IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 26, 2020, accepted August 12, 2020, date of publication August 19, 2020, date of current version September 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017819

Deep Networks With Detail Enhancement for Infrared Image Super-Resolution

YIFAN YANG, QI LI, CHENWEI YANG, YANNIAN FU, HUAJUN FENG, ZHIHAI XU, AND YUETING CHEN

State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310000, China

Corresponding author: Qi Li (liqi@zju.edu.cn)

This work was supported by the Equipment Pre-Research Key Laboratory Fund Project under Grant 61240802214.

ABSTRACT Due to the limitation of hardware, infrared (IR) images have low-resolution (LR) and poor visual quality. Image super-resolution (SR) is a good solution to this problem. In this paper, we present a new convolution network (CNN) to improve the spatial resolution of infrared (IR) images. Our network is able to restore fine details by decomposing the input image into low-frequency and high-frequency domains. In low-frequency domains, we reconstruct image structure by deep networks. In high frequency domains, we reconstruct IR image details. Furthermore, we proposed another network to remove artifacts. Additionally, we propose a new loss function using visible (VIS) images to enhance the details of IR images. In training phase, we use VIS images to guide IR image restoration and in testing phase we get SR IR images with LR IR images input only. We optimize our deep network with a targeted function which penalizes images at different semantic levels using the corresponding terms. Besides, we build a dataset where paired LR-VIS images on the same scene are captured by a camera with both infrared and visible light sensors which both sensors have the same optical axis. Extensive experiments demonstrate that the proposed algorithm achieves superior performance and visual improvements against the state-of-the-arts.

INDEX TERMS Neural networks, infrared imaging, detail enhancement, super resolution.

I. INTRODUCTION

Infrared (IR) images provide valuable information for many applications such as thermal analysis, video surveillance, medical diagnosis, and remote sensing. The main reason for the quality and resolution degradation of an IR image is blurring effects due to non-ideal optics and finite detector size. Generally, IR images have poor quality and limited spatial resolution compared with visible (VIS) images. To achieve high-accuracy thermal measurement, infrared detectors are encapsulated in individual vacuum packages, which is a time-consuming and expensive process [1]. Given low-resolution (LR) infrared images, we focus on developing effective algorithms to restore details through solving an ill-posed inverse problem, which is essential to enable reliable target detection and recognition tasks but only available in high-resolution (HR) infrared images [2], [3].

Super-resolution (SR) method is a technique to reconstruct a HR image of a single LR image or multiple LR images [4]–[8]. And SR method is one of the best solutions to improve resolution of LR IR images. For example, Yao *et al.* [9] presented a reconstruction method of super-resolving IR images based on sparse representation. Still, there are limits to improve the resolution only by IR images.

Due to the great performances achieved by deep learning based methods, many researchers start to design deep neural networks to map LR images to HR images [10]–[14]. Although, many successful SR methods are proposed to increase resolution of VIS images and work well on IR images, but VIS images and IR images have different characters and it is not clear what is the optimal strategy to migrate a deep-learning-based method from VIS spectrum to IR [15].

Compared with VIS images, IR images have many limits. The most important one is the low Signal-to-Noise Ratio (SNR), because there are various noises in infrared images due to imperfection in infrared imaging systems and various disturbances in the environment. Besides, the contrast ratio of IR images is lower than that of VIS images, since objects and surroundings are subjected to heat exchange, heat radiation and absorption all the time, but temperature difference is not too much generally.

The associate editor coordinating the review of this manuscript and approving it for publication was Chih-Yu Hsu.

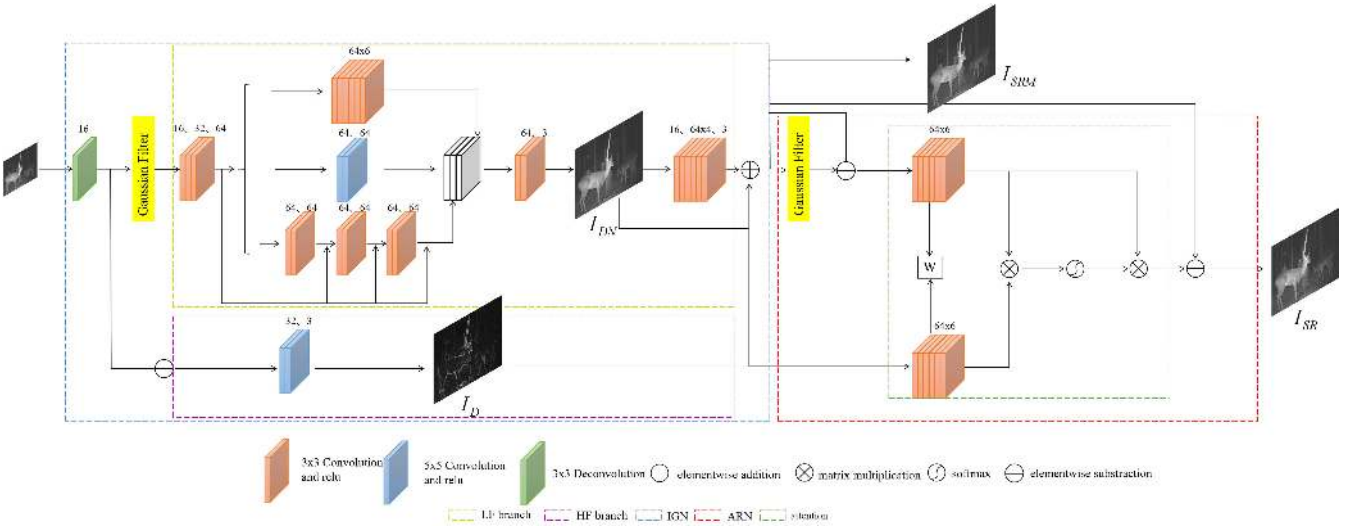


FIGURE 1. Our proposed architecture of deep networks, I_{DN} is the denoise image output, and I_D is the detail output,I_{SRM} denotes the output of IGN. I_{SR} denotes the final output. W denotes filters share the same weight. The filters number is shown on the top of every unit.

In [16], they did experiments to analyze the relationship between the objective/perceptual image quality, and their results showed that the low-frequency (LF) sub-band has a significant effect on the objective quality of the image, while the high-frequency (HF) sub-band affects the perceptual quality significantly. Owing to this concept, we designed a new convolution network (CNN) with two-branch cascaded architecture, as illustrated in Fig.1. Our network is composed by image generation network (IGN) and artifacts remove network (ARN). IGN is used to generate HR IR images. But perceptual learning strategies can maintain the visual authenticity of the generated images but with a large number of artifacts and indeterminate details. Additionally, we proposed a ARN to remove artifacts.

Considering the low SNR of IR images, we recover image structures and remove noise in the first stage for the low-frequency (LF) part. Then we recover image details with a relatively smaller receptive field. For the high-frequency(HF) part, we restore IR images details guided by VIS images as a part of our loss function. We get the HR image by adding the output of those two parts. Learning from [17] and. [18], we designed a artifacts remove network based on attention module. The correlation of HF part of HR image and LR image is calculated through attention module, and artifacts is removed by this way. Our experiments demonstrate that the proposed deep networks can achieve better performance compared with state-of-the-art SR methods.

Regarding the perceptual function, state-of-the-art approaches use different levels of features to restore the original image; this choice determines whether they focus on local information such as edges, mid-level features such as textures or high-level features corresponding to semantic information [19]–[21]. To apply it on IR images SR tasks, we propose a novel method benefited from loss designed by characters of IR images. Figure 12 shows an overview of our proposed loss function.

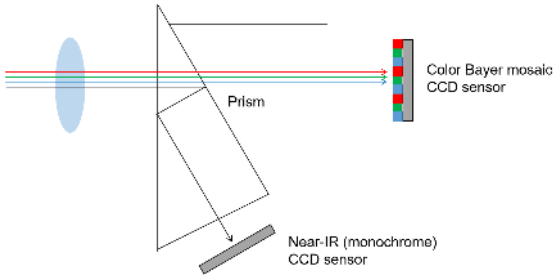


FIGURE 2. The structure of our VIS/IR camera. The color Bayer mosaic CCD sensor and Near-IR (monochrome) CCD sensor have the same optical path and divided by prism.

It is thus highly desired that we need a training dataset consisting of paired IR and VIS images. Due to the fact that existing IR and VIS dataset like CVC14 [22], IR images and VIS images are unpaired, we should register two images before inputting into network. And TNO [23], IR and VIS images are blurry and hard to be used on network training. In this paper, we aim to construct a paired and high-resolution IR/VIS dataset. Specifically, we capture images of the same scene using a camera with both IR sensor and VIS sensor, as shown in Fig.2. These two camera systems have the same optical axis, and a dichroic prism is used to split light which projects into two imaging devices. IR and VIS image pairs are collected in this way. The dataset contains various indoor and outdoor scenes, providing a good benchmark for training and evaluating IR and VIS image algorithms in practical applications.

The main idea of our article is to improve image spatial resolution of images while reconstructing rich details. Our purpose is not to reconstruct images as the same as the ground truth IR images. Considering the low-quality of ground truth IR images, we try to restore clear IR images of the real scene as shown in Fig.5. Additionally, we analyze the relationship between IR/VIS images in Sec.IV-A. Besides, we do not want

to impose much information of VIS images on IR images. We just use VIS images boundary to guide IR images restore clear details, in our loss function. VIS images are just used as a part of loss function in training phase, and in testing phase, we get the SR IR images with only IR images input. Considering the IR/VIS images pairs are difficult to acquire in real applications, our method is more effective than image fusion methods, and the images we recovered are closer to the results obtained by high-quality real infrared cameras.

Major contribution of our the proposed method includes:

- We show the relationship between IR/VIS images high-frequency (HF) and the low-frequency(LF) sub-bands, which lays an important foundation to push forward the super-resolution of IR images.
- We proposed a network designed by IR image characteristics, which uses VIS images to guide IR images to restore more high-frequency details. VIS images are just used as a part of loss function in training phase and we only input IR images in testing phase. Besides, we presented a stage network to recover the image structure as well as remove noise. The deep supervision with the loss function proposed by us improved the performance of our network.
- We built a paired IR/VIS dataset consisting of precisely aligned IR and VIS image pairs, providing a general purpose benchmark for IR super-resolution model training and evaluation.

II. RELATED WORK

A. SINGLE-IMAGE SUPER-RESOLUTION

Single image SR is an under-determined inverse problem. Most conventional CNN-based SR techniques are developed for VIS images only, they are meaningful because they can be directly applied to IR images. Classic learning based methods such as neighbor embedding(NE) [24], [25],anchored neighborhood regression(ANR) [26], sparse coding(SCSR) [27] attempt to constrain the solution space with prior information. Timofte *et al.* [26] utilized a number of linear regressors to anchor the neighborhood embedding of a LR patch the nearest atom in the dictionary and to pre-compute the corresponding embedding matrix. Yang *et al.* [27] assumed that LR patched share the same sparse representation with corresponding HR counterparts. Then the LR dictionary are passed to corresponding HR dictionary for HR patches reconstruction. Besides Glasner *et al.* [2] exploited the self-similarity prior that patches in a natural image tend to recur within and across scales of the same image. Although self-similarity based approaches do not require a training process, they involve time-consuming internal patch searching processes.

In recent years, deep learning has been successfully applied in various computer vision tasks (e.g., object classification [28], pedestrian detection [29], and image de-nosing [30]) and achieves breakthrough improvements. Dong *et al.* [12] proposed SRCNN, applied CNN technique to SR for the first time. SRCNN directly learned an end-to-end mapping between LR and HR images represented as a deep CNN that takes the LR images as the input and outputs the HR images. The same author also developed a fast version(FSRCNN) to accelerate SRCNN [31] and achieved a real-time speed. In 2015, Kim *et al.* [32] introduced a very deep CNN-based SR(VDSR) with deep network structure by employing visual geometry group(VGG) network. And they used residual-learning and high learning rates to optimize the network, applied gradient clipping to ensure training stability. As a result, they achieved the best result on network at that time. But VDSR contains a large number of model parameters which are impractical for real-time implementation. The efficient sub-pixel convolution layer was proposed by ESPCN [33] to upscale the final LR feature maps into HR output to solve the problem of over-smooths and blurs in the original LR image. But these methods do not consider images of different frequencies separately and tend to produce a blurred result.

B. MULTI-IMAGE SUPER-RESOLUTION

The first multi-image super-resolution work was proposed by Tsai [34], they used a frequency-domain technique to combine multiple under-sampled images with sub-pixel displacements to improve the spatial resolution. Farsiu *et al.* [35] presented an algorithm to enhance the quality of a set of noisy blurred images and produce a HR image with less noise and blur effects. Their method removes outliers efficiently, resulting in images with sharp edges. Kawulok *et al.* [36] proposed a method which can be highly optimized to benefit from parallel processing performed, and their method is sufficient for some real-time applications. Molini *et al.* [37] proposed a novel framework integrates the spatial registration task directly inside the CNN, and allows one to exploit the representation learning capabilities of the network to enhance registration accuracy. And the whole network can be trained end-to-end to recover a single high-resolution image from multiple unregistered low-resolution images.

C. OBJECTIVE FUNCTION

Despite variant architectures proposed for the SISR task,the behavior of optimization-based methods is principally driven by the choice of the objective function. The objective functions used by these works mostly contain a loss term with the pixel-wise distance between the super-resolved and the ground-truth HR images. However, using this function alone leads to blurry and over-smoothed super-resolved images due to the pixel-wise average of all plausible solutions.

Based on the idea of perceptual similarity [38], Johnson *et al.* [39] proposed perceptual loss to minimize the error in feature space. After that, a number of papers have used this optimization to generate images [40]–[42]. Similarly, contextual loss [43] is proposed to generate images with natural image statistics, which focuses on the feature distribution rather than merely comparing the appearance. Although these works generate near-photorealistic results, they can not be used on IR image super-resolution because they do not take the characters of IR image into consideration and could not improve the visual quality.

III. IR/VIS DATASET

As we stated earlier, existing IR/VIS datasets do not contain high-resolution IR and VIS image pairs, and we should use additional program to pair IR and VIS images to improve image details, which will cause other problems.

To build a dataset of real-world IR/VIS models, we proposed to collect images by multi-spectral prism camera with two CCDs. This camera can simultaneously capture visible and IR images through the same optical path by two CCDs with 1296×966 active pixels per channel. Figure 2 shows the basic structure of our camera, the color Bayer mosaic CCD sensor and IR (monochrome) CCD sensor have the same optical axis, using prism to split light which projects into two imaging sensors.

We use aperture priority mode and adjust aperture according to the depth-of-field (DoF). Basically, we select enough aperture value to make DoF cover the scene and avoid severe diffraction. The white balance is set to automatic mode. As the VIS and IR imaging sensors share the same optical axis, we can acquire both VIS and IR images of the same scene at the same time.

To ensure the generality of our dataset, we take photos in both indoor and outdoor environments. For each scene, we record VIS and IR images of the same scene for a period, and select two best paired images as our dataset. Finally, we select 100 VIS and IR image pairs as our dataset.

IV. PROPOSED METHOD

In this section, we first describe the relationship between IR/VIS images detail, and additionally, we described the low-frequency and high-frequency sub-bands of IR images, analyzing the importance of those sub-bands in super-resolution tasks. Finally we proposed an IR image super-resolution model architecture and suggested the network we used and the loss function design.

A. RELATIONSHIP BETWEEN IR/VIS IMAGE DETAIL

VIS images contain the sunlight reflection information of scenes. It is characterized by clear details and rich color information, which is more conducive to the visual observation of the human eyes. However, it is also easy to be affected by environments during the imaging process. The working principle of infrared sensor is to obtain infrared images formed by different thermal infrared rays by infrared difference between itself and the background. Although the position and shape of the target can be roughly sketched, the details of the target cannot be clearly expressed. Due to its special imaging principle and special use environment, obtained images have poor visual effect, and the resolution of images is obviously lower than that of visible light. Therefore, in IR image super-resolution tasks, we propose to use VIS images to guide the restoration of IR images. Additionally, we improve the detail performance of image SR task which affect the perceptual quality significantly.

Comparing with IR images, VIS images contain more details which means the boundary of VIS images is more obvious than IR images. As shown in Fig.3, VIS image boundary contains more information than IR image. That is, we can reconstruct rich detail information when we restore IR images guided by VIS images.

TABLE 1. The SSIM and correlation coefficient between IR and VIS images boundary in our dataset.

SSIM	correlation coefficient
0.7773	0.6289

Our IR/VIS dataset was taken based on the same scene, and hence the boundary of VIS images has high correlation with boundary of IR images. To prove this, we did a experiment by calculating SSIM [44] of images boundary and the correlation coefficient of their frequency spectrum.

SSIM [44] is a measure of the similarity between two images. This indicator was first proposed by the Laboratory for Image and Video Engineering at the University of Texas at Austin. Structural similarity ranges from 0 to 1. When the two images are exactly the same, the value of SSIM is equal to 1.

Correlation coefficient is a statistical index used to reflect the closeness of the correlation between variables. It is calculated by the product difference method, which is also based on the deviation of the two variables and their respective averages, and the degree of correlation between these two variables is reflected by multiplying the two deviations, focusing on the linear single correlation coefficient. It is conducted by

$$r = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\delta_x}\sqrt{\delta_y}}, \tag{1}$$

where X, Y denote two images and δ denotes the variance of image. And r range from -1 to 1.

We use SSIM [44] to measure the similarity between IR and VIS images boundary and use correlation coefficient to represent the correlation of its frequency spectrum. We calculate SSIM [44] and correlation coefficient of dataset we proposed in Sec.II-A. The result is shown in Tab.1, we can find that there is a strong correlation between IR and VIS images boundary. As a result, we can use VIS image to guide the restoration of IR image super-resolution without causing too much error.

B. LOW-FREQUENCY AND HIGH-FREQUENCY SUB-BANDS OF IR IMAGES

For an image, low-frequency component means that the color changes slowly. that is, the grayscale changes slowly, which means it is a continuously gradual area. Generally, the content inside the edge is most of the information of the image which is low frequency. That is the general overview of the image and contour.

The high-frequency component corresponds to the part where the image changes drastically, such as edge, noise and details of the image.

Figure 4 shows a HF component and LF component of one image from CVC14 [22]. The middle image is the LF

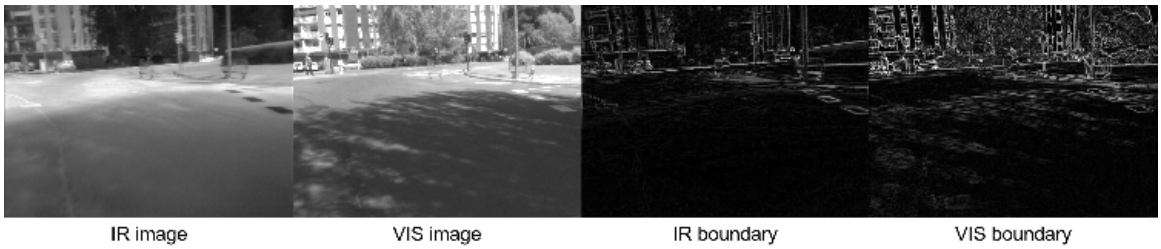


FIGURE 3. The boundary comparisons between IR and VIS image. The IR image, VIS image, IR boundary image and VIS boundary image are arranged from left to right.

FIGURE 4. Comparison of the high-frequecy (HF) and low-frequency (LF) part of infrared (IR) image. The left image is original IR image, and the middle image is LF part of IR image extracted by gaussian filters with 30×30 kernel and σ equal to 8. The right image is the correspond HF part by the same gaussian filter.

component of this image extracted by gaussian filers with kernel size 30×30 and $\sigma = 8$. The HF component which is shown in the right line on the figure shows more detail information with much noise.

Considering this, we designed a two-branch cascaded neural network. In the first branch, we recover the LF sub-bands of IR image to suppress noise, and in the subsequent stage, we reconstruct part of detail information. Generally, we will lose much detail information when we suppress noise, and details play an important role in super-resolution tasks. In the second branch, we extract HF sub-bands of image and use VIS image to guide its reconstruction, And then the restored HF image is added into the final stage pixel-by-pixel to preserve information integrity and reduce noise components in image.

C. PROPOSED NETWORK

In this paper we present a cascaded architecture of deep networks to address the challenging problem of infrared image SR. Figure1 shows the overall structure of the proposed algorithm. Our algorithm is composed by image generation network (IGN) and artifacts remove network (ARN).

1) IGN

Our IGN consists of three steps: First, we extract LF information and use it to restore image structure I_{DN}. We remove much noise of IR images by this step. Then we use some convolution blocks to reconstruct detail information of IR image. Futhermore, HF information of IR image I_D is reconstructed by VIS image. The reconstructed HF part is added onto LF branch and outputs the super-resolution image I_{SR}. Note that we train the entire network jointly, end-to-end.

As shown in Fig.1, we first use a de-convolution block to increase image size by

$$D_I = DConv_3(x), \tag{2}$$

where D_I denotes the de-convolution image, $DConv_k(\cdot)$ denotes de-convolution with kernel $k \times k$. The image is decomposed into HF and LF part by a gaussian filter with kernel size 30×30 and $\sigma = 8$.

Early layers of a network return low-level spatial information regarding local relations, such as information about edges. Hence we use two convolution blocks to extract detail of images. One convolution block contains a convolution and relu [45] as illustrated in Fig.1. The reconstructed detail image can be expressed by

$$I_D = Conv_5^1(Conv_5^1(D_I^{HF})), \tag{3}$$

where D_I^{HF} denotes HF part of D_I, $Conv_k^n$ means n stacked convolution layers with kernel $k \times k$ and relu.

For denoise images in the middle network, we consider using three parallel lines to extract different features of image. In the parallel network, feature extraction is realized by passing the state of preceding layer to six convolution blocks with 3×3 kernels. In the top line, we use small amount of convolution blocks with 5×5 kernel to repair part of image details, preserving image information completeness. In the second line, we just use two convolutions to extract underlying image information. Additionally, we use global residual learning to fuse multi-layer features. Those module can be expressed by

$$I_{DN} = concat\{F_H, F_L, F_G\}, \tag{4}$$

where F_H is high-level features extracted by six convolution blocks, F_L is low-level features extracted by two convolution blocks and F_G represent the global residual learning mechanism.

We concatenate those three parallel lines to make use of different information of image to rebuild I_{DN}. Further, we use six convolution blocks and add the HF part to get the final output of our network by

$$I_{SRM} = Conv_3^6(I_{DN}) + I_D. \tag{5}$$

2) ARN

It is generally known that the loss designed for perceptual improvement can push the generator to generate results

keeping in line with the true manifold but also add less meaningful high-frequency artifacts that is irrelevant to the input image. To address this issue, we design an effective subnetwork ARN to extract edge features that are useful for identification, so that the network can focus on the real edge information to achieve the purpose of removing artifacts.

As shown in Fig.1, the ARN module takes I_{SRM} and I_D obtained from IGN as input. We use Gaussian filter to detect and extract the edges of I_{SRM}, then six convolution blocks with 3×3 kernels are used to extract the feature by

$$F_{SRM} = Conv_3^6(I_{SRM} - G(I_{SRM})), \tag{6}$$

where F_{SRM} is the features of I_{SRM} edge. $G(\cdot)$ denotes gaussian procedure.

The features of I_D are extracted by the same way without gaussian filter, besides the filters share the same wight. We get the output image through

$$I_{SR} = I_{SRM} + F_{SRM} \otimes (softmax(F_{SRM} \otimes F_D)), \tag{7}$$

D. PROPOSED NETWORK

E. LOSS FUNCTION

The state-of-the-art approaches such as [20] and [21] estimate perceptual similarity by comparing the ground-truth and the predicted super-resolved image in a deep feature domain by mapping both HR and SR images into a feature space using a pre-trained classification network. The output of a specific convolution layer is used as the feature map.

Let x be the input LR image and θ be the set of network parameters to be optimized. General super-resolution tasks is to learn a mapping function f for generating a high-resolution image $I_{SR} = f(x; \boldsymbol{\theta})$ closed to the ground truth HR image y. But the ground truth of IR images is blurred, as a result, we propose to design a network to learn a mapping function close to IR image convoluted by matrix shown in Fig.12. And we present to use VIS images information to generate IR images with more details. We do not want to introduce additional information of VIS images on IR images, which is different from images fusion methods. As stated previously, IR images generally contain less detail information compared with VIS images and early layers of a CNN return low-level spatial information regarding local relations, such as information about edges and blobs. As shown in Fig. 12, we use VIS image to enhance IR image detail by estimating the feature distance of an early CNN layer between I_D and VIS images, which focuses more on low-level spatial information.Besides, VIS images is just used in our training phase as a part of loss function, in testing phase, we get SR IR images with LR IR images input only. Mid-level features are mostly representing textures and high-level features amount to the global semantic meaning. We compute mid-level CNN features to estimate the perceptual similarity between IR and SR images. Considering that there are not many high-frequency information components in the IR images, we use a filter shown on Fig. 12 to convolute HR IR image and get IR image with clear edges, then we compute mix-level CNN features to estimate the similarity between them.

The overall loss function is given as:

$$\mathcal{L} = \alpha \cdot \mathcal{G}_s\left(I_{SR}, F(I_{IR})\right) + \beta \cdot \mathcal{G}_b\left(I_D, I_{VIS}\right) + \gamma \cdot \mathcal{G}_l\left(I_{DN}, I_{IR}\right), \tag{8}$$

where α, β and γ are the corresponding weights of the loss terms used for the overall, detail and LF loss. $F(\cdot)$ denotes the filter operator. $\mathcal{G}_s$, $\mathcal{G}_b$ and $\mathcal{G}_l$ are the functions to calculate different feature space distances between two given images. In our experiments, we use $L = l_1$ norm for those three functions, because l_1 loss function does not over-penalize larger errors, and proved to be more powerful for performance and convergence [46].

By this way, we remove noise in denoise network and noise cannot be reproduced in following steps, the network can provide more image details of image and better visual effects.

V. EXPERIMENTAL RESULTS

In this section, first, we describe the training parameters and dataset in details, then we evaluate our proposed method in terms of qualitative and quantitative analysis.

A. IMPLEMENTATION AND TRAINING DETAILS

In our proposed method, the size and number of filters are shown in Fig.1. We initialize the convolution filters using the method of He *et al.* [47]. All the convolutional layers are followed by ReLU [45] and we pad zeros around the boundaries before applying convolution to keep the size of all features maps the same as the output of de-convolutional layers.

We first use CVC14 [22] as our training set for 50 epochs and use IR images as the detail restoration network ground truth. Then we use our dataset proposed in Sec.II-A, and use VIS images as the detail restoration network ground truth for 100 epochs. In each training batch, we randomly sample 64 patches with the size of 128×128. An epoch has 1,000 iterations of back-propagation. We augment the training data in three ways: Scaling, Rotation and Flipping. Following the protocol of existing methods [12], we generate the LR training patches using the bicubic downsampling. Draw on the experience of [39],the weights of each term in our loss function α, β and γ were set to 100, 20 and 50. The Adam optimizer [48] was used during both steps. The learning rate was set to 1×10^{-3} and then decayed by a factor of 10 every 30 epochs.

B. IMAGE QUALITY ASSESSMENT METRICS

General image quality assessment metrics like PSNR and SSIM [49] are not very well matched to perceived visual quality. If we use those methods as image quality assessment metrics of our results, we should reconstruct images as the same as ground truth. But as shown in Fig.5, our results are better than ground truth images perceptually. Besides our method tries to produce an image with a clear image edge

TABLE 2. Results obtained on 2× scaling factor for quantitative comparison with the current state-of-the-art methods.

Metric	Bicubic	SRCNN	VDSR	LapSRN	IDN	hr	Ours
Entropy	7.3496	7.3552	7.3577	7.3516	7.3545	7.3625	**7.3982**
Brenner	22.4665	29.1125	29.1328	29.0204	29.1528	31.4152	**143.2546**
SMD2	21.6517	29.7028	29.8714	29.1125	29.9224	40.6292	**352.1785**
Tenengrad	12.8245	14.3598	14.3357	14.4649	14.5025	14.9205	**27.2537**
NRSS	0.1408	0.2065	0.2133	0.2082	0.2186	0.2875	**0.8542**

TABLE 3. Results obtained on 4× scaling factor for quantitative comparison with the current state-of-the-art methods.

Metric	Bicubic	SRCNN	VDSR	LapSRN	IDN	hr	Ours
Entropy	7.3477	7.3564	7.3759	7.3545	7.3574	7.3625	**7.4211**
Brenner	13.3105	19.7165	22.0568	22.1700	22.2274	31.4152	**153.6875**
SMD2	11.8575	19.2680	22.3724	20.4395	21.0987	40.6292	**332.1795**
Tenengrad	9.8868	11.9927	12.6636	12.7060	12.7529	14.9205	**33.3584**
NRSS	0.0401	0.0662	0.0717	0.0714	0.0915	0.2875	**0.7325**

which cause pixel difference with ground truth, considering IR image is generally blurred, which means the SSIM of our method is lower than other method.

Entropy is a concept which originally arose from the study of the physics of heat engines. It can be described as a measure of the amount of disorder in a system. In the case of an image, these states correspond to the gray levels which the individual pixels can adopt. As the entropy of the image is decreased, so is its information content, which means the image with high entropy can express rich details.

In the spatial domain, the grayscale difference between adjacent pixels of images is large, that is, the edges are sharp and the gradient is large. The Brenner metric [50] is a form of a spatial domain derivative:

$$F_{\text{Brenner}} = \sum_i \sum_j \left(G_{ij} - G_{i+2j}\right)^2 \| i < N_x, \quad j+2 < N_y, \quad (9)$$

where G_{ij} is the grayscale intensity at pixel position ij, N_x and N_y are the image width and height. It has been shown to be a very robust autofocus metric [51].

Reference [52] proposed a new evaluation function named Sum Modulus Difference 2 (SMD2) based on gray scale difference using product of adjacent pixels difference in the horizontal and vertical direction as core function by

$$\mathrm{D(f)} = \sum_{\mathrm{y}} \sum_{\mathrm{x}} |\mathrm{f(x, y)} - \mathrm{f(x+1, y)}| * |\mathrm{f(x, y)} - \mathrm{f(x, y+1)}|, \quad (10)$$

where $f(x, y)$ denotes the pixel value of point (x, y) and $D(x, y)$ denotes result value.

As introduced in [53], Tenengrad gradient function uses Sobel operator to extract gradient values in horizontal and vertical directions by

$$\mathrm{D(f)} = \Sigma_{\mathrm{y}} \Sigma_{\mathrm{x}} |G(\mathrm{x}, \mathrm{y})| \quad (\mathrm{G(x, y)} > \mathrm{T}), \quad (11)$$

$$G(x, y) = \sqrt{G_x^2(x, y) + G_y^2(x, y)}, \quad (12)$$

where is the detection threshold, G_x and G_y is gradient values in horizontal and vertical directions.

Reference [54] introduced a novel no-reference image quality assessment index called No-Reference Structured Sharpness (NRSS) for quality evaluation of blurred images. This method constructed a reference image by a low-pass filter, and assessed the image quality by computing the SSIM between the original image and the reference one, thus considering the mathematical model of imaging system as well as the advantages of SSIM. NRSS is calculated by

$$\mathrm{NRSS} = 1 - \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{SSIM}\left(\mathrm{x_i}, \mathrm{y_i}\right), \quad (13)$$

where x_i and y_i is different image patch of image x and y.

In our article, we use image quality assessment above-mentioned as our image quality evaluation metrics.

C. QUALITATIVE RESULTS

We compare the proposed method with 4 state-of-the-art SR algorithm: SRCNN [12], VDSR [32], LapSRN [55], IDN [56]. The source codes of those methods are provided by their authors. All of these methods are trained using the same dataset to ensure fair comparison. We carry out extensive experiments using real IR dataset: CVC14 [22] and our dataset. CVC14 [22] is composed by two sets of sequences. These sequences are named as the day and night sets, which refers to the moment of the day they were acquired, and Visible and IR depending the camera that was user to record the sequences. For training 3695 images during the day, and 3390 images during night, with around 1500 mandatory pedestrian annotated for each sequence. For testing around 700 images for both sequences with around 2000 pedestrian during day, and around 1500 pedestrian during night.

Figure5 and 6 shows the output of our model, along with the outputs of other state-of-the-art models for visual comparison. It can be observed that the up-sampled image generated by our model contains more details and is visually more similar to the ground truth than other images. It is observed that both methods output blurry SR results except ours. That

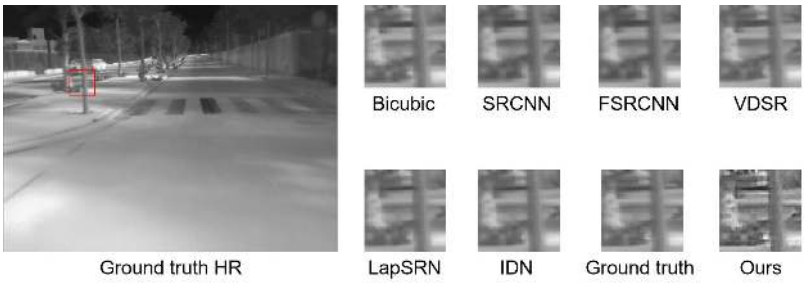


FIGURE 5. Visualisation of the upsampled images for qualitative comparison with the current state-of-the-art methods on 2× scaling factor.

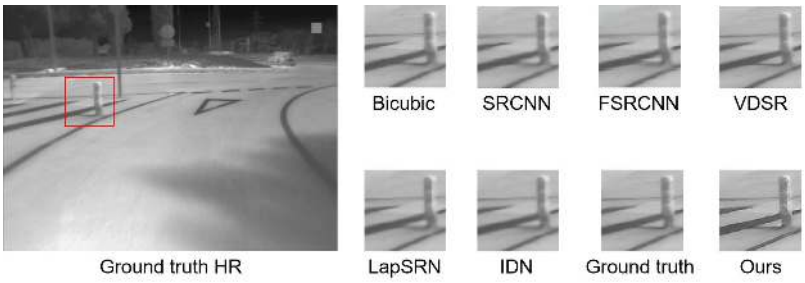


FIGURE 6. Visualisation of the upsampled images for qualitative comparison with the current state-of-the-art methods on 4× scaling factor.

is mainly because the ground truth of IR image is usually blurred and those SR methods cannot exceed this limit.

To quantitatively compare the performance of our model with other state-of-the-art models, we computed the average of image quality evaluation metrics values stated in Sec.V-B between all the predicted images and their corresponding ground-truths.

Table 2 and 3 shows the results on CVC14 [22] with different scaling factors, obtained by current state-of-the-art methods and ours. It is evident that our method outperforms all the cited methods. On average, our method outperforms other state-of-the-art SR methods by large margins. Moreover, the performance of our method is very stable and it achieves the best SR results.

To prove our method is effective on real images, we evaluate our data generation method as well as the proposed network on real captured images. As shown in Fig.8 and 7, those methods cannot generate clear images. By contrast, we achieve better results with sharper edges and finer details, which demonstrates the effectiveness of our method.

D. USER STUDY

Drawing on the user study method of [57],we performed a user study to compare the reconstruction quality of different approaches to see which images are more appealing to users. Five methods were used in the study: SRCNN [12], VDSR [32], LapSRN [55], IDN [56] and ours. During the experiment, super-resolution images reconstructed by the mentioned approaches were shown to each user. Users were requested to vote for the most appealing images. All images

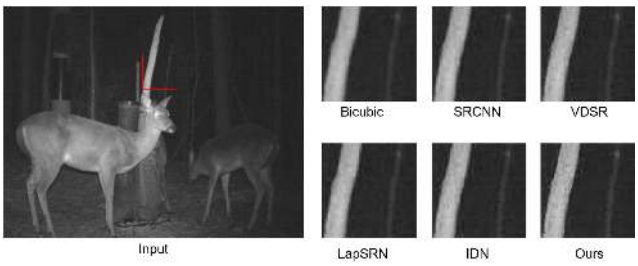


FIGURE 7. Results on real images with 4× scaling factor.

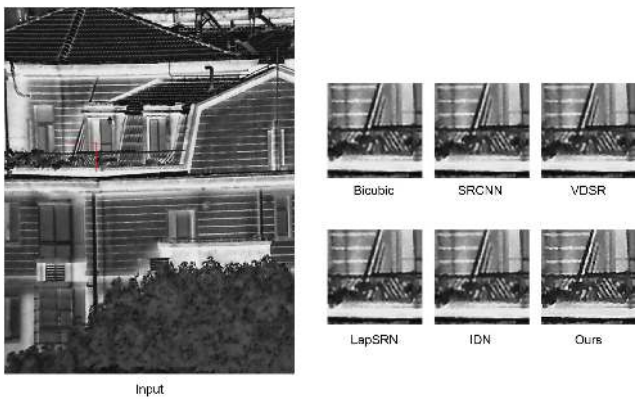


FIGURE 8. Results on real images with 2× scaling factor.

were presented in a randomized fashion to each person. In order to maximize the number of participants, we created our online assessment tool for this purpose. In total, 56 per-

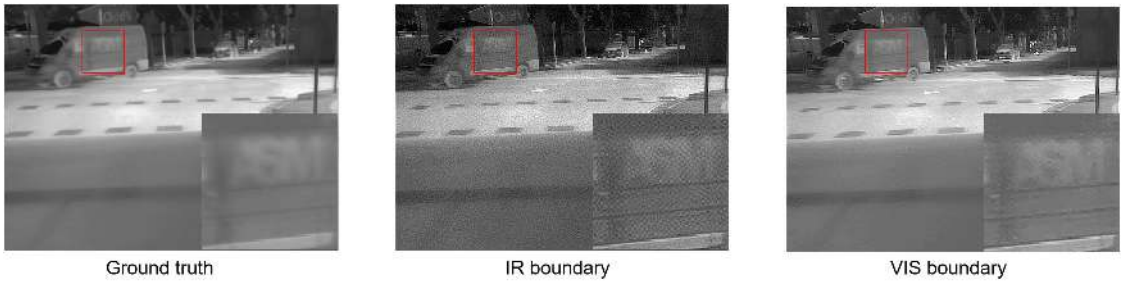


FIGURE 9. Comparison on our method trained with IR boundary and VIS boundary.

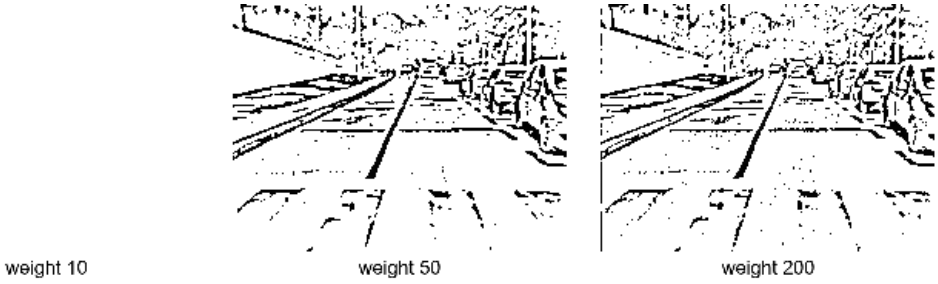


FIGURE 10. Visualisation of the boundary output images with different weight, we reverse the background from 0 to 255.

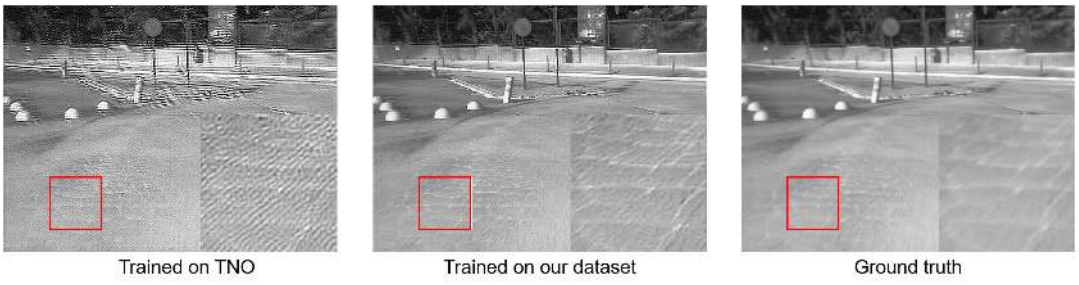


FIGURE 11. SR results on our model trained on different datasets.

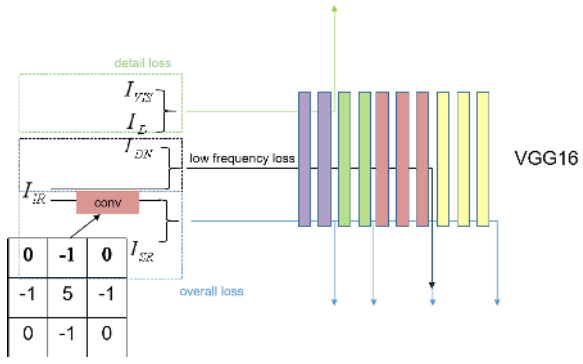


FIGURE 12. Loss function overview. I_{VIS} denotes VIS image, I_{SR} denotes IR image, I_D is the detail output of our network, I_{DN} is the denoise output, *conv* in this figure is used to convolute IR image. We use a pretrained VGG16 network for image classification to measure perceptual differences in detail, low frequency and overall part between images. The loss network remains fixed during the training process.

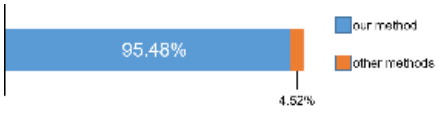


FIGURE 13. The results of the user study. Our method produces visual results that are the preferred choice for the users by a large margin in terms of percentage of votes.

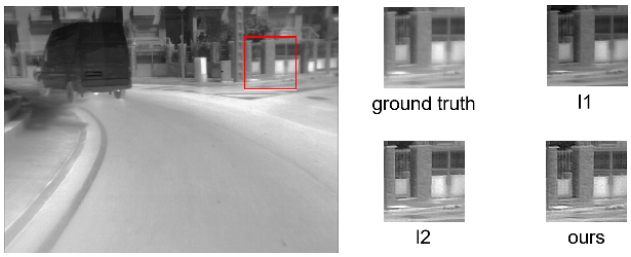


FIGURE 14. SR results on our model trained on different loss function.

sons participated in the survey. Figure 13 illustrates that the images reconstructed by our approach are more appealing to the users by a large margin. These results confirm that our approach reconstructs visually more convincing images compared to mentioned methods for the users.

E. ABLATION STUDY

1) DATASET

To demonstrate the advantages of our dataset, we conduct experiments to compare the super-resolution performance of our model trained on TNO [23] and our dataset and we use

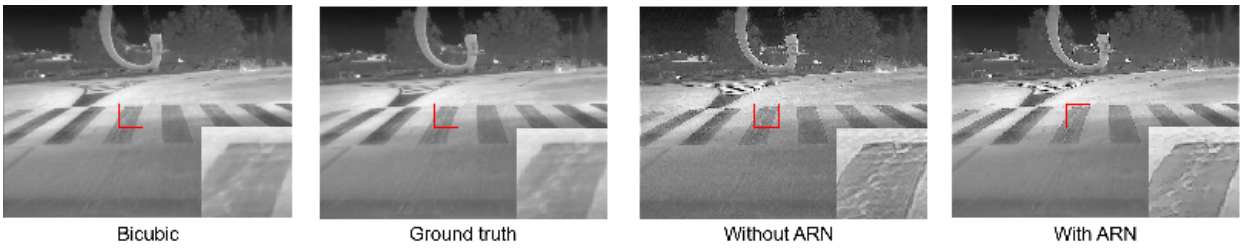


FIGURE 15. SR results on our model trained with/without ARN.

CVC14 [22] as test dataset for fair comparison. The result is shown in Fig.11. One can see that, our model trained on TNO dataset cannot generate clear images because IR and VIS images of this dataset is blurry and hard to be used on network training. Training on our dataset brings much improvement over ground truth.

2) LOSS FUNCTION

To validate the effectiveness of our loss function, we calculate the loss of the SR images and HR images and train the proposed network with l_1 and l_2 loss function. As illustrated in Fig.14, SR images reconstructed by our full model contain relatively clean and sharp details. Both methods trained by our network show better results compared with ground truth, which demonstrate the effectiveness of our model.

3) EFFECTIVENESS OF THE VIS IMAGES

To analyse the efficacy of VIS images, we trained the network after replacing the VIS images I_{VIS} with IR images. As shown in Fig.9, the image trained with IR images generates much noise, result from the inherent noise properties of IR images. As a comparison, the image trained with VIS images suppressed noise greatly and generated clear image edge (e.g. character A in Fig.9)

4) ANALYSIS OF THE WEIGHT OF DETAIL LOSS

As stated in Sec.IV-E, our loss function is composed of three part: the overall, detail and LF loss. The overall loss is a data fidelity term to make sure the generated image is similar to the original image. The detail part is to generate more fine details of the VIS image. If the weight of the detail loss is too large, it will introduce noise. Contrarily, if the weight of detail loss is too small, it will be useless.

Figure.10 shows the detail output images with different weights. In order to display the detail of images better, we reverse the background from 0 to 255. As shown in Fig.10, the detail images with weight 20 cannot predict images. Images with weight 200 will introduce much noise on the road. Images with weight 50 can predict image edge accurately and suppress noise.

5) ARN

For a analysis on proposed ARN, we remove the ARN module and re-train the network. We use the output of module IGN as the final super-resolution output. As shown in Fig.15, our method trained with ARN could generate image with rich details while less artifacts and noise.

VI. CONCLUSION

We proposed a new network to generate clear IR images guided by VIS images. In addition, we constructed paired IR/VIS images dataset to ensure accurate pixel-wise alignment between image pairs. The proposed algorithm compares favorably against state-of-the-art methods both quantitatively and qualitatively, and more importantly, our method breaks the limitation caused by infrared image blur. In the future, we will enlarge the dataset by collecting more image pairs with more types of cameras, and investigate new SR model training strategies on it.

REFERENCES

[1] A. Rogalski, P. Martyniuk, and M. Kopytko, "Challenges of small-pixel infrared detectors: A review," *Rep. Prog. Phys.*, vol. 79, no. 4, Apr. 2016, Art. no. 046501.
[2] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 349–356.
[3] K. Zeng, J. Yu, R. Wang, C. Li, and D. Tao, "Coupled deep autoencoder for single image super-resolution," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 27–37, Jan. 2017.
[4] J. B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. Comput. Vis. Pattern Recognit.*, 2015, pp. 5197–5206.
[5] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.
[6] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.
[7] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
[8] W. Sun and Z. Chen, "Learned image downscaling for upscaling using content adaptive resampler," *IEEE Trans. Image Process.*, vol. 29, pp. 4027–4040, Feb. 2020.
[9] Y. Zhao, Q. Chen, X. Sui, and G. Gu, "A novel infrared image super-resolution method based on sparse representation," *Infr. Phys. Technol.*, vol. 71, pp. 506–513, Jul. 2015.
[10] D. Chao, C. L. Chen, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. ECCV*, 2014, pp. 184–199.
[11] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3194–3207, Jul. 2016.
[12] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[13] Y. Wang, L. Wang, H. Wang, and P. Li, "End-to-end image super-resolution via deep and shallow convolutional networks," *IEEE Access*, vol. 7, pp. 31959–31970, 2019.

[14] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang, "Non-local recurrent network for image restoration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1673–1682.
[15] Z. He, S. Tang, J. Yang, Y. Cao, M. Ying Yang, and Y. Cao, "Cascaded deep networks with multiple receptive fields for infrared image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2310–2322, Aug. 2019.
[16] X. Deng, R. Yang, M. Xu, and P. L. Dragotti, "Wavelet domain style transfer for an effective perception-distortion tradeoff in single image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3076–3085.
[17] L. Wang, Y. Wang, Z. Liang, Z. Lin, J. Yang, W. An, and Y. Guo, "Learning parallax attention for stereo image super-resolution," 2019, *arXiv:1903.05784*. [Online]. Available: http://arxiv.org/abs/1903.05784
[18] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019.
[19] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Mar. 2016, pp. 694–711.
[20] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4681–4690.
[21] M. S. M. Sajjadi, B. Schlkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," 2017, *arXiv:1612.07919*. [Online]. Available: https://arxiv.org/abs/1612.07919
[22] *CVC14*. Accessed: Dec. 18, 2019. [Online]. Available: http://adas.cvc.uab.es/elektra/enigma-portfolio/cvc-14-visible-fir-day-night-pedestrian-sequence-dataset/
[23] *TNO*. Accessed: Jun. 9, 2020. [Online]. Available: https://figshare.com/articles/TN_Image_Fusion_Dataset/1008029
[24] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jul. 2004, p. 1.
[25] C. G. M. Bevilacqua, A. Roumy, and M.-L. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. BMVC*, 2012, pp. 1, 2, 6, and 8.
[26] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.
[27] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[29] W. Ouyang and X. Wang, "Joint deep learning for pedestrian detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2056–2063.
[30] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 633–640.
[31] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," 2016, *arXiv:1608.00367*. [Online]. Available: https://arxiv.org/abs/1608.00367
[32] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1646–1654.
[33] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," 2016, *arXiv:1609.05158*. [Online]. Available: http://arxiv.org/abs/1609.05158
[34] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," *Adv. Comput. Vis. Image Process.*, vol. 1, no. 2, pp. 317–339, 1984.
[35] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.
[36] M. Kawulok, P. Benecki, K. Hrynczenko, D. Kostrzewa, S. Piechaczek, J. Nalepa, and B. Smolka, "Deep learning for fast super-resolution reconstruction from multiple images," *Proc. SPIE*, vol. 10996, May 2019, Art. no. 109960B.
[37] A. Bordone Molini, D. Valsesia, G. Fracastoro, and E. Magli, "DeepSUM: Deep neural network for super-resolution of unregistered multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3644–3656, May 2020.
[38] J. Bruna, P. Sprechmann, and Y. Lecun, "Super-resolution with deep convolutional sufficient statistics," 2015, *arXiv:1511.05666*. [Online]. Available: https://arxiv.org/abs/1511.05666
[39] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.
[40] L. A. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," 2015, *arXiv:1505.07376*. [Online]. Available: https://arxiv.org/abs/1505.07376
[41] S. Vasu, N. T. Madam, and R. A. N, "Analyzing perception-distortion tradeoff using enhanced perceptual super-resolution network," 2018, *arXiv:1811.00344*. [Online]. Available: https://arxiv.org/abs/1811.00344
[42] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, "Understanding neural networks through deep visualization," 2015, *arXiv:1506.06579*. [Online]. Available: https://arxiv.org/abs/1506.06579
[43] R. Mechrez, I. Talmi, and L. Zelnik-Manor, "The contextual loss for image transformation with non-aligned data," 2018, *arXiv:1803.02077*. [Online]. Available: https://arxiv.org/abs/1803.02077
[44] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[45] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, vol. 15, Jan. 2010, pp. 315–323.
[46] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 136–144.
[47] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[50] J. F. Brenner, B. S. Dew, J. B. Horton, T. King, P. W. Neurath, and W. D. Selles, "An automated microscope for cytologic research a preliminary evaluation.," *J. Histochemistry Cytochemistry*, vol. 24, no. 1, pp. 100–111, Jan. 1976.
[51] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston, "Comparison of autofocus methods for automated microscopy," *Cytometry*, vol. 12, no. 3, pp. 195–206, 1991.
[52] Z. Cheng, "Fast and high sensitivity focusing evaluation function," *Appl. Res. Comput.*, vol. 27, no. 4, pp. 1534–1536, 2010.
[53] Q. li and R. Dai, "Digital image sharpness evalution function," *ACTA Photonica Sinica*, vol. 31, no. 6, pp. 736–738, 2002.
[54] X. Xie, J. Zhou, and Q. Wu, "No-reference quality index for image blur," *J. Comput. Appl.*, vol. 30, no. 4, pp. 71–74, 2010.
[55] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, Nov. 2019.
[56] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 723–731.
[57] M. S. Rad, B. Bozorgtabar, U.-V. Marti, M. Basler, H. K. Ekenel, and J.-P. Thiran, "SROBB: Targeted perceptual loss for single image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2710–2719.

YIFAN YANG received the B.E. degree from the School of Informatics, Xiamen University, in 2016. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Modern Optical Instrumentation, Zhejiang University. His research interests include computer vision, deep learning, and image processing.

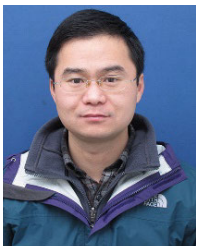
QI LI received the Ph.D. degree in optical engineering from Zhejiang University (ZJU), in 2004. He is currently an Associate Professor with the State Key Laboratory of Modern Optical Instrumentation, ZJU. His research interests include optical system design and imaging techniques.

CHENWEI YANG received the B.E. degree from the College of Optical Science and Engineering, Zhejiang University (ZJU), in 2015, where he is currently pursuing the Ph.D. degree with the State Key Laboratory of Modern Optical Instrumentation. His research interest includes image signal processing such as image deblurring, noise reduction, and HDR.

YANNIAN FU received the B.E. degree from the School of Informatics, Nanjing University, in 2018. He is currently pursuing the master's degree with the State Key Laboratory of Modern Optical Instrumentation, Zhejiang University. His research interests include computer vision, deep learning, and image processing.

HUAJUN FENG received the B.S. and master's degrees from Zhejiang University (ZJU), in 1983. He is currently a Professor with the State Key Laboratory of Modern Optical Instrumentation, ZJU. His research interests include imaging techniques, imaging processing, precision testing technology, and optical system design.

ZHIHAI XU received the bachelor's, master's, and Ph.D. degrees from Zhejiang University (ZJU), in 1986, 1989, and 1996, respectively. He is currently a Professor with the State Key Laboratory of Modern Optical Instrumentation, ZJU. His research interests include optical remote sensing and the imaging chain of cameras.



YUETING CHEN received the bachelor's and Ph.D. degrees from Zhejiang University (ZJU), in 2004 and 2009, respectively. He is currently a Lecturer with the State Key Laboratory of Modern Optical Instrumentation, ZJU. His research interests include computational imaging and optical imaging.