Journal of Beijing Institute of Technology, 2026, Vol. 35, No. 5

An Integrated Algorithm of Image Fusion and Super-Resolution Reconstruction for Dual-Band Micro-Scanning Microscope Imaging System

Meijing Gao[1]✉, Qitai Sun[1], Yufan Du[1], Haozheng Yin[2], Sibo Chen[3], Huanyu Sun[1], Xu Chen[1]

(1. College of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China; 2. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, Hebei, China; 3. Beijing Institute of Aerospace Intelligence and Information, Beijing 100854, China)

Abstract: Micro-scanning thermal microscope imaging system (MTMIS) enables non-destructive observation of temperature variations in fine targets and has important applications in microelectronic detection and biomedical diagnosis. However, existing dual-band systems suffer from low resolution, poor fusion quality, and noise interference, limiting effective utilization of complementary infrared and visible information. To address these issues, this paper proposes a dual-band image fusion and super-resolution (SR) reconstruction network (TCD-KDnet) based on knowledge distillation. A teacher network generates high-resolution fused images and intermediate feature maps as pseudo-labels to guide the student network, which integrates a fusion module and an SR reconstruction module. In particular, a Transformer-CNN (convolutional neural network) dual-branch fusion method is designed to capture global cross-modal context and local texture details, enhanced by hierarchical fusion and attention mechanisms. Furthermore, an improved lightweight RCAN is adopted for efficient SR reconstruction with reduced parameters. Experimental results on public datasets and real MTMIS data demonstrate that the proposed method significantly improves image quality, detail preservation, and reconstruction efficiency.

Keywords: dual-band micro-scanning microscope imaging system; TCD-KDnet; image fusion; super-resolution (SR) reconstruction; knowledge distillation

1 Introduction

Since the 1990s, with the strong demand for night vision and combat in harsh conditions in the military field, thermal imaging technology has developed rapidly [1–3]. However, most of these systems are telescopic, while in practical applications, micro-scanning thermal microscope imaging systems (MTMIS) with optical amplification mode are often used to analyze the fine temperature distribution of samples. According to statistics, about 70% of the physical and chemical changes of objects are related to temperature. MTMIS can observe, record, and analyze the temperature variation process of subtle targets in detail due to its high sensitivity to temperature. Therefore, MTMIS has broad application prospects in nondestructive testing fields such as microelectronic devices, scientific research, public security criminal investigation [4,5], biomedical diagnosis [6–9], and thermophysics and materials science.

In 2005, the research group first developed

Manuscript received Apr. 23, 2026; revised Apr. 26, 2026; accepted Apr. 30, 2026. The associate editor coordinating the review of this manuscript was Dr. Yixiao Yang. This work was supported by National Natural Science Foundation of China (No. 62471034), Hebei Natural Science Foundation (No. F2023105001) and Sichuan Science and Technology Education Joint Foundation (No. 25 LHJJ0222).

✉ Corresponding author. Email: gaomeijing@126.com

DOI: 10.15918/j.jbit1004-0579.2026.029

an MTMIS based on an uncooled focal plane detector, and subsequently constructed three sets of uncooled MTMIS, leading the development of this field [10–19]. Several micro-scanning thermal image processing algorithms were also proposed for this system [10–17]. However, previous MTMIS mainly relied on single-band infrared detectors, which suffer from limited depth of field and low resolution, resulting in unclear texture details and strong noise interference.

To overcome these limitations, a parallel optical path infrared/visible fusion MTMIS was proposed in 2022, enabling simultaneous acquisition of infrared and visible microscopic images. By leveraging image fusion techniques, complementary information from both modalities can be effectively integrated, improving image clarity and overcoming the limitations of single-source imaging in terms of resolution, information content, and physical properties. This also enhances the accuracy of fault detection and localization. However, due to the optical diffraction limit, spatial discrete sampling of focal plane detectors, and spatial integration effects of detector units, micro-scanning thermal imaging often suffers from under-sampling. In addition, the long working wavelength further aggravates diffraction effects, leading to spectral aliasing and image quality degradation.

Super-resolution (SR) reconstruction techniques can recover high-resolution images with rich high-frequency details and clear edges from low-resolution inputs. Therefore, this paper investigates SR reconstruction methods based on deep learning, image fusion, and micro-scanning theory to improve the performance of MTMIS.

Specifically, an integrated framework for image fusion and SR reconstruction is proposed. A knowledge distillation-based infrared/visible image fusion and SR reconstruction network (TCD-KDnet) is constructed, which achieves lightweight design and efficient training. The teacher network generates high-resolution fused images as pseudo-labels to guide the student network, enabling simultaneous low-resolution image fusion and SR reconstruction.

The remainder of this paper is organized as follows. Section 2 reviews the related work on image fusion and super-resolution reconstruction. Section 3 introduces the dual-band micro-scanning microscope imaging system and the overall framework of the proposed TCD-KDNet. Section 4 presents the proposed network architecture, feature extraction and fusion modules, knowledge distillation strategy, and loss functions. Section 5 reports the experimental evaluation and provides qualitative and quantitative analyses of the proposed method. Section 6 further validates the effectiveness of the proposed method on an actual dual-band fusion micro-scanning microscope imaging system. Finally, Section 7 concludes the paper.

2 Related work

2.1 Construction of the Dataset

With the rapid development of deep learning and high-performance GPUs, deep learning-based methods have been widely applied in image SR reconstruction. Two mainstream approaches for infrared image SR are convolutional neural network (CNN)-based and generative adversarial network (GAN)-based methods.

In CNN-based SR algorithms, Dong et al. [20] proposed the super-resolution convolutional neural network (SRCNN) in 2014, which employs a three-layer CNN for image SR reconstruction. However, its shallow structure limits feature extraction capability. In 2021, Prajapati et al. [21] introduced a channel-splitting CNN for infrared image SR, but this approach may fragment feature representations and increase fusion complexity, affecting reconstruction consistency. In the same year, Zou et al. [22] incorporated skip connections and channel fusion to enhance detail restoration; however, training remains difficult and incomplete feature integration may still degrade reconstruction fidelity.

To further improve visual quality, Ledig et al. [23] introduced the generative adversarial network into SR reconstruction (SRGAN), which enhances perceptual details but leads to lower PSNR and SSIM. In 2021, Keqing et al. proposed the Wasserstein GAN (WGAN), improving PSNR and SSIM while reconstructing richer high-frequency textures, but at the cost of increased memory and inference time [24]. Huang et al. [25] developed the heterogeneous SR WGAN (HetSRWGAN), which employs heterogeneous convolutions and adversarial learning for infrared feature extraction; however, it may introduce artifacts due to adversarial training. In 2023, Liang et al. [26] applied dual-attention mechanisms to jointly model spatial and channel features using CNN and GAN, yet performance remains limited. In 2024, Yuan et al. [27] proposed the graph recurrent attention network (GRAN), which effectively integrates deep and fine-grained features, but tends to produce overly smooth results.

Most existing SR methods improve performance by increasing network depth, which leads to higher computational cost and memory consumption, limiting practical deployment. Furthermore, performing image fusion prior to SR reconstruction introduces noise and interference into the SR process, thereby degrading reconstruction performance.

3 Dual-Band Fusion Microscope Imaging System and the Overall Framework of Fusion/SR Integrated Algorithm

3.1 System Composition and Working Principle of the Imaging System

Based on the uncooled focal plane detector module, the precision optical flat micro-scanning system, and the dual-beam microscope head, an integrated micro-scanning dual-beam microscope imaging system is constructed, as shown in Fig. 1.

The overall workflow of the system is illustrated in Fig. 2. Specifically, the target object passes through two parallel micro-optical paths: one for micro-scanning thermal imaging and the

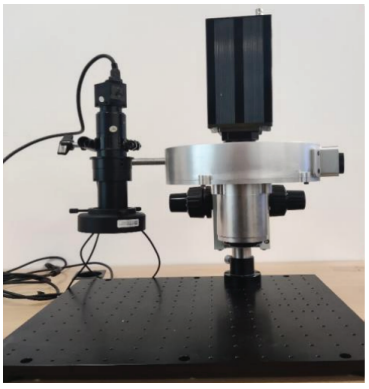


Fig. 1 Infrared-visible fusion micro-scanning microscope imaging system

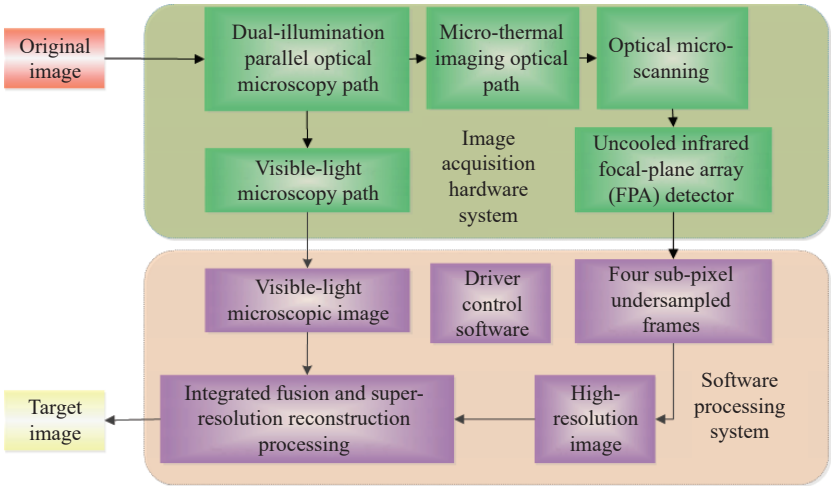


Fig. 2 Configuration of the dual-optical-path micro-scanning microscope imaging system

other for visible light imaging. First, infrared micro-rays are approximately collimated by the optical MTMIS and then captured by an uncooled detector, generating four undersampled images I_1–I_4. These images are reconstructed using a standard 2×2 micro-scanning scheme to obtain a high-resolution image. Meanwhile, the visible light path captures corresponding microscopic images. Finally, the reconstructed high-resolution thermal image and the visible light image are fused and enhanced through the SR reconstruction algorithm to obtain a dual-modality fused SR microscopic image.

3.2 Overall Framework of TCD-KDnet: An Integrated Fusion and SR Reconstruction Algorithm for Dual-Band Micro-Scanning Microscope Imaging System

The TCD-KDnet designed in this paper consists of two parts: the teacher network and the student network. The network structure is shown in Fig. 3. Among them, the student network consists of a fusion part and an SR reconstruction part, which are responsible for the fusion of low-resolution images and the reconstruction of SR images, respectively. The teacher's network is responsible for fusing high-resolution images as labels, and then guiding the student's network to fuse low-resolution images and reconstruct SR.

4 Overall Architecture of TCD-KDnet: An Integrated Fusion and SR Reconstruction Framework

4.1 Design of the Teacher Network

As shown in Fig. 3, the teacher network designed in this paper consists of three main components: an encoder, a fusion module, and adecoder. The encoder comprises a 3 × 3 convolutional layer followed by a densely connected block. The convolutional layer is responsible for extracting shallow features. The densely connected block contains three stacked 3×3 convolutional layers, where the output of each layer is concatenated and passed as input to the next layer, effectively preserving feature information and enhancing high-frequency details.

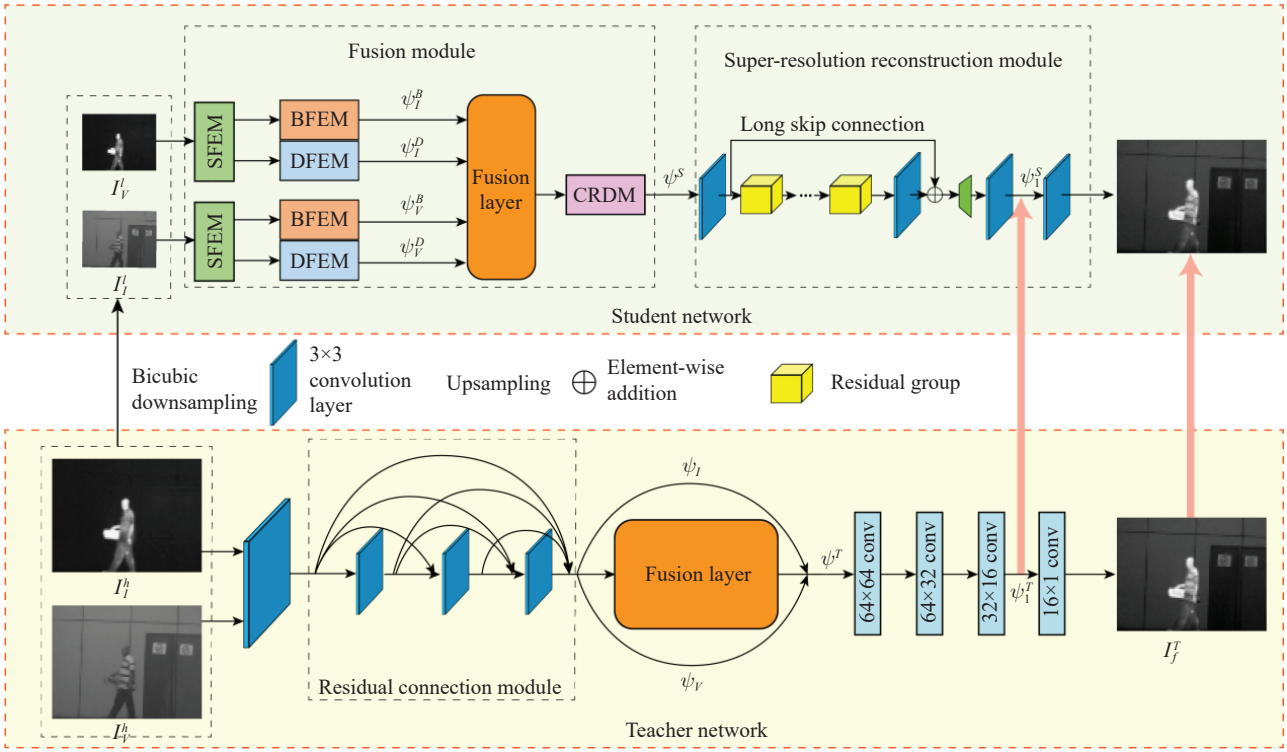


Fig. 3 TCD-KDnet network architecture

The fusion module aggregates the features extracted by the encoder and combines them with the infrared features ψ_1 and visible features ψ_v to form the fused representation ψ^T, which is then fed into the decoder. The decoder is composed of four convolutional layers and produces the final high-resolution fused image I_f^T. Additionally, the intermediate output ψ_1^T from the penultimate convolutional layer and the final output I_f^T are used as supervisory signals to guide the training of the student network.

The student network is divided into two subnetworks: a fusion network and an SR reconstruction network, which are introduced in detail below.

4.2 Design of the Feature Fusion Module

4.2.1 Design of the Shared Feature Extraction Module (SFEM)

As shown in Fig. 4, this paper constructs an SFEM based on a Transformer backbone to apply attention along the channel dimension and capture global context from input images. The module consists of two key components: MDTA [28] and GDFN, which are connected by residuals, and can extract cross-modal shallow features without increasing computational overhead significantly, thus improving the fusion effect.

To keep the parameter scale and computational complexity close to that of conventional fusion networks (FN), this paper makes two improvements to the FN of the GDFN module. First, it reduces the feature expansion ratio, which improves the feature modeling ability while still maintaining high computational efficiency. Second, it introduces a nonlinear activation function into the hidden layer.

4.2.2 Design of the Detail Feature Extraction Module (DFEM)

The input and output features of the invertible neural network (INN) are generated symmetrically, which helps better preserve the input information. Therefore, this paper constructs a DFEM module based on the INN block [29] that includes affine coupling layers. To ensure feature preservation, we adopt a densely connected reversible feature fusion module composed of multiple reversible units. The composition of the DFEM module is shown in Fig. 5, where each basic unit utilises an affine coupling layer.

To further enhance the representational

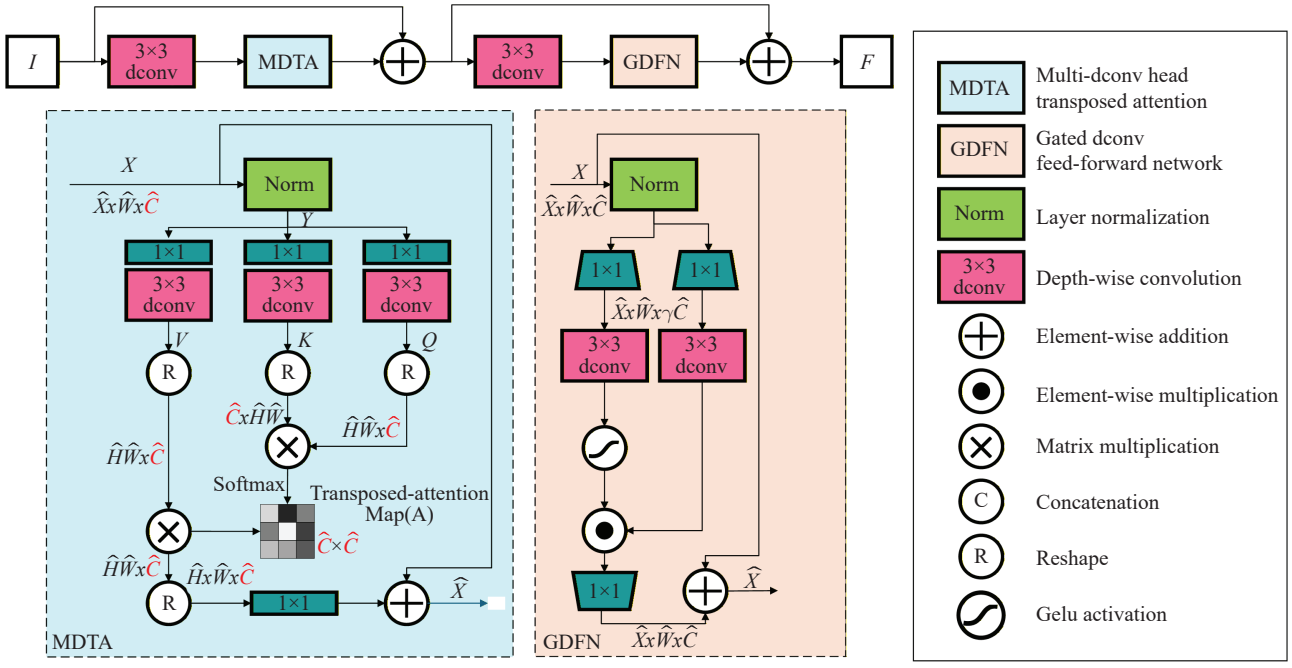


Fig. 4 Structure of the shared feature extraction module

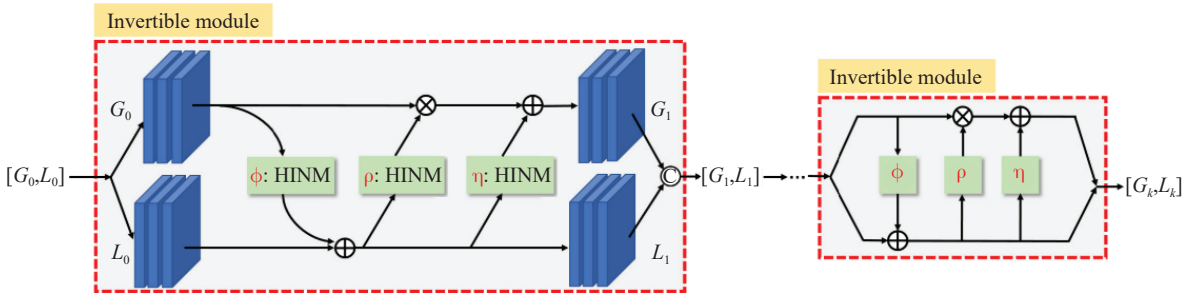


Fig. 5 Schematic of the detail feature extraction module

capacity of the network, this paper makes the following improvements. First, the direct outputs of each reversible unit are propagated to the final reversible unit via skip connections, and then concatenated to further strengthen the representation. Second, an effective interaction mechanism between two reversible units is designed, specifically using additive coupling for the long-range branch and an enhanced affine coupling for the local branch.

4.2.3 Design of the Background Feature Extraction Module (BFEM)

The BFEM designed in this paper is composed of multiple LT blocks connected in series, as illustrated in Fig. 6. To further reduce the computational cost, the standard convolution operations within the LT block are replaced with depthwise separable convolutions.

4.2.4 Design of the Feature Fusion Module (GFS)

Aiming at the challenge of simultaneously preserving multi-source feature information in infrared and visible images, we design a hierarchical fusion strategy called GFS. As shown in Fig. 7, GFS is divided into two levels. The first level contains two branches: one for background feature fusion and the other for detail feature fusion. The background features of dual-modality images are input to the background feature fusion layer for fusion, and the detail texture features of dual-modality images are input to the detail feature fusion layer for fusion. This design avoids feature degradation or distortion caused by excessive convolution without increasing the parameter count.

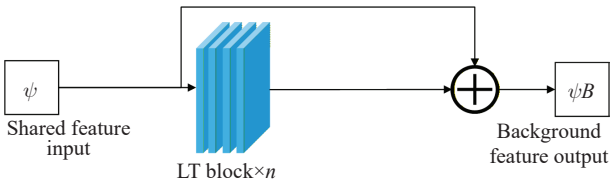


Fig. 6 Schematic of the background feature extraction module

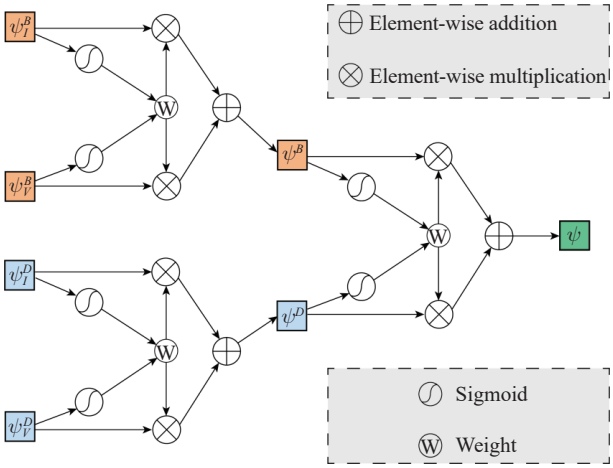


Fig. 7 Schematic of the feature fusion module (GFS)

4.2.5 Design of the Cross-Fusion Refinement Module (CFRM)

The CFRM module is designed to decode the fused-layer output features and produce the fused image. To enhance both image quality and feature reconstruction efficiency, a convolutional block abttention module (CBAM) is incorporated into the feature reconstruction stage [30]. Its architecture is illustrated in Fig. 8.

4.3 Design of the SR Reconstruction Network

In the SR reconstruction network, by feeding the fused features into the SR module, a high-resolution fused image can be reconstructed. To ensure the effectiveness of SR reconstruction while controlling the overall network parameters, this paper improves the RCAN [31], as shown in the

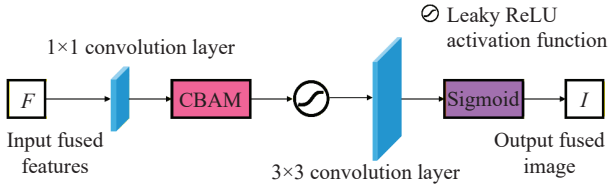


Fig. 8 Structure diagram of the cross-fusion refinement module

SR reconstruction section of Fig. 3. The original architecture of 12 stacked residual groups has been reduced to 6 stacked groups.

RCAN uses a single convolutional layer to extract shallow features from the input image I^{LR}. Let $H_{\mathrm{SF}}(\cdot)$ denote the shallow feature extraction operator and F_0 the resulting feature map output from this layer, expressed as

$$F_0 = H_{\mathrm{SF}}(I^{\mathrm{LR}}) \tag{1}$$

The extraction of deep features F_{DF} employs a nested residual structure known as residual in residual (RIR), where short skip connections are added within the residual blocks. To reduce the number of parameters and prevent overfitting, this paper adopts an RIR composed of six serially connected residual groups to extract deep image features. Its specific expression is

$$F_{\mathrm{DF}} = H_{\mathrm{RIR}}(F_0) \tag{2}$$

where $H_{\mathrm{RIR}}(\cdot)$ denotes the RIR operator.

Next, F_{DF} is processed by an upsampling operation. In this paper, the upsampling method is sub-pixel convolution. The expression is

$$F_{\mathrm{UP}} = H_{\mathrm{UP}}(F_{\mathrm{DF}}) \tag{3}$$

where F_{UP} represents the feature map obtained after upsampling, and $H_{\mathrm{UP}}(\cdot)$ corresponds to the sub-pixel convolution operation.

Finally, F_{UP} is fed into the reconstruction layer to obtain the final SR reconstructed image I_{SR}, defined by the expression:

$$I_{\mathrm{SR}} = \mathrm{Conv}(F_{\mathrm{UP}}) \tag{4}$$

4.4 Design of the Loss Functions

4.4.1 Design of the Teacher Network Loss Function

A hybrid loss function consisting of pixel loss, structural similarity (SSIM) loss, and gradient loss is constructed to guide the optimisation of the teacher network. This ensures consistency between the input and reconstructed images across pixel intensity, structural similarity, and gradient variation, thereby preserving textural details more effectively. The total loss of the teacher network $\mathcal{L}_t$ can be expressed as

$$\mathcal{L}_t = \mathcal{L}_{\mathrm{pixel}} + \alpha\,\mathcal{L}_{\mathrm{ssim}} + \mathcal{L}_{\mathrm{grad}} \tag{5}$$

where $\mathcal{L}_{\mathrm{pixel}}$ denotes the pixel loss, $\mathcal{L}_{\mathrm{ssim}}$ is the SSIM loss, $\mathcal{L}_{\mathrm{grad}}$ represents the gradient loss, and α is a weight coefficient. These three loss functions are defined as follows

$$\mathcal{L}_{\mathrm{pixel}} = \|x - y\|_2 \tag{6}$$

$$\mathcal{L}_{\mathrm{ssim}} = 1 - \mathrm{SSIM}(x, y) \tag{7}$$

$$\mathcal{L}_{\mathrm{grad}} = \|\nabla x - \nabla y\|_2 \tag{8}$$

where $\|\cdot\|_2$ represents the norm of ℓ_2, and $\mathrm{SSIM}(\cdot)$ denotes the structural similarity function.

4.4.2 Design of the Student Network Loss Function

After training the teacher network, the student network is trained by inputting low-resolution infrared and visible images into both the student and teacher networks. The ℓ_1 loss between the high-resolution fusion images generated by the teacher network and the student network is computed. Additionally, to enable the student network to learn richer high-resolution fusion features, ℓ_1 loss is also applied to the intermediate layer features of both networks. The total loss function $\mathcal{L}_s$ for the student network is expressed as

$$\mathcal{L}_s = \left\|I_f^T - I_f^S\right\|_1 + \varphi\left\|\psi_1^T - \psi_1^S\right\|_1, \tag{9}$$

where I_f^T represents the high-resolution fusion image generated by the teacher network, I_f^S denotes the high-resolution fusion image produced by the student network, ψ_1^T refers to the penultimate layer features of the teacher network, ψ_1^S indicates the penultimate layer fea-

tures of the student network, and φ signifies the weight coefficient applied to the intermediate feature loss.

5 Experimental Evaluation and Results Analysis

The algorithm in this paper is based on the fusion module TCDFusion, adding a lightweight RCAN image SR network, and integrated training by knowledge distillation. To evaluate the performance of the proposed method, five fusion algorithms were combined with the RCAN SR reconstruction algorithm, and their output results were compared and analysed. All comparative experiments strictly followed the parameter configuration requirements specified in the original literature. Considering the scarcity of microscope image datasets and the unique characteristics of actual acquisition scenarios, a phased validation strategy was adopted. First, simulation validation was conducted based on two general image fusion benchmark datasets, TNO and RoadScene. Subsequently, experimental validation was performed using images collected from the actual MTMIS.

5.1 Experimental Setup, Training Process, and Training Datasets

The experimental setup used in this chapter is detailed in Tab. 1. A total of 1 079 pairs of infrared and visible images were selected as the training set, and 360 pairs as the test set. During training, the data were loaded in HDF5 format. Due to varying image sizes in the dataset, preprocessing was performed by cropping the images into 128×128 patches to serve as ground truth labels. The Adam optimiser was employed with a batch size of 2.

Tab. 1 Hardware and software environment used in this study

Software/hardware	Version/model
Host system	Windows 11
DL framework	PyTorch 1.7.0
CPU	Intel Core i9-9900k
GPU	NVIDIA RTX 2080Ti
CUDA	11.7
Host memory	16 GB

In the knowledge distillation training process, the teacher network was trained first for 200 epochs. Subsequently, the student network was trained for a total of 300 epochs. The first 60 epochs focused solely on training the fusion network to establish basic fusion capabilities. The remaining 240 epochs involved training the entire student network, resulting in a fused infrared and visible image SR reconstruction network. The loss function curve is depicted in Fig. 9, showing a gradual decrease in loss with stabilisation around the 140th epoch.

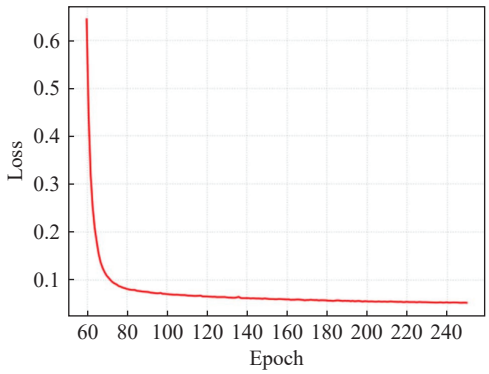
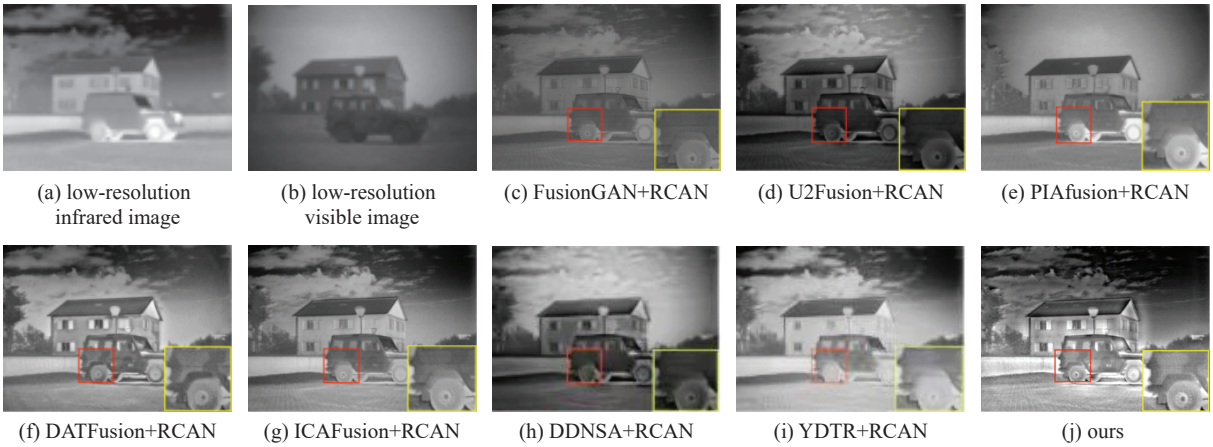


Fig. 9 Loss function variation curve during training

5.2 Subjective Evaluation

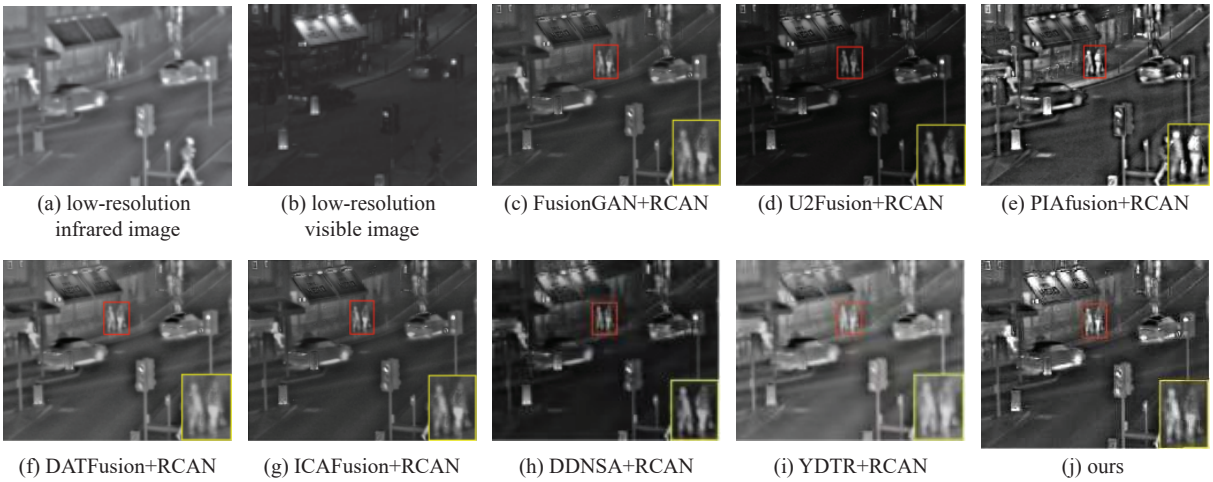
Figs. 10–15 present the fusion and SR reconstruction results across six different scene types. Figs. 10–12 demonstrate improvements in texture detail enhancement, integration of infrared salient targets with visible light texture details, and preservation of infrared texture information, respectively. Figs. 13–15 show the fusion effect of infrared target retention under complex lighting conditions, overexposure and strong light irradiation, respectively. In each group of images, (a) and (b) represent low-resolution infrared and visible images, respectively; (c) to (i) display fusion and SR reconstruction results from five comparative algorithms; and (j) presents the fusion and SR reconstruction result from the proposed method. To better illustrate target retention



(a) low-resolution infrared image (b) low-resolution visible image (c) FusionGAN+RCAN (d) U2Fusion+RCAN (e) PIAfusion+RCAN

(f) DATFusion+RCAN (g) ICAFusion+RCAN (h) DDNSA+RCAN (i) YDTR+RCAN (j) ours

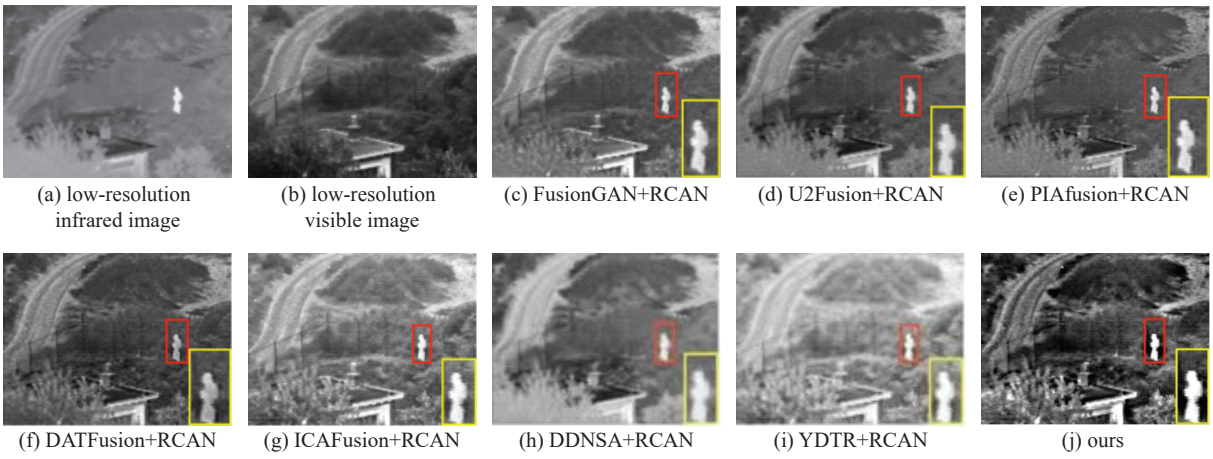
Fig. 10 Fusion and SR reconstruction experiment of the first group of images



(a) low-resolution infrared image (b) low-resolution visible image (c) FusionGAN+RCAN (d) U2Fusion+RCAN (e) PIAfusion+RCAN

(f) DATFusion+RCAN (g) ICAFusion+RCAN (h) DDNSA+RCAN (i) YDTR+RCAN (j) ours

Fig. 11 Fusion and SR reconstruction experiment of the second group of images



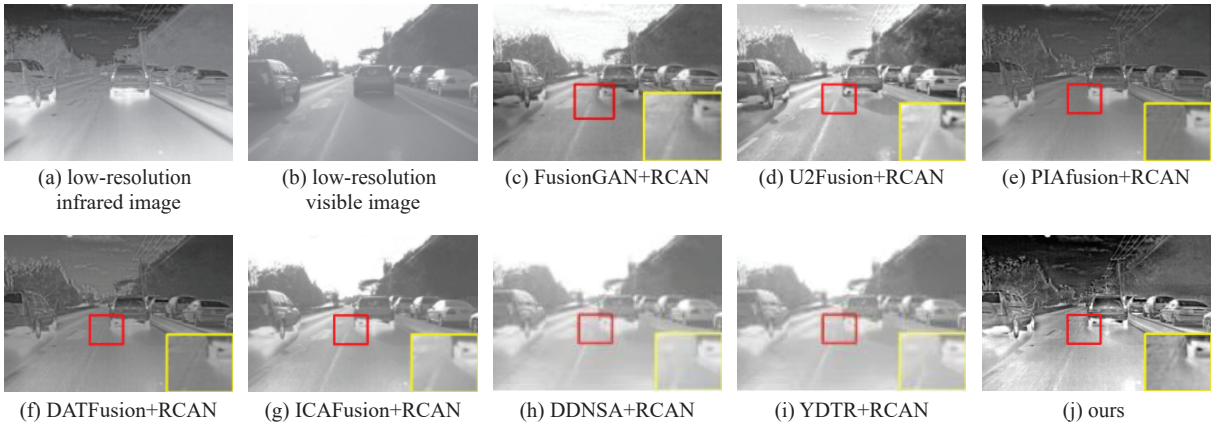
(a) low-resolution infrared image (b) low-resolution visible image (c) FusionGAN+RCAN (d) U2Fusion+RCAN (e) PIAfusion+RCAN

(f) DATFusion+RCAN (g) ICAFusion+RCAN (h) DDNSA+RCAN (i) YDTR+RCAN (j) ours

Fig. 12 Fusion and SR reconstruction experiment of the third group of images

capabilities, zoomed-in regions of interest are displayed in the bottom-right corners of the images.

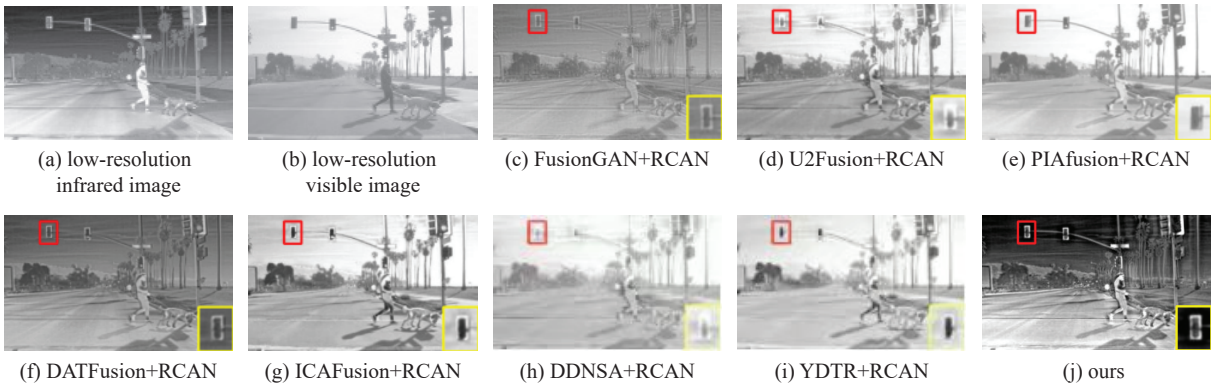
Upon analysing the results of Fig. 10, it is observed that FusionGAN + RCAN and YDTR + RCAN produce images with unclear edge information. U2Fusion + RCAN and DDNSA + RCAN yield relatively balanced results; however, the overall images appear darker, leading to a loss of some infrared information. PIAFusion + RCAN results in significant loss of infrared infor-



(a) low-resolution infrared image (b) low-resolution visible image (c) FusionGAN+RCAN (d) U2Fusion+RCAN (e) PIAfusion+RCAN

(f) DATFusion+RCAN (g) ICAFusion+RCAN (h) DDNSA+RCAN (i) YDTR+RCAN (j) ours

Fig. 13 Fusion and SR reconstruction experiment of the fourth group of images



(a) low-resolution infrared image (b) low-resolution visible image (c) FusionGAN+RCAN (d) U2Fusion+RCAN (e) PIAfusion+RCAN

(f) DATFusion+RCAN (g) ICAFusion+RCAN (h) DDNSA+RCAN (i) YDTR+RCAN (j) ours

Fig. 14 Fusion and SR reconstruction experiment of the fifth group of images



(a) low-resolution infrared image (b) low-resolution visible image (c) FusionGAN+RCAN (d) U2Fusion+RCAN (e) PIAfusion+RCAN

(f) DATFusion+RCAN (g) ICAFusion+RCAN (h) DDNSA+RCAN (i) YDTR+RCAN (j) ours

Fig. 15 Fusion and SR reconstruction experiment of the sixth group of images

mation, with elements like clouds in the sky disappearing entirely. DATFusion + RCAN achieves good visual effects overall, but still experiences some loss of infrared information, such as clouds on rooftops. ICAFusion + RCAN maintains a good sense of layering but suffers from low contrast. In contrast, the image processed by the proposed method exhibits clearer contours of the vehicle body and wheels, better contrast, and appropriate overall brightness, demonstrating superior performance.

Fig. 11 illustrates that pedestrian targets in the images processed by FusionGAN + RCAN, U2Fusion + RCAN, and DDNSA + RCAN algorithms appear relatively blurred. Although ICAFusion + RCAN and PIAFusion + RCAN algorithms preserve significant infrared information, the edges of the targets are not clearly defined,

resulting in an overall blurry visual effect. DATFusion + RCAN and YDTR + RCAN algorithms lose a considerable amount of visible image information, leading to poor overall image quality. In contrast, the image processed by the proposed method exhibits distinct texture details; the letters on the canopy are discernible, pedestrian targets are clearer, and the image maintains good contrast and appropriate overall brightness. Therefore, under complex lighting conditions and subtle texture variations, the proposed method demonstrates superior performance.

Fig. 12 demonstrates that while the overall visual effects of all algorithms are acceptable, certain methods exhibit specific limitations. FusionGAN + RCAN, U2Fusion + RCAN, PIAFusion + RCAN, and DDNSA + RCAN algorithms fail to preserve the visible light texture information of the tree in the lower-left corner. DATFusion + RCAN algorithm leans more towards visible image information, resulting in less distinct infrared target contours. ICAFusion + RCAN and YDTR + RCAN algorithms have overall low contrast, with insufficient edge information and a lack of texture details in the background. In contrast, the image processed by the proposed method exhibits clearer background texture details, well-defined human contours, and appropriate image contrast, demonstrating superior performance in infrared and visible image fusion and SR reconstruction.

Fig. 13 illustrates that the visible image in this group suffers from overexposure due to strong lighting and the high reflectivity of the car's surface. Although the visual effect appears smooth, it lacks clear texture details and prominent edge structures. To better showcase the details of the infrared and visible image fusion and SR reconstruction, local enlargements of the tire and lane markings are provided, with the zoomed-in regions placed in the lower-right corner of the image. In this scenario, FusionGAN + RCAN, ICAFusion + RCAN, DDNSA + RCAN, and YDTR + RCAN algorithms fail to extract texture information from the infrared image, such as the clouds present in the original infrared image, which are barely visible in the fused images. U2Fusion emphasises visible image information more, resulting in a higher overall brightness with richer texture details. However, thermal target information is not sufficiently prominent. PIAFusion and DATFusion algorithms produce images that are generally darker, with insufficient extraction of visible image texture details, making the lane markings on the road less distinct. Although the image processed by the proposed method shows some quality degradation under low resolution, the overall texture details remain visible, the lane markings are clearly marked, the edges of thermal targets are distinct, and the image maintains good contrast and appropriate overall brightness. Therefore, the proposed method demonstrates superior performance in infrared and visible image fusion and SR reconstruction.

Fig. 14 illustrates that the FusionGAN + RCAN, PIAFusion + RCAN, and DDNSA + RCAN algorithms exhibit overall lack of clarity in fine details, especially in the edge regions of objects, making it difficult to distinguish target objects from the background. The U2Fusion + RCAN algorithm, due to lighting effects, results in noticeable halo artefacts around traffic lights, pedestrians, and the background in the processed images. The ICAFusion + RCAN and YDTR + RCAN algorithms present clear object edges but suffer from overexposure, causing the background to lose texture details. In contrast, both the DATFusion + RCAN and the proposed method maintain relatively clear contours for traffic lights, pedestrians, and the background, effectively avoiding halo artefacts, and the images exhibit appropriate overall brightness. Thus, the proposed method demonstrates superior performance compared to other algorithms.

Fig. 15 illustrates that the images processed by the U2Fusion + RCAN, FusionGAN + RCAN, DDNSA + RCAN, and YDTR + RCAN algorithms completely lose the pedestrian information present in the infrared images. The image processed by the DATFusion + RCAN algorithm exhibits significant noise, resulting in a poor visual effect. The PIAFusion + RCAN algorithm fails to delineate the edges of distant buildings clearly. In contrast, the images processed by the ICAFusion + RCAN algorithm and the proposed method demonstrate overall better visual effects; the former provides superior processing of trees, while the latter offers clearer edge texture details of vehicles. Furthermore, the proposed method effectively eliminates the strong light interference from oncoming vehicles, maintains good image contrast, and ensures appropriate overall brightness. In summary, the proposed method exhibits superior comprehensive performance.

In summary, the algorithm proposed in this paper effectively integrates information from infrared and visible images and enhances their resolution under various experimental scenarios and complex interference conditions, demonstrating excellent performance. The proposed algorithm not only significantly improves the texture details in low-brightness areas but also enhances the edge structures of visible images, effectively increasing the resolution of the fused images and ultimately producing high-resolution fused infrared and visible images. To objectively evaluate the algorithm's performance, eight evaluation metrics were employed, as shown in Tab. 2, with bold fonts indicating the best results among the experimental outcomes. As observed in Tab. 2, compared to the other seven algorithms, the proposed method achieves optimal values in average gradient (AG), effective number of independent pixels (EIN), entropy (EN), feature difference (FD), spatial frequency (SF), visual information fidelity for fusion (VIFF), structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR), with improvements of 6.1%, 7.3%, 1.8%, 2.9%, 8.0%, 14.7%, 4.2%, and 17.7% over the second-best values, respectively. These results, combined with the visual effects of the proposed algorithm, fully demonstrate its capability to effectively reduce the impact of adverse factors such as human interference, thereby exhibiting superior performance.

6 Validation of the Actual Dual-Band Fusion Micro-scanning Microscope Imaging System

We collected four groups of dual-light images by using the dual-band microscope imaging system built by our previous research group, and used the algorithm proposed in this paper for fusion and SR reconstruction. Due to space limitations, only one group of reconstruction results is presented here. As shown in Fig. 16, image (a) is the low-resolution thermal microscope image, image (b) is the low-resolution visible light microscope image, and images (c) to (j) sequen-

Tab. 2 Quantitative evaluation results on the TNO dataset

Method	AG ↑	EIN ↑	EN ↑	FD ↑	SF ↑	VIFF ↑	SSIM ↑	PSNR ↑
FusionGAN + RCAN	3.044	30.689	4.947	3.718	6.009	0.851	0.450	14.732
U2Fusion + RCAN	3.795	29.847	5.498	4.451	5.770	0.603	0.507	15.457
PIAFusion + RCAN	3.368	33.985	5.705	5.135	8.755	0.682	0.450	15.860
DATFusion + RCAN	3.591	34.288	5.340	5.087	10.893	0.735	0.532	16.885
ICAFusion + RCAN	3.089	37.816	5.360	5.791	10.548	0.849	0.595	18.769
DDNSA + RCAN	4.026	34.026	5.268	5.074	6.578	0.687	0.578	17.621
YDTR + RCAN	3.996	36.682	5.199	5.617	10.232	0.824	0.577	18.206
Ours	4.271	40.580	5.808	5.960	11.766	0.976	0.620	22.090

tially represent the fusion and SR reconstruction results of various comparison algorithms and the proposed algorithm. Visually, the image processed by the proposed algorithm exhibits a brighter field of view, more prominent edge textures, and a stronger overall sense of layering, with clearer and more distinct targets.

In summary, the algorithm proposed in this paper effectively integrates information from infrared and visible images and enhances their resolution under various experimental scenarios and complex interference conditions, demonstrating excellent performance. The proposed algorithm not only significantly improves the texture details in low-brightness areas but also enhances the edge structures of visible images, effectively increasing the resolution of the fused images and ultimately producing high-resolution fused infrared and visible images. To objectively evaluate the algorithm's performance, eight evaluation metrics were employed, as shown in Tab. 3, with bold fonts indicating the best results among the experimental outcomes. Compared with the other algorithms, the proposed method achieves optimal values in average gradient (AG), effective number of independent pixels (EIN), entropy (EN), feature difference (FD), spatial frequency (SF), visual information fidelity (VIFF), structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR). Specifically, the proposed method improves upon the second-best algorithms by 5%, 6.5%, 3.4%, 9.3%, 14.7%, 15.3%, 11.2%, and 25%, respectively. These quantitative results, combined with the visual effects observed, fully demonstrate that the proposed algorithm exhibits superior performance in the joint task of microscope image fusion and SR reconstruction.



(a) low-resolution thermal micro-scope image (b) low-resolution visible light microscope image (c) FusionGAN+RCAN (d) U2Fusion+RCAN (e) PIAfusion+RCAN



(f) DATFusion+RCAN (g) ICAFusion+RCAN (h) DDNSA+RCAN (i) YDTR+RCAN (j) ours

Fig. 16 Fusion and SR reconstruction experiment of the first group of low-resolution microscope images

Tab. 3 Average evaluation results of fusion and SR reconstruction experiments on four groups of low-resolution microscope images

Method	AG ↑	EIN ↑	EN ↑	FD ↑	SF ↑	VIFF ↑	SSIM ↑	PSNR ↑
FusionGAN + RCAN	3.047	31.274	4.990	3.735	6.074	0.808	0.398	14.981
U2Fusion + RCAN	3.814	30.414	5.552	4.483	5.831	0.555	0.456	15.721
PIAfusion + RCAN	3.378	34.640	5.764	5.182	8.878	0.635	0.399	16.133
DATFusion + RCAN	3.605	34.948	5.391	5.133	11.061	0.689	0.482	17.179
ICAFusion + RCAN	3.092	38.551	5.412	5.852	10.709	0.805	0.546	19.103
DDNSA + RCAN	4.348	34.672	6.029	5.111	6.647	0.633	0.520	17.922
YDTR + RCAN	4.308	41.250	5.791	5.262	9.459	0.861	0.504	17.440
Ours	4.566	43.933	6.232	6.397	12.692	0.993	0.607	23.885

7 Conclusions

In this study, we propose a knowledge distillation network, TCD-KDnet, that unifies image fusion and SR reconstruction tasks. This network utilizes a teacher network to generate high-resolution fused images and intermediate features, which serve as pseudo-labels to supervise the student network. This approach ensures efficient training while achieving model lightweighting and high-performance training, effectively solving the spatial resolution limitations of thermal microscope imaging systems and providing robust technical support for high-resolution microscope thermal analysis.

In the network architecture design, the teacher network is solely responsible for fusing high-resolution infrared and visible images, while the student network comprises two modules: the fusion module and the SR reconstruction module. Considering the characteristics of infrared and visible light microscope images—such as low resolution and significant texture differences —we propose an image fusion algorithm based on Transformer/CNN dual-branch feature interaction. Through the complementarity of the two networks in receptive field and feature expression, the texture information of infrared thermal target and visible light is preserved.

We first construct a feature extraction encoder that employs three modules to separately extract background and detail features from infrared and visible images through dual branches: a shared feature extraction module (SFEM) designed with a Transformer-CNN hybrid structure to mine cross-modal shallow features; a detail feature extraction module (DFEM) based on an invertible network design to retain high-frequency texture information; and a background feature extraction module (BFEM) where a lightweight Transformer further refines global background information.

Subsequently, we design a graded fusion strategy (GFS) to effectively address the challenge of simultaneously preserving different types of feature information from infrared and visible images. Furthermore, by introducing channel and spatial attention mechanisms, we design a channel and spatial feature reconstruction module (CFRM) to optimize the quality of the fused image. Finally, considering the frequency domain distribution characteristics of infrared and visible light signals, we propose a correlation loss function based on the "low-frequency correlation-high-frequency independence" theory. This function, when combined with pixel loss and gradient loss, guides the network to maintain image structure and details better.

In the SR reconstruction section, we improve the RCAN architecture by using only six residual groups to achieve high-fidelity amplification of the fused features, thereby controlling the model parameter count while ensuring reconstruction quality. Experimental results demonstrate that the images processed by our algorithm exhibit a brighter field of view, more prominent edge textures, and a stronger overall sense of layering, with clearer and more distinct targets.

Subjective visual effects and quantitative metrics from various experiments confirm the effectiveness of the integrated fusion and SR reconstruction network proposed in this paper. When performing joint fusion and SR reconstruction of infrared and visible images in microscope scenes, our algorithm can better preserve infrared thermal information and visible light texture details, achieving higher spatial resolution and contrast.

References:

[1] Y. Cao, Y. Zhang, and J. Yang,"Research on key technologies of uncooled infrared thermal imaging temperature measurement," *Journal of Vibration, Measurement and Diagnosis*, vol. 44, no. 2, pp. 209-216, 2024.

[2] X. Chen, L. Chen, L. Chen, P. Chen, G. Sheng, X. Yu, and Y. Zou,"Modeling thermal infrared image

degradation and real-world super-resolution under background thermal noise and streak interference," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 7, pp. 6194-6206, 2024.

[3] Z. Wei, C. Huiyuan, W. Yanqi, T. Jie, and Y. Yaping, "Application of infrared thermography in measuring the focal plane of a focused ultrasound transducer, " *Metrology Science and Technology*, vol. 68, no. 10, pp. 73-78, 2024. https://jlkj.nim.ac.cn/en/article/doi/10.12338/j.issn.2096-9015.2023.0280

[4] C. Ferrante, L. Lucchesi, A. Cemmi, I. Di Sarcina, J. Scifo, A. Verna, A. Taschin, L. Senni, M. Beghini, B. D. Monelli, and F. Raffaelli, "Gamma irradiation effect on polymeric chains of epoxy adhesive, " *Polymers*, vol. 16, no. 9, 2024. https://www.mdpi.com/2073-4360/16/9/1202

[5] G. Tu, Y. Wu, and L. Yong,"Research on an effective defect localization technique with packaging," *China Integrated Circuit*, vol. 31, no. 9, pp. 79-84, 2022.

[6] P. Zelger, A. Brunner, B. Zelger, E. Willenbacher, S. H. Unterberger, R. Stalder, C. W. Huck, W. Willenbacher, and J. D. Pallua,"Deep learning analysis of mid-infrared microscopic imaging data for the diagnosis and classification of human lymphomas," *Journal of Biophotonics*, vol. 16, no. 11, pp. e202300015, 2023.

[7] J. Ding and S. Peng,"Photothermal microimaging: a non-invasive and high-resolution imaging technique," *Laser & Optoelectronics Progress*, vol. 60, no. 22, pp. 2200001, 2023.

[8] E. Staffa, V. Bernard, L. Kubicek, R. Vlachovsky, D. Vlk, V. Mornstein, A. Bourek, and R. Staffa,"Infrared thermography as option for evaluating the treatment effect of percutaneous transluminal angioplasty by patients with peripheral arterial disease," *Vascular*, vol. 25, no. 1, pp. 42-49, 2017.

[9] Q. Xia, J. Yin, Z. Guo, and J. X. Cheng,"Mid-infrared photothermal microscopy: Principle, instrumentation, and applications," *The Journal of Physical Chemistry B*, vol. 126, no. 43, pp. 8597-8613, 2022.

[10] M. Gao, W. Jin, X. Wang, and J. Yu,"Research on zero-point calibration method for optical microscanning thermal microscopic imaging system," *Acta Optica Sinica*, vol. 29, no. 8, pp. 2175-2182, 2009.

[11] M. Gao, W. Jin, X. Wang, C. Xu, and Y. Chen,"Design and implementation of a high-resolution optical microscanning thermal imaging system," *Journal of Instrumentation*, vol. 30, no. 5, pp. 1037-1041, 2009.

[12] M. Gao, H. Gu, C. Guan, and W. Wu,"Adaptive position calibration for thermal microscopic imaging system," *Acta Optica Sinica*, vol. 33, no. 1, pp. 0111002, 2013.

[13] M. Gao, J. Wang, W. Xu, and C. Guan, "High-resolution over-sampling reconstruction algorithm for a microscanning thermal microscope imaging system, " *Infrared Physics & Technology*, vol. 76, pp. 661-666, 2016. https://www.sciencedirect.com/science/article/pii/S1350449515300700

[14] M. Gao, J. Xu, A. Tan, Z. Zu, M. Yang, and J. Wang, "Error correction based on micro-scanning preprocessing for an optical micro-scanning thermal microscope imaging system, " *Infrared Physics & Technology*, vol. 83, pp. 252-258, 2017. https://www.sciencedirect.com/science/article/pii/S1350449516307010

[15] B. Zhang, M. Gao, P. L. Rosin, X. Sun, Q. Chang, C. Peng, Q. Yan, P. Chen, and Y. Shang, "Adaptive position calibration technique for an optical micro-scanning thermal microscope imaging system, " *Infrared Physics & Technology*, vol. 105, pp. 103186, 2020. https://www.sciencedirect.com/science/article/pii/S1350449519309272

[16] B. Zhang, M. Gao, P. Chen, Y. Shang, S. Li, Y. Bai, H. Liao, Z. Liu, and Z. Li, "Infrared and visible image fusion based on contrast enhancement guided filter and infrared feature decomposition, " *Infrared Physics & Technology*, vol. 127, pp. 104404, 2022. https://www.sciencedirect.com/science/article/pii/S1350449522003851

[17] M. Gao, Y. Bai, Y. Xie, B. Zhang, S. Li, and Z. Li, "SMC-SRGAN-lightning super-resolution algorithm based on optical micro-scanning thermal microscope image, " *Visual Computer*, vol. 40, no. 12, pp. 8441-8454, 2024.

[18] M. Gao, W. Jin, X. Wang, L. Wang, Y. Chen, and S. Du, "Study on the noise equivalent temperature difference and noise equivalent irradiation difference mathematical models for micro thermal imaging systems, " *Transactions of Beijing Institute of Technology*, vol. 1, no. 1, pp. 50-54, 2007. http://journal.bit.edu.cn/zr/cn/article/id/20070113

[19] B. Zhang, M. Gao, P. L. Rosin, X. Sun, Q. Chang, Q. Yan, and Y. Shang, "Research on performance evaluation and optimization theory for thermal

microscope imaging systems, " *Applied Sciences*, vol. 11, no. 13, 2021. https://www.mdpi.com/2076-3417/11/13/5897

[20] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution, " in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham, pp. 184-199, 2014.

[21] K. Prajapati, V. Chudasama, H. Patel, A. Sarvaiya, K. Upla, K. Raja, R. Ramachandra, and C. Busch, "Channel split convolutional neural network (chasnet) for thermal image super-resolution, " in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4363-4372, 2021.

[22] Y. Zou, L. Zhang, C. Liu, B. Wang, Y. Hu, and Q. Chen, "Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections, " *Optics and Lasers in Engineering*, vol. 146, pp. 106717, 2021. https://www.sciencedirect.com/science/article/pii/S0143816621001871

[23] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network, " in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105-114, 2017.

[24] K. Ning, Z. Zhang, K. Han, S. Han, and X. Zhang,"Multi-frame super-resolution algorithm based on a WGAN," *IEEE Access*, vol. 9, pp. 85839-85851, 2021.

[25] Y. Huang, Z. Jiang, Q. Wang, Q. Jiang, and G. Pang, "Infrared image super-resolution via heterogeneous convolutional wgan, " in *PRICAI 2021: Trends in Artificial Intelligence*, D. N. Pham, T. Theeramunkong, G. Governatori, and F. Liu, eds., Cham, pp. 461-472, 2021.

[26] S. Liang, K. Song, W. Zhao, S. Li, and Y. Yan, "DASR: Dual-attention transformer for infrared image super-resolution, " *Infrared Physics & Technology*, vol. 133, pp. 104837, 2023. https://www.sciencedirect.com/science/article/pii/S1350449523002955

[27] X. Yuan, B. Zhang, J. Zhou, C. Lian, Q. Zhang, and J. Yue, "Gradient residual attention network for infrared image super-resolution, " *Optics and Lasers in Engineering*, vol. 175, pp. 107998, 2024. https://www.sciencedirect.com/science/article/pii/S0143816623005274

[28] D. Lopez-Pajares, E. Rojas, M. P. Mishra, P. Jindgar, J. Alvarez-Horcajo, N. Manso, and J. Desmarais, "MDTA: An efficient, scalable and fast multiple disjoint tree algorithm for dynamic environments, " *Computer Communications*, vol. 229, pp. 107989, 2025. https://www.sciencedirect.com/science/article/pii/S0140366424003360

[29] M. Wan, G. Gu, W. Qian, K. Ren, Q. Chen, and X. Maldague, "Infrared image enhancement using adaptive histogram partition and brightness correction, " *Remote Sensing*, vol. 10, no. 5, 2018. https://www.mdpi.com/2072-4292/10/5/682

[30] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module, " in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, eds., Cham, pp. 3-19, 2018.

[31] L. Tang, J. Yuan, H. Zhang, X. Jiang, and J. Ma, "Piafusion: A progressive infrared and visible image fusion network based on illumination aware, " *Information Fusion*, vol. 83-84, pp. 79-92, 2022. https://www.sciencedirect.com/science/article/pii/S156625352200032X



Meijing Gao received the Ph.D. degree in optical engineering from Beijing Institute of Technology, Beijing, China, in 2008. She has been a Professor with the School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing, China, since 2021. Her current research interests include image and video processing, optoelectronic imaging, and photoelectric detection technologies.



Qitai Sun received the B.S. degree in Robotics Engineering from Hunan University, Changsha, China, in 2025. He is currently pursuing the M.S. degree with Beijing Institute of Technology, Beijing, China. His research interests include multimodal perception, multimodal object detection, and multi-object tracking.



Yufan Du received the M.S. degree from Beijing Institute of Technology, Beijing, China. His research interests include image fusion, image super-resolution reconstruction, optoelectronic imaging, and intelligent image processing.



Haozheng Yin received the M.S. degree from Yanshan University, Qinhuangdao, China. His research interests include image fusion, image super-resolution reconstruction, computer vision, and optoelectronic imaging.



Sibo Chen is currently pursuing the Ph.D. degree with Beijing Institute of Technology, Beijing, China. His research interests include image processing, computer vision, intelligent image analysis, and deep learning.



Huanyu Sun received the B.S. degree in Computer Science and Technology from Northeastern University at Qinhuangdao, Qinhuangdao, China, in 2024. She is currently pursuing the Ph.D. degree with Beijing Institute of Technology, Beijing, China. Her research interests include optoelectronic imaging target detection and recognition.



Xu Chen received the B.S. degree from Hebei University of Technology, Tianjin, China, in 2022. He is currently pursuing the M.S. degree with Beijing Institute of Technology, Beijing, China. His research interests include multimodal perception, multimodal object detection, and multi-object tracking.