Letter

Progressive Fusion Network Based on Infrared Light Field Equipment for Infrared Image Enhancement

Yong Ma, Xinya Wang, Wenjing Gao, You Du, Jun Huang, and Fan Fan

Dear Editor,

Infrared imaging, generally, of low quality, plays an important role in security surveillance and target detection. In this letter, we improve the quality of infrared images by combining both hardware and software. To this end, an infrared light field imaging enhancement system is built for the first time, including a 3×3 infrared light field imaging device, a large-scale infrared light field dataset (IRLF-WHU), and a progressive fusion network for infrared image enhancement (IR-PFNet). The proposed algorithm leverages rich angular views among the infrared light field image to explore and fuse auxiliary information for infrared image enhancement. Given an infrared light field image, multiple views are first divided into four groups according to the angle and each group contains parallax shifts along the same direction. As strong spatial-angular correlations are existing in each group, we customize a progressive pyramid deformable fusion (PPDF) module for intra-group fusion without explicit alignment. In the PPDF module, the deformation and parallax are modeled in a progressive pyramid way. To integrate the supplementary information from all directions, we further propose a recurrent attention fusion (RAF) module, which constructs attention fusion block to learn the residual recurrently and provides several intermediate results for multi-supervision. Experiments on our proposed IRLF-WHU dataset demonstrate that IR-PFNet can achieve state-of-the-art performance on different degradations, yielding satisfying results. The dataset is available at: https://github.com/wxywhu/IRLF-WHU, and the code is available at: https://github.com/wxywhu/IR-PFNet.

Infrared imaging, which could capture the thermal radiation information emitted from the objects, is an important technology in many fields, such as surveillance systems, security monitoring, and military target detection [1]−[3]. In order to improve the quality of infrared imaging, various methods have been developed for imaging systems or infrared images. Few methods are dedicated to improving the infrared detectors [4]−[6], which are usually expensive and not suitable for wide applications. Other algorithms designed to improve the infrared image quality are served as post-processing techniques. Based on the convolutional neural network (CNN), some approaches were proposed for single thermal image enhancement [7], [8]. Several methods also registered and fused multiple infrared images for quality enhancement [9], [10]. However, the former only relies on a single input, and thus, has limited performance. The latter performance depends heavily on image registration, which is itself a difficult problem. Therefore, it is desirable to combine software and hardware techniques to design an efficient system for infrared image enhancement. To address this issue, we design an infrared light field imaging equipment and a progressive fusion algorithm to enhance the quality of infrared image.

Light field imaging that records different angular views in sub-aperture images has been developed well in RGB images. The strong correlations among sub-aperture images could provide abundant supplementary information, which has proven to be beneficial for super-resolution (SR) techniques [11]−[13]. Due to this advantage, we design infrared light field imaging equipment composed of nine infrared cameras arranged in a regular grid of 3×3. As shown in Fig. 1, in a one-shot, the sub-aperture images captured by the device exhibit sub-pixel shifts in fixed directions corresponding to the central image. Thereby, we expect to use rich angular information to improve the quality of the central view. Based on this equipment, we construct an infrared light field dataset, including four scenes: building, people, car, and others. The whole dataset contains 1132 infrared light field images in a size of $3\times3\times384\times280$ pixels.

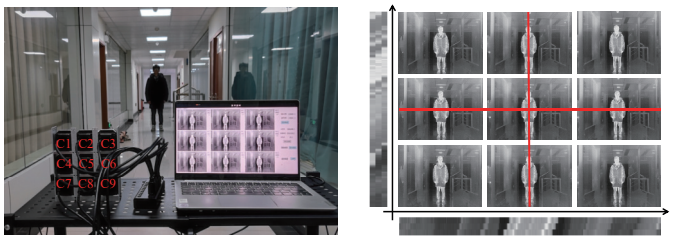

Fig. 1. Infrared light field imaging equipment. (a) The designed equipment. This device is composed of 9 infrared cameras arranged in a regular grid of 3×3. (b) The infrared sub-images. To show differences between sub-images, we extracted lines horizontally (bottom of sub-images) and vertically (left of sub-images) along the red lines for visualization.

Considering the rich angular information existing in infrared light field images, we custom a novel progressive fusion network for central infrared image enhancement, termed as IR-PFNet. As shown in Fig. 2, we divide the infrared sub-images into four groups according to the direction. Each group is fed into the PPDF module for intra-group fusion, which avoids explicit alignment. Subsequently, the intermediate features from different directions are utilized to provide auxiliary information for the central infrared image in the RAF module. In this way, our proposed IR-PFNet could make full use of rich angular information to improve the center image quality.

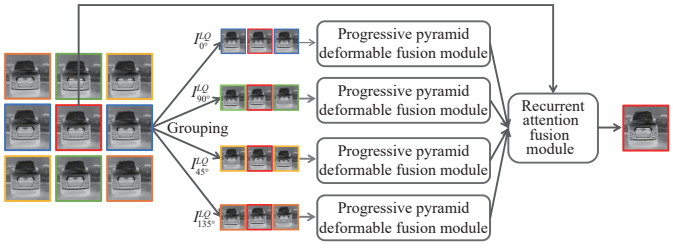

Fig. 2. The whole network of our proposed IR-PFNet.

The main contributions of this letter can be summarized as follows. 1) For the first time, an infrared light field imaging equipment is built and the infrared light field dataset is constructed, named IRLF-WHU. 2) Without image registration, a novel progressive fusion network is designed for infrared image enhancement, termed as IR-PFNet, which achieves state-of-the-art performance.

Methodology: Given the input low-quality (LQ) infrared light field image, $L^{LQ}\in\mathbb{R}^{U\times V\times H\times W}$, where U, V denote the angular resolution and H, W denote the width and height of each infrared view, the goal of our method is to restore the high-quality (HQ)

Corresponding author: Fan Fan.

Citation: Y. Ma, X. Y. Wang, W. J. Gao, Y. Du, J. Huang, and F. Fan, "Progressive fusion network based on infrared light field equipment for infrared image enhancement," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 9, pp. 1687–1690, Sept. 2022.

The authors are with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: mayong@whu.edu.cn; wangxinya@whu.edu.cn; wenjinggao@whu.edu.cn; dy0815@whu.edu.cn; junhwong@whu.edu.cn; fanfan@whu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2022.105812

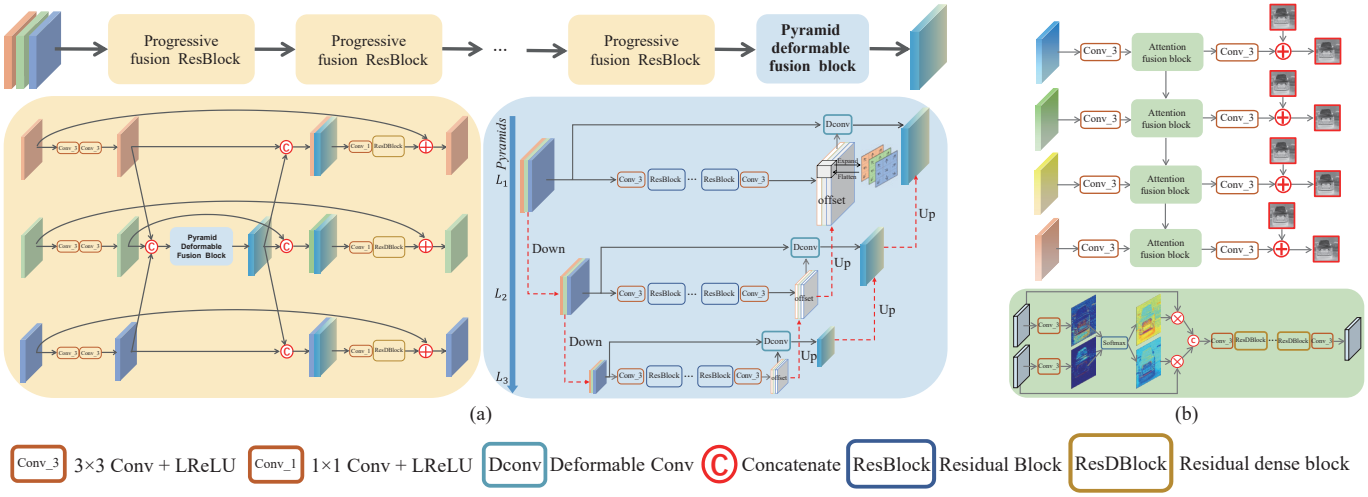

Fig. 3. The detailed architecture in our IR-PFNet. (a) PPDF module. This module consists of a series of PFRBs followed by a PDFB; (b) RAF module.

central infrared view $\hat{I}_c^{HQ}$ from reference LQ viewpoint I_c^{LQ} with the help of the whole infrared light field image. Hence, the estimated $\hat{I}_c^{HQ}$ is expected to be as close to the ground-truth HQ infrared image I_c^{HQ} as possible. Due to the strictly equidistant arrangement of the infrared cameras in the infrared light-field device, the sub-infrared images exhibit parallax in a narrow baseline along a different direction. Therefore, as illustrated in Fig. 2, the infrared light field image is divided into four groups $\{I_{0^\circ}^{LQ}, I_{45^\circ}^{LQ}, I_{90^\circ}^{LQ}, I_{135^\circ}^{LQ}\}$ according to relative angular position and each group contains three sub-infrared images. Therefore, our IR-PFNet adopts four LQ stacks as auxiliary information to predict the central HQ view

$$\hat{I}_c^{HQ} = \mathcal{F}_{IR-PFNet}(I_{0^\circ}^{LQ}, I_{45^\circ}^{LQ}, I_{90^\circ}^{LQ}, I_{135^\circ}^{LQ}; \theta) \tag{1}$$

where θ represents the parameters of our method. Note that in this letter we focus on improving the quality of the central infrared view whereas our proposed IR-PFNet could deal with each sub-infrared view because of the flexible network design.

Network design: As illustrated in Fig. 2, according to the direction, our IR-PFNet groups the input LQ infrared light field image into four stacks. Taking one stack containing sub-pixel shifts from a specific direction as input, the PPDF module is supposed to fully extract spatial information and fuse the angular information to capture intra-group spatial-parallax correlations. Hence, we employ a progressive fusion structure, where the deep features are gradually extracted and fused. For the angular information fusion, since the traditional regular convolution cannot well capture spatial deformation and angular dynamics in multiple views, we use spatial-angular deformable convolution to learn the kernel offset, avoiding explicit alignment in angular information fusion. In specific, the spatial-angular deformable convolution is deployed in a pyramid structure to model the deformation at multiple levels for obtaining the most angular-correlated information to the center view. After acquiring the supplementary information from different directions, the RAF module with attention fusion block is designed to integrate all auxiliary information for the central image in a recurrent way, which benefits our task by enabling the model to pay different attention across angles.

1) Progressive pyramid deformable fusion: As shown in Fig. 3(a), a series of progressive fusion residual blocks (PFRBs) are deployed in the PPDF module, which is supposed to make full extraction of intra-group spatial-parallax correlations. After that, the learned features from each direction are merged by a pyramid deformable fusion block (PDFB).

We show the detailed structure of PFRB in the bottom left of Fig. 3(a). It takes sub-views as input in three paths. Sharing the same insight as [14], we first conduct two 3×3 convolutional layers to extract self-independent features with N maps from each infrared view. Later, these feature maps are concatenated and merged into one part, the depth of which is $3\times N$. Then, we feed the concatenated features into the PDFB, intending to fuse features that are most angular-correlated to the center view. After that, the fused feature is further concatenated to all the previous maps and we further adopt one 1×1 convolutional layer followed by a dense residual block to extract both spatial and angular information for residual learning, respectively. In the PFRB, we share weights across paths for both convolutional layers and residual blocks.

In order to fully fuse intra-group views to obtain spatial-angular correlations, we introduce the PDFB, the detailed structure of which is shown in the bottom right of Fig. 3(a). In this block, to avoid explicit alignment, spatial-angular deformable convolution is deployed in pyramidal processing and cascading refinement for extracting the most angular-correlated information to the center view. We first introduce deformable convolution [15] tailored in this letter. Supposing the concatenated features F_a with the depth of $3\times N$ is the input, for arbitrary spatial position $\mathbf{p}$, the merged feature M_a with N maps by the deformable operation can be expressed as

$$M_a(\mathbf{p}) = \sum_{n=1}^{3}\sum_{k=1}^{K\times K} w_{n,k}\cdot F_a(\mathbf{p}+\mathbf{p}_k+\triangle\mathbf{p}_{(n,\mathbf{p}),k}),\ \ a\in\{0^\circ,45^\circ,90^\circ,135^\circ\} \tag{2}$$

where K is the size of convolution kernel, $w_{n,k}$ represents the convolutional filter and $\mathbf{p}_k$ represents the regular sampling offsets. When $K=3$, $\mathbf{p}_k\in\{(-1,-1),(-1,0),\ldots,(0,1),(1,1)\}$. Specifically, the learnable offsets $\triangle\mathbf{p}_k\in\mathbb{R}^{2K^2}$ are position-specific. That is, the individual offset $\mathbf{p}_{n,\mathbf{p}}$ will be assigned for each convolution window centered at angular-spatial position $(n,\mathbf{p})$. Thus, spatial deformations as well as angular dynamics within the concatenated features could be simultaneously modeled.

To cope with complex scenes and occlusion problems, we employ the spatial-angular deformable convolution in a pyramid way and the fusing features are cascading refined pyramidally. Specifically, as displayed with red dash lines in Fig. 3(a), to generate the merged feature M_a^l at the l-th level, $l=1,2$, we use a strided convolution with the factor 2 to downsample the features into the $(l+1)$-th pyramid level, obtaining three-level pyramids of feature representation. At the l-th level, the offsets $\triangle\mathbf{P}_a^l=\{\triangle\mathbf{p}\}$ and merged features M_a^l are also estimated with the upsampled offsets and merged features from the upper $(l+1)$-th level, respectively (red dash lines in Fig. 3(a))

$$\triangle\mathbf{P}_a^l = \mathcal{H}(F_a,(\triangle\mathbf{P}_a^{l+1})\uparrow^2) \tag{3}$$

$$M_a^l = f(Dconv(F_a^l,\triangle\mathbf{P}_a^l),(F_a^{l+1})\uparrow^2) \tag{4}$$

in which $(\cdot)\uparrow^2$ represents upsampling at the factor 2, implemented by bilinear interpolation as in [16], $Dconv$ refers to the deformable convolution described in (2), $\mathcal{H}$ denotes convolutional layers and residual blocks, and f is one convolutional layer. To reduce parameters, we share weights across three levels for both convolutional layers and residual blocks.

2) Recurrent attention fusion: After obtaining the most spatial-

angular correlated features from different directions by the PPDF module, we design the RAF module to integrate all auxiliary information for the central image. As depicted in Fig. 3(b), in the light of the complementarity of input features from different directions, a recurrent structure is adopted in the RAF module to fuse features step by step. For each step, the merged feature is first processed by one 3×3 convolutional layer. Later, we design the attention fusion block for fusing two features from different directions. Then, the output of attention fusion block is fed back to itself in the next step and simultaneously sent into one 3×3 convolutional layer to learn the residual information for the HQ infrared image.

Concretely, in the attention fusion block displayed at the bottom of Fig. 3(b), we expect this attention fusion block could benefit our task by enabling the model to pay different attention across angles. Taking two angular features as input, we adopt the 3×3 convolutional layer to generate the 2D feature map respectively. For producing the attention mask, the Softmax function is utilized on two maps along the angular dimension. Since angular attention is supposed to work as a guide to efficiently integrate angular features, we apply the attention mask on the input feature through element-wise multiplication. Then, the weighted features are concatenated and sent into deep feature extracting blocks which consist of four dense residual blocks with two 3×3 convolutional layers at the head and the tail positions.

Loss function: Since our method integrates auxiliary information in a recurrent way during the RAF module demonstrated in Fig. 3(b), we adopt a multi-supervised approach to train the whole network, which can be formulated as

$$\mathcal{L}=\frac{1}{B}\sum_{b=1}^{B}\sum_{a}\|(\hat{I}_c^{HQ})^a-I_c^{HQ}\|_1,\ \ a\in\{0^\circ,45^\circ,90^\circ,135^\circ\} \tag{5}$$

in which B is the batch size, $(\hat{I}_c^{HQ})^a$ is the enhanced central image for angular group a and I_c^{HQ} is the ground truth HQ image. With multiple supervision, the training difficulty of the network is alleviated. We take the final output as the result of quality improvement.

Experimental setup: We train our network through adaptive moment estimation (ADAM) optimizer with $\beta_1=0.9$, $\beta_2=0.999$, and $\epsilon=10^{-8}$. During training, we set the batch size as 16, and the input LQ infrared light field image is of size $3\times3\times32\times32$. The learning rate is initialized as 3×10^{-4} and then further decreased to 10^{-5} gradually. The whole training process is stopped after $240K$ iterations with an NVIDIA GTX 3090Ti GPU by Pytorch.

Previous studies on image enhancement [8], [13], [14] are rarely developed for infrared images. The lack of standard and large-scale infrared image datasets limits the development of the field. In this letter, we divide our proposed IRLF-WHU dataset into training, validation and testing sets according to Table 1. For quantitative evaluation, the original images are served as ground truth HQ images and we generate the LQ infrared images array by applying degraded operations such as reducing details, blurring, and adding noise.

Since we pioneer in designing the infrared light field device, the comparative methods only can be either single image enhancement or light field image enhancement algorithms. Thereby, we compare our proposed method with five algorithms, including two state-of-the-art deep single image enhancement methods, i.e., TherISuRNet [8], and MIRNet [17], and three current advanced deep light field image enhancement methods, i.e., resLF [13], LF-ATO [12] and DPT [11]. For a fair comparison, these methods are trained from scratch on the same dataset to achieve their best performance.

Experiments on different blur degradations: We consider three kinds of blur widely existing in the infrared image, i.e., Gaussian blur, disk blur, and motion blur. Specifically, the Gaussian blur kernel is of 7×7 size with a width of 0.6 and the radius of the disk blur is 2.5. For the motion blur, we set the angle to 10 degrees, and the motion displacement to 10 pixels. To produce LQ images, we first blur the HQ image and use bicubic downsampling and upsampling with a scale factor of 4 to reduce details. Small white Gaussian noise with standard deviation is also added to model real LQ degradation.

The quantitative results evaluated on different blur types are reported in Table 2. As can be observed, our proposed IR-PFNet achieves the best performance in terms of average peak signal to noise ratio (PSNR) and structural similarity (SSIM) for all scenes on different blur degradations. More specifically, our method outperforms the second-best method by an average of 0.2 dB on the PSNR value. We note that our IR-PFNet take angular information from different directions into consideration, unlike DPT [11] which utilizes the whole light field image as input, thereby making full extraction of spatial-angular correlations. As multiple views are grouped in resLF [13] and LF-ATO [12], the proposed IR-PFNet manages to integrate the angular information in a pyramid way, and thus further improves the quality of infrared images.

Table 1. The Dividing Way of The Proposed IRLF-WHU Dataset

Class	Training	Validation	Testing
Building	89	10	10
People	425	50	50
Car	257	32	32
Others	147	15	15
Total number	918	107	107

Table 2. Quantitative Results (PSNR/SSIM) on Different Blur Types

Method	Building	Car	People	Others
Blur type		Gaussian		
TherISuRNet	29.89/0.8274	31.75/0.8693	32.54/0.8854	32.06/0.8788
MIRNet	29.95/0.8247	31.90/0.8696	32.49/0.8841	32.06/0.8775
resLF	30.83/0.8503	32.98/0.8882	33.49/0.9000	32.75/0.8887
LF-ATO	31.32/0.8583	33.21/0.8921	33.88/0.9047	33.23/0.8967
DPT	31.40/0.8585	33.36/0.8921	34.00/0.9042	33.33/0.8968
IR-PFNet	**31.56/0.8633**	**33.70/0.8981**	**34.27/0.9099**	**33.59/0.9016**
Blur type		Disk		
TherISuRNet	29.73/0.8185	31.43/0.8609	32.11/0.8768	31.63/0.8704
MIRNet	29.89/0.8226	31.85/0.8671	32.37/0.8815	31.86/0.8749
resLF	30.27/0.8349	32.38/0.8755	32.88/0.8880	32.25/0.8774
LF-ATO	30.55/0.8415	32.65/0.8820	32.99/0.8937	32.43/0.8852
DPT	30.67/0.8413	32.83/0.8817	33.28/0.8936	32.74/0.8749
IR-PFNet	**30.94/0.8466**	**33.01/0.8861**	**33.51/0.8983**	**32.92/0.8906**
Blur type		Motion		
TherISuRNet	28.37/0.7834	30.68/0.8447	31.05/0.8565	30.69/0.8478
MIRNet	28.78/0.7910	30.98/0.8496	31.29/0.8612	30.92/0.8533
resLF	28.73/0.7922	31.32/0.8558	31.41/0.8643	30.87/0.8506
LF-ATO	29.43/0.8036	31.83/0.8652	32.02/0.8744	31.60/0.8639
DPT	29.68/0.8172	31.99/0.8666	32.21/0.8761	31.79/0.8670
IR-PFNet	**29.88/0.8221**	**32.27/0.8724**	**32.48/0.8816**	**32.01/0.8720**

Fig. 4 provides the qualitative results on three test images blurred by different degradations. Although image quality enhancement can decently reduce those artifacts, the resulting images usually become over-smoothed and lack details. It can be seen that compared to the single image enhancement methods (TherISuRNet and MIRNet), the light field image enhancement methods could produce better results with the help of auxiliary views. All in all, our IR-PFNet can restore sharper structural details and be more robust to blur degradations because of better exploration of spatial-angular information.

Experiments on different noise degradations: We also conduct experiments on the degradation of different types of noise including Salt&Pepper noise with density 0.06, additive Gaussian white noise with zero mean and standard deviation 10, Poisson noise whose level is determined by the brightness of the input image. After adding noise, we use bicubic downsampling and upsampling with a scale factor of 4 to reduce details, and blur HQ images through a Gaussian kernel with 5×5 size and 0.5 width to produce the LQ images.

Table 3 reports PSNR and SSIM results for different types of noise on the test set. As demonstrated in Table 3, our model IR-PFNet is superior in all scenes degraded by different types of noise. Due to progressive pyramid integration and recurrent attention fusion in our

Table 3. Quantitative Results (PSNR/SSIM) on Different Noise Types

Method	Building	Car	People	Others
Noise type		Gaussian		
TherISuRNet	29.71/0.8168	31.45/0.8595	32.16/0.8753	31.72/0.8693
MIRNet	29.98/0.8223	31.86/0.8661	32.40/0.8803	31.97/0.8740
resLF	30.58/0.8410	32.64/0.8822	33.18/0.8945	32.49/0.8831
LF-ATO	30.93/0.8436	32.78/0.8833	33.35/0.8961	32.79/0.8872
DPT	31.05/0.8472	32.90/0.8843	33.48/0.8965	32.91/0.8889
IR-PFNet	**31.30/0.8547**	**33.33/0.8927**	**33.88/0.9043**	**33.30/0.8969**
Noise type		Salt & Pepper		
TherISuRNet	30.79/0.8497	32.69/0.8893	33.55/0.9034	32.91/0.8970
MIRNet	30.98/0.8526	33.12/0.8950	33.79/0.9072	33.16/0.9003
resLF	31.79/0.8757	33.99/0.9088	34.60/0.9186	33.75/0.9081
LF-ATO	32.04/0.8773	34.00/0.9086	34.70/0.9193	33.97/0.9110
DPT	32.32/0.8825	34.36/0.9120	35.09/0.9224	34.34/0.9147
IR-PFNet	**32.61/0.8890**	**34.86/0.9195**	**35.52/0.9283**	**34.70/0.9202**
Noise type		Poisson		
TherISuRNet	29.65/0.8156	31.38/0.8583	32.08/0.8751	31.66/0.8688
MIRNet	29.97/0.8218	31.79/0.8650	32.37/0.8803	31.96/0.8741
resLF	30.59/0.8408	32.62/0.8823	33.18/0.8951	32.53/0.8841
LF-ATO	30.76/0.8413	32.65/0.8820	33.23/0.8952	32.71/0.8863
DPT	30.93/0.8468	32.82/0.8836	33.41/0.8962	32.90/0.8892
IR-PFNet	**31.13/0.8533**	**33.28/0.8916**	**33.78/0.9036**	**33.22/0.8959**

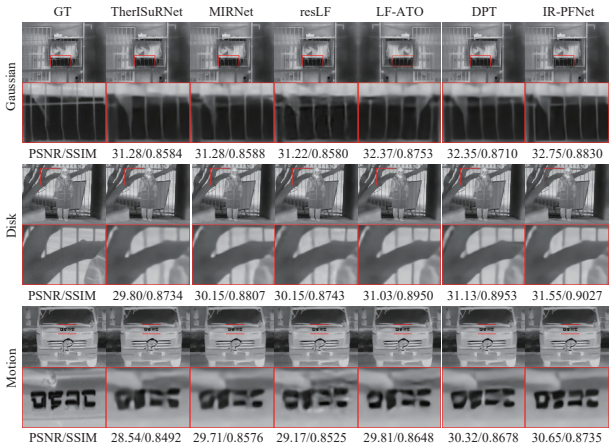

Fig. 4. Qualitative results of three blur types.

proposed model, the IR-PFNet can fully exploit spatial-temporal information, thereby leading to better enhancement results.

Fig. 5 shows the enhanced images of all comparing methods for three different types of noise. From the visual enhancement results, we can see that most other methods generate blurry or misleading structures. In contrast, our method is able to reconstruct clear and right textures closer to the ground truth ones, which demonstrates higher reconstruction quality.

Conclusion: In this work, we improve the quality of infrared image by our proposed infrared light field imaging enhancement system. We first build a 3×3 infrared light field imaging equipment and a large-scale infrared light field dataset, namely IRLF-WHU, is collected. Relying on this dataset, we tailor the PFNet to exploit spatial-angular correlations and integrate the auxiliary information for the final result. Specifically, spatial-angular correlations are fully fused by progressive fusion residual block coupling with pyramid deformable convolution to avoid registration. The attention fusion module is deployed in a recurrent manner for feature integration from all directions. Based on two fusion modules, our proposed PFNet can achieve better results compared with state-of-the-art approaches.

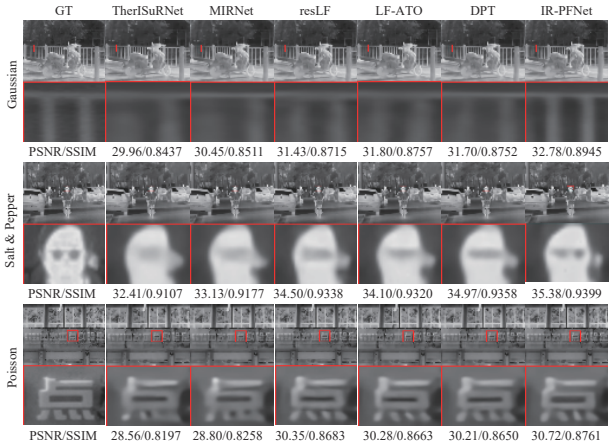

Fig. 5. Results of three noise types generated by competing methods.

References

[1] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, 2019.

[2] J. A. Sobrino, F. Del Frate, M. Drusch, J. C. Jiménez-Muñoz, P. Manunta, and A. Regan, "Review of thermal infrared applications and requirements for future high-resolution sensors," *IEEE Trans. Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2963–2972, 2016.

[3] Y. Ma, Z. Wang, H. Yang, and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: A survey," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 2, pp. 315–329, 2020.

[4] C.-C. Hsieh, C.-Y. Wu, F.-W. Jih, and T.-P. Sun, "Focal-plane-arrays and CMOS readout techniques of infrared imaging systems," *IEEE Trans. Circ. Syst. Video Technol.*, vol. 7, no. 4, pp. 594–605, 1997.

[5] J. Liang, W. Hu, Z. Ye, L. Liao, Z. Li, X. Chen, and W. Lu, "Improved performance of HgCdTe infrared detector focal plane arrays by modulating light field based on photonic crystal structure," *J. Applied Physics*, vol. 115, no. 18, p. 184504, 2014. DOI: 10.1063/1.4876227.

[6] J. Ma, W. Gao, Y. Ma, J. Huang, and F. Fan, "Learning spatialparallax prior based on array thermal camera for infrared image enhancement," *IEEE Trans. Ind. Informat*. DOI: 10.1109/TII.2021.3118682.

[7] Y. Choi, N. Kim, S. Hwang, and I. S. Kweon, "Thermal image enhancement using convolutional neural network," in *Proc. IEEE Int. Conf. Intelligent Robots and Systems*, 2016, pp. 223–230.

[8] V. Chudasama, H. Patel, K. Prajapati, K. P. Upla, R. Ramachandra, K. Raja, and C. Busch, "TherisurNet−A computationally efficient thermal image super-resolution network," in *Proc. CVPRW*, 2020, pp. 86–87.

[9] E. Mandanici, L. Tavasci, F. Corsini, and S. Gandolfi, "A multiimage super-resolution algorithm applied to thermal imagery," *Applied Geomatics*, vol. 11, no. 3, pp. 215–228, 2019.

[10] P. Cascarano, F. Corsini, S. Gandolfi, E. L. Piccolomini, E. Mandanici, L. Tavasci, and F. Zama, "Super-resolution of thermal images using an automatic total variation based method," *Remote Sensing*, vol. 12, no. 10, p. 1642, 2020. DOI: 10.3390/rs12101642.

[11] S. Wang, T. Zhou, Y. Lu, and H. Di, "Detail-preserving transformer for light field image super-resolution," in *Proc. AAAI*, 2022, pp. 2522–2530.

[12] J. Jin, J. Hou, J. Chen, and S. Kwong, "Light field spatial superresolution via deep combinatorial geometry embedding and structural consistency regularization," in *Proc. CVPR*, 2020, pp. 2260–2269.

[13] S. Zhang, Y. Lin, and H. Sheng, "Residual networks for light field image super-resolution," in *Proc. CVPR*, 2019, pp. 11046–11055.

[14] P. Yi, Z. Wang, K. Jiang, J. Jiang, and J. Ma, "Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations," in *Proc. ICCV*, 2019, pp. 3106–3115.

[15] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable convnets v2: More deformable, better results," in *Proc. CVPR*, 2019, pp. 9308–9316.

[16] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. ICCV*, 2017, pp. 764–773.

[17] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *Proc. Europ. Conf. Comput. Vis.*, 2020, pp. 492–511.