Rotation-Equivariant Self-Supervised Method in Image Denoising

Hanze Liu[1] Jiahong Fu[1] Qi Xie[1,†] Deyu Meng[1,2,3]
[1] Xi'an Jiaotong University, Xi'an, China
[2] Pengcheng Laboratory, Shenzhen, China
[3] Macau University of Science and Technology, Taipa, Macao

Abstract

Self-supervised image denoising methods have garnered significant research attention in recent years, for this kind of method reduces the requirement of large training datasets. Compared to supervised methods, self-supervised methods rely more on the prior embedded in deep networks themselves. As a result, most of the self-supervised methods are designed with Convolution Neural Networks (CNNs) architectures, which well capture one of the most important image prior, translation equivariant prior. Inspired by the great success achieved by the introduction of translational equivariance, in this paper, we explore the way to further incorporate another important image prior. Specifically, we first apply high-accuracy rotation equivariant convolution to self-supervised image denoising. Through rigorous theoretical analysis, we have proved that simply replacing all the convolution layers with rotation equivariant convolution layers would modify the network into its rotation equivariant version. To the best of our knowledge, this is the first time that rotation equivariant image prior is introduced to self-supervised image denoising at the network architecture level with a comprehensive theoretical analysis of equivariance errors, which offers a new perspective to the field of self-supervised image denoising. Moreover, to further improve the performance, we design a new mask mechanism to fusion the output of rotation equivariant network and vanilla CNN-based network, and construct an adaptive rotation equivariant framework. Through extensive experiments on three typical methods, we have demonstrated the effectiveness of the proposed method. The code is available at: https://github.com/liuhanze623/AdaReNet.

1. Introduction

During the process of capturing and transmitting images, unexpected noises are frequently introduced [16, 28]. Such noise can severely degrade image quality and disrupt subsequent image processing tasks. Image denoising is a technique designed to address the standard inverse problem in image processing and is widely utilized across various fields. With the rapid advancement of deep learning (DL), denoisers based on different deep networks have achieved outstanding results.

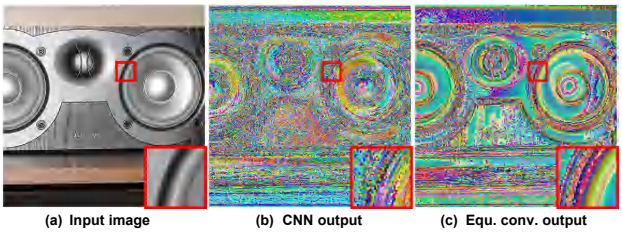


Figure 1. Illustration of the output feature map of a typical image obtained by standard CNN and our used rotation equivariant convolution neural network. Both networks are initialized randomly.

Early, it was believed that the success of deep learning is primarily due to the exhaustive utilization of training data. Thus most early DL-based image denoising are in supervised manner [2, 7, 14, 20, 34, 42, 56, 57], where models are trained on extensive datasets that include pairs of clean and noisy images, thereby learning the transformation from noisy to clean images. However, in reality, compiling such comprehensive datasets is both costly and time-intensive, posing substantial challenges [1, 35, 50].

Self-supervised approaches thus achieved significant research attention in recent years. These approaches rely less on extensive supervised datasets, but pay more attention to the full utilization of the prior information inherent in deep networks [3, 6, 18, 21, 26, 32, 33, 44, 51, 55]. Notably, Lehtinen *et al*. introduced the innovative self-supervised learning approach called Noise2Noise [26], which enables training on pairs of noisy images depicting the identical scene. [26] was the first self-supervised algorithm that achieved performance comparable to supervised image denoising. Other methods involve training denoising models on single noisy images by designing blind-spot networks [3, 21, 22, 48], while some approaches devise strategies to generate training image pairs from noisy im-

† Corresponding author.

ages [6, 18, 32, 33, 51] and have achieved excellent results.

Nowadays, it has become common sense that the prior information inherent in deep networks plays an important role in improving the performance of DL-based methods in image processing tasks [11, 39, 46, 49]. The most classic example should be the CNN, which well captures the translation-equivariant prior of natural images. In CNNs, shifting an input image of CNN is equivalent to shifting all of its intermediate feature maps and the output image. Compared with fully-connected neural networks, this translational equivariance property brings in rational weight sharing, makes the network parameters being used more efficiently, and thus leads to substantially better performance. Inspired by the significant success brought about by the introduction of translational equivariance, another essential image prior (As illustrated in Fig. 1), rotation-equivariant prior has to be taken into consideration very recently. Rotation-equivariant CNNs have been applied and achieved notable performance improvement in multiple supervised image processing tasks [11, 49].

Compared with supervised learning, self-supervised learning approaches indeed depend more on the prior knowledge embedded within the network, for the information that can be learned from the dataset is less. As a result, most current self-supervised image denoising methods are all based on the CNN architectures, which is largely due to the reliance on the prior of translational equivariance. Therefore, it holds even greater significance to incorporate rotational equivariant design into the deep networks for self-supervised methods.

However, there are two critical issues that need to be addressed when introducing equivariant priors into self-supervised image denoising. On the one hand, although, Xie *et al*. have recently constructed a rotation equivariant CNN suitable for image processing [49] and Fu *et al*. further analyzed the global rotation equivariance error of the entire network [11], current rotation equivariant designs for image processing are usually in ResNet[15] structure without upsampling or downsampling. For example, for the super-resolution network in [49] where upsampling module is inevitable, only the portion of the network before upsampling is rotation equivariant and the equivariance for the entire network can not be guaranteed. Since self-supervised networks often utilize U-Net structures, which include multiple upsampling and downsampling modules, it is imperative to first explore the impact of these modules on rotation equivariance of the network.

On the other hand, rotation equivariant design for deep networks inevitably introduces parameter sharing and convolution kernel parameterization, which usually degrades the representation accuracy of the network. However, image denoising tasks often have high requirements for the network's representation accuracy, since important high-frequency components need to be reconstructed. Besides, not all areas of natural images strictly comply with rigid rotation equivariance, indiscriminately adopting a rotation equivariant network for the entire image can often be detrimental to the reconstruction performance of images.

This study primarily focuses on integrating rotation equivariant prior to existing self-supervised denoising techniques while solving the aforementioned issues. The key contributions can be summarized as follows:

- We explore the way for introducing rotation equivariant prior into self-supervised image denoising frameworks at the network architecture level. Particularly, for the first time, we rigorously analyze the impact of upsampling and downsampling on equivariant networks theoretically. The equivariant errors of upsampling and downsampling layer indeed approach zero when the resolution of the input image increases, though the approach rate ($O(h)$, h denoted the mesh size of image) is slower than equivariant convolutional layer (whose equivariant error approaches zero at a rate of $O(h^2)$). Then, by taking U-Net as an example, we further analyze the rotation equivariant error of the entire network, showing that by simply replacing all the convolution layers with rotation equivariant convolutions [49], we can indeed achieve a reliable rotation equivariant network for self-supervised image denoising.
- We have further developed an adaptive rotation equivariant network to enhance the representation accuracy. Specifically, we design a fusion module for integrating the advantages of both rotation-equivariant and non-equivariant networks. The module can automatically determine which regions of the image would benefit more from a rotation-equivariant network compared to a normal CNN-based network. This design provides greater flexibility and yields improved denoising results.
- We conducted comprehensive experiments across various self-supervised denoising methods, demonstrating the effectiveness of integrating rotation-equivariant image priors into neural networks for self-supervised techniques. Our approach provides a novel perspective in the field of self-supervised image denoising.

2. Related Work and Prior Knowledge

2.1. Image Denoising

Non-learning Denoising Methods: Conventional denoising methods predominantly depend on the statistical characteristics of images and mathematical modeling, often employing image priors instead of learned denoisers. Notably, NLM [4, 5] and BM3D [9, 29] have been proposed that can effectively remove noise based on the exploitation of image self-similarity. WNNM [12, 13] treated image denoising as a low-rank matrix approximation problem and achieved great performance.

Supervised Denoising Methods: The majority of DL-based denoising algorithms are supervised [2, 7, 14, 52, 56, 57]. DnCNN [56] employed an end-to-end training strategy to learn the mapping between noisy images and the residual components, markedly improving denoising efficacy. FFDNet [57] proposed a versatile denoising architecture by incorporating the noise level as a network parameter. Nevertheless, to train a proficient model, an extensive collection of paired noisy and clean images is necessary to comprehensively capture the range of image contents and noise variations [2, 14, 19, 25, 37, 53]. The acquisition of such training datasets can be prohibitively costly and challenging, and in the case of certain medical images, it may be virtually unattainable.

Self-supervised Denoising Methods: Self-supervised deep learning methods have garnered significant interest due to their independence from clean reference images [3, 6, 18, 21, 26, 32, 33, 44, 51, 55]. Noise2Noise [26], proposed by Lehtinen *et al*., was the first self-supervised method that achieved performance comparable to supervised methods only using paired noisy images. Noise2Void [21] and Noise2Self [3] used a single noisy image and employed a blind-spot network to predict the clean pixel values based on neighboring pixels, thereby avoiding collapse to an identity mapping. [6, 18, 32, 33, 51] devise strategies to generate training image pairs from noisy images to train a network. R2R [33] has demonstrated that the cost function defined on the noisy/noisy image pairs generated by this method is statistically equivalent to the supervised method. Single-image denoising methods [36, 41] capitalize on the statistical properties of the image itself, showing powerful denoising capabilities without extensive training datasets. These methods lack clean images for supervision and rely more on the prior knowledge embedded within the network, which is the reason for the success of these CNN-based self-supervised image denoising methods. Inspired by this, we explore the way to further incorporate rotation equivariant image prior to these methods with rigorous theoretical analysis.

2.2. Rotation Equivariance

Equivariant Convolution Neural Networks (ECNNs) have drawn substantial interest within the field of computer vision these years [8, 17, 30, 39, 40, 45–47, 58]. Their key strength stems from their ability to effectively handle image rotations through architectural design, which significantly enhances the model's generalization and robustness.

Cohen *et al*. introduced Group Equivariant Convolutional Networks (G-CNNs) [8] and first integrated $\frac{\pi}{2}$ degree rotation equivariance into the neural network. Recently, the filter parametrization technique has been widely employed and Weiler *et al*. used harmonics as steerable filters to achieve exact equivariance [45, 46]. However, a notable limitation of these filter parameterization techniques is the inaccuracy in their representation, which significantly affects low-level vision tasks. Fortunately, Xie *et al*. addressed this issue by proposing Fourier series expansion-based filter parametrization, which has relatively high expression accuracy [49]. The proposed Fconv exhibits precise equivariance in the continuous domain, degrading to approximation only after discretization.

Equivariant convolutions, distinct from data augmentation techniques, inherently integrate rotational symmetry image priors into the network architecture, thereby guaranteeing the network's inherent equivariance and offering superior interpretability and generalizability. However, in the U-Net structure, there is no theoretical guarantee for the design of rotation equivariance. To the best of our knowledge, this study represents the first application of rotation equivariant convolution in the field of self-supervised image denoising and offers theoretical assurance of equivariance for networks based on the U-Net architecture.

2.3. Prior Knowledge about Equivariance

Equivariance to a transformation indicates that the application of a transformation to the input results in a corresponding, predictable transformation of the output [39, 45, 49]. Concretely, consider a mapping Ψ that transforms the input feature space to the output feature space, and let G denote a set of transformations. For any $g \in G$, the following relationship holds:

$$\Psi\left[\pi_g[f]\right] = \pi'_g\left[\Psi[f]\right], \tag{1}$$

where f represents any feature map within the input feature space, and π_g and π'_g describe the action of the transformation g on the input and output features, respectively.

3. Proposed Method

3.1. ECNNs for Self-supervised Image Denoising

Self-supervised image denoising methods lack clean images for supervision and rely more on the prior knowledge embedded within the network. Therefore, introducing rotation equivariant image prior to this field is reasonable, which can be achieved by ECNNs [49]. The U-Net can effectively restore high-frequency details in images with an encoder-decoder architecture, making it widely used in this task. Consequently, to construct an equivariant self-supervised denoising network, we need to discuss the impact of the essential upsampling and downsampling operators in the U-Net network on equivariance.

In the subsequent section, we first present some notations and concepts, and then theoretically analyze the impact of upsampling and downsampling operators on equivariant networks. Furthermore, we deduce the equivariant error for the complete U-Net network and provide theoretical guarantees for the implementation of equivariance in U-Net architectures.

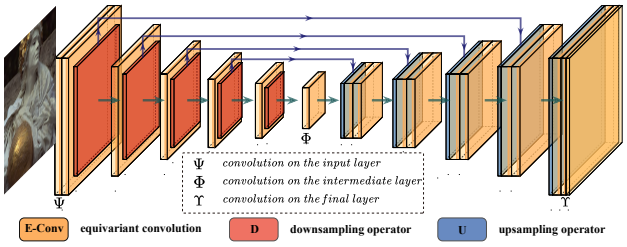


Figure 2. The network architecture of the equivariant N2N method. The network can be divided into multiple upsampling and downsampling blocks. Each downsampling block (DB) consists of one E-Conv layer and a downsampling operator, while each upsampling block (UB) is composed of an upsampling operator and two E-Conv layers.

3.1.1. Notations and Concepts

We first introduce some necessary notations and preliminaries as follows.

We consider the equivariance on the orthogonal group $O(2)$. Formally, $O(2) = \{A \in \mathbb{R}^{2\times 2} | A^T A = I_{2\times 2}\}$, which contains all rotation and reflection matrices. Without ambiguity, we use A to parameterize $O(2)$. The Euclidean group $E(2) = \mathbb{R}^2 \rtimes O(2)$ ($\rtimes$ is a semidirect-product), whose element is represented as (x, A). Restricting the domain of A and x, we can also use this representation to parameterize any subgroup of $E(2)$. In practice, the subgroup is usually assumed to contain t rotations with $\frac{2\pi}{t}$ degree for an integer $t \in \mathbb{N}_+$.

An image $I \in R^{n\times n}$ is viewed as a two-dimensional discretization of a smooth function $r : \mathbb{R}^2 \to \mathbb{R}$, at the cell-center of a regular grid with $n \times n$ cells, *i.e.*, for $i, j = 1, 2, \cdots, n$,

$$I_{ij} = r(x_{ij}), \tag{2}$$

where $x_{ij} = \left(\left(i - \frac{n+1}{2}\right) h, \left(j - \frac{n+1}{2}\right) h\right)^T$, h is the mesh size.

An intermediate feature map $F \in \mathbb{R}^{n\times n\times t}$ in equivariant networks is a multi-channel tensor, which can be viewed as the discretization of a continuous function defined on $\tilde{E} = \mathbb{R}^2 \rtimes S$, where S is a subgroup of $O(2)$ and t is the number of elements in S. Formally, F can be represented as a three-dimensional grid tensor sampled from a smooth function $e : \mathbb{R}^2 \times S \to \mathbb{R}$, *i.e.*, for $i, j = 1, 2, \cdots, n$,

$$F^A_{ij} = e(x_{ij}, A), \tag{3}$$

where x_{ij} is defined in (2) and $A \in S$.

With above notations, the transformations on the input and feature maps can be mathematically formulated. Specifically, in the continuous domain, for an input $r \in C^\infty(\mathbb{R}^2)$ and feature map $e \in C^\infty(E(2))$, the transformation $\tilde{A} \in O(2)$ acts on r and e respectively by:

$$\begin{aligned} \pi^R_{\tilde{A}}[r](x) &= r(\tilde{A}^{-1}x), \forall x \in \mathbb{R}^2, \\ \pi^E_{\tilde{A}}[e](x, A) &= e(\tilde{A}^{-1}x, \tilde{A}^{-1}A), \forall (x, A) \in E(2). \end{aligned} \tag{4}$$

In particular, if $A_\theta \in O(2)$ is the rotation matrix $\begin{bmatrix} \cos\theta, \sin\theta \\ -\sin\theta, \cos\theta \end{bmatrix}$, then the corresponding rotation operators can be expressed by π^R_θ and π^E_θ.

Besides, in the discrete domain, we can also define the transformation $\tilde{A} \in S$ on the input image and feature map as followings:

$$\begin{aligned} \left(\tilde{\pi}^R_{\tilde{A}}(I)\right)_{ij} &= \pi^R_{\tilde{A}}[r](x_{ij}), \\ \left(\tilde{\pi}^E_{\tilde{A}}(F)\right)^A_{ij} &= \pi^E_{\tilde{A}}[e](x_{ij}, A), \\ \forall i, j &= 1,2, \cdots, n, A \in S. \end{aligned} \tag{5}$$

Similarly, rotation operators can be denoted as $\tilde{\pi}^R_\theta$ and $\tilde{\pi}^E_\theta$.

3.1.2. Equivariance of Downsampling and Upsampling

As shown in Fig. 2, the U-Net architecture commonly used in self-supervised image denoising typically integrates upsampling and downsampling operators. In specific applications, these layers will inevitably affect the equivariance of the network. Recent studies [11, 49] have analyzed the equivariance properties of the group convolution layer. However, the effects of upsampling and downsampling on network equivariance remain unexplored.

First of all, We provide the definitions of commonly used downsampling methods in the continuous domain.

Maxpooling Downsampling. Maxpooling is a commonly used downsampling method in CNNs, which reduces the spatial dimensions of feature maps by sliding a fixed-size window over the feature map and selecting the maximum value within each region as the output [23]. In the continuous domain, we can define maxpooling operator $M(\cdot)$ as follows,

$$[M(F)](x, A) = max_{\Omega_x} F^A_{ij}, \tag{6}$$

where $x = [x_1, x_2]^T \in \mathbb{R}^2$ denotes the spatial coordinates, and $x_1 \in [i, i+1]$, $x_2 \in [j, j+1]$, $\Omega_x = \{(i, j), (i+1, j), (i, j+1), (i+1, j+1)\}$.

Stride Downsampling. Stride Downsampling is also a widely used downsampling operator which reduce the size of the feature map by adjusting the stride of the convolution operation [24]. In the continuous domain, we can define stride downsampling operator $S(\cdot)$ as follows,

$$[S(F)](x, A) = F^A_{i,j+1}, \tag{7}$$

where $x = [x_1, x_2]^T \in \mathbb{R}^2$ denotes the spatial coordinates, and $x_1 \in [i, i+1]$, $x_2 \in [j, j+1]$. $\Omega_x = \{(i, j), (i+1, j), (i, j+1), (i+1, j+1)\}$.

For the above two common downsampling operators, it is of great significance to analyze their impact on the equivariant network. Therefore, we constructed their equivariant errors under the rotational equivariant structure.

Theorem 1 *Assume that a feature map $F \in \mathbb{R}^{n\times n\times t}$ is discretized from the smooth function $e : \mathbb{R}^2 \times S \to \mathbb{R}$, $|S| = t$, the mesh size is h, $D(\cdot)$ is the downsampling operator. If for any $A, B \in S, x \in \mathbb{R}^2$, the following conditions are satisfied:*

$$\| \nabla e(x, A) \| \leq G, \tag{8}$$

then the following results are satisfied:

$$|D\left[\tilde{\pi}_B^E\right](F)(x, A) - \pi_B^E[D(F)](x, A)| \leq 2\sqrt{2}Gh. \tag{9}$$

The downsampling operator $D(\cdot)$ can be either $M(\cdot)$ or $S(\cdot)$. Theorem 1 reveals that the equivariant error of the downsample operator is primarily influenced by the mesh size h, and it indeed approach zero when the resolution of the input image increases with the approach rate $O(h)$.

Then, we provide the definitions of commonly used upsampling methods in the continuous domain.

Nearest Neighbor Upsampling. Nearest neighbor interpolation is an image scaling method that fills the pixels of the interpolated image by selecting the original pixel value closest to the target pixel position. In the continuous domain, we can define the nearest neighbor operator $N(\cdot)$ as follows,

$$[N(F)](x, A) = F^A_{i^* j^*}, \tag{10}$$

where $(i^*, j^*) = arg\min_{ij} ||x_{ij} - x||_2^2$.

Bilinear Upsampling. Bilinear interpolation calculates the new pixel value by taking the weighted average of the four surrounding known pixel values. In the continuous domain, we can define the bilinear interpolation operator $B(\cdot)$ as follows,

$$[B(F)](x, A) = \sum_{i=1}^{2}\sum_{j=1}^{2} \lambda_{ij} f(Q_{ij}), \tag{11}$$

where λ_{ij} are the coefficients of bilinear interpolation and $f(Q_{ij})$ represent the grid points, $x = [x_1, x_2]^T \in \mathbb{R}^2$ denotes the 2D spatial coordinates, $x_1 \in [i, i + 1]$, $x_2 \in [j, j + 1]$.

Both of the aforementioned upsampling operators are widely utilized across various network architectures, making it essential to analyze their mathematical properties within rotational equivariant networks. Accordingly, we evaluated their equivariant errors under a rotational equivariant framework.

Theorem 2 *Assume that a feature map $F \in \mathbb{R}^{n\times n\times t}$ is discretized from the smooth function $e : \mathbb{R}^2 \times S \to \mathbb{R}$, $|S| = t$, the mesh size is h, $U(\cdot)$ is the upsampling operator. If for any $A, B \in S, x \in \mathbb{R}^2$, the following conditions are satisfied:*

$$\| \nabla e(x, A) \| \leq G, \tag{12}$$

then the following results are satisfied:

$$|U\left[\tilde{\pi}_B^E\right](F)(x, A) - \pi_B^E[U(F)](x, A)| \leq 2(\sqrt{2} + 1)Gh. \tag{13}$$

The upsampling operator $U(\cdot)$ can be either $N(\cdot)$ or $B(\cdot)$. It is worth noting that the above conclusions indicate that the error introduced by the upsampling operator in the rotational equivariant network is related to h, aligning with established understanding.

3.1.3. Analysis of Complete U-Net Network

We take the network design of the N2N method [26] as an example to give the derivation of the rotation equivariant error for the entire U-Net architecture. As shown in Fig. 2, we decompose the network into multiple upsampling and downsampling blocks. Each downsampling block (DB) consists of one E-Conv layer (Equivariant Convolution) and a downsampling operator, while each upsampling block (UB) is composed of an upsampling operator and two E-Conv layers, we provide the equivariant error for each block and subsequently derive the equivariant error bounds for the complete network.

Theorem 3 demonstrates the equivariance error of the complete U-Net network under discrete angles, and the corollary further provides the equivariance error of the complete U-Net network under any rotation angle. Note that the mesh size of upsampling and downsampling are different, we define the mesh size of the original picture to be h, the mesh size after a $\times 2$ downsampling is $2h$, and so on.

Theorem 3 *For an image X with size $H \times W \times n_0$, and a N-layer rotation equivariant U-Net network $\mathrm{UNet}_{eq}(\cdot)$, whose channel number of the l^{th} layer is n_l, rotation equivariant subgroup is $S \leqslant O(2), |S| = t$, and activation function is set as ReLU. If the latent continuous function of the c^{th} channel of X denoted as $r_c : \mathbb{R}^2 \to \mathbb{R}$, and the latent continuous function of any convolution filters in the l^{th} layer denoted as $\phi^l : \mathbb{R}^2 \to \mathbb{R}$, DB_i and UB_i represent the downsampling block and the upsampling block, respectively. $\hat{\Psi}$, $\hat{\Phi}$, and $\hat{\Upsilon}$ represent the convolutional layers in the input, middle, and output stages, respectively. We define:*

$$\mathrm{UNet}_{eq}(\cdot) = \hat{\Upsilon}\left[\hat{UB}_m \cdots \hat{UB}_1\left[\hat{\Phi}\left[\hat{DB}_m \cdots \hat{DB}_1\left[\hat{\Psi}\right]\right]\cdots\right]\right](\cdot), \tag{14}$$

the following conditions are satisfied:

$$\begin{aligned} &|r_c(x)| \leq F_0, \|\nabla_x r_c(x)\| \leq G_0, \|\nabla_x^2 r_c(x)\| \leq H_0, \\ &|\phi^l(x)| \leq F_l, \|\nabla_x \phi^l(x)\| \leq G_l, \|\nabla_x^2 \phi^l(x)\| \leq H_l, \\ &\forall \|x\| \geq (p+1)h/2, \phi_l(x) = 0, \end{aligned} \tag{15}$$

where p is the filter size, h is the mesh size, $\theta_k = \frac{2k\pi}{t}$, $k = 1, 2, \cdots, t$. ∇_x and ∇_x^2 denote the operators of gradient and Hessian matrix, respectively. We have

$$\left|\mathrm{UNet}_{eq}\left[\tilde{\pi}_{\theta_k}^R\right](X) - \tilde{\pi}_{\theta_k}^R\left[\mathrm{UNet}_{eq}\right](X)\right| \leq R_1 h + R_2 h^2. \tag{16}$$

where R_1, R_2 are two constants with respect to N, n_l and the upper bound in (15), their specific values can be found in the supplementary materials.

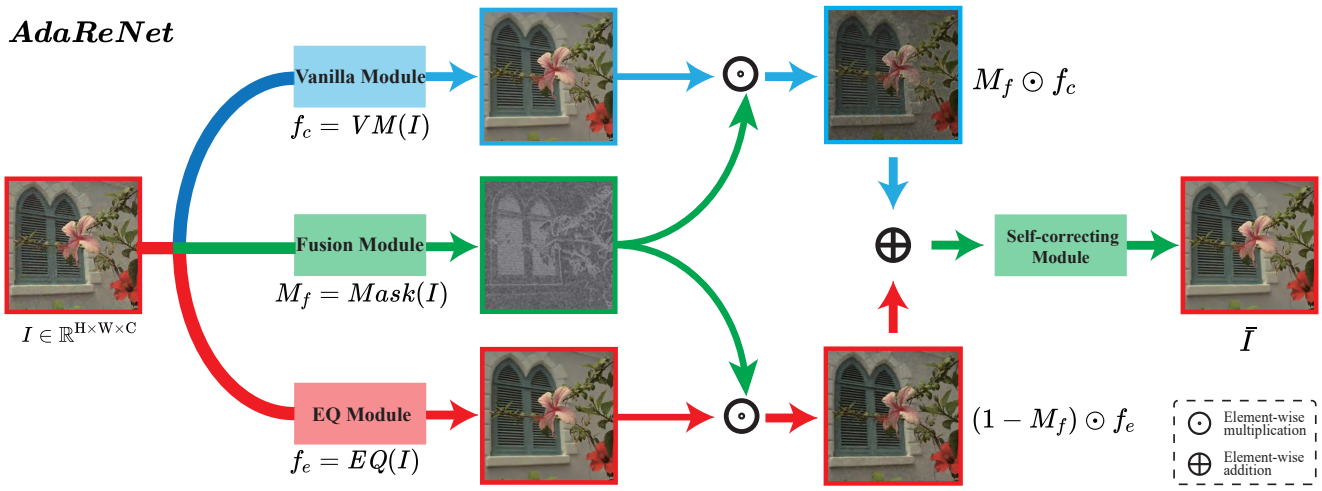


Figure 3. Illustrations of our proposed adaptive network AdaReNet. Specifically, $I \in \mathbb{R}^{H\times W\times C}$ represents a noisy image, where H and W represent the spatial dimensions, and C denotes the channel dimension. The Vanilla Module and EQ Module each produce their respective preliminary denoising results, denoted as f_c and f_e. The Fusion Module $Mask(\cdot)$ automatically decides which areas of the image to use more EQ Module would gain more benefit. After adaptive fusion by $Mask(\cdot)$ and correction by the Self-correcting Module $S_c(\cdot)$, the final denoised image $\bar{I}$ is output.

Corollary 1 *Under the same condition as Theorem 3, for an arbitrary $\theta \in [0, 2\pi]$, let π_θ denote the rotation transformation, then $\forall\theta$ we have*

$$\left|\mathrm{UNet}_{eq}\left[\tilde{\pi}_\theta^R\right](X) - \tilde{\pi}_\theta^R\left[\mathrm{UNet}_{eq}\right](X)\right| \leq R_1 h + R_2 h^2 + R_3 t^{-1} h, \tag{17}$$

where R_1, R_2, R_3 are constants that can be found in the supplementary materials.

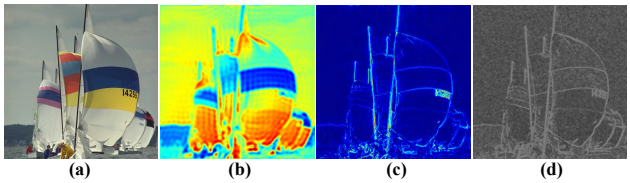


Figure 4. (a) An image from the Kodak dataset, (b) the heatmap of the low-frequency component, (c) the heatmap of the high-frequency component, (d) the output of our proposed MaskNetwork (the brighter area indicates the use of more Vanilla Module).

3.2. Proposed Adaptive Network AdaReNet

Rotation equivariant design usually degrades the representation accuracy of the network because of parameter sharing and filter parameterization and not all areas of natural images strictly comply with rigid rotation equivariance. In order to solve the above problems, the proposed network predominantly comprises four primary modules, as depicted in Fig. 3, which are the Vanilla Module, EQ Module, Fusion Module, and Self-correcting Module.

Vanilla Module. This module utilizes a conventional CNN architecture, maintaining the original design where only translation equivariance is incorporated into the network. We define the result of the input image I after passing through the Vanilla Module as f_c.

$$f_c = \mathrm{VM}(I). \tag{18}$$

EQ Module. This module employs the same architecture as the previous one, with the conventional convolution replaced by the rotation equivariant network to incorporate rotation equivariance prior into the architecture. We define the result of input image I after passing through the EQ Module as f_e,

$$f_e = \mathrm{EQ}(I). \tag{19}$$

Fusion Module. This module introduces a MaskNetwork $Mask(\cdot)$, a specialized network designed to merge the outputs from the Vanilla Module and the EQ Module. It employs several layers of standard convolutions to automatically determine which regions of the image would benefit more from a rotation-equivariant network compared to a normal CNN-based network. The output of $Mask(\cdot)$ is shown in Fig. 4(d). We define the result of input image I after passing through MaskNetwork as M_f,

$$M_f = \mathrm{Mask}(I). \tag{20}$$

Self-correcting Module. Following the Fusion Module, a self-correcting module is applied to refine the fused result. This is achieved through the use of simple ResNet Blocks. We define the Self-correcting Module as $S_c(\cdot)$.

Based on the network presented in Fig. 3, we can derive the following mathematical representation to encapsulate the training and inference processes of the entire network:

$$\hat{I} = M_f \odot f_c + (1 - M_f) \odot f_e, \bar{I} = S_c(\hat{I}), \tag{21}$$

where $\hat{I}$ denotes the results of adaptive fusion, and $\bar{I}$ represents the final restoration result after going through the Self-correcting Module, $\odot$ denote element-wise multiplication.

Loss Function. We incorporate the loss of the two subnetworks as regularization terms into the main loss. The loss function is defined as follows:

$$\mathrm{L} = \|\bar{I} - \mathrm{target}\|_2 + \alpha_1 \|f_c - \mathrm{target}\|_2 + \alpha_2 \|f_e - \mathrm{target}\|_2, \tag{22}$$

where α_1, α_2 are hyperparameters and we empirically set $\alpha_1 = \alpha_2 = 0.1$.

Remark. Embedding rotational equivariance into the network is highly effective for self-supervised image denoising task. Since natural images do not adhere to strict rotational equivariance, our proposed adaptive rotational equivariant network, named AdaReNet, can automatically decide which regions of the image to apply the rotation-equivariant network, thereby further enhancing performance. By observing the output of the Fusion Module in Fig. 4(d), it is noted that the mask values are larger at the edge details of the pattern, indicating that the adaptive equivariant network tends to use the outputs from the Vanilla Module more frequently in these areas. Conversely, for the restoration of the majority of low-frequency components in the image, the network predominantly uses the outputs from the EQ Module. This aligns with the common sense that convolutions are more adept at fitting high-frequency information[43].

4. Experiments

In this section, we conducted experiments based on the existing setup and validated our method in three classic approaches. For the Noise2Noise [26] and Noise2Void [21] methods, we showed experiments on U-Net [38], which can significantly speed up the training process while maintaining acceptable performance, and demonstrated the effectiveness of our method. Besides, we also conducted the R2R experiments based on DnCNN [33], further validating the superiority of the proposed method. Due to space limitations, we only present partial results. Implementation details and further experiments concerning different models, various datasets, model parameter counts, as well as experiments in the field of self-supervised fluorescence microscopy denoising[27] can be found in supplementary materials.

4.1. Experiments on Compared Methods

Rotation Equivariant N2N: We selected the most classical Gaussian noise for our experiments, and randomized the standard deviation $\sigma \in [0, 50]$ of noise for each training example individually. N2N-EQ denotes the rotation equivariant network, while N2N-EQ$^+$ serves as the adaptive rotation equivariant network. The notation in other experiments is similar. The results are shown in Tab. 1. The superiority of our method can also be observed from Fig. 5.

Dataset	Gaussian25			Gaussian50		
	N2N [26]	N2N-EQ	N2N-EQ$^+$	N2N [26]	N2N-EQ	N2N-EQ$^+$
Kodak [10]	31.47/0.874	31.60/0.878	**31.72/0.880**	28.29/0.778	28.58/0.790	**28.69/0.791**
BSD300 [31]	30.18/0.869	30.28/0.872	**30.36/0.873**	27.02/0.762	27.24/0.771	**27.31/0.772**
Set14 [54]	30.02/0.851	30.06/0.854	**30.19/0.855**	27.16/0.768	27.32/0.775	**27.44/0.777**

Table 1. N2N: three networks with U-Net architecture were tested under conditions of Gaussian noise at levels 25 and 50.

Rotation Equivariant N2V: We conducted experiments with U-Net architectures. The results are presented in Tab. 2 and Fig. 6. The equivariant error[1] of networks N2V, N2V-EQ, and N2V-EQ$^+$ are 0.233, 0.068, and 0.076, respectively. This verifies that the improvements are achieved by reducing equivariant errors. Besides, Tab. 3 further performed diverse scenarios trained on the BSD500 dataset. Our method consistently achieved superior results.

Dataset	Gaussian25			Gaussian50		
	N2V [21]	N2V-EQ	N2V-EQ$^+$	N2V [21]	N2V-EQ	N2V-EQ$^+$
BSD500 [31]	28.17/0.820	29.05/0.834	**29.12/0.845**	26.07/0.725	26.38/0.735	**26.82/0.755**
Kodak24 [10]	28.86/0.811	29.78/0.825	**29.93/0.836**	26.75/0.716	27.15/0.732	**27.72/0.754**
Set14 [54]	27.22/0.800	28.04/0.806	**28.09/0.816**	25.40/0.715	25.65/0.726	**26.23/0.749**
Average	28.08/0.811	28.96/0.822	**29.05/0.832**	26.07/0.719	26.39/0.731	**26.92/0.753**

Table 2. N2V: three networks were tested under conditions of Gaussian noise at levels 25 and 50.

Method	Poisson	Poisson30	Poisson&Gaussian	Peppersalt	Speckle
N2V	30.82/0.912	32.38/0.961	27.91/0.821	23.93/0.782	26.03/**0.775**
N2V-EQ(ours)	31.22/0.904	33.59/0.966	28.84/0.835	24.53/0.792	**26.42**/0.736
N2V-EQ$^+$(ours)	**32.19/0.924**	**35.96/0.976**	**29.34/0.850**	**24.83/0.811**	24.85/0.676

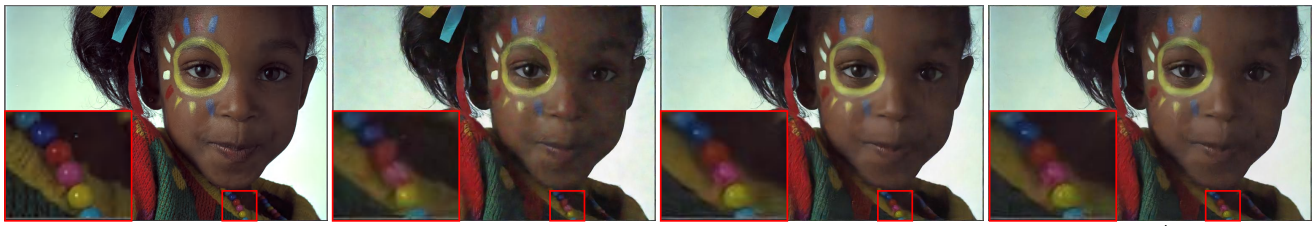
Table 3. Quantitative comparison on diverse scenarios.

Rotation Equivariant R2R: The experimental results are shown in Tab. 5. The parameter counts for methods R2R, R2R-EQ, and R2R-EQ$^+$ are 0.67M, 0.17M and 0.84M, respectively. Due to the larger number of adaptive network parameters, it is expected that the performance is inferior to our EQ version when the training dataset is small.

4.2. Ablation Study

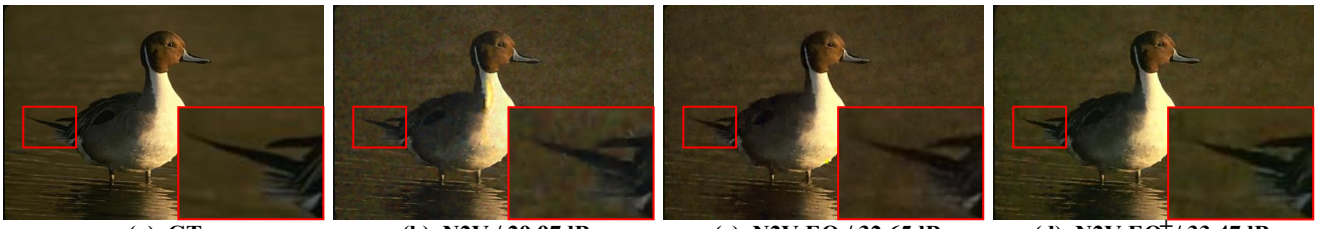
Selection of rotation-equivariant networks: We conducted ablation experiments with mainstream rotation-equivariant networks, including Cohen *et al*. [8], Weiler *et al*. [45] and Shen *et al*. [39]. Tab. 4 confirms that the chosen Fconv [49] approach is the most effective method in the field of low-level vision.

[1] $\|[L_R\Phi(f) - \Phi L_R(f)\|_2^2 / \|L_R\Phi(f)\|_2^2$, where $\Phi(\cdot)$ represents the network and $L_R(\cdot)$ denotes the rotation transformation.



(a) GT (b) N2N / 28.82dB (c) N2N-EQ / 29.26dB (d) N2N-EQ$^+$/ 29.41dB

Figure 5. N2N: image denoising results of one image from kodak with $\sigma = 50$.



(a) GT (b) N2V / 29.97dB (c) N2V-EQ / 32.65dB (d) N2V-EQ$^+$/ 33.47dB

Figure 6. N2V: image denoising results of one image from BSD500 with $\sigma = 25$.

Method	Gaussian25								Gaussian50							
	Kodak24 [10]		BSDS300 [31]		Set14 [54]		Average		Kodak24 [10]		BSDS300 [31]		Set14 [54]		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CNN	31.47	0.874	30.18	0.869	30.02	0.851	30.56	0.865	28.29	0.778	27.02	0.762	27.16	0.768	27.49	0.769
G-CNN	31.39	0.872	30.23	0.869	30.02	0.852	30.55	0.864	28.04	0.774	26.85	0.757	26.88	0.762	27.26	0.764
E2-CNN	31.23	0.869	30.02	0.864	29.70	0.845	30.32	0.859	28.04	0.768	26.86	0.754	26.83	0.758	27.24	0.760
PDO-eConv	31.42	0.874	30.15	0.869	29.83	0.849	30.47	0.864	28.34	0.783	27.06	0.765	27.05	0.768	27.48	0.772
Fconv	**31.60**	**0.878**	**30.28**	**0.872**	**30.06**	**0.854**	**30.65**	**0.868**	**28.58**	**0.790**	**27.24**	**0.771**	**27.32**	**0.775**	**27.71**	**0.779**

Table 4. Ablation on rotation-equivariant networks: PSNR and SSIM results for Gaussian25 and Gaussian50.

Dataset	σ	R2R [33]	R2R-EQ	R2R-EQ$^+$
BSD68	25	29.03/0.822	**29.14**/0.825	29.10/**0.826**
BSD68	50	26.01/0.700	**26.14/0.711**	26.00/0.700

Table 5. R2R: three networks were tested under conditions of Gaussian noise at levels 25 and 50.

Dataset	Gaussian25			Gaussian50		
	N2V [21]	N2V-EQ	N2V-EQ$^+$	N2V [21]	N2V-EQ	N2V-EQ$^+$
BSD500 [31]	23.09/0.758	28.19/0.811	**29.05/0.847**	22.20/0.663	26.06/0.718	**26.47/0.745**
Kodak24 [10]	22.77/0.758	28.98/0.809	**29.82/0.840**	22.72/0.657	26.96/0.721	**27.32/0.743**
Set14 [54]	22.03/0.736	27.07/0.787	**28.04/0.820**	19.65/0.609	25.27/0.707	**25.67/0.731**
Average	22.63/0.751	28.08/0.802	**28.97/0.836**	21.52/0.643	26.10/0.715	**26.49/0.740**

Table 6. N2V w/o rotation augmentation: three networks were tested under conditions of Gaussian noise at levels 25 and 50.

Data rotation augmentation: Rotational data augmentation on training data is a commonly used method to enhance model performance and robustness. We conducted ablation experiments with rotation augmentation in the N2V method. Tab. 2 shows the experiments on the U-Net network with rotation augmentation, while Tab. 6 presents our ablation experiments without augmentation.

5. Conclusion

In this work, we first explore and introduce rotation equivariant image prior into the self-supervised image denoising task at the network architecture level. Through rigorous theoretical analysis, we prove that simply replacing the convolution layers with ECNNs leads to a rotational equivariant version of the network. Building on this, we propose AdaReNet, an adaptive rotation equivariant network, which further enhances performance. Extensive experiments demonstrate the effectiveness of our approach, confirming that incorporating rotation equivariant prior significantly improves denoising results. Overall, our work underscores the importance of leveraging rotation invariance for self-supervised learning and sets a foundation for future research in this field.

Acknowledgement. This research was supported by NSFC project under contract U21A6005; the Major Key Project of PCL under Grant PCL2024A06; Tianyuan Fund for Mathematics of the National Natural Science Foundation of China (Grant No.12426105) and Key Research and Development Program (Grant No. 2024YFA1012000).

References

[1] Josue Anaya and Adrian Barbu. Renoir–a dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation*, 51:144–154, 2018. 1

[2] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3155–3164, 2019. 1, 3

[3] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pages 524–533. PMLR, 2019. 1, 3

[4] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, pages 60–65. Ieee, 2005. 2

[5] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. Non-local means denoising. *Image Processing On Line*, 1: 208–212, 2011. 2

[6] Sungmin Cha, Taeeon Park, and Taesup Moon. Gan2gan: Generative noise learning for blind image denoising with single noisy images. *arXiv preprint arXiv:1905.10488*, 3, 2019. 1, 2, 3

[7] Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu. Spatial-adaptive network for single image denoising. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXX 16*, pages 171–187. Springer, 2020. 1, 3

[8] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016. 3, 7

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 2

[10] Rich Franzen. Kodak lossless true color image suite. *source: http://r0k. us/graphics/kodak*, 4(2):9, 1999. 7, 8

[11] Jiahong Fu, Qi Xie, Deyu Meng, and Zongben Xu. Rotation equivariant proximal operator for deep unfolding methods in image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2, 4

[12] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014. 2

[13] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International journal of computer vision*, 121:183–208, 2017. 2

[14] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1712–1722, 2019. 1, 3

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[16] R Mark Henkelman. Measurement of signal intensities in the presence of noise in mr images. *Medical physics*, 12(2): 232–233, 1985. 1

[17] Emiel Hoogeboom, Jorn WT Peters, Taco S Cohen, and Max Welling. Hexaconv. In *International Conference on Learning Representations*, 2018. 3

[18] Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14781–14790, 2021. 1, 2, 3

[19] Xixi Jia, Sanyang Liu, Xiangchu Feng, and Lei Zhang. Focnet: A fractional optimal control network for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6054–6063, 2019. 3

[20] Yoonsik Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3482–3492, 2020. 1

[21] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2129–2137, 2019. 1, 3, 7, 8

[22] Alexander Krull, Tomáš Vičar, Mangal Prakash, Manan Lalit, and Florian Jug. Probabilistic noise2void: Unsupervised content-aware denoising. *Frontiers in Computer Science*, 2:5, 2020. 1

[23] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989. 4

[24] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 4

[25] Stamatios Lefkimmiatis. Universal denoising networks: a novel cnn architecture for image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3204–3213, 2018. 3

[26] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 1, 3, 5, 7

[27] Jizhihui Liu, Qixun Teng, and Junjun Jiang. Fm2s: Self-supervised fluorescence microscopy denoising with single noisy image. *arXiv preprint arXiv:2412.10031*, 2024. 7

[28] Zhiliang Liu and Liyuan Ren. Shaking noise exploration and elimination for detecting local flaws of steel wire ropes based on magnetic flux leakages. *IEEE Transactions on Industrial Electronics*, 70(4):4206–4216, 2022. 1

[29] Ymir Mäkinen, Lucio Azzari, and Alessandro Foi. Exact transform-domain noise variance for collaborative filtering of stationary correlated noise. In *2019 IEEE international conference on image processing (ICIP)*, pages 185–189. IEEE, 2019. 2

[30] Diego Marcos, Michele Volpi, Nikos Komodakis, and Devis Tuia. Rotation equivariant vector field networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5048–5057, 2017. 3

[31] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, pages 416–423. IEEE, 2001. 7, 8

[32] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2noise: Learning to denoise from unpaired noisy data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12064–12072, 2020. 1, 2, 3

[33] Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: unsupervised deep learning for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2043–2052, 2021. 1, 2, 3, 7, 8

[34] Bumjun Park, Songhyun Yu, and Jechang Jeong. Densely connected hierarchical network for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019. 1

[35] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1586–1595, 2017. 1

[36] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1890–1898, 2020. 3

[37] Yuhui Quan, Yixin Chen, Yizhen Shao, Huan Teng, Yong Xu, and Hui Ji. Image denoising using complex-valued deep cnn. *Pattern Recognition*, 111:107639, 2021. 3

[38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 7

[39] Zhengyang Shen, Lingshen He, Zhouchen Lin, and Jinwen Ma. Pdo-econvs: Partial differential operator based equivariant convolutions. In *International Conference on Machine Learning*, pages 8697–8706. PMLR, 2020. 2, 3, 7

[40] Zhengyang Shen, Tiancheng Shen, Zhouchen Lin, and Jinwen Ma. Pdo-es2cnns: Partial differential operator based equivariant spherical cnns. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9585–9593, 2021. 3

[41] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 3

[42] Raviteja Vemulapalli, Oncel Tuzel, and Ming-Yu Liu. Deep gaussian conditional random field network: A model-based deep network for discriminative denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4801–4809, 2016. 1

[43] Haohan Wang, Xindi Wu, Zeyi Huang, and Eric P Xing. High-frequency component helps explain the generalization of convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8684–8694, 2020. 7

[44] Zejin Wang, Jiazheng Liu, Guoqing Li, and Hua Han. Blind2unblind: Self-supervised image denoising with visible blind spots. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2027–2036, 2022. 1, 3

[45] Maurice Weiler and Gabriele Cesa. General e (2)-equivariant steerable cnns. *Advances in neural information processing systems*, 32, 2019. 3, 7

[46] Maurice Weiler, Fred A Hamprecht, and Martin Storath. Learning steerable filters for rotation equivariant cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 849–858, 2018. 2, 3

[47] Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5028–5037, 2017. 3

[48] Xiaohe Wu, Ming Liu, Yue Cao, Dongwei Ren, and Wangmeng Zuo. Unpaired learning of deep image denoising. In *European conference on computer vision*, pages 352–368. Springer, 2020. 1

[49] Qi Xie, Qian Zhao, Zongben Xu, and Deyu Meng. Fourier series expansion based filter parametrization for equivariant convolutions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4537–4551, 2022. 2, 3, 4, 7

[50] Jun Xu, Hui Li, Zhetong Liang, David Zhang, and Lei Zhang. Real-world noisy image denoising: A new benchmark. *arXiv preprint arXiv:1804.02603*, 2018. 1

[51] Jun Xu, Yuan Huang, Ming-Ming Cheng, Li Liu, Fan Zhu, Zhou Xu, and Ling Shao. Noisy-as-clean: Learning self-supervised denoising from corrupted image. *IEEE Transactions on Image Processing*, 29:9316–9329, 2020. 1, 2, 3

[52] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16*, pages 41–58. Springer, 2020. 3

[53] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020. 3

[54] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and*

Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7, pages 711–730. Springer, 2012. 7, 8

[55] Dan Zhang, Fangfang Zhou, Felix Albu, Yuanzhou Wei, Xiao Yang, Yuan Gu, and Qiang Li. Unleashing the power of self-supervised image denoising: A comprehensive review. *arXiv preprint arXiv:2308.00247*, 2023. 1, 3

[56] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 1, 3

[57] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 1, 3

[58] Yanzhao Zhou, Qixiang Ye, Qiang Qiu, and Jianbin Jiao. Oriented response networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 519–528, 2017. 3