International Journal of Computer Vision
https://doi.org/10.1007/s11263-019-01285-y

Gated Fusion Network for Degraded Image Super Resolution

Xinyi Zhang[1] · Hang Dong[2] · Zhe Hu[3] · Wei-Sheng Lai[4] · Fei Wang[2] · Ming-Hsuan Yang[4]

Received: 21 January 2019 / Accepted: 15 December 2019

Abstract
Single image super resolution aims to enhance image quality with respect to spatial content, which is a fundamental task in computer vision. In this work, we address the task of single frame super resolution with the presence of image degradation, e.g., blur, haze, or rain streaks. Due to the limitations of frame capturing and formation processes, image degradation is inevitable, and the artifacts would be exacerbated by super resolution methods. To address this problem, we propose a dual-branch convolutional neural network to extract base features and recovered features separately. The base features contain local and global information of the input image. On the other hand, the recovered features focus on the degraded regions and are used to remove the degradation. Those features are then fused through a recursive gate module to obtain sharp features for super resolution. By decomposing the feature extraction step into two task-independent streams, the dual-branch model can facilitate the training process by avoiding learning the mixed degradation all-in-one and thus enhance the final high-resolution prediction results. We evaluate the proposed method in three degradation scenarios. Experiments on these scenarios demonstrate that the proposed method performs more efficiently and favorably against the state-of-the-art approaches on benchmark datasets.

Keywords Super resolution · Image restoration · Deep learning

1 Introduction

Single image super resolution (SISR) aims to restore a high-resolution (HR) image from a low-resolution (LR) one, such as those captured from surveillance and mobile cameras. The generated HR image can improve the performance of the numerous high-level vision tasks, e.g., object detection (Zhang et al. 2011), face recognition (Bai et al. 2018), and surveillance applications (Zhang et al. 2010; Zou and Yuen 2012). However, image degradation is often inevitable due to the limitations of the imaging processors and complex capturing scenes. For example, motion blur, as well as hazy and rainy weather would introduce undesired artifacts in the captured LR images. Those artifacts cannot be fully removed by the imaging formation pipeline and would adversely affect the super resolution algorithms and the following high-level tasks. The problems of super resolution and image restoration from degradation are often dealt with separately, as each one is known to be ill-posed. However, such a strategy is neither optimal nor efficient due to error accumulation.

In this work, we address the joint problem of single image super resolution and restoration. We evaluate the proposed super resolution architecture on images with three representative image degradations: motion blur, rain streaks, and haze.

Communicated by Yasuyuki Matsushita.

Xinyi Zhang and Hang Dong have equally contributed to this work.

✉ Fei Wang
wfx@mail.xjtu.edu.cn

Xinyi Zhang
jacqueline@stu.xjtu.edu.cn

Hang Dong
dhunter@stu.xjtu.edu.cn

Zhe Hu
zhe.hu@hikvision.com

Wei-Sheng Lai
wlai24@ucmerced.edu

Ming-Hsuan Yang
mhyang@ucmerced.edu

[1] School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China

[2] College of Artificial Intelligence, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China

[3] Hikvision Research America, Santa Clara, CA, USA

[4] Electrical Engineering and Computer Science, University of California, Merced, CA, USA

Published online: 13 January 2020

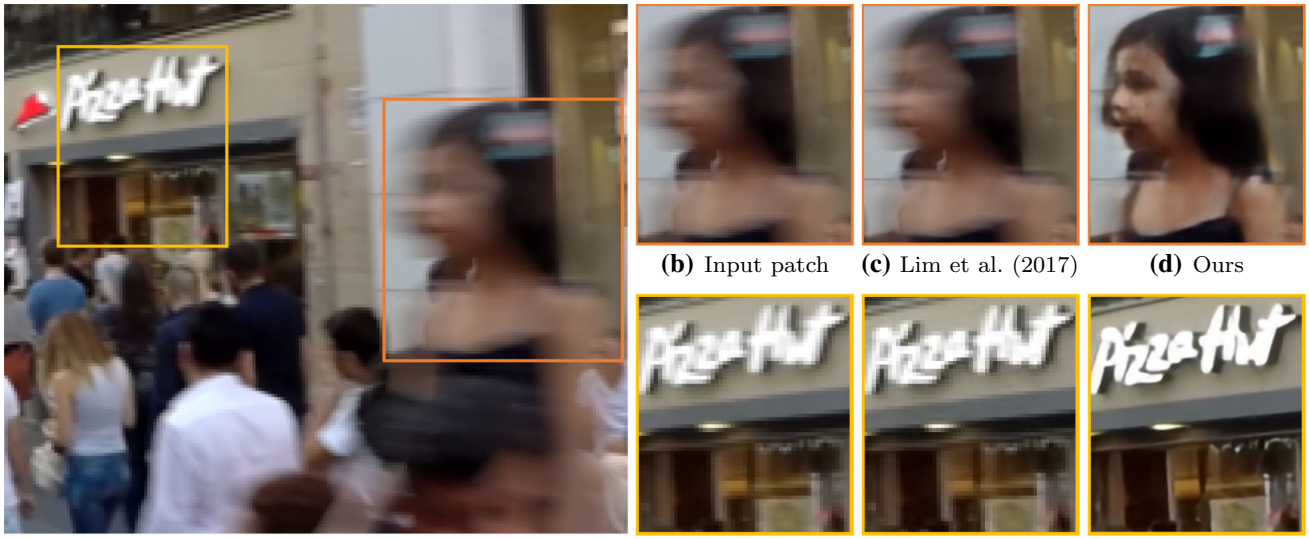


Fig. 1 Joint image deblurring and super resolution. While the state-of-the-art super resolution algorithm by Lim et al. (2017) does not reduce the non-uniform blur in the input image due to the assumption of bicubic downsampling, the top-performing non-uniform deblurring algorithm by Nah et al. (2017) generates sharp results but with few details. In contrast, the proposed model generates a sharp HR image with more details

Here, we use super resolution of a blurred image as the example to illustrate this joint task. Motion blur is often caused by camera shake, object motion, and scene depth variation. Figure 1 shows one blurry LR image, which contains non-uniform blur. As the existing super resolution algorithms (Lim et al. 2017; Ledig et al. 2017; Lai et al. 2019; Kim et al. 2016a) are not designed to handle motion blur explicitly, the resulting HR image is still blurry (see Fig. 1b, c). On the other hand, the state-of-the-art non-uniform deblurring methods (Noroozi et al. 2017; Gong et al. 2017; Nah et al. 2017; Kupyn et al. 2018) generate sharp image but cannot restore fine details or enlarge the spatial resolution (see Fig. 1e, f).

With the advances of deep Convolutional Neural Networks (CNNs), the state-of-the-art image super resolution (Lim et al. 2017; Ledig et al. 2017; Lai et al. 2019) and image restoration (Nah et al. 2017; Kupyn et al. 2018; Mei et al. 2018; Zhang et al. 2017; Zhang and Patel 2018b; Li et al. 2018d) methods are developed based on end-to-end networks and achieve promising performance. To jointly handle the image super resolution and degradation restoration, a straightforward approach is to solve the two sub-problems sequentially, i.e., performing image restoration followed by super resolution, or vice versa. However, there are numerous issues within such an approach. First, a simple concatenation of two models is prone to error accumulation. That is, the estimation error of the first model will be propagated and exacerbated in the second model. Second, the two-step network does not fully exploit the dependence between the two tasks. For example, the feature extraction and image reconstruction steps are performed twice and result in computational redundancy. As both the training and inference processes are memory and time consuming, these approaches cannot be applied to resource-constrained real-time applications, e.g., autonomous driving and video surveillance.

Several recent methods (Xu et al. 2017; Yu et al. 2018; Zhang et al. 2018a; Bao et al. 2017) jointly solve the degraded image super resolution problem using end-to-end networks. However, these methods focus on either domain-specific inputs, e.g., face and text (Xu et al. 2017; Yu et al. 2018) images, or extending the existing architecture to a particular degradation (Bao et al. 2017).

Zhang et al. (2018a) propose a network with two output branches to solve the joint deblurring and super resolution task on natural images. Although this method can be extended to super resolve other degraded images by changing the loss functions and training data, it does not perform well when severe degradation exists, e.g., non-uniform blur, heavy rain, or uneven haze. In this work, we aim to handle these severe degradations for natural images, which is more challenging.

We use a common image in a dynamic scene to illustrate the motivation of the proposed method. The blurry LR input is mixed with degraded (motion blur in this example) regions and relatively sharper regions. If we extract features from these two regions in one single branch, the training data contains noisy samples and thus makes it difficult to learn an effective model for deblurring. To address this problem, we propose a Gated Fusion Network (GFN) which consists

of two branches: a restoration branch to extract features for recovering the sharp LR image, and a base branch to extract features for fusing. We adopt a recursive gate module to adaptively fuse the features from two branches for super resolution. The fused features are then fed into an image reconstruction module to generate the sharp HR image. Extensive evaluations demonstrate that the proposed model performs favorably against the combination of the state-of-the-art super resolution and image restoration methods as well as the existing joint models in different applications.

The contributions of this work are threefold:

- To the best of our knowledge, the proposed method is the first generic deep learning architecture for image super resolution under different degradations.
- We decouple the joint problem into two sub-tasks for better network regularization. We propose a dual-branch network to extract the base features and recovered features separately and learn a recursive gate module for adaptive feature fusion.
- The proposed model entails low computational cost as most operations are performed in the LR space. Our model performs more efficiently than the combinations of the state-of-the-art super resolution and image restoration methods while achieving significant performance improvement.

2 Related Work

Both image super resolution and image restoration are fundamental problems in computer vision. In this section, we discuss image super resolution and restoration methods closely related to this work.

2.1 Image Super Resolution

Single image super resolution is an ill-posed problem as there are multiple HR images corresponding to the same LR input image. Conventional approaches learn the LR-to-HR mappings using sparse dictionaries (Timofte et al. 2014), random forest (Schulter et al. 2015), or self-similarity (Huang et al. 2015). In recent years, the CNN-based methods (Dong et al. 2016; Kim et al. 2016a) have demonstrated significant improvement against conventional super resolution approaches. Several techniques have been developed based on recursive learning (Tai et al. 2017; Kim et al. 2016b), pixel shuffling (Shi et al. 2016; Lim et al. 2017), Laplacian pyramid (Lai et al. 2019), back-projection (Haris et al. 2018), and channel attention (Zhang et al. 2018b). In addition, several approaches use the adversarial loss (Ledig et al. 2017), perceptual loss (Johnson et al. 2016), and texture loss (Sajjadi et al. 2017) to generate super resolution images. As most super resolution algorithms assume that the LR images are generated by a simple downsampling kernel, e.g., bicubic kernel, they do not perform well when the input images suffer from other unexpected degradation. In contrast, the proposed model is able to super resolve LR images with severe degradation.

2.2 Motion Deblurring

Most existing image deblurring approaches (Cho and Lee 2009; Xu et al. 2013; Pan et al. 2016; Shan et al. 2008; Schmidt et al. 2011, 2013) assume that the blur is uniform and spatially invariant across the entire image. However, due to depth variation and object motion, real-world images typically contain non-uniform blur. Several approaches address the non-uniform deblurring problem by jointly estimating blur kernels with scene depth (Paramanand and Rajagopalan b; Hu et al. 2014) or segmentation (Kim et al. 2013). As the kernel estimation step is computationally expensive, recent methods (Hradiš et al. 2015; Noroozi et al. 2017; Nah et al. 2017; Nimisha et al. 2017) learn deep CNNs to bypass the kernel estimation and efficiently solve the non-uniform deblurring problem. Kupyn et al. (2018) adopt the Wasserstein generative adversarial network (GAN) to generate realistic deblurred images and facilitate the object detection task.

2.3 Image Dehazing

Existing single image dehazing methods often rely on strong image priors or statistical assumptions (Fattal 2008; Tan 2008; He et al. 2011). Tan (2008) assumes that haze-free images should have higher contrast compared with corresponding hazy images. He et al. (2011) propose the dark channel prior for haze-free outdoor images and achieve impressive results. Recent algorithms (Ren et al. 2016; Zhang and Patel 2018a; Zhang et al. 2017) adopt deep CNNs to estimate the transmission map, a major component in the haze model, for reconstructing the haze-free outputs. However, inaccurate transmission maps often adversely affect the dehazing results (Zhang and Patel 2018a). Therefore, end-to-end architectures have been proposed (e.g., (Ren et al. 2018; Mei et al. 2018) to directly recover the haze-free image without estimating the transmission map.

2.4 Image Deraining

It is challenging to develop restoration algorithms to deal with images captured from outdoor scenes as the contents are complex, dynamic, and with large lighting variations. Existing deraining methods can be categorized as video-based (Li et al. 2018c; Liu et al. 2018; Jiang et al. 2017) or image-based

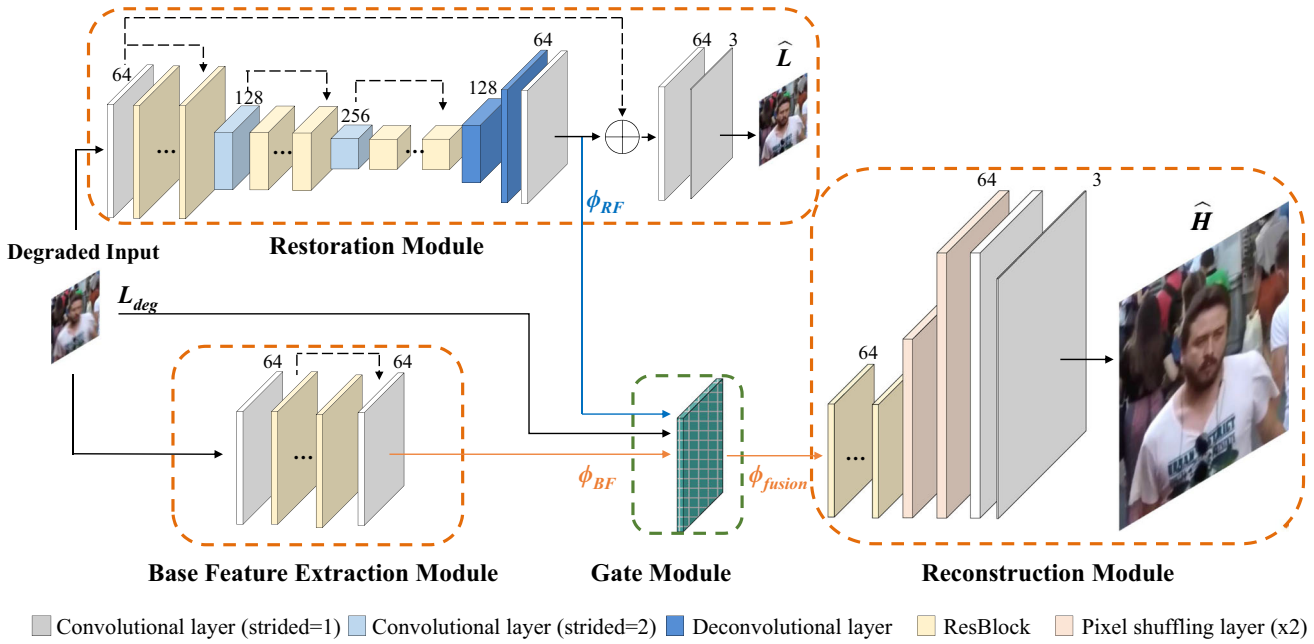


Fig. 2 Architecture of the proposed GFN model. Our model consists of four major modules: restoration module G_{res}, base feature extraction module G_{base}, gate module G_{gate}, and reconstruction module G_{recon}. The features extracted by G_{res} and G_{base} are fused by G_{gate} and then fed into G_{recon} to reconstruct the HR output image

(Mairal et al. 2009; Liu et al. 2013; Reynolds et al. 2000; Zhang and Patel 2018b; Zhang et al. 2019; Li et al. 2018b, d; Yang et al. 2017; Fan et al. 2017). Although video-based algorithms perform better by exploiting the temporal information, the single image deraining problem receives much research attention because of its flexibility and generality. Early methods rely on handcrafted low-level features and prior information, e.g., sparse coding and dictionary learning (Mairal et al. 2009), low-rank representation (Liu et al. 2013), and Gaussian mixture models (Reynolds et al. 2000). However, these schemes are prone to failures of recovering high-frequency details and removing the rain streaks completely. Recent approaches show promising improvement based on deep CNNs (Zhang and Patel 2018b; Zhang et al. 2019; Li et al. 2018b; Fan et al. 2017), recurrent neural network (RNNs) (Li et al. 2018d), and iterative networks (Yang et al. 2017). Zhang and Patel (2018b) propose a multi-streaming network for joint rain event detection and deraining. Recently, Zhang et al. (2019) introduce the conditional adversarial loss to recover high-frequency details and a refined loss to suppress the artifacts. Li et al. (2018d) utilize a deep convolutional RNN to remove the overlap rain streaks with multiple stages.

2.5 Degraded Image Super Resolution

Most super resolution methods in the literature operate on images without significant degradation caused by noise or blur. Some approaches (Yamaguchi et al. 2010; Park and Lee 2017; Bascle et al. 1996) aim to solve the joint task of super resolution and deblurring by exploiting temporal information from the videos. As these methods depend on the optical flow estimation, such schemes cannot be applied to the case of single input images. Xu et al. (2017) train a generative adversarial network to super resolve blurry face and text images. As face and text images have distinct structured properties that can be exploited, compact models can be developed to address the joint task of super resolution and deblurring for specific object categories. Zhang et al. (2018a) propose a deep encoder-decoder network (ED-DSRN) for joint image deblurring and super resolution. However, the HR images are directly reconstructed from the inputs, which tend to generate unexpected structures in severely degraded regions. In this work, we design the network architecture to better extract features in the presence of complex degradations. The proposed model has fewer parameters than those of (Zhang et al. 2018a) and can generate sharp HR images under different degradations.

3 Gated Fusion Network

In this section, we describe the architecture design, training loss functions, and implementation details of the proposed GFN for super resolution on degraded images.

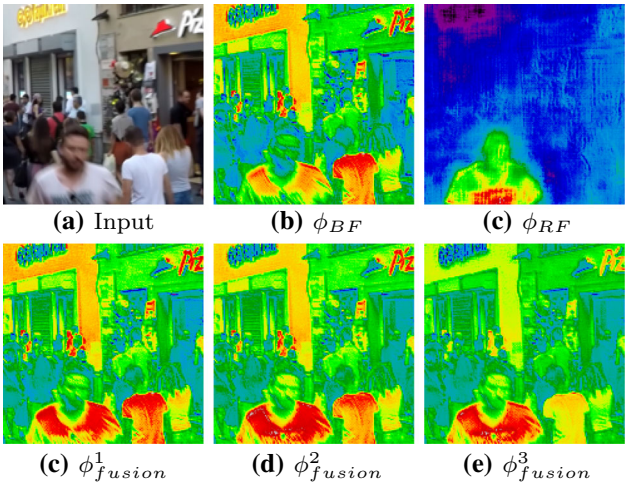


Fig. 3 Feature responses of the base features ϕ_{BF}, recovered (deblurring) features ϕ_{RF}, and fused features from different gate blocks ϕ^1_{fusion}, ϕ^2_{fusion}, and ϕ^3_{fusion}. The base features contain unclear contours around the degraded (blurry) regions, while the recovered (deblurring) features have strong responses on those regions. The fused features restore sharp structure and contours information progressively by selectively merging ϕ_{RF} into ϕ_{BF} in a recursive way. We normalize the feature maps for better visualization

3.1 Network Architecture

Given a degraded LR image L_{deg} as the input, our goal is to recover a sharp HR image $\widehat{H}$. In this work, we consider the case of 4 times super resolution, i.e., the width and height of $\widehat{H}$ are 4 times larger than those of L_{deg}. The proposed model has a dual-branch architecture and consists of four major modules: (i) a restoration module G_{res} for recovering a sharp LR image $\widehat{L}$, (ii) a base feature extraction module G_{base} to extract visual information from the blurry LR input, (iii) a gate module G_{gate} for merging the features from the restoration and base feature extraction modules, and (iv) a reconstruction module G_{recon} to reconstruct the final HR output image. An overview of the proposed model is illustrated in Fig. 2.

3.1.1 Restoration Module

The goal of this restoration module is to extract features for recovering a sharp LR image $\widehat{L}$ from the degraded LR input L_{deg}. We use an asymmetric residual encoder-decoder architecture to enlarge the receptive field. The encoder consists of three scales, where each scale has a residual group (six residual blocks as proposed by Lim et al. (2017)) and the first two residual groups are followed by a strided convolutional layer to downsample the feature maps by 1/2 times. The decoder has two deconvolutional layers to enlarge the spatial resolution of feature maps. Finally, we use two additional convolutional layers to reconstruct a sharp LR image $\widehat{L}$. We denote the output features of the decoder by ϕ_{RF}, which are fed into the gate module for feature fusion.

3.1.2 Base Feature Extraction Module

We use eight residual blocks (Lim et al. 2017) to extract base features from the degraded input L_{deg}. To retain the spatial information, we do not use any pooling or strided convolutional layers. We denote the base features by ϕ_{BF}.

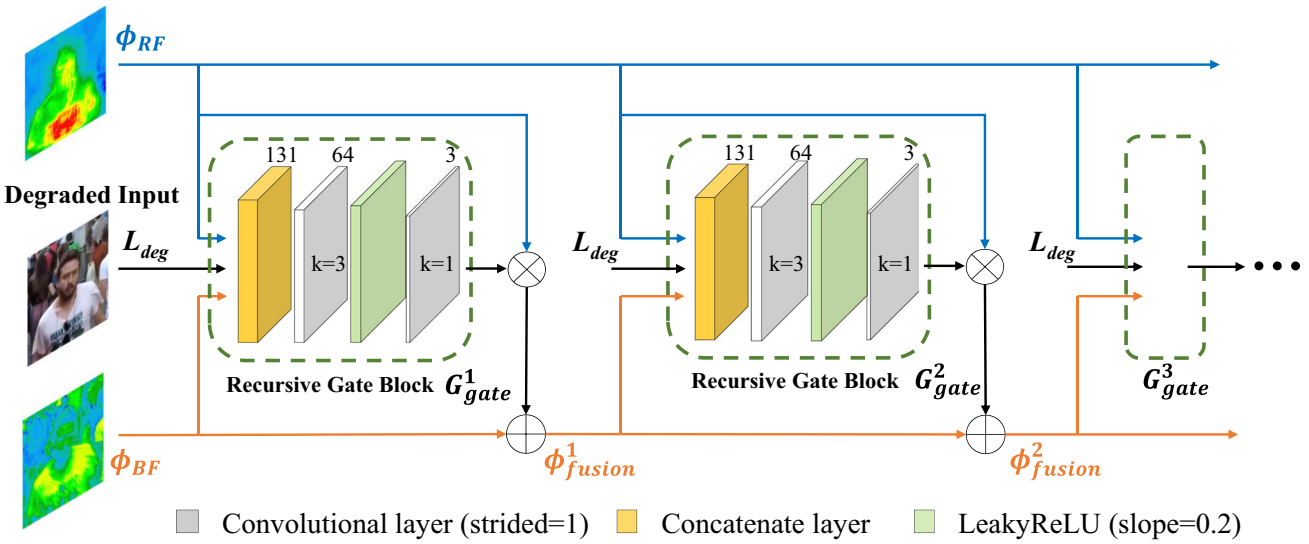


Fig. 4 Structure of the recursive gate module. We use recursive gate blocks to fully exploit the correlation between the features from two independent branches and fuse them progressively. Since each block serves as the same purpose of adaptively merging the recovered features ϕ_{RF} into the main stream ϕ^n_{fusion} (ϕ_{BF} for the first block), the parameters are shared among blocks

3.1.3 Gate Module

In Fig. 3, we show the responses of ϕ_{RF}, ϕ_{BF} and fused features ϕ^n_{fusion} from a blurry LR input. While the base features ϕ_{BF} contain both sharp and unclear contours (as shown on the wall of Fig. 3b), the recovered (deblurring) features ϕ_{RF} have high response on the regions with large motion (as shown by the pixels of the moving person in Fig. 3c). Thus, the responses of ϕ_{RF} and ϕ_{BF} complement each other, especially on the degraded (blurry) regions. To better extract features for super resolution, we adaptively merge the recovered features and base features by learning a gate module, which has been shown effective to discover feature importance for multi-modal fusion (Hochreiter and Schmidhuber 1997; Ren et al. 2018). We apply a basic gate block and adopt a recursive merging strategy to progressively fuse the features.

As shown in Fig. 4, each gate block consists of a concatenation layer, two convolutional layers with the filter size of 3×3 and 1×1, and a leaky rectified linear unit (LReLU) between the two convolutional layers. The first recursive gate block, G^1_{gate}, takes ϕ_{RF}, ϕ_{BF}, and the degraded LR input L_{deg} as input, and generates a pixel-wise weight map. The fused features can be formulated as,

$$\phi^1_{fusion} = G^{1\Delta}_{gate}(\phi_{RF}, L_{deg}, \phi_{BF}) \otimes \phi_{RF} + \phi_{BF}, \tag{1}$$

where $\otimes$ denotes the element-wise multiplication.

We propose a recursive strategy to exploit the dependence of two independent branches for feature fusion. We stack N gate blocks, where each block serves as the same purpose of adaptively merging the recovered features ϕ_{RF} into the main stream ϕ^n_{fusion} (ϕ_{BF} for the first block). The parameters are shared among the recursive gate blocks, and the output of the previous block ϕ^{k-1}_{fusion}, $k = 2, \ldots, N$ is used as the base features in the next block. Figure 3 shows the proposed recursive fusion process. Compared to the base features ϕ_{BF}, the features after the first fusion ϕ^1_{fusion} contain sharp contours of the moving person. The fused features after the second and third fusion steps contain clearer and finer information of the person, especially on the chest region, which is useful for HR image reconstruction.

The DGFN (Ren et al. 2018) method trains a network to predict confidence maps for three hand-crafted enhanced images derived from the input hazy image and then uses a gate module to combine them for generating the sharp image without haze. This method is specifically developed for the single-image dehazing task, which cannot be straightforwardly extended to other restoration tasks due to the usage of the hand-crafted enhanced images. In contrast, the proposed method is a generic framework for the joint image restoration and super-resolution problem and does not involve any heuristic process. Our gate module is designed to predict the confidence maps to adaptively fuse the features from two sub-networks, where one extracts features for restoration and another one extracts features from the input image. The fused features are then fed into an image reconstruction module to generate the sharp HR output.

To retrieve more contextual information from hand-crafted enhanced images, their gate module is constructed with 3 dilated convolutional blocks and 3 deconvolutional blocks. Since our gate module only aims to fuse two extracted features, our gate module only consists of two convolutional layers and a leaky rectified linear unit (LReLU) to maintain simplicity.

3.1.4 Reconstruction Module

In the final stage, the fused features ϕ^N_{fusion} are fed into eight residual blocks (Lim et al. 2017) and two pixel-shuffling layers (Shi et al. 2016) to enlarge the spatial resolution by 4 times. We then use two final convolutional layers to reconstruct an HR output image $\widehat{H}$. We note that most of the network operations are performed in the LR feature space. Thus, the proposed model entails low computational cost in both training and inference phases.

3.2 Loss Functions

The proposed network generates two output images: a recovered LR image $\widehat{L}$ and a sharp HR image $\widehat{H}$. In our training data, each degraded LR image L_{deg} has a corresponding ground truth HR image H and a ground truth LR image L, which is bicubic-downsampled from H. Thus, we train our network by jointly optimizing a super resolution loss and a recovering loss:

$$\min \mathcal{L}_{SR}(\widehat{H}, H) + \alpha \mathcal{L}_{recover}(\widehat{L}, L), \tag{2}$$

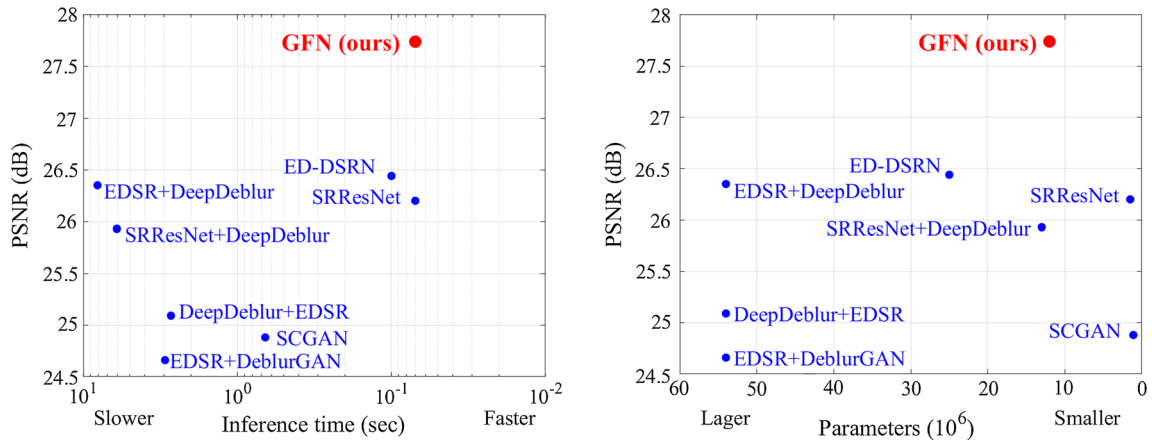
where α is a weight to balance the two loss terms.

Without the recovering loss, both the base feature extraction module and restoration module are solely guided by the super resolution loss. In this case, there is no guarantee that the dual-branch architecture can learn to extract recovered features. We have trained GFN without recovering loss and found that its result in PSNR is worse than the proposed GFN (27.69 vs. 27.91) on the joint deblurring and super resolution problem. Therefore, we impose a guidance on the restoration branch using the recovering loss to encourage the branch to extract recovered features for the restoration task. We use the pixel-wise L2 loss function for both $\mathcal{L}_{SR}$ and $\mathcal{L}_{recover}$, and empirically set α to 0.5.

Table 1 Quantitative comparisons with the state-of-the-art methods on super resolving the blurry images

Method	#Params	LR-GOPRO 4× PSNR/SSIM/time (s)	LR-Köhler 4× PSNR/SSIM/time (s)
SCGAN	1.1M	22.74/0.783/0.66	23.19/0.763/0.45
SRResNet	1.5M	24.40/0.827/**0.07**	24.81/0.781/**0.05**
EDSR	43M	24.52/0.836/2.10	24.86/0.782/1.43
RCAN	16M	24.54/0.836/1.76	24.87/0.782/1.17
SCGAN⋆	1.1M	24.88/0.836/0.66	24.82/0.795/0.45
SRResNet⋆	1.5M	26.20/0.818/**0.07**	25.36/0.803/**0.05**
ED-DSRN⋆	25M	26.44/0.873/0.10	25.17/0.799/0.08
DeepDeblur + SRResNet	13M	24.99/0.827/0.66	25.12/0.800/0.55
SRResNet + DeepDeblur	13M	25.93/0.850/6.06	25.15/0.792/4.18
DeblurGAN + SRResNet	13M	21.71/0.686/0.14	21.10/0.628/0.12
SRResNet + DeblurGAN	13M	24.44/0.807/0.91	24.92/0.778/0.54
DeblurGAN + EDSR	54M	21.53/0.682/2.18	20.74/0.625/1.57
EDSR + DeblurGAN	54M	24.66/0.827/2.95	25.00/0.784/1.92
DeepDeblur + EDSR	54M	25.09/0.834/2.70	25.16/0.801/2.04
EDSR + DeepDeblur	54M	26.35/0.869/8.10	25.24/0.795/5.81
DeepDeblur + RCAN	37M	25.10/0.833/3.91	25.16/0.801/3.52
RCAN + DeepDeblur	37M	26.34/0.870/5.39	25.24/0.794/4.67
SRN + RCAN	17M	25.62/0.867/3.10	25.18/0.798/1.66
RCAN + SRN	17M	26.00/0.874/5.76	25.20/0.799/4.49
GFN⋆ (ours)	12M	**27.91/0.902/0.07**	**25.79/ 0.818/0.05**

The evaluated methods include **super resolution methods**, SRResNet (Ledig et al. 2017), EDSR (Lim et al. 2017), RCAN (Zhang et al. 2018b), **image deblurring methods**, DeepDeblur (Nah et al. 2017), DeblurGAN (Kupyn et al. 2018), SRN (Tao et al. 2018), and **joint approaches**, SCGAN (Xu et al. 2017), ED-DSRN (Zhang et al. 2018a). The methods with a ⋆ sign are trained on our LR-GOPRO training set. Bold values indicate the best performance. The proposed GFN model performs favorably against existing methods while maintaining a small model size and fast inference speed



(a) PSNR vs. inference time **(b)** PSNR vs. number of parameters

Fig. 5 Performance versus inference time and model parameters. The results are evaluated on the LR-GOPRO dataset

3.3 Implementation Details

In the proposed network, the filter size is set as 7×7 in the first and the last convolutional layers, 4×4 in the deconvolutional layers, 1×1 in the last convolutional layers of the gate blocks, and 3×3 in all the other convolutional layers. We randomly initialize all the trainable parameters by using the method of (He et al. 2015). We use the leaky rectified linear unit (LReLU) with a negative slope of 0.2 as the activation function. As suggested in (Lim et al. 2017), we do

not use any batch normalization layers in order to retain the range flexibility of features. To facilitate the training process, we use skip connections in the restoration module and base feature extraction module (refer to the dashed lines in Fig. 2). From quantitative evaluations (see Table 4), we find that the gate module with 3 recursive gate blocks, i.e., $N = 3$, achieves the best performance on all three applications. Thus we set N to 3 as the default parameter of the proposed GFN model. We use the ADAM solver (Kingma and Ba 2015) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ to optimize the network. All the training and evaluation processes are conducted on an NVIDIA 1080Ti GPU. The source code can be found at https://github.com/BookerDeWitt/GFN-IJCV.

4 Experimental Results

In this section, we evaluate the proposed GFN model on super resolving blurry, hazy, and rainy images. We present quantitative and qualitative comparisons with state-of-the-art approaches. In addition, we carry out ablation studies to analyze several design choices of the proposed model.

4.1 Super Resolving Blurry Image

4.1.1 Training Dataset and Details

We use the GOPRO (Nah et al. 2017) dataset to generate the training data for the joint super resolution and deblurring problem. The GOPRO dataset contains 2103 blurry and sharp HR image pairs for training. To augment the training data, we resize each HR image pair with three random scales within the scales of 0.5 and 1.0. We then crop the HR images into several patches with a size of 256×256 and a stride of 128. We downsample the blurry HR patch H_{blur} and sharp HR patch H by 4 times using bicubic downsampling to generate the blurry LR patches L_{blur} and sharp LR patches L. We obtain 107,584 triplets of $\{L_{blur}, L, H\}$ for training (the blurry HR patches H_{blur} are discarded during training). The generated dataset is referred to as LR-GOPRO in the following.

To facilitate the training process, we use a three-stage training strategy. First, we pre-train the network without the gate module by simply fusing ϕ_{RF} and ϕ_{BF} via addition. Therefore, only the restoration module, base feature extraction module, and reconstruction modules are updated in this stage. We use an initial learning rate of 10^{-4} with a decay rate of 0.5 every 6 epochs. The network is trained from scratch for 25 epochs. We note that the rapidly decaying pre-training without the gate module is important in the early stage as it helps avoid the exploding gradient issues. In the second stage, we continue training the models without the gate module for 60 epoch. The learning rate is reset to 10^{-4} and multiplied by 0.1 for every 30 epochs. Finally, we include the gate module and train the entire network for 60 epochs. The learning rate is set to 5×10^{-5} and multiplied by 0.1 for every 25 epochs. We use a batch size of 16.

4.1.2 Performance Evaluation

We evaluate the proposed GFN model with the state-of-the-art super resolution methods (Ledig et al. 2017; Lim et al. 2017; Zhang et al. 2018b), joint image deblurring and super resolution approaches (Xu et al. 2017; Zhang et al. 2018a), and straightforward combinations of super resolution and non-uniform deblurring schemes (Nah et al. 2017; Kupyn et al. 2018; Tao et al. 2018). For fair comparisons, we re-train the SCGAN (Xu et al. 2017), SRResNet (Ledig et al. 2017), and ED-DSRN (Zhang et al. 2018a) models on the same training dataset discussed above. Other super resolution methods are trained on the DIV2K dataset (Agustsson and Timofte 2017) and deblurring methods are trained on the GOPRO dataset (Nah et al. 2017).

We use bicubic downsampling to generate blurry LR images from the test set of the GOPRO (Nah et al. 2017) and Köhler (Köhler et al. 2012) datasets for evaluation. Table 1 shows the quantitative evaluation in terms of PSNR, SSIM, and average inference time. The tradeoff between image quality and efficiency is better visualized in Fig. 5. The proposed GFN model performs favorably against the existing methods on both datasets and maintains a low computational cost and execution time. While the re-trained SCGAN and SRResNet perform better than their pre-trained models, both methods do not handle the complex non-uniform blur well due to their small model capacity. It is noted that the SCGAN takes bicubic upsampled images as the inputs, and most operations are performed in the HR feature space. In contrast, the ED-DSRN and our GFN take LR images as the inputs, and most operations are performed in the LR feature space. Therefore, the SCGAN runs slower than others even with fewer parameters. The ED-DSRN method performs well using a large model with more parameters. However, the single-branch architecture of ED-DSRN is less effective than the proposed dual-branch network.

The straightforward approaches by combining super resolution and deblurring methods are generally less effective due to the error accumulation. We note that the approaches first using super resolution (i.e., performing super resolution followed by image deblurring) typically perform better than the alternatives (i.e., performing image deblurring followed by super resolution). However, the strategy by first performing super resolution entails heavy computational cost as the time-consuming image deblurring step is performed in the HR image space. Compared with the best-performing combination of EDSR and DeepDeblur methods, the proposed GFN model executes

Fig. 6 Visual comparisons on the LR-GOPRO dataset. The evaluated methods include SRResNet (Ledig et al. 2017), EDSR (Lim et al. 2017), SCGAN (Xu et al. 2017), ED-DSRN (Zhang et al. 2018a), and DeepDeblur (Nah et al. 2017). The methods with a ⋆ sign are trained on our LR-GOPRO training set. The proposed method generates sharper HR images with more details

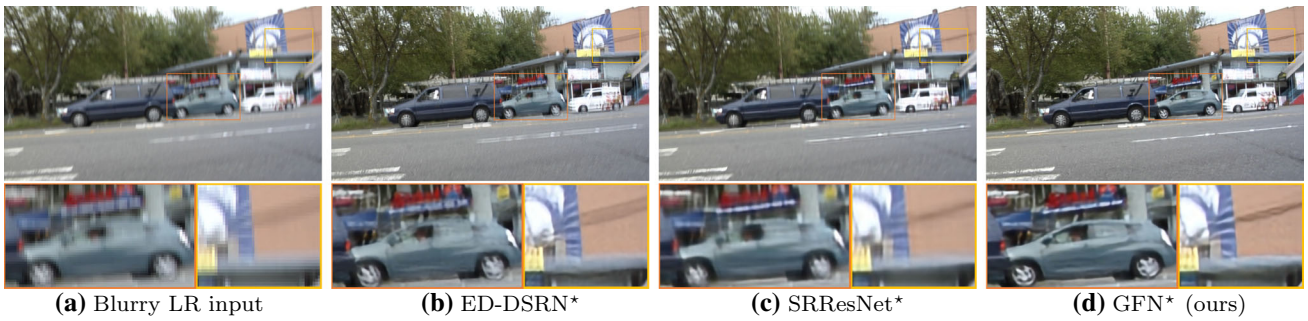


Fig. 7 Visual comparisons on the real blurry image dataset (Su et al. 2017). The methods with a ⋆ sign are trained on our LR-GOPRO training set. The proposed GFN model is more robust to outliers in real images and generates sharper results than the re-trained state-of-the-art methods ED-DSRN (Zhang et al. 2018a) and SRResNet (Ledig et al. 2017)

116 times faster and uses 78% fewer model parameters.

We present the qualitative results of the LR-GOPRO dataset in Fig. 6 and a real blurry image in Fig. 7. The methods using the combination scheme, e.g., DeepDeblur + EDSR and EDSR + DeepDeblur, often introduce undesired artifacts due to the error accumulation problem. Existing joint super resolution and deblurring methods (ED-DSRN and SCGAN) do no handle non-uniform blur well. In contrast, the proposed algorithm generates sharp HR images with more details.

4.2 Super Resolving Hazy Image

4.2.1 Training Dataset and Details

We use the RESIDE (Li et al. 2018a) dataset to generate the training data for the joint super resolution and dehazing problem. For training, we randomly select 5005 outdoor hazy and sharp HR image pairs in 35 different haze concentrations and 5000 indoor HR pairs in 10 different haze concentrations from RESIDE training sets. We apply the same procedure as the LR-GOPRO dataset to generate the training triplets

Table 2 Quantitative comparisons with the state-of-the-art methods on super resolving the hazy images

Method	#Params	LR-RESIDE 4× PSNR/SSIM/time (s)
SRResNet	1.5M	13.29/0.566/**0.02**
EDSR	43M	13.71/0.650/0.12
RCAN	16M	13.72/0.652/0.30
SRResNet⋆	1.6M	23.58/0.791/**0.02**
ED-DSRN⋆	25M	24.89/0.813/0.04
DGFN + SRResNet	2.2M	19.62/0.618/0.03
SRResNet + DGFN	2.2M	19.34/0.607/0.08
AODN + SRResNet	1.5M	16.49/0.597/**0.02**
SRResNet + AODN	1.5M	17.07/0.563/0.05
EDSR + DCP	16M	17.46/0.572/24.0
EDSR + NLD	16M	17.70/0.576/10.6
AODN + EDSR	43M	17.05/0.702/0.12
EDSR + AODN	43M	18.30/0.713/0.13
DGFN + EDSR	44M	21.03/0.740/0.13
EDSR + DGFN	44M	21.89/0.775/0.18
DGFN + RCAN	16M	21.04/0.740/0.30
RCAN + DGFN	16M	21.92/0.777/0.36
PFFNet + RCAN	38M	20.55/0.678/0.31
RCAN + PFFNet	38M	23.76/0.795/0.31
RCAN + GCANet	17M	22.93/0.786/1.2
GFN⋆ (ours)	12M	**25.77/0.830/0.02**

The evaluated methods include **super resolution methods**, SRResNet (Ledig et al. 2017), EDSR (Lim et al. 2017), RCAN (Zhang et al. 2018b), **image dehazing methods**, DCP (He et al. 2011), NLD (Berman et al. 2016), DGFN (Ren et al. 2018), GCANet (Chen et al. 2019), PFFNet (Mei et al. 2018), AODN (Li et al. 2017), and **joint approaches**, SCGAN (Xu et al. 2017), ED-DSRN (Zhang et al. 2018a). The methods with a ⋆ sign are trained on our LR-RESIDE training set. Bold values indicate the best performance. The proposed GFN performs favorably against existing methods while maintaining a small model size and fast inference speed

of $\{L_{haze}, L, H\}$. We refer to the generated dataset as LR-RESIDE in the following.

Since the training process of joint dehazing and super resolution is more stable compared with that of joint deblurring and super resolution, we simplify the training process into two stages. First, we train the network without the gate module from scratch for 25 epochs. The learning rate is set to 10^{-4} and multiplied by 0.5 for every 7 epochs. Second, we enable the gate module and train the complete model for 60 epochs where the learning rate is set to 10^{-4} and multiplied by 0.1 for every 25 epochs. The other settings are the same as those for blurry image super resolution.

4.2.2 Performance Evaluation

We choose 500 indoor image pairs and 500 outdoor image pairs from the test set of the RESIDE dataset for evaluation. We compare the proposed GFN model with the state-of-the-art super resolution methods (Ledig et al. 2017; Lim et al. 2017; Zhang et al. 2018b), joint image deblurring and super resolution approaches (Zhang et al. 2018a; Xu et al. 2017), and combinations of super resolution algorithms and dehazing schemes (He et al. 2011; Berman et al. 2016; Ren et al. 2018; Mei et al. 2018; Li et al. 2017; Chen et al. 2019). For fair comparisons, we re-train the models of SRResNet (Ledig et al. 2017), ED-DSRN (Zhang et al. 2018a), and PFFNet (Mei et al. 2018) on our training set.[1] Other super resolution methods are trained on the DIV2K dataset (Agustsson and Timofte 2017) and deep learning-based dehazing methods are trained on the RESIDE dataset (Li et al. 2018a).

The quantitative evaluations in Table 2 show that the proposed GFN model performs well in terms of PSNR and SSIM with shorter inference time. We present qualitative results on the LR-RESIDE dataset in Fig. 8. The state-of-the-art super resolution method (RCAN) does not remove the haze from the hazy input, and the straightforward combination schemes, e.g., PFFNet + RCAN and RCAN + PFFNet, generate undesired artifacts and distorted colors on the flat regions due to

[1] Since the pre-trained model of the PFFNet is not available, we train the network directly on the RESIDE dataset and achieve quantitative results on the RESIDE dataset better than the reported results. We use this model in the following experiments.

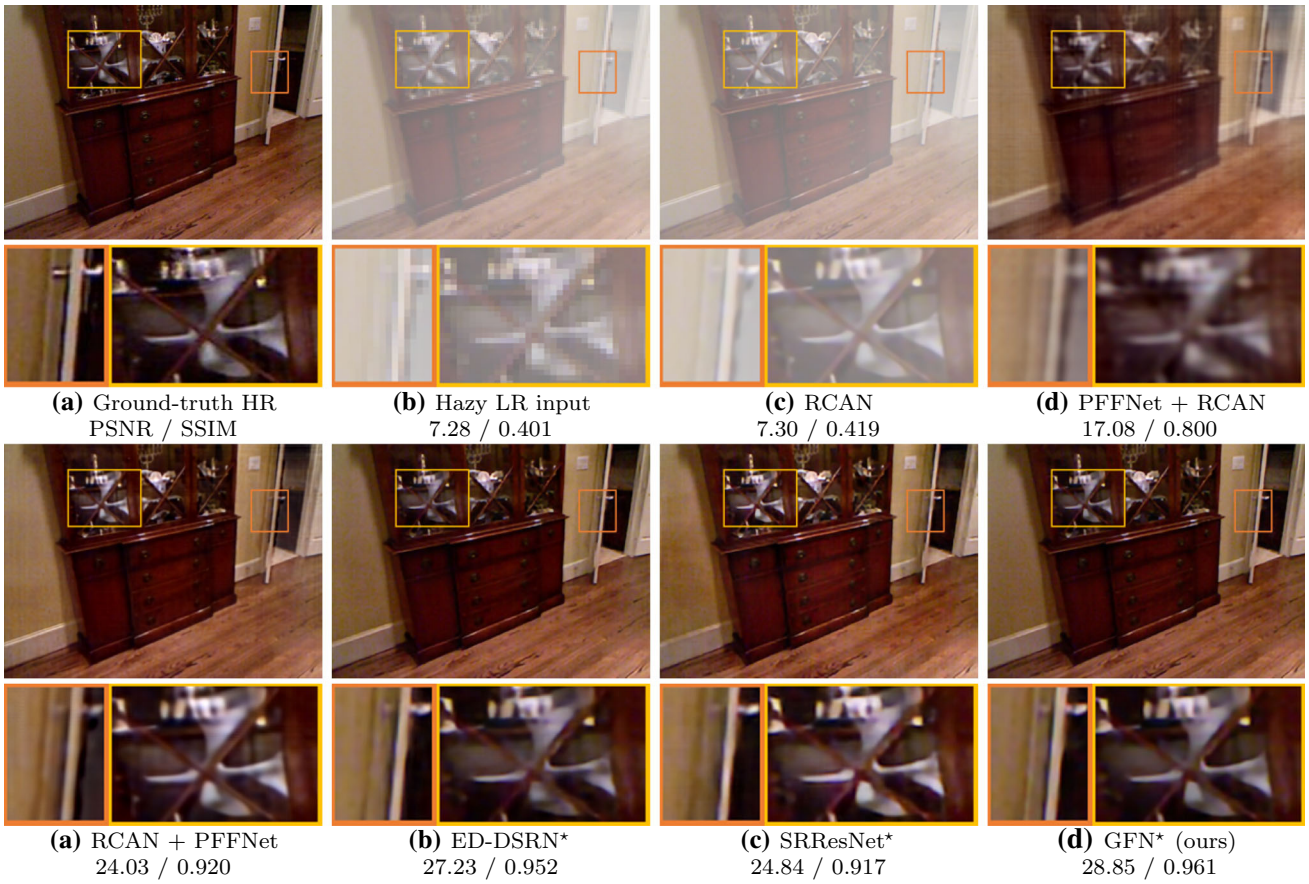


Fig. 8 Visual comparisons on the LR-RESIDE dataset. The evaluated methods include SRResNet (Ledig et al. 2017), RCAN (Zhang et al. 2018b), ED-DSRN (Zhang et al. 2018a), and PFFNet (Mei et al. 2018). The methods with a ⋆ sign are trained on our LR-RESIDE training set. The proposed model generates sharper HR images with more details

the error accumulation problem. The re-trained SRResNet and ED-DSRN methods do not recover the details well. In contrast, the proposed model generates better results with more details.

4.3 Super Resolving Rainy Image

4.3.1 Training Dataset and Details

Since there is no off-the-shelf dataset for rainy image super resolution, we use the Rain1200 (Zhang and Patel 2018b) dataset to generate rainy LR images. We note that directly applying bicubic downsampling on the rainy HR images tends to remove rain streaks as this operator, similar to low-pass filtering, reduces high-frequency details such as thin structures. As shown in Fig. 9, the LR image directly downsampled from rainy HR image does not contain many rain streaks and is similar to the LR image downsampled from the clean HR image (see Fig. 9b, c). In order to obtain more realistic LR inputs, we generate rainy LR images by synthesizing rainy streaks on downsampled sharp images. We first apply bicubic downsampling on the sharp HR image H to generate the sharp LR image L. Similar to (Zhang and Patel 2018b), we use Photoshop to synthesize rain streaks on L to generate the rainy LR image L_{rain}. After data augmentation, we obtain 24,000 triplets of $\{L_{rain}, L, H\}$ for training and 1200 triplets for testing. We refer to the generated dataset as LR-Rain1200 in the following.

To remove long streaks in the rainy images, we modify the network structure of the restoration module to enlarge the receptive field. Specifically, we use the structure in (Mei et al. 2018) as the encoder-decoder architecture of the restoration module. Since the resolution of a rainy LR input in this dataset is relatively low (128×128), we remove the last strided convolutional layer and set the output channels of the rest three scales to 64, 128, and 256 respectively. Moreover, we apply the residual learning scheme in the reconstruction module G_{recon} to accelerate the training process. We use a deconvolutional layer with the filter size of 4×4 to upsample the rainy LR input before merging with the output of the reconstruction module G_{recon}. The training processes are mostly the same as the one for the blurry image super resolution

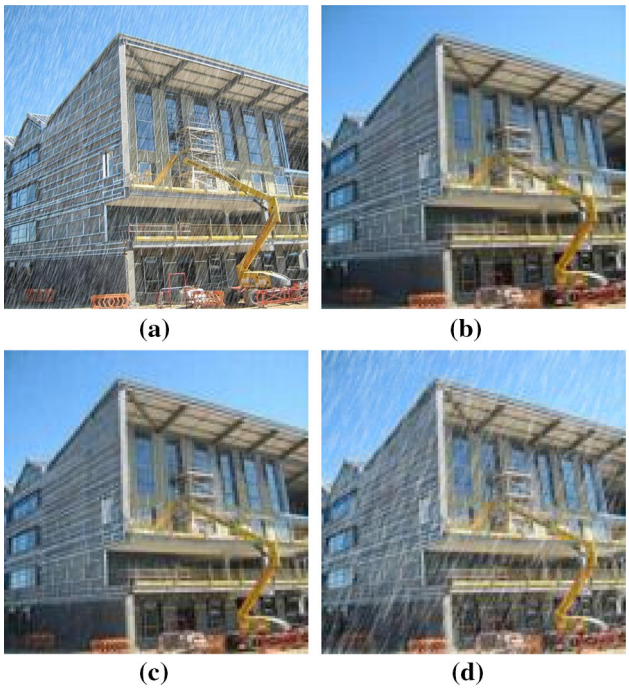


Fig. 9 Examples of the generated low-resolution image for super resolving rainy images. To generate training and testing dataset for super resolving rainy images, we use rainy/sharp image pairs from Rain1200 dataset (Zhang and Patel 2018b) as the HR images. Simply applying bicubic downsampling on the rainy HR image (**a**) results in the LR image (**c**), where many rain streaks are removed. Thus, we first obtain a sharp LR image (**b**) by applying bicubic downsampling on the sharp HR image and generate the rainy LR image (**d**) by synthesizing rain streaks using the approach in (Zhang and Patel 2018b)

task except that we use a batch size of 6 due to limited GPU memory.

4.3.2 Performance Evaluation

Table 3 shows the quantitative results in terms of PSNR, SSIM, and average inference time. Since there exists no approach for joint image deraining and super resolution, we evaluate our method against the state-of-the-art super resolution algorithms (Ledig et al. 2017; Lim et al. 2017; Zhang et al. 2018b), joint image deblurring and super resolution approaches (Zhang et al. 2018a; Xu et al. 2017), and straightforward combinations of super resolution and deraining schemes (Zhang and Patel 2018b; Zhang et al. 2019; Li et al. 2018d). For fair comparisons, we re-train the SRResNet (Ledig et al. 2017) and ED-DSRN (Zhang et al. 2018a) models on our training set. The other super resolution methods are trained on the DIV2K dataset (Agustsson and Timofte 2017) and deep learning-based deraining methods are trained on the Rain800 dataset (Zhang et al. 2019) (RESN and IDGAN) and Rain1200 dataset (Zhang and Patel 2018b) (DID-MDN). As shown in Table 3, the proposed model with a large receptive field, denoted by GFN-Large, achieves better performance with shorter inference time and fewer model parameters than the evaluated methods. Some deblurred results are shown in Fig. 10. Although the RCAN method recovers some high-frequency details, it does not remove the rain streaks on the image. The re-trained SRResNet model and straightforward combination approaches, DID-MDN + RCAN and RCAN + DID-MDN, do not remove long rain streaks and often introduce unexpected artifacts on the rich texture regions due to the error accumulation problem. The re-trained ED-DSRN model removes most rain streaks but does not restore clear contours and high-frequency details. The proposed GFN-Large model accurately removes the rain streaks while preserving the structural information and recovering more details.

4.4 Ablation Study and Analysis

The proposed GFN consists of four modules: a restoration module to extract recovered features; a dual-branch architecture instead of a concatenation of base and restoration modules; a fusion approach on the feature level; and a gate module to adaptively fuse base and recovered features. To further analyze the components, we train the combination of the base feature extraction module and reconstruction module ($G_{base} + G_{recon}$) as the baseline and introduce other modules progressively to evaluate them. All the models in this section are trained from scratch using the same settings for fair comparisons. Without loss of generality, we conduct these experiments on two applications, blurry image super resolution and hazy image super resolution. The evaluated network architectures are illustrated in Figure 11, and the results are shown in Table 4.

4.4.1 Effect of Restoration Module

We use the restoration module and baseline model in two ways: restoration first (Model-1) and SR first (Model-2). The restoration module shows significant performance improvement over the baseline on both applications (0.62 dB and 1.24 dB for blurry and hazy images, respectively). The SR-first combination scheme achieves better performance but has slower execution speed.

4.4.2 Effect of Dual-Branch Architecture

We use a dual-branch structure to extract the base and recovered features separately (Model-3 listed in Table 4). The outputs of the two modules are fused by direct addition, and the recovering loss is used to guide this process. Compared with the sequential restoration and super resolution method (Model-1), it achieves 0.19 dB and 0.28 dB improvement on blurry and hazy images, respectively. Furthermore,

Table 3 Quantitative comparisons with the state-of-the-art methods on super resolving the rainy images.

Method	#Params	LR-Rain1200 4× PSNR/SSIM/time (s)
SRResNet	1.5M	16.27/0.341/0.02
EDSR	43M	19.88/0.548/0.12
RCAN	16M	19.94/0.548/0.30
SRResNet⋆	1.6M	23.29/0.624/**0.02**
ED-DSRN⋆	25M	23.86/0.694/0.04
RESN + SRResNet	1.7M	19.01/0.497/0.05
SRResNet + RESN	1.7M	17.75/0.386/0.72
IDGAN + SRResNet	1.8M	18.28/0.451/0.30
SRResNet + IDGAN	1.8M	17.25/0.407/0.37
RESN + EDSR	43M	20.71/0.617/0.15
EDSR + RESN	43M	22.37/0.644/0.82
IDGAN + EDSR	43M	19.57/0.581/0.40
EDSR + IDGAN	43M	19.96/0.606/0.47
RESN + RCAN	16M	20.74/0.618/0.33
RCAN + RESN	16M	22.53/0.650/1.00
DID-MDN + RCAN	82M	22.31/0.610/0.31
RCAN + DID-MDN	82M	23.50/0.685/0.35
GFN⋆ (ours)	12M	24.64/0.705/0.02
GFN-large⋆ (ours)	24M	**25.24/0.709/0.02**

The evaluated methods include **super resolution methods**, SRResNet (Ledig et al. 2017), EDSR (Lim et al. 2017), RCAN (Zhang et al. 2018b), **image deraining methods**, RESN (Li et al. 2018d), IDGAN (Zhang et al. 2019), DID-MDN (Zhang and Patel 2018b), and **joint approaches**, SCGAN (Xu et al. 2017), ED-DSRN (Zhang et al. 2018a). The methods with a ⋆ sign are trained on our LR-Rain1200 training set. Bold values indicate the best performance. The GFN-Large scheme performs favorably against existing methods while maintaining a small model size and fast inference speed

the Model-3 performs comparably with the SR-first method (Model-2) but more efficiently. This is due to the heavy computation load of the restoration process carried out in the HR feature space for Model-2.

4.4.3 Effect of Feature Level Fusion

Since the recovering loss in the Model-3 is computed after fusion, it does not provide explicit guidance on each branch. In the Model-4, we impose the recovering loss on the restoration branch as explicit regularization. Furthermore, to reduce the computational redundancy and avoid error accumulation, we fuse the branches on the feature level, instead of fusing them on the pixel level. Compared with the Model-3, the Model-4 achieves 0.52 dB and 0.11 dB performance improvement and lower computational cost on two tasks.

4.4.4 Effect of Gate Module

We introduce the gate module to enable local and channel-wise feature fusion from two branches. The gate module also helps exploit the dependence between the features. Here, we evaluate the gate module in terms of the number of recursive blocks N. As shown in Table 4, the gate module with 3 recursive blocks performs best, with improvements of 0.38 dB and 0.58 dB over the Model-4 on two tasks. We note that the gate module with more than 3 recursive blocks does not perform well.

4.4.5 Generalizability of GFN

To show that the proposed architecture is a generic framework, we replace the original restoration and reconstruction modules with more advanced network architectures and show that it can obtain further performance gains. We use the Residual Dense Block (RDB) in the RDN (Zhang et al. 2018c) to replace the ResBlock in the reconstruction module and use the dilation architecture in the GCANet (Chen et al. 2019) to replace the classical encoder-decoder architecture in the restoration module. As shown in Table 5, using the more advanced structure improves the performance under the same training settings. More ablation study are included in the appendix.

Fig. 10 Visual comparisons on the LR-Rain1200 dataset. The evaluated methods include SRResNet (Ledig et al. 2017), RCAN (Zhang et al. 2018b), ED-DSRN (Zhang et al. 2018a), and DID-MDN (Zhang and Patel 2018b). The methods with a $\star$ sign are trained on our LR-Rain1200 training set. The proposed model is able to remove rain streaks and generates sharper HR images with more details

Table 4 Analysis of key components in the proposed GFN

Modifications	Baseline	Model-1	Model-2	Model-3	Model-4	$GFN_{N=1}$	$GFN_{N=2}$	$GFN_{N=3}$	$GFN_{N=4}$
Restoration module		✓	✓	✓	✓	✓	✓	✓	✓
Dual-branch				✓	✓	✓	✓	✓	✓
Feature level					✓	✓	✓	✓	✓
Gate module						✓	✓	✓	✓
SR-first			✓						
DB + SR PSNR(dB)	26.20	26.82	27.00	27.01	27.53	27.74	27.87	27.91	27.86
Time (s)	0.07	0.09	0.57	0.10	0.07	0.07	0.07	0.07	0.07
DH + SR PSNR(dB)	23.58	24.82	25.12	25.10	25.21	25.44	25.69	25.77	25.72
Time (s)	0.02	0.03	0.16	0.03	0.02	0.02	0.02	0.02	0.02

All models are trained on the LR-GOPRO dataset or the LR-RESIDE dataset with the same hyper-parameters. The baseline method is the network with the base feature extraction module and the reconstruction module. DB is the abbreviation for "deblurring", and DH is the abbreviation for "dehazing"

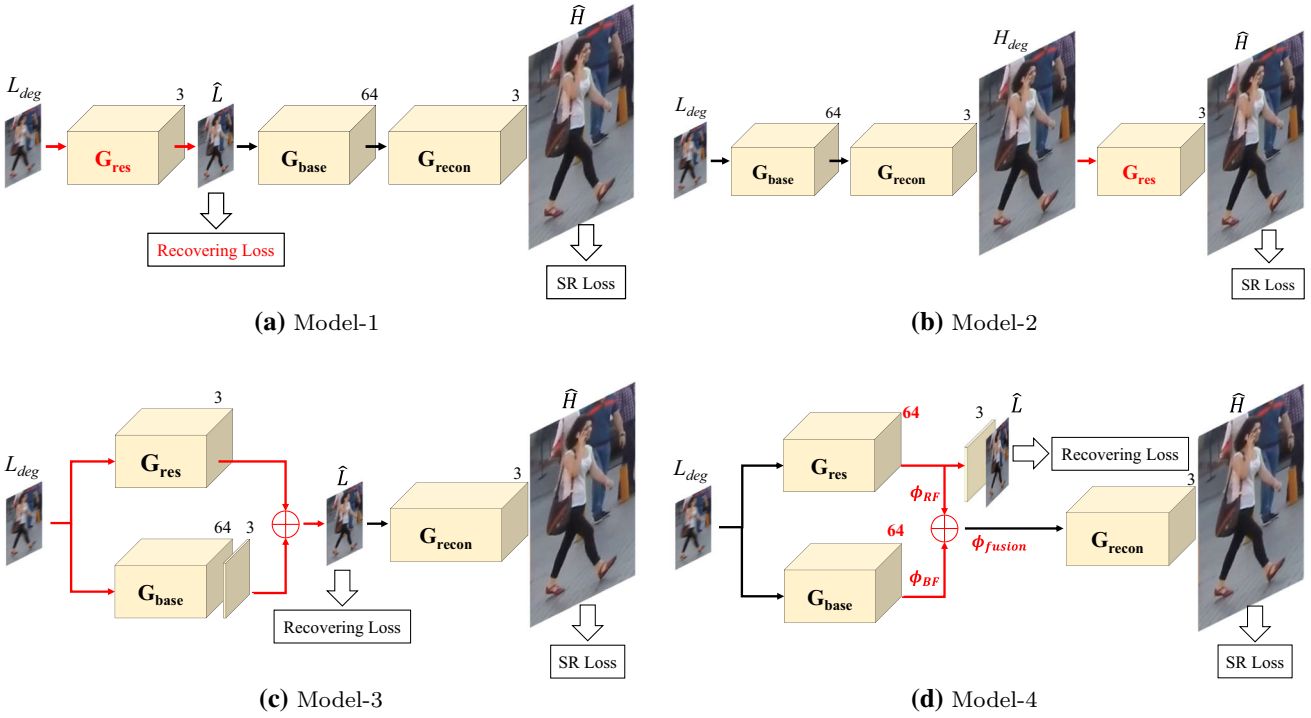


Fig. 11 Network structure of the models in the ablation study. G_{res}, G_{base}, G_{recon} represent the restoration module, base feature extraction module, and reconstruction module in GFN

Table 5 Qualitative results using different restoration and reconstruction modules on the LR-RESIDE dataset

	GFN	GFN_RDN	GFN_GCANet	GFN_RDN_GCANet
PSNR/SSIM	25.77/0.830	25.88/0.833	25.83/0.831	25.90/0.833

We evaluate the following methods: replacing the ResBlock in the proposed GFN with the Residual Dense Block (RDB) in the RDN (Zhang et al. 2018c) (GFN_RDN), replacing the restoration module in the proposed GFN with the dilation architecture in the GCANet (Chen et al. 2019) (GFN_GCANet), and replacing both of them (GFN_RDN_GCANet). All the models are trained using the same setting

4.5 Limitations

To remove non-local degradation, such as haze or long rain streaks, we use an encoder-decoder architecture to extract global and contextual information. However, this approach does not effectively extract local features commonly used for super resolution (Lim et al. 2017; Ledig et al. 2017). As a result, the proposed method tends to generate over-smoothed results compared to the other schemes, as shown in some regions of Figs. 8 and 10. For future work, we will explore more effective architectures to better exploit both global and local visual information for super resolution on degraded images.

5 Conclusions

In this paper, we propose an end-to-end architecture to recover a sharp HR image from a degraded LR input. The proposed network consists of two branches to extract recovered and base features separately. The extracted features are fused through a recursive gate module and used to reconstruct the final results. The network design decouples the joint problem into two restoration tasks and enables efficient training and inference. Extensive evaluations of different restoration tasks demonstrate that the proposed model is effective for super resolving degraded images.

Acknowledgements X. Zhang, H. Dong, and F. Wang are supported in part by National Major Science and Technology Projects of China Grant under No. 2019ZX01008103, National Natural Science Foundation of China (61603291), Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2018JM6057), and the Fundamental Research Funds for the Central Universities. W.-S. Lai and M.-H. Yang are supported in part by NSF CAREER Grant #1149783 and Gifts from Verisk, Adobe and Google.

Appendix

Network Configuration

We present the detailed configuration of the proposed network in Table 6, with respect to the four modules in the network: the **deblurring module**, **SR feature extraction module**, **recursive gate module**, and **reconstruction module**.

List of the Evaluated Methods

All the the evaluated methods in Sect. 4 are listed in Table 7.

Table 6 Configuration of the proposed network. The values in the **skip** row are **layer** names, indicating whose outputs are added to the outputs of the corresponding layers

Layer	Output size	Kernel	LReLU	Skip
Restoration module				
Input_1	$3 \times h \times w$			
conv1	$64 \times h \times w$	7		
Resblock 1–6	$64 \times h \times w$	3		conv1
conv2	$128 \times \frac{h}{2} \times \frac{w}{2}$	3		
Resblock 7–12	$128 \times \frac{h}{2} \times \frac{w}{2}$	3		conv2
conv3	$256 \times \frac{h}{4} \times \frac{w}{4}$	3		
Resblock 13–18	$256 \times \frac{h}{4} \times \frac{w}{4}$	3		conv3
deconv1	$128 \times \frac{h}{2} \times \frac{w}{2}$	4	✓	
deconv2	$64 \times h \times w$	4	✓	
conv4	$64 \times h \times w$	7		conv1
conv5	$64 \times h \times w$	3	✓	
conv6	$3 \times h \times w$	3		
Base feature extraction module				
Input_1	$3 \times h \times w$			
conv7	$64 \times h \times w$	7		
Resblock 19–26	$64 \times h \times w$	3		
conv8	$64 \times h \times w$	3		conv7
Gate module				
Input_2.0	$131 \times h \times w$			
conv9	$64 \times h \times w$	3	✓	
conv10	$64 \times h \times w$	1		
Elementwise mul	$64 \times h \times w$			conv8
Input_2.1	$131 \times h \times w$			
conv9	$64 \times h \times w$	3	✓	
conv10	$64 \times h \times w$	1		
Elementwise mul	$64 \times h \times w$			input_2.1
Input_2.2	$131 \times h \times w$			
conv9	$64 \times h \times w$	3	✓	
conv10	$64 \times h \times w$	1		
Elementwise mul	$64 \times h \times w$			input_2.2
Reconstruction module				
Input_3	$64 \times h \times w$			
Resblock 27-34	$64 \times h \times w$	3		
conv11	$256 \times h \times w$	3		
pixel shuffle	$64 \times 2h \times 2w$		✓	
conv12	$256 \times 2h \times 2w$	3		
pixel shuffle	$64 \times 4h \times 4w$		✓	
conv13	$64 \times 4h \times 4w$	3	✓	
conv14	$3 \times 4h \times 4w$	3		

Table 7 List of the evaluated methods in Sect. 4

Method	Reference
SRResNet	MSE-based SRResNet in "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network" by Ledig et al. (2017)
EDSR	EDSR in "Enhanced deep residual networks for single image super-resolution" by Lim et al. (2017)
RDN	RDN (D=20, C=6, and G=32) in "Residual dense network for image super-resolution" by Zhang et al. (2018c)
RCAN	"Image super-resolution using very deep residual channel attention networks" by Zhang et al. (2018b)
SCGAN	MSE-based SCGAN in "Learning to super-resolve blurry face and text images" by Xu et al. (2017)
ED-DSRN	"A deep encoder-decoder networks for joint deblurring and super-resolution" by Zhang et al. (2018a)
DeepDeblur	"Deep multi-scale convolutional neural network for dynamic scene deblurring" by Nah et al. (2017)
DeblurGAN	"DeblurGAN: Blind motion deblurring using conditional adversarial networks" by Kupyn et al. (2018)
SRN	"Scale-recurrent network for deep image deblurring" by Tao et al. (2018)
DCP	"Single image haze removal using dark channel prior" by He et al. (2011)
NLD	"Non-local image dehazing" by Berman et al. (2016)
AODN	"Aod-net: All-in-one dehazing network" by Li et al. (2017)
DGFN	MSE-based DGFN in "Gated fusion network for single image dehazing" by Ren et al. (2018)
GCANet	"Gated Context Aggregation Network for Image Dehazing and Deraining" by Chen et al. (2019)
PFFNet	"Progressive feature fusion network for realistic image dehazing" by Mei et al. (2018)
RESN	"Recurrent squeeze-and-excitation context aggregation net for single image deraining" by Li et al. (2018d)
IDGAN	"Image de-raining using a conditional generative adversarial network" by Zhang et al. (2019)
DID-MDN	"Density-aware single image de-raining using a multi-stream dense network" by Zhang and Patel (2018b)

Additional Visual Results

In this section, we present more qualitative comparisons on the LR-RESIDE in Fig. 12, which includes the combinations of the SR algorithm (Lim et al. 2017) and dehazing algorithms (He et al. 2011; Berman et al. 2016; Ren et al. 2018).

Additional Ablation Study and Analysis

To further demonstrate the importance of the dual-branch architecture and gate module, more ablation study and visual results are presented in this section. We first compare the restoration module with the state-of-the-art restoration methods to evaluate the performance contribution brought by the image restoration module. Then, the qualitative results of the ablation study are presented to demonstrate how other modules help to improve the performance.

Performance of Restoration Module

We provide the quantitative results from the state-of-the-art restoration methods and the proposed restoration module in Table 8. The restoration methods include deblurring algorithms [DeepDeblur Nah et al. (2017), DeblurGAN Kupyn et al. (2018), and SRN Tao et al. (2018)], dehazing algorithms [DGFN Ren et al. (2018), GCANet Chen et al. (2019), and PFFNet Mei et al. (2018)), and deraining algorithms (IDGAN Zhang et al. (2019), RESN Li et al. (2018d), and DID-MDN Zhang and Patel (2018b)]. Since these restoration methods are trained on the high-resolution images (GOPRO, RESIDE, Rain1200 datasets), we re-train the restoration module on the same high-resolution datasets for fair comparisons. As shown in Table 8, in none of these three datasets does our restoration module acquire the best results, while the proposed GFN still performs favorably on all the three datasets as shown in Tables 1, 2, 3 of the manuscript. Therefore, the favorable performance of the proposed method comes from the architecture designs, such as the dual-branch architecture and the gate module.

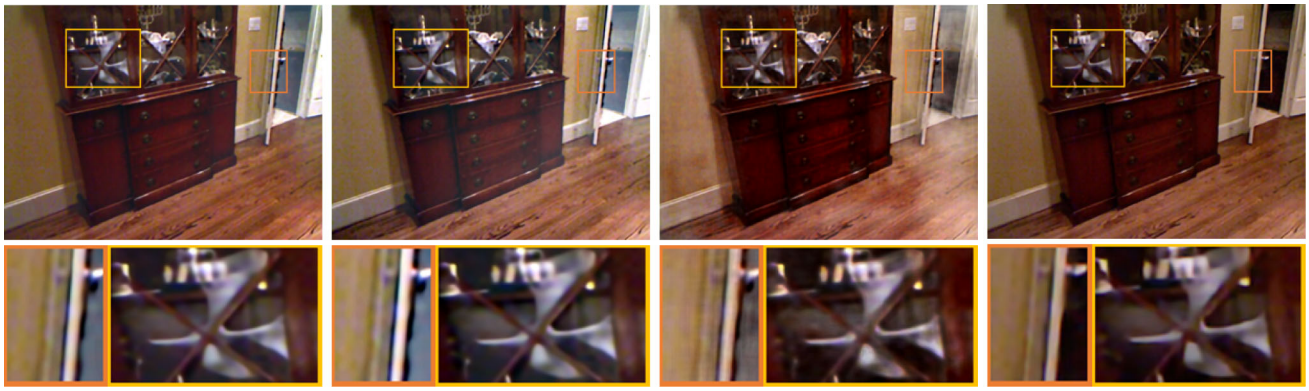
Effect of Dual-Branch Architecture and Gate Module

To further demonstrate the benefits of the dual-branch architecture and gate module, we present an example in Fig. 13. Figure 13b, c show the outputs of the restoration module G_{res} and Model-1 (G_{res} + G_{base} + G_{recon}) in Fig. 11a. Since the artifacts in the G_{res} are propagated to the G_{base} and G_{recon}, the Model-1 generates less satisfactory results as shown in Fig. 13c. Figure 13d shows the output of the Model-4 in Fig. 11d, which adopts the dual-branch architecture without the gate module G_{gate}. Figure 13d contains fewer artifacts than Fig. 13c, especially on the regions that are relatively sharper in the input image. This is because the dual-branch architecture combines features from both input images and recovered images and, therefore, avoids error propagation from only the recovered images. Fig. 13e shows the output of the proposed GFN introducing the gate module to adaptively fuse the features. By exploiting the confidence of the features from two branches (ϕ_{RF} into ϕ_{BF}), the gate module manages to suppress the artifacts and blurry features via local and channel-wise feature fusion. Figure 13f–j shows that our model progressively fuses features and suppresses artifacts through the gate module.

Applications on Detection Task

To demonstrate that the proposed method can help the following high-level tasks, we compare the proposed GFN with state-of-the-art methods on the object detection task. We first generate two datasets from the KITTI dataset (Geiger et al. 2012), one blurry low-resolution dataset and a hazy low-resolution dataset. For the blurry dataset, we apply the single image non-uniform blurry synthesis method in Lai et al. (2016) to generate the blurry HR images and use the bicubic downsampling to generate the blurry LR images as the inputs. We then generate recovered HR images with the following methods: the bicubic upsampling, deblurring method SRN (Tao et al. 2018 with super resolution method RCAN (Zhang et al. 2018b), joint restoration and super-resolution method ED-DSRN (Zhang et al. 2018a), and the proposed GFN. For the hazy dataset, we first apply the single image depth estimation method, the Monodepth2 (Godard et al. 2019), to predict a depth map for each image and then synthesize the hazy image following the instruction of the RESIDE dataset (Li et al. 2018a). We compare the proposed GFN with the following approaches: the bicubic upsampling, dehazing method PFFNet (Mei et al. 2018) with super resolution method RCAN (Zhang et al. 2018b), and joint restoration and super-resolution method ED-DSRN (Zhang et al. 2018a). We use the above methods to recover HR images and then use the YOLOv3 (Redmon and Farhadi 2018) to evaluate the detection accuracy.

We show the detection accuracy in Tables 9 and 10. The HR images restored from the proposed GFN obtain the best detection accuracy in both applications. The qualitative results in Figs. 14 and 15 demonstrate that our GFN not only generates clean HR outputs but also improves the detection algorithm to recognize the cars and pedestrians.



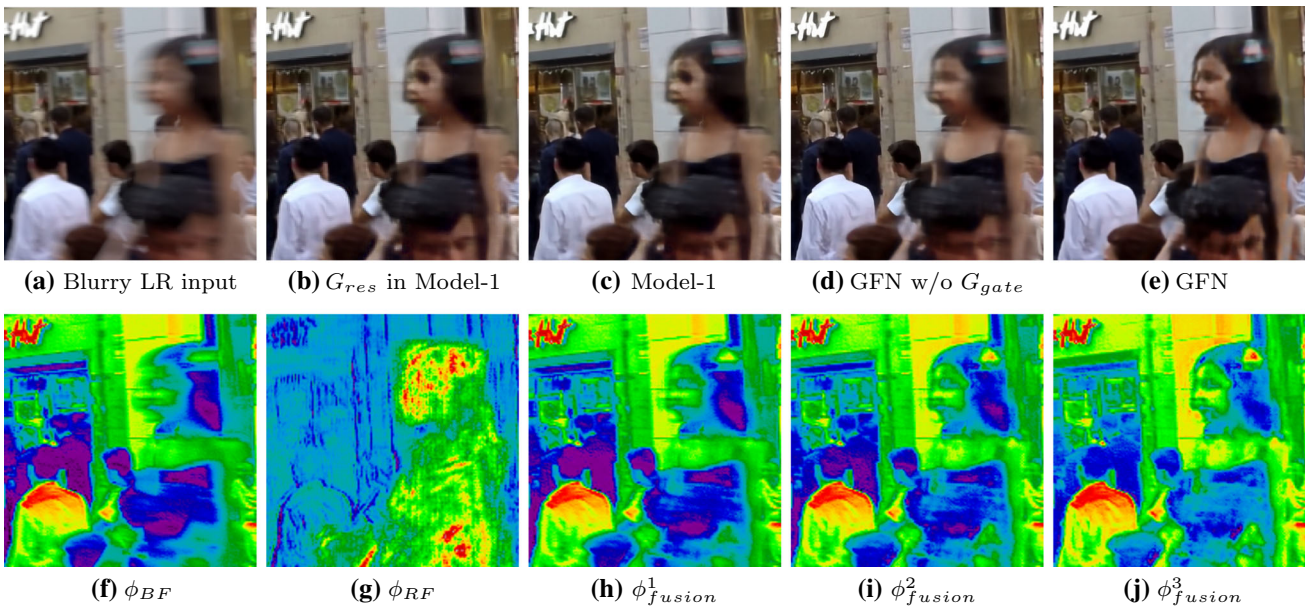
(a) EDSR + DCP 18.68 / 0.823 **(b)** EDSR + NLD 16.04 / 0.761 **(c)** EDSR + DGFN 17.54 / 0.875 **(d)** GFN (ours) 28.85 / 0.961

Fig. 12 More visual comparisons on the LR-RESIDE dataset. The evaluated methods include EDSR (Lim et al. 2017), DCP (He et al. 2011), NLD (Berman et al. 2016), and DGFN (Ren et al. 2018). The proposed model generates sharper HR images with more details

Table 8 Quantitative comparison with the state-of-the-art restoration methods on three applications

	Restoration module	DeepDeblur	DeblurGAN	SRN
GOPRO dataset				
Deblurring PSNR	*29.16*	27.48	27.02	**30.26**
	Restoration module	DGFN	GCANet	PFFNet
RESIDE dataset				
Dehazing PSNR	24.46	23.47	*26.32*	**28.20**
	Restoration module	IDGAN	RESN	DID-MDN
Rain1200 dataset				
Deraining PSNR	*29.36*	27.50	29.12	**30.10**

All the comparison methods for each application are trained using the same setting
Bold values and italic values indicate the best and the second-best performance respectively



(a) Blurry LR input **(b)** G_{res} in Model-1 **(c)** Model-1 **(d)** GFN w/o G_{gate} **(e)** GFN

(f) ϕ_{BF} **(g)** ϕ_{RF} **(h)** ϕ^1_{fusion} **(i)** ϕ^2_{fusion} **(j)** ϕ^3_{fusion}

Fig. 13 Qualitative results of the ablation study. ϕ_{BF} denotes the base features from the base module G_{base} and ϕ_{RF} denotes the features from the restoration module G_{res}. All the models are trained on the LR-GOPRO dataset with the same training settings as the proposed GFN

Table 9 Objects detection results on the KITTI detection dataset (Geiger et al. 2012) with non-uniform motion blur

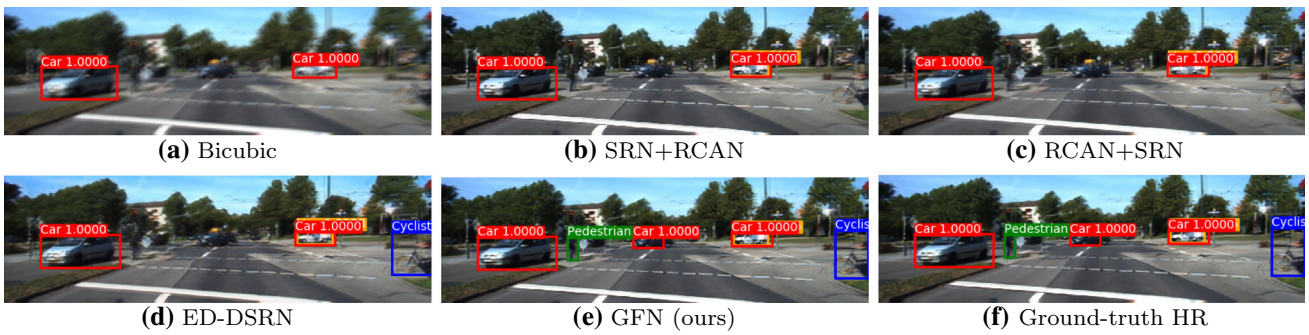
YOLOv3	Bicubic	SRN + RCAN	RCAN + SRN	ED − DSRN	GFN	Ground-truth HR
Car	0.258	0.481	0.481	0.416	0.499	0.812
Van	0.149	0.358	0.392	0.298	0.406	0.724
Truck	0.208	0.558	0.578	0.499	0.584	0.842
Pedestrian	0.164	0.327	0.329	0.305	0.370	0.604
Person Sitting	0.028	0.187	0.105	0.122	0.163	0.436
Cyclist	0.026	0.203	0.171	0.158	0.283	0.592
Tram	0.108	0.331	0.314	0.272	0.383	0.796
mAP	0.120	0.316	0.308	0.267	**0.352**	0.646

We test different joint deblurring and super-resolution methods, and use YOLOv3 (Redmon and Farhadi 2018) as the detection algorithm. The comparison methods include bicubic upsampling, deblurring method SRN (Tao et al. 2018) + super resolution method RCAN (Zhang et al. 2018b), joint restoration and super-resolution method ED-DSRN (Zhang et al. 2018a), and the proposed GFN. We also show the detection result using the ground-truth sharp HR image. The mAP is the abbreviation of mean average precision. Bold values indicate the best detection precision except for the Ground-truth HR

Table 10 Objects detection results on the KITTI detection dataset (Geiger et al. 2012) with haze degradation

YOLOv3	Bicubic	PFFNet + RCAN	RCAN + PFFNet	ED − DSRN	GFN	Ground-truth HR
Car	0.146	0.073	0.254	0.431	0.505	0.812
Van	0.053	0.033	0.113	0.228	0.301	0.724
Truck	0.033	0.008	0.047	0.119	0.197	0.842
Pedestrian	0.178	0.067	0.297	0.406	0.461	0.604
Person Sitting	0.000	0.000	0.021	0.204	0.327	0.436
Cyclist	0.055	0.027	0.179	0.196	0.279	0.592
Tram	0.000	0.000	0.007	0.097	0.239	0.796
mAP	0.058	0.026	0.117	0.219	**0.303**	0.646

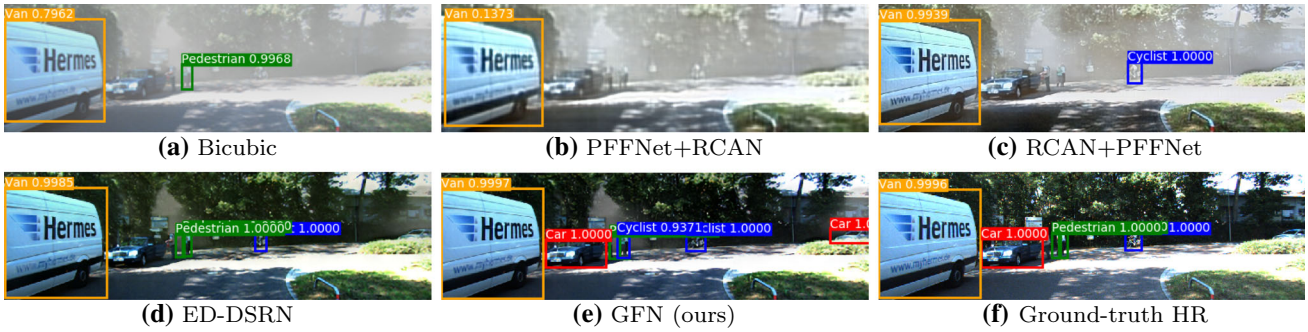
We test different joint dehazing and super-resolution methods, and use YOLOv3 (Redmon and Farhadi 2018) as the detection algorithm. The comparison methods include bicubic upsampling, dehazing method PFFNet (Mei et al. 2018) + super resolution method RCAN (Zhang et al. 2018b), joint restoration and super-resolution method ED-DSRN (Zhang et al. 2018a), and the proposed GFN. We also show the detection result using the ground-truth sharp HR image. The mAP is the abbreviation of mean average precision. Bold values indicate the best detection precision except for the Ground-truth HR



(a) Bicubic **(b)** SRN+RCAN **(c)** RCAN+SRN

(d) ED-DSRN **(e)** GFN (ours) **(f)** Ground-truth HR

Fig. 14 Detection results using the recovered images from different methods. We compare the following methods: bicubic upsampling, deblurring method SRN (Tao et al. 2018) + super resolution method RCAN (Zhang et al. 2018b), joint restoration and super-resolution method ED-DSRN (Zhang et al. 2018a), and the proposed GFN



(a) Bicubic **(b)** PFFNet+RCAN **(c)** RCAN+PFFNet

(d) ED-DSRN **(e)** GFN (ours) **(f)** Ground-truth HR

Fig. 15 Detection results using the recovered images from different methods. We compare the following methods: bicubic upsampling, dehazing method PFFNet (Mei et al. 2018) + super resolution method RCAN (Zhang et al. 2018b), joint restoration and super-resolution method ED-DSRN (Zhang et al. 2018a), and the proposed GFN

References

Agustsson, E., & Timofte, R. (2017). Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPR workshops*

Bai, Y., Zhang, Y., Ding, M., & Ghanem, B. (2018). Finding tiny faces in the wild with generative adversarial network. In *IEEE conference on computer vision and pattern recognition*.

Bao, W., Zhang, X., Yan, S., & Gao, Z. (2017). Iterative convolutional neural network for noisy image super-resolution. In *IEEE international conference on image processing*.

Bascle, B., Blake, A., & Zisserman, A. (1996). Motion deblurring and super-resolution from an image sequence. In *European conference on computer vision*.

Berman, D., Avidan, S., et al. (2016). Non-local image dehazing. *IEEE conference on computer vision and pattern recognition* (pp. 1674–1682).

Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G. (2019). Gated context aggregation network for image dehazing and deraining. In *IEEE winter conference on applications of computer vision* (pp. 1375–1383). IEEE.

Cho, S., & Lee, S. (2009). Fast motion deblurring. *ACM Transactions on Graphics*, *28*(5), 145.

Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307.

Fan, Z., Wu, H., Fu, X., Huang, Y., & Ding, X. (2017). Residual-guide network for single image deraining. In *ACM international conference on multimedia*.

Fattal, R. (2008). Single image dehazing. *ACM Transactions on Graphics*, *27*(3), 72.

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The kitti vision benchmark suite. In *IEEE conference on computer vision and pattern recognition*.

Godard, C., Mac Aodha, O., Firman, M., & Brostow, G. J. (2019). Digging into self-supervised monocular depth prediction. In *IEEE international conference on computer vision*.

Gong, D., Yang, J., Liu, L., Zhang, Y., Reid, I. D., Shen, C., Van Den Hengel, A., & Shi, Q. (2017). From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *IEEE conference on computer vision and pattern recognition*.

Haris, M., Shakhnarovich, G., & Ukita, N. (2018). Deep backprojection networks for super-resolution. In *IEEE conference on computer vision and pattern recognition*.

He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(12), 2341–2353.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *IEEE conference on computer vision and pattern recognition*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hradiš, M., Kotera, J., Zemcík, P., & Šroubek, F. (2015). Convolutional neural networks for direct text deblurring. In *British machine vision conference*.

Huang, J. B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In *IEEE conference on computer vision and pattern recognition*.

Hu, Z., Xu, L., & Yang M. H. (2014). Joint depth estimation and camera shake removal from single blurry image. In *IEEE conference on computer vision and pattern recognition*.

Jiang, T. X., Huang, T. Z., Zhao, X. L., Deng, L. J., & Wang, Y. (2017). A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In *IEEE conference on computer vision and pattern recognition*.

Johnson, J., Alahi, A., & Li, F. F. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*.

Kim, J., Lee, J. K., & Lee, K. M. (2016a). Accurate image super-resolution using very deep convolutional networks. In *IEEE conference on computer vision and pattern recognition*.

Kim, J., Lee, J. K., & Lee, K. M. (2016b). Deeply-recursive convolutional network for image super-resolution. In *IEEE conference on computer vision and pattern recognition*.

Kim, T. H., Ahn, B., & Lee, K. M. (2013). Dynamic scene deblurring. In *IEEE international conference on computer vision*.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *International conference on learning representations*.

Köhler, R., Hirsch, M., Mohler, B., Schölkopf, B., & Harmeling, S. (2012). Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *European conference on computer vision*.

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., & Matas, J. (2018). DeblurGAN: Blind motion deblurring using conditional adversarial networks. In *IEEE conference on computer vision and pattern recognition*.

Lai, W. S., Huang, J. B., Ahuja, N., & Yang, M. H. (2019). Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(11), 2599–2613.

Lai, W. S., Huang, J. B., Hu, Z., Ahuja, N., & Yang, M. H. (2016). A comparative study for single image blind deblurring. In *IEEE conference on computer vision and pattern recognition*.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A. P., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE conference on computer vision and pattern recognition*.

Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). Aod-net: All-in-one dehazing network. In *IEEE international conference on computer vision*.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., et al. (2018a). Reside: A benchmark for single image dehazing. *IEEE Transactions on Image Processing*, *28*(1), 492–505.

Li, G., He, X., Zhang, W., Chang, H., Dong, L., & Lin, L. (2018b). Non-locally enhanced encoder-decoder network for single image de-raining. arXiv:1808.01491.

Li, M., Xie, Q., Zhao, Q., Wei, W., Gu, S., Tao, J., & Meng, D. (2018c). Video rain streak removal by multiscale convolutional sparse coding. In *IEEE conference on computer vision and pattern recognition*.

Li, X., Wu, J., Lin, Z., Liu, H., & Zha, H. (2018d). Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *European conference on computer vision*.

Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution. In *CVPR workshops*.

Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., & Ma, Y. (2013). Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 171–184.

Liu, J., Yang, W., Yang, S., & Guo, Z. (2018). Erase or fill? deep joint recurrent rain removal and reconstruction in videos. In *IEEE conference on computer vision and pattern recognition*.

Mairal, J., Bach, F., Ponce, J., Sapiro, G. (2009). Online dictionary learning for sparse coding. In *International conference on machine learning*.

Mei, K., Jiang, A., Li, J., & Wang, M. (2018). Progressive feature fusion network for realistic image dehazing. In *Asian conference on computer vision*.

Nah, S., Kim, T. H., & Lee, K. M. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. In *IEEE conference on computer vision and pattern recognition*.

Nimisha, T. M., Singh, A. K., & Rajagopalan, A. N. (2017). Blur-invariant deep learning for blind-deblurring. In *IEEE international conference on computer vision*.

Noroozi, M., Chandramouli, P., & Favaro, P. (2017). Motion deblurring in the wild. In *German conference on pattern recognition*.

Pan, J., Sun, D., Pfister, H., & Yang, M. H. (2016). Blind image deblurring using dark channel prior. In *IEEE conference on computer vision and pattern recognition*.

Paramanand, C., & Rajagopalan, A. N. (2013). Non-uniform motion deblurring for bilayer scenes. In *IEEE conference on computer vision and pattern recognition*.

Park, H., & Lee, K. M. (2017). Joint estimation of camera pose, depth, deblurring, and super-resolution from a blurred image sequence. In *IEEE international conference on computer vision*.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv:1804.02767.

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M. H. (2016). Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*.

Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., & Yang, M. H. (2018). Gated fusion network for single image dehazing. In *IEEE conference on computer vision and pattern recognition*.

Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, *10*(1–3), 19–41.

Sajjadi, M. S., Schölkopf, B., & Hirsch, M. (2017). Enhancenet: Single image super-resolution through automated texture synthesis. In *IEEE international conference on computer vision*.

Schmidt, U., Rother, C., Nowozin, S., Jancsary, J., & Roth, S. (2013). Discriminative non-blind deblurring. In *IEEE conference on computer vision and pattern recognition*.

Schmidt, U., Schelten, K., & Roth, S. (2011). Bayesian deblurring with integrated noise estimation. In *IEEE conference on computer vision and pattern recognition*.

Schulter, S., Leistner, C., & Bischof, H. (2015). Fast and accurate image upscaling with super-resolution forests. In *IEEE conference on computer vision and pattern recognition*.

Shan, Q., Jia, J., & Agarwala, A. (2008). High-quality motion deblurring from a single image. *ACM Transactions on Graphics*, *27*(3), 73.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *IEEE conference on computer vision and pattern recognition*.

Su, S., Delbracio, M., Wang, J., Sapiro, G., Heidrich, W., & Wang, O. (2017). Deep video deblurring for hand-held cameras. In *IEEE conference on computer vision and pattern recognition*.

Tai, Y., Yang, J., & Liu, X. (2017). Image super-resolution via deep recursive residual network. In *IEEE conference on computer vision and pattern recognition*.

Tan, R. T. (2008). Visibility in bad weather from a single image. In *IEEE conference on computer vision and pattern recognition*.

Tao, X., Gao, H., Wang, Y., Shen, X., Wang, J., & Jia, J. (2018). Scale-recurrent network for deep image deblurring. In *IEEE conference on computer vision and pattern recognition*.

Timofte, R., De Smet, V., & Van Gool, L. (2014). A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian conference on computer vision*.

Xu, L., Zheng, S., & Jia, J. (2013). Unnatural l0 sparse representation for natural image deblurring. In *IEEE conference on computer vision and pattern recognition*.

Xu, X., Sun, D., Pan, J., Zhang, Y., Pfister, H., & Yang, M. H. (2017). Learning to super-resolve blurry face and text images. In *IEEE international conference on computer vision*.

Yamaguchi, T., Fukuda, H., Furukawa, R., Kawasaki, H., & Sturm. P. (2010). Video deblurring and super-resolution technique for multiple moving objects. In *Asian conference on computer vision*.

Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z., & Yan, S. (2017). Deep joint rain detection and removal from a single image. In *IEEE conference on computer vision and pattern recognition*.

Yu, X., Fernando, B., Hartley, R., & Porikli, F. (2018). Super-resolving very low-resolution face images with supplementary attributes. In *IEEE conference on computer vision and pattern recognition*.

Zhang, H., & Patel, V. M. (2018a). Densely connected pyramid dehazing network. In *IEEE conference on computer vision and pattern recognition*.

Zhang, H., & Patel V. M. (2018b). Density-aware single image deraining using a multi-stream dense network. In *IEEE conference on computer vision and pattern recognition*.

Zhang, H., Yang, J., Zhang, Y., Nasrabadi, N. M., & Huang, T. S. (2011). Close the loop: Joint blind image restoration and recognition with sparse representation prior. In *IEEE international conference on computer vision*.

Zhang, H., Sindagi, V., & Patel, V. M. (2017). Joint transmission map estimation and dehazing using deep networks. arXiv:1708.00581.

Zhang, H., Sindagi, V., & Patel, V. M. (2019). Image de-raining using a conditional generative adversarial network. In *Transactions on*

circuits and systems for video technology. IEEE. https://doi.org/10.1109/TCSVT.2019.2920407.

Zhang, L., Zhang, H., Shen, H., & Li, P. (2010). A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, *90*(3), 848–859.

Zhang, X., Wang, F., Dong, H., & Guo, Y. (2018a). A deep encoder-decoder networks for joint deblurring and super-resolution. In *IEEE international conference on acoustics, speech, and signal processing*.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018b). Image super-resolution using very deep residual channel attention networks. In *European conference on computer vision*.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018c). Residual dense network for image super-resolution. In *IEEE conference on computer vision and pattern recognition* (pp. 2472–2481).

Zou, W. W., & Yuen, P. C. (2012). Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, *21*(1), 327–340.