A Regularization-Guided Equivariant Approach for Image Restoration

Supplementary Material

1. Beneficial of Equivariance for IR

The benefits of equivariance have actually been validated in image restoration (IR) tasks. A classical example is the convolutional neural network (CNN), which has been shown to outperform fully connected networks in IR tasks. One of the fundamental reasons lies in its reasonable implementation of translation equivariance. In other words, the translation symmetry prior of images is well captured. Since IR is often ill-posed and suffers from information insufficiency, embedding translation symmetry prior knowledge is naturally beneficial.

For the rotation equivariance involved in this paper, it can further capture the rotation symmetry prior in images. Similar to the benefits of translation equivariance in CNNs, the reasonable utilization of rotation equivariance can also enhance performance in IR tasks. Previous works, such as in [1] (Fig.1), have shown that rotational symmetry in local features is a fundamental characteristic present in nearly all types of images. Since IR tasks are highly correlated to the representation of local features, the exploration of rotation equivariance is important for IR tasks. For instance, data augmentation through random rotation has been proven to boost performance in almost all IR tasks. Additionally, rotation equivariant CNNs often outperform standard CNNs in IR tasks, as evidenced by [1, 3] and our experiments.

Previous works [1, 3] have provided two intuitive reasons why rotation equivariance benefits IR. First, incorporating rotation equivariance improves the preservation of nonlocal orientational similarity, which ensures a more faithful recovery of local image features. This is clearly demonstrated by a comparative observation of the enlarged view of the plate's rim in Fig.1(b) and (c). Second, rotation equivariance aids in better preserving local isotropic symmetry during high-resolution recovery, as clearly shown by the comparison of the enlarged view of the phone in Fig.1(b) and (c). This property ensures that features with similar characteristics in all directions, such as repetitive textures and isotropic patterns, can be reconstructed consistently without introducing directional artifacts.

2. Features Visualizations

In this section, we provide additional visualizations of feature maps produced by different frameworks, as shown in Fig.2. From left to right, the feature maps are visualized from shallow layer to deep layer. It can be observed that the equivariant constraints imposed by the proposed method align more effectively with the intrinsic symmetry of the data, both in shallow and deep layers. In contrast, the data augmentation strategy does not effectively supervise the intermediate layers of the CNN network. As a result, when the input is rotated 90 degrees, the feature map exhibits unpredictable changes. Consequently, when a circular pattern is input, the CNN's feature map fails to preserve the circular symmetry. This could undermine performance in image restoration tasks. For strictly equivariant networks (EQ-CNN [3]), the feature maps show a perfectly symmetric pattern, but the features are no longer circular. Specifically, due to the strict enforcement of rotational symmetry at specific angles, the feature maps are constrained to exhibit four distinct directions, resulting in an inaccuracy representation of the underlying patterns. In conclusion, the proposed method applies more reasonable rotation-equivariant supervision to the intermediate layers, enabling the network to capture the proper image symmetry priors directly from the data.

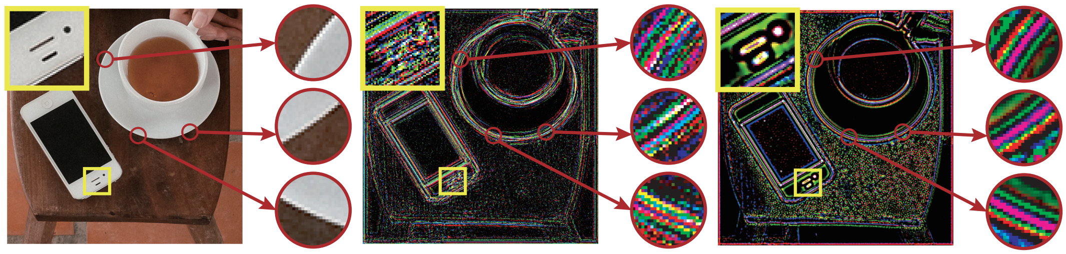
3. Error analysis

In this section, we further discuss and analyze the equivariant error and reconstruction error of the models. As shown in Fig.3, for the equivariant error (left), standard convolutional networks (CNN) fail to capture the rotational equivariance in the data, leading to notable equivariant errors in the trained model. Conversely, strictly equivariant networks (EQ-CNN) eliminate these errors for specific strict rotations, such as 90 degrees. The proposed method (EQ-Reg), however, significantly reduces the equivariant error after training, effectively demonstrating that the model learns the rotational symmetry present in the data. For the reconstruction error (right) between the restoration results of different models and the ground truth, both CNN without rotation equivariance and strict rotation equivariant networks(EQ-CNN) exhibit reconstruction errors that cannot be ignored. However, the proposed method in this paper effectively learns the rotation symmetry inherent in the data, while also achieving results closer to the ground truth.

4. More Experimental Results

4.1. Image SR

We compare the proposed method with classical rotation equivariant methods, including GCNN, E2-CNN and PDO-eConv using SOTA network RDN. We replace the original convolutions in RDN with the competing convolutions respectively. The training set consists of 800 images from the DIV2K dataset. For testing, we use four standard benchmark datasets. All experiments are performed under a bicu-



(a) Input Image (b) CNN output (c) EQ-CNN output

Figure 1. Illustration of the output feature map obtained by standard CNN and rotation equivariant CNN(EQ-CNN) [1].

bic degradation model. The results are shown in Table 1.

Table 1. PSNR/SSIM of competing methods for RDN with x4 SR.

Method	Urban100	BSD100	Set14	Set5
RDN	26.27/0.795	27.63/0.746	28.70/0.790	**32.30/0.899**
RDN-$gcnn$	26.15/0.792	27.63/0.746	28.66/0.789	32.26/0.899
RDN-$e2cnn$	26.06/0.789	27.59/0.744	28.61/0.787	32.09/0.897
RDN-$pdoe$	24.37/0.723	26.85/0.721	27.42/0.759	30.42/0.864
RDN-reg	**26.35/0.797**	**27.66/0.747**	**28.71/0.791**	32.26/**0.899**

4.2. Image Denoising

We designed simulation denoising experiments. The competing methods include DnCNN, Swin-Conv UNet (SCUNet), EDSR and classical rotation equivariant methods. All methods are trained with 800 samples from DIV2K dataset, and the simulated Gaussian noise has a standard deviation of 50. The results are shown in Table 2.

Table 2. PSNR/SSIM of competing methods on image denoising.

Method	Urban100	BSD100	Set14	Set5
DnCNN	28.71/0.844	28.96/0.796	29.01/0.790	30.95/0.857
SCUNet	30.50/0.889	29.55/0.818	29.96/0.816	31.79/0.881
EDSR	30.57/0.888	29.56/0.817	29.86/0.811	31.84/0.881
EDSR-$gcnn$	30.75/**0.891**	29.60/**0.820**	29.89/0.811	31.91/0.881
EDSR-$e2cnn$	30.66/0.889	29.58/0.818	29.79/0.809	31.90/0.882
EDSR-$pdoe$	29.55/0.863	29.19/0.800	29.50/0.803	31.47/0.870
EDSR-reg	**30.77/0.891**	**29.61/0.820**	**29.98/0.814**	**31.92/0.883**

4.3. Generalization Results in Metal Artifact Reduction

To demonstrate the generalization capabilities of the proposed method, we evaluate it on another public dataset, CLINIC-metal [2]. As illustrated in Fig. 4-6, the proposed method surpasses both the baseline and strictly equivariant methods in removing shading and streaking artifacts, while more accurately reconstructing human tissue structures. These results provide a compelling visual validation of the advantages of the proposed method in the generalization ability.

4.4. Image Inpainting Visualizations

In this section, we present additional visual results for the image inpainting task. As shown in Fig.7, increasing the intensity of Poisson noise progressively disrupts the inherent symmetries in the image. Under these conditions, recovery results using strictly equivariant networks become suboptimal, as they enforce rigid equivariant constraints which fail to align with the image's actual symmetry characteristics. In contrast, the proposed method demonstrates superior performance by effectively removing noise and restoring intricate image details.

4.5. Single Image Rain removal Visualizations

In this section, we present additional visual results for the Single Image Rain removal task. As shown in Figure 8, the rain removal model based on convolutional sparse coding often mistakenly identifies white stripes and grids as rain stripes, but the proposed method can effectively overcome this shortcoming.

References

[1] 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. 1, 2

[2] Pengbo Liu, Hu Han, Yuanqi Du, Heqin Zhu, Yinhao Li, Feng Gu, Honghu Xiao, Jun Li, Chunpeng Zhao, Li Xiao, et al. Deep learning to segment pelvic bones: large-scale ct datasets and baseline models. *International Journal of Computer Assisted Radiology and Surgery*, 16:749–756, 2021. 2, 3, 4

[3] 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. 1, 3

[4] Wenhan Yang, Robby T Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(6):1377–1393, 2019. 5

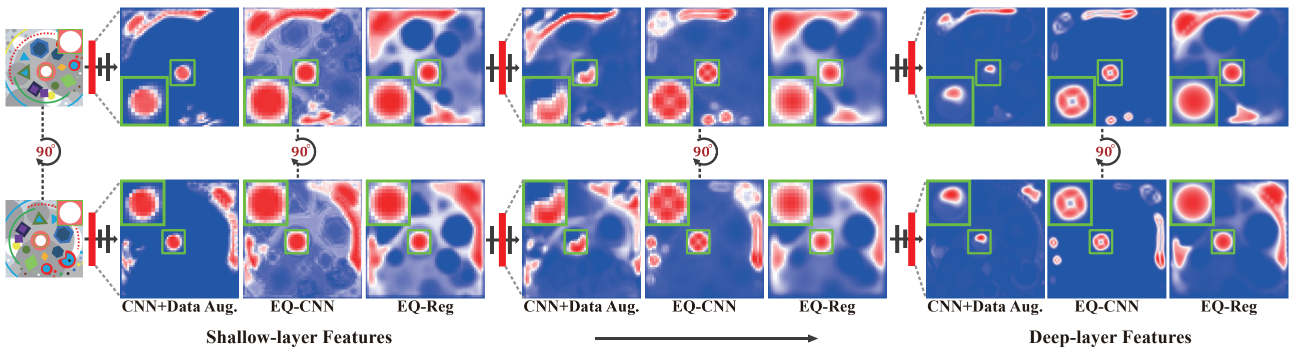


Figure 2. The feature map in the network from shallow layer to deep layer of the trained neural network for CNN+Data Aug., EQ-CNN[3], and the proposed EQ-Reg.

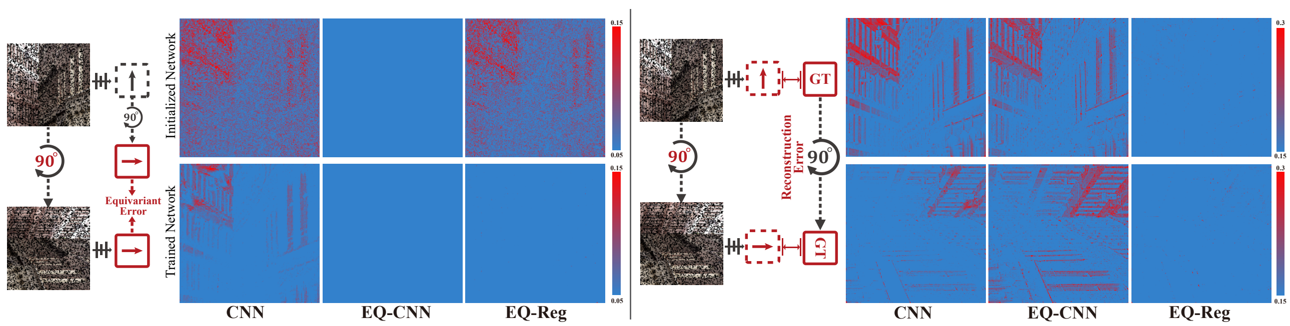


Figure 3. Illustration of (left) rotation equivariant errors($\left|\Phi[\pi^I_{\tilde{A}}](I) - \pi^F_{\tilde{A}}[\Phi](I)\right|$ for image I and network Φ) and (right) reconstruction errors.

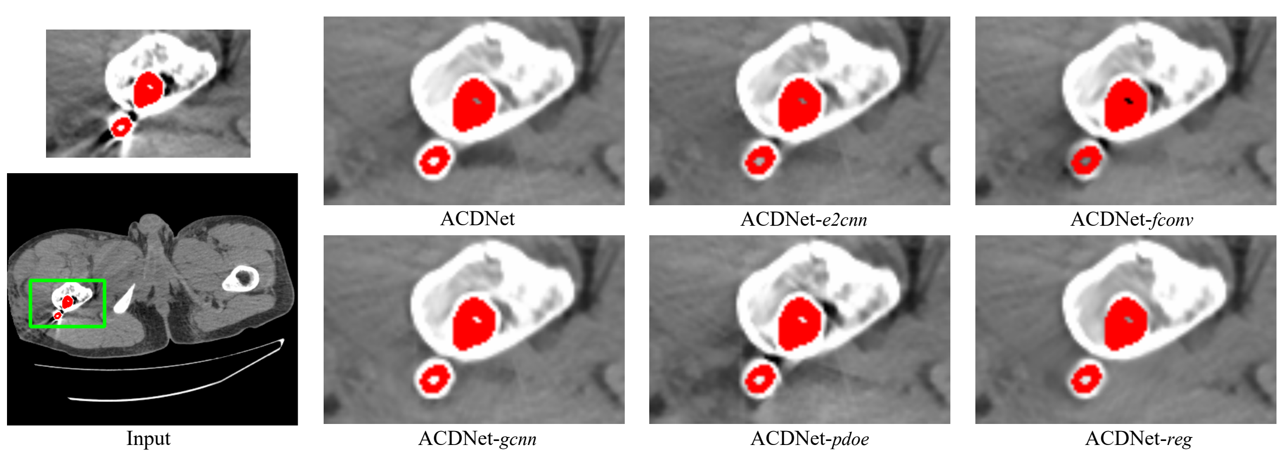


Figure 4. Generalization results on a real clinical metal-affected CT image from CLINIC-metal [2]. The red pixels stand for metallic implants, which are segmented with the thresholding of 2500 HU.

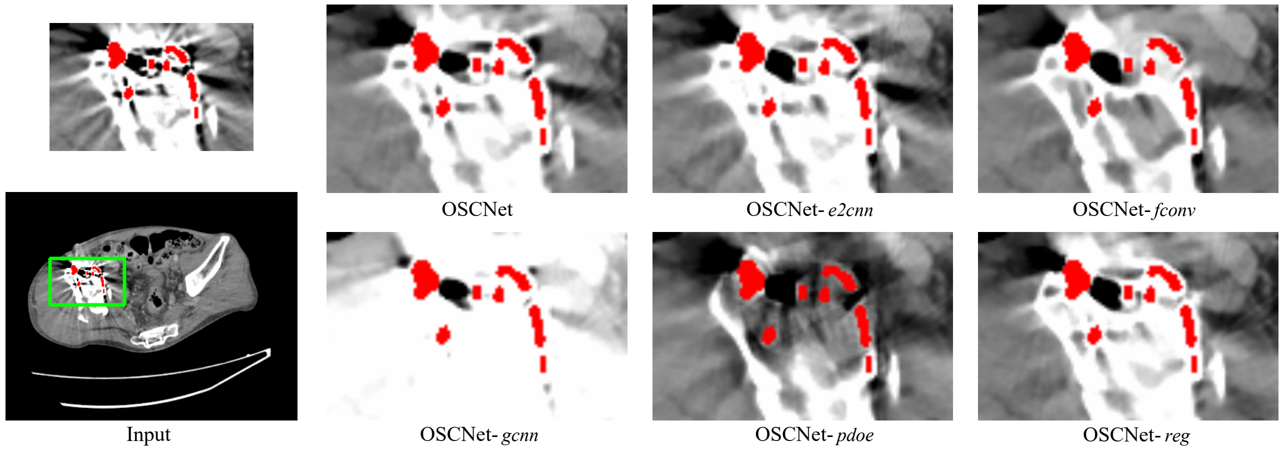


Figure 5. Generalization results on a real clinical metal-affected CT image from CLINIC-metal [2]. The red pixels stand for metallic implants, which are segmented with the thresholding of 2500 HU.

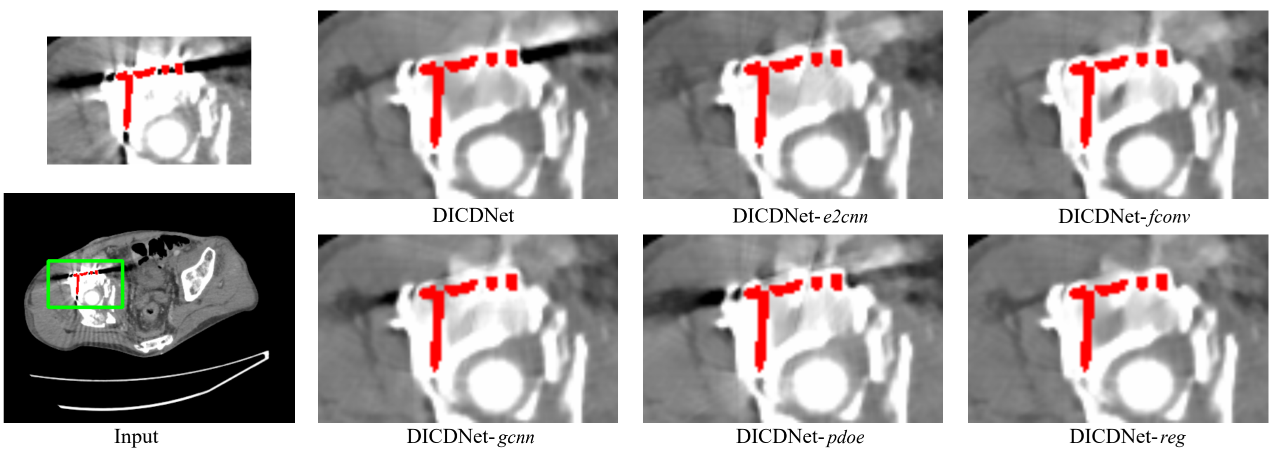


Figure 6. Generalization results on a real clinical metal-affected CT image from CLINIC-metal [2]. The red pixels stand for metallic implants, which are segmented with the thresholding of 2500 HU.

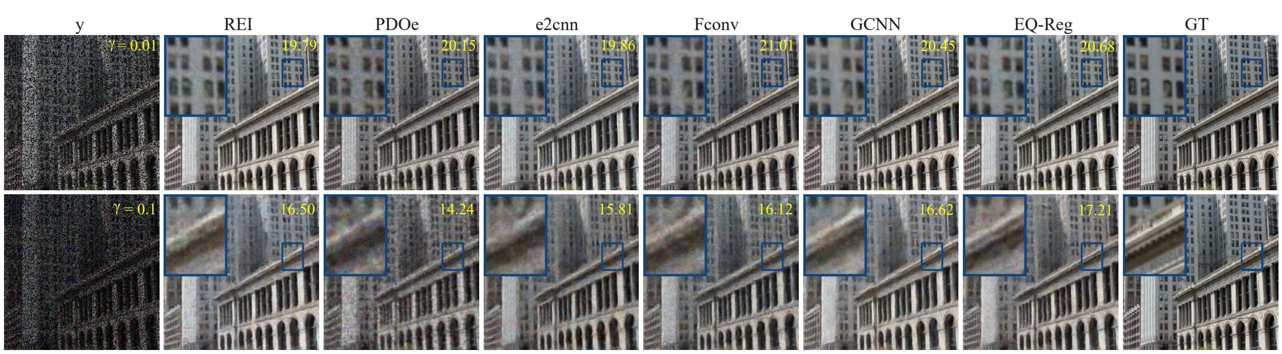


Figure 7. Inpainting reconstructions on test images with Poisson noise (from top to bottom: γ = 0.01, 0.1) and 30% mask rate. PSNR values are shown in the top right corner of the images.

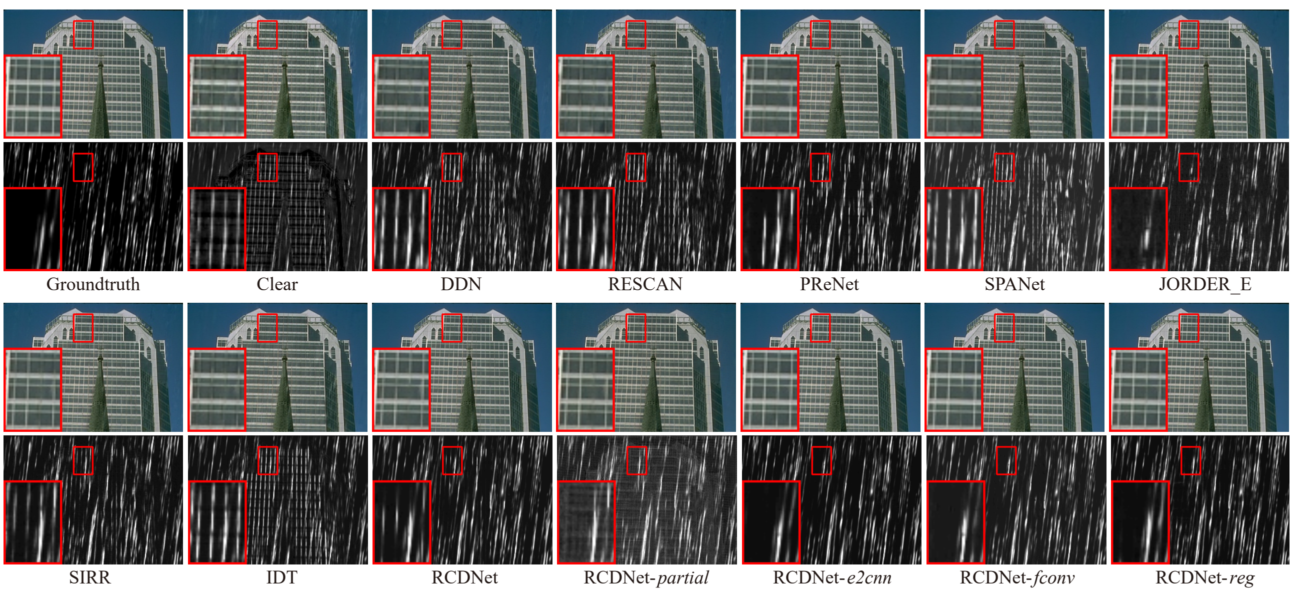


Figure 8. The 1^{st} column: a typical ground truth sample in Rain100L [4] dataset (upper) and its ground truth rain layer (lower). The $2^{nd} - 14^{th}$ columns: derained results (upper) and extracted rain layers (lower) by all competing methods