DRCNet: Dynamic Image Restoration Contrastive Network

Fei Li[1,3,4,*], Lingfeng Shen[2,*], Yang Mi[1,4], and Zhenbo Li[1,3,4,5,6]✉

[1] College of Info. and Electr. Eng., China Agricultural University, China
{leefly072,miy,lizb}@cau.edu.cn
[2] Department of Computer Science, Johns Hopkins University, USA
lshen30@jh.edu
[3] National Innovation Center for Digital Fishery, Ministry of Agriculture and Rural Affairs, China
[4] Key Lab. of Agricultural Info. Acquisition Tech., Ministry of Agriculture and Rural Affairs, China
[5] Key Lab. of Smart Breeding Tech., Ministry of Agriculture and Rural Affairs, China
[6] Precision Agriculture Techn. Integrated Scientific Experiment Base, Ministry of Agriculture and Rural Affairs, China

Abstract. Image restoration aims to recover images from spatially-varying degradation. Most existing image-restoration models employed static CNN-based models, where the fixed learned filters cannot fit the diverse degradation well. To this end, we propose a novel **D**ynamic Image **R**estoration **C**ontrastive **Net**work (DRCNet) to address this issue. The principal block in DRCNet is the **D**ynamic **F**ilter **R**estoration module (DFR), which mainly consists of the spatial filter branch and the energy-based attention branch. Specifically, the spatial filter branch suppresses spatial noise for varying spatial degradation; the energy-based attention branch guides the feature integration for better spatial detail recovery. To make degraded images and clean images more distinctive in the representation space, we develop a novel **Intra**-class **C**ontrastive **R**egularization (Intra-CR) to serve as a constraint in the solution space for DRCNet. Meanwhile, our theoretical derivation proved Intra-CR owns less sensitivity towards hyper-parameter selection than previous contrastive regularization. DRCNet outperforms previous methods on the ten widely used benchmarks in image restoration. Besides, the ablation studies investigate the impact of the DFR module and Intra-CR, respectively.

Keywords: Image restoration, Dynamic convolution, Contrastive regularization.

1 Introduction

Image restoration (IR) is one of the basic tasks in computer vision, which recovers clean images from degraded versions, typically caused by rain [12], noise [25]

* Equal contribution. ✉ Corresponding author.

and blur [23]. It is imperative to restore such degraded images to improve their visual quality. Among the models for IR, most of the achieved progress [24,29] is primarily attributed to static Convolutional Neural Networks (CNN) [16]. However, the image degradation is spatially varying [33], which is incompatible with static CNN that are in a filter sharing manner across spatial domains [6].

Therefore, static CNN-based approaches perform imperfectly when the input image contains noise pixels, as well as severe intensity distortions in different spatial regions [33]. To be specific, static CNN-based [17,14] models have some drawbacks. First, the typical CNN filter is spatial-invariance and content-agnostic, leading to the sub-optimal in IR [67,33]. Second, the fixed learned filters can not automatically fit the diverse input degraded images [17,48]. Considering the limitations mentioned above, we need to design a module to dynamically restore the degraded images since each input image has a variable degree of distortion and specific spatial distribution.

Recently, some efforts [6,48,67] have been made to compensate for the drawbacks of static convolution, enabling the model to flexibly adjust the structure and parameters to be suitable for diverse task demands. Few works [33] have employed dynamic convolution for region-level restoration, which may not effectively reconstruct the fine-grained pixels. To solve this, we propose a new model called **D**ynamic Image **R**estoration **C**ontrastive **Net**work (DRCNet), which consists of two key components: Dynamic Filter Restoration module (DFR) and Intra-Class Contrastive Regularization (Intra-CR). The core component of DRCNet is DFR, which effectively restores the pixel-level spatial details by using the dynamic mask to suppress spatial noise and applying feature integration. Specifically, there are two principal designs in DFR. One is a spatial filter branch, which masks the noise pixels and applies adaptive feature normalization. The other one is the energy-based attention branch, which is designed to calibrate features dynamically. Moreover, to make degraded images and clean images more distinctive in the representation space, we propose a new contrastive regularization called Intra-CR, serving as a constraint in the solution space. Specifically, Intra-CR constructs negative samples through mixup [62] while existing Contrastive Regularization (CR) construct negative samples by random sampling. Its effectiveness is validated through theoretical derivation and empirical studies.

To summarize, the main contributions of this study are as follows:

- We propose a Dynamic Filter Restoration module **(DFR)** that is adaptive in various image restoration scenes. Such a block enables DRCNet to handle spatial-varying image degradation.
- A novel contrastive regularization is proposed, dubbed **Intra-CR**, to construct intra-class negative samples through mixup. Empirical studies show its superiority over vanilla contrastive regularization, and our theoretical results show that Intra-CR is less sensitive to hyper-parameter selection.
- Extensive experimental results on ten image restoration baselines demonstrate the efficacy of the proposed DRCNet, which achieves state-of-the-art performance.

2 Related Work

2.1 Image Restoration

Early image restoration approaches are based on prior-based models [16,46], sparse models [28], and physical models [4]. Recently, the significant performance improvements in image restoration can be attributed to the architecture of Convolution Neural Networks (CNN) [32,55]. Most CNN-based methods focus on elaborating architecture designs, such as, multi-stage networks [55,60], dense connections[32], and Neural Architecture Search (NAS) [57]. Due to the spatial-varying image degradation, static CNN-based models are less capable than desired to handle this issue [33]. In contrast, we propose the DFR module with dynamic spatial filter and energy-based attention (EA), which is more effective than static CNN.

The most relevant work to our work is SPAIR [33]. However, there are several principal differences between DRCNet and SPAIR. **Operations:** SPAIR is a two-stage framework for IR. In contrast, DFR is a plug-in module that can be easily inserted into any CNN. **Mask construction:** A pre-trained network generates the mask of SPAIR, and it mainly captures the location information of degraded pixels. In contrast, DFR directly generates a mask based on a spatial response map, which also detects degradation intensity automatically. **Spatial adaptability:** The sparse convolution receptive field of SPAIR is limited, and adaptive global context aggregation is performed on degraded pixel locations. In contrast, DFR utilizes adaptive feature normalization and a set of learnable affine parameters to gather relevant features from the whole image. **Attention weight:** SPAIR can not directly build connections between two spatial pixels. It produces weights by conducting a pairwise similar process from four directions. In contrast, DRF obtains weights by utilizing EA (i.e., considering both spatial and channel dimensions) and calibrates features dynamically. **Loss function:** SPAIR utilized $\mathcal{L}_{CE}$ and $\mathcal{L}_1$. In contrast, DRCNet proposes a new Intra-CR loss which utilizes the negative sample and outperforms SPAIR.

2.2 Dynamic Filter

Compared to standard convolution, the dynamic filters can achieve dynamic restoration towards different input features. With the key idea of adaptive inference, dynamic filters are applied to various tasks, such as image segmentation [41], super-resolution [47] and restoration [33]. Dynamic filters can be divided into scale-adaptive [66] and spatially-adaptive filters [32,33]. DFR belongs to the spatially-adaptive category, which can adjust filter values to suit different input features. In particular, dynamic spatially-adaptive filters, such as DRConv [6], DynamicConv [8] and DDF [67], can automatically assign multiple filters to corresponding spatial regions. However, most dynamic filters are not specifically designed for image restoration, which results in imperfect performance.

2.3 Contrastive Regularization

Contrastive learning (CL) is a self-supervised representation learning paradigm [26] which is based on the assumption that good representation should bring similar images closer while pushing away dissimilar ones. Most existing works often use CL in high-level vision tasks [15]. While some works [44] have demonstrated that contrastive learning can be used as a regularization to remove the haze. Such CR considers all other images in the batch as negative samples, which may lead to sub-optimal performance. Further, only a few works consider that Intra-class CR can improve the generalization. Therefore, in this paper, we construct a new CR method to improve the model generalization for image restoration. The essential distinction between Intra-CR and existing CR is how the negative examples are constructed. Specifically, we construct negative samples by a mixup [62] operation between the clean image and its degraded version.

3 Methods

In this section, we first provide an overview of DRCNet. Then, we detail the proposed DFR module and Intra-CR.

3.1 Dynamic Filter Restoration Network

Due to two- and multi-stage UNet are proven to be effective in encoding broad contextual information [7,18,29,55]. Dynamic Image Restoration Contrastive Net (DRCNet) is designed of two encoder-decoder sub-networks with four down-sampling and up-sampling operations. The overview of DRCNet is shown in Fig. 1. The sub-network first adopts a 3×3 convolution to extract features. Then, the features are processed with four DFR modules for suppressing the degraded pixels and extracting the clean feature in encoders. We employ three ResBlocks [17] in the decoder to reconstruct images with fine spatial details. The restored images are obtained by using a 3×3 convolution to process the decoder output. To link the two sub-networks, we utilize the Cross-Stage Feature Fusion (CSFF) module and Supervised Attention Module (SAM) [55] to fuse the features, which are highlighted by the red dotted lines and green line as illustrated in Fig. 1. Finally, we propose Intra-CR, which serves as a regularization to pull away degraded images and get close to clean images in the representation space.

3.2 Dynamic Filter Restoration Module

The structure of DFR is shown in Fig. 1. The DFR module aims to automatically suppress potential degraded pixels and generate better spatial detail recovery with fewer parameters. Generally, it achieves such goals by constructing three different branches for inputs: (1) spatial filter branch, (2) energy-based attention branch, (3) identity branch. In spatial filter branch, we first utilize a 3×3 convolution to refine the input feature $F \in \mathbb{R}^{C_{in} \times H \times W}$ where C_{in}, H, and W

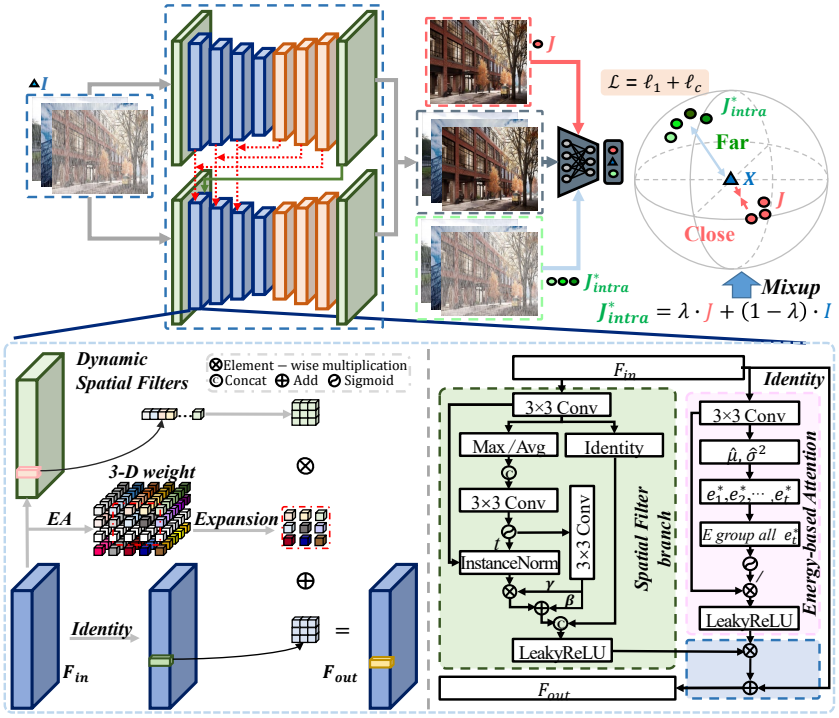


Fig. 1. The architecture of the proposed DRCNet. It consists of two sub-networks and employs the encoder-decoder paradigm to restore images. The core components of Dynamic Filter Restoration module are the spatial filters branch in green color region, and the energy-based attention branch in the pink color region. Moreover, we minimize the L1 reconstruction loss (ℓ_1) with CR (ℓ_c) to better pull the restored image (i.e. anchor, X) to the clear (i.e. positive, J) image and push the restored image away from the degraded (i.e. negative, J^*_{intra}) images.

denote the input channel, height, width of the feature maps. Then, we randomly divided the feature map into two parts: one part utilizes our proposed adaptive feature normalization to mask degraded signals, the other to keep context information [7,45]. Finally, we concatenate the two parts to aggregate the features. This operation enables DFR to suppress noise for adaptability to varying spatial degradation, leading to sacrifice of texture details. Therefore, we design an EA branch that focuses on generating the texture details. Besides, the EA branch guides the feature integration between the EA branch and the spatial filter branch [49]. Moreover, the identity branch launches a vanilla transformation with 1×1 convolution to the inputs, which helps maintain the features from the original images. Overall, the whole DFR module can be defined as follows:

$$F'_{(r,i)} = \sum_{j \in \Omega(i)} D_i^{sp}[p_i - p_j] \mathcal{W}_i^{ea}[p_i - p_j] F_{(r,j)} \tag{1}$$

where $F^{'}_{(r,i)}, F_{(r,j)} \in \mathbb{R}$ denotes the output/input feature value at the i^{th}, j^{th} pixel of r^{th} channel. $\Omega(i)$ denotes the k × k convolution window around i^{th} pixel. $D^{sp} \in \mathbb{R}^{h\times w\times k\times k}$ is the spatial dynamic filter with $D^{sp}_i \in \mathbb{R}^{k\times k}$ denoting the filter at i^{th} pixel. $\mathcal{W}^{ea} \in \mathbb{R}^{h\times w\times k\times k}$ is the dynamic attention weights with $\mathcal{W}^{ea}_i \in \mathbb{R}^{k\times k}$ denoting the 3-D attention weights value at i^{th} pixel. To delve into the details of DFR, we detail the two principal modules of DFR: spatial filter branch and energy-based attention branch.

Spatial Filter Branch. Since previous spatially-adaptive IR methods without considering the degraded pixel intensity change, we set out to design an adaptive feature normalization to detect intensity changes and recover them. We first perform convolution on input feature $F_{in} \in \mathbb{R}^{C\times H\times W}$ to extract initial feature and employ max-pooling and average-pooling $F_{max}, F_{avg} \in \mathbb{R}^{1\times H\times W}$ along the channel to obtain an efficient feature descriptor [56]. The spatial response map $\mathcal{M}_{sr}$ is obtained by a convolution on F_{max}, F_{avg} with sigmoid function, which represents local representation [52] and can be defined as follows:

$$F_{max}, F_{avg} = Conv(F_{in}) \tag{2}$$

$$\mathcal{M}_{sr} = sigmoid\left(Conv\left([F_{max}, F_{avg}]\right)\right) \tag{3}$$

The threshold t in Fig. 1 aims to detect the degraded pixels with a soft distinction. The mask $\mathcal{M}_p \in \mathbb{R}^{1\times H\times W}$ is 1 when $\mathcal{M}_{sr}$ greater than t, and is 0 otherwise. Specifically, $p \in (h, w)$ represents 2D pixel location. Considering the spatial relationship of $\mathcal{M}_{sr}$, we utilize a convolution layer to obtain a set of learnable parameters expanded along the channel dimension $\gamma^i_c \in \mathbb{R}^{1\times H\times W}$ and bias $\beta^i_c \in \mathbb{R}^{1\times H\times W}$, which enhances the feature representation. The computation for γ^i_c, β^i_c is formulated as follows:

$$\gamma^i_c, \beta^i_c = Conv\left(\mathcal{M}_{sr}\right) \tag{4}$$

The μ^i_c and σ^i_c are the channel-wise mean and variance of the features in i-th layer, which relate to global semantic information and local texture [19]:

$$\mu^i_c = \frac{1}{\sum_p \mathcal{M}^p_i} \sum_p F_{in} \odot \mathcal{M}_p \tag{5}$$

$$\sigma^i_c = \sqrt{\frac{1}{\sum_p \mathcal{M}_p} \sum_p \left(F_{in} \odot \mathcal{M}_p - \mu^i_c + \varepsilon\right)} \tag{6}$$

where $\sum_p \mathcal{M}_p$ indicates the number of masked pixels, $\odot$ represents element-wise product, and ε is a small constant to avoid σ^i_c equal to 0. The final feature output of the spatial filter branch is obtained as follows:

$$F^i_{h,w,c} = \gamma^i_c \cdot \frac{F_{in} - \mu^i_c}{\sigma^i_c} + \beta^i_c \tag{7}$$

Energy-based Attention Branch. The spatial filter branch with adaptive feature normalization suppresses degraded pixels, which may impede the restoration in texture areas. Thus, we introduce the energy-based attention branch to remedy the deficiency of spatial information, which considers the 3-D weights and preserves the details of the textures in the heavily degraded image [49]. Different from the previous works only refined features along either channel or spatial dimensions, we integrate 3D attention in the DFR to directly infer attention weights and calibrate the pixel. Moreover, EA can also leave the clean pixel features and guide the feature integration by calculating the importance score for each pixel.

Specifically, we first obtain the initial feature from a convolution operation. Then, we calculate the mean $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$ and variance $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{\mu}\right)^2$ over all neurons ($N = H \times W$) in that channel. $\hat{\mu}$ and $\hat{\sigma}^2$ are used for calculating the energy function for each pixel, which is the same as re-weighting the input feature map. We minimize the energy of target neuron t and formulate as follows:

$$e_t^* = \frac{4\left(\hat{\sigma}^2 + \delta\right)}{\left(t - \hat{\mu}\right)^2 + 2\hat{\sigma}^2 + 2\delta} \tag{8}$$

where δ is the coefficient hyper-parameter. The refined features $\tilde{X}$ as follows:

$$\tilde{X} = sigmoid\left(\frac{1}{E}\right) \odot X \tag{9}$$

where E is obtained by grouping all e_t^* across the channel and spatial dimensions. As for the identity branch, it generates identity features by a 1x1 Convolution.

The final feature output of DFR is obtained as follows: (1) we multiply the features of spatial filter branch and EA to obtain the intermediate features (2) we add the intermediate features with identity features to obtain the integrated features, as highlighted by the blue region in Fig. 1. The final features output by the encoder will be fed to the decoder for restored image generation.

3.3 Contrastive Regularization

Previous contrastive regularization [44] simply selected other haze images as negative samples from the same batch, namely **Extra**-class **CR** (Extra-CR), which may result in sub-optimal performance. Thus, we propose a new contrastive regularization method called **Intra**-class **CR** (Intra-CR), which constructs negative samples through mixup [62] between clean images and degraded images.

In a classical IR scenario, a degraded image I is transformed to the restored image X to approximate its clean image J. Specifically, we denote $s = (X, J)$ as the pair of X and J, and $s^* = (X, J^*)$ as the pair of negative sample X and J^*. $\ell_1(s, \theta)$ and $\ell_c(s^*, \theta)$ represent L1 reconstruction loss and contrastive regularizer, where θ represents the model's parameters. Then the empirical risk minimizer for model optimization is given by:

$$\theta_{\alpha, s^*} = \text{argmin}_{\theta \in \Theta} \sum_{s \in S} \left[\ell_1\left(s, \theta\right) + \alpha \cdot \ell_c(s^*, \theta)\right] \tag{10}$$

where α is the weight to control the balance the reconstruction loss and contrastive regularization. Besides, Intra-CR constructs the negative samples J^* through a mixup operation between degraded images I and clean images J, defined as follows:

$$J^*_{intra} = \lambda \cdot J + (1 - \lambda) \cdot I \tag{11}$$

where λ is the hyper-parameter in mixup operation, we choose different λ to construct different negative samples. Then we give the theoretical analysis between Intra-CR and Extra-CR. The idea is to compute the parameter change as the weight α changes.

We define the sensitivity of performance towards α as follows:

$$\mathcal{R}_{sen}(s^*) = \left| \lim_{\alpha \to 0} \frac{d\theta_{\alpha,s^*}}{d\alpha} \right| \tag{12}$$

The sensitivity $\mathcal{R}_{sen}$ is a metric that reflects the how sensitive the model's performance towards α change. Then we give our main theorem:

Theorem 1. *Let s^*_{intra} and s^*_{extra} denote the negative pairs in Intra-CR and Extra-CR, then we obtain $\mathcal{R}_{sen}(s^*_{intra}) < \mathcal{R}_{sen}(s^*_{extra})$.*

The detailed proof is deferred to the supplementary materials. The above theorem indicates that Intra-CR is more stable towards hyper-parameter α than Extra-CR, and such a sensitivity can be reflected through performance changes [61], which will be shown in Sec. 4.6. Then the training objective $\mathcal{L}$ in Intra-CR can be formulated as follows:

$$\mathcal{L} = \ell_1(s, \theta) + \alpha \cdot \frac{\ell_1(G(X), G(J))}{\ell_1(G(X), G(J^*_{intra}))} \tag{13}$$

where G is a fixed pre-trained VGG19 [39]. $G(\cdot)$ aims to extract hidden features of the images, and we leverage $G(\cdot)$ to compare the common intermediate features between a pair of images. Note that our method is different from the perceptual loss [19], which only adds with positive-pair regularization, but Intra-CR also adopts negative pairs. Besides, our Intra-CR is different from the Extra-CR [44] on negative sample construction. Experiments demonstrate that Intra-CR outperforms Extra-CR in image restoration tasks.

4 Experiments and Analysis

For comprehensive comparisons, the proposed DRCNet is contrasted on three IR tasks in this section: image deraining, denoising, and deblurring.

4.1 Benchmarks and Evaluation

We evaluate our method by Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM)[42]. As in [55], we report (in parenthesis) the reduce in error for each model relative to the best performing method

by RMSE $\left(RMSE \propto \sqrt{10^{-PSNR/10}}\right)$ and DSSIM ($DSSIM = (1 - SSIM)/2$). Meanwhile, qualitative evaluation is shown through the visualization of different benchmarks. The benchmarks are listed as follows: **Image Deraining.** We employ the same training data as MPRNet [55] which consists of 13,712 clean-rain image pairs, and that of Test100 [59], Rain100H [50], Rain100L [50], Test2800 [13], and Test1200 [58] as testing sets. **Image Denoising**. We train DRCNet on the SIDD medium version [1] dataset with 320 high-resolution images and directly test it on the DND [31] dataset with 50 pairs of real-world noisy images. **Image Deblurring.** We train on the GoPro[29] dataset that contains 2,103 image pairs for training and 1,111 pairs for evaluation and directly apply it to HIDE[38] and RealBlur[36] to demonstrate generalization.

4.2 Implementation Details

Our DRCNet is trained with Adam optimizer [21], and the learning rate is set to 2×10^{-4} by default, and decreased to 1×10^{-7} with cosine annealing strategy [27]. δ in the Eq. (8) is set to 1×10^{-6}, the degraded pixel mask threshold t in Fig. 1 is set to 0.75. Detailed analysis of δ and t will be discussed in our supplementary materials. For Intra-CR, α is set to 0.04, and the number n of negative samples is set to 3. The mixup parameters are selected as $0.90, 0.95, 1.00$. We train our model on 256×256 patches with a batch size of 32 for 4×10^5 iterations. Specifically, we apply random rotation, cropping, and flipping to the images to augment the training data.

4.3 Image Deraining Results

We conduct experiments to illustrate that the proposed DRCNet outperforms prior approaches in visual results and achieves competitive performance in terms of PSNR/SSIM scores on all derain benchmarks, shown in Table 1. For the image deraining task, consistent with prior work [55], Table 1 illustrates that our method significantly advances state-of-the-art by consistently achieving better PSNR/SSIM scores on all derain benchmarks. Compared to the state-of-the-art model HINet [7], we obtain significant performance gains of 0.7dB in PSNR and 0.011 in SSIM, and a 7.76% and 2.7% error reduction averaged across all derain benchmarks. Specifically, the improvement on Rain100L can reach 0.95 dB, which well demonstrates that our model can effectively remove rain streaks. Meanwhile, the qualitative results on image derain samples are illustrated in Fig. 2, which demonstrates that DRCNet produces better visual qualities with fine-detailed structures. In contrast, the output restored images of other comparison methods fail to recover complex textures. Due to the effectiveness of DRF, DRCNet can faithfully recover the texture and structures with fewer parameters.

4.4 Image Denoising Results

Table 2 and Fig. 3 show quantitative and qualitative comparisons with other denoising models on the SIDD [1] and DND [31] datasets. DRCNet achieves

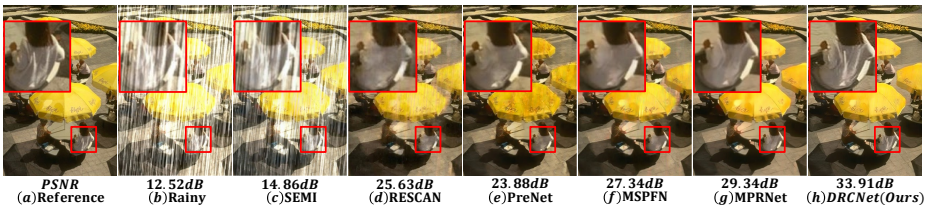


Fig. 2. Visual comparisons on the derain test set. Our DRCNet obtains better visual results with more natural details while removing rain.

Table 1. Quantitative results of image deraining.The best and second-best scores are bolden and underlined, respectively. Our DRCNet achieves substantial improvements in PSNR over HINet [7]. 'Params' means the number of parameters (Millions). ↑ denotes higher is better.

	Test100[59]		Rain100H[50]		Rain100L[50]		Test2800[13]		Test1200[58]		Average		Params
Methods	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	(M)
DerainNet[12]	22.77	0.810	14.92	0.592	27.03	0.884	24.31	0.861	23.38	0.835	22.48	0.796	-
SEMI[43]	22.35	0.788	16.56	0.486	25.03	0.842	24.43	0.782	26.05	0.822	22.88	0.744	-
DIDMDN[58]	22.56	0.818	17.35	0.524	25.23	0.741	28.13	0.867	29.65	0.901	24.58	0.770	0.37
UMRL[51]	24.41	0.829	26.01	0.832	29.18	0.923	29.97	0.905	30.55	0.910	28.02	0.880	0.9
RESCAN[24]	25.00	0.835	26.36	0.786	29.80	0.881	31.29	0.904	30.51	0.882	28.59	0.857	0.15
PreNet[35]	24.81	0.851	26.77	0.858	32.44	0.950	31.75	0.916	31.36	0.911	29.42	0.897	0.16
MSPFN[18]	27.50	0.876	28.66	0.860	32.40	0.933	32.82	0.930	32.39	0.916	30.75	0.903	21
MPRNet[55]	30.27	0.897	30.41	0.890	36.40	0.965	33.64	0.938	32.91	0.916	32.73	0.921	3.64
SPAIR[33]	30.35	0.909	30.95	0.892	36.93	0.969	33.34	0.936	33.04	0.922	32.91	0.926	-
HINet[7]	30.29	0.906	30.65	0.894	37.28	0.970	33.91	0.941	33.05	0.919	33.03	0.926	88.7
DRCNet(Ours)	**32.18**	**0.917**	**30.96**	**0.895**	**38.23**	**0.976**	**33.89**	**0.946**	**33.40**	**0.94**	**33.73**	**0.933**	18.9

consistently better PSNR and SSIM. The results show that DRCNet outperforms the state-of-the-art denoising approaches, i.e., 0.37 dB and 0.05 dB over MPRNet on SIDD and DND. In the SSIM metric, DRCNet also has a performance rise compared to MPRNet, boosting from 0.958 to 0.972. It means that the proposed model can successfully restore the detailed regional textures. As the DND dataset does not provide any training images, DRCNet can achieve impressive results, indicating it has good generalization capability. As no training images in DND dataset, DRCNet can achieve impressive results, indicating it has good generalization capability. Generally, our DRCNet provides better image denoise performance on the denoising task, which effectively removes the noise and artifacts while preserving the main structure and contents.

4.5 Image Deblurring Results

Table 3 and Fig. 4 report the image deblurring performance on GoPro [29] and HIDE [38] dataset. Our method achieves 32.82 PSNR and 0.961 in SSIM on the GoPro [29] dataset and achieves 31.08 PSNR and 0.94 SSIM on the HIDE dataset. It is worth mentioning that DRCNet is trained only on the GoPro dataset and obtains outstanding performance on the HIDE dataset, validating that the proposed method has good generalization. Moreover, we directly evaluate the GoPro trained model on RealBlur-J, which can further test the

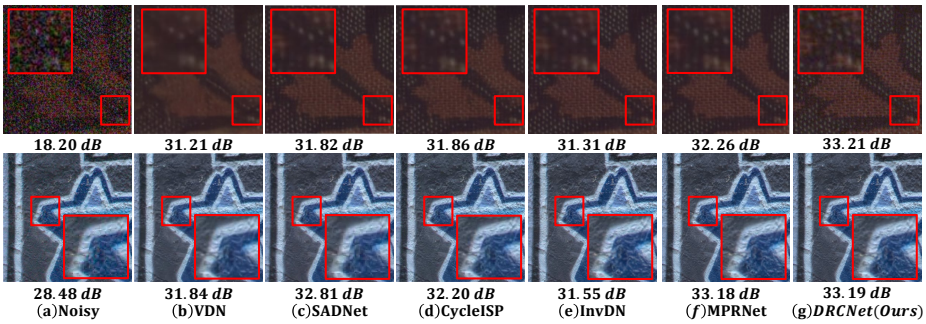


Fig. 3. Qualitative comparisons with the existing methods on the denoising datasets. The top row is from SIDD[1] and the down row is from DND[31]. The proposed DRCNet can produce fine-grained texture and high-frequency details.

Table 2. Quantitative results of image denoising on SIDD [1] and DND [31] datasets. We denote the comparison methods using additional training data with *. Following [55], we perform the reduction in error relative to the best-performing algorithm in parenthesis (see Sec. 4.1 for calculation).

	SIDD[1]		DND[31]		Params
Method	PSNR↑	SSIM↑	PSNR↑	SSIM↑	(M)
DnCNN[64]	23.66 (84.90%)	0.583 (93.29%)	32.43 (57.44%)	0.790 (79.05%)	0.56
MLP[3]	24.71 (82.96%)	0.641 (92.20%)	34.23 (47.64%)	0.833 (73.65%)	-
BM3D[10]	25.65 (81.01%)	0.685 (91.11%)	34.51 (45.93%)	0.851 (70.47%)	-
CBDNet*[14]	30.78 (65.72%)	0.801 (85.93%)	38.06 (18.62%)	0.942 (24.14%)	4.36
RIDNet*[2]	38.71 (14.59%)	0.951 (42.86%)	39.26 (6.57%)	0.953 (6.38%)	1.49
AINDNet*[20]	38.95 (12.20%)	0.952 (41.67%)	39.37 (5.38%)	0.951 (10.20%)	13.76
VDN[53]	39.28 (8.80%)	0.956 (36.36%)	39.38 (5.27%)	0.952 (8.33%)	7.81
SADNet*[5]	39.46 (6.89%)	0.957 (34.88%)	39.59 (2.95%)	0.952 (8.33%)	0.42
DANet+*[54]	39.47 (6.78%)	0.957 (34.88%)	39.58 (3.06%)	0.955 (2.22%)	9.1
CycleISP*[56]	39.52 (6.25%)	0.957 (34.88%)	39.56 (3.29%)	0.956 (0.00%)	2.8
InvDN [25]	39.28 (8.80%)	0.955 (37.78%)	39.57 (3.17%)	0.952 (8.33%)	2.64
MPRNet [55]	39.71 (4.17%)	0.958 (33.33%)	39.80 (0.58%)	0.954 (4.35%)	15.7
DRCNet(Ours)	**40.08** (0.00%)	**0.972** (0.00%)	**39.85** (0.00%)	**0.956** (0.00%)	18.9

generalization of models. Table 3 also shows the experimental results of the DRCNet training and testing on the RealBlur-J dataset. Our DRCNet obtains a performance gain of 0.06 dB on the RealBlur-J subset over the other comparison methods. Overall, the proposed DRCNet outputs restored images with fine-detailed structures and has better visual results than competing methods.

4.6 Ablation study

To demonstrate the effectiveness of the proposed DRCNet, we conduct ablation studies to analyze the effectiveness of crucial components of DRCNet, including the DFR module and Intra-CR.

Comparison to Other Dynamic Filters. Since there are other proposed dynamic convolution filters, we conduct experiments to compare them in image

Table 3. Deblurring results. Our method is trained only on the GoPro dataset [29] and directly tested to the HIDE dataset [38] and RealBlur-J [36] datasets. The scores in the PSNR ‡ column were obtained after training and testing on RealBlur-J dataset.

	GoPro [29]		HIDE [38]		RealBlur-J [36]			Params
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR‡	(M)
DeblurGAN[22]	28.70	0.858	24.51	0.871	27.97	0.834		-
Nah et al.[29]	29.08	0.914	25.73	0.874	27.87	0.827		11.7
Zhang et al.[63]	29.19	0.913	-	-	27.80	0.847		9.2
DeblurGAN-v2[23]	29.55	0.934	26.61	0.875	28.70	0.866	29.69	60.9
SRN [34]	30.26	0.934	28.36	0.915	28.56	0.867	31.38	6.8
Shen *et al.*[38]	30.26	0.940	28.89	0.930	-	-		100
DBGAN[65]	31.10	0.942	28.94	0.915	-	-		11.6
MT-RNN[30]	31.15	0.945	29.15	0.918	-	-		2.6
DMPHN[60]	31.20	0.940	29.09	0.924	28.42	0.860		21.7
RADN[32]	31.76	0.952	29.68	0.927	-	-		-
SAPHNet [40]	31.85	0.948	29.98	0.930	-	-		-
SPAIR[33]	32.06	0.953	30.29	0.931	28.81	0.875	31.82	-
MPRNet[55]	32.66	0.959	30.96	0.939	28.70	0.873	31.76	20.1
MIMO-UNet[9]	32.45	0.957	29.99	0.930	27.63	0.873		16.1
HINet[7]	32.71	0.959	30.32	0.932	-	-		88.7
DRCNet(Ours)	**32.82**	**0.961**	**31.08**	**0.940**	**28.87**	**0.881**	**31.85**	18.9

restoration tasks, as shown in Table 5. We replace the DFR module with three dynamic filters: SACT [11], CondConv [48], UDVD[47], and DDF [67].

Table 5 compares the performance and the parameters of the whole network in various restoration tasks. The experimental results show that models with other dynamic filters obtain significantly worse performance and have more parameters than DFR in the image restoration tasks. Such results validate that DFR is suitable for image restoration tasks.

Effectivenes of DFR. We first construct our base network as baseline, which mainly consists of normal UNet [37] with Resblock [17] in encoding and decoding phrases with SAM and CSFF [55]. Subsequently, we replace DFR module and add the Intra-CR scheme into the base network as follows: (1) **base+SF**: only add spatial filter branch (SF) into baseline. (2) **base+EA**: only add energy-based attention branch into baseline. (3) **base+SF+identity branch**: add spatial filter branch and identity branch. (4) **base+SF+EA**: add spatial filter branch and energy-based attention branch. (5) **base+DFR**: add combination of three branches as DFR module. (6) **base+DFR+CR**: add DFR module and Extra-CR. (7) **DRCNet**: the combination of DFR module and Intra-CR for training. The performance of these models is summarized in Table 4.

As shown in Table 4, the spatial filter branch can strengthen DRCNet with more representation power than the base model. Besides, the energy-based attention also improves the model's restoration capacity by dynamically guiding feature integration. Besides, Table 4 shows a significant performance drop in PSNR from 40.01 dB to 33.25 dB by removing the whole DFR, which shows that DFR is a successful and crucial module in DRCNet. Specifically, we show the visualization of intermediate features produced by different branches of DFR. As shown in Fig. 5 (a) denotes that spatial filter can effectively reduce noise, the

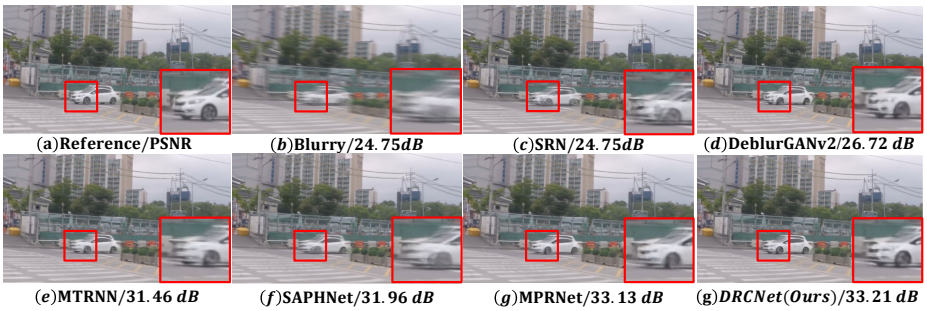


Fig. 4. Qualitative comparisons on GoPro [29] test dataset. The deblurred results listed from left to right are from SRN [34], DeblurGANv2 [23], MTRNN [30], SAPHNet [40], MPRNet [55] and ours, respectively.

Table 4. Ablation studies on DRCNet on SIDD benchmark.

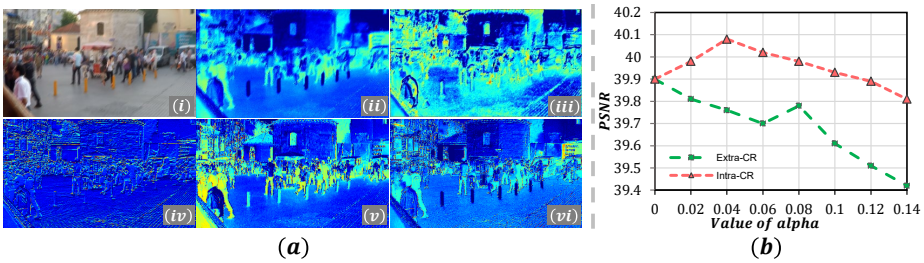
Model	CR	PSNR	SSIM	Params(M)	Times(ms)
base	-	33.25	0.812	12.7	15.2
base+SF	-	37.76	0.933	15.4	20.1
base+EA	-	37.12	0.945	13.1	17.7
base+SF+identity branch	-	38.79	0.925	15.8	21.2
base+SF+EA	-	38.89	0.965	18.6	22.7
base+DFR	-	40.01	0.971	18.9	23.9
base+DFR	Extra-CR	39.95	0.956	-	-
base+DFR	Intra-CR	**40.08**	**0.972**	-	-

energy-based attention branch focuses on the textures and sharpness in terms of SSIM. Besides, the identity branch can further enhance the feature integration. Overall, the combination of the three branches achieves the best results.

Table 5. Comparison of the parameter number and PSNR (dB).

Filter		SACT[11]	CondConv[48]	UDVD[47]	DDF [67]	**DFR**
Params		103.1M	165.7M	95.2M	87.4M	**18.9M**
PSNR	Derain [55]	19.28	23.53	21.72	25.12	**32.39**
	SIDD[1]	27.28	39.43	37.21	39.04	**40.01**
	GoPro [29]	19.97	23.09	27.45	29.01	**32.21**

Effect of Contrastive Regularization. In Table 4, Extra-CR empirically impairs the model performance, indicating that using extra-class images as negative samples in denoising is not beneficial due to easy negative samples. This section illustrates the effectiveness of our Intra-CR. Specifically, we apply Intra-CR and Extra-CR on DRCNet, respectively. Moreover, we vary the value of weight α in Eq. (13) and observe the tendency of Intra-CR and Extra-CR. As shown in Fig. 5 (b), Intra-CR outperforms Extra-CR as α varies from 0 to 0.14. Moreover, Intra-CR achieves more stable results towards α, which matches our



(*a*) (*b*)

Fig. 5. (a) Visualization of intermediate features on images from the GoPro test set [29]. (i-ii) Input blurred image and feature map. (iii-v) Comparisons among the feature map by using spatial filter, energy-based attention and DFR module, respectively. (vi) ground truth feature map; (b) Ablation experiment of comparison between Intra-CR and Extra-CR on SIDD benchmark.

theoretical analysis that Intra-CR is less sensitive to α. Overall, the results validate the superiority of our Intra-CR.

5 Conclusion

In this paper, we propose a dynamic restoration contrastive network (DRCNet) for image restoration with two principal components: Dynamic Filter Restoration module (DFR) and Intra-class contrastive regularization (Intra-CR). The DFR module, built on a spatial filter branch and an energy-based attention branch, benefits from being dynamically adaptive toward spatially varying image degradation. The key insight of Intra-CR is to construct intra-class negative samples, which is accomplished through mixup operations. Through comprehensive evaluation of the performance of DRCNet on various benchmarks, we validate that the DRCNet achieves state-of-the-art results on ten datasets across various restoration tasks. Although DRCNet shows superior performance on three types of degradation, it needs to be trained for each type of degradation with a separate model, which limits the practical utility of the proposed approach. In the future, we will develop an all-in-one model for various restoration tasks.

Acknowledgement

The authors acknowledge the financial support from the Key Research and Development Plan Project of Guangdong Province ($No.2020B0202010009$) and the National Key R&D Program of China ($No.2020YFD0900204, 2021ZD0113805$). We appreciate the comments from Dr. Linfeng Zhang and the seminar participants at the Center for Deep Learning of Computer vision Research at China Agricultural University, which improves the manuscript significantly.

References

1. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1692–1700 (2018)
2. Anwar, S., Barnes, N.: Real image denoising with feature attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3155–3164 (2019)
3. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with bm3d? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2392–2399 (2012)
4. Cao, X., Chen, Y., Zhao, Q., Meng, D., Wang, Y., Wang, D., Xu, Z.: Low-rank matrix factorization under general mixture noise distributions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1493–1501 (2015)
5. Chang, M., Li, Q., Feng, H., Xu, Z.: Spatial-adaptive network for single image denoising. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 171–187 (2020)
6. Chen, J., Wang, X., Guo, Z., Zhang, X., Sun, J.: Dynamic region-aware convolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8064–8073 (2021)
7. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: Hinet: Half instance normalization network for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 182–192 (2021)
8. Chen, Y., Dai, X., Liu, M., Chen, D., Yuan, L., Liu, Z.: Dynamic convolution: attention over convolution kernels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11030–11039 (2020)
9. Cho, S.J., Ji, S.W., Hong, J.P., Jung, S.W., Ko, S.J.: Rethinking coarse-to-fine approach in single image deblurring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4641–4650 (2021)
10. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on Image Processing **16**(8), 2080–2095 (2007)
11. Figurnov, M., Collins, M.D., Zhu, Y., Zhang, L., Huang, J., Vetrov, D., Salakhutdinov, R.: Spatially adaptive computation time for residual networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1039–1048 (2017)
12. Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.: Clearing the skies: A deep network architecture for single-image rain removal. IEEE Transactions on Image Processing **26**(6), 2944–2956 (2017)
13. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3855–3863 (2017)
14. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1712–1722 (2019)
15. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9729–9738 (2020)

16. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(12), 2341–2353 (2010)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
18. Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., Ma, J., Jiang, J.: Multi-scale progressive fusion network for single image deraining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8346–8355 (2020)
19. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 694–711 (2016)
20. Kim, Y., Soh, J.W., Park, G.Y., Cho, N.I.: 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 (CVPR). pp. 3482–3492 (2020)
21. Kingma, D., Ba, J.: Adam: A method for stochastic optimization (2014)
22. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: Deblurgan: Blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8183–8192 (2018)
23. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 8878–8887 (2019)
24. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 254–269 (2018)
25. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13365–13374 (2021)
26. Lo, Y.C., Chang, C.C., Chiu, H.C., Huang, Y.H., Chang, Y.L., Jou, K.: Clcc: Contrastive learning for color constancy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8053–8063 (2021)
27. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. In: Proceedings of the International Conference on Learning Representations (ICLR) (2016)
28. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2272–2279 (2009)
29. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3883–3891 (2017)
30. Park, D., Kang, D.U., Kim, J., Chun, S.Y.: Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 327–343 (2020)
31. Plotz, T., Roth, S.: Benchmarking denoising algorithms with real photographs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1586–1595 (2017)

32. Purohit, K., Rajagopalan, A.: Region-adaptive dense network for efficient motion deblurring. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 11882–11889 (2020)
33. Purohit, K., Suin, M., Rajagopalan, A.N., Boddeti, V.N.: Spatially-adaptive image restoration using distortion-guided networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2309–2319 (2021)
34. Ren, D., Shang, W., Zhu, P., Hu, Q., Meng, D., Zuo, W.: Single image deraining using bilateral recurrent network. IEEE Transactions on Image Processing **29**, 6852–6863 (2020)
35. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: A better and simpler baseline. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3937–3946 (2019)
36. Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 184–201 (2020)
37. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention. pp. 234–241 (2015)
38. Shen, Z., Wang, W., Lu, X., Shen, J., Ling, H., Xu, T., Shao, L.: Human-aware motion deblurring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5572–5581 (2019)
39. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations (ICLR) (2015)
40. Suin, M., Purohit, K., Rajagopalan, A.: Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3606–3615 (2020)
41. Thomas, H., Qi, C.R., Deschaud, J.E., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6411–6420 (2019)
42. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
43. Wei, W., Meng, D., Zhao, Q., Xu, Z., Wu, Y.: Semi-supervised transfer learning for image rain removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3877–3886 (2019)
44. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10551–10560 (2021)
45. Xie, C., Tan, M., Gong, B., Wang, J., Yuille, A.L., Le, Q.V.: Adversarial examples improve image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
46. Xu, L., Zheng, S., Jia, J.: Unnatural l0 sparse representation for natural image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1107–1114 (2013)
47. Xu, Y.S., Tseng, S.Y.R., Tseng, Y., Kuo, H.K., Tsai, Y.M.: Unified dynamic convolutional network for super-resolution with variational degradations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12496–12505 (2020)

48. Yang, B., Bender, G., Le, Q.V., Ngiam, J.: Condconv: Conditionally parameterized convolutions for efficient inference. In: Advances in Neural Information Processing Systems. pp. 1307–1318 (2019)
49. Yang, L., Zhang, R.Y., Li, L., Xie, X.: Simam: A simple, parameter-free attention module for convolutional neural networks. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 11863–11874 (2021)
50. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1357–1366 (2017)
51. Yasarla, R., Patel, V.M.: Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8405–8414 (2019)
52. Yu, T., Guo, Z., Jin, X., Wu, S., Chen, Z., Li, W., Zhang, Z., Liu, S.: Region normalization for image inpainting. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 12733–12740 (2020)
53. 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 (2019)
54. Yue, Z., Zhao, Q., Zhang, L., Meng, D.: Dual adversarial network: Toward real-world noise removal and noise generation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 41–58 (2020)
55. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14821–14831 (2021)
56. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Shao, L.: Cycleisp: Real image restoration via improved data synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2696–2705 (2020)
57. Zhang, H., Li, Y., Chen, H., Shen, C.: Memory-efficient hierarchical neural architecture search for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3657–3666 (2020)
58. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 695–704 (2018)
59. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. IEEE Transactions on Circuits and Systems for Video Technology **30**(11), 3943–3956 (2019)
60. Zhang, H., Dai, Y., Li, H., Koniusz, P.: Deep stacked hierarchical multi-patch network for image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5978–5986 (2019)
61. Zhang, H., Yu, Y., Jiao, J., Xing, E., El Ghaoui, L., Jordan, M.: Theoretically principled trade-off between robustness and accuracy. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 7472–7482 (2019)
62. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: Mixup: Beyond empirical risk minimization. In: Proceedings of the International Conference on Learning Representations (ICLR) (2018)
63. Zhang, J., Pan, J., Ren, J., Song, Y., Bao, L., Lau, M.H.: Dynamic scene deblurring using spatially variant recurrent neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2521–2529 (2018)

64. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing **26**(7), 3142–3155 (2017)
65. Zhang, K., Luo, W., Zhong, Y., Ma, L., Stenger, B., Liu, W., Li, H.: Deblurring by realistic blurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2737–2746 (2020)
66. Zhang, R., Tang, S., Zhang, Y., Li, J., Yan, S.: Scale-adaptive convolutions for scene parsing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2031–2039 (2017)
67. Zhou, J., Jampani, V., Pi, Z., Liu, Q., Yang, M.H.: Decoupled dynamic filter networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6647–6656 (2021)