

Article

Not peer-reviewed version

Thermal Single-Image Super-Resolution: A Review and Demosaicking-Inspired Deep Fusion Framework

[Armine Bayramyan](#)*, [Sos Agaian](#), [Shreyas Kamath K. M.](#)

Posted Date: 16 September 2026

Keywords: thermal imaging; infrared image super-resolution; review; demosaicking; datasets; image quality assessment; adaptive fusion; hypercomplex; quaternion; RLKA-Net



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Thermal Single-Image Super-Resolution: A Review and Demosaicking-Inspired Deep Fusion Framework

Armine Bayramyan ^{1,*}, Sos Agaian ² and Shreyas Kamath K. M. ³

¹ Mathematics and Mechanics Department, Yerevan State University, Yerevan, Armenia

² Computer Science Department, Graduate Center, College of Staten Island (CSI), City University of New York, New York, NY, USA

³ Independent Researcher, Sunnyvale, USA

* Correspondence: armine.bayramyan@edu.ysu.am

Abstract

Thermal image super-resolution (SR) addresses the limited spatial resolution of infrared sensors, improving the visibility of fine details and supporting applications such as inspection, monitoring, and detection. This capability is particularly valuable in resource-constrained environments where sensor cost, storage, computational resources, and communication bandwidth limit high-resolution thermal imaging. Unlike interpolation, which estimates missing pixels from local image structure, SR aims to recover high-frequency spatial details while mitigating degradations such as blur, noise, and aliasing. This paper presents a comprehensive review of thermal-image SR, covering thermal imaging characteristics, publicly available datasets, evaluation metrics, and representative classical and deep learning methods. It also introduces a demosaicking-inspired deep fusion framework for single-image thermal SR. The proposed approach generates complementary reconstruction candidates using three conventional methods: bilinear interpolation, Malvar–He–Cutler symmetry-based demosaicking, and Menon-based interpolation. A lightweight FusionNet combines these outputs with an RLKA-Net reconstruction by learning spatially adaptive, pixel-wise fusion weights. Experiments on a thermal building-inspection dataset of 306 images demonstrate that FusionNet achieves a PSNR of 37.69 dB and an SSIM of 0.9746, outperforming the individual classical and deep-learning methods. Additionally, a quaternion-based FusionNet reduces model parameters by approximately 75% while maintaining comparable performance, and a higher-capacity quaternion variant slightly surpasses the real-valued model. The paper concludes by discussing remaining challenges and future directions for robust thermal-image SR in real-world applications.

Keywords: thermal imaging; infrared image super-resolution; review; demosaicking; datasets; image quality assessment; adaptive fusion; hypercomplex; quaternion; RLKA-Net

1. Introduction

Thermal imaging has become an important sensing technology in building inspection, structural health monitoring, industrial monitoring, surveillance, autonomous navigation, and security applications. Unlike conventional RGB cameras, thermal cameras capture infrared radiation emitted by objects. Therefore, thermal images can reveal temperature-related information that is not directly visible in the visible spectrum, such as heat leakage, moisture accumulation, insulation failure, and abnormal thermal patterns.

Thermal imaging is particularly important for inspection tasks because many defects are not visible in the RGB domain. Regions that appear visually homogeneous in RGB images may contain hidden thermal patterns caused by moisture, heat leakage, insulation failure, or structural inconsistency [1,2].

As illustrated in Figure 1 and summarized in Table 1, thermal imaging provides complementary information to RGB imaging by revealing temperature distributions and hidden defects that are often invisible in the visible spectrum. These characteristics make thermal imaging particularly valuable

for building inspection, where moisture intrusion, heat leakage, insulation failure, and structural inconsistencies must be accurately identified.

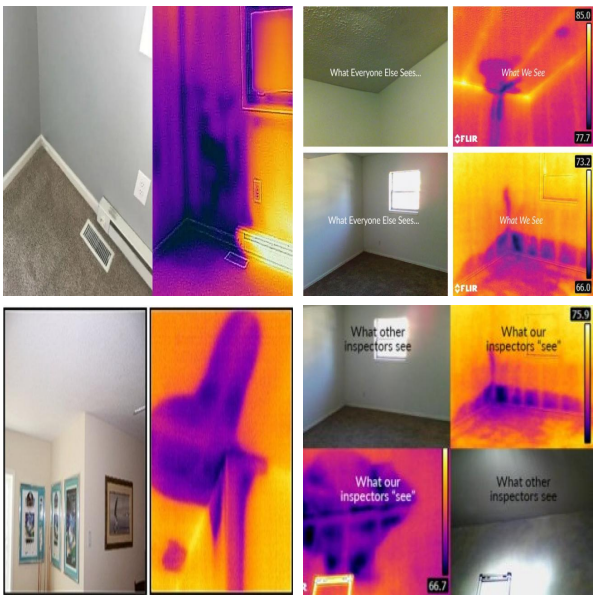


Figure 1. Comparison between RGB and thermal imaging samples [3].

Table 1. Comparison between RGB and thermal imaging modalities.

Characteristic	RGB Imaging	Thermal Imaging
Imaging principle	Captures reflected visible light	Captures emitted infrared radiation
Operating conditions	Requires sufficient illumination	Operates in both daylight and darkness
Information provided	Color, texture, and appearance	Surface temperature distribution and heat patterns
Ability to detect hidden defects	Limited for concealed defects	Can reveal moisture, heat leakage, insulation defects, and abnormal thermal patterns
Texture richness	Rich textures and sharp edges	Low contrast with weak textures and blurred boundaries
Sensitivity to lighting	Highly affected by illumination	Largely independent of visible lighting
Typical applications	Object recognition, scene understanding, photography	Building inspection, predictive maintenance, surveillance, industrial monitoring
Challenges for super-resolution	Texture recovery and realistic detail reconstruction	Low spatial resolution, sensor noise, weak textures, limited datasets, and preservation of temperature-related information

Despite these advantages, thermal imaging is still constrained by the relatively low spatial resolution of infrared sensors. High-resolution infrared detectors are expensive and technically difficult to manufacture, making low-resolution thermal cameras the predominant choice in many practical applications. Consequently, improving the spatial resolution of thermal images has become an important research problem, motivating the development of thermal image super-resolution methods.

Image super-resolution (SR) aims to reconstruct a high-resolution (HR) image from a low-resolution (LR) observation. In visible-light imaging, SR has progressed from classical interpolation methods to convolutional neural networks (CNNs), generative adversarial networks (GANs),

transformers, state-space models, hypercomplex neural networks, and diffusion-based methods [4,5]. However, thermal image SR remains more challenging because infrared images have different statistical properties from RGB images, including weaker textures, lower contrast, sensor-dependent noise, and fewer large-scale public datasets [6,7].

This paper reviews thermal and infrared image super-resolution from the perspectives of datasets, evaluation metrics, reconstruction methods, and open challenges. In addition to reviewing existing methods, this paper introduces a demosaicking-inspired upsampling framework for thermal SR and presents FusionNet as an adaptive fusion case study. The proposed direction is motivated by the similarity between demosaicking and image upsampling: in both cases, only a subset of samples is known, while missing values must be reconstructed from neighboring observations.

The main contributions of this paper are summarized as follows:

- **Comprehensive Review:** A comprehensive review of thermal image super-resolution (SR) is presented, covering thermal imaging characteristics, publicly available datasets, evaluation metrics, and representative classical and deep-learning-based SR methods.
- **Demosaicking-Inspired Deep Fusion Framework:** A novel thermal-image SR framework is proposed that reformulates bilinear interpolation, Menon interpolation, and the Malvar–He–Cutler demosaicking method using a zero-insertion strategy. FusionNet fuses the resulting reconstruction candidates with the RLKA-Net output, learning spatially adaptive, pixel-wise fusion weights.
- **Performance Evaluation and Hypercomplex Fusion:** Extensive experiments demonstrate that FusionNet consistently outperforms the individual reconstruction methods. In addition, quaternion-based variants, QFusionNet and QFusionNet+, reduce parameters by about 75% while maintaining comparable reconstruction quality, with QFusionNet+ achieving slight performance improvements over the real-valued model.

The remainder of this paper is organized as follows. Section 2 introduces the background and problem formulation. Section 3 reviews datasets and assessment metrics. Section 4 presents a taxonomy of thermal SR methods. Section 5 describes the proposed demosaicking-inspired framework and FusionNet. Section 6 presents the experimental evaluation. Section 7 discusses open challenges and future directions. Section 8 concludes the paper.

2. Background

2.1. Thermal Imaging

Thermal cameras measure infrared radiation emitted by objects and convert it into image intensities. Since emitted radiation is related to temperature, thermal images provide information that is complementary to visible-light images. This makes thermal imaging valuable in applications where lighting conditions are poor or where the target information is temperature-related rather than appearance-related.

In building inspection, thermal imaging can help identify heat loss, water leakage, insulation defects, cracks, and other hidden structural problems. In surveillance and autonomous systems, thermal cameras can operate at night or in low-light environments. In industrial monitoring, thermal images can reveal overheating components and abnormal operating conditions.

However, thermal imaging has several limitations. Thermal sensors often have lower spatial resolution than RGB sensors. Thermal images may also contain sensor noise, non-uniformity, low contrast, and blurred edges. These limitations motivate the development of thermal image enhancement and super-resolution methods.

2.2. Image Super-Resolution

Image super-resolution (SR) aims to reconstruct a high-resolution image from one or more low-resolution observations. Figure 2 illustrates the objective of image super-resolution. Starting from a

low-resolution thermal image, the image is progressively reconstructed at higher spatial resolutions, enabling finer structural details to become visible.

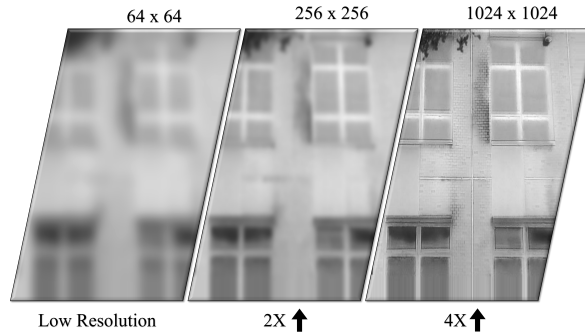


Figure 2. Illustration of the image super-resolution process.

Depending on the number of input images, SR methods can be broadly categorized into *single-image super-resolution* (SISR) and *multi-frame super-resolution* (MFSR). SISR reconstructs a high-resolution image from a single low-resolution input, whereas MFSR exploits information from multiple frames or images. Since the reconstruction setting considered in this work uses a single low-resolution thermal image as input, this work focuses on single-image super-resolution.

Figure 3 summarizes the taxonomy of image super-resolution methods reviewed in this paper. This work focuses on the single-image super-resolution branch.

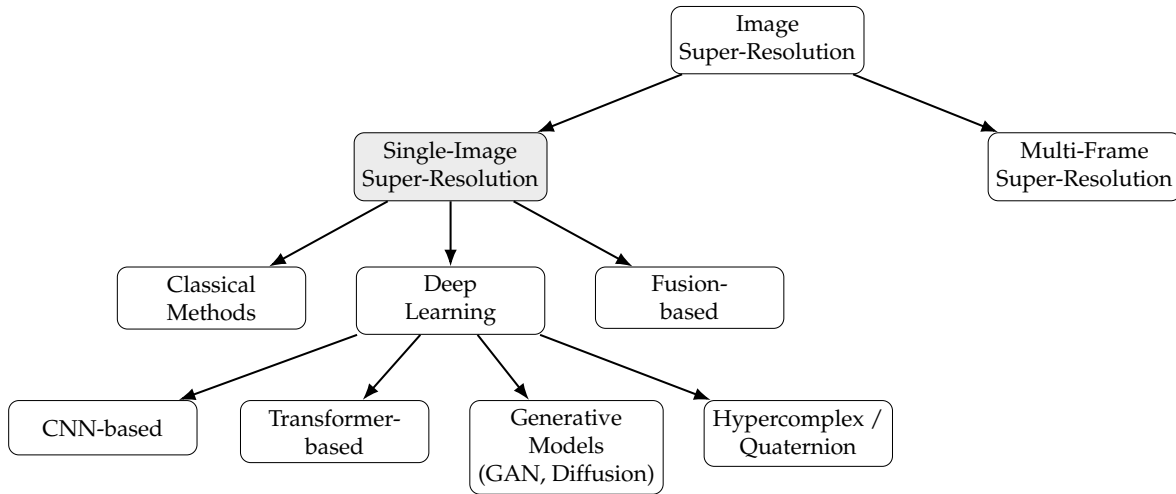


Figure 3. Taxonomy of image super-resolution methods reviewed in this paper. The shaded branch highlights the focus of this survey: single-image super-resolution.

As shown in Figure 3, the reviewed single-image SR methods include classical interpolation, CNN-based methods, GAN-based methods, transformer-based methods, state-space models, diffusion/autoregressive methods, and fusion-based approaches. The proposed framework later builds on this branch by adapting demosaicking algorithms to thermal upsampling and combining them with deep learning through adaptive fusion.

A common degradation model for single-image super-resolution is

$$I_{LR} = D(I_{HR}; \theta_D), \quad (1)$$

where I_{HR} is the original high-resolution image, I_{LR} is the degraded low-resolution image, $D(\cdot)$ is the degradation operator, and θ_D represents degradation parameters such as the blur kernel, scale factor, noise level, and downsampling operator.

The goal of single-image super-resolution is to learn or define a reconstruction function $R(\cdot)$ such that

$$\hat{I}_{HR} = R(I_{LR}; \theta_R), \quad (2)$$

where $\hat{I}_{HR}$ is the reconstructed high-resolution image and θ_R denotes the parameters of the reconstruction method.

In many super-resolution benchmarks, I_{LR} is generated by bicubic downsampling. Although this setting allows controlled evaluation, it does not fully represent real thermal degradation. Real infrared images may suffer from optical blur, motion blur, sensor noise, temperature drift, compression artifacts, and non-uniform degradation.

2.3. Challenges in Thermal Image Super-Resolution

Thermal SR is more challenging than conventional RGB SR for several reasons:

- **Low contrast:** Thermal images often contain small intensity differences, making edge and texture recovery difficult.
- **Weak textures:** Infrared images usually contain fewer high-frequency textures than RGB images.
- **Sensor noise:** Thermal cameras may introduce non-uniform noise and sensor-specific artifacts.
- **Dataset scarcity:** Large paired LR-HR thermal datasets are limited compared with RGB SR datasets.
- **Real-world degradation:** Real thermal degradation is often different from synthetic bicubic downsampling.
- **Physical meaning:** Thermal intensities may correspond to temperature-related information; therefore, hallucinated details can be misleading in inspection applications.

3. Datasets and Assessment Metrics

Datasets and evaluation metrics play a fundamental role in the development and comparison of image super-resolution algorithms. While RGB image super-resolution benefits from several standardized benchmarks, including Set5, Set14, BSD100, Urban100, Manga109, and DIV2K [7], thermal image super-resolution has comparatively fewer publicly available datasets specifically designed for this task. Consequently, many existing thermal datasets were originally developed for object detection, multispectral fusion, surveillance, or autonomous driving and are later adapted for super-resolution by synthetically generating low-resolution images through bicubic downsampling.

3.1. Thermal and Infrared Image Datasets

Table 2 summarizes representative datasets that are frequently used in thermal and infrared image processing research and have been employed, either directly or indirectly, for thermal image super-resolution.

Compared with RGB image super-resolution, thermal image super-resolution still lacks large standardized benchmarks. Consequently, most existing studies construct low-resolution inputs by bicubic downsampling of high-resolution thermal images. The PBVS Thermal Image Super-Resolution Challenge series represents one of the few benchmarks specifically developed for thermal SR. Earlier editions focused on thermal-only reconstruction, whereas recent editions additionally investigate RGB-guided thermal super-resolution, reflecting the growing interest in cross-spectral reconstruction methods.

The dataset used in this work, Thermal Imaging in Building Inspection, was originally developed for thermal building inspection rather than super-resolution. Therefore, low-resolution inputs are generated using bicubic downsampling with a scaling factor of $\times 2$, and the original thermal images are used as high-resolution ground-truth references.

Table 2. Representative thermal and infrared datasets commonly used in image super-resolution research.

Dataset	Year	Size	Resolution	Primary Use	Description
FLIR ADAS Thermal Dataset [8]	2018	~10,000 aligned thermal-visible image pairs	640 × 512	Detection, guided SR	Large automotive thermal-visible dataset captured for autonomous driving. Although designed for object detection, it is frequently adopted for thermal image restoration and RGB-guided thermal super-resolution.
KAIST Multispectral Pedestrian Dataset [9]	2015	~95,000 aligned thermal-visible image pairs	Visible: 640 × 480 Thermal: 320 × 256	Detection, fusion	Daytime and nighttime multispectral pedestrian dataset containing aligned visible and thermal images. Frequently used for multispectral learning and cross-modal image reconstruction.
CATS (Color and Thermal Stereo) [10]	2017	Stereo thermal pairs	visible-image 640 × 480	Stereo, fusion	Contains synchronized stereo visible and thermal images together with disparity information, making it useful for multimodal reconstruction and geometry-aware image processing.
LLVIP [11]	2021	15,488 aligned visible-infrared image pairs	Camera dependent	Detection, fusion	Large-scale paired visible-infrared dataset collected under extremely low-light conditions. It is commonly used for cross-spectral detection, fusion, and RGB-guided thermal reconstruction.
PBVS Thermal Image Super-Resolution Challenge [12]	2021	1,021 base images (3,063 LR/HR pairs)	Multiple resolutions	Thermal SR benchmark	Community benchmark dedicated to thermal image super-resolution with ×2, ×3, and ×4 enlargement factors.
PBVS Thermal Image Super-Resolution Challenge [13]	2024	1,000 registered thermal-RGB pairs	640 × 480	Cross-spectral SR	Recent benchmark introducing RGB-guided thermal super-resolution with larger scaling factors and paired thermal-visible images.
Thermal Imaging in Building Inspection (this work) [3]	2024	306 thermal images (277 train, 14 validation, 15 test)	640 × 640	Building inspection	Thermal building-inspection dataset employed in this work. Since no paired low-resolution images are provided, LR images are generated using bicubic ×2 downsampling, while the original images serve as HR references.

3.2. Assessment Metrics

Thermal image super-resolution methods are commonly evaluated using full-reference image quality metrics when ground-truth high-resolution images are available. The most frequently reported metrics are Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). In addition, perceptual and efficiency metrics are often reported to provide a more comprehensive evaluation.

3.2.1. Peak Signal-to-Noise Ratio [7]

PSNR measures the pixel-wise fidelity between the reconstructed image and its ground-truth counterpart. It is defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{L^2}{\text{MSE}} \right), \tag{3}$$

where L denotes the maximum possible pixel value. The mean squared error (MSE) is given by

$$\text{MSE} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (I_{\text{HR}}(i, j) - \hat{I}_{\text{HR}}(i, j))^2. \tag{4}$$

Higher PSNR values indicate smaller reconstruction errors.

3.2.2. Structural Similarity Index Measure [7]

SSIM evaluates perceptual similarity by jointly considering luminance, contrast, and structural information:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \tag{5}$$

Here, μ_x and μ_y denote the local means, σ_x^2 and σ_y^2 represent the local variances, σ_{xy} is the covariance, and C_1 and C_2 are small constants introduced for numerical stability. SSIM values range from 0 to 1, where values closer to 1 indicate greater structural similarity.

3.2.3. Additional Evaluation Metrics

Although PSNR and SSIM remain the dominant evaluation metrics, they do not always correlate well with perceptual image quality. Therefore, recent studies increasingly report additional objective and perceptual metrics. Table 3 summarizes the most commonly used evaluation metrics, their categories, purposes, and preferred optimization directions.

Table 3. Common evaluation metrics for thermal image super-resolution.

Metric	Category	Description	Preferred
PSNR	Full-reference	Pixel-wise reconstruction fidelity.	Higher
SSIM	Full-reference	Structural similarity between reconstructed and reference images.	Higher
RMSE	Full-reference	Root mean squared reconstruction error.	Lower
LPIPS [14]	Perceptual	Learned perceptual similarity using deep neural network features.	Lower
NIQE [15]	No-reference	Blind perceptual image quality without requiring ground truth.	Lower
PI [16]	No-reference	Perceptual Index combining no-reference image quality measures.	Lower

For application-oriented thermal imaging, image quality metrics alone may not fully reflect practical performance. Therefore, many recent studies additionally evaluate downstream tasks such as object detection, semantic segmentation, or defect localization to determine whether the reconstructed thermal images improve real-world inspection performance.

4. Taxonomy of Thermal Image Super-Resolution Methods

Thermal image super-resolution (SR) methods can be broadly divided into *classical*, *learning-based*, and *fusion-based* approaches. Classical methods estimate unknown samples through fixed mathemati-

cal operations and do not require training data. Learning-based methods instead learn a nonlinear mapping from low-resolution (LR) images to high-resolution (HR) images and can be further classified according to their main architectural principle, including convolutional neural networks (CNNs), transformers, state-space models, and autoregressive models. Fusion-based methods combine the outputs or features of multiple reconstruction approaches to exploit their complementary characteristics.

Although several influential SR architectures were originally developed for visible-spectrum images, they are relevant to thermal SR because they are frequently adopted as benchmark models or adapted to single-channel infrared data. Therefore, this taxonomy distinguishes between general-image SR architectures used as comparison baselines and methods developed specifically for infrared or thermal imagery.

4.1. Classical Methods

Classical SR methods calculate the intensity of an unknown output pixel from the available LR samples using a predefined interpolation rule. They are deterministic, computationally inexpensive, and independent of training data. However, because they do not learn image-specific priors, their ability to reconstruct missing high-frequency information is limited.

4.1.1. Interpolation-Based Methods [17]

Nearest-neighbor interpolation assigns each output location the value of the closest input pixel. It is extremely fast but commonly produces block-like regions, jagged boundaries, and abrupt transitions.

Bicubic interpolation applies cubic convolution over a larger neighborhood, typically involving 4×4 input samples. It generally preserves gradual intensity variations more effectively than bilinear interpolation. However, it still estimates new pixels only from nearby intensity values and cannot recover genuinely missing structures.

Lanczos interpolation approximates ideal sinc interpolation using a finite-support windowed sinc kernel. It can produce sharper boundaries than bilinear or bicubic interpolation, but oscillations around strong intensity transitions may lead to ringing artifacts.

These limitations are particularly important in thermal inspection images, where defects such as cracks may appear as thin, low-contrast structures. Excessive smoothing can remove or weaken such structures, while ringing can introduce artificial thermal boundaries.

4.1.2. Demosaicking-Derived Methods

Demosaicking algorithms were originally designed to estimate missing color components from sparsely sampled color filter array measurements. Although demosaicking and image super-resolution are distinct reconstruction problems, both require the estimation of unavailable samples from spatially incomplete observations. This similarity motivates the adaptation of demosaicking principles to thermal image upsampling.

Bilinear interpolation constitutes the simplest reconstruction strategy in this group. In conventional demosaicking, missing samples are estimated through local weighted averaging of neighboring observations. In this work, the same principle is reformulated for single-channel thermal images within a zero-insertion upsampling framework, where only the inserted spatial locations are reconstructed while the original thermal samples are preserved.

The Malvar–He–Cutler method improves basic bilinear demosaicking by applying larger-support linear filters together with Laplacian correction terms [18,19]. These correction terms use information from neighboring samples to reduce interpolation errors around edges and local intensity variations.

The Menon method performs directional interpolation and applies an a posteriori decision to determine the more reliable interpolation direction [20]. By distinguishing between horizontal and vertical structures, it reduces averaging across strong edges and can preserve direction-dependent image details more effectively than nondirectional interpolation.

The original Bilinear, Malvar–He–Cutler, and Menon formulations are associated with missing-sample reconstruction in demosaicking and cannot be transferred unchanged to the single-channel

thermal SR setting considered here. Accordingly, all three are reformulated within the proposed zero-insertion upsampling procedure. Their mathematical definitions and processing steps are presented in Section 5.

Although interpolation- and demosaicking-derived methods are computationally efficient and require no training data, their reconstruction capability remains limited by fixed mathematical assumptions and local pixel relationships. Figure 4 summarizes their principal limitations.

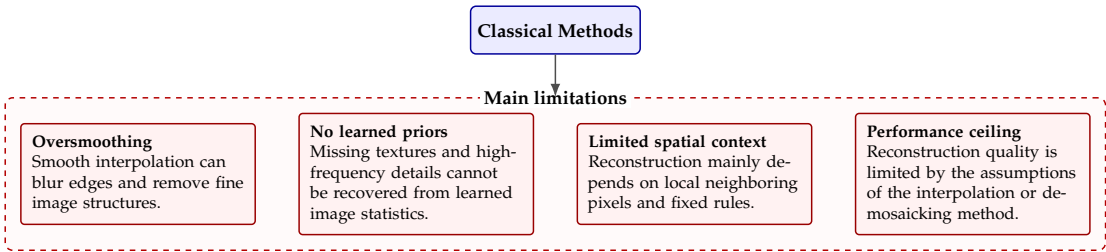


Figure 4. Main limitations of classical interpolation- and demosaicking-based super-resolution methods.

4.2. Learning-Based Methods

Learning-based methods estimate an HR image using parameters learned from training examples. Compared with fixed interpolation, these methods can model complex nonlinear relationships between LR and HR images and can reconstruct stronger edges and finer structures. Their performance, however, depends on the quantity and representativeness of the training data and on the degradation model used to generate LR–HR pairs.

4.2.1. CNN-Based Methods

CNN-based SR methods extract hierarchical spatial features through learned convolutional filters [21–25]. Their local receptive fields, parameter sharing, and translation-equivariant processing make them suitable for image reconstruction.

General-image CNN baselines.

The Deep Recursive Residual Network (DRRN) combines recursive learning with local and global residual connections [26]. Recursive parameter sharing allows DRRN to construct a deep network while limiting the number of independent parameters. Residual learning facilitates the optimization of the deep recursive architecture. Although DRRN was designed for general-image SISR, it can be applied to single-channel thermal images and is therefore used as a comparison baseline in this study.

The Enhanced Deep Super-Resolution Network (EDSR) modifies conventional residual architectures by removing modules that are unnecessary for SR, particularly batch-normalization layers, and by increasing the capacity of the remaining residual blocks [27]. Removing batch normalization reduces memory consumption and avoids normalization of the image intensity range. EDSR achieved strong reconstruction performance on natural-image SR benchmarks and has subsequently become a widely used baseline.

The Information Multi-Distillation Network (IMDN) was proposed as a lightweight SR architecture [28]. Its Information Multi-Distillation Blocks progressively extract selected feature channels and aggregate hierarchical representations. A contrast-aware channel attention mechanism emphasizes informative channels during feature fusion. IMDN provides a favorable trade-off between reconstruction accuracy, parameter count, and computational cost, making it relevant to resource-constrained thermal imaging applications.

These networks were originally developed using visible-spectrum datasets and do not explicitly model infrared sensor characteristics. Their inclusion in thermal SR experiments should therefore be interpreted as an evaluation of their architectural transferability rather than as evidence that they are infrared-specific models.

Infrared-specific CNN and adversarial methods.

PSRGAN was proposed to address infrared SR under limited infrared training data through progressive reconstruction and transfer learning [29]. Its architecture contains a main path for infrared feature reconstruction and a branch path used to transfer information learned from visible images. The main path employs depth-wise residual blocks, whereas the branch path uses shallow lightweight distillation residual blocks. A multistage transfer-learning strategy is used to reduce the discrepancy between visible and infrared feature spaces.

Because PSRGAN incorporates adversarial learning, it can produce perceptually sharper reconstructions than methods optimized only with pixel-wise losses. However, adversarially generated details are not guaranteed to represent physically observed thermal structures. This limitation is important in inspection and measurement applications, where visually plausible patterns may be misinterpreted as actual defects.

RLKA-Net, the Recurrent Large Kernel Attention Network, was developed specifically for efficient single infrared image SR [30]. It employs recurrent learning units to reuse network parameters and progressively refine feature representations. Large-kernel attention enlarges the effective receptive field and enables the model to capture spatial information at different granularities. An attention gate is used to regulate feature propagation and reduce unnecessary parameters.

The recurrence in RLKA-Net represents repeated internal feature refinement and does not imply that the model processes a thermal video sequence. RLKA-Net is a single-image method. Its relatively lightweight design makes it suitable for the proposed fusion framework. Nevertheless, like other supervised networks, its performance depends on the correspondence between the training degradation and the degradation present in real thermal images.

CNN-based super-resolution methods substantially improve reconstruction quality by learning hierarchical image features. Nevertheless, their performance often depends on the availability of representative training data, and their ability to generalize across sensors and imaging conditions may remain limited. Figure 5 summarizes the common limitations of this method family.

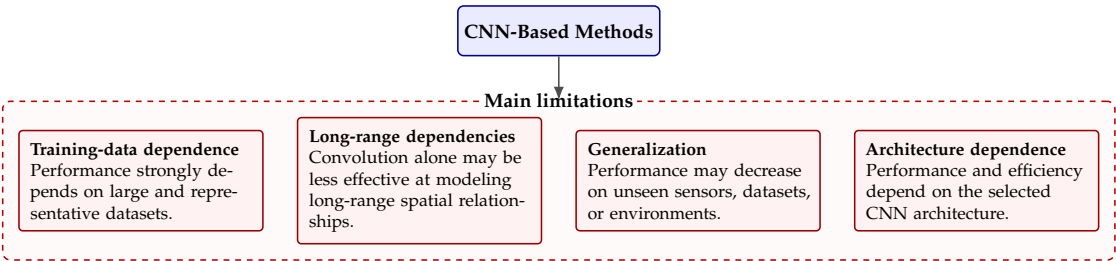


Figure 5. Common limitations of CNN-based super-resolution methods, including DRRN, EDSR, IMDN, PSRGAN, and RLKA-Net.

4.2.2. Transformer-Based Methods

Transformer-based SR methods aim to model broader spatial dependencies than conventional small-kernel CNNs [31–33]. Their attention or attention-inspired operations allow distant image regions to contribute to the reconstruction of a local area.

General-image transformer baseline.

SwinIR is an image-restoration architecture based on the Swin Transformer [34]. It consists of shallow feature extraction, deep feature extraction, and image reconstruction stages. Its deep feature extractor contains Residual Swin Transformer Blocks, each of which combines several Swin Transformer layers with a residual connection. Shifted-window self-attention enables information exchange between neighboring windows while avoiding the full computational cost of global self-attention.

SwinIR was evaluated on classical, lightweight, and real-world image SR, as well as denoising and JPEG artifact reduction. Although it was not designed specifically for thermal imagery, its ability to model spatial dependencies makes it an important transformer-based comparison method.

Infrared-specific transformer model.

LKFormer is a large-kernel transformer-style network developed for infrared image SR [35]. Instead of conventional self-attention, it uses a Large Kernel Residual Depth-wise Convolutional Attention module. Large-kernel depth-wise convolutions model nonlocal spatial relationships while maintaining linear computational complexity with respect to the number of pixels.

LKFormer also introduces a Gated-Pixel Feed-Forward Network to control feature flow while retaining spatially localized information. This design is intended to address common infrared-image characteristics, including relatively homogeneous regions, sparse textures, and a limited gradient range.

Although LKFormer avoids the quadratic complexity of global self-attention, its large-kernel processing remains more complex than basic interpolation and small lightweight CNNs. Its practical performance also depends on whether the training degradation accurately represents the optical blur, sensor noise, and preprocessing of the target thermal camera.

4.2.3. State-Space Methods

State-space architectures have recently been introduced into image restoration as an alternative to convolutional processing and self-attention. Selective state-space models can represent long-range dependencies with computational complexity that grows approximately linearly with sequence length.

IRSRMamba applies Mamba-based state-space modeling to infrared image SR [36]. The method combines state-space blocks with wavelet-transform feature modulation. Wavelet decomposition separates low-frequency structural information from high-frequency detail components, allowing the network to adaptively process information at different spatial frequencies.

The state-space blocks model long-range contextual relationships, whereas the wavelet components provide multi-scale local information. IRSRMamba also introduces a semantic-consistency objective intended to reduce fragmented contextual representations produced during block-wise processing.

This combination is well suited to infrared images, which often contain broad homogeneous regions together with weak but diagnostically important boundaries. However, the method remains dependent on supervised training and on the assumed LR-generation process. Furthermore, state-space image models are newer and currently have less extensive cross-dataset and cross-sensor validation than established CNN architectures.

4.2.4. Autoregressive Methods

Autoregressive SR methods reconstruct image information progressively rather than predicting the complete HR output through a single feed-forward pass. They can model dependencies between reconstruction stages or image representations but generally require sequential computation.

Real-IISR was introduced for real-world infrared image SR and is accompanied by the FLIR-IISR benchmark [37]. In contrast to studies that construct LR images using only bicubic downsampling, FLIR-IISR contains paired LR and HR infrared observations acquired using focus variation and motion-induced blur.

Real-IISR performs scale-by-scale reconstruction using thermal-structural guided visual autoregression. Its Thermal-Structural Guidance module encodes thermal-distribution and structural information to account for the possible misalignment between thermal radiation patterns and visible object edges. A Condition-Adaptive Codebook adjusts discrete image representations according to degradation-aware thermal priors.

The method additionally introduces a Thermal Order Consistency Loss. Rather than enforcing exact temperature values, this loss preserves the relative ordering between image intensities, thereby accounting for spatial misalignment and thermal drift in real image pairs.

Real-IISR directly addresses several limitations of synthetic infrared SR benchmarks. However, its autoregressive reconstruction is sequential and is therefore likely to require greater inference time than a compact single-pass CNN. Because both the framework and FLIR-IISR dataset are recent, broader

independent evaluation across thermal cameras, spectral ranges, environments, and acquisition conditions is still required.

Recent learning-based super-resolution methods use Transformer architectures, state-space models, and autoregressive reconstruction to improve long-range feature modeling and structural recovery. However, these methods may introduce greater computational complexity, slower inference, and continued dependence on representative training data. Their common limitations are summarized in Figure 6.

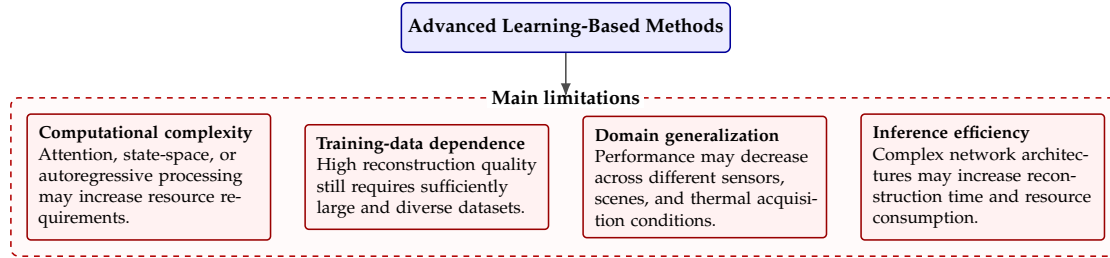


Figure 6. Common limitations of Transformer-, state-space-, and autoregressive-based super-resolution methods, including SwinIR, LKFormer, IRSRMamba, and Real-IISR.

4.2.5. Hypercomplex-Based Methods

While conventional convolutional neural networks process multi-channel images (such as RGB or multi-spectral visible-thermal data) by treating the spectral bands as separate, independent feature maps, hypercomplex neural networks represent multi-channel observations as unified multidimensional entities. In particular, Quaternion Neural Networks (QNNs) extend standard real-valued operations to the four-dimensional quaternion space $\mathbb{H}$ [38–43].

A quaternion $q \in \mathbb{H}$ consists of a real part and three imaginary parts:

$$q = q_r + q_i \mathbf{i} + q_j \mathbf{j} + q_k \mathbf{k}, \quad (6)$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are imaginary units satisfying the fundamental multiplication rules:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1. \quad (7)$$

For a four-channel image—such as cross-spectral RGB-Thermal (RGB-T) data—the thermal channel I_T and the visible color channels I_R , I_G , and I_B can be naturally mapped to the real and imaginary components of a quaternion matrix:

$$q = I_T + I_R \mathbf{i} + I_G \mathbf{j} + I_B \mathbf{k}. \quad (8)$$

Similarly, standard three-channel RGB color images can be represented by setting the real part to the grayscale version of the image I_{gray} , and the imaginary parts to the red, green, and blue components:

$$q = I_{\text{gray}} + I_R \mathbf{i} + I_G \mathbf{j} + I_B \mathbf{k}. \quad (9)$$

This formulation preserves inter-channel spectral correlations and prevents independent channel degradation during feature learning.

The core operation in these networks is the quaternion convolution (QConv) [44–46]. Given a quaternion input $x = x_r + x_i \mathbf{i} + x_j \mathbf{j} + x_k \mathbf{k}$ and a quaternion filter weight $W = w_r + w_i \mathbf{i} + w_j \mathbf{j} + w_k \mathbf{k}$, the convolution is defined using the Hamilton product:

$$y = W \otimes x. \quad (10)$$

By expanding the Hamilton product, the forward convolution can be represented in matrix form as:

$$\begin{bmatrix} y_r \\ y_i \\ y_j \\ y_k \end{bmatrix} = \begin{bmatrix} w_r & -w_i & -w_j & -w_k \\ w_i & w_r & -w_k & w_j \\ w_j & w_k & w_r & -w_i \\ w_k & -w_j & w_i & w_r \end{bmatrix} * \begin{bmatrix} x_r \\ x_i \\ x_j \\ x_k \end{bmatrix}. \quad (11)$$

where $*$ denotes the spatial convolution operation. The weight-sharing structure in the Hamilton matrix forces cross-spectral interactions among the channels while reducing the number of independent parameters by 75% (a $4\times$ reduction) compared to a real-valued convolution layer with the same input-output channel configuration.

In the context of super-resolution, the QSRNet framework [46] demonstrates the advantages of hypercomplex representations by adapting real-valued architectures to the quaternion domain. By replacing standard convolutions with quaternion convolutions, QSRNet processes structural features, gradient-based soft edges, and multi-scale details as unified multidimensional entities. By exploiting hypercomplex algebra, the model achieves competitive restoration fidelity with an approximate 75% reduction in network parameters, making it highly suitable for computationally constrained edge-computing devices in thermal and multispectral inspection systems.

4.3. Fusion-Based Methods

Classical and learning-based SR methods have different reconstruction characteristics, and no individual approach performs optimally in all image regions. Bilinear interpolation generally produces smooth and stable results but may weaken narrow boundaries and fine details. Larger-support Malvar–He–Cutler filtering can introduce additional correction information, whereas Menon interpolation can better preserve direction-dependent structures. Learning-based methods such as RLKA-Net can reconstruct richer features through learned representations, although their performance remains dependent on the characteristics of the training data.

Fusion-based methods aim to exploit these complementary properties by combining the outputs or feature representations of multiple reconstruction approaches. A simple fixed average assigns the same importance to every input method and therefore cannot adapt to spatial variations within an image. Adaptive fusion instead allows the contribution of each reconstruction method to vary according to the local image content, such as homogeneous regions, object boundaries, and fine structural details.

In this work, a fusion-based model named FusionNet is proposed to combine the outputs of Bilinear interpolation, the adapted Malvar–He–Cutler method, the adapted Menon method, and RLKA-Net. The model learns spatially varying contributions for the four reconstructed images and produces a single fused high-resolution output. This approach is intended to preserve the stable reconstruction of classical methods while benefiting from the richer details recovered by the learning-based model.

The architecture, mathematical formulation, training procedure, and implementation details of FusionNet are presented in Section 5.

4.4. Comparison of Representative Methods

Table 4 summarizes the principal characteristics of the methods discussed in this section. DRRN, EDSR, IMDN, and SwinIR are identified as general-image methods because they were not originally designed for infrared imaging. They are nevertheless included because they provide representative CNN and transformer baselines in the experimental comparison.

Table 4. Taxonomy and comparison of representative methods relevant to thermal and infrared single-image super-resolution.

Method	Year	Category	Main principle	Main advantages	Main limitations
Nearest neighbor		Classical Interpolation	Copies the value of the spatially nearest LR sample.	Extremely fast, deterministic, and training-free.	Produces block-like regions and jagged boundaries.
Bilinear		Classical Interpolation	Uses weighted averaging of the four nearest samples.	Fast, stable, and simple to implement.	Oversmooths edges and cannot reconstruct missing high-frequency information.
Bicubic		Classical Interpolation	Applies cubic convolution over a larger local neighborhood.	Usually smoother and sharper than bilinear interpolation.	Still blurs fine structures and cannot learn image priors.
Lanczos		Classical Interpolation	Uses a finite-support windowed sinc kernel.	Can preserve relatively sharp transitions.	May introduce ringing near strong boundaries.
Malvar–He–Cutler	2004	Demosaicking-derived	Uses linear filters with Laplacian correction terms.	Improves reconstruction around local structures compared with simple averaging.	Originally designed for Bayer demosaicking and requires adaptation for single-channel thermal images.
Menon	2007	Demosaicking-derived	Performs directional interpolation followed by an a posteriori directional decision.	Reduces interpolation across strong directional edges.	Directional estimation may be unreliable in noisy or weak-gradient regions.
DRRN	2017	General-image CNN	Combines recursive parameter sharing with local and global residual learning.	Constructs a deep network using a comparatively limited number of independent parameters.	Processes interpolated images in HR space and is not infrared-specific.
EDSR	2017	General-image CNN	Uses high-capacity residual blocks without batch normalization.	Strong reconstruction baseline with effective residual learning.	Large configurations have substantial computational and memory requirements.
IMDN	2019	Lightweight general-image CNN	Uses information multi-distillation blocks and contrast-aware channel attention.	Provides a favorable accuracy–efficiency trade-off.	Has less representational capacity than substantially larger models and is not infrared-specific.
PSRGAN	2021	Infrared adversarial CNN	Combines progressive reconstruction, transfer learning, residual feature extraction, and adversarial training.	Addresses limited infrared training data and can produce visually sharp outputs.	Adversarial learning may generate visually plausible but physically unsupported details.
SwinIR	2021	General-image transformer	Uses Residual Swin Transformer Blocks with shifted-window self-attention.	Models broader spatial dependencies and supports several image-restoration tasks.	More computationally complex than lightweight CNNs and is not infrared-specific.
RLKA-Net	2023	Infrared recurrent-attention CNN	Uses recurrent learning, large-kernel attention, and attention gates.	Infrared-specific, parameter-efficient, and capable of broad spatial feature modeling.	Depends on the training degradation and performs sequential internal feature refinement.
LKFormer	2024	Infrared transformer-style model	Replaces standard self-attention with large-kernel residual depth-wise convolutional attention and uses a gated pixel feed-forward network.	Models nonlocal spatial information with linear-complexity operations.	More complex than lightweight CNNs and remains dependent on synthetic training degradations.
IRSRMamba	2025	Infrared state-space model	Combines Mamba-based state-space blocks, wavelet feature modulation, and semantic-consistency supervision.	Integrates long-range context with multi-scale frequency information.	State-space image models are relatively new, and broader real-sensor validation is needed.
Real-IISR	2026	Infrared autoregressive model	Uses thermal-structural guidance, a condition-adaptive codebook, scale-by-scale autoregression, and thermal-order consistency.	Explicitly addresses real optical and sensing degradations.	Sequential reconstruction can increase inference cost; cross-camera validation remains limited.
FusionNet		Proposed Adaptive fusion	Predicts pixel-wise softmax weights for Bilinear, Malvar, Menon, and RLKA-Net outputs.	Exploits complementary classical, directional, and learned reconstruction behavior.	Requires the computation of all component outputs and cannot reconstruct information absent from every input.

4.5. Discussion

The development of thermal image SR demonstrates a gradual transition from fixed local interpolation to data-driven reconstruction with increasingly broad spatial modeling. Classical interpolation remains attractive because it is transparent, inexpensive, and does not require a training dataset. Demosaicking-derived methods introduce larger-support and directional estimation but must be modified before they can be applied to single-channel thermal upsampling.

CNN-based methods learn stronger nonlinear reconstruction mappings. DRRN, EDSR, and IMDN represent recursive, high-capacity residual, and lightweight information-distillation designs, respectively. Although these architectures were developed for visible-spectrum data, their inclusion makes it possible to evaluate whether general SR principles transfer effectively to thermal imagery. PSRGAN and RLKA-Net introduce infrared-specific transfer learning, adversarial learning, recurrent refinement, and large-kernel attention.

Transformer and state-space methods further enlarge the modeled spatial context. SwinIR uses shifted-window attention, while LKFormer replaces conventional self-attention with large-kernel depth-wise processing. IRSRMamba combines long-range state-space modeling with wavelet-domain modulation. Real-IISR extends the field toward real-world acquisition by introducing paired data with optical and motion degradations and by preserving relative thermal intensity ordering.

Despite this progress, supervised SR models remain sensitive to dataset bias and degradation mismatch. A network trained using bicubic downsampling may perform well on similarly generated test data while generalizing poorly to optical blur, motion, nonuniform sensor noise, fixed-pattern noise, or camera-specific image processing. Generative models require additional caution because perceptually sharp details are not necessarily physically accurate.

The proposed FusionNet follows a complementary direction. Instead of assuming that one reconstruction mechanism is optimal for every image region, it combines stable interpolation, correction-based filtering, directional estimation, and learned infrared reconstruction. The spatially varying fusion weights allow different methods to dominate in homogeneous regions, along structural boundaries, and around fine thermal details.

5. Proposed Demosaicking-Inspired Thermal SR Framework

This section presents the proposed thermal image super-resolution framework. The method reformulates image upsampling as a missing-sample reconstruction problem and adapts three interpolation strategies—Bilinear, Malvar–He–Cutler-style, and Menon directional interpolation—to thermal image super-resolution. Their outputs are subsequently combined with the reconstruction produced by RLKA-Net using the proposed FusionNet.

5.1. Demosaicking-Inspired Thermal Upsampling

The proposed approach is motivated by the similarity between demosaicking and image upsampling. In both cases, unavailable pixel values must be estimated from observed neighboring samples. For a low-resolution thermal image

$$g \in \mathbb{R}^{H \times W}, \quad (12)$$

a zero-inserted image

$$u \in \mathbb{R}^{2H \times 2W} \quad (13)$$

is constructed as

$$u(2i, 2j) = g(i, j), \quad 0 \leq i < H, \quad 0 \leq j < W. \quad (14)$$

All remaining positions are initialized to zero and form the set of unknown samples Ω_0 . During reconstruction, the original thermal samples are preserved and only the inserted locations are estimated.

5.1.1. Bilinear Interpolation Adaptation

The first reconstruction branch adapts Bilinear interpolation to the zero-inserted thermal image. The interpolation kernel is

$$H_{\text{bil}} = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \quad (15)$$

Algorithm 1 summarizes the complete procedure.

Algorithm 1: Demosaicking-Inspired Bilinear Thermal Upsampling

Input: Low-resolution thermal image $g \in \mathbb{R}^{H \times W}$
Output: Upsampled thermal image $I_{\text{bil}} \in \mathbb{R}^{2H \times 2W}$
Construct the zero-inserted image u using Equation (14);
Define the set of inserted samples Ω_0 ;
Apply the Bilinear kernel:

$$\tilde{u} = u * H_{\text{bil}}$$

foreach (x, y) in the $2H \times 2W$ image **do**

if $(x, y) \in \Omega_0$ **then**
| $I_{\text{bil}}(x, y) \leftarrow \tilde{u}(x, y);$
else
| $I_{\text{bil}}(x, y) \leftarrow u(x, y);$

return $I_{\text{bil}};$

This procedure fills the missing positions using local weighted averages while preserving the original thermal samples.

5.1.2. Malvar–He–Cutler-Style Adaptation

The Malvar–He–Cutler method was originally developed for color demosaicking. Since thermal images are single-channel, its cross-channel correction mechanism is not applicable. Instead, the proposed adaptation retains the larger-support filtering principle and applies it to the zero-inserted thermal image.

The first interpolation filter is

$$K_{\text{diag}} = \frac{1}{8} \begin{bmatrix} 0 & 0 & -\frac{3}{2} & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 \\ -\frac{3}{2} & 0 & 6 & 0 & -\frac{3}{2} \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & -\frac{3}{2} & 0 & 0 \end{bmatrix} \quad (16)$$

and the refinement filter is

$$K_{\text{main}} = \frac{1}{8} \begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ -1 & 2 & 4 & 2 & -1 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \quad (17)$$

The proposed adaptation is summarized in Algorithm 2.

Algorithm 2: Malvar–He–Cutler-Style Thermal Upsampling

Input: Low-resolution thermal image $g \in \mathbb{R}^{H \times W}$
Output: Upsampled thermal image $I_{\text{mal}} \in \mathbb{R}^{2H \times 2W}$
Construct zero-inserted image u using Equation (14);
Define inserted sample locations Ω_0 ;
Compute the first-stage filtered image:

$$T_{\text{diag}} = u * K_{\text{diag}}$$

Initialize $u^{(1)} \leftarrow u$;
foreach $(x, y) \in \Omega_0$ **do**
 $u^{(1)}(x, y) \leftarrow T_{\text{diag}}(x, y)$;
Compute the second-stage filtered image:

$$T_{\text{main}} = u^{(1)} * K_{\text{main}}$$

Initialize $I_{\text{mal}} \leftarrow u^{(1)}$;
foreach (x, y) *such that* $u^{(1)}(x, y) = 0$ **do**
 $I_{\text{mal}}(x, y) \leftarrow T_{\text{main}}(x, y)$;
Preserve all remaining values from $u^{(1)}$;
return I_{mal} ;

The wider spatial support allows the method to incorporate more neighboring information than simple Bilinear interpolation.

5.1.3. Menon Directional Interpolation Adaptation

The Menon adaptation exploits local edge direction by generating horizontal and vertical interpolation candidates and selecting the direction with lower local variation.

The one-dimensional interpolation filters are

$$h_0 = \left[0, \frac{1}{2}, 0, \frac{1}{2}, 0\right], \quad h_1 = \left[-\frac{1}{4}, 0, \frac{1}{2}, 0, -\frac{1}{4}\right] \quad (18)$$

The horizontal and vertical candidates are

$$H = (u * h_0)_H + (u * h_1)_H \quad (19)$$

$$V = (u * h_0)_V + (u * h_1)_V \quad (20)$$

and the corresponding local variations are

$$D_H = |H - H(\cdot, \cdot + 2)| \quad (21)$$

$$D_V = |V - V(\cdot + 2, \cdot)| \quad (22)$$

The decision mask is

$$k = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \quad (23)$$

with directional costs

$$d_H = D_H * k, \quad d_V = D_V * k^T \quad (24)$$

Algorithm 3 summarizes the adaptation.

Algorithm 3: Menon Directional Thermal Upsampling

Input: Low-resolution thermal image $g \in \mathbb{R}^{H \times W}$

Output: Upsampled thermal image $I_{\text{men}} \in \mathbb{R}^{2H \times 2W}$

Construct zero-inserted image u using Equation (14);

Define inserted sample locations Ω_0 ;

Compute horizontal candidate H ;

Compute vertical candidate V ;

Compute directional variations D_H and D_V ;

Aggregate the variations:

$$d_H = D_H * k, \quad d_V = D_V * k^T$$

Initialize $u_{\text{dir}} \leftarrow u$;

foreach $(x, y) \in \Omega_0$ **do**

if $d_H(x, y) \leq d_V(x, y)$ **then**

$u_{\text{dir}}(x, y) \leftarrow H(x, y)$;

else

$u_{\text{dir}}(x, y) \leftarrow V(x, y)$;

Apply the refinement filter

$$h_r = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \end{bmatrix}$$

along rows and columns using mirror padding, updating only the interpolated samples;

Preserve all original thermal samples;

return I_{men} ;

Unlike non-directional interpolation, this procedure allows reconstruction to adapt to local horizontal and vertical structures.

5.2. RLKA-Net Reconstruction

RLKA-Net is used as the learning-based reconstruction branch of the proposed framework. It was designed for infrared image super-resolution and combines recurrent feature refinement with large-kernel attention. In this work, its super-resolved output, denoted by I_{RLKA} , is used directly as the fourth input to FusionNet.

The architecture of RLKA-Net is shown in Figure 7. Since RLKA-Net is an existing method, only its role within the proposed framework is described here; further architectural details are available in the original work [30].

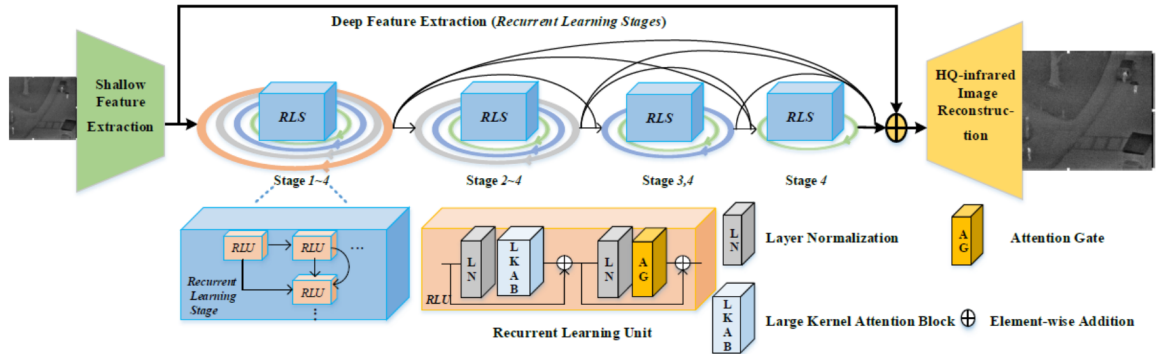


Figure 7. Architecture of RLKA-Net, reproduced from [30].

5.3. Proposed FusionNet

The three demosaicking-inspired reconstructions and the RLKA-Net output exhibit complementary reconstruction characteristics. FusionNet is therefore proposed to combine them adaptively using spatially varying learned weights. Figure 8 illustrates the proposed architecture.

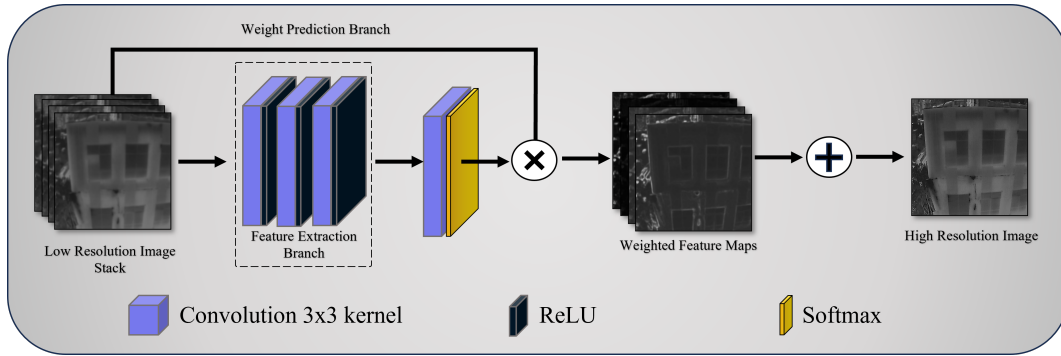


Figure 8. Architecture of the proposed FusionNet.

Let

$$I_1 = I_{bil}, \quad I_2 = I_{mal}, \quad I_3 = I_{men}, \quad I_4 = I_{RLKA} \quad (25)$$

denote the four reconstruction candidates. They are concatenated along the channel dimension as

$$X = \text{Concat}(I_1, I_2, I_3, I_4) \in \mathbb{R}^{4 \times H \times W}. \quad (26)$$

FusionNet consists of three 3×3 convolutional layers with ReLU activations followed by a four-channel weight-prediction layer, as summarized in Table 5.

Table 5. Architecture of the proposed FusionNet.

Layer	Kernel	Channels	Activation
Conv1	3×3	$4 \rightarrow 32$	ReLU
Conv2	3×3	$32 \rightarrow 32$	ReLU
Conv3	3×3	$32 \rightarrow 32$	ReLU
Weight prediction	3×3	$32 \rightarrow 4$	Softmax

The predicted weight maps are normalized at each spatial position using Softmax:

$$w_i(x, y) = \frac{\exp(z_i(x, y))}{\sum_{j=1}^4 \exp(z_j(x, y))}, \quad i = 1, \dots, 4, \quad (27)$$

where $z_i(x, y)$ denotes the unnormalized score for the i th reconstruction. Consequently,

$$\sum_{i=1}^4 w_i(x, y) = 1. \quad (28)$$

The final super-resolved image is obtained by pixel-wise weighted fusion:

$$I_{\text{fused}}(x, y) = \sum_{i=1}^4 w_i(x, y) I_i(x, y). \quad (29)$$

The complete procedure is summarized in Algorithm 4.

The learned spatial weights allow the contribution of each reconstruction method to vary across the image, enabling FusionNet to adapt the combination of classical and learning-based reconstructions to local image content.

5.4. Training Objective

FusionNet is trained using the Charbonnier loss,

$$\mathcal{L}_{\text{char}} = \frac{1}{N} \sum_{n=1}^N \sqrt{(\hat{x}_n - y_n)^2 + \epsilon}, \quad (30)$$

where $\hat{x}_n$ and y_n denote the reconstructed and ground-truth pixel values, respectively, and $\epsilon = 10^{-6}$ is used for numerical stability.

6. Experimental Evaluation

6.1. Dataset

Experiments are conducted on the Thermal Imaging in Building Inspection dataset [3]. The dataset contains 306 thermal images and is divided into 277 training images, 14 validation images, and 15 test images. Since paired LR-HR images are not directly provided, LR inputs are generated by bicubic downsampling with scale factor $\times 2$. The original thermal images are used as HR ground truth.

6.2. Training Details

RLKA-Net is fine-tuned on the thermal building-inspection dataset. FusionNet is trained using the outputs of Bilinear interpolation, Malvar-style interpolation, Menon-style interpolation, and RLKA-Net. The fusion network is optimized using the Adam optimizer with an initial learning rate of 10^{-4} . Training is performed using image patches of size 128×128 and batch size 8.

Algorithm 4: Proposed FusionNet Adaptive Reconstruction

Input: $I_{\text{bil}}, I_{\text{mal}}, I_{\text{men}},$ and I_{RLKA}

Output: Final super-resolved image I_{fused}

Concatenate the four reconstructions:

$$X \leftarrow \text{Concat}(I_{\text{bil}}, I_{\text{mal}}, I_{\text{men}}, I_{\text{RLKA}})$$

Extract features using three 3×3 convolutional layers with ReLU activations;

Predict four spatial weight maps;

Normalize the predicted weights using Softmax according to Equation (27);

Compute the pixel-wise weighted reconstruction using Equation (29);

return I_{fused} ;

6.3. Hypercomplex Analysis

To evaluate the potential of hypercomplex representations for computationally constrained edge devices, we draw inspiration from quaternion-based frameworks such as QSRNet [46] and replace the

real-valued convolutions in FusionNet with quaternion convolutions. Because quaternion convolutions intrinsically reduce the parameter count by approximately 75% compared to real-valued counterparts (while maintaining the same number of output channels), we evaluate two quaternion-based variants: QFusionNet, which matches the channel dimensions of the baseline, and QFusionNet⁺, which increases the channel capacity to try to match the original parameter count.

Table 6 presents the quantitative comparison. The baseline FusionNet contains 20.8K parameters. QFusionNet achieves competitive performance (37.6149 dB PSNR and 0.9742 SSIM) despite utilizing only 5.3K parameters. Furthermore, the scaled QFusionNet⁺ (19.8K parameters) outperforms the real-valued baseline, achieving 37.7958 dB PSNR and 0.9749 SSIM. These results demonstrate that the structural priors introduced by quaternion networks can lead to improved thermal image super-resolution performance given a similar parameter budget.

Table 6. Effect of hypercomplex representations in the proposed fusion network. QFusionNet matches the channel dimensions of the baseline, whereas QFusionNet⁺ tries to match its parameter count.

Model	Parameters	PSNR	SSIM
FusionNet	20.8K	37.6891	0.9746
QFusionNet	5.3K	37.6149	0.9742
QFusionNet ⁺	19.8K	37.7958	0.9749

6.4. Compared Methods

The taxonomy presented in Section 4 includes a broad range of thermal image super-resolution methods. Ideally, all reviewed methods would be evaluated under identical experimental conditions. In practice, this is not always feasible because some methods cannot be fairly reproduced or compared within the scope of this work.

The experimental evaluation therefore includes representative methods from each major category:

- **Classical interpolation:** Nearest Neighbor, Bicubic, and Lanczos;
- **Demosaicking-inspired methods:** the proposed Bilinear, Malvar–He–Cutler-style, and Menon adaptations;
- **CNN-based methods:** DRRN, IMDN, EDSR, PSRGAN, and RLKA-Net;
- **Transformer-based methods:** SwinIR;
- **State-space models:** IRSRMamba;
- **Proposed method:** FusionNet, QFusionNet, and QFusionNet⁺.

LKFormer is omitted because the authors have not officially released pretrained weights, making a reproducible evaluation difficult. Real-IISR is not included in the quantitative comparison because it targets real-world infrared super-resolution using a degradation model and paired real LR–HR benchmark that differ substantially from the synthetic bicubic ×2 protocol adopted in this study. Its autoregressive reconstruction and degradation-aware preprocessing therefore do not constitute a directly matched experimental setting.

6.5. Quantitative Results

Table 7 reports the average quantitative performance on the thermal building-inspection test set using PSNR, SSIM, RMSE, LPIPS, and NIQE.

Table 7. Average quantitative performance on the thermal building-inspection test set. Higher PSNR and SSIM are better, whereas lower RMSE, LPIPS, and NIQE are better.

Method	Params	PSNR	SSIM	RMSE	LPIPS	NIQE
Nearest Neighbor	-	30.4059	0.9170	8.0050	0.1228	11.7693
Bilinear	-	27.8215	0.8961	11.2472	0.1143	10.0420
Bicubic	-	30.4862	0.9111	8.4390	0.2306	7.2846
Lanczos	-	30.7143	0.9147	8.3081	0.2486	7.8083
Malvar-style [18,19]	-	25.4356	0.8807	14.2709	0.1620	10.3072
Menon-style [20]	-	28.7640	0.8758	10.5547	0.1312	9.4547
DRRN [26]	298K	32.8461	0.9402	6.6318	0.1304	7.3889
IMDN [28]	694K	33.4405	0.9387	6.6029	0.1123	6.8863
EDSR -baseline [27]	1.3M	33.4950	0.9405	6.4790	0.1209	7.0529
PSRGAN [29]	312K	30.3675	0.9056	9.1533	0.1239	6.3242
IRSRMamba [36]	26.46M	33.5411	0.9429	6.2702	0.1213	8.4404
SwinIR [34]	878K	33.6335	0.9414	6.3936	0.1080	6.9993
RLKA-Net [30]	167K	36.3688	0.9644	4.2540	0.0563	10.1624
FusionNet	20.8K	37.6891	0.9746	4.1379	0.0536	8.9763
QFusionNet	5.3K	37.6149	0.9742	3.6627	0.0391	8.8797
QFusionNet ⁺	19.8K	37.7958	0.9749	3.5934	0.0380	9.0544

Among the conventional interpolation baselines, Lanczos achieves the highest PSNR, while Nearest Neighbor provides the highest SSIM and the lowest RMSE and LPIPS. Bicubic obtains the lowest NIQE within this group. Among the demosaicking-derived adaptations, Menon-style reconstruction achieves the highest PSNR, lowest RMSE, and lowest NIQE, whereas the Bilinear adaptation provides the highest SSIM and lowest LPIPS.

The learning-based methods generally provide stronger full-reference reconstruction quality than the classical and demosaicking-derived approaches. Among the individual learned models, RLKA-Net achieves the highest PSNR and SSIM and the lowest RMSE and LPIPS. However, PSRGAN obtains the lowest NIQE, showing that rankings based on no-reference perceptual quality do not necessarily follow the same trend as full-reference fidelity measures.

The proposed FusionNet further improves upon RLKA-Net, achieving a PSNR of 37.6891 dB and an SSIM of 0.9746. The quaternion-based variants demonstrate two complementary advantages. QFusionNet reduces the parameter count from 20.8K to 5.3K while maintaining comparable reconstruction quality, achieving 37.6149 dB PSNR and 0.9742 SSIM. In contrast, QFusionNet⁺ uses a parameter budget comparable to FusionNet (19.8K) and achieves the highest PSNR and SSIM among the evaluated methods, reaching 37.7958 dB and 0.9749, respectively. These results indicate that quaternion-based representations can substantially improve parameter efficiency while preserving reconstruction fidelity and, when additional channel capacity is used, can slightly improve reconstruction performance over the real-valued FusionNet. PSRGAN nevertheless retains the best NIQE, highlighting the difference between distortion-based, learned perceptual, and no-reference quality measures.

6.6. Qualitative Discussion

Figures 9 and 10 present a qualitative comparison of the evaluated classical interpolation, demosaicking-inspired, learning-based, and fusion-based super-resolution methods on representative thermal images. The full images on the left indicate the selected regions of interest, while the enlarged crops on the right facilitate a detailed comparison of local structures, edges, textures, and thermal details.

Among the classical interpolation methods, Nearest Neighbor preserves sharp transitions but introduces noticeable aliasing, whereas Bilinear, Bicubic, and Lanczos produce smoother reconstructions with progressively improved edge continuity. The proposed demosaicking-inspired methods further enhance local structure preservation, with the Menon adaptation generally producing sharper directional features than the Bilinear and Malvar-style variants.

The learning-based methods consistently outperform the interpolation-based approaches by recovering clearer structures and more accurate thermal boundaries. Among the individual learning-based SR models shown in Figures 9 and 10, RLKA-Net provides the strongest reconstruction. FusionNet further combines the complementary information provided by the interpolation-based branches and RLKA-Net, producing sharp edges, well-preserved thermal structures, and fewer visible reconstruction artifacts.

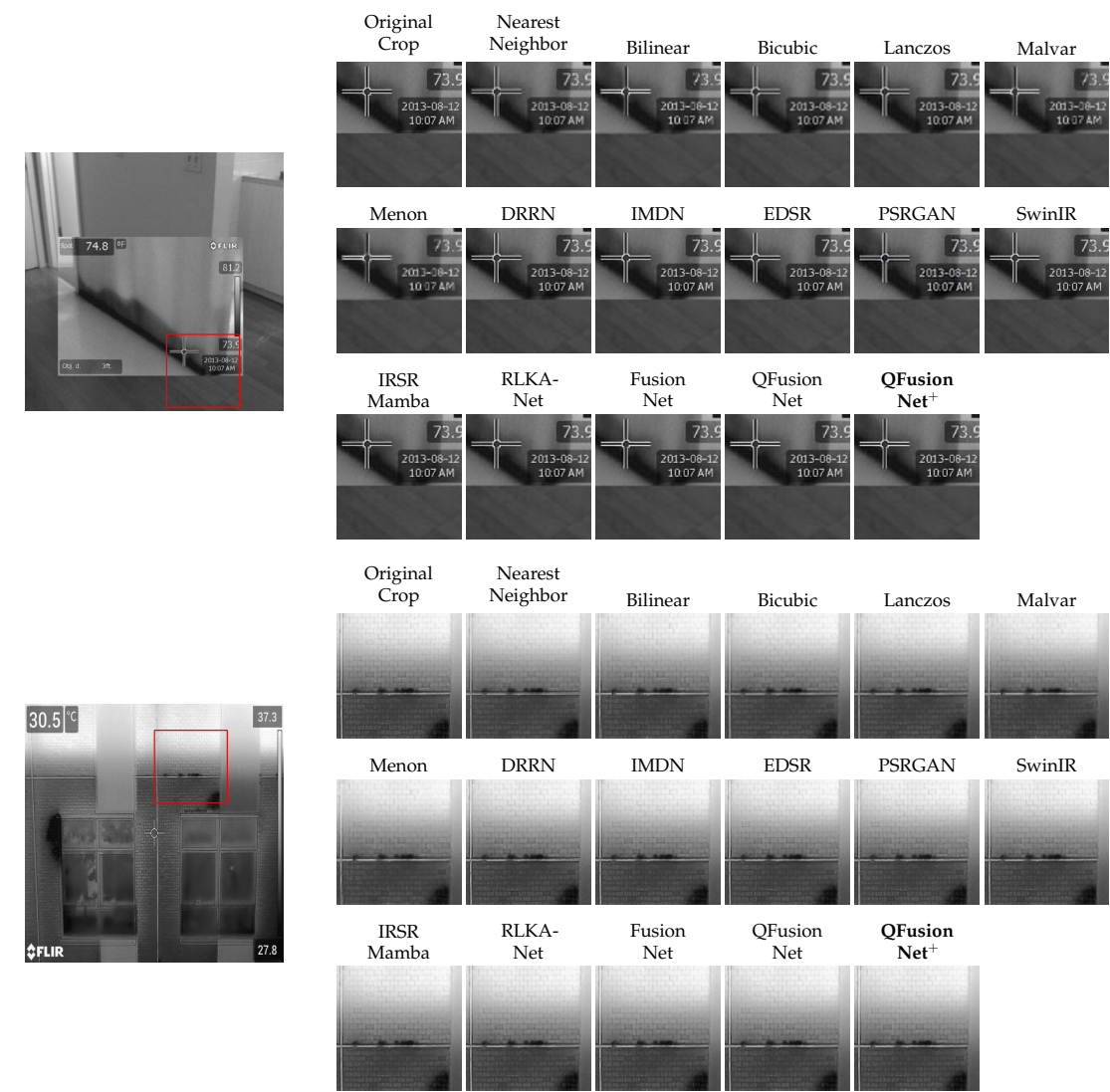


Figure 9. Visual comparison of evaluated super-resolution methods (Part 1). Full images (left) show evaluated regions; enlarged crops (right) highlight differences in fine-detail reconstruction.

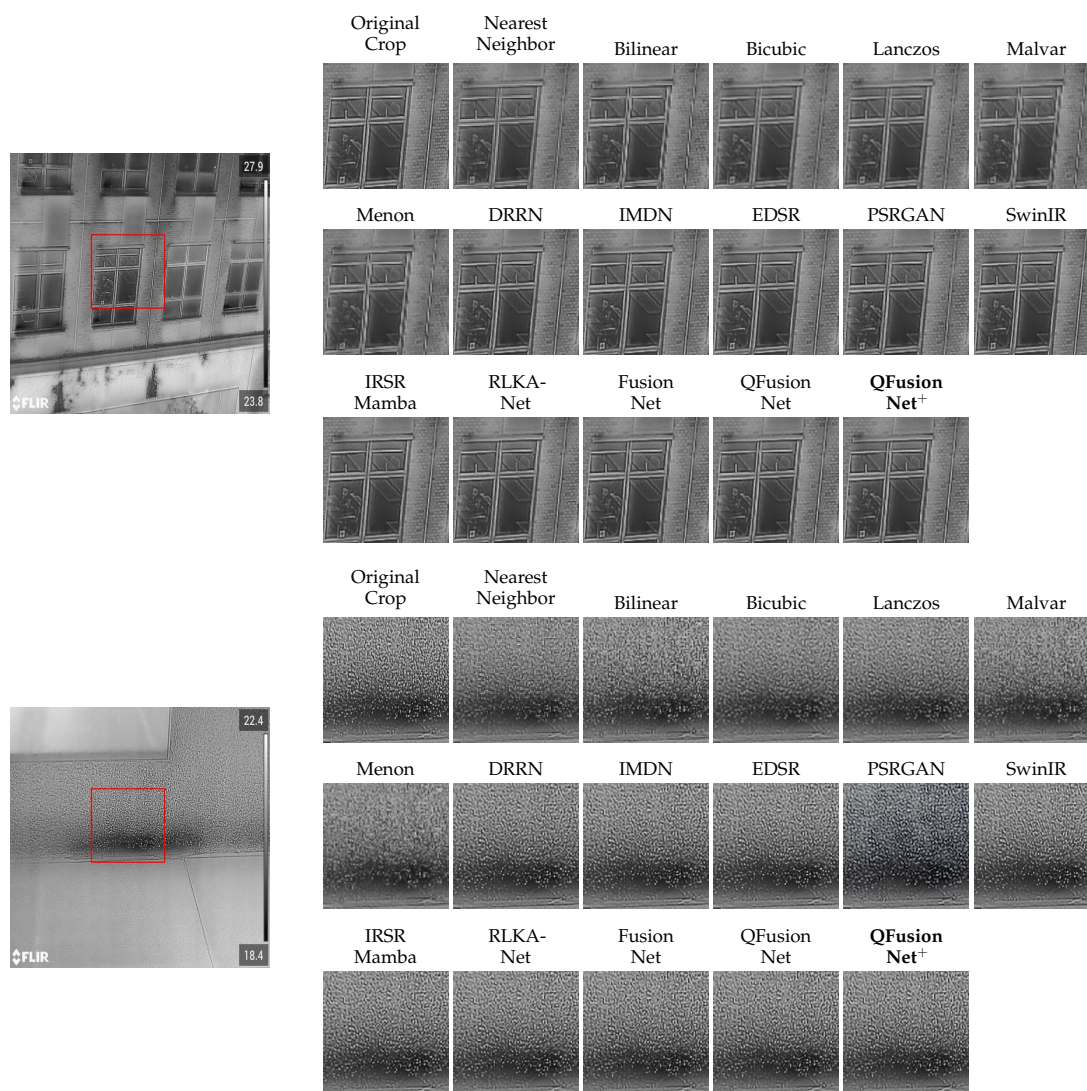


Figure 10. Visual comparison of evaluated super-resolution methods (Part 2). Full images (left) show evaluated regions; enlarged crops (right) highlight differences in fine-detail reconstruction.

7. Open Challenges and Future Directions

7.1. Limited Thermal SR Datasets

A major limitation in thermal SR is the lack of large paired LR–HR datasets. Many existing thermal datasets were originally developed for detection, fusion, surveillance, or autonomous driving rather than super-resolution. Although synthetic bicubic downsampling enables controlled evaluation, it does not fully represent the degradation produced by real thermal imaging systems.

7.2. Real-World Degradation Modeling

Many thermal SR studies rely on simplified degradation models, particularly bicubic downsampling. In practice, thermal images may be affected by optical blur, motion blur, sensor noise, non-uniformity, temperature drift, compression, and camera-specific processing. Future research should therefore focus on realistic degradation models and paired real-world thermal LR–HR datasets.

7.3. Cross-Spectral Guided Super-Resolution

Recent PBVS challenge settings demonstrate increasing interest in RGB-guided thermal super-resolution. Since RGB images generally contain sharper boundaries and richer spatial textures, they can provide useful structural guidance. However, cross-spectral reconstruction requires accurate

registration and careful fusion because visible-spectrum structures do not always correspond directly to thermal patterns.

Furthermore, standard cross-spectral fusion approaches treat visible and thermal channels as separate, independent feature maps, which can lead to feature degradation and loss of spatial-spectral consistency. A promising future direction to address this is the adoption of hypercomplex representations, such as Quaternion Super-Resolution (QSR) [47]. By encoding the thermal band along with the three visible channels (red, green, blue) as a single quaternion entity (e.g., $q = I_T + I_R\mathbf{i} + I_G\mathbf{j} + I_B\mathbf{k}$), hypercomplex networks can leverage the Hamilton product to model cross-spectral correlations directly within the algebraic structure, preserving the boundaries between color and thermal signals without treating them as independent feature maps.

7.4. Efficient Models for Practical Deployment

Thermal cameras are frequently deployed in embedded, mobile, or real-time systems. Therefore, reconstruction accuracy should be balanced with computational cost, memory consumption, and inference speed. Lightweight CNNs, efficient transformers, state-space models, pruning, quantization, and knowledge distillation represent promising directions for practical thermal SR.

7.5. Physical and Thermal Consistency

Thermal image intensities encode temperature-related radiometric information. Consequently, super-resolution methods should avoid generating visually plausible but physically misleading structures. Future methods should better preserve thermal ordering, temperature-related relationships, structural boundaries, and defect-relevant information.

7.6. Uncertainty-Aware Fusion

FusionNet predicts pixel-wise softmax weights for combining multiple reconstruction candidates. Future fusion methods could additionally incorporate uncertainty estimation, confidence-aware weighting, attention mechanisms, or iterative refinement. Such strategies may improve robustness in regions where the individual reconstruction methods provide conflicting estimates.

7.7. Evaluation Beyond PSNR and SSIM

PSNR and SSIM remain widely used, but they do not fully characterize perceptual quality or practical thermal-image fidelity. Future evaluation should combine full-reference, perceptual, no-reference, efficiency, and task-level measures, while also considering whether improvements in reconstructed images benefit downstream applications such as defect detection, segmentation, or inspection.

8. Conclusions

This paper reviewed thermal and infrared image super-resolution from the perspectives of datasets, evaluation metrics, reconstruction methods, and open research challenges. The reviewed approaches include conventional interpolation, demosaicking-derived reconstruction, CNNs, GANs, transformers, state-space models, diffusion and autoregressive methods, and fusion-based approaches. Thermal SR remains challenging because of low spatial resolution, weak textures, sensor-dependent degradation, limited paired datasets, and the need to preserve physically meaningful thermal information.

In addition to the review, this paper introduced a demosaicking-inspired thermal super-resolution framework. Bilinear, Malvar–He–Cutler-style, and Menon directional reconstruction were reformulated for single-channel thermal upsampling using zero insertion. Their outputs were combined with RLKA-Net using the proposed FusionNet, which learns spatially adaptive pixel-wise fusion weights.

Experiments on a thermal building-inspection dataset showed that FusionNet achieved 37.6891 dB PSNR and 0.9746 SSIM, outperforming the individual reconstruction methods used in the fusion framework. The hypercomplex analysis further demonstrated two complementary advantages of

quaternion-based fusion. QFusionNet reduced the parameter count from 20.8K to 5.3K while maintaining comparable reconstruction quality, whereas QFusionNet⁺ achieved the best overall PSNR, SSIM, RMSE, and LPIPS among the evaluated methods, reaching 37.7958 dB PSNR and 0.9749 SSIM. PSRGAN achieved the lowest NIQE, illustrating that full-reference and no-reference quality measures may rank reconstruction methods differently. These results highlight the potential of quaternion-based fusion for both parameter-efficient and high-fidelity thermal super-resolution. **Author Contributions:**

All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Vollmer, M.; Möllmann, K.P. *Infrared Thermal Imaging: Fundamentals, Research and Applications*; John Wiley & Sons, 2011. <https://doi.org/10.1002/9783527630868.ch1>.
2. Meola, C., Ed. *Infrared thermography: recent advances and future trends*; Bentham Science: New York, 2012. <https://doi.org/10.2174/97816080514341120101>.
3. Thermal Imaging in Building Inspection Dataset. <https://universe.roboflow.com/test-x7ahc/thermal-imaging-in-building>, 2024.
4. Dong, C.; Loy, C.C.; Tang, X. Accelerating the Super-Resolution Convolutional Neural Network. In Proceedings of the Computer Vision – ECCV 2016. Springer, 2016, pp. 391–407. https://doi.org/10.1007/978-3-319-46475-6_25.
5. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 105–114.
6. Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.H.; Liao, Q. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia* **2019**, *21*, 3106–3121.
7. Bashir, S.M.; Wang, Y.; Khan, M.; Niu, Y. A comprehensive review of deep learning-based single image super-resolution. *PeerJ Computer Science* **2021**, *7*, e621. <https://doi.org/10.7287/peerj-cs.621v0.1/reviews/1>.
8. Teledyne FLIR. FLIR Releases Starter Thermal Imaging Dataset for Machine Learning Advanced Driver Assistance Development. Available online: <https://www.flir.com/news-center/press-releases/flir-releases-starter-thermal-imaging-dataset-for-machine-learning-advanced-driver-assistance-development/>, 2018.
9. Choi, Y.; Kim, N.; Hwang, S.; Park, K.; Yoon, J.S.; An, K.; Kweon, I.S. KAIST multi-spectral day/night data set for autonomous and assisted driving. *IEEE Transactions on Intelligent Transportation Systems* **2018**, *19*, 934–948. <https://doi.org/10.1109/tits.2018.2791533>.
10. Treible, W.; Saponaro, P.; Sorensen, S.; Kolagunda, A.; O’Neal, M.; Phelan, B.; Sherbondy, K.; Kambhamettu, C. Cats: A color and thermal stereo benchmark. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2961–2969. <https://doi.org/10.1109/cvpr.2017.22>.
11. Jia, X.; Zhu, C.; Li, M.; Tang, W.; Zhou, W. LLVIP: A visible-infrared paired dataset for low-light vision. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). IEEE, Oct 2021, pp. 3489–3497. <https://doi.org/10.1109/iccvw54120.2021.00389>.
12. Rivadeneira, R.E.; Sappa, A.D.; Vintimilla, B.X.; Nathan, S.; Kansal, P.; Mehri, A.; Ardakani, P.B.; Dalal, A.; Sharma, D.; Pandey, S. Thermal image super-resolution challenge-pbvs 2021. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Jun 2021, pp. 4354–4362. <https://doi.org/10.1109/cvprw53098.2021.00492>.
13. Rivadeneira, R.E.; Sappa, A.D.; Wang, C.; Jiang, J.; Zhong, Z.; Chen, P.; Wang, S. Thermal image super-resolution challenge results-pbvs 2024. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Jun 2024, pp. 3113–3122.

14. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Jun 2018, pp. 586–595.
15. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters* **2012**, *20*, 209–212. <https://doi.org/10.1109/lsp.2012.2227726>.
16. Blau, Y.; Mechrez, R.; Timofte, R.; Michaeli, T.; Zelnik-Manor, L. The 2018 PIRM challenge on perceptual image super-resolution. In Proceedings of the European Conference on Computer Vision. Springer, Sep 2018, pp. 334–355. https://doi.org/10.1007/978-3-030-11021-5_21.
17. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 4th ed.; Pearson, 2018.
18. Getreuer, P. Malvar-he-cutler linear image demosaicking. *Image Processing on Line* **2011**, *1*, 83–89. https://doi.org/10.5201/ipol.2011.g_mhcd.
19. Malvar, H.S.; He, L.W.; Cutler, R. High-quality linear interpolation for demosaicking of Bayer-patterned color images. In Proceedings of the Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). IEEE, May 2004, Vol. 3, pp. iii–485. <https://doi.org/10.1109/icassp.2004.1326587>.
20. Menon, D.; Andriani, S.; Calvagno, G. Demosaicking with directional filtering and a posteriori decision. *IEEE Transactions on Image Processing* **2006**, *16*, 132–141. <https://doi.org/10.1109/tip.2006.884928>.
21. Chudasama, V.; Patel, H.; Prajapati, K.; Upla, K.P.; Ramachandra, R.; Raja, K.; Busch, C. TherISuRNet: A Computationally Efficient Thermal Image Super-Resolution Network. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 388–397. <https://doi.org/10.1109/cvprw50498.2020.00051>.
22. Kansal, P.; Nathan, S. A Multi-Level Supervision Model: A Novel Approach for Thermal Image Super Resolution. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 426–431. <https://doi.org/10.1109/cvprw50498.2020.00055>.
23. Deepak, S.; Sahoo, S.; Patra, D. Superresolution of Thermal Images Using GAN Network. In Proceedings of the 2021 Advanced Communication Technologies and Signal Processing (ACTS). IEEE, 2021, pp. 1–5. <https://doi.org/10.1109/acts53447.2021.9708340>.
24. Cheng, L.; Kersemans, M. Dual-IRT-GAN: A Defect-Aware Deep Adversarial Network to Perform Super-Resolution Tasks in Infrared Thermographic Inspection. *Composites Part B: Engineering* **2022**, *247*, 110309. <https://doi.org/10.1016/j.compositesb.2022.110309>.
25. Liu, Q.M.; Jia, R.S.; Liu, Y.B.; Sun, H.B.; Yu, J.Z.; Sun, H.M. Infrared Image Superresolution Reconstruction by Using Generative Adversarial Network with an Attention Mechanism. *Applied Intelligence* **2021**, *51*, 2018–2030. <https://doi.org/10.1007/s10489-020-01987-8>.
26. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 3147–3155. <https://doi.org/10.1109/cvpr.2017.298>.
27. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced deep residual networks for single image super-resolution. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 136–144. <https://doi.org/10.1109/cvprw.2017.151>.
28. Hui, Z.; Gao, X.; Yang, Y.; Wang, X. Lightweight image super-resolution with information multi-distillation network. In Proceedings of the Proceedings of the ACM International Conference on Multimedia, Oct 2019, pp. 2024–2032. <https://doi.org/10.1145/3343031.3351084>.
29. Huang, Y.; Jiang, Z.; Lan, R.; Zhang, S.; Pi, K. Infrared image super-resolution via transfer learning and PSRGAN. *IEEE Signal Processing Letters* **2021**, *28*, 982–986. <https://doi.org/10.1109/lsp.2021.3077801>.
30. Liu, G.; Zhou, S.; Chen, X.; Yue, W.; Ke, J. Recurrent large kernel attention network for efficient single infrared image super-resolution. *IEEE Access* **2023**, *12*, 923–935. <https://doi.org/10.1109/access.2023.3344830>.
31. Liang, S.; Song, K.; Zhao, W.; Li, S.; Yan, Y. DASR: Dual-Attention Transformer for Infrared Image Super-Resolution. *Infrared Physics & Technology* **2023**, *133*, 104837. <https://doi.org/10.1016/j.infrared.2023.104837>.
32. Hu, L.; Hu, L.; Chen, M. Edge-Enhanced Infrared Image Super-Resolution Reconstruction Model Under Transformer. *Scientific Reports* **2024**, *14*, 15585. <https://doi.org/10.1038/s41598-024-66302-8>.
33. Jiang, H.; Chen, Z. Flexible Window-Based Self-Attention Transformer in Thermal Image Super-Resolution. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2024, pp. 3076–3085. <https://doi.org/10.1109/cvprw63382.2024.00313>.

34. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. SwinIR: Image restoration using Swin Transformer. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 1833–1844. <https://doi.org/10.1109/iccvw54120.2021.00210>.
35. Qin, F.; Yan, K.; Wang, C.; Ge, R.; Peng, Y.; Zhang, K. LKFormer: Large kernel transformer for infrared image super-resolution. *Multimedia Tools and Applications* **2024**, *83*, 72063–72077. <https://doi.org/10.1007/s11042-024-18409-3>.
36. Huang, Y.; Miyazaki, T.; Liu, X.; Omachi, S. Irsrmamba: Infrared image super-resolution via mamba-based wavelet transform feature modulation model. *IEEE Transactions on Geoscience and Remote Sensing* **2025**, *63*, 1–6. <https://doi.org/10.1109/tgrs.2025.3584385>.
37. Zou, Y.; Ma, J.; Jiao, Z.; Li, X.; Jiang, Z.; Liu, J. Toward real-world infrared image super-resolution: A unified autoregressive framework and benchmark dataset. *arXiv preprint arXiv:2603.04745* **2026**. <https://doi.org/10.1109/iccv.2019.00318>.
38. Hamilton, W.R. On quaternions; or on a new system of imaginaries in algebra. *Letter to John T. Graves* **1843**.
39. Le Bihan, N.; Sangwine, S.J. Quaternion principal component analysis of color images. In Proceedings of the Proceedings of the International Conference on Image Processing (ICIP). IEEE, 2003, Vol. 1, pp. I–809.
40. Greenblatt, C.; Mosquera-Lopez, C.; Agaian, S.S. Quaternion Neural Networks Applied to Prostate Cancer Gleason Grading. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 2013; pp. 1144–1149. <https://doi.org/10.1109/SMC.2013.199>.
41. Grigoryan, A.M.; Agaian, S.S. *Quaternion and Octonion Color Image Processing with MATLAB*; SPIE Press, 2018; p. 404.
42. Frants, V.; Agaian, S.; Panetta, K. QSAM-Net: Rain Streak Removal by Quaternion Neural Network With Self-Attention Module. *IEEE Transactions on Multimedia* **2023**, *26*, 789–798. <https://doi.org/10.1109/TMM.2023.3271829>.
43. Frants, V.; Agaian, S.; Panetta, K.; Grigoryan, A. QWR-Dec-Net: A Quaternion-Wavelet Retinex Framework for Low-Light Image Enhancement with Applications to Remote Sensing. *Information* **2026**, *17*, 89. <https://doi.org/10.3390/info17010089>.
44. Gaudet, C.J.; Maida, A.S. Deep quaternion networks. In Proceedings of the Proceedings of the International Joint Conference on Neural Networks (IJCNN). IEEE, 2018, pp. 1–8.
45. Zhu, X.; Xu, Y.; Xu, H.; Chen, C. Quaternion convolutional neural networks. In Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 631–647.
46. KM, S.K.; Rao, S.P.; Panetta, K.; Agaian, S.S. QSRNet: Towards quaternion-based single image super-resolution. In Proceedings of the Multimodal Image Exploitation and Learning 2022. SPIE, 2022, Vol. 12100, pp. 192–205.
47. Parcollet, T.; Morchid, M.; Linares, G. Quaternion convolutional neural networks for heterogeneous image processing. In Proceedings of the Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 8514–8518.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.