Learning Enriched Features for Real Image Restoration and Enhancement

Syed Waqas Zamir[1], Aditya Arora[1], Salman Khan[1,2], Munawar Hayat[1,2], Fahad Shahbaz Khan[1,2], Ming-Hsuan Yang[3,4], and Ling Shao[1,2]

[1] Inception Institute of Artificial Intelligence, UAE
[2] Mohamed bin Zayed University of Artificial Intelligence, UAE
[3] University of California, Merced, USA
[4] Google Research

Abstract. With the goal of recovering high-quality image content from its degraded version, image restoration enjoys numerous applications, such as in surveillance, computational photography, medical imaging, and remote sensing. Recently, convolutional neural networks (CNNs) have achieved dramatic improvements over conventional approaches for image restoration task. Existing CNN-based methods typically operate either on full-resolution or on progressively low-resolution representations. In the former case, spatially precise but contextually less robust results are achieved, while in the latter case, semantically reliable but spatially less accurate outputs are generated. In this paper, we present a novel architecture with the collective goals of maintaining spatially-precise high-resolution representations through the entire network, and receiving strong contextual information from the low-resolution representations. The core of our approach is a multi-scale residual block containing several key elements: (a) parallel multi-resolution convolution streams for extracting multi-scale features, (b) information exchange across the multi-resolution streams, (c) spatial and channel attention mechanisms for capturing contextual information, and (d) attention based multi-scale feature aggregation. In a nutshell, our approach learns an enriched set of features that combines contextual information from multiple scales, while simultaneously preserving the high-resolution spatial details. Extensive experiments on five real image benchmark datasets demonstrate that our method, named as MIRNet, achieves state-of-the-art results for a variety of image processing tasks, including image denoising, super-resolution and image enhancement. The source code and pre-trained models are available at https://github.com/swz30/MIRNet.

Keywords: Image denoising, super-resolution, and image enhancement.

1 Introduction

Image content is exponentially growing due to the ubiquitous presence of cameras on various devices. During image acquisition, degradations of different severity are often introduced. It is either because of the physical limitations of cameras or due to inappropriate lighting conditions. For instance, smartphone cameras

come with a narrow aperture and have small sensors with limited dynamic range. Consequently, they frequently generate noisy and low-contrast images. Similarly, images captured under the unsuitable lighting are either too dark or too bright. The art of recovering the original clean image from its corrupted measurements is studied under the image restoration task. It is an ill-posed inverse problem, due to the existence of many possible solutions.

Recently, deep learning models have made significant advancements for image restoration and enhancement, as they can learn strong (generalizable) priors from large-scale datasets. Existing CNNs typically follow one of the two architecture designs: 1) an encoder-decoder, or 2) high-resolution (single-scale) feature processing. The encoder-decoder models [84,59,17,124] first progressively map the input to a low-resolution representation, and then apply a gradual reverse mapping to the original resolution. Although these approaches learn a broad context by spatial-resolution reduction, on the downside, the fine spatial details are lost, making it extremely hard to recover them in the later stages. On the other side, the high-resolution (single-scale) networks [27,120,127,50] do not employ any downsampling operation, and thereby produce images with spatially more accurate details. However, these networks are less effective in encoding contextual information due to their limited receptive field.

Image restoration is a position-sensitive procedure, where pixel-to-pixel correspondence from the input image to the output image is needed. Therefore, it is important to remove only the undesired degraded image content, while carefully preserving the desired fine spatial details (such as true edges and texture). Such functionality for segregating the degraded content from the true signal can be better incorporated into CNNs with the help of large context, *e.g.*, by enlarging the receptive field. Towards this goal, we develop a new *multi-scale* approach that maintains the original high-resolution features along the network hierarchy, thus minimizing the loss of precise spatial details. Simultaneously, our model encodes multi-scale context by using *parallel convolution streams* that process features at lower spatial resolutions. The multi-resolution parallel branches operate in a manner that is complementary to the main high-resolution branch, thereby providing us more precise and contextually enriched feature representations.

The main difference between our method and existing multi-scale image processing approaches is the way we aggregate contextual information. First, the existing methods [97,71,37] process each scale in isolation, and exchange information only in a top-down manner. In contrast, we progressively fuse information across all the scales at each resolution-level, allowing both top-down and bottom-up information exchange. Simultaneously, both fine-to-coarse and coarse-to-fine knowledge exchange is laterally performed on each stream by a new *selective kernel* fusion mechanism. Different from existing methods that employ a simple concatenation or averaging of features coming from multi-resolution branches, our fusion approach dynamically selects the useful set of kernels from each branch representations using a self-attention approach. More importantly, the proposed fusion block combines features with varying receptive fields, while preserving their distinctive complementary characteristics.

Our main contributions in this work include:

- A novel feature extraction model that obtains a complementary set of features across multiple spatial scales, while maintaining the original high-resolution features to preserve precise spatial details.
- A regularly repeated mechanism for information exchange, where the features across multi-resolution branches are progressively fused together for improved representation learning.
- A new approach to fuse multi-scale features using a selective kernel network that dynamically combines variable receptive fields and faithfully preserves the original feature information at each spatial resolution.
- A recursive residual design that progressively breaks down the input signal in order to simplify the overall learning process, and allows the construction of very deep networks.
- Comprehensive experiments are performed on five real image benchmark datasets for different image processing tasks including, image denoising, super-resolution and image enhancement. Our method achieves state-of-the-results on *all* five datasets. Furthermore, we extensively evaluate our approach on practical challenges, such as generalization ability across datasets.

2 Related Work

With the rapidly growing image content, there is a pressing need to develop effective image restoration and enhancement algorithms. In this paper, we propose a new method capable of performing image denoising, super-resolution and image enhancement. Unlike existing works for these problems, our approach processes features at the original resolution in order to preserve spatial details, while effectively fuses contextual information from multiple parallel branches. Next, we briefly describe the representative methods for each of the studied problems.

Image denoising. Classic denoising methods are mainly based on modifying transform coefficients [115,30,90] or averaging neighborhood pixels [91,98,78,86]. Although the classical methods perform well, the self-similarity [31] based algorithms, *e.g.*, NLM [10] and BM3D [21], demonstrate promising denoising performance. Numerous patch-based algorithms that exploit redundancy (self-similarity) in images are later developed [28,38,70,43]. Recently, deep learning-based approaches [11,5,9,35,39,80,120,121,119] make significant advances in image denoising, yielding favorable results than those of the hand-crafted methods.

Super-resolution (SR). Prior to the deep-learning era, numerous SR algorithms have been proposed based on the sampling theory [55,53], edge-guided interpolation [4,122], natural image priors [58,110], patch-exemplars [15,33] and sparse representations [114,113]. Currently, deep-learning techniques are actively being explored, as they provide dramatically improved results over conventional algorithms. The data-driven SR approaches differ according to their architecture designs [106,6,13]. Early methods [26,27] take a low-resolution (LR) image as input and learn to directly generate its high-resolution (HR) version. In contrast to directly producing a latent HR image, recent SR networks [56,95,94,48]

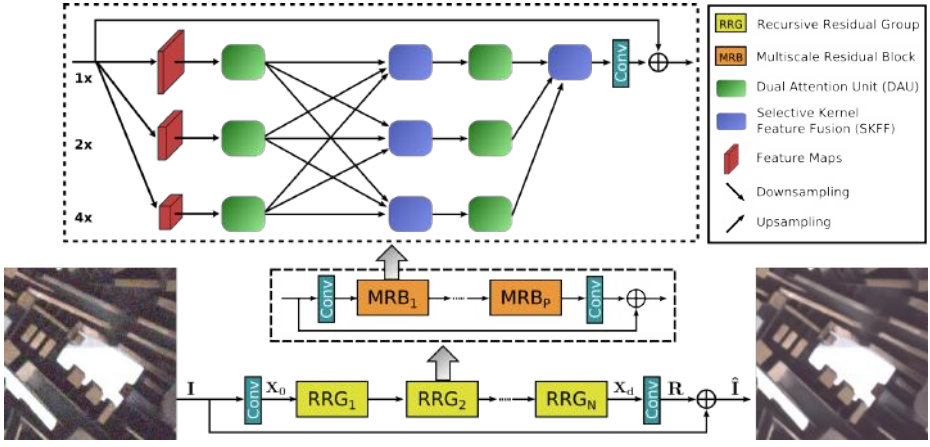


Fig. 1: Framework of the proposed network MIRNet that learns enriched feature representations for image restoration and enhancement. MIRNet is based on a recursive residual design. In the core of MIRNet is the multi-scale residual block (MRB) whose main branch is dedicated to maintaining spatially-precise high-resolution representations through the entire network and the complimentary set of parallel branches provide better contextualized features. It also allows information exchange across parallel streams via selective kernel feature fusion (SKFF) in order to consolidate the high-resolution features with the help of low-resolution features, and vice versa.

employ the residual learning framework [42] to learn the high-frequency image detail, which is later added to the input LR image to produce the final super-resolved result. Other networks designed to perform SR include recursive learning [57,41,3], progressive reconstruction [105,60], dense connections [99,104,127], attention mechanisms [125,23,126], multi-branch learning [60,66,22,64], and generative adversarial networks (GANs) [104,76,87,63].

Image enhancement. Oftentimes, cameras generate images that are less vivid and lack contrast. A number of factors contribute to the low quality of images, including unsuitable lighting conditions and physical limitations of camera devices. For image enhancement, histogram equalization is the most commonly used approach. However, it frequently produces under- or over-enhanced images. Motivated by the Retinex theory [61], several enhancement algorithms mimicking human vision have been proposed in the literature [8,74,54,83]. Recently, CNNs have been successfully applied to general, as well as low-light, image enhancement problems [52]. Notable works employ Retinex-inspired networks [89,107,124], encoder-decoder networks [18,68,81], and GANs [19,51,25].

3 Proposed Method

In this section, we first present an overview of the proposed MIRNet for image restoration and enhancement, illustrated in Fig. 1. We then provide details of the *multi-scale residual block*, which is the fundamental building block of our

method, containing several key elements: **(a)** parallel multi-resolution convolution streams for extracting (fine-to-coarse) semantically-richer and (coarse-to-fine) spatially-precise feature representations, **(b)** information exchange across multi-resolution streams, **(c)** attention-based aggregation of features arriving from multiple streams, **(d)** dual-attention units to capture contextual information in both spatial and channel dimensions, and **(e)** residual resizing modules to perform downsampling and upsampling operations.

Overall Pipeline. Given an image $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$, the network first applies a convolutional layer to extract low-level features $\mathbf{X_0} \in \mathbb{R}^{H \times W \times C}$. Next, the feature maps $\mathbf{X_0}$ pass through N number of recursive residual groups (RRGs), yielding deep features $\mathbf{X_d} \in \mathbb{R}^{H \times W \times C}$. We note that each RRG contains several multi-scale residual blocks, which is described in Section 3.1. Next, we apply a convolution layer to deep features $\mathbf{X_d}$ and obtain a residual image $\mathbf{R} \in \mathbb{R}^{H \times W \times 3}$. Finally, the restored image is obtained as $\hat{\mathbf{I}} = \mathbf{I} + \mathbf{R}$. We optimize the proposed network using the Charbonnier loss [16]:

$$\mathcal{L}(\hat{\mathbf{I}}, \mathbf{I}^*) = \sqrt{\|\hat{\mathbf{I}} - \mathbf{I}^*\|^2 + \varepsilon^2}, \tag{1}$$

where $\mathbf{I}^*$ denotes the ground-truth image, and ε is a constant which we empirically set to 10^{-3} for all the experiments.

3.1 Multi-scale Residual Block (MRB)

In order to encode context, existing CNNs [84,72,73,109,7,77] typically employ the following architecture design: **(a)** the receptive field of neurons is fixed in *each* layer/stage, **(b)** the spatial size of feature maps is *gradually* reduced to generate a semantically strong low-resolution representation, and **(c)** a high-resolution representation is *gradually* recovered from the low-resolution representation. However, it is well-understood in vision science that in the primate visual cortex, the sizes of the local receptive fields of neurons in the same region are different [47,82,88,49]. Therefore, such a mechanism of collecting multi-scale spatial information in the same layer needs to be incorporated in CNNs [46,92,32,93]. In this paper, we propose the multi-scale residual block (MRB), as shown in Fig. 1. It is capable of generating a spatially-precise output by maintaining high-resolution representations, while receiving rich contextual information from low-resolutions. The MRB consists of multiple (three in this paper) fully-convolutional streams connected in parallel. It allows information exchange across parallel streams in order to consolidate the high-resolution features with the help of low-resolution features, and vice versa. Next, we describe the individual components of MRB.

Selective kernel feature fusion (SKFF). One fundamental property of neurons present in the visual cortex is to be able to change their receptive fields according to the stimulus [65]. This mechanism of adaptively adjusting receptive fields can be incorporated in CNNs by using multi-scale feature generation (in the same layer) followed by feature aggregation and selection. The most commonly used approaches for feature aggregation include simple concatenation

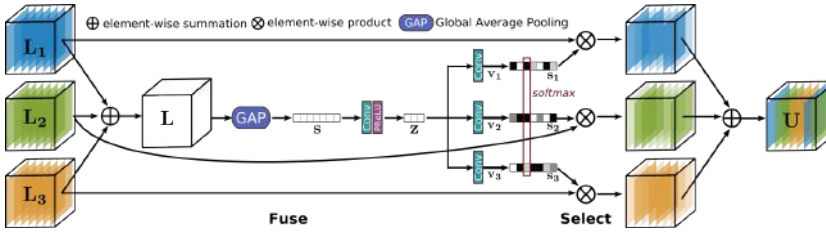


Fig. 2: Schematic for selective kernel feature fusion (SKFF). It operates on features from multiple convolutional streams, and performs aggregation based on self-attention.

or summation. However, these choices provide limited expressive power to the network, as reported in [65]. In MRB, we introduce a nonlinear procedure for fusing features coming from multiple resolutions using a self-attention mechanism. Motivated by [65], we call it selective kernel feature fusion (SKFF).

The SKFF module performs dynamic adjustment of receptive fields via two operations –*Fuse* and *Select*, as illustrated in Fig. 2. The *fuse* operator generates global feature descriptors by combining the information from multi-resolution streams. The *select* operator uses these descriptors to recalibrate the feature maps (of different streams) followed by their aggregation. Next, we provide details of both operators for the three-stream case, but one can easily extend it to more streams. **(1) Fuse:** SKFF receives inputs from three parallel convolution streams carrying different scales of information. We first combine these multi-scale features using an element-wise sum as: $\mathbf{L} = \mathbf{L_1} + \mathbf{L_2} + \mathbf{L_3}$. We then apply global average pooling (GAP) across the spatial dimension of $\mathbf{L} \in \mathbb{R}^{H \times W \times C}$ to compute channel-wise statistics $\mathbf{s} \in \mathbb{R}^{1 \times 1 \times C}$. Next, we apply a channel-downscaling convolution layer to generate a compact feature representation $\mathbf{z} \in \mathbb{R}^{1 \times 1 \times r}$, where $r = \frac{C}{8}$ for all our experiments. Finally, the feature vector $\mathbf{z}$ passes through three parallel channel-upscaling convolution layers (one for each resolution stream) and provides us with three feature descriptors $\mathbf{v_1}, \mathbf{v_2}$ and $\mathbf{v_3}$, each with dimensions $1 \times 1 \times C$. **(2) Select:** this operator applies the softmax function to $\mathbf{v_1}, \mathbf{v_2}$ and $\mathbf{v_3}$, yielding attention activations $\mathbf{s_1}, \mathbf{s_2}$ and $\mathbf{s_3}$ that we use to adaptively recalibrate multi-scale feature maps $\mathbf{L_1}, \mathbf{L_2}$ and $\mathbf{L_3}$, respectively. The overall process of feature recalibration and aggregation is defined as: $\mathbf{U} = \mathbf{s_1} \cdot \mathbf{L_1} + \mathbf{s_2} \cdot \mathbf{L_2} + \mathbf{s_3} \cdot \mathbf{L_3}$. Note that the SKFF uses $\sim 6\times$ fewer parameters than aggregation with concatenation but generates more favorable results (an ablation study is provided in the experiments section).

Dual attention unit (DAU). While the SKFF block fuses information across multi-resolution branches, we also need a mechanism to share information within a feature tensor, both along the spatial and the channel dimensions. Motivated by the advances of recent low-level vision methods [125,5,23,126] based on the attention mechanisms [44,103], we propose the dual attention unit (DAU) to extract features in the convolutional streams. The schematic of DAU is shown in Fig. 3. The DAU suppresses less useful features and only allows more informative ones to pass further. This feature recalibration is achieved by using channel attention [44] and spatial attention [108] mechanisms. **(1) Channel attention**

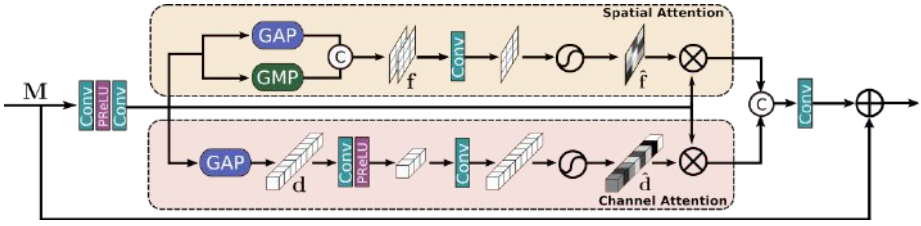
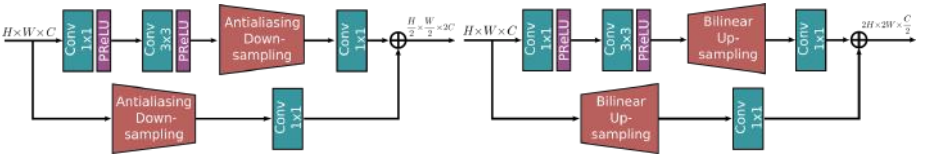


Fig. 3: Dual attention unit incorporating spatial and channel attention mechanisms.



(a) Downsampling module (b) Upsampling module

Fig. 4: Residual resizing modules to perform downsampling and upsampling.

(CA) branch exploits the inter-channel relationships of the convolutional feature maps by applying *squeeze* and *excitation* operations [44]. Given a feature map $\mathbf{M} \in \mathbb{R}^{H\times W\times C}$, the squeeze operation applies global average pooling across spatial dimensions to encode global context, thus yielding a feature descriptor $\mathbf{d} \in \mathbb{R}^{1\times 1\times C}$. The excitation operator passes $\mathbf{d}$ through two convolutional layers followed by the sigmoid gating and generates activations $\hat{\mathbf{d}} \in \mathbb{R}^{1\times 1\times C}$. Finally, the output of CA branch is obtained by rescaling $\mathbf{M}$ with the activations $\hat{\mathbf{d}}$.
(2) Spatial attention (SA) branch is designed to exploit the inter-spatial dependencies of convolutional features. The goal of SA is to generate a spatial attention map and use it to recalibrate the incoming features $\mathbf{M}$. To generate the spatial attention map, the SA branch first independently applies global average pooling and max pooling operations on features $\mathbf{M}$ along the channel dimensions and concatenates the outputs to form a feature map $\mathbf{f} \in \mathbb{R}^{H\times W\times 2}$. The map $\mathbf{f}$ is passed through a convolution and sigmoid activation to obtain the spatial attention map $\hat{\mathbf{f}} \in \mathbb{R}^{H\times W\times 1}$, which we then use to rescale $\mathbf{M}$.

Residual resizing modules. The proposed framework employs a recursive residual design (with skip connections) to ease the flow of information during the learning process. In order to maintain the residual nature of our architecture, we introduce residual resizing modules to perform downsampling (Fig. 4a) and upsampling (Fig. 4b) operations. In MRB, the size of feature maps remains constant along convolution streams. On the other hand, across streams the feature map size changes depending on the input resolution index i and the output resolution index j. If $i < j$, the input feature tensor is downsampled, and if $i > j$, the feature map is upsampled. To perform $2\times$ downsampling (halving the spatial dimension and doubling the channel dimension), we apply the module in Fig. 4a only once. For $4\times$ downsampling, the module is applied twice, consecutively. Similarly, one can perform $2\times$ and $4\times$ upsampling by applying the module in Fig. 4b once and twice, respectively. Note in Fig. 4a, we integrate anti-aliasing downsampling [123] to improve the shift-equivariance of our network.

4 Experiments

In this section, we perform qualitative and quantitative assessment of the results produced by our MIRNet and compare it with the previous best methods. Next, we describe the datasets, and then provide the implementation details. Finally, we report results for **(a)** image denoising, **(b)** super-resolution and **(c)** image enhancement on five real image datasets.

4.1 Real Image Datasets

Image denoising. (1) DND [79] consists of 50 images captured with four consumer cameras. Since the images are of very high-resolution, the dataset providers extract 20 crops of size 512×512 from each image, yielding 1000 patches in total. All these patches are used for testing (as DND does not contain training or validation sets). The ground-truth noise-free images are not released publicly, therefore the image quality scores in terms of PSNR and SSIM can only be obtained through an online server [24]. **(2) SIDD** [1] is particularly collected with smartphone cameras. Due to the small sensor and high-resolution, the noise levels in smartphone images are much higher than those of DSLRs. SIDD contains 320 image pairs for training and 1280 for validation.

Super-resolution. (1) RealSR [14] contains real-world LR-HR image pairs of the same scene captured by adjusting the focal-length of the cameras. RealSR has both indoor and outdoor images taken with two cameras. The number of training image pairs for scale factors $\times 2$, $\times 3$ and $\times 4$ are 183, 234 and 178, respectively. For each scale factor, 30 test images are also provided in RealSR.

Image enhancement. (1) LoL [107] is created for low-light image enhancement problem. It provides 485 images for training and 15 for testing. Each image pair in LoL consists of a low-light input image and its corresponding well-exposed reference image. **(2) MIT-Adobe FiveK** [12] contains 5000 images of various indoor and outdoor scenes captured with DSLR cameras in different lighting conditions. The tonal attributes of all images are manually adjusted by five different trained photographers (labelled as experts A to E). Same as in [45,75,100], we also consider the enhanced images of expert C as the ground-truth. Moreover, the first 4500 images are used for training and the last 500 for testing.

4.2 Implementation Details

The proposed architecture is end-to-end trainable and requires no pre-training of sub-modules. We train three different networks for three different restoration tasks. The training parameters, common to all experiments, are the following. We use 3 RRGs, each of which further contains 2 MRBs. The MRB consists of 3 parallel streams with channel dimensions of $64, 128, 256$ at resolutions $1, \frac{1}{2}, \frac{1}{4}$, respectively. Each stream has 2 DAUs. The models are trained with the Adam optimizer ($\beta_1 = 0.9$, and $\beta_2 = 0.999$) for 7×10^5 iterations. The initial learning rate is set to 2×10^{-4}. We employ the cosine annealing strategy [69] to steadily

Table 1: Denoising comparisons on the SIDD dataset [1].

Method	DnCNN [120]	MLP [11]	GLIDE [96]	TNRD [20]	FoE [85]	BM3D [21]	WNNM [38]	NLM [10]	KSVD [2]	EPLL [128]	CBDNet [39]	RIDNet [5]	VDN [118]	MIRNet (Ours)
PSNR ↑	23.66	24.71	24.71	24.73	25.58	25.65	25.78	26.76	26.88	27.11	30.78	38.71	39.28	**39.72**
SSIM ↑	0.583	0.641	0.774	0.643	0.792	0.685	0.809	0.699	0.842	0.870	0.754	0.914	0.909	**0.959**

Table 2: Denoising comparisons on the DND dataset [79].

Method	EPLL [128]	TNRD [20]	MLP [11]	BM3D [21]	FoE [85]	WNNM [38]	KSVD [2]	MCWNNM [112]	FFDNet+ [121]	TWSC [111]	CBDNet [39]	RIDNet [5]	VDN [118]	MIRNet (Ours)
PSNR ↑	33.51	33.65	34.23	34.51	34.62	34.67	36.49	37.38	37.61	37.94	38.06	39.26	39.38	**39.88**
SSIM ↑	0.824	0.831	0.833	0.851	0.885	0.865	0.898	0.929	0.942	0.940	0.942	0.953	0.952	**0.956**

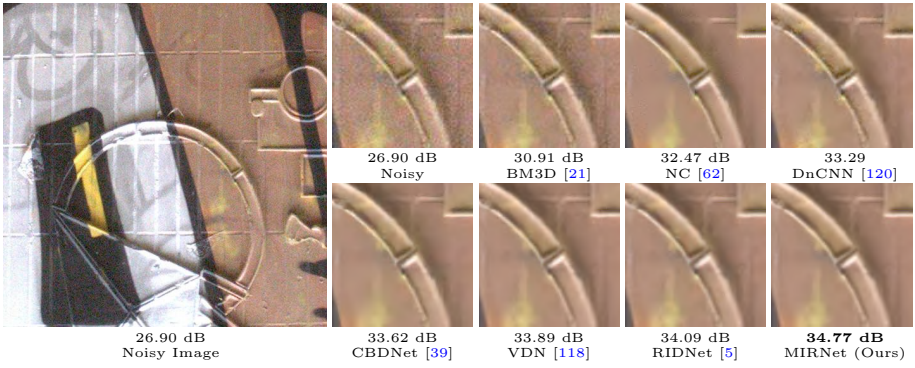


Fig. 5: Denoising example from DND [79]. Our MIRNet generates visually-pleasing and artifact-free results.

decrease the learning rate from initial value to 10^{-6} during training. We extract patches of size 128×128 from training images. The batch size is set to 16 and, for data augmentation, we perform horizontal and vertical flips.

4.3 Image Denoising

In this section, we demonstrate the effectiveness of the proposed MIRNet for image denoising. We train our network only on the training set of the SIDD [1] and directly evaluate it on the test images of both SIDD and DND [79] datasets. Quantitative comparisons in terms of PSNR and SSIM metrics are summarized in Table 1 and Table 2 for SIDD and DND, respectively. Both tables show that our MIRNet performs favourably against the data-driven, as well as conventional, denoising algorithms. Specifically, when compared to the recent best method VDN [118], our algorithm demonstrates a performance gain of 0.44 dB on SIDD and 0.50 dB on DND. Furthermore, it is worth noting that CBDNet [39] and RIDNet [5] use additional training data, yet our method provides significantly better results. For instance, our method achieves 8.94 dB improvement over CBDNet [39] on the SIDD dataset and 1.82 dB on DND.

In Fig. 5 and Fig. 6, we present visual comparisons of our results with those of other competing algorithms. It can be seen that our MIRNet is effective in removing real noise and produces perceptually-pleasing and sharp images. More-

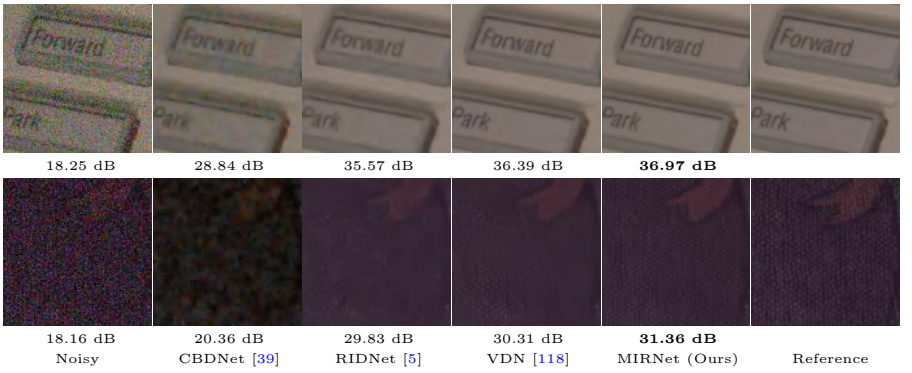


Fig. 6: Denoising examples from SIDD [1]. Our method effectively removes real noise from challenging images, while better recovering structural content and fine texture.

Table 3: Super-resolution evaluation on the RealSR [14] dataset. Compared to the state-of-the-art, our method consistently yields significantly better image quality scores for all three scaling factors.

Scale	Bicubic		VDSR [56]		SRResNet [63]		RCAN [125]		LP-KPN [14]		MIRNet (Ours)	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×2	32.61	0.907	33.64	0.917	33.69	0.919	33.87	0.922	33.90	0.927	**34.35**	**0.935**
×3	29.34	0.841	30.14	0.856	30.18	0.859	30.40	0.862	30.42	0.868	**31.16**	**0.885**
×4	27.99	0.806	28.63	0.821	28.67	0.824	28.88	0.826	28.92	0.834	**29.14**	**0.843**

over, it is capable of maintaining the spatial smoothness of the homogeneous regions without introducing artifacts. In contrast, most of the other methods either yield over-smooth images and thus sacrifice structural content and fine textural details, or produce images with chroma artifacts and blotchy texture.

Generalization capability. The DND and SIDD datasets are acquired with different sets of cameras having different noise characteristics. Since the DND benchmark does not provide training data, setting a new state-of-the-art on DND with our SIDD trained network indicates the good generalization capability of our approach.

4.4 Super-Resolution (SR)

We compare our MIRNet against the state-of-the-art SR algorithms (VDSR [56], SRResNet [63], RCAN [125], LP-KPN [14]) on the testing images of the RealSR [14] for upscaling factors of ×2, ×3 and ×4. Note that all the benchmarked algorithms are trained on the RealSR [14] dataset for a fair comparison. In the experiments, we also include bicubic interpolation [55], which is the most commonly used method for generating super-resolved images. Here, we compute the PSNR and SSIM scores using the Y channel (in YCbCr color space), as it is a common practice in the SR literature [125,14,106,6]. The results in Table 3 show that the bicubic interpolation provides the least accurate results, thereby indicating its low suitability for dealing with real images. Moreover, the same

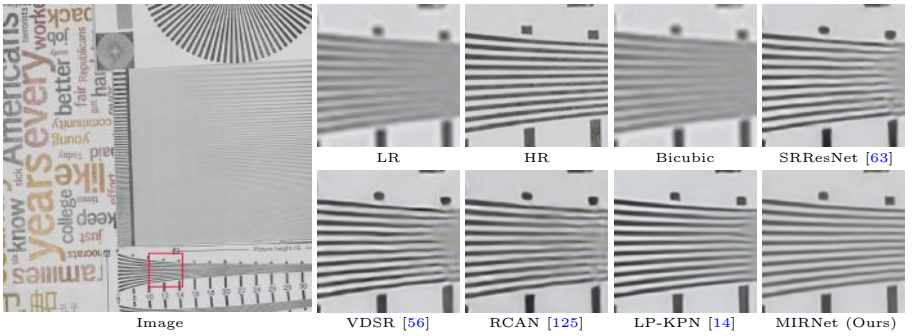


Fig. 7: Comparisons for ×4 super-resolution from the RealSR [14] dataset. The image produced by our MIRNet is more faithful to the ground-truth than other competing methods (see lines near the right edge of the crops).

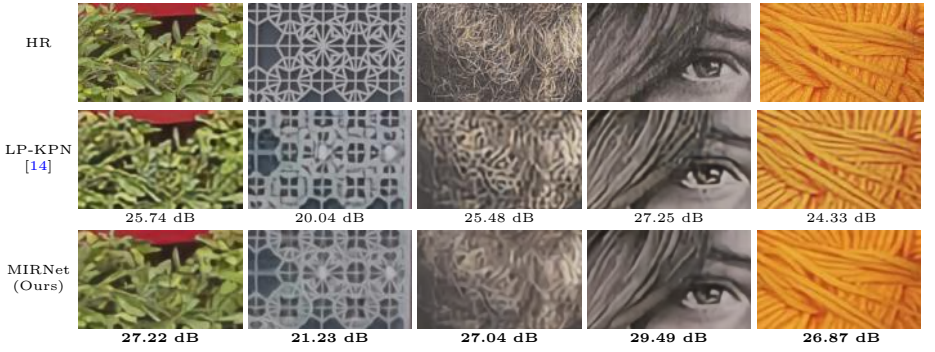


Fig. 8: Additional visual examples for ×4 super-resolution, comparing our MIRNet against the previous best approach [14]. Note that all example crops are taken from different images. The full-resolution versions (and many more examples) are provided in the supplementary material.

table shows that the recent method LP-KPN [14] provides marginal improvement of only ∼ 0.04 dB over the previous best method RCAN [125]. In contrast, our method significantly advances state-of-the-art and consistently yields better image quality scores than other approaches for all three scaling factors. Particularly, compared to LP-KPN [14], our method provides performance gains of 0.45 dB, 0.74 dB, and 0.22 dB for scaling factors ×2, ×3 and ×4, respectively. The trend is similar for the SSIM metric as well.

Visual comparisons in Fig. 7 show that our MIRNet recovers content structures effectively. In contrast, VDSR [56], SRResNet [63] and RCAN [125] reproduce results with noticeable artifacts. Furthermore, LP-KPN [14] is not able to preserve structures (see near the right edge of the crop). Several more examples are provided in Fig. 8 to further compare the image reproduction quality of our method against the previous best method [14]. It can be seen that LP-KPN [14] has a tendency to over-enhance the contrast (cols. 1, 3, 4) and in turn causes loss of details near dark and high-light areas. In contrast, the proposed MIR-

Table 4: Cross-camera generalization test for super-resolution. Networks trained for one camera are tested on the other camera. Our MIRNet shows good generalization for all possible cases.

Tested on	Scale	Bicubic	RCAN [125] (Trained on)		LP-KPN [14] (Trained on)		MIRNet (Trained on)	
			Canon	Nikon	Canon	Nikon	Canon	Nikon
Canon	×2	33.05	34.34	34.11	34.38	34.18	**35.41**	**35.14**
	×3	29.67	30.65	30.28	30.69	30.33	**31.97**	**31.56**
	×4	28.31	29.46	29.04	29.48	29.10	**30.35**	**29.95**
Nikon	×2	31.66	32.01	32.30	32.05	32.33	**32.58**	**33.19**
	×3	28.63	29.30	29.75	29.34	29.78	**29.71**	**30.05**
	×4	27.28	27.98	28.12	28.01	28.13	**28.16**	**28.37**

Table 5: Low-light image enhancement evaluation on the LoL dataset [107]. The proposed method significantly advances the state-of-the-art.

Method	BIMEF [116]	CRM [117]	Dong [29]	LIME [40]	MF [34]	RRM [67]	SRIE [34]	Retinex-Net [107]	MSR [54]	NPE [101]	GLAD [102]	KinD [124]	MIRNet (Ours)
PSNR	13.86	17.20	16.72	16.76	18.79	13.88	11.86	16.77	13.17	16.97	19.72	20.87	**24.14**
SSIM	0.58	0.64	0.58	0.56	0.64	0.66	0.50	0.56	0.48	0.59	0.70	0.80	**0.83**

Table 6: Image enhancement comparisons on the MIT-Adobe FiveK dataset [12].

Method	HDRNet [36]	W-Box [45]	DR [75]	DPE [19]	DeepUPE [100]	MIRNet (Ours)
PSNR	21.96	18.57	20.97	22.15	23.04	**23.73**
SSIM	0.866	0.701	0.841	0.850	0.893	**0.925**

Net successfully reconstructs structural patterns and edges (col. 2) and produces images that are natural (cols. 1, 4) and have better color reproduction (col. 5).

Cross-camera generalization. The RealSR [14] dataset consists of images taken with Canon and Nikon cameras at three scaling factors. To test the cross-camera generalizability of our method, we train the network on the training images of one camera and directly evaluate it on the test set of the other camera. Table 4 demonstrates the generalization of competing methods for four possible cases: (a) training and testing on Canon, (b) training on Canon, testing on Nikon, (c) training and testing on Nikon, and (d) training on Nikon, testing on Canon. It can be seen that, for all scales, LP-KPN [14] and RCAN [125] shows comparable performance. In contrast, our MIRNet exhibits more promising generalization.

4.5 Image Enhancement

In this section, we demonstrate the effectiveness of our algorithm by evaluating it for the image enhancement task. We report PSNR/SSIM values of our method and several other techniques in Table 5 and Table 6 for the LoL [107] and MIT-Adobe FiveK [12] datasets, respectively. It can be seen that our MIRNet achieves significant improvements over previous approaches. Notably, when compared to the recent best methods, MIRNet obtains 3.27 dB performance gain over KinD [124] on the LoL dataset and 0.69 dB improvement over DeepUPE[5] [100] on the Adobe-Fivek dataset.

[5] Note that the quantitative results reported in [100] are incorrect. The correct scores are later released by the original authors [link].

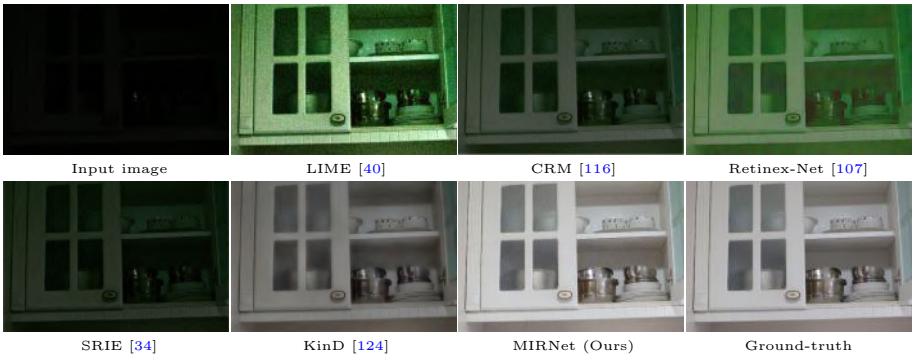


Fig. 9: Visual comparison of low-light enhancement approaches on the LoL dataset [107]. Our method reproduces image that is visually closer to the ground-truth in terms of brightness and global contrast.

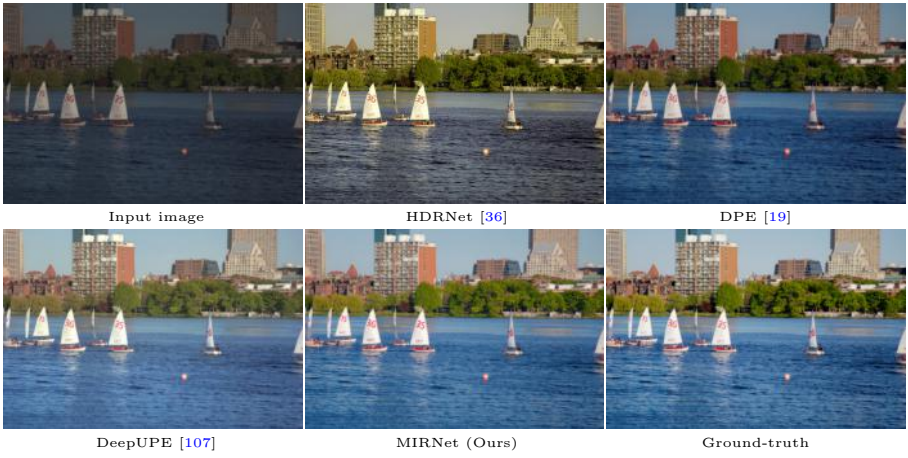


Fig. 10: Visual results of image enhancement on the MIT-Adobe FiveK [12] dataset. Compared to the state-of-the-art, our MIRNet makes better color and contrast adjustments and produces image that is vivid, natural and pleasant in appearance.

We show visual results in Fig. 9 and Fig. 10. Compared to other techniques, our method generates enhanced images that are natural and vivid in appearance and have better global and local contrast.

5 Ablation Studies

We study the impact of each of our architectural components and design choices on the final performance. All the ablation experiments are performed for the super-resolution task with $\times 3$ scale factor. Table 7 shows that removing skip connections causes the largest performance drop. Without skip connections, the network finds it difficult to converge and yields high training errors, and consequently low PSNR. Furthermore, the information exchange among parallel

Table 7: Impact of individual components of MRB.

Skip connections		✓	✓	✓	✓
DAU	✓		✓		✓
SKFF intermediate	✓	✓			✓
SKFF final	✓	✓	✓	✓	✓
PSNR (in dB)	27.91	30.97	30.78	30.57	**31.16**

Table 8: Feature aggregation. Our SKFF uses $\sim 6\times$ fewer parameters than concat, but generates better results.

Method	Sum	Concat	SKFF
PSNR (in dB)	30.76	30.89	**31.16**
Parameters	0	12,288	2,049

Table 9: Ablation study on different layouts of MRB. *Rows* denote the number of parallel resolution streams, and *Cols* represent the number of columns containing DAUs.

	Rows = 1			Rows = 2			Rows = 3		
	Cols = 1	Cols = 2	Cols = 3	Cols = 1	Cols = 2	Cols = 3	Cols = 1	Cols = 2	Cols = 3
PSNR	29.92	30.11	30.17	30.15	30.83	30.92	30.24	31.16	31.18

convolution streams via SKFF is helpful and leads to improved performance. Similarly, DAU also makes a positive influence to the final image quality.

Next, we analyze the feature aggregation strategy in Table 8. It shows that the proposed SKFF generates favorable results compared to summation and concatenation. Moreover, it can be seen that our SKFF uses $\sim 6\times$ fewer parameters than concatenation. Finally, in Table 9 we study how the number of convolutional streams and columns (DAU blocks) of MRB affect the image restoration quality. We note that increasing the number of streams provides significant improvements, thereby justifying the importance of multi-scale features processing. Moreover, increasing the number of columns yields better scores, thus indicating the significance of information exchange among parallel streams for feature consolidation. Additional ablation studies and qualitative results are provided in the supplementary material.

6 Concluding Remarks

Conventional image restoration and enhancement pipelines either stick to the full resolution features along the network hierarchy or use an encoder-decoder architecture. The first approach helps retain precise spatial details, while the latter one provides better contextualized representations. However, these methods can satisfy only one of the above two requirements, although real-world image restoration tasks demand a combination of both conditioned on the given input sample. In this work, we propose a novel architecture whose main branch is dedicated to full-resolution processing and the complementary set of parallel branches provides better contextualized features. We propose novel mechanisms to learn relationships between features within each branch as well as across multi-scale branches. Our feature fusion strategy ensures that the receptive field can be dynamically adapted without sacrificing the original feature details. Consistent achievement of state-of-the-art results on five datasets for three image restoration and enhancement tasks corroborates the effectiveness of our approach.

Acknowledgments. Ming-Hsuan Yang is supported by the NSF CAREER Grant 1149783.

References

1. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: CVPR (2018) 8, 9, 10
2. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. Trans. Sig. Proc. (2006) 9
3. Ahn, N., Kang, B., Sohn, K.A.: Fast, accurate, and lightweight super-resolution with cascading residual network. In: ECCV (2018) 4
4. Allebach, J., Wong, P.W.: Edge-directed interpolation. In: ICIP (1996) 3
5. Anwar, S., Barnes, N.: Real image denoising with feature attention. ICCV (2019) 3, 6, 9, 10
6. Anwar, S., Khan, S., Barnes, N.: A deep journey into super-resolution: A survey. arXiv (2019) 3, 10
7. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. TPAMI (2017) 5
8. Bertalmío, M., Caselles, V., Provenzi, E., Rizzi, A.: Perceptual color correction through variational techniques. TIP (2007) 4
9. Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: Unprocessing images for learned raw denoising. In: CVPR (2019) 3
10. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: CVPR (2005) 3, 9
11. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with BM3D? In: CVPR (2012) 3, 9
12. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In: CVPR (2011) 8, 12, 13
13. Cai, J., Gu, S., Timofte, R., Zhang, L.: Ntire 2019 challenge on real image super-resolution: Methods and results. In: CVPRW (2019) 3
14. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model. In: ICCV (2019) 8, 10, 11, 12
15. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: CVPR (2004) 3
16. Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: ICIP (1994) 5
17. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: CVPR (2018) 2
18. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: ECCV (2018) 4
19. Chen, Y.S., Wang, Y.C., Kao, M.H., Chuang, Y.Y.: Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In: CVPR (2018) 4, 12, 13
20. Chen, Y., Yu, W., Pock, T.: On learning optimized reaction diffusion processes for effective image restoration. In: CVPR (2015) 9
21. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. TIP (2007) 3, 9
22. Dahl, R., Norouzi, M., Shlens, J.: Pixel recursive super resolution. In: ICCV (2017) 4

23. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: CVPR (2019) 4, 6
24. https://noise.visinf.tu-darmstadt.de/benchmark/ (2017), [Online; accessed 29-Feb-2020] 8
25. Deng, Y., Loy, C.C., Tang, X.: Aesthetic-driven image enhancement by adversarial learning. In: ACM Multimedia (2018) 4
26. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV (2014) 3
27. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. TPAMI (2015) 2, 3
28. Dong, W., Shi, G., Li, X.: Nonlocal image restoration with bilateral variance estimation: a low-rank approach. TIP (2012) 3
29. Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., Lu, Y.: Fast efficient algorithm for enhancement of low lighting video. In: ICME (2011) 12
30. Donoho, D.L.: De-noising by soft-thresholding. Trans. on information theory (1995) 3
31. Efros, A.A., Leung, T.K.: Texture synthesis by non-parametric sampling. In: ICCV (1999) 3
32. Fourure, D., Emonet, R., Fromont, É., Muselet, D., Trémeau, A., Wolf, C.: Residual conv-deconv grid network for semantic segmentation. In: BMVC (2017) 5
33. Freedman, G., Fattal, R.: Image and video upscaling from local self-examples. TOG (2011) 3
34. Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: CVPR (2016) 12, 13
35. Gharbi, M., Chaurasia, G., Paris, S., Durand, F.: Deep joint demosaicking and denoising. TOG (2016) 3
36. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. TOG (2017) 12, 13
37. Gu, S., Li, Y., Gool, L.V., Timofte, R.: Self-guided network for fast image denoising. In: ICCV (2019) 2
38. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: CVPR (2014) 3, 9
39. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: CVPR (2019) 3, 9, 10
40. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. TIP (2016) 12, 13
41. Han, W., Chang, S., Liu, D., Yu, M., Witbrock, M., Huang, T.S.: Image super-resolution via dual-state recurrent networks. In: CVPR (2018) 4
42. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016) 4
43. Hedjam, R., Moghaddam, R.F., Cheriet, M.: Markovian clustering for the non-local means image denoising. In: ICIP (2009) 3
44. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: CVPR (2018) 6, 7
45. Hu, Y., He, H., Xu, C., Wang, B., Lin, S.: Exposure: A white-box photo post-processing framework. TOG (2018) 8, 12
46. Huang, G., Chen, D., Li, T., Wu, F., van der Maaten, L., Weinberger, K.Q.: Multi-scale dense networks for resource efficient image classification. In: ICLR (2018) 5

47. Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology (1962) 5
48. Hui, Z., Wang, X., Gao, X.: Fast and accurate single image super-resolution via information distillation network. In: CVPR (2018) 3
49. Hung, C.P., Kreiman, G., Poggio, T., DiCarlo, J.J.: Fast readout of object identity from macaque inferior temporal cortex. Science (2005) 5
50. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: DSLR-quality photos on mobile devices with deep convolutional networks. In: ICCV (2017) 2
51. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: Wespe: weakly supervised photo enhancer for digital cameras. In: CVPRW (2018) 4
52. Ignatov, A., Timofte, R.: Ntire 2019 challenge on image enhancement: Methods and results. In: CVPRW (2019) 4
53. Irani, M., Peleg, S.: Improving resolution by image registration. CVGIP (1991) 3
54. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. TIP (1997) 4, 12
55. Keys, R.: Cubic convolution interpolation for digital image processing. TASSP (1981) 3, 10
56. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: ICCV (2016) 3, 10, 11
57. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: CVPR (2016) 4
58. Kim, K.I., Kwon, Y.: Single-image super-resolution using sparse regression and natural image prior. TPAMI (2010) 3
59. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: ICCV (2019) 2
60. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate superresolution. In: CVPR (2017) 4
61. Land, E.H.: The retinex theory of color vision. Scientific american (1977) 4
62. Lebrun, M., Colom, M., Morel, J.M.: The noise clinic: a blind image denoising algorithm. IPOL (2015) 9
63. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR (2017) 4, 10, 11
64. Li, J., Fang, F., Mei, K., Zhang, G.: Multi-scale residual network for image super-resolution. In: ECCV (2018) 4
65. Li, X., Wang, W., Hu, X., Yang, J.: Selective kernel networks. In: CVPR (2019) 5, 6
66. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: CVPRW (2017) 4
67. Liu, Y., Wang, R., Shan, S., Chen, X.: Structure inference net: Object detection using scene-level context and instance-level relationships. In: CVPR (2018) 12
68. Lore, K.G., Akintayo, A., Sarkar, S.: LLNet: a deep autoencoder approach to natural low-light image enhancement. Pattern Recognition (2017) 4
69. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. In: ICLR (2017) 8
70. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: ICCV (2009) 3
71. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: CVPR (2017) 2

72. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: ECCV (2016) 5
73. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: ICCV (2015) 5
74. Palma-Amestoy, R., Provenzi, E., Bertalmío, M., Caselles, V.: A perceptually inspired variational framework for color enhancement. TPAMI (2009) 4
75. Park, J., Lee, J.Y., Yoo, D., So Kweon, I.: Distort-and-recover: Color enhancement using deep reinforcement learning. In: CVPR (2018) 8, 12
76. Park, S.J., Son, H., Cho, S., Hong, K.S., Lee, S.: SRFEAT: Single image super-resolution with feature discrimination. In: ECCV (2018) 4
77. Peng, X., Feris, R.S., Wang, X., Metaxas, D.N.: A recurrent encoder-decoder network for sequential face alignment. In: ECCV (2016) 5
78. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. TPAMI (1990) 3
79. Plotz, T., Roth, S.: Benchmarking denoising algorithms with real photographs. In: CVPR (2017) 8, 9
80. Plötz, T., Roth, S.: Neural nearest neighbors networks. In: NeurIPS (2018) 3
81. Ren, W., Liu, S., Ma, L., Xu, Q., Xu, X., Cao, X., Du, J., Yang, M.H.: Low-light image enhancement via a deep hybrid network. TIP (2019) 4
82. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. Nature neuroscience (1999) 5
83. Rizzi, A., Gatta, C., Marini, D.: From retinex to automatic color equalization: issues in developing a new algorithm for unsupervised color equalization. Journal of Electronic Imaging (2004) 4
84. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: MICCAI (2015) 2, 5
85. Roth, S., Black, M.J.: Fields of experts. IJCV (2009) 9
86. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena (1992) 3
87. Sajjadi, M.S., Scholkopf, B., Hirsch, M.: Enhancenet: Single image super-resolution through automated texture synthesis. In: ICCV (2017) 4
88. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Robust object recognition with cortex-like mechanisms. TPAMI (2007) 5
89. Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., Ma, J.: Msr-net: Low-light image enhancement using deep convolutional network. arXiv (2017) 4
90. Simoncelli, E.P., Adelson, E.H.: Noise removal via bayesian wavelet coring. In: ICIP (1996) 3
91. Smith, S.M., Brady, J.M.: SUSANa new approach to low level image processing. IJCV (1997) 3
92. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: CVPR (2019) 5
93. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: CVPR (2015) 5
94. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: CVPR (2017) 3
95. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: ICCV (2017) 3
96. Talebi, H., Milanfar, P.: Global image denoising. TIP (2013) 9
97. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: CVPR (2018) 2

98. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: ICCV (1998) 3
99. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: ICCV (2017) 4
100. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: CVPR (2019) 8, 12
101. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. TIP (2013) 12
102. Wang, W., Wei, C., Yang, W., Liu, J.: Gladnet: Low-light enhancement network with global awareness. In: FG (2018) 12
103. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: CVPR (2018) 6
104. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: ESRGAN: enhanced super-resolution generative adversarial networks. In: ECCVW (2018) 4
105. Wang, Z., Liu, D., Yang, J., Han, W., Huang, T.: Deep networks for image super-resolution with sparse prior. In: ICCV (2015) 4
106. Wang, Z., Chen, J., Hoi, S.C.: Deep learning for image super-resolution: A survey. TPAMI (2019) 3, 10
107. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. BMVC (2018) 4, 8, 12, 13
108. Woo, S., Park, J., Lee, J.Y., So Kweon, I.: CBAM: Convolutional block attention module. In: ECCV (2018) 6
109. Xiao, B., Wu, H., Wei, Y.: Simple baselines for human pose estimation and tracking. In: ECCV (2018) 5
110. Xiong, Z., Sun, X., Wu, F.: Robust web image/video super-resolution. TIP (2010) 3
111. Xu, J., Zhang, L., Zhang, D.: A trilateral weighted sparse coding scheme for real-world image denoising. In: ECCV (2018) 9
112. Xu, J., Zhang, L., Zhang, D., Feng, X.: Multi-channel weighted nuclear norm minimization for real color image denoising. In: ICCV (2017) 9
113. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: CVPR (2008) 3
114. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. TIP (2010) 3
115. Yaroslavsky, L.P.: Local adaptive image restoration and enhancement with the use of DFT and DCT in a running window. In: Wavelet Applications in Signal and Image Processing IV (1996) 3
116. Ying, Z., Li, G., Gao, W.: A bio-inspired multi-exposure fusion framework for low-light image enhancement. arXiv preprint arXiv:1711.00591 (2017) 12, 13
117. Ying, Z., Li, G., Ren, Y., Wang, R., Wang, W.: A new image contrast enhancement algorithm using exposure fusion framework. In: CAIP (2017) 12
118. Yue, Z., Yong, H., Zhao, Q., Meng, D., Zhang, L.: Variational denoising network: Toward blind noise modeling and removal. In: NeurIPS (2019) 9, 10
119. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: CycleISP: Real image restoration via improved data synthesis. In: CVPR (2020) 3
120. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. TIP (2017) 2, 3, 9
121. Zhang, K., Zuo, W., Zhang, L.: FFDNet: Toward a fast and flexible solution for CNN-based image denoising. TIP (2018) 3, 9

122. Zhang, L., Wu, X.: An edge-guided image interpolation algorithm via directional filtering and data fusion. TIP (2006) 3
123. Zhang, R.: Making convolutional networks shift-invariant again. In: ICML (2019) 7
124. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: MM (2019) 2, 4, 12, 13
125. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV (2018) 4, 6, 10, 11, 12
126. Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Residual non-local attention networks for image restoration. In: ICLR (2019) 4, 6
127. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image restoration. TPAMI (2020) 2, 4
128. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: ICCV (2011) 9