Pixel-Guided Dual-Branch Attention Network for Joint Image Deblurring and Super-Resolution

Si Xi[1] Jia Wei[1] Weidong Zhang
Netease Games AI Lab, Hangzhou, China
xisi789@126.com, wzwrwj@163.com, zagwin@gmail.com

Figure 1. We train the Pixel-Guided Dual-Branch Attention Network for super-resolution and deblurring, successfully restore fine details. The left image is input, and the right one is output.

Abstract

Image deblurring and super-resolution (SR) are computer vision tasks aiming to restore image detail and spatial scale, respectively. Besides, only a few recent works of literature contribute to this task, as conventional methods deal with SR or deblurring separately. We focus on designing a novel Pixel-Guided dual-branch attention network (PDAN) that handles both tasks jointly to address this issue. Then, we propose a novel loss function better focus on large and medium range errors. Extensive experiments demonstrated that the proposed PDAN with the novel loss function not only generates remarkably clear HR images and achieves compelling results for joint image deblurring and SR tasks. In addition, our method achieves second place in NTIRE 2021 Challenge on track 1 of the Image Deblurring Challenge.

[1]These authors contributed equally to this work and should be considered co-first authors.

1. Introduction

Deep Neural Network (DNN) has promoted many practical applications, including image classification, video understanding, and many other applications. Recently, Image deblurring and super-resolution (SR) [12] have become important tasks that aim at recovering a sharp latent image with more detailed information and finer picture quality.

Real-world blurs typically have unknown blur kernels, and the downsampling function from HR to LR is uncertain [8, 5]. Image deblurring and SR are challenging than Image interpolation. Although significant progress has been made recently based on Deep Neural Network techniques, it is still an open problem.

Previous works developed which solely focus on deblurring or SR [22, 3]. However, the tasks of image deblurring and SR are highly correlated. Firstly, the feature information can be shared between them. In addition, these two tasks can complement each other: better deblurring improves SR sharpness, and the SR information can refine deblurring results vice versa.

Therefore, we propose a new network with the dual-branches architecture to solve image deblurring and super-resolution. The pixel-guided hard example mining loss fo-

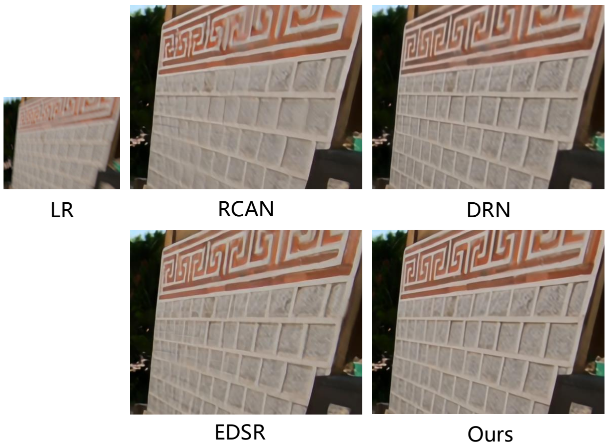

Figure 2. Visualization comparison among RCAN, EDSR, DRN, and our method.

cuses on the hard pixel. And auto-focusing evaluation functions are used to ensembles, different models. In the experiments section, our method generates high-resolution results with clear details, especially in NTIRE 2021 Image Deblurring Track 1 [13]. Low Resolution, Our method ranks second in the resulting PSNR measurement. (see Figure 2)

2. Related Work

Image deblurring and image super-resolution have been the topic of extensive research in the past several decades. In this section, we will present the early approaches to solve image deblurring and image super-resolution and some methods to solve the two tasks jointly.

2.1. Image Deblurring

The complex, uneven blur is caused by camera shake, depth changes, object movement, and defocusing, making it a difficult task in computer vision. Conventional deblurring methods[17, 19, 6] focus on estimating the blur kernel corresponding to each pixel, which is a severe ill-conditioned problem. They usually make some assumptions about the blur kernel. Methods[1, 23] uses a simple parametric prior model to estimate the local linear blur kernel quickly.

Early learning-based methods [4, 26] mainly use a convolutional neural network to estimate the unknown blur kernel to improve the accuracy of blind recovery, and then use the conventional deconvolution method to recover the blurred image.

Recently, some end-to-end deep learning-based image deblurring algorithms [15, 7, 8] have been proposed, inspired by research work, such as image transmission based on Generative Adversarial Network (GAN) [9]. Experimental results show that, compared with conventional methods, these methods have achieved good results in subjective and objective quality.

However, most conventional algorithms mainly focus on how to solve the motion blur caused by the movement of simple target objects, camera translation, rotation, and other factors. Simultaneously, the blurry images of real dynamic scenes suffer from complex and uneven blur degradation. Therefore, traditional methods are difficult to solve the problem of non-uniform blur in real scenes effectively. They usually involve iteration, which leads to time-consuming and limited performance.

2.2. Image Super Resolution

Image SR focuses on recovering HR images from LR images. In order to solve the task of image super-resolution, many methods have been proposed in the computer vision field. Over the years, the powerful capability of convolutional neural networks can significantly improve SR tasks. The basic idea of those kinds of approaches is to extract features from the original LR images and increasing the spatial resolution at the end of the network. By removing unnecessary modules in the traditional residual network, Lim et al.[11] proposed EDSR and MDSR, and they have made significant improvement. Furthermore, in RCAN[29], attention-based learning was used in combination with residual learning, L1 pixel loss function, and subpixel upsampling method to achieve the state-of-the-art results in image SR.

2.3. Multi-Branch Network in Image Restoration Task

In many deep learning-based algorithms, the multi-branch network architecture has been widely used, and some image restoration attempts have also been made. Different branches are usually designed as different architectures for specific tasks. Li et al.[10] proposed a deep guided network for image deblurring tasks, including image deblurring branches and scene depth feature extraction branches. The scene depth feature extraction branch guides the deblurring image branch to restore a clear image. The image restoration task usually contains two parts of information, that is, the image structure and details. Pan et al. [16] combining these functions, proposed a parallel convolutional neural network for image restoration tasks. The network includes two parallel branches to estimate the image structure and detailed information and restore them in an end-to-end manner. Therefore, combining certain features of the image itself can help improve the quality of restoration.

From the point of view of signal processing, global uniform blur is linear movement invariant processing. In a local non-uniform blur, the blur kernel will change with respect to the spatial position. Therefore, different network mechanisms should be considered to handle the two types of ambiguity separately.

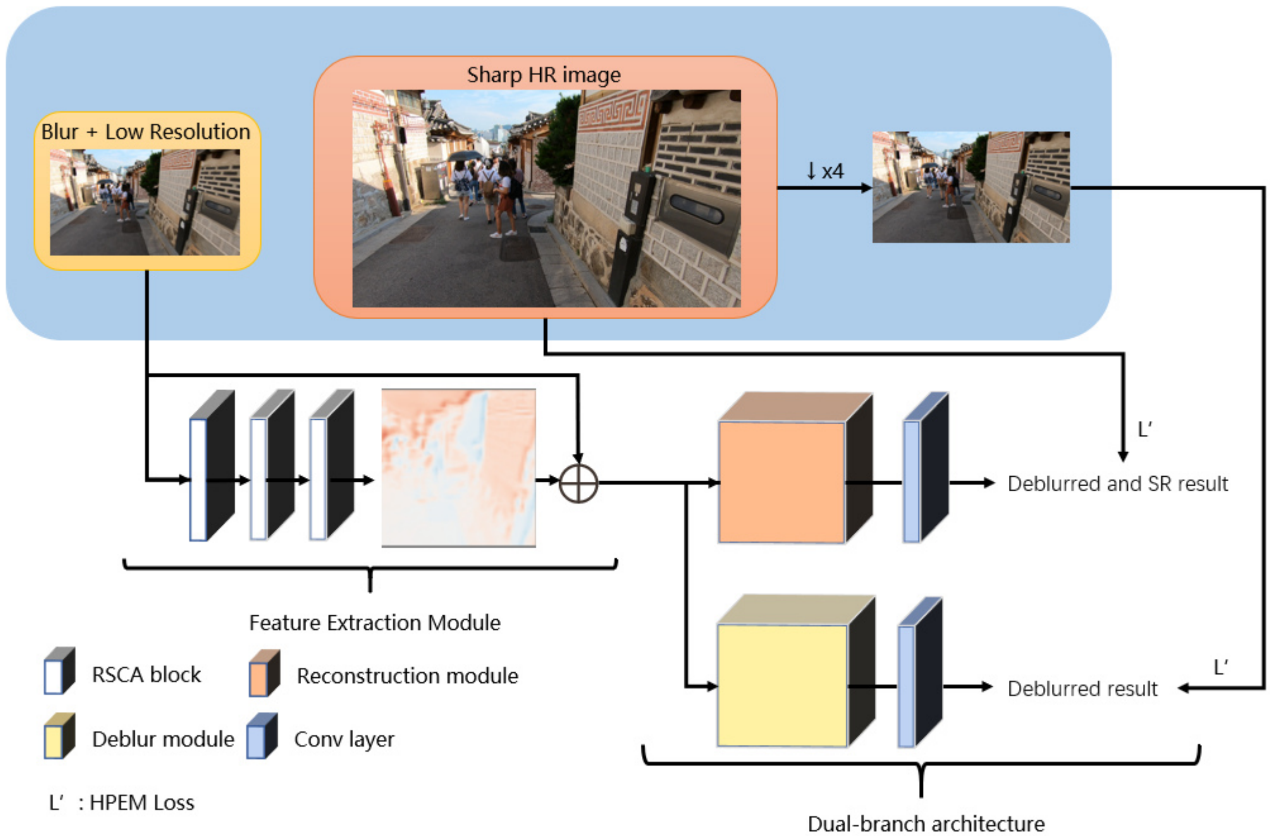

Figure 3. The pipeline of our pixel-guided dual-branch attention network.

Since the input image is extremely blurry and low resolution, the joint problem is more challenging than the individual problem. Recently, some algorithms [2, 18, 25] recover HR images from LR and blurred video sequences based on the neighboring information of the video. However, these kinds of algorithms cannot be applied to the case where a single image is used as input. Zhang et al. [28] focused on solving LR images degraded by uniform gaussian blur. Zhang et al. [27] adopt gated fusion network using dual-branch design.

3. Proposed Method

This section introduces the entire network architecture and then defines the loss function we proposed to optimize the model better. We will finally show our optimization scheme.

3.1. Network Architecture

The framework of our network is shown in Figure 3. Our network takes a single blurred LR image L_{blur} as input, and the network can better restore clear HR images H_{sharp}. According to the competition requirements, the spatial resolution of L_{sharp} is 4x larger than L_{blur}.

$$L_{blur} = F_{blurry}(H_{sharp}) \downarrow_s \tag{1}$$

First, the network extracts the blurry LR features of the input. Our pixel-guided dual-branch attention network (PDAN) consists of three major modules: (i) a feature extraction module to blurry blur and LR features, (ii) a deblurring module to predict a sharp LR image, (iii) a reconstruction module to reconstruct the final sharp HR output image.

Residual Spatial and Channel Attention Module. Recently, in many deep neural networks, the application of the channel attention module to improve the performance of the network has proven to be very effective. However, the channel attention layer in Residual Channel Attention Network(RCAN) [29] is too simple to achieve better performance. Here, we proposed a novel module for the better extracting feature of blurry LR image.

Inspired by the successful case of residual blocks in [11], we propose residual spatial and channel attention module (RSCA) (see Figure 4). We are more focused on extracting information features better by fusing cross-channel and spatial information.

Deblurring Module. This module aims to restore a sharp LR image from the blurry LR features extracted by

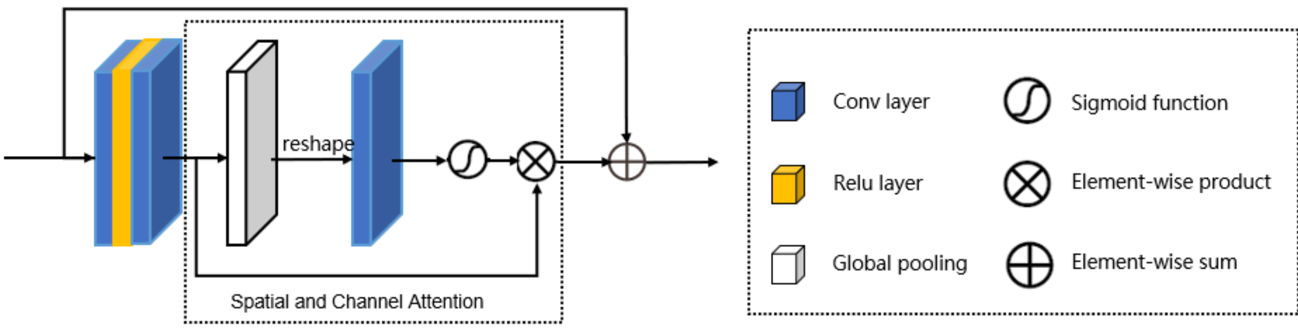

Figure 4. Residual Spatial and Channel Attention module. (RSCA)

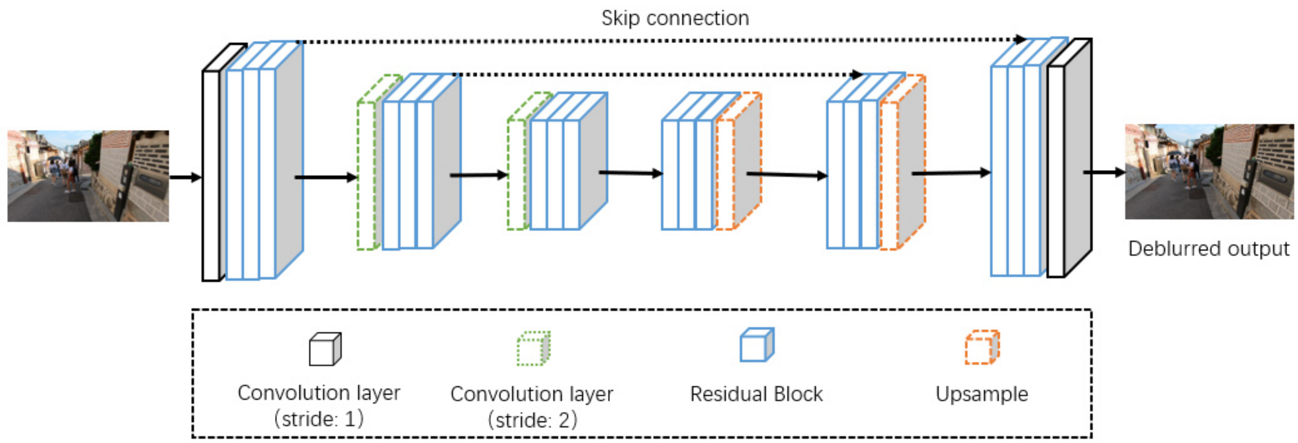

Figure 5. Deblurring module.

the previous module. In order to enlarge the receptive field of the deblurring module, we adopted a residual encoder-decoder architecture. First, the encoder aims to down-sample the feature map twice. Each downsampling uses a ResBlocks, followed by a stride convolutional layer with LeakyReLU. The decoder aims to increase the spatial resolution of the feature map and is completed by two deconvolution layers. Finally, two additional residual blocks are used to reconstruct a sharp LR image L_{sharp} (see Figure 5).

As described in Section 2, there is a lot of abundant information in the blurry and LR images. The deblurring module aims to recover more useful deblurring information. The abundant deblurring features can be learned through dual-branch architecture. In the test phase, the deblurring module is not used for computational efficiency.

Reconstruction Module. The shallow feature from our feature extraction module is fed into convolutional layers and pixel-shuffling layers [20] to increase the spatial resolution by 4x. Then reconstruct the upscaled features through a conv layer to further reconstruct the HR output image. With dual-branch attention architecture, more abundant deblurring and SR information are easier learned in the training process.

Based on the dual-branch attention architecture and RSCA module, we can better address the joint image deblurring and SR tasks. We construct a very deep DAN optimized with the novel loss function shown in Section 3.2 and achieves notable performance improvements.

3.2. Loss Function

Given a training image L_{blur} and a network Φ, we can predict the final sharp HR image H'_{sharp}. Similarly, we formalize this process as $H'_{sharp} = \Phi(L_{blur})$. The loss is defined as:

$$Loss(H_{sharp}, H'_{sharp}) = \sum_{i=1}^{h \times w} f(h_i, h'_i) \tag{2}$$

where H_{sharp} is the ground-truth clear HR images, in general, for $f(x)$ in the above equation, the L1 function has been used for perceptual SR. The L1 loss function is defined as $L1(x) = |x|$, which meant the magnitude of the gradient is the same for all the points, but larger errors disproportionately influence the step size.

For joint image deblurring and SR tasks, the blurry image is generated by merging subsequent frames. We found

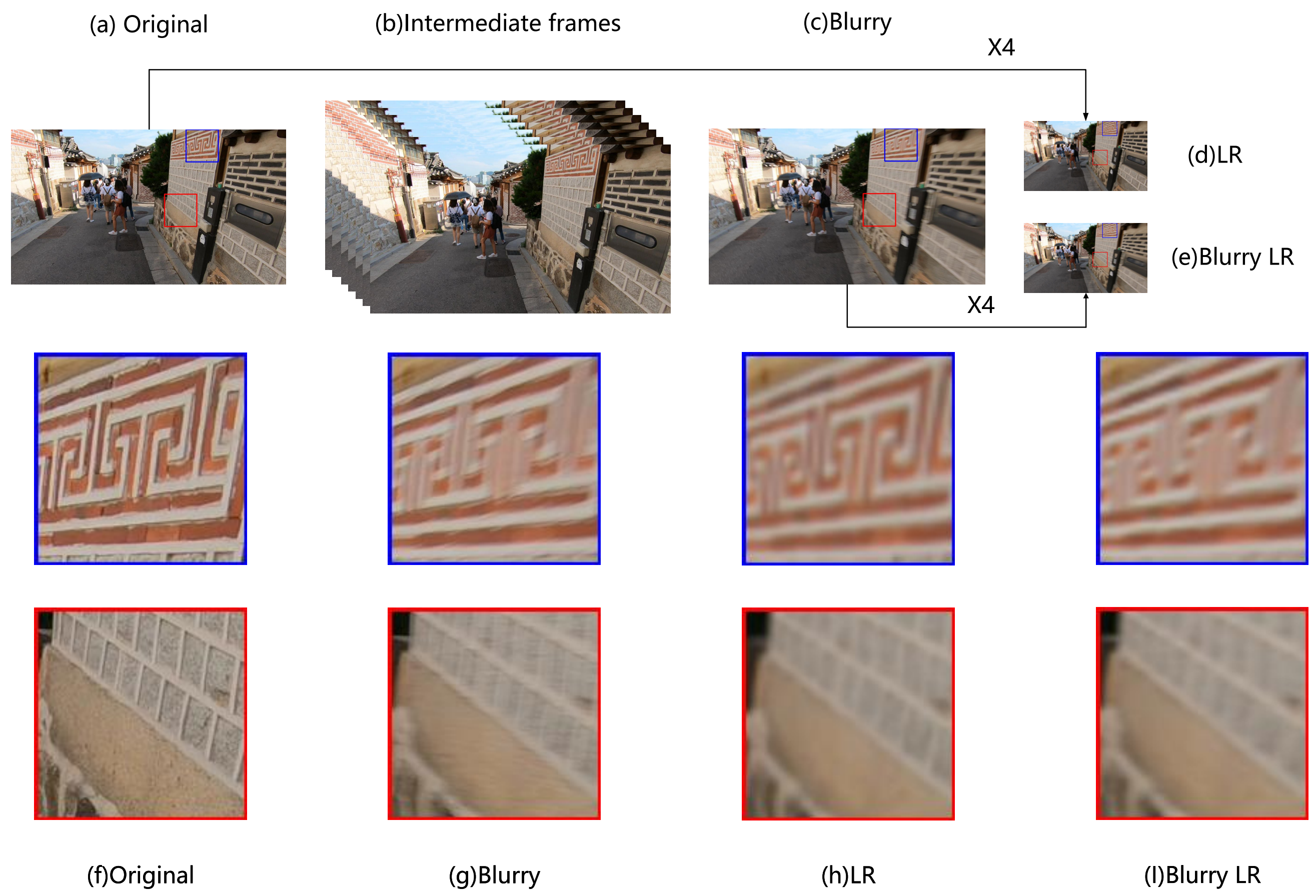

Figure 6. Data generation process.

that many fixed objects are clear, and only part of the object motion caused the blur phenomenon. (see Figure 7). Therefore, in our training images L_{blur}, each pixel's composition is not uniform. Some pixels are generated through a large amount of blur and LR, while other pixels may only be generated through LR. When compounding the contributions from multiple pixels, the gradient will be dominated by small errors, but the step size by larger errors. This will cause the L1 loss function to be unable to optimize each pixel effectively.

Therefore, we need to balance large errors and small errors better to optimize the network. To solve this problem, inspired by [24, 21], first we use the L1 loss function training model from scratch. We further use a hard example mining strategy to adaptively focus more on the difficult pixels, thus encouraging the model to restore image detail and spatial scale, which is called Hard Pixel Example Mining loss (HPEM). Specifically, we calculate the error of all pixels and sort all pixels in descending order, and then set the top p percent of pixels as hard samples according to the sorting. Finally we increase their losses through the weight w to force the model to pay more attention to hard pixels. Similarly, we formalize this process as:

$$Loss_{Hard} = \sum_{i}^{h \times w} \left\| M_i^{hard} \cdot (s_i - s_i') \right\|_1 \tag{3}$$

where M^{hard} is the binary mask which mean the hard pixels. We fix p as 50% and w as 2. The total loss function is the weighted sum of the following two losses:

$$Loss_{Blurry,LR} = L1 + w \times Loss_{Hard} \tag{4}$$

3.3. Optimization Scheme

The proposed pixel-guided dual-branch attention network contains two parallel branches, which solve image deblurring and image super-resolution, respectively. To better optimize the model, the training process is divided into three phases:

- Train the dual-branch attention network with the REalistic and Dynamic Scenes dataset (REDS [12]) the L1 loss function.

Modifications	Models			
	baseline	model 1	model 2	model 3
baseline	✓	✓	✓	✓
attention		✓	✓	✓
dual-branch			✓	✓
loss				✓
PSNR	27.61	27.68	27.84	27.89

Table 1. Ablation study. This table reports the average PSNR obtained by different methods of our model on the REDS validation data set.

- Fine-tune the pixel-guided dual-branch attention network with the REDS [12] dataset and the novel loss function $Loss_{Blurry,LR}$.
- Fine-tune the pixel-guided dual-branch attention network with the REDS [12] and extra dataset generated by us, and use $Loss_{Blurry,LR}$ to optimize the model.

4. Experiments

In this section, we first introduce experiment datasets, make a qualitative and quantitative comparison with existing approaches, and finally discuss the influence of our design on the model.

4.1. Dataset

For training, we use REDS [12] dataset, which contains 24,000 training images. It includes a variety of scenes and can be used for SR and deblurring. We randomly crop patches of size 256 × 256 from the training H_{sharp} images and corresponding crop 64 × 64 size patches from the training L_{blur} images. Rotation and flip are applied to augment the training data. The validation set contains 3000 images with corresponding ground truth. Therefore, metrics can be calculated by ground truth. The reference image of the deblurring branch is obtained by downsampling high-resolution images.

4.2. Synthesizing External Dataset

As shown in Figure 6, we synthesize 72,000 external blurry LR images through the following three steps:

- **Frame Interpolation:** We used the REDS 120fps [12] to obtain the extra data. A higher frame rate can get smoother blurred images. We first increase the frame rate by inter-frame interpolation. And use open-source frame interpolation methods [14] to create intermediate frames. In this way, we increased the virtual frame rate to 1920 fps.
- **Blur Synthesis:** We average 1920 fps images in the signal space and get 24 fps blurred images.
- **Downscale:** We use the MATLAB function to downsample images.

Options	DRN	EDSR*	RCAN*	PDAN(ours)
# Parameters	10.5M	43M	15.5M	61M
# Flops	1.433T	2.896T	0.919T	3.669T
Dual-Branch	No	No	No	Yes
Loss function	L1	L1	L1	L1+HPEM
SSIM	0.7726	0.7732	0.7737	**0.7798**
PSNR	27.53	27.55	27.61	**27.89**

Table 2. Quantitative results on the REDS validation dataset compared with EDSR, DRN, and RCAN. * indicates that the patch size is 256×256.

4.3. Implementation Details

During the training, we performed a lot of data enhancement, such as rotation, mirroring, brightness. The size of the input patches is set to 256 × 256. L1 loss and hard pixel example mining loss we proposed were used and improved PSNR and SSIM.

We implement the proposed network with the PyTorch 1.7 framework and using eight Tesla V100 running for distributed training. Use the ADAM optimizer for training and set the initial learning rate to 10^{-4}. When the loss stagnated 300,000 iterations, the learning rate drops by 50 percent. The batch size is set to 128.

We feed a full-size RGB image (with a typical resolution of 320 × 180) into the model to obtain a high-resolution image. The model takes 1.7 seconds per image during inference to joint deblur and SR a 1280-by-720 pixels image.

Evaluation Metric: We use PSNR, SSIM, and LPIPS as evaluation indicators. PSNR and SSIM pay more attention to fidelity, and LPIPS focuses on visual quality.

4.4. Evaluation on REDS and Ablation Study

Our proposed method aims to solve joint image deblurring and SR tasks, so we evaluate PDAN on the REDS dataset. As shown in Table 2 and Figure 7, our PDAN achieves the best PSNR performance, indicating our results can better deal with image deblurring and super-resolution tasks at the same time. Since we have adopted the dual-branch attention architecture, PDAN has more parameters (61 M) than EDSR (43 M) and RCAN (15.5M), which can achieve better performance in joint image deblurring and SR tasks.

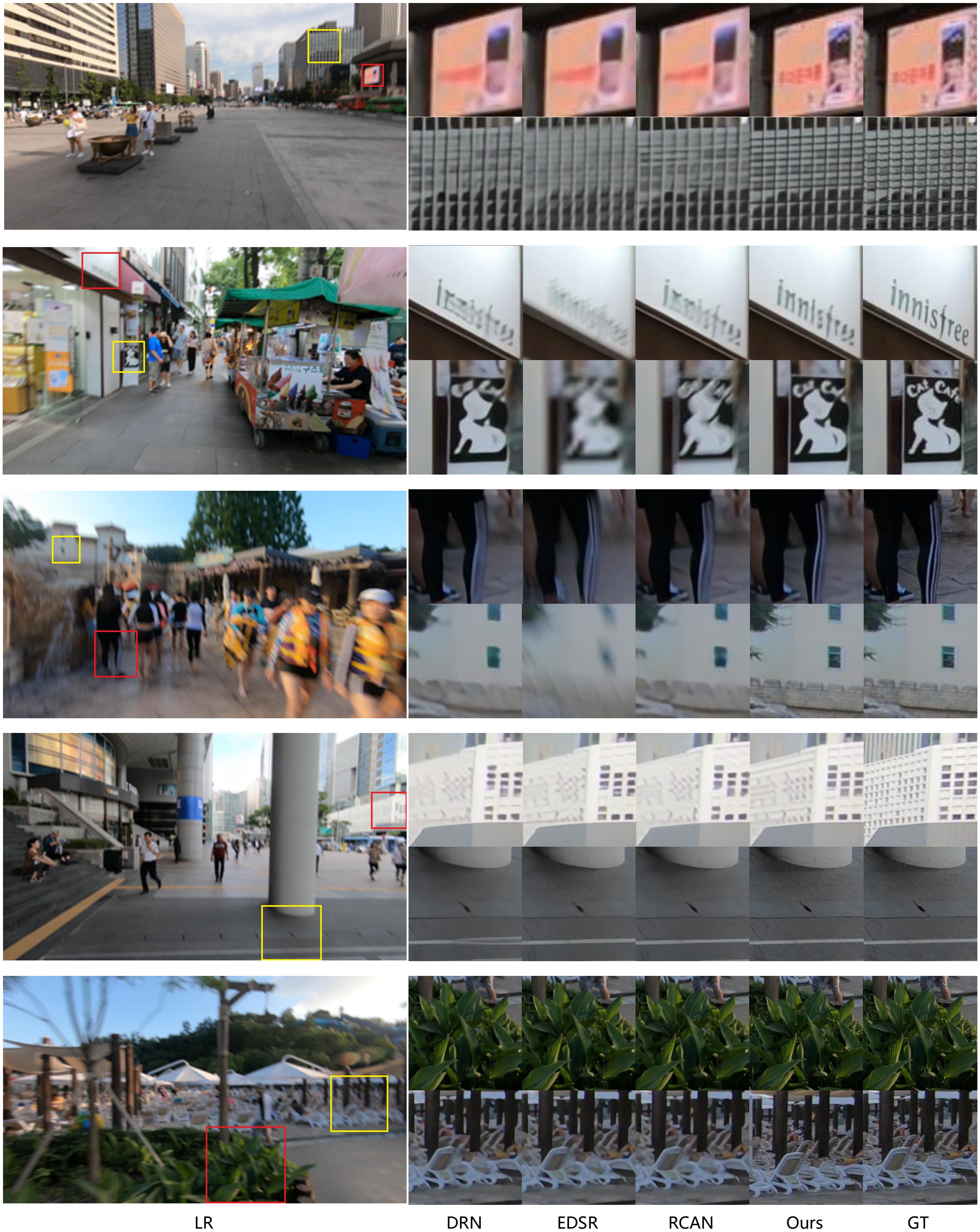

Figure 7. Qualitative results with other methods on the REDS test sets. DRN and EDSR models have limitations in restoring blurry and sharpness. RCAN achieves better results than DRN and EDSR, but sometimes the edges of objects are not sharp enough and accompanied by artifacts. Our method enhances image details by focusing on difficult pixels.

Finally, we do ablation experiments on the REDS dataset. As shown in Table 1, this table reports the PSNR improvement obtained through different modifications. Our baseline only adopts RCAN, which the patch size is set to 256. The number of filters in each layer is set to 128. Next, we employ residual spatial and channel attention modules to extract better blurry LR images' features (Model 1), which achieves 0.07dB performance improvement. We also adopt a dual-branch attention architecture and fuse the intensity level features, achieving an improvement of 0.16dB (Model 2). Finally, with our proposed loss, model 3 is 0.05dB better than Model 2, demonstrating the effectiveness of our attention dual-branch attention architecture and loss function.

4.5. NTIRE 2021 Challenge on Image Deblurring

Our method is the second place of NTIRE 2021 Challenge on track 1 Low Resolution of Image Deblurring Challenge [13]. The task of deblurring is based on a set of continuous frames of sharp video using an upsampling factor of x4. The goal of the challenge is to obtain a solution that can restore sharp results with the best fidelity (PSNR, SSIM) to the ground truth.

We applied the proposed method training with an extra dataset and achieved the second place results on track 1 as shown in Table 3. Note that we also applied geometric self-ensemble to improve performance. Then, we ensemble three models on the REDS test dataset with a novel method. Divide the image into 16 patches and calculate each area's sharpness using the auto-focusing evaluation functions without reference image. Perform a weighted average of the corresponding patches according to the sharpness to get the final image. This results in clearer areas being weighted higher, improving the overall accuracy of the image. Finally, we achieved 28.91dB, which outperforms 3rd approaches by a large margin (+0.4dB).

Team	Track 1. Low Resolution			
	PSNR	SSIM	LPIPS	Rank
VIDAR	29.04	0.8416	0.2397	1
netai(ours)	**28.91**	**0.8246**	**0.2569**	**2**
NJUST-IMAG	28.51	0.8172	0.2547	3
SRC-B	28.44	0.8158	0.2531	4
Baidu	28.44	0.8135	0.2704	5
MMM	28.42	0.8132	0.2685	6
Imagination (nju)	28.36	0.813	0.2666	7
Noah_CVlab	28.33	0.8132	0.2606	8
TeamInception	28.28	0.811	0.2651	9
ZOCS_Team	28.25	0.8108	0.2636	10
Mier	28.21	0.8109	0.2646	11
INFINITY	28.11	0.8064	0.2734	12
DMLAB	27.87	0.8009	0.283	13
RTQSA-Lab	27.78	0.7968	0.283	14
Yonsei-MCML	27.64	0.7956	0.273	15
SCUT-ZS	27.61	0.7936	0.2885	16
SNAP	27.55	0.7935	0.2785	17
Expasoft team	27.44	0.7902	0.285	18

Table 3. NTIRE 2021 Image Deblurring Challenge results on REDS test dataset. The scores were provided in [13]

5. Conclusions

This paper proposes a pixel-guided dual-branch attention network with hard pixel example mining loss (PDAN) to restore potential high-resolution images from blurred low-resolution images. PDAN uses two branches to extract the latent features effectively. Through a phased training strategy, the network learns to analyze blurry and low-resolution features. The residual spatial and channel attention module is used to extract features from inputs efficiently. The experimental results show that our method significantly improved the subjective and objective performances for joint image deblurring and SR tasks compared with the existing methods. Furthermore, our method is also the second place of NTIRE 2021 Challenge on track 1 Low Resolution of Image Deblurring Challenge. [13]

References

[1] S Derin Babacan, Rafael Molina, Minh N Do, and Aggelos K Katsaggelos. Bayesian blind deconvolution with general sparse image priors. In *European conference on computer vision*, pages 341–355. Springer, 2012. 2

[2] Benedicte Bascle, Andrew Blake, and Andrew Zisserman. Motion deblurring and super-resolution from an image sequence. In *European conference on computer vision*, pages 571–582. Springer, 1996. 3

[3] D. Chao, C. L. Chen, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, 2014. 1

[4] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian Reid, Chunhua Shen, Anton Van Den Hengel, and Qinfeng Shi. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2319–2328, 2017. 2

[5] Y. Guo, J. Chen, J. Wang, Q. Chen, J. Cao, Z. Deng, Y. Xu, and M. Tan. Closed-loop matters: Dual regression networks for single image super-resolution. *IEEE*, 2020. 1

[6] Adam Kaufman and Raanan Fattal. Deblurring using analysis-synthesis networks pair. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5811–5820, 2020. 2

[7] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018. 2

[8] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude)

faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8878–8887, 2019. 1, 2

[9] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2

[10] Lerenhan Li, Jinshan Pan, Wei-Sheng Lai, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. Dynamic scene deblurring by depth guided model. *IEEE Transactions on Image Processing*, 29:5273–5288, 2020. 2

[11] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 2, 3

[12] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 1, 5, 6

[13] Seungjun Nah, Sanghyun Son, Suyoung Lee, Radu Timofte, Kyoung Mu Lee, et al. NTIRE 2021 challenge on image deblurring. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2, 8

[14] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive separable convolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 261–270, 2017. 6

[15] Mehdi Noroozi, Paramanand Chandramouli, and Paolo Favaro. Motion deblurring in the wild. In *German conference on pattern recognition*, pages 65–77. Springer, 2017. 2

[16] Jinshan Pan, Sifei Liu, Deqing Sun, Jiawei Zhang, Yang Liu, Jimmy Ren, Zechao Li, Jinhui Tang, Huchuan Lu, Yu-Wing Tai, et al. Learning dual convolutional neural networks for low-level vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3070–3079, 2018. 2

[17] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, 2016. 2

[18] Haesol Park and Kyoung Mu Lee. Joint estimation of camera pose, depth, deblurring, and super-resolution from a blurred image sequence. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4613–4621, 2017. 3

[19] Uwe Schmidt, Carsten Rother, Sebastian Nowozin, Jeremy Jancsary, and Stefan Roth. Discriminative non-blind deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 604–611, 2013. 2

[20] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 4

[21] Abhinav Shrivastava, Abhinav Gupta, and Ross Girshick. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 761–769, 2016. 5

[22] X. Tao, H. Gao, Y. Wang, X. Shen, J. Wang, and J. Jia. Scale-recurrent network for deep image deblurring. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 1

[23] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *European conference on computer vision*, pages 157–170. Springer, 2010. 2

[24] Rui Xu, Xiaoxiao Li, Bolei Zhou, and Chen Change Loy. Deep flow-guided video inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3723–3732, 2019. 5

[25] Takuma Yamaguchi, Hisato Fukuda, Ryo Furukawa, Hiroshi Kawasaki, and Peter Sturm. Video deblurring and super-resolution technique for multiple moving objects. In *Asian conference on computer vision*, pages 127–140. Springer, 2010. 3

[26] Ruomei Yan and Ling Shao. Blind image blur estimation via deep learning. *IEEE Transactions on Image Processing*, 25(4):1910–1921, 2016. 2

[27] Xinyi Zhang, Hang Dong, Zhe Hu, Wei-Sheng Lai, Fei Wang, and Ming-Hsuan Yang. Gated fusion network for joint image deblurring and super-resolution. *arXiv preprint arXiv:1807.10806*. 3

[28] Xinyi Zhang, Fei Wang, Hang Dong, and Yu Guo. A deep encoder-decoder networks for joint deblurring and super-resolution. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1448–1452. IEEE, 2018. 3

[29] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 2, 3