 18 cannot be analytically solved in many circumstances. To address this issue, authors adopted Monte-Carlo SURE [176] to approximate the divergence term with the following:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mathcal{F}(\mathbf{z};\ \theta)}{\partial z_i} \approx \frac{1}{\epsilon n}\hat{\boldsymbol{\eta}}^T\left(\mathcal{F}(\mathbf{z}+\hat{\boldsymbol{\eta}};\ \boldsymbol{\theta}) - \mathbf{z}\right) \tag{19}$$

Provided that $\hat{\boldsymbol{\eta}}$ is a random vector from normal distribution $\hat{\boldsymbol{\eta}} \sim N(0,1)$ and, ϵ is a fixed small positive value. The DnCNN [51] network is trained with the

Table 4: Summary of some representative methods in self-supervised and unsupervised denoising. The "BS", "SU", "PR", "IO" represent networks based on blind-spot networks, SURE-based losses, image prior, and input output custimization, respectively.

Ref	Publication	Framework				Technique
		BS	SU	PR	IO	
[151]	2018, CVPR	-	-	✓	-	captures image statistics via network structure
[152]	2020, ECCV	-	-	✓	-	combines deep image prior with neural architecture search
[153]	2019, ICASSP	-	-	✓	-	combines deep image prior with total variation regularization
[154]	2018, ICASSP	✓	✓	-	-	denoiser based on SURE-like estimated loss and blind-spot nets
[155]	2020, CVPR	-	-	-	✓	adds noise to input and learns mapping between noisier and original noisy images
[156]	2019, CVPR	-	-	-	✓	leverages surrounding context to predict noise-free estimates of a central pixel.
[157]	2018, ICML	-	-	-	✓	learns mapping between pairs of noisy images of the same clean image.
[158]	2019, NeurIPS	-	-	-	✓	formulates blind-spot denoising in a Bayesian framework.
[159]	2021, CVPR	-	-	-	✓	assumes neighboring pixels as different noisy realizations of same signal.
[160]	2020, NeurIPS	-	-	-	✓	provides a novel training and loss which exploits the entire noisy image for updates.
[161]	2018, NeurIPS	-	✓	-	-	uses SURE-based loss for unsupervised training of denoising networks
[162]	2019, NeurIPS	-	✓	-	-	combines SURE-based loss and *noise2noise* training.
[163]	2021, CVPR	-	-	-	✓	unpaired noisy input outputs
[159]	2021, CVPR	-	-	-	✓	uses neighboring pixels as target noisy outputs
[164]	2021, ICCV	-	-	✓	-	designs a new stopping criterion for image prior scheme.
[165]	2021, NeurIPS	-	-	-	✓	provides a Bayesian solution based on Tweedie's formula.

proposed MC-SURE objective function for simulated Gaussian noise removal. Zhussip et al. [162] extended the SURE-based method to train denoising networks when two uncorrelated noise realizations per clean image are available. Furthermore, they investigated the feasibility of using imperfect clean ground truths for supervision. The Monte-Carlo approximation for SURE involves a hyper-parameter for optimal performance that is hard to select. To address this, Soltanayev et al. [173] proposed a new approximation for divergence term without any hyper-parameter.

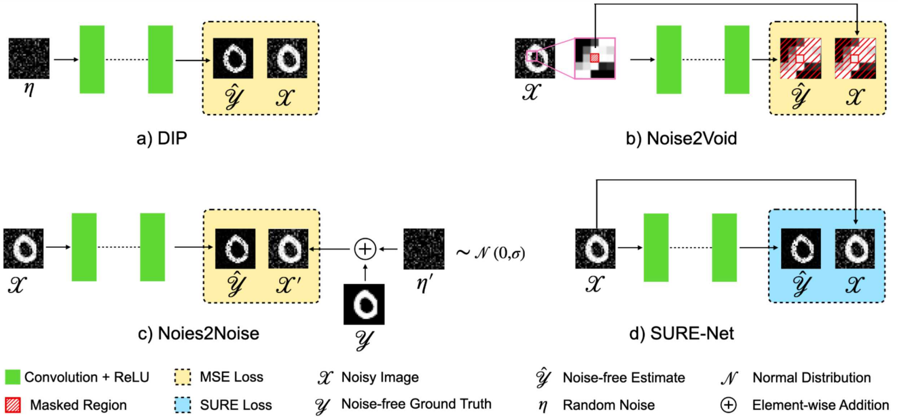
η
$\hat{\mathcal{Y}}$
$\mathcal{X}$
a) DIP
b) Noise2Void
$\mathcal{X}'$
η'
$\sim \mathcal{N}(0,\sigma)$
$\mathcal{Y}$
c) Noies2Noise
d) SURE-Net
Convolution + ReLU
Masked Region
MSE Loss
SURE Loss
$\mathcal{X}$ Noisy Image
$\mathcal{Y}$ Noise-free Ground Truth
$\hat{\mathcal{Y}}$ Noise-free Estimate
η Random Noise
$\mathcal{N}$ Normal Distribution
$\oplus$ Element-wise Addition

Fig. 5: A glimpse of the some methods representing different network architectures for supervised image denoising.

5.2 Image Prior

Some researchers have examined the benefits of data-driven or hand-crafted priors for unsupervised denoising. [151, 152, 166, 167]. Ulyanov et al. [151] showed that the structure of the neural networks itself is able to capture a great portion of the image statistics prior. Specifically, an image-specific network is firstly initialized with random weights and then a random uniform vector **u** is fed into the input layer and, the parameters of the network are optimized to match the output of the network to the observed noisy image using L2 loss. i.e.

$$\theta^* = \arg\min_{\theta} \mathcal{F}_i(\mathbf{u};\ \theta) - x_i; \qquad \forall i \in \{1, 2, ..., n\} \tag{20}$$

In a clean image, different regions are spatially coherent and therefore the network can rapidly capture this prior and reconstruct smooth estimates. Conversely, less spatial coherency makes the perfect reconstruction of the noise to be time-consuming. Accordingly, the implicit regularization imposed by the structure of the image and the early stopping of the optimization implies generating clean estimates. Mataev et al. [167] combined the implicit regularization captured by the CNN structure in deep image prior [151] with the explicit regularization paradigm in Regularization by Denoising [177] to improve the overall regularization effect and enhance image denoising. Similarly, Liu et al. [153] proposed to combine the implicit CNN regularization with an explicit total variation penalty to improve the denoising power of deep image prior [151]. Chen et al. [152] took deep image prior idea one step further and employed neural search algorithm [178] to optimize the CNN structure by searching for both the upsampling units in the decoder and skip connection patterns between encoder and decoder layers. Jo et al. [164] designed a novel metric based on the loss value to improve the stopping criterion in deep image prior training.

5.3 Noise Statistics

One of the well-known properties of the noise component in digital images imply that the noise pixels at different spatial locations are independent given the clean pixel values, i.e.

$$P(\eta_i \mid y_i) \perp_{i \neq j} P(\eta_j \mid y_j); \qquad \forall i, j \in \{1, 2, ..., n\} \tag{21}$$

Furthermore, the noise is assumed to be zero-mean distributed, i.e.

$$\mathbb{E}[\eta_i] = 0; \qquad \forall i \in \{1, 2, ..., n\} \tag{22}$$

Many works have recently relied on these statistical properties of the noise and have employed neural networks for the denoising task in absence of clean images[156, 157, 168], showing a great advance in performance with respect to both the reconstruction error and perceptual quality. We briefly discuss the representative methods related to this line in the following.

Noise2Noise Learning. Lehtinen et al. [157] introduced *noise2noise* a training scheme in which the parameters of the network are optimized to learn a mapping function $\Phi^{-1}(\mathbf{X}, \theta)$ between pairs of independent corrupted images $\mathbf{X} = \{y_i + \eta_i\}$ and $\mathbf{X\prime} = \{y_i + \eta_i'\}$. In other words, two images $\mathbf{X}$ and $\mathbf{X\prime}$ in the training pairs are identical as they share the same underlying clean image $\mathbf{Y}$, but per pixel noise realizations, i.e. $\{\eta_i\}$ and $\{\eta_i\prime\}$, are independent and different. With such training pairs, the network minimizes the MSE loss between the original noisy input and the noisy target,

$$\theta^* = \arg\min_{\theta} \sum_{i}^{n} \|x_i - x_i'\|^2 \tag{23}$$

Obviously, it is impossible for the learned network to predict a different noisy image from another one. Therefore, the network inevitably converges to output the arithmetic mean of inputs for each pixel, i.e. $\mathbb{E}[x_i]$. Given that the noise is assumed to be zero-mean, the learned network converges the clean image, as shown below:

$$\begin{aligned} \mathbb{E}[x_i] &= \mathbb{E}[y_i + \eta_i] && (24) \\ &= \mathbb{E}[y_i] + \mathbb{E}[\eta_i] && (25) \\ &= \mathbb{E}[y_i] && (26) \end{aligned}$$

This training framework allows the network to be trained only based on the noisy images without access to the clean ground truth. Even though this learning strategy may ask for multiple noisy images during training, however, Lehtinen et al. demonstrated that even one additional noisy image is sufficient to achieve reasonable denoising performance. Recently,

Blind-spot Networks. Despite the unprecedented success of noise2noise, requiring different noisy pairs during training is a significant shortcoming of *noise2noise*. To solve this problem, a line of works [154, 156, 158, 168, 171, 179] pioneered by *noise2void* [156] and *noise2self* [168] proposed the idea of *blind-spot* networks to train denoisers by using single noisy images without ground truth. Considering a patch $\{x_i\}_{i=1}^{k^2}$ of size $k \times k$ centered at location i, the central pixel x_i is excluded from the receptive of the network through a masking scheme. Then, the network is trained to predict the value at location i while the original pixel value x_i is utilized as the ground truth for loss calculation. Due to the lack of information about the x_i in feed-forward, the network fails to learn an identity mapping between the input and output and, unavoidably produces an estimate consistent with the surrounding. According to Eq. 21 and Eq. 24, the neighbouring pixels carry no information about the noise part η_i and therefore the networks produce the expected value of inputs at convergence. i.e. $\mathbb{E}[x_i]$. The key idea of blind-spot networks has been recently expanded by Laine et al. [158] who incorporated the blind receptive field in architecture design rather than masking scheme on input patches. In particular, four rotated version of the input image is fed into a network with directional receptive filed to build exclude the central pixel from the surrounding context.

Other Improvements. Along this direction, Quan et al. [169] proposed *self2self* which trains a single network for every noisy image. They exploit the dropout operations [59] to randomly mask a subset of pixels during training. At inference, multiple random masks are applied to the input image to produce a set of clean estimates followed by an averaging to generate a robust clean estimate with less variance. Moran et al. [155] built on noise2noise and proposed a novel learning strategy called *noisier2noise*. Given a statistical model for the noise, a synthetic noise is drawn and added to the original noisy image to generate a doubly-noisy image. Then, the network is trained to predict the original noisy image from the doubly-noisy image. After convergence, the clean estimate is obtained by a set of simple mathematical operations. Similar to noisier2noise, Xu et al. [170] proposed to add synthetic noise to the original noisy image in the input, however, they train a distinct network for every individual image to be denoised. Additionally, Huang et al. [159] introduced *neighbor2neighbor* learning to train a network on different versions of the individual noisy image collected by a novel random neighbour sub-sampler. Since neighbour pixels are very similar in terms of underlying clean content and, the noise is independent, such a learning strategy approximates the noise2noise learning using only single noisy images. Furthermore, the authors proposed a novel regularization in the loss function to improve the denoising performance. Pang et al. [163] extended the *noise2noise* and proposed *Recorrupted2Recorrupted* scheme where only a set unorganized noisy images without pair-wise correspondences are used during training. Lastly, Kim et al. [165] introduced a novel Bayesian framework called *noise2score* that provides the posterior mean of canonical parameters from noisy images based on Tweedie's formula [180].

6 Denoising Applications

Thanks to the ubiquitous demand for denoising algorithms based on deep learning and the vigorous advances of denoising techniques in recent years, the ideas of DL-based denoising are being widely adopted in various applications such as video and, burst images.

6.1 Joint Demosaicing and Denoising

Image demosaicing and denoising typically form the first two components in the image processing pipeline of the camera systems, which bring the most data loss and perturbation. The sequential nature of these two modules results in accumulative errors introduced by either and it is therefore sub-optimal. Recent data-driven approaches have been developed to mitigate this challenge by joint demosaicing and denoising [76, 181–184]. Ehret et al. [183] presented the first attempt to leverage a deep neural network for joint learning of demosaicing and denoising through fine-tuning on RAW bursts. Inspired by the emergence of content-adaptive networks, Liu et al. [182] proposed a self-guided network for joint demosaicing and denoising. In particular, an initial estimate of the green channel is used to guide the improve content recovery for all channels in the main branch. Moreover, a density map is utilized in the main branch to help the network deal better with different difficulty levels of regions. Xing et al. [184] focused on developing a joint solution for three fundamental image restoration problems – demosaicing, denoising, and super-resolution. Their proposed network is universal in the sense that each of the modules can be eliminated from the process, which results in the output for the remaining modules.

6.2 Burst Denoising

In recent years, mobile photography on handheld mobiles has been re-targeted to the task of burst denoising. A burst captures a sequence of short exposure with small cross-frame motion and strong in-frame noise. Given that the noise is independent across different frames, burst denoising relies on the assumption that averaging multiple noisy images lead to a more accurate estimate of the clean image. Many recently proposed burst denoising techniques employ deep neural networks to improve the state-of-the-art [76, 92, 185–192].

Mildenhall et al. [185] adopted kernel-prediction networks (KPN) for burst denoising in which pixel-wise kernels are predicted by the network and convolved with the sequence of frames to obtain a clean frame. Doing so, The averaging weights in every window centred at pixels are predicted from the noisy images to address the cross-frame motion and in-frame image discontinuities. Later, Marinc et al. [191] proposed an extended version of KPNs with multiple kernels of different sizes. Additionally, Zhang et al. [92] equipped KPNs with an attention mechanism to account for inter-frame and intra-frame relationships for better denoising performance. Taking one step further, Xia et al. [186] proposed a basis prediction network (BPN) that, given a sequence of the burst images, produces a set of basis 3D kernels and per-pixel mixing coefficients. Basis kernels and coefficients are then combined to generate per-pixel kernels which are then convolved to estimate the clean image. Liang et al. [188] designed a model to decouple the learning of motion from the learning of noise statistics for burst denoising. Most recently, Bhat et al. proposed a deep reparametrization of the MAP formulation for burst image super-resolution and denoising. Their method learns a error metric and a feature space for the target clean image. The learned feature space can then be used to directly image formation process and to integrate image priors in to the clean estimate.

6.3 Video Denoising

Until recently, video denoising with neural networks had been largely underexplored. Similar to burst denoising, video data typically contain a strong correlation along the temporal dimension that could help the restoration process. Therefore, existing works on video denoising mainly focus on making full use of the spatio-temporal correlations between consecutive frames via recurrent methods [193–195], explicit motion estimation and warping [196–198] and, implicit motion compensation [199–203]. The first attempt to denoise videos was proposed by Chen et al. [195] who leveraged recurrent neural networks to learn the mapping between noisy and clean video sequences. Among the optical flow-based methods, Tassano et al. [197] proposed to align individually denoised frames with respect to the reference frame, followed by a temporal denoiser to operate on a sequence of aligned frames. Similarly, Claus et al. [200] proposed the idea of decomposing video denoising into two steps – frame alignment and temporal filtering. Vaksman et al. [204] introduced the idea of generating artificial frames based on patch-crafting, which are used to augment video sequences. The enlarged video sequence is then processed by applying spatial and temporal filtering to yield the denoised video.

Due to the heavy computational costs associated with motion estimation, several attempts have tried to deal with motion in an implicit manner. Claus et al. [200] suggested sequentially chain spatial denoising and temporal fusion by processing three frames at a time to get the clean estimate of the middle frame. Similarly, Tassano et al. [201] extended their previous work by replacing the optical flow alignment with an implicit motion compensation integrated by network architecture. Maggioni et al. [194] adopted a multi-stage framework to perform video denoising. The temporal coherency across frames is firstly aggregated by a fusion stage and, then a spatial denoising stage discards the leftover noise in the fused image. Lastly, a spatio-temporal refinement step restores more high-frequency details.

More recently, some works have attempted to take advantage of the temporal redundancy in videos in designing self-supervised denoising solutions [198, 203, 205]. Inspired by *noise2noise* [157], Ehret et al. [198] proposed a frame-to-frame training scheme for blind video denoising to adapt a generic pre-trained denoiser for different noise models and data. Lee et al. [206]

proposed a simple yet effective self-supervised training scheme in which a pre-trained denoiser is fine-tuned for every individual input test sequence. During fine-tuning, the initial output of the pre-trained network is considered as the pseudo clean ground truth for loss calculation. Given a sequence of noisy frames surrounding frame at time t, Dewil et al.[203] adopted a solution similar to blind-spot networks and proposed to withhold frame at time *t-1* from the inputs to the network and use it as the ground truth to penalize network's output for the frame at time t during training. Lastly,

6.4 Medical Imaging

Medical imaging analysis has witnessed rapid development in recent decades and has become a crucial factor in disease diagnosis. Medical images are often used to provide an accurate internal view of the human body which is subsequently assessed by diagnostic techniques to identify tissues or body organs requiring treatment. In medical imaging, precise and accurate information extraction is of paramount importance for disease diagnosis, staging and treatment. However, noise artifacts may degrade the visual representation of the medical images during the process of acquisition and/or later processing steps. On the other hand, low-quality medical images baffle the identification of the disease and may hamper the patient's care and treatment. Hence, denoising of medical images is indispensable and has become a mandatory pre-processing stage in medical imaging systems. In this section, we firstly provide a brief introduction to most prevailing medical imaging modalities followed by recent advances in denoising algorithms for each of them.

Ultrasound Imaging. Ultrasound provides an efficient and non-invasive medical imaging modality and is widespread in medical diagnosis for muscle-skeletal, cardiac, and obstetrical diseases. One of the main issues with ultrasound images is the presence of noise artifacts introduced during the process of acquisition, transmission and analysis, which complicates the diagnosis of diseases by clinicians or computer-aided diagnosis (CAD) systems liu2019deep. In recent years, attempts have been made to leverage the strength of CNNs with application in ultrasound denoising, a.k.a despeckling [207–209]. Lan et al. [207] proposed a novel residual network based on UNet architecture for ultrasound image despeckling. Furthermore, they equipped the network with both spatial and channel-wise attention mechanisms to enhance the feature learning of the network for improved noise removal. Cammarasana et al. [209] trained a network in which the inputs are noisy images and the ground truths are denoised counterparts obtained from the parameters-tuned WNNM algorithm.

Magnetic Resonance Imaging. Magnetic resonance (MR) imaging is a widely used non-invasive imaging technique that provides high-resolution visualization of the anatomical structure, tissues and organs. However, MR images may inevitably be captured along with noise artifacts caused by physiological motion, instabilities of the MR imaging scanning hardware, and short acquisition time. As a result, noise removal or attenuation is essential for the comprehension and evaluation of MR images. In recent years, a number of deep learning-based methods have been proposed for image denoising [210–213]. Jian et al. [210] proposed to extend DnCNN [51] for multi-channel MR inputs in two training strategies: with and without noise model. Ran et al. [211] exploited a residual encoder-decoder coupled with adversarial and perceptual loss [214] to outperform state-of-the-art methods in both simulated and real clinical data. You et al. [212] used a wide architecture design to address the vanishing gradient issue and facilitate capturing more structural features. Most recently, Aetesam et al. [213] proposed to incorporate the prior information about the image degradation in form of loss functions to improve the learning performance of the network. They also adopted the Bayesian maximum a posteriori (MAP) estimator to further improve the quality of the restored images.

Given the advanced in self-supervised denoising approaches [156, 157, 168], Xu et al. [215] proposed a denoising framework for dynamic MRI images. This approach only single noisy images along with a few auxiliary observatories from different time frames to optimize the parameters of the network. To further improve the quality of restored image, single-image and multi-image denoising schemes are aggregated in an end-to-end trainable network. Furthermore, spatial transformer networks [216] are utilized to approximate the motion between slices.

Computed Tomography. Computed tomography (CT) imaging is a widespread imaging modality that allows high-resolution visualization of anatomical structures. However, a major concern inherent to CT acquisition is about the potential health hazard related to the ionizing radiation [217]. A common strategy to alleviate radiation exposure is to lower the operating current in CT examinations. However, a potential drawback of dose reduction is the introduction of noise artifacts in reconstructed CT images. A line of work to improve the quality of low-dose CT images is to post-process the obtained images after reconstruction. Recently, DL-based denoising methods have shown promising performance to remove the unwanted artifacts in low-dose CT.

Early deep learning methods leveraged feedforward and residual architectures to improve the feature extraction capability of networks for low-dose denoising [218–220]. Chen et al. [218] adopted a CNN to learn the mapping between low-dose and normal-dose CT in a patch-by-patch manner. Concurrently, they continued their efforts by proposing a residual encoder-decoder network for patch-based low-dose denoising. Kang et al., [220] proposed to apply a directional wavelet transform on low-dose CT images prior to feeding them into CNNs. By doing so, the network can efficiently exploit the intra- and inter-band correlations to suppress the noise artifacts. In order to capture structural information across large regions, Li et al. [221] introduced a 3D self-attention module to benefit from spatial information both within CT slices and between CT slices.

With the increased popularity of GANs in medical imaging [222], many researchers attempted to boost the performance of DL-based low-dose CT denoising methods [223–226]. Wolterink et al. [225] was the first who adopted GANs for low-dose CT denoising. Next, Yi et al. [226] used a UNet-like architecture for the generator and demonstrated improved denoising performance, thanks to its multi-scale encoding and decoding structure. Yang et al. [227] further augmented the loss functions in a Wasserstein GAN with perceptual loss [214] to replace noise artifacts with more plausible recovered details. Most recently, Zhang et al. [224] further improved the low-dose denoising performance by adding edge-aware and noise-aware attention mechanisms in the generator. They also adopted a multi-scale discriminator to expand its receptive field and improve its judgemental capabilities. A recent work by Li et al. [228] investigated the applicability of cycleGAN [229] to train a low-dose CT denoising network based on unpaired image-to-image translation.

Built around successful use of self-supervised denoisers on the natural images, several works have tried to remove the need for normal-dose CT images during trains of deep networks [230–232]. Hendriksen et al. [232] Noise2Inverse where a CNN is trained to transform an image reconstructed from a sub-sinogram to another from the complementary sub-sinogram. The key idea of Noise2Inverse is to partition the data in the sinogram domain and train the CNN in the image domain. After the training, the network is applied to perform denoising only in the image domain. Inspired by Noise2Noise, Hasan et al. [231] introduced a collaborative technique to map many low-dose CT images to the normal-dose CT counterpart through joint training of multiple generators. The difference between any two generated outputs is further incorporated in the overall loss function to provide the collaborative circumstance between generators. The most recent work on self-supervised CT denoising was proposed by Won et al. [230] who developed a novel training strategy based on the pre-trained noise model and denoiser. For a new test low-dose CT image, the pre-trained denoiser is further fine-tuned through back-propagating the loss between the output and Pseudo-CT, which is simply a noise difference map predicted by a pre-trained noise model.

Positron Emission Tomography. Positron emission tomography (PET) imaging is one of the leading imaging modalities for quantitative *in vivo* measurement of physiological and biochemical processes with application in oncology [233], cardiology [234] and neurology [235]. However, high noise levels are one of the main shortcomings of PET compared to CT or MR. The amount of noise artifact in PET directly depends on two factors: 1) amount of injected tracer and 2) duration of scanning. On the other hand, patients' exposure to radiation has been a major concern in recent years [236]. Therefore, a significant amount of researches has been devoted to reconstructing normal-dose PET images from low-dose counterparts by removing the noise artifacts.

The DL-based denoising methods for PET imaging can be divided into two categories. The first category only works on the PET image for noise removal [237–241]. Xu et al. [239] proposed to leverage a UNet to learn the mapping between the reconstructed PET images with 1/200 injection and normal-dose ground truths. They further offered a multi-slice input strategy to improve the robustness of the network. Gong et al. [237] proposed the idea to pre-train a CNN with simulated data and fine-tune the last few layers using real data. They also adopted perceptual loss [214] to improve details of the restored image. Later, Wang et al. [238] exploited a 3D GAN framework to estimate high-quality normal-dose PET images from their corresponding low-dose PET images. Zhou et al.[241] adopted cycleGAN [229] to learn an enhanced mapping between low-dose and full-dose PET mapping. Most recently, Gong et al. [242] used a Wasserstein GAN to perform denoising on low-dose PET images. They further adopted a task-specific initialization to transfer the weights from a pre-trained model for improved training. The other category encompasses the works that receive PET and MR images as the input to the networks [243, 244]. Xiang et al. [243] designed a CNN with two input channels of low-dose PET and the accompanying T1-weighted acquisition from the MR modality. Then, the network learned to combine these two different inputs to help better remove the noise. Intending to incorporate more structural information in the network, Chen et al. [244] proposed a network to receive multi-contrast MR images along with low-dose PET in the input.

Another emerging family of low-dose PET denoising methods is based on self-supervised or unsupervised training. Cui et al. [245] conducted the first investigation of performing low-dose PET denoising without full-dose clean ground truth. Inspired by deep image prior [151], they designed a network whose input is CT or MR images of the same patient and used original low-dose PET image for loss computation against the loss between network output. The Noise2Noise [157] formed the motivation for Yi et al. [246] to propose a self-supervised method for low-dose PET denoising. In particular, they proposed to employ clinical list-mode PET data allowing for the generation of the real statistically independent noisy image with various noise levels. This data then were used to train CNNs over pairs of noisy images.

Fluorescence Microscopy. Fluorescence microscopy (FM) has become an indispensable tool in cell biology that provides visualization of living cells and tissues, hence forming the basis for the analysis of their morphological and structural characteristics. However, due to the weak signals and diffraction limit, FM images suffer from high noise artifacts. There are several methods focused on the development of DL-based denoising schemes in FM imaging [247–249]. Weigert et al. [247] applied a data generation technique to collect semi-synthetic FM images followed by training a UNet model for image restoration. A step into combining optimization scheme and deep learning was made by Pronina et al. [248] by aggregating learnable regularizers into the Wiener-Kolmogorov filter.

A common practice to collect pairs of noisy and clean images in FM is to simulate noise from models and overlay it onto the synthetic clean images. Zhong et al. [250] followed the same strategy and adopted a GAN framework to synthesize synthetic noisy and clean pairs to train a denoising network.

Zhang et al. [251] introduced the first dataset for CM imaging where the clean images are obtained by averaging multiple noisy captures. Averaging process, however, is not an effective way to obtain the clean images as it only weakens the noise artifact rather than eliminating them. On the other hand, the process of collecting the clean ground truths is cumbersome.

To lift the requirement of the clean image during training, a variety of self-supervised and unsupervised schemes for CM denoising were proposed in recent years [250, 252–258]. Izadi et al. [252] developed a disentangling network that was able to separate the noise and signal parts of the noisy image using formulated prior information about noise and desired clean output in the loss function. Later, they integrated classic the patch-based non-local Bayesian filtering algorithm into deep networks [253]. Goncharova et al.[256] built upon the success of blind-spot networks and proposed to include additional knowledge about the structure of the signal into the self-supervised training. They added a convolution operation between the network output and a point spread function (PSF) to account for the diffraction limitation in light microscopy. Krull et al. [255] extended Noise2Void [156] by computing a posterior distribution based on the sampling-based noise model and prior distribution over the true pixel intensities. The clean estimate for each pixel is then obtained with an arbitrary statistical estimator. The most recent work was developed by Byun et a. [258] with the focus on improving the computational burden and inference speed of blind-spot networks.

7 Future Direction

Very recently and thanks to the advent of deep learning techniques, especially the strong learning capacity of convolutional neural networks, recent literature has witnessed promising progress of denoising in both methodology and applications. However, there still exists open problems, challenges, and limitations because of the intrinsic difficulty of solving ill-posed problems. In this section, we summarize a few challenges together with possible future directions in this field.

Theoretical Analysis. Most of the existing works in DL-based image denoising lack a theoretical foundation to endorse the design choices. Particularly, the proposed methods are often mainly designed by intuition and empirically evaluated on benchmark datasets. In the era of deep learning, it is of the highest significance to bridge the gap between traditional image denoising techniques and neural networks through establishing a solid theoretical foundation in architecture designs, loss functions, and even training strategies.

Universality and Robustness The universality mentioned here is twofold: generalization of the denoising algorithms against 1) different types of noise and 2) different strengths of the noise of the same type. Among the studied denoising algorithms, most of them train distinct networks for different noise strengths and noise types represented by the statistical distributions and their parameters. However, many extrinsic factors from the scene and/or intrinsic parameters from the camera can dynamically influence the nature of the noise in practice. Therefore, improving the robustness of the model against various noise characteristics is substantially meaningful.

Interpretability. DL-based denoising approaches inherit the black-box nature of deep learning models and often aim to reach higher performance on benchmark datasets, ignoring the explainability of the learned representations and results. We believe that the literature demands more thorough efforts to make the models more transparent to humans by illustrating why the found setting and design performs better.

Computational Efficiency. Since DL-based denoising research has been focused on improving the state of the art, progressive improvements on benchmark datasets have been correlated with an increase in network complexity, power consumption and execution time. Accordingly, such powerful denoising models might not necessarily be efficient enough for direct deployment in the real world. For instance, one of the essential uses cases of denoising algorithms in smartphones ISP and other embedded devices with limited computational power that demand highly efficient and fast models for real-time execution. As such, improving the computational burden of DL-based denoising approaches to make them more compatible with existing real-world compute constrained hardware and software is a timely yet challenging topic.

8 Conclusion

Image denoising has played a key role in steadily improving the acquisition quality of cameras and delivering high-quality content to customers and/or other downstream tasks in computer vision. In this paper, we started with revisiting the fundamental concepts and mathematical definition of image denoising and later provided an in-depth review of existing benchmark datasets and widely used evaluation metrics. Then we laid out a novel categorization of supervised and unsupervised techniques and systematically highlighted the improvements and new trends in each category. The novel taxonomy introduced in the paper is systematic and comprehensive and may help the reader appreciate the multiple focus areas of training strategies, loss functions, and architecture designs. We further discussed the denoising problems in burst images and videos and elaborated on the important future research directions in image denoising. This survey provided a unique view of the recent progress in image denoising based on deep learning, which we hope will drive further interest in image denoising problems and address their limitations.

9 Acknowledgements

For their funding support, we are grateful to the Natural Sciences and Engineering Research Council of Canada (NSERC), and its Discovery and Collaborative Health Research Projects (CHRP) programs, and the Canadian

Institutes of Health Research (CIHR). We also thank Compute Canada and NVIDIA for the provision of computational resources.

References

[1] Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018). IEEE

[2] Plotz, T., Roth, S.: Benchmarking denoising algorithms with real photographs. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017). IEEE

[3] Zhang, Y., *et al.*: A poisson-gaussian denoising dataset with real fluorescence microscopy images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). IEEE

[4] Liu, D., Wen, B., Jiao, J., Liu, X., Wang, Z., Huang, T.S.: Connecting image denoising and high-level vision tasks via deep learning. IEEE Transactions on Image Processing (TIP) **29**, 3695–3706 (2020). IEEE

[5] Bertero, M., Boccacci, P.: Introduction to Inverse Problems in Imaging, (1998). CRC press

[6] Vogel, C.R.: Computational Methods for Inverse Problems vol. 23, (2002). SIAM

[7] Buades, A., Coll, B., Morel, J.-.: A non-local algorithm for image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 60–652 (2005). IEEE

[8] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transaction on Image Processing (TIP) **16**(8), 2080–2095 (2007). IEEE

[9] Mairal, J., *et al.*: Non-local sparse models for image restoration. In: IEEE International Conference on Computer Vision (ICCV) (2009). IEEE

[10] Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Transaction on Image Processing (TIP) **18**(11), 2419–2434 (2009). IEEE

[11] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**(1), 259–268 (1992)

[12] Osher, S., *et al.*: An iterative regularization method for total variation-based image restoration. Multiscale Modeling & Simulation **4**(2), 460–489 (2005)

[13] Lan, X., *et al.*: Efficient belief propagation with learned higher-order markov random fields. In: European Conference on Computer Vision (ECCV), pp. 269–282 (2006). Springer

[14] Roth, S., Black, M.J.: Fields of experts. International Journal of Computer Vision (IJCV) **82**(2), 205 (2009). Springer

[15] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems (NeurIPS) **25**, 1097–1105 (2012). Curran Associates, Inc.

[16] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[17] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015). IEEE

[18] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). IEEE

[19] Lemarchand, F., Montesuma, E.F., Pelcat, M., Nogues, E.: Opendenoising: an extensible benchmark for building comparative studies of image denoisers. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2648–2652 (2020). IEEE

[20] Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W., Lin, C.-W.: Deep learning on image denoising: An overview. Neural Networks (2020). Elsevier

[21] Thakur, R.S., Yadav, R.N., Gupta, L.: State-of-art analysis of image denoising methods using convolutional neural networks. IET Image Processing **13**(13), 2367–2380 (2019). IET

[22] Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with bm3d? In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2012). IEEE

[23] Jain, V., Seung, S.: Natural image denoising with convolutional networks. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 769–776 (2009). Curran Associates, Inc.

[24] Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: IEEE International Conference on Computer Vision (ICCV) (2017). IEEE

[25] Chen, J., Chen, J., Chao, H., Yang, M.: Image blind denoising with generative adversarial network based noise modeling. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018). IEEE

[26] Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks **3**(1), 47–57 (2017). IEEE

[27] Hasinoff, S.W.: Photon, Poisson Noise. (2014)

[28] Leyris, C., Hoffmann, A., Valenza, M., Vildeuil, J.-C., Roy, F.: Trap competition inducing rts noise in saturation range in n-mosfets. In: Noise in Devices and Circuits III, vol. 5844, pp. 41–51 (2005). International Society for Optics and Photonics

[29] Konnik, M., Welsh, J.: High-level numerical simulations of noise in ccd and cmos photosensors: review and tutorial. arXiv preprint arXiv:1412.4031 (2014)

[30] Liu, C., Szeliski, R., Bing Kang, S., Zitnick, C.L., Freeman, W.T.: Automatic estimation and removal of noise from a single image. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **30**(2), 299–314 (2008). IEEE

[31] Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K.: Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. IEEE Transaction on Image Processing (TIP) **17**(10), 1737–1754 (2008). IEEE

[32] Martin, D., *et al.*: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: IEEE International Conference on Computer Vision (ICCV) (2001). IEEE

[33] Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: CVPR Workshop (2017). IEEE

[34] Roth, S., Black, M.J.: Fields of experts. International Journal of Computer Vision (IJCV) **82**(2), 205 (2009). Springer

[35] Huang, J.-B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) (2015). IEEE

[36] Kodak24. http://r0k.us/graphics/kodak/

[37] Fujimoto, A., *et al.*: Manga109 dataset and creation of metadata. In: International Workshop on coMics ANalysis, Processing and Understanding, pp. 2–125 (2016)

[38] Zeyde, R., *et al.*: On single image scale-up using sparse-representations. In: Curves and Surfaces, pp. 711–730 (2012). IEEE

[39] Bevilacqua, M., *et al.*: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: British Machine Vision Conference, pp. 135–113510 (2012). BMVA press

[40] Lebrun, M., Colom, M., Morel, J.-M.: The Noise Clinic: a Blind Image Denoising Algorithm. Image Processing On Line **5**, 1–54 (2015)

[41] Anaya, J., Barbu, A.: Renoir – a dataset for real low-light image noise reduction. Journal of Visual Communication and Image Representation **51**, 144–154 (2018). Elsevier

[42] Nam, S., *et al.*: A holistic approach to cross-channel image noise modeling and its application to image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016). IEEE

[43] Xu, J., et al.: Real-world noisy image denoising: A new benchmark. arXiv:1804.02603 (2018)

[44] Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3291–3300 (2018). IEEE

[45] Brummer, B., De Vleeschouwer, C.: Natural image noise dataset. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). IEEE

[46] Wang, Z., Bovik, A.C., Lu, L.: Why is image quality assessment so difficult? In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2002). IEEE

[47] Blau, Y., *et al.*: The 2018 pirm challenge on perceptual image super-resolution. In: European Conference on Computer Vision Workshops (ECCVW) (2018). Springer

[48] Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Asilomar Conference on Signals, Systems Computers (2003)

[49] Zhou Wang, Bovik, A.C.: A universal image quality index. IEEE Signal

Processing Letters **9**(3), 81–84 (2002). IEEE

[50] Zhou Wang, Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transaction on Image Processing (TIP) **13**(4), 600–612 (2004). IEEE

[51] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transaction on Image Processing (TIP) **26**(7), 3142–3155 (2017). IEEE

[52] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **43**(7), 2480–2495 (2020). IEEE

[53] Liu, D., *et al.*: Non-local recurrent network for image restoration. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 1673–1682 (2018). Curran Associates, Inc.

[54] Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2862–2869 (2014). IEEE

[55] Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: International Conference on Machine Learning (ICML), pp. 807–814 (2010). PMLR

[56] Clevert, D.-A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). In: International Conference on Learning Representations (ICLR) (2016)

[57] Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (ICML), vol. 37, pp. 448–456 (2015). PMLR

[58] Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)

[59] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research **15**(56), 1929–1958 (2014)

[60] Krizhevsky, A., *et al.*: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 1097–1105 (2012). Curran Associates, Inc.

[61] Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 341–349. Curran Associates, Inc.

[62] Schmidt, U., Roth, S.: Shrinkage fields for effective image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2774–2781 (2014). IEEE

[63] Hel-Or, Y., Shaked, D.: A discriminative approach for wavelet denoising. IEEE Transaction on Image Processing (TIP) **17**(4), 443–457 (2008). IEEE

[64] Chen, Y., Pock, T.: Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **39**(6), 1256–1272 (2017). IEEE

[65] Chen, Y., Wei Yu, Pock, T.: On learning optimized reaction diffusion processes for effective image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5261–5269 (2015). IEEE

[66] Kim, Y., Jung, H., Min, D., Sohn, K.: Deeply aggregated alternating minimization for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6419–6427 (2017). IEEE

[67] Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. Journal on Imaging Sciences **1**(3), 248–272 (2008). SIAM

[68] Zhang, K., Zuo, W., Zhang, L.: Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. IEEE Transaction on Image Processing (TIP) **27**(9), 4608–4622 (2018). IEEE

[69] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). IEEE

[70] Srivastava, R.K., Greff, K., Schmidhuber, J.: Training very deep networks. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 2377–2385 (2015). Curran Associates, Inc.

[71] Srivastava, R.K., Greff, K., Schmidhuber, J.: Highway networks. International Conference on Machine Learning (ICML) Deep Learning workshop (2015). PMLR

[72] Bae, W., Yoo, J., Chul Ye, J.: Beyond deep residual learning for

image restoration: Persistent homology-guided manifold simplification. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 145–153 (2017). IEEE

[73] Jiao, J., Tu, W.-C., He, S., Lau, R.W.: Formresnet: Formatted residual learning for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 38–46 (2017). IEEE

[74] Liu, X., Suganuma, M., Sun, Z., Okatani, T.: Dual residual networks leveraging the potential of paired operations for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7007–7016 (2019). IEEE

[75] Song, Y., Zhu, Y., Du, X.: Dynamic residual dense network for image denoising. Sensors **19**(17), 3809 (2019). Multidisciplinary Digital Publishing Institute

[76] Kokkinos, F., Lefkimmiatis, S.: Iterative joint image demosaicking and denoising using a residual denoising network. IEEE Transactions on Image Processing (TIP) **28**(8), 4177–4188 (2019). IEEE

[77] Wang, T., Sun, M., Hu, K.: Dilated deep residual network for image denoising. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1272–1279 (2017). IEEE

[78] Ren, H., El-Khamy, M., Lee, J.: Dn-resnet: Efficient deep residual network for image denoising. In: Asian Conference on Computer Vision (ACCV), pp. 215–230 (2018). Springer

[79] Santhanam, V., Morariu, V.I., Davis, L.S.: Generalized deep image to image regression. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5609–5619 (2017). IEEE

[80] Mao, X., Shen, C., Yang, Y.-B.: Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: Advances in Neural Information Processing Systems (NeurIPS), (2016). Curran Associates, Inc.

[81] Meinhardt, T., Moller, M., Hazirbas, C., Cremers, D.: Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. In: IEEE International Conference on Computer Vision (ICCV), pp. 1781–1790 (2017). IEEE

[82] Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep cnn denoiser prior for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2808–2817 (2017). IEEE

[83] Remez, T., Litany, O., Giryes, R., Bronstein, A.M.: Class-aware fully convolutional gaussian and poisson denoising. IEEE Transaction on Image Processing (TIP) **27**(11), 5707–5722 (2018). IEEE

[84] Huang, G., Liu, Z., v. d. Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269 (2017). IEEE

[85] Vaswani, A., *et al.*: Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008 (2017). Curran Associates, Inc.

[86] Chaudhari, S., Polatkan, G., Ramanath, R., Mithal, V.: An attentive survey of attention models. arXiv preprint arXiv:1904.02874 (2019)

[87] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141 (2018). IEEE

[88] Hu, J., *et al.*: Gather-excite: Exploiting feature context in convolutional neural networks. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 9401–9411 (2018). Curran Associates, Inc.

[89] Anwar, S., Barnes, N.: Real image denoising with feature attention. In: IEEE International Conference on Computer Vision (ICCV), pp. 3155–3164 (2019). IEEE

[90] Gu, S., *et al.*: Self-guided network for fast image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). IEEE

[91] Mou, C., Zhang, J.: Graph attention neural network for image restoration. In: 2021 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2021). IEEE

[92] Zhang, B., Jin, S., Xia, Y., Huang, Y., Xiong, Z.: Attention mechanism enhanced kernel prediction networks for denoising of burst images. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2083–2087 (2020). IEEE

[93] Cheng, S., Wang, Y., Huang, H., Liu, D., Fan, H., Liu, S.: Nbnet: Noise basis learning for image denoising with subspace projection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4896–4906 (2021). IEEE

[94] Hu, X., Ma, R., Liu, Z., Cai, Y., Zhao, X., Zhang, Y., Wang, H.: Pseudo 3d auto-correlation network for real image denoising. In: IEEE

Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16175–16184 (2021). IEEE

[95] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: European Conference on Computer Vision (ECCV), pp. 492–511 (2020). Springer

[96] Suganuma, M., Liu, X., Okatani, T.: Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). IEEE

[97] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H., Shao, L.: Multi-stage progressive image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14821–14831 (2021). IEEE

[98] Clevert, D.-A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus) (2016)

[99] Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-normalizing neural networks. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 972–981 (2017). Curran Associates, Inc.

[100] Misra, D.: Mish: A self regularized non-monotonic neural activation function. arXiv preprint arXiv:1908.08681 **4**, 2 (2019)

[101] Kligvasser, I., Shaham, T.R., Michaeli, T.: xunit: Learning a spatial activation function for efficient image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2433–2442 (2018). IEEE

[102] Gu, S., Li, W., Gool, L.V., Timofte, R.: Fast image restoration with multi-bin trainable linear units. In: IEEE International Conference on Computer Vision (ICCV) (2019). IEEE

[103] Lefkimmiatis, S.: Non-local color image denoising with convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017). IEEE

[104] Xia, Z., Chakrabarti, A.: Identifying recurring patterns with deep neural networks for natural image denoising. In: IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 2426–2434 (2020). IEEE

[105] Plötz, T., Roth, S.: Neural nearest neighbors networks. In: Advances

in Neural Information Processing Systems (NeurIPS), pp. 1087–1098 (2018). Curran Associates, Inc.

[106] Zhang, Y., *et al.*: Residual non-local attention networks for image restoration. In: International Conference on Learning Representations (ICLR) (2019)

[107] Guo, L., Zha, Z., Ravishankar, S., Wen, B.: Self-convolution: A highly-efficient operator for non-local image restoration. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1860–1864 (2021). IEEE

[108] Xu, X., Li, M., Sun, W., Yang, M.-H.: Learning spatial and spatio-temporal pixel aggregations for image and video denoising. IEEE Transactions on Image Processing (TIP) **29**, 7153–7165 (2020). IEEE

[109] Lefkimmiatis, S.: Universal denoising networks : A novel cnn architecture for image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018). IEEE

[110] Chang, M., Li, Q., Feng, H., Xu, Z.: Spatial-adaptive network for single image denoising. In: European Conference on Computer Vision (ECCV), pp. 171–187 (2020). Springer

[111] Tachella, J., Tang, J., Davies, M.: The neural tangent link between cnn denoisers and non-local filters. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8618–8627 (2021). IEEE

[112] Zhu, X., Hu, H., Lin, S., Dai, J.: Deformable convnets v2: More deformable, better results. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9308–9316 (2019). IEEE

[113] Valsesia, D., Fracastoro, G., Magli, E.: Deep graph-convolutional image denoising. IEEE Transactions on Image Processing (TIP) **29**, 8226–8237 (2020). IEEE

[114] Brooks, T., *et al.*: Unprocessing images for learned raw denoising. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). IEEE

[115] Guo, S., *et al.*: Toward convolutional blind denoising of real photographs. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). IEEE

[116] Wei, K., Fu, Y., Yang, J., Huang, H.: A physics-based noise formation model for extreme low-light raw denoising. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2758–2767 (2020).

IEEE

[117] Wang, Y., Huang, H., Xu, Q., Liu, J., Liu, Y., Wang, J.: Practical deep raw image denoising on mobile devices. In: European Conference on Computer Vision (ECCV), pp. 1–16 (2020). Springer

[118] Liu, J., Wu, C.-H., Wang, Y., Xu, Q., Zhou, Y., Huang, H., Wang, C., Cai, S., Ding, Y., Fan, H., *et al.*: Learning raw image denoising with bayer pattern unification and bayer preserving augmentation. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2019). IEEE

[119] Liu, Y., Qin, Z., Anwar, S., Ji, P., Kim, D., Caldwell, S., Gedeon, T.: Invertible denoising network: A light solution for real noise removal. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13365–13374 (2021). IEEE

[120] Kim, Y., Soh, J.W., Park, G.Y., Cho, N.I.: Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3482–3492 (2020). IEEE

[121] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H., Shao, L.: Cycleisp: Real image restoration via improved data synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2696–2705 (2020). IEEE

[122] Yue, Z., Zhao, Q., Zhang, L., Meng, D.: Dual adversarial network: Toward real-world noise removal and noise generation. In: European Conference on Computer Vision (ECCV), pp. 41–58 (2020). Springer

[123] Chen, C., et al.: Real-world image denoising with deep boosting. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 1–1 (2019). IEEE

[124] Choi, J.H., Elgendy, O.A., Chan, S.H.: Optimal combination of image denoisers. IEEE Transaction on Image Processing (TIP) **28**(8), 4016–4031 (2019). IEEE

[125] Yue, Z., Yong, H., Zhao, Q., Meng, D., Zhang, L.: Variational denoising network: Toward blind noise modeling and removal. In: Advances in Neural Information Processing Systems (NeurIPS) 32, pp. 1690–1701 (2019). Curran Associates, Inc.

[126] Chang, K.-C., Wang, R., Lin, H.-J., Liu, Y.-L., Chen, C.-P., Chang, Y.-L., Chen, H.-T.: Learning camera-aware noise models. In: European Conference on Computer Vision (ECCV), pp. 343–358 (2020). Springer

[127] Li, Y., Fu, X., Zha, Z.-J.: Cross-patch graph convolutional network for image denoising. In: IEEE Conference on Computer Vision (ICCV), pp. 4651–4660 (2021). IEEE

[128] Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7794–7803 (2018). IEEE

[129] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6450–6458 (2017). IEEE

[130] Valsesia, D., Fracastoro, G., Magli, E.: Image denoising with graph-convolutional neural networks. In: IEEE International Conference on Image Processing (ICIP) (2019). IEEE

[131] Mou, C., Zhang, J., Wu, Z.: Dynamic attentive graph learning for image restoration. In: IEEE Conference on Computer Vision (ICCV), pp. 4328–4337 (2021). IEEE

[132] Moseley, B., Bickel, V., Lopez-Francos, I.G., Rana, L.: Extreme low-light environment-driven image denoising over permanently shadowed lunar regions with a physical noise model. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6317–6327 (2021). IEEE

[133] Pan, Z., Li, B., Cheng, H., Bao, Y.: Deep residual network for msfa raw image denoising. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2413–2417 (2020). IEEE

[134] Jaroensri, R., Biscarrat, C., Aittala, M., Durand, F.: Generating training data for denoising real rgb images via camera pipeline simulation. arXiv preprint arXiv:1904.08825 (2019). IEEE

[135] Zhang, Y., Qin, H., Wang, X., Li, H.: Rethinking noise synthesis and modeling in raw denoising. In: IEEE Conference on Computer Vision (ICCV), pp. 4593–4601 (2021). IEEE

[136] Jang, G., Lee, W., Son, S., Lee, K.M.: C2n: Practical generative noise modeling for real-world denoising. In: IEEE Conference on Computer Vision (ICCV), pp. 2350–2359 (2021). IEEE

[137] Talebi, H., Zhu, X., Milanfar, P.: How to saif-ly boost denoising performance. IEEE Transaction on Image Processing (TIP) **22**(4), 1470–1485 (2013). IEEE

[138] Romano, Y., Elad, M.: Boosting of image denoising algorithms. Journal

on Imaging Sciences **8**(2), 1187–1219 (2015). https://doi.org/10.1137/140990978. SIAM

[139] Chen, C., *et al.*: Deep boosting for image denoising. In: Computer Vision – ECCV 2018, pp. 3–19 (2018). Springer

[140] Abdelhamed, A., Brubaker, M.A., Brown, M.S.: Noise flow: Noise modeling with conditional normalizing flows. In: IEEE International Conference on Computer Vision (ICCV), pp. 3165–3173 (2019). IEEE

[141] Kingma, D.P., Dhariwal, P.: Glow: Generative flow with invertible 1x1 convolutions. In: Advances in Neural Information Processing Systems (NeurIPS) 31, pp. 10215–10224 (2018). Curran Associates, Inc.

[142] Rezende, D.J., Mohamed, S.: Variational inference with normalizing flows. In: International Conference on Machine Learning (ICML), vol. 37, pp. 1530–1538 (2015). PMLR

[143] Dinh, L., Krueger, D., Bengio, Y.: Nice: Non-linear independent components estimation. arXiv preprint arXiv:1410.8516 (2014)

[144] Goodfellow, I., *et al.*: Generative adversarial nets. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 2672–2680 (2014). Curran Associates, Inc.

[145] Kim, D.-W., Ryun Chung, J., Jung, S.-W.: Grdn:grouped residual dense network for real image denoising and gan-based real-world noise modeling. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). IEEE

[146] Lin, K., Li, T.H., Liu, S., Li, G.: Real photographs denoising with noise domain adaptation and attentive generative adversarial network. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2019). IEEE

[147] Marras, I., Chrysos, G.G., Alexiou, I., Slabaugh, G., Zafeiriou, S.: Reconstructing the noise manifold for image denoising. arXiv preprint arXiv:2002.04147 (2020)

[148] Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8798–8807 (2018). IEEE

[149] Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815–823 (2015). IEEE

[150] Li, C., Xu, K., Zhu, J., Zhang, B.: Triple generative adversarial nets. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 4091–4101 (2017). Curran Associates, Inc.

[151] Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018). IEEE

[152] Chen, Y.-C., Gao, C., Robb, E., Huang, J.-B.: Nas-dip: Learning deep image prior with neural architecture search. In: European Conference on Computer Vision (ECCV), pp. 442–459 (2020). Springer

[153] Liu, J., Sun, Y., Xu, X., Kamilov, U.S.: Image restoration using total variation regularized deep image prior. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7715–7719 (2019). IEEE

[154] Cha, S., Moon, T.: Neural adaptive image denoiser. In: (ICASSP (2018). IEEE

[155] Moran, N., Schmidt, D., Zhong, Y., Coady, P.: Noisier2noise: Learning to denoise from unpaired noisy data. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12064–12072 (2020). IEEE

[156] Krull, A., Buchholz, T.-O., Jug, F.: Noise2void - learning denoising from single noisy images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). IEEE

[157] Lehtinen, J., *et al.*: Noise2Noise: Learning image restoration without clean data. In: Dy, J., Krause, A. (eds.) International Conference on Machine Learning (ICML), vol. 80, pp. 2965–2974 (2018). PMLR

[158] Laine, S., *et al.*: High-quality self-supervised deep image denoising. In: Advances in Neural Information Processing Systems (NeurIPS) 32, pp. 6968–6978 (2019). Curran Associates, Inc.

[159] Huang, T., Li, S., Jia, X., Lu, H., Liu, J.: Neighbor2neighbor: Self-supervised denoising from single noisy images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14781–14790 (2021). IEEE

[160] Xie, Y., Wang, Z., Ji, S.: Noise2same: Optimizing a self-supervised bound for image denoising. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 20320–20330 (2020). Curran Associates, Inc.

[161] Soltanayev, S., Chun, S.Y.: Training deep learning based denoisers without ground truth data. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 3257–3267 (2018). Curran Associates, Inc.

[162] Zhussip, M., Soltanayev, S., Chun, S.Y.: Extending steins unbiased risk estimator to train deep denoisers with correlated pairs of noisy images. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 1463–1473 (2019). Curran Associates, Inc.

[163] Pang, T., Zheng, H., Quan, Y., Ji, H.: Recorrupted-to-recorrupted: unsupervised deep learning for image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2043–2052 (2021). IEEE

[164] Jo, Y., Chun, S.Y., Choi, J.: Rethinking deep image prior for denoising. In: IEEE Conference on Computer Vision (ICCV), pp. 5087–5096 (2021). IEEE

[165] Kim, K., Ye, J.C.: Noise2score: tweedie's approach to self-supervised image denoising without clean images. Advances in Neural Information Processing Systems (NeurIPS) **34**, 864–874 (2021). Curran Associates, Inc.

[166] Izadi, S., *et al.*: Whitenner-blind image denoising via noise whiteness priors. In: IEEE International Conference on Computer Vision Workshops (ICCVW) (2019). IEEE

[167] Mataev, G., Milanfar, P., Elad, M.: Deepred: Deep image prior powered by red. In: CVPR Workshops (2019). IEEE

[168] Batson, J., Royer, L.: Noise2Self: Blind denoising by self-supervision. In: International Conference on Machine Learning (ICML), vol. 97, pp. 524–533 (2019). PMLR

[169] Quan, Y., Chen, M., Pang, T., Ji, H.: Self2self with dropout: Learning self-supervised denoising from single image. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1890–1898 (2020). IEEE

[170] Xu, J., Huang, Y., Cheng, M.-M., Liu, L., Zhu, F., Xu, Z., Shao, L.: Noisy-as-clean: learning self-supervised denoising from corrupted image. IEEE Transactions on Image Processing (TIP) **29**, 9316–9329 (2020). IEEE

[171] Cha, S., Moon, T.: Fully convolutional pixel adaptive image denoiser. In: IEEE International Conference on Computer Vision (ICCV) (2019). IEEE

[172] Stein, C.M.: Estimation of the mean of a multivariate normal distribution. The annals of Statistics, 1135–1151 (1981). JSTOR

[173] Soltanayev, S., Giryes, R., Chun, S.Y., Eldar, Y.C.: On divergence approximations for unsupervised training of deep denoisers based on stein's unbiased risk estimator. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3592–3596 (2020). IEEE

[174] Nguyen, M.P., Chun, S.Y.: Bounded self-weights estimation method for non-local means image denoising using minimax estimators. IEEE Transactions on Image Processing (TIP) **26**(4), 1637–1649 (2017). IEEE

[175] Van De Ville, D., Kocher, M.: Sure-based non-local means. IEEE Signal Processing Letters **16**(11), 973–976 (2009). IEEE

[176] Ramani, S., Blu, T., Unser, M.: Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms. IEEE Transactions on Image Processing (TIP) **17**(9), 1540–1554 (2008). IEEE

[177] Romano, Y., Elad, M., Milanfar, P.: The little engine that could: Regularization by denoising (red). Journal on Imaging Sciences **10**(4), 1804–1844 (2017). https://doi.org/10.1137/16M1102884. SIAM

[178] Zoph, B., Le, Q.V.: Neural architecture search with reinforcement learning. In: International Conference on Learning Representations (ICLR) (2017)

[179] Wu, X., Liu, M., Cao, Y., Ren, D., Zuo, W.: Unpaired learning of deep image denoising. In: European Conference on Computer Vision (ECCV), pp. 352–368 (2020). Springer

[180] Efron, B.: Tweedie's formula and selection bias. Journal of the American Statistical Association **106**(496), 1602–1614 (2011). Taylor & Francis

[181] Zhou, R., Achanta, R., Süsstrunk, S.: Deep residual network for joint demosaicing and super-resolution. In: Color and Imaging Conference, vol. 2018, pp. 75–80 (2018). Society for Imaging Science and Technology

[182] Liu, L., Jia, X., Liu, J., Tian, Q.: Joint demosaicing and denoising with self guidance. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2240–2249 (2020). IEEE

[183] Ehret, T., *et al.*: Joint demosaicking and denoising by fine-tuning of bursts of raw images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). IEEE

[184] Xing, W., Egiazarian, K.: End-to-end learning for joint image demosaicing, denoising and super-resolution. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3507–3516 (2021). IEEE

[185] Mildenhall, B., Barron, J.T., Chen, J., Sharlet, D., Ng, R., Carroll, R.: Burst denoising with kernel prediction networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018). IEEE

[186] Xia, Z., Perazzi, F., Gharbi, M., Sunkavalli, K., Chakrabarti, A.: Basis prediction networks for effective burst denoising with large kernels. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11844–11853 (2020). IEEE

[187] Rong, X., Demandolx, D., Matzen, K., Chatterjee, P., Tian, Y.: Burst denoising via temporally shifted wavelet transforms. In: European Conference on Computer Vision (ECCV), 2020, Proceedings, Part XIII 16, pp. 240–256 (2020). Springer

[188] Liang, Z., Guo, S., Gu, H., Zhang, H., Zhang, L.: A decoupled learning scheme for real-world burst denoising from raw images. In: European Conference on Computer Vision (ECCV), pp. 150–166 (2020). Springer

[189] Godard, C., *et al.*: Deep burst denoising. In: European Conference on Computer Vision (ECCV), pp. 560–577 (2018). Springe

[190] Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M.: Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics (ToG) **35**(6), 1–12 (2016). ACM

[191] Marinč, T., Srinivasan, V., Gül, S., Hellge, C., Samek, W.: Multi-kernel prediction networks for denoising of burst images. In: 2019 IEEE International Conference on Image Processing (ICIP), pp. 2404–2408 (2019). IEEE

[192] Bhat, G., Danelljan, M., Yu, F., Van Gool, L., Timofte, R.: Deep reparametrization of multi-frame super-resolution and denoising. In: IEEE Conference on Computer Vision (ICCV), pp. 2460–2470 (2021). IEEE

[193] Wang, W., Chen, X., Yang, C., Li, X., Hu, X., Yue, T.: Enhancing low light videos by exploring high sensitivity camera noise. In: IEEE International Conference on Computer Vision (ICCV), pp. 4111–4119 (2019). IEEE

[194] Maggioni, M., Huang, Y., Li, C., Xiao, S., Fu, Z., Song, F.: Efficient multi-stage video denoising with recurrent spatio-temporal fusion. In:

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3466–3475 (2021). IEEE

[195] Chen, X., Song, L., Yang, X.: Deep rnns for video denoising. In: Applications of Digital Image Processing, vol. 9971, p. 99711 (2016). International Society for Optics and Photonics

[196] Xue, T., Chen, B., Wu, J., Wei, D., Freeman, W.T.: Video enhancement with task-oriented flow. International Journal of Computer Vision **127**(8), 1106–1125 (2019). Springer

[197] Tassano, M., Delon, J., Veit, T.: Dvdnet: A fast network for deep video denoising. In: 2019 IEEE International Conference on Image Processing (ICIP), pp. 1805–1809 (2019). IEEE

[198] Ehret, T., Davy, A., Morel, J.-M., Facciolo, G., Arias, P.: Model-blind video denoising via frame-to-frame training. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11369–11378 (2019). IEEE

[199] Yue, H., Cao, C., Liao, L., Chu, R., Yang, J.: Supervised raw video denoising with a benchmark dataset on dynamic scenes. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2301–2310 (2020). IEEE

[200] Claus, M., van Gemert, J.: Videnn: Deep blind video denoising. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2019). IEEE

[201] Tassano, M., Delon, J., Veit, T.: Fastdvdnet: Towards real-time deep video denoising without flow estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1354–1363 (2020). IEEE

[202] Chen, C., Chen, Q., Do, M.N., Koltun, V.: Seeing motion in the dark. In: IEEE International Conference on Computer Vision (ICCV), pp. 3185–3194 (2019). IEEE

[203] Dewil, V., Anger, J., Davy, A., Ehret, T., Facciolo, G., Arias, P.: Self-supervised training for blind multi-frame video denoising. In: IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 2724–2734 (2021). IEEE

[204] Vaksman, G., Elad, M., Milanfar, P.: Patch craft: Video denoising by deep modeling and patch matching. In: IEEE Conference on Computer Vision (ICCV), pp. 2157–2166 (2021). IEEE

[205] Lee, S., Cho, D., Kim, J., Kim, T.H.: Self-supervised fast adaptation for denoising via meta-learning. arXiv preprint arXiv:2001.02899 (2020)

[206] Lee, S., Cho, D., Kim, J., Kim, T.H.: Restore from restored: Video restoration with pseudo clean video. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3537–3546 (2021). IEEE

[207] Lan, Y., Zhang, X.: Real-time ultrasound image despeckling using mixed-attention mechanism based residual unet. IEEE Access **8**, 195327–195340 (2020). IEEE

[208] Karaoğlu, O., Bilge, H.Ş., Uluer, İ.: Removal of speckle noises from ultrasound images using five different deep learning networks. Engineering Science and Technology, an International Journal (2021). Elsevier

[209] Cammarasana, S., Nicolardi, P., Patanè, G.: A universal deep learning framework for real-time denoising of ultrasound images. arXiv preprint arXiv:2101.09122 (2021)

[210] Jiang, D., Dou, W., Vosters, L., Xu, X., Sun, Y., Tan, T.: Denoising of 3d magnetic resonance images with multi-channel residual learning of convolutional neural network. Japanese Journal of Radiology **36**(9), 566–574 (2018). Springer

[211] Ran, M., Hu, J., Chen, Y., Chen, H., Sun, H., Zhou, J., Zhang, Y.: Denoising of 3d magnetic resonance images using a residual encoder–decoder wasserstein generative adversarial network. Medical image analysis **55**, 165–180 (2019). Elsevier

[212] You, X., Cao, N., Lu, H., Mao, M., Wanga, W.: Denoising of mr images with rician noise using a wider neural network and noise range division. Magnetic resonance imaging **64**, 154–159 (2019). Elsevier

[213] Aetesam, H., Maji, S.K.: Noise dependent training for deep parallel ensemble denoising in magnetic resonance images. Biomedical Signal Processing and Control **66**, 102405 (2021). Elsevier

[214] Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision (ECCV), pp. 694–711 (2016). Springer

[215] Xu, J., Adalsteinsson, E.: Deformed2self: Self-supervised denoising for dynamic medical imaging. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 25–35 (2021). Springer

[216] Jaderberg, M., Simonyan, K., Zisserman, A., *et al.*: Spatial transformer networks. Advances in Neural Information Processing Systems (NeurIPS) **28**, 2017–2025 (2015). Curran Associates, Inc.

[217] Brenner, D.J., Hall, E.J.: Computed tomography—an increasing source of radiation exposure. New England journal of medicine **357**(22), 2277–2284 (2007). Mass Medical Soc

[218] Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., Wang, G.: Low-dose ct via convolutional neural network. Biomedical Optics Express **8**(2), 679–694 (2017). Optical Society of America

[219] Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J., Wang, G.: Low-dose ct with a residual encoder-decoder convolutional neural network. IEEE Transactions on Medical Imaging (TMI) **36**(12), 2524–2535 (2017). IEEE

[220] Kang, E., Min, J., Ye, J.C.: A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. Medical physics **44**(10), 360–375 (2017). Wiley Online Library

[221] Li, M., Hsu, W., Xie, X., Cong, J., Gao, W.: Sacnn: Self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network. IEEE Transactions on Medical Imaging (TMI) **39**(7), 2289–2301 (2020). IEEE

[222] Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical image analysis **58**, 101552 (2019). Elsevier

[223] Shan, H., Zhang, Y., Yang, Q., Kruger, U., Kalra, M.K., Sun, L., Cong, W., Wang, G.: 3-d convolutional encoder-decoder network for low-dose ct via transfer learning from a 2-d trained network. IEEE Transactions on Medical Imaging (TMI) **37**(6), 1522–1534 (2018). IEEE

[224] Zhang, X., Han, Z., Shangguan, H., Han, X., Cui, X., Wang, A.: Artifact and detail attention generative adversarial networks for low-dose ct denoising. IEEE Transactions on Medical Imaging (TMI) (2021). IEEE

[225] Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Generative adversarial networks for noise reduction in low-dose ct. IEEE Transactions on Medical Imaging (TMI) **36**(12), 2536–2545 (2017). IEEE

[226] Yi, X., Babyn, P.: Sharpness-aware low-dose ct denoising using conditional generative adversarial network. Journal of digital imaging **31**(5), 655–669 (2018). Springer

[227] Yang, Q., *et al.*: Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE Transactions on Medical Imaging (TMI) **37**(6), 1348–1357 (2018). IEEE

[228] Li, Z., Zhou, S., Huang, J., Yu, L., Jin, M.: Investigation of low-dose ct image denoising using unpaired deep learning methods. IEEE Transactions on Radiation and Plasma Medical Sciences **5**(2), 224–234 (2020). IEEE

[229] Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision (ICCV), pp. 2223–2232 (2017). IEEE

[230] Won, D., Jung, E., An, S., Chikontwe, P., Park, S.H.: Self-supervised learning based ct denoising using pseudo-ct image pairs. arXiv preprint arXiv:2104.02326 (2021)

[231] Hasan, A.M., Mohebbian, M.R., Wahid, K.A., Babyn, P.: Hybrid-collaborative noise2noise denoiser for low-dose ct images. IEEE Transactions on Radiation and Plasma Medical Sciences **5**(2), 235–244 (2020). IEEE

[232] Hendriksen, A.A., Pelt, D.M., Batenburg, K.J.: Noise2inverse: Self-supervised deep convolutional denoising for tomography. IEEE Transactions on Computational Imaging **6**, 1320–1335 (2020). IEEE

[233] Rohren, E.M., Turkington, T.G., Coleman, R.E.: Clinical applications of pet in oncology. Radiology **231**(2), 305–332 (2004). Radiological Society of North America

[234] Dilsizian, V., Bacharach, S.L., Beanlands, R.S., Bergmann, S.R., Delbeke, D., Dorbala, S., Gropler, R.J., Knuuti, J., Schelbert, H.R., Travin, M.I.: Asnc imaging guidelines/snmmi procedure standard for positron emission tomography (pet) nuclear cardiology procedures. Journal of Nuclear Cardiology **23**(5), 1187–1226 (2016). Springer

[235] Herholz, K., Heiss, W.-D.: Positron emission tomography in clinical neurology. Molecular Imaging & Biology **6**(4), 239–269 (2004). Elsevier

[236] Nievelstein, R., van Ufford, H.Q., Kwee, T., Bierings, M., Ludwig, I., Beek, F., de Klerk, J., Mali, W.T.M., de Bruin, P., Geleijns, J.: Radiation exposure and mortality risk from ct and pet imaging of patients with malignant lymphoma. European radiology **22**(9), 1946–1954 (2012). Springer

[237] Gong, K., Guan, J., Liu, C.-C., Qi, J.: Pet image denoising using a deep

neural network through fine tuning. IEEE Transactions on Radiation and Plasma Medical Sciences **3**(2), 153–161 (2018). IEEE

[238] Wang, Y., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., Zhou, L.: 3d conditional generative adversarial networks for high-quality pet image estimation at low dose. Neuroimage **174**, 550–562 (2018). Elsevier

[239] Xu, J., Gong, E., Pauly, J., Zaharchuk, G.: 200x low-dose pet reconstruction using deep learning. arXiv preprint arXiv:1712.04119 (2017)

[240] Ouyang, J., Chen, K.T., Gong, E., Pauly, J., Zaharchuk, G.: Ultra-low-dose pet reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. Medical physics **46**(8), 3555–3564 (2019). Wiley Online Library

[241] Zhou, L., Schaefferkoetter, J.D., Tham, I.W., Huang, G., Yan, J.: Supervised learning with cyclegan for low-dose fdg pet image denoising. Medical image analysis **65**, 101770 (2020). Elsevier

[242] Gong, Y., Shan, H., Teng, Y., Tu, N., Li, M., Liang, G., Wang, G., Wang, S.: Parameter-transferred wasserstein generative adversarial network (pt-wgan) for low-dose pet image denoising. IEEE Transactions on Radiation and Plasma Medical Sciences **5**(2), 213–223 (2020). IEEE

[243] Xiang, L., Qiao, Y., Nie, D., An, L., Lin, W., Wang, Q., Shen, D.: Deep auto-context convolutional neural networks for standard-dose pet image estimation from low-dose pet/mri. Neurocomputing **267**, 406–416 (2017). Elsevier

[244] Chen, K.T., Gong, E., de Carvalho Macruz, F.B., Xu, J., Boumis, A., Khalighi, M., Poston, K.L., Sha, S.J., Greicius, M.D., Mormino, E., *et al.*: Ultra–low-dose 18f-florbetaben amyloid pet imaging using deep learning with multi-contrast mri inputs. Radiology **290**(3), 649–656 (2019). Radiological Society of North America

[245] Cui, J., Gong, K., Guo, N., Wu, C., Meng, X., Kim, K., Zheng, K., Wu, Z., Fu, L., Xu, B., *et al.*: Pet image denoising using unsupervised deep learning. European journal of nuclear medicine and molecular imaging **46**(13), 2780–2789 (2019). Springer

[246] Yie, S.Y., Kang, S.K., Hwang, D., Lee, J.S.: Self-supervised pet denoising. Nuclear Medicine and Molecular Imaging **54**(6), 299–304 (2020). Springer

[247] Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., *et al.*: Content-aware

image restoration: pushing the limits of fluorescence microscopy. Nature methods **15**(12), 1090–1097 (2018). Nature Publishing Group

[248] Pronina, V., Kokkinos, F., Dylov, D.V., Lefkimmiatis, S.: Microscopy image restoration with deep wiener-kolmogorov filters. In: European Conference on Computer Vision (ECCV), 2020, Proceedings, Part XX 16, pp. 185–201 (2020). Springer

[249] Khademi, W., Rao, S., Minnerath, C., Hagen, G., Ventura, J.: Self-supervised poisson-gaussian denoising. In: IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 2131–2139 (2021). IEEE

[250] Zhong, L., Liu, G., Yang, G.: Blind denoising of fluorescence microscopy images using gan-based global noise modeling. In: International Symposium on Biomedical Imaging (ISBI), pp. 863–867 (2021). IEEE

[251] Zhang, Y., Zhu, Y., Nichols, E., Wang, Q., Zhang, S., Smith, C., Howard, S.: A poisson-gaussian denoising dataset with real fluorescence microscopy images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11710–11718 (2019). IEEE

[252] Izadi, S., Mirikharaji, Z., Zhao, M., Hamarneh, G.: Whitenner-blind image denoising via noise whiteness priors. In: IEEE International Conference on Computer Vision Workshops (ICCVW) (2019). IEEE

[253] Izadi, S., Hamarneh, G.: Patch-based non-local bayesian networks for blind confocal microscopy denoising. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 46–55 (2020). Springer

[254] Wang, F., Henninen, T.R., Keller, D., Erni, R.: Noise2atom: unsupervised denoising for scanning transmission electron microscopy images. Applied Microscopy **50**(1), 1–9 (2020). SpringerOpen

[255] Krull, A., Vičar, T., Prakash, M., Lalit, M., Jug, F.: Probabilistic noise2void: Unsupervised content-aware denoising. Frontiers in Computer Science **2**, 5 (2020). https://doi.org/10.3389/fcomp.2020.00005

[256] Goncharova, A.S., Honigmann, A., Jug, F., Krull, A.: Improving blind spot denoising for microscopy. In: European Conference on Computer Vision (ECCV), pp. 380–393 (2020). Springer

[257] Lequyer, J., Philip, R., Sharma, A., Pelletier, L.: Noise2fast: Fast self-supervised single image blind denoising. arXiv preprint arXiv:2108.10209 (2021)

[258] Byun, J., Cha, S., Moon, T.: Fbi-denoiser: Fast blind image denoiser for

poisson-gaussian noise. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5768–5777 (2021). IEEE