

Dual-Input Frequency-Aware Network for High-Quality Thermal Image Super-Resolution

Priya Kansal*
 Couger Inc
 Tokyo, Japan

priya@couger.co.jp

Sabari Nathan*
 Couger Inc
 Tokyo, Japan

sabari@couger.co.jp

Abstract

Thermal image super-resolution (SR) faces critical challenges in recovering fine details from low-resolution inputs due to inherent infrared imaging limitations specifically low contrast between objects, thermal diffusion-induced blurring, and sparse high-frequency components. We propose a computationally efficient dual-input network that leverages frequency-domain analysis with attention-based feature fusion to address these $\times 2/\times 4$ SR challenges. The architecture first decomposes input images into low- and high-frequency components via Gaussian-blur differencing, explicitly isolating edge information often lost in thermal data. These components undergo parallel processing through a Self-Dual Calibrated Projection Attention (SD-CPA) module, enhancing structural coherence while suppressing noise. A multi-scale learning stage then combines three complementary pathways, followed by a Dual Attention Module for dynamic feature fusion before sub-pixel convolution-based reconstruction. Extensive experiments on multiple thermal datasets (PBVS 2020, FLIR, OSU, CVC09 and Thermal6 datasets) demonstrate that our approach outperforms existing state-of-the-art methods in both $\times 2$ and $\times 4$ upscaling scenarios, achieving superior PSNR and SSIM metrics, while maintaining computational efficiency. This work establishes a practical framework for deploying thermal SR in resource-constrained applications like night vision systems and medical thermography, effectively balancing computational efficiency with high-fidelity reconstruction.

1. Introduction

Thermal infrared (TIR) imaging captures heat radiation emitted by objects, enabling visibility in darkness and through visual obscurants such as smoke and fog [47]. This capability is invaluable for surveillance [27], autonomous

driving [32], industrial inspection [29], and medical diagnostics [10]. However, thermal imagery’s practical utility is often constrained by low spatial resolution, stemming from sensor technology limitations and manufacturing constraints [42]. Single-image super-resolution (SISR) techniques offer a solution by computationally reconstructing high-resolution images from low-resolution counterparts. Applying these techniques to thermal images presents unique challenges, as thermal imagery exhibits lower contrast, fewer textural details, more homogeneous regions, and lacks color information [13]. These characteristics complicate the super-resolution process, which typically relies on rich textural features and sharp edges. Deep learning has significantly advanced thermal image super-resolution. Traditional CNNs have been superseded by architectures tailored to thermal imagery, including Liu et al.’s [14] STEN-TSRN with edge enhancement and Cai et al.’s [30] feature representation learning model. Attention-based methods [26], [33], transformer [31] based architectures have also recently shown promise in thermal image enhancement. Despite these advances, existing methods struggle to balance computational efficiency with reconstruction quality, limiting deployment in resource-constrained environments [15]. We propose a lightweight dual-input thermal image super-resolution network to overcome these limitations. Our approach utilizes a high-frequency extraction module to separate edge information from the input image, followed by parallel processing of both frequency components through Self-Dual Calibrated Projection Attention. The architecture then employs a multi-scale feature learning framework with three specialized branches to capture diverse thermal characteristics, before a Dual Attention Module refines these features for upsampling. The main contributions of our model are:

- A dedicated high-frequency extraction module that isolates and preserves crucial edge and texture information typically lost in thermal super-resolution.
- Self-Dual Calibrated Projection Attention (SDCPA) to process low-frequency and high-frequency components

*These authors contributed equally to this work

separately, enabling more effective feature representation.

- A hierarchical multi-scale feature learning framework with parallel branches—Normalized Attention Feedforward Block (NAFBlock), Residual Dense Attention Block (RDAB), and Enhanced Cross-Covariant Attention (EXCA)—each capturing complementary thermal features at different scales.
- A Dual Attention Module that refines multi-scale features prior to upsampling.

Experiments on standard benchmarks, including PBVS 2020 Challenge [40], FLIR [1], OSU [8], CVC09 [35] and Thermal6 [38] datasets, demonstrate that our lightweight method consistently outperforms existing approaches in both $\times 2$ and $\times 4$ upscaling scenarios. The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 details our methodology, Section 4 presents the experimental setup, Section 5 offers results and analysis, and Section 6 concludes with a summary and future directions.

2. Related Works

Early approaches to thermal image super-resolution primarily relied on bicubic interpolation, which provided computationally efficient upsampling but often resulted in blurred details and artifacts along thermal boundaries. Despite its simplicity, bicubic interpolation struggles with preserving the critical edge information that represents temperature gradients, leading to smoothed boundaries and loss of thermal detail.

The advancements made by deep learning in Visible Image Super-Resolution (VISR) have driven similar progress in Thermal Image Super-Resolution (TISR). Alongside the FLIR dataset [1] and PBVS 2020 Challenge dataset [40], several other notable contributions have emerged in this domain. These datasets served as the benchmark and facilitated the transition to more advanced methods. For example, GAN-based approaches have been widely explored in TISR. FEGAN [53] introduced a feature-oriented enhanced GAN that preserves thermal signatures during super-resolution. Similarly, Dual-IRT-GAN [4] incorporated defect awareness for industrial inspection applications. Other GAN implementations such as [9], [46] have focused on thermal image enhancement and edge preservation, which are critical for maintaining temperature gradient boundaries. The GAN-based vision transformer [31] combined generative capabilities with long-range modeling, though these approaches often struggle with computational efficiency and training stability.

Furthermore, the multiscale spatio-temporal feature fusion network [31], [51] has shown promise by leveraging features across multiple spatial scales while exploiting temporal consistency when available. CIPPSRNet [48] introduced a camera internal parameters perception network us-

ing contrastive learning to understand the imaging process better, while the multi-path residual attention network [50] combined path diversity with attention mechanisms for enhanced feature representation.

Attention mechanisms have proven particularly effective for thermal imagery. Basic attention implementations [28] demonstrated that focusing on thermally significant regions improves reconstruction quality. More sophisticated approaches include flexible window-based self-attention transformers [17], which adaptively adjust attention based on thermal content, and DASR [24], which employs dual-attention transformers to capture cross-dimensional dependencies. Pyramidal edge-maps with attention-based guidance [12] have explicitly incorporated edge information through multi-scale structures, significantly enhancing boundary reconstruction.

Multi-scale feature extraction has emerged as another valuable strategy, with channel transformer approaches [37] establishing relationships between features at different scales. Lightweight architectures like TheriSURNet [6] have achieved competitive performance with reduced model complexity, addressing efficiency concerns for real-world deployment. The recurrent large kernel attention network [26] has shown effectiveness by combining recurrent processing with large receptive fields, while MSMA-Net [30] employed multiscale super-resolution enhancement with multilevel attention fusion. The edge-enhanced infrared image super-resolution model [14] specifically focused on preserving thermal boundaries through transformer architectures. Recent benchmarks [41], [40] have facilitated a rigorous evaluation of thermal super-resolution methods. Novel training strategies, including multi-level supervision models [19] and multi-scale backbones with multilevel supervision [33], have improved feature representation through guidance signals at various network depths. Lightweight approaches targeting adverse weather conditions [43] have been optimized for both efficiency and robustness.

Despite these advancements, a significant gap remains in balancing computational efficiency with reconstruction quality, particularly in preserving fine details and thermal signatures. Our proposed lightweight dual-input network addresses these limitations by separating and processing high-frequency and low-frequency components through specialized attention mechanisms and a parallel multi-scale feature learning framework.

3. Proposed Architecture

The proposed dual-input image enhancement network (Figure 1) reconstructs a high-resolution image X_{up} from a low-resolution input X_{low} while preserving fine textures and structural details. First, a high-frequency extraction module extracts important edge details by computing the difference between the original image and its Gaussian-blurred

version. The extracted low-frequency and high-frequency features are processed separately through a Self-Dual Calibrated Projection Attention (SDCPA) module, followed by element-wise fusion. The fused features are then fed into a multi-scale feature learning module, consisting of three parallel branches: a Normalized Attention Feedforward Block (NAFBlock), a Residual Dense Attention Block (RDAB), and an Enhanced Cross-Covariant Attention (EXCA) module. Finally, a Dual Attention Module refines the multi-scale features, which are then upsampled using subpixel convolution and combined with a residual connection for high-resolution reconstruction. Following subsections provides the details of the modules:

3.1. High-Frequency Feature Extraction

To retain edge information and texture details, we extract a high-frequency (HF) image from the input using a Gaussian blur operation:

$$X_{\text{blur}} = G(X_{\text{low}}, k) \quad (1)$$

where $G(\cdot, k)$ is a Gaussian smoothing filter with kernel size k . The high-frequency component is obtained by subtracting the blurred image from the original input:

$$X_{\text{HF}} = X_{\text{low}} - X_{\text{blur}} \quad (2)$$

Here, X_{HF} captures fine structural details lost during down-sampling. The model uses both X_{low} and X_{HF} as inputs to improve feature extraction and high-resolution reconstruction.

3.2. SDPCA for Initial Feature Extraction

The Self-Dual Calibrated Projection Attention (SDCPA) module enhances feature representation by integrating dual attention mechanisms through Pixel Attention Convolutions. It processes both low-resolution input features and high-frequency components to capture both spatial and contextual dependencies, ensuring improved local and global feature refinement.

Mathematically, given an input feature F_n , the first pathway applies a 1×1 convolution for channel reduction, followed by a 3×3 convolution to extract local patterns:

$$F'_{n-1} = \mathcal{F}_{1 \times 1}(F_{n-1}) \quad (3)$$

$$F'_n = \mathcal{F}_{3 \times 3}(F'_{n-1}) \quad (4)$$

Simultaneously, the second pathway applies a separate 1×1 convolution, followed by two pixel attention mechanisms to enhance feature importance. The first attention mechanism (PAConv) integrates sigmoid activation to emphasize key spatial regions:

$$F_{\text{PA1}} = \sigma(\mathcal{F}_{1 \times 1}(F_{n-1})) \cdot \mathcal{F}_{3 \times 3}(F_{n-1}) \quad (5)$$

The second attention mechanism (PAConv2) employs softmax activation, capturing complementary feature interactions:

$$F_{\text{PA2}} = \text{softmax}(\mathcal{F}_{1 \times 1}(F_{n-1})) + \mathcal{F}_{3 \times 3}(\mathcal{F}_{1 \times 1}(F_{n-1})) \quad (6)$$

These refined feature maps are concatenated and projected using a 1×1 convolution to merge multi-level representations:

$$F_{\text{SDCPA}} = \mathcal{F}_{1 \times 1}([F'_n, F_{\text{PA1}}, F_{\text{PA2}}]) \quad (7)$$

A residual connection is applied to ensure gradient flow and feature preservation:

$$F_n = F_{\text{SDCPA}} + F_{n-1} \quad (8)$$

This module is applied separately to both low-resolution input images and their high-frequency counterparts for feature extraction:

$$F_{\text{low}} = \mathcal{F}_{\text{SDCPA}}(X_{\text{low}}) \quad (9)$$

$$F_{\text{HF}} = \mathcal{F}_{\text{SDCPA}}(X_{\text{HF}}) \quad (10)$$

where $\mathcal{F}_{\text{SDCPA}}(\cdot)$ denotes the dual projection attention mechanism used for learning. These feature maps are then fused via concatenation to preserve structural and high-frequency details effectively:

$$F_{\text{concat}} = \text{Concat}(F_{\text{low}}, F_{\text{HF}}) \quad (11)$$

This fusion strategy enables the network to leverage both coarse and fine details, improving the robustness and quality of extracted features for subsequent multi-scale processing.

3.3. Multi-Residual Dense Attention Block (MRDAB) Overview

The Multi-Residual Dense Attention Block (MRDAB) is an advanced deep learning architecture designed to enhance feature extraction and representation learning by integrating residual connections, dense connectivity, and channel attention mechanisms. This structure is particularly effective in image processing tasks such as super-resolution, image denoising, and feature enhancement, where hierarchical and multi-scale features need to be efficiently captured.

Given an input feature map F_{in} , the residual dense connection in MRDAB consists of multiple convolutional layers with feature concatenation:

$$F_{l+1} = \sigma(W_l * [F_0, F_1, \dots, F_l] + b_l) \quad (12)$$

where W_l and b_l are the weights and biases of the l -th convolutional layer, $[F_0, F_1, \dots, F_l]$ represents the concatenation of feature maps from all previous layers, $\sigma(\cdot)$ is the

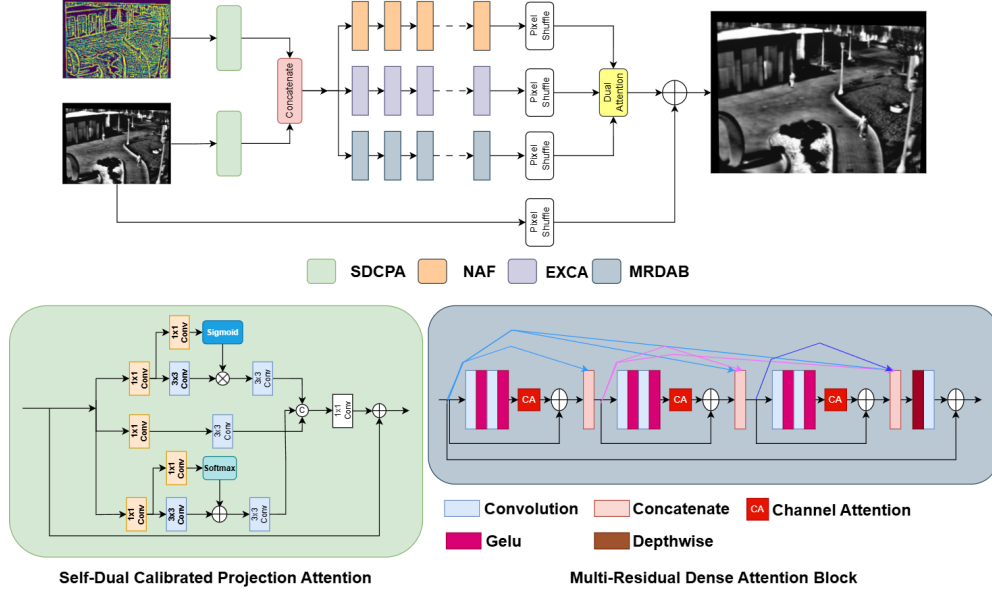


Figure 1. Architecture of the proposed network. The top diagram illustrates the proposed network for thermal image super-resolution. The bottom left depicts the Self-Dual Calibrated Projection Attention (SDCPA) block, while the bottom right shows the Multi-Residual Dense Attention Block (MRDAB).

activation function (GeLU), and $*$ denotes the convolution operation.

A residual connection is introduced to preserve low-level features:

$$F_{out} = F_{in} + \alpha \cdot F_{dense} \quad (13)$$

where α is a scaling factor, ensuring stable gradient flow and feature reuse.

To enhance feature selectivity, a channel attention mechanism is incorporated using global average pooling (GAP):

$$A_c = \sigma(W_c \cdot GAP(F) + b_c) \quad (14)$$

where:

$$GAP(F) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F(i, j) \quad (15)$$

computes the spatial average of feature maps, W_c and b_c are learnable parameters, and A_c represents the attention weight per channel.

The attention-enhanced output is given by:

$$F_{att} = A_c \odot F \quad (16)$$

where $\odot$ represents element-wise multiplication, allowing the model to focus on the most important channels dynamically.

By integrating residual connections, dense blocks, and attention mechanisms, the complete MRDAB output is formulated as:

$$F_{proj} = W_p * DwConv(\beta \cdot g([F_{RD_1}, F_{RD_2}, \dots, F_{RD_n}])) + b_p \quad (17)$$

$$F_{MRDAB} = F_{in} + F_{proj} \quad (18)$$

In these equations, F_{RD_i} represents the feature maps obtained from the i -th residual dense sub-block. The function $g(\cdot)$ performs feature aggregation by concatenating the feature maps and then applying a projection operation. The parameter β acts as a residual scaling factor that controls the contribution of the aggregated features. The operation $DwConv(F)$ applies spatial filtering independently across each channel. The term W_p denotes the weights associated with the pointwise convolution, while b_p represents the corresponding bias. The refined feature map before the final residual connection is denoted as F_{proj} . This residual connection enhances gradient flow, stabilizes training, and allows the network to maintain fine details while learning high-level representations. The MRDAB thus integrates multi-scale feature learning, attention-driven refinement, and efficient residual learning, making it well-suited for tasks requiring enhanced feature discrimination and detail preservation.

3.4. Hierarchical Multi-Scale Feature Extraction and Upsampling with Dual Attention

The multi-scale feature extraction module processes the concatenated feature map F_{concat} through three parallel branches, each responsible for learning distinct feature representations. Instead of performing direct fusion within the feature extraction blocks, the concatenated feature map is used as input to the sub-feature extraction function. This ensures independent yet complementary feature learning, leveraging Normalized Attention Feedforward (NAF) Blocks [3], MRDAB [34], and Enhanced Cross-Covariant Attention (EXCA) [18] Modules.

3.4.1. Parallel Feature Extraction

Given the input feature map:

$$F_{\text{concat}} = \text{Concat}(F_{\text{low}}, F_{\text{HF}}) \quad (19)$$

The extracted features are computed as follows:

$$F_{\text{NAF}}, F_{\text{MRDAB}}, F_{\text{EXCA}} = \mathcal{F}_{\text{sub_feature}}(F_{\text{concat}}) \quad (20)$$

Each branch learns a unique feature representation:

$$F_{\text{NAF}} = \mathcal{F}_{\text{NAF}}(F_{\text{concat}}) \quad (21)$$

$$F_{\text{MRDAB}} = \mathcal{F}_{\text{MRDAB}}(F_{\text{concat}}) \quad (22)$$

$$F_{\text{EXCA}} = \mathcal{F}_{\text{EXCA}}(F_{\text{concat}}) \quad (23)$$

where $\mathcal{F}_{\text{NAF}}(\cdot)$ applies Normalized Attention Feedforward Blocks (NAFBlockSR) to refine attention-based representations, $\mathcal{F}_{\text{MRDAB}}(\cdot)$ applies MRDAB for hierarchical feature learning, and $\mathcal{F}_{\text{EXCA}}(\cdot)$ applies Enhanced Cross-Covariant Attention Modules (EXCA) to improve spatial cross-attention.

3.4.2. Upsampling and Feature Fusion

Each extracted feature is upsampled independently using subpixel convolution to ensure high-resolution feature reconstruction:

$$F_{\text{NAF-up}} = \mathcal{F}_{\text{subpixel}}(F_{\text{NAF}}, s) \quad (24)$$

$$F_{\text{MRDAB-up}} = \mathcal{F}_{\text{subpixel}}(F_{\text{MRDAB}}, s) \quad (25)$$

$$F_{\text{EXCA-up}} = \mathcal{F}_{\text{subpixel}}(F_{\text{EXCA}}, s) \quad (26)$$

where $\mathcal{F}_{\text{subpixel}}(\cdot, s)$ represents the subpixel convolution up-sampling function, and s is the scaling factor used for resolution enhancement. The upsampled feature maps are then concatenated to integrate multi-scale information:

$$F_{\text{concat-up}} = \text{Concat}(F_{\text{NAF-up}}, F_{\text{MRDAB-up}}, F_{\text{EXCA-up}}) \quad (27)$$

To further refine feature representations, a dual attention module is applied:

$$F_{\text{att}} = \mathcal{F}_{\text{DualAttention}}(F_{\text{concat-up}}) \quad (28)$$

where $\mathcal{F}_{\text{DualAttention}}(\cdot)$ applies channel attention and spatial attention to enhance feature discrimination. A residual connection is incorporated to maintain fine details:

$$F_{\text{final}} = F_{\text{att}} + F_{\text{concat-up}} \quad (29)$$

3.4.3. Final Output with Residual Learning

A final convolutional layer is applied to reconstruct the output:

$$X_{\text{up}} = \mathcal{F}_{\text{conv}}(F_{\text{final}}) \quad (30)$$

where $\mathcal{F}_{\text{conv}}(\cdot)$ applies a 5×5 convolution for output refinement. To further enhance structural preservation, a final residual connection is applied between the upsampled low-resolution input X_{low} and the reconstructed high-resolution image X_{up} :

$$X_{\text{final}} = X_{\text{up}} + \mathcal{F}_{\text{subpixel}}(X_{\text{low}}, s) \quad (31)$$

where $\mathcal{F}_{\text{subpixel}}(X_{\text{low}}, s)$ upsamples the original low-resolution input. The residual connection ensures low-frequency content is preserved, reducing over-smoothing effects.

4. Experimental Setup

4.1. Dataset

We utilized several datasets to validate our proposed model's efficacy. Data preprocessing methods were synchronized with previous research to facilitate meaningful comparisons. Details are given below:

PBVS 2020: Our network training utilized the dataset shared in PBVS 2020 [39], specifically the Flir (HR) and Axis (MR) components arranged according to the PBVS-2020 challenge specifications. This dataset consists of 951 HR images for training, 50 images for validation, and 10 images for testing. For the x4 resolution task, the low-resolution images are obtained by downsampling the HR images by a factor of 4 and then adding Gaussian noise with a mean of 0 and Standard Deviation of 10. For x2 resolution, due to the different camera system in input and output images, SURF feature based registration is applied for proper alignment as outlined in [36] for training purposes. However, for testing purposes the HR images are downsampled by a factor of 2 and used as LR.

FLIR dataset: The FLIR [1] dataset consists of 2,000 training images and 200 validation images.

CVC09 dataset: CVC09 dataset [35] offers 4,000 training images and 400 test images.

OSU dataset: OSU [8] collection comprises 270 images for training purposes and 27 for testing.

When developing x2 and x4 models for FLIR, CVC09 and OSU dataset, we downsampled high-resolution images randomly using bicubic or Lanczos method to obtain the input images in accordance with [26].

Thermal6: To evaluate the resilience of our proposed architecture and trained model, we conducted additional testing on the Thermal6 [38] dataset. This dataset was captured using a Tau2 camera with a 640 x 512 pixel resolution and consists of 101 images encompassing both indoor and outdoor environments during daytime and nighttime conditions. This dataset is used for testing purpose to evaluate the generalization capacity of the proposed method.

4.2. Training Details

The Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$, is used for model optimization. The initial learning rate is set at $\times 10^{-3}$ and reduced after every 4th epoch if PSNR does not improve. All the experiments are conducted on the Colab A100 GPU. The model is set to train for 1000 epochs. However, training is stopped if PSNR does not improve for 30 epochs.

4.3. Loss Function

The proposed loss function, denoted as $\mathcal{L}_{FV2}$, is formulated as:

$$\mathcal{L}_{FV2} = 0.75\mathcal{L}_{SSIM} + \mathcal{L}_{MAE} + 0.75\mathcal{L}_{Sobel} + \epsilon \quad (32)$$

where $\mathcal{L}_{SSIM}$ represents the **Structural Similarity Index (SSIM) loss**, ensuring the preservation of perceptual quality and structural details in the reconstructed image. It is defined as:

$$\mathcal{L}_{SSIM} = 1 - SSIM(Y, \hat{Y}) \quad (33)$$

The term $\mathcal{L}_{MAE}$ refers to the **Mean Absolute Error (MAE)**, which minimizes the absolute difference between the pixel intensities of the predicted image $\hat{Y}$ and the ground truth image Y , improving pixel-wise fidelity. It is given by:

$$\mathcal{L}_{MAE} = \frac{1}{N} \sum_{i=1}^N |Y_i - \hat{Y}_i| \quad (34)$$

The **Sobel edge loss** ($\mathcal{L}_{Sobel}$) is incorporated to enhance edge details by computing the absolute difference between the Sobel gradients of $\hat{Y}$ and Y , ensuring sharpness and texture preservation:

$$\mathcal{L}_{Sobel} = \frac{1}{N} \sum_{i=1}^N |\nabla_{Sobel}(\hat{Y}_i) - \nabla_{Sobel}(Y_i)| \quad (35)$$

where $\nabla_{Sobel}(Y)$ represents the Sobel gradient of the image Y , and N is the total number of pixels.

The coefficient **0.75** balances the contributions of **SSIM loss** and **Sobel loss** to prevent overemphasis on either structural integrity or edge fidelity. Finally, ϵ is a small constant added for numerical stability, preventing division by zero.

This formulation integrates **perceptual quality, pixel-wise accuracy, and edge preservation**, making it well-suited for tasks such as **image super-resolution and enhancement**.

5. Results and Analysis

5.1. Quantitative Analysis

The proposed method was quantitatively evaluated by comparing it with existing state-of-the-art methods using PSNR and SSIM metrics. To ensure comprehensive benchmarking, we included leading methods from the RGB spectrum, although we did not train these methods by ourselves but rather utilized results from existing studies after careful verification of training and testing data processing protocols.

For the PBVS 2020 dataset, all these comparative results were adopted directly from [36]. Table 1 presents the PSNR and SSIM values obtained using these methods on the PBVS-2020 validation dataset for two specific tracks: Track 1 (thermal image super-resolution using upscaling factor $\times 4$ with noise elimination) and Track 2. For Track 2, due to the absence of true registered LR-HR pairs, we utilize LR images generated through bicubic down-sampling as described in [36]. We also include CIPPSR [48] in our comparison, reporting its performance on 50 validation images. However the 60 percent central cropping mechanism is not followed as specified by the authors. Table 1 represents the comparative results on PBVS 2020 dataset.

Method	PSNR/SSIM (Flir x4)	PSNR/SSIM (Axis x2)
Bicubic	32.66/0.8625	34.74/0.9200
SRResNet [21]	33.12/0.9018	33.66/0.9229
MSRN [22]	34.47/0.9076	36.96/0.9471
SRFeat [26]	34.12/0.9007	–
EDSR [25]	34.48/0.9068	36.91/0.9466
RCAN [52]	34.42/0.9072	36.67/0.9438
TEN [5]	33.62/0.8910	36.10/0.9392
CNN-IR [2]	33.77/0.8938	36.66/0.9438
PBVS-2020 winner [39]	34.49/0.9073	34.17/0.9116
TherISuRNet [7]	34.49/0.9101	36.76/0.9450
Chasnet [36]	34.90/0.9134	37.49/0.9518
CIPPSRNet [48]	35.08/0.9166	37.78/0.9543
ProposedNet	36.09/0.9534	38.45/0.9785

Table 1. Quantitative performance of SoTA SISR and TISR models for PBVS 2020 dataset.

For the FLIR, OSU, and CV09 datasets, we adopted results from [26]. Table 2 and Table 3 display the PSNR and SSIM metrics for these methods as reported in [26]. Similar to the PBVS 2020 dataset, our method outperforms the PSNR and SSIM metrics on these datasets for both x4 and x2 scale. Despite using more parameters than RLKA-Net [26], our model’s substantial quality improvements justify this modest increase, while still maintaining significantly lower complexity than all other heavyweight models.

Method	Params	OSU [8]	FLIR [1]	CVC09 [35]
Bicubic	-	29.35/0.6717	29.72/0.6981	38.77/0.9116
FSRCNN [11]	56 K	30.98/0.6974	31.93/0.7461	39.65/0.9341
EDSR [25]	1518 K	32.34/0.7445	32.57/0.7803	41.70/0.9472
DRCN [20]	1774 K	32.23/0.7387	32.31/0.7714	41.58/0.9401
DRRN [44]	297 K	32.29/0.7397	32.53/0.7726	41.63/0.9415
MemNet [45]	667 K	32.21/0.7382	33.46/0.7765	41.67/0.9459
RCAN [52]	15472 K	32.46/0.7457	33.97/0.8163	42.06/0.9509
IMDN [16]	715 K	32.58/0.7513	34.67/0.8066	41.89/0.9458
MAN [49]	8712 K	32.80/0.7606	34.89/0.8419	42.23/0.9541
SwinIR [23]	11942 K	32.93/0.7618	34.82/0.8417	42.21/0.9537
RLKA Net [26]	167 K	33.11/0.7621	34.94/0.8457	42.45/0.9665
Proposed Net	372k	34.18/0.8482	36.14/9007	42.96/0.9794

Table 2. Quantitative performance of SoTA SISR and TISR models for OSU, FLIR and CVC09 datasets for x2 scale.

For the x4 scale, our proposed model significantly outperforms all existing state-of-the-art methods. 3 Similarly,

Method	Params	OSU [8]	FLIR [1]	CVC09 [35]
Bicubic	-	27.85/0.5744	28.01/0.5871	34.04/0.8429
FSRCNN [11]	56 K	29.16/0.5998	28.89/0.6016	34.62/0.8673
EDSR [25]	1518 K	30.57/0.6398	29.78/0.6207	36.55/0.9001
DRCN [20]	1774 K	30.49/0.6346	29.63/0.6192	36.59/0.9016
DRRN [44]	297 K	30.52/0.6353	29.74/0.6201	36.67/0.9011
MemNet [45]	667 K	30.39/0.6398	29.57/0.6169	36.30/0.8939
RCAN [52]	15472 K	30.62/0.6407	30.49/0.6385	36.97/0.9035
IMDN [16]	882 K	30.49/0.6371	30.53/0.6419	36.42/0.8996
MAN [49]	1537 K	30.98/0.6478	31.07/0.6489	37.21/0.9262
SwinIR [23]	14367 K	30.70/0.6439	30.86/0.6472	37.26/0.9287
RLKA-Net [26]	193 K	31.06/0.6504	31.12/0.6506	37.31/0.9294
Proposed Net	372k	32.16/0.7831	33.636/0.8318	38.72/0.9378

Table 3. Quantitative performance of SoTA SISR and TISR models for OSU, FLIR and CVC09 datasets for x4 scale.

for the Thermal6 dataset, the results of SOTA methods are adopted from [33]. The model trained on the PBVS 2020 dataset is used for testing Thermal6 dataset. The results in Table 4 confirm our architecture’s superior capability at more challenging higher upscaling factors like x4 where detail preservation is critical. The superior results on x4 dataset proves the generalization capacity of our proposed method.

5.2. Qualitative Analysis

Figure 2 demonstrates the effectiveness of the proposed super-resolution method across multiple thermal datasets.

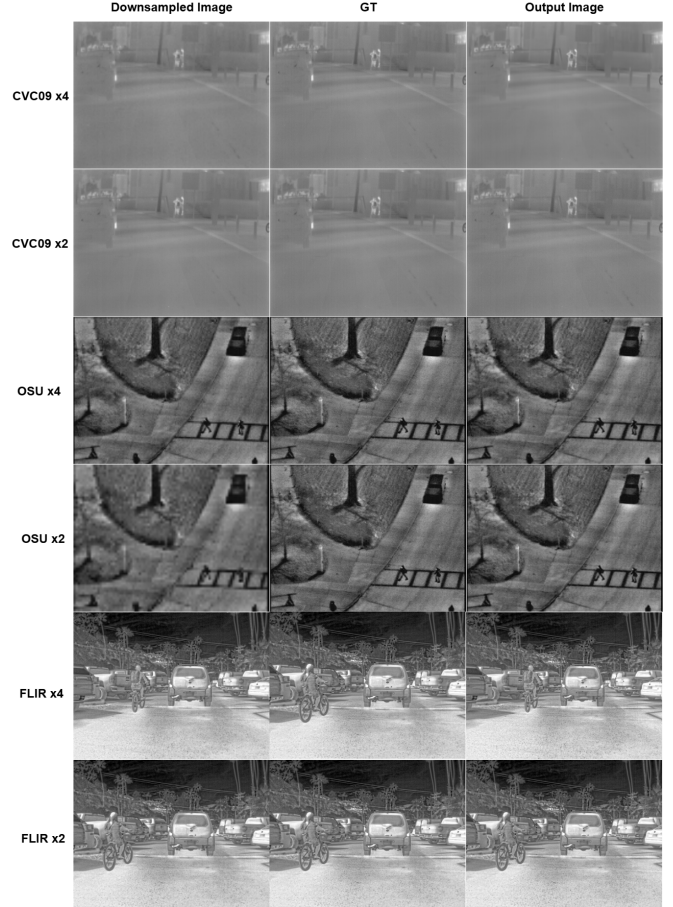


Figure 2. Comparison of thermal image super-resolution performance across different datasets. Each row represents a dataset (CVC09, OSU, FLIR) with downsampling factors ($\times 2$ and $\times 4$), while the columns depict the downsampled input, ground truth (GT), and super-resolved output.

Scale	Bicubic Model	TISR [21]	MLSM [10]	MSBM [33]	ProposedNet
x2 scale	39.59	41.24	40.89	39.82	40.328
x4 scale	34.98	37.85	37.6	37.47	38.2453

Table 4. Table 4: Quantitative performance of SoTA SISR and TISR models for Thermal6 dataset.

The output images exhibit enhanced structural details and reduced blurring compared to the downsampled inputs, closely matching the ground truth (GT). Notably, finer textures and object boundaries, especially in the OSU and FLIR datasets, are well-preserved, indicating the robustness of the approach in recovering high-frequency details.

Figure 3 illustrates a qualitative comparison of super-resolution results on thermal images from the FLIR and CVC09 datasets. The proposed method effectively restores fine details and preserves structural consistency, as highlighted in the red-boxed regions. Compared to ex-

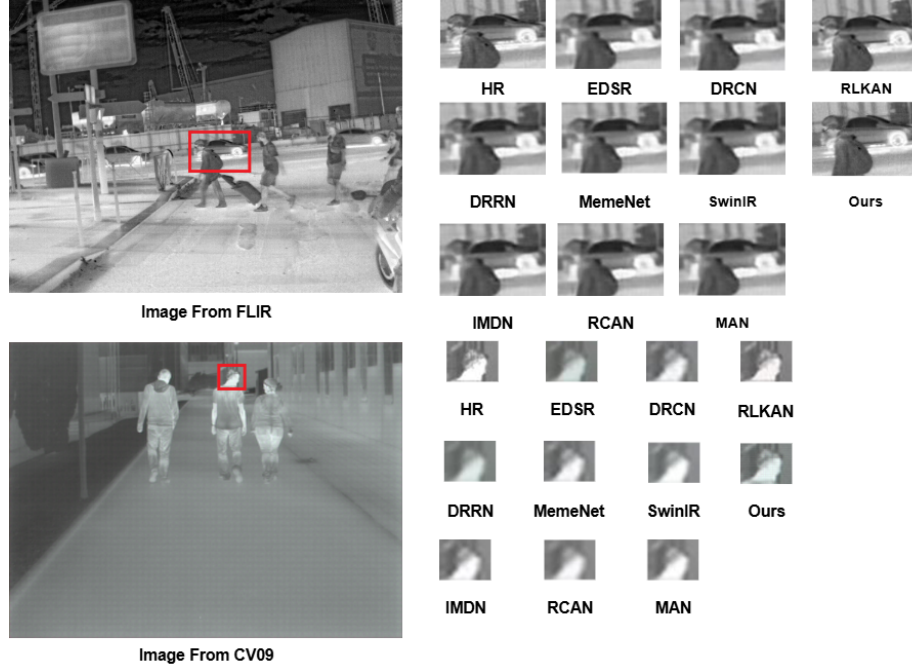


Figure 3. Qualitative comparison of super-resolution results on thermal images from the FLIR and CVC09 datasets. The proposed method preserves structural consistency, enhances contrast, and restores fine details better than state-of-the-art methods.

isting state-of-the-art approaches, such as EDSR, SwinIR, and RCAN, our method produces sharper reconstructions with reduced blurring and enhanced contrast, particularly in high-frequency regions. Notably, the edges of objects, including human silhouettes and vehicle contours, are well-defined in our results, whereas competing methods suffer from excessive smoothing or texture loss. The ability of our approach to retain intricate thermal patterns while mitigating noise artifacts underscores its advantage in real-world thermal imaging applications.

5.3. Ablation Study

Table 5 presents an ablation study on the OSU dataset ($\times 2$ scale) evaluating the contributions of the Self-Dual Calibrated Projection Attention (SDCPA) and Multi-Residual Dense Attention Block (MRDAB). The results indicate that incorporating both modules significantly improves performance, achieving the highest PSNR (34.18 dB) and SSIM (0.8482). When either module is removed, performance degrades, with the absence of MRDAB leading to a sharper decline in PSNR (30.6487 dB) and SSIM (0.7219). This highlights the complementary roles of SDCPA and MRDAB in enhancing structural details and restoring high-frequency information in super-resolution tasks.

6. Conclusion

This paper presents a novel lightweight dual-input thermal image super-resolution network that effectively addresses

SDCPA	MRDAB	PSNR (dB)	SSIM
Yes	No	31.1861	0.7318
No	Yes	30.4226	0.7149
No	No	30.6487	0.7219
Yes	Yes	34.18	0.8482

Table 5. Ablation study on the OSU dataset ($\times 2$ scale). The impact of SDCPA and MRDAB is evaluated based on PSNR (dB) and SSIM. The best results are obtained when both modules are incorporated.

the unique challenges inherent to thermal imagery. Our approach preserves critical edge information and thermal signatures by decomposing and processing input images through specialized attention mechanisms. Experiments across multiple datasets demonstrate that our method consistently outperforms existing state-of-the-art approaches in both $\times 2$ and $\times 4$ upscaling scenarios, achieving superior metrics while maintaining lower computational complexity. This work establishes a practical framework for deploying thermal SR in resource-constrained real-world applications such as night vision systems and medical thermography. By effectively balancing computational efficiency with high-fidelity reconstruction, our approach addresses an important gap in the field of thermal image enhancement. Future work will explore the integration of temporal information for video super-resolution and the adaptation of this architecture to other thermal imaging applications such as detection and segmentation tasks.

References

- [1] Flir adas dataset. <https://www.flir.com/oem/adas/adasdataset-form/>. Description of the ADAS dataset. 2, 6, 7
- [2] Purbaditya Bhattacharya, Jörg Riechen, and Udo Zölzer. Infrared image enhancement in maritime environment with convolutional neural networks. In *VISIGRAPP (4: VISAPP)*, pages 37–46, 2018. 6
- [3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 5
- [4] Liangliang Cheng and Mathias Kersemans. Dual-irt-gan: A defect-aware deep adversarial network to perform super-resolution tasks in infrared thermographic inspection. *Composites Part B: Engineering*, 247:110309, 2022. 2
- [5] Yukyung Choi, Namil Kim, Soonmin Hwang, and In So Kweon. Thermal image enhancement using convolutional neural network. In *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 223–230. IEEE, 2016. 6
- [6] Vishal Chudasama, Heena Patel, Kalpesh Prajapati, Kishor P Upla, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. Therisurnet-a computationally efficient thermal image super-resolution network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 86–87, 2020. 2
- [7] Vishal Chudasama, Heena Patel, Kalpesh Prajapati, Kishor P Upla, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. Therisurnet-a computationally efficient thermal image super-resolution network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 86–87, 2020. 6
- [8] James W Davis and Mark A Keck. A two-stage template approach to person detection in thermal imagery. In *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1*, pages 364–369. IEEE, 2005. 2, 6, 7
- [9] Shashikant Deepak, Sanuj Sahoo, and Dipti Patra. Super-resolution of thermal images using gan network. In *2021 Advanced Communication Technologies and Signal Processing (ACTS)*, pages 1–5. IEEE, 2021. 2
- [10] Nilanjan Dey, Amira S Ashour, and Afnan S Althoupety. Thermal imaging in medical science. *Recent Advances in Applied Thermal Imaging for Industrial Applications*, pages 87–117, 2017. 1
- [11] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*, pages 391–407. Springer, 2016. 7
- [12] Honey Gupta and Kaushik Mitra. Pyramidal edge-maps and attention based guided thermal super-resolution. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 698–715. Springer, 2020. 2
- [13] Yunze He, Baoyuan Deng, Hongjin Wang, Liang Cheng, Ke Zhou, Siyuan Cai, and Francesco Ciampa. Infrared machine vision and infrared thermography with deep learning: A review. *Infrared physics & technology*, 116:103754, 2021. 1
- [14] Lei Hu, Long Hu, and Minghui Chen. Edge-enhanced infrared image super-resolution reconstruction model under transformer. *Scientific Reports*, 14(1):15585, 2024. 1, 2
- [15] Yongsong Huang, Tomo Miyazaki, Xiaofeng Liu, and Shinichiro Omachi. Infrared image super-resolution: Systematic review, and future trends. *arXiv preprint arXiv:2212.12322*, 2022. 1
- [16] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia*, pages 2024–2032, 2019. 7
- [17] Hongcheng Jiang and ZhiQiang Chen. Flexible window-based self-attention transformer in thermal image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3076–3085, 2024. 2
- [18] Siyuan Jiang, Senyan Xu, and Xingfu Wang. Rbsformer: Enhanced transformer network for raw image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6479–6488, 2024. 5
- [19] Priya Kansal and Sabari Nathan. A multi-level supervision model: A novel approach for thermal image super resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 94–95, 2020. 2
- [20] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016. 7
- [21] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 6
- [22] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *Proceedings of the European conference on computer vision (ECCV)*, pages 517–532, 2018. 6
- [23] Jingyun Liang, Jiezhong Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 7
- [24] ShuBo Liang, Kechen Song, Wenli Zhao, Song Li, and Yunhui Yan. Dasr: Dual-attention transformer for infrared image super-resolution. *Infrared Physics & Technology*, 133: 104837, 2023. 2
- [25] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 6, 7

- [26] Gangping Liu, Shuaijun Zhou, Xiaxu Chen, Wenjie Yue, and Jun Ke. Recurrent large kernel attention network for efficient single infrared image super-resolution. *IEEE Access*, 12:923–935, 2023. 1, 2, 6, 7
- [27] Jingjing Liu, Shaoting Zhang, Shu Wang, and Dimitris N Metaxas. Multispectral deep neural networks for pedestrian detection. *arXiv preprint arXiv:1611.02644*, 2016. 1
- [28] Qing-Ming Liu, Rui-Sheng Jia, Yan-Bo Liu, Hai-Bin Sun, Jian-Zhi Yu, and Hong-Mei Sun. Infrared image super-resolution reconstruction by using generative adversarial network with an attention mechanism. *Applied Intelligence*, 51: 2018–2030, 2021. 2
- [29] Zheng Liu, Hiroyuki Ukida, Pradeep Ramuhalli, and Kurt Niel. Integrated imaging and vision techniques for industrial inspection. *Advances in Computer Vision and Pattern Recognition*, 2015. 1
- [30] Tianlei Ma, Hao Wang, Jing Liang, Jinzhu Peng, Qi Ma, and Zhiqiang Kai. Msma-net: An infrared small target detection network by multiscale super-resolution enhancement and multilevel attention fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–20, 2023. 1, 2
- [31] Mohamed Amine Marnissi and Abir Fathallah. Gan-based vision transformer for high-quality thermal image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 817–825, 2023. 1, 2
- [32] Farzeen Munir, Shoaib Azam, Muhammd Aasim Rafique, Ahmad Muqeem Sheri, Moongu Jeon, and Witold Pedrycz. Exploring thermal images for object detection in underexposure regions for autonomous driving. *Applied Soft Computing*, 121:108793, 2022. 1
- [33] Sabari Nathan and Priya Kansal. Leveraging multi scale backbone with multilevel supervision for thermal image super resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4332–4338, 2021. 1, 2, 7
- [34] Sabari Nathan and Priya Kansal. Skeletonnetv2: A dense channel attention blocks for skeleton extraction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2142–2149, 2021. 5
- [35] Karol Piniarski and Paweł Pawłowski. Multi-branch classifiers for pedestrian detection from infrared night and day images. In *2016 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages 248–253. IEEE, 2016. 2, 6, 7
- [36] Kalpesh Prajapati, Vishal Chudasama, Heena Patel, Anjali Sarvaiya, Kishor P Upla, Kiran Raja, Raghavendra Ramachandra, and Christoph Busch. Channel split convolutional neural network (chasnet) for thermal image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4368–4377, 2021. 5, 6
- [37] Raghunath Sai Puttagunta, Birendra Kathariya, Zhu Li, and George York. Multi-scale feature fusion using channel transformers for guided thermal image super resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3086–3095, 2024. 2
- [38] Rafael E Rivadeneira, Patricia L Suárez, Angel D Sappa, and Boris X Vintimilla. Thermal image superresolution through deep convolutional neural network. In *Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, ON, Canada, August 27–29, 2019, Proceedings, Part II 16*, pages 417–426. Springer, 2019. 2, 6
- [39] Rafael E. Rivadeneira, Angel D. Sappa, Boris X. Vintimilla, Lin Guo, Jiankun Hou, Armin Mehri, Parichehr Behjati Ardakani, Heena Patel, Vishal Chudasama, Kalpesh Prajapati, Kishor P. Upla, Raghavendra Ramachandra, Kiran Raja, Christoph Busch, Feras Almasri, Olivier Debeir, Sabari Nathan, Priya Kansal, Nolan Gutierrez, Bardia Mojra, and William J. Beksi. Thermal image super-resolution challenge - pbvs 2020. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020. 5, 6
- [40] Rafael E Rivadeneira, Angel D Sappa, Boris X Vintimilla, Sabari Nathan, Priya Kansal, Armin Mehri, Parichehr Behjati Ardakani, Anurag Dalal, Aparna Akula, Darshika Sharma, et al. Thermal image super-resolution challenge-pbvs 2021. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4359–4367, 2021. 2
- [41] Rafael E Rivadeneira, Angel D Sappa, Chenyang Wang, Junjun Jiang, Zhiwei Zhong, Peilin Chen, and Shiqi Wang. Thermal image super-resolution challenge results-pbvs 2024. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3113–3122, 2024. 2
- [42] Antoni Rogalski. Infrared detectors: status and trends. *Progress in Quantum Electronics*, 27(2-3):59–210, 2003. 1
- [43] Pranjay Shyam and HyunJin Yoo. Lightweight thermal super-resolution and object detection for robust perception in adverse weather conditions. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7471–7482, 2024. 2
- [44] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017. 7
- [45] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE international conference on computer vision*, pages 4539–4547, 2017. 7
- [46] Nguyen Duc Thuan, Trinh Phuong Dong, Bui Quang Manh, Hoang Anh Thai, Tran Quang Trung, and Hoang Si Hong. Edge-focus thermal image super-resolution using generative adversarial network. In *2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, pages 1–6. IEEE, 2022. 2
- [47] Michael Vollmer and Klaus-Peter Möllmann. *Infrared thermal imaging: fundamentals, research and applications*. John Wiley & Sons, 2018. 1
- [48] Kai Wang, Qigong Sun, Yicheng Wang, Huiyuan Wei, Chonghua Lv, Xiaolin Tian, and Xu Liu. Cippsnet: A camera internal parameters perception network based contrastive learning for thermal image super-resolution. In *Proceedings*

- of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 342–349, 2022. [2](#), [6](#)
- [49] Yan Wang, Yusen Li, Gang Wang, and Xiaoguang Liu. Multi-scale attention network for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5950–5960, 2024. [7](#)
- [50] Haikun Zhang, Yueli Hu, Ming Yan, and Bin Ma. Thermal image super-resolution via multi-path residual attention network. *Signal, Image and Video Processing*, 17(5):2073–2081, 2023. [2](#)
- [51] Wenhui Zhang, Xiubao Sui, Guohua Gu, Qian Chen, and Heyang Cao. Infrared thermal imaging super-resolution via multiscale spatio-temporal feature fusion network. *IEEE Sensors Journal*, 21(17):19176–19185, 2021. [2](#)
- [52] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. [6](#), [7](#)
- [53] Linzhen Zhu, Renjie Wu, Boon-Giin Lee, Lionel Nkenyer-eye, Wan-Young Chung, and Gen Xu. Fegan: A feature-oriented enhanced gan for enhancing thermal image super-resolution. *IEEE Signal Processing Letters*, 31:541–545, 2024. [2](#)