

Infrared Image Super-Resolution: Systematic Review, and Future Trends

Yongsong Huang^a, Tomo Miyazaki^a, Xiaofeng Liu^b, Shinichiro Omachi^a

^a*Graduate School of Engineering, Tohoku University, Sendai, 9808579, Japan*

^b*Department of Radiology and Biomedical Imaging, Yale University, New Haven, 06519, USA*

Abstract

Image Super-Resolution (SR) is essential for a wide range of computer vision and image processing tasks. Investigating infrared (IR) image (or thermal images) super-resolution is a continuing concern within the development of deep learning. This survey aims to provide a comprehensive perspective of IR image super-resolution, including its applications, hardware imaging system dilemmas, and taxonomy of image processing methodologies. In addition, the datasets and evaluation metrics in IR image super-resolution tasks are also discussed. Furthermore, the deficiencies in current technologies and possible promising directions for the community to explore are highlighted. To cope with the rapid development in this field, we intend to regularly update the relevant excellent work at https://github.com/yongsongH/Infrared_Image_SR_Survey.

Keywords:

Image Super-resolution, Deep Learning, Convolutional Neural Network, Generative Adversarial Nets, Pattern Recognition

1. Introduction

Image super-resolution reconstruction (SR) is a fundamental low-level vision task in image processing [117, 80, 16, 99, 83]. The objective of infrared (IR) image super-resolution is to reconstruct high-resolution (HR) IR images from low-resolution (LR) IR inputs, a problem that remains ill-posed due to diverse degradation paths [117, 116, 126, 127, 80]. Mathematically, the degradation process can be formulated as:

$$I_{LR} = \mathbb{D}(I_{HR}; \delta) \quad (1)$$

where \mathbb{D} denotes a degradation function, I_{HR} is the high-resolution IR image, I_{LR} is the low-resolution IR image, and δ represents the degradation parameters. The degradation process generally consists of three key factors: 1) downsampling, 2) noise, and 3) compression, as defined in Eq. 2.

$$\mathbb{D}(I_{HR}; \delta) = (I_{HR} \otimes \kappa) \downarrow_d + n_\varsigma, \{\kappa, \varsigma\} \subset \delta \quad (2)$$

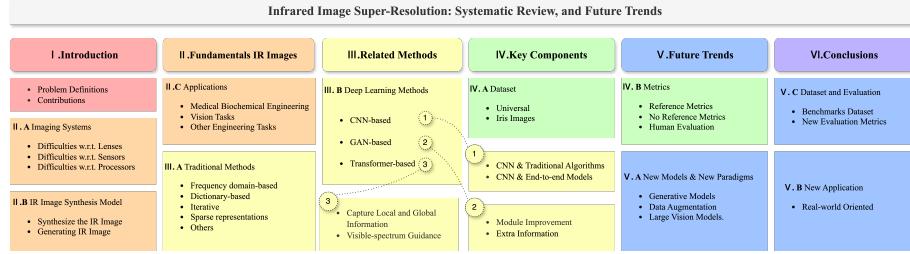


Figure 1: Hierarchical and structural taxonomy of this survey.

where $I_{HR} \otimes \kappa$ represents the convolution between a blur kernel κ and the HR image I_{HR} . The degradation process includes noise and compression artifacts. The downsampling factor \downarrow_d (e.g., $4\times$ or $8\times$) further reduces the resolution, while n_ς represents additive white Gaussian noise with standard deviation ς . Given this degradation model, the goal of IRSR can be formulated as an optimization problem:

$$\hat{\theta} = \arg \min_{\theta} \mathcal{L}(I_{HR}, I_{SR}) + \lambda \Phi(\theta) \quad (3)$$

where \mathcal{L} is the loss function measuring the discrepancy between the HR image I_{HR} and the super-resolved image I_{SR} . The regularization term $\Phi(\theta)$ and weighting parameter λ control the model complexity.

Early approaches applied general SR techniques designed for visible-light images to IRSR, utilizing convolutional neural networks (CNNs) [22, 140], generative adversarial networks (GANs) [47, 117, 114], and Transformers [53]. However, these methods often fail to generalize due to the distinct physical properties of IR imaging. Unlike visible-light images, IR images exhibit sensor-induced noise, spectral dependency, and contrast degradation, making traditional SR strategies less effective.

The degradation process in IR images is more complex due to several reasons. First, the blur kernel commonly used in visible-light image reconstruction, often modeled as Gaussian blur with JPEG compression [27, 114, 64, 136, 84], is insufficient to represent real-world IR imaging conditions, which are affected by sensor limitations and environmental factors. Second, IR images typically exhibit low color contrast, weak edge gradients, and overlapping information between high- and low-frequency components [115, 107, 130, 4]. Additionally, many IR cameras are deployed in outdoor environments, further increasing the complexity of degradation. These fundamental differences necessitate domain-specific super-resolution reconstruction approaches for IR images, which we discuss in subsequent sections.

In this paper, we systematically review the applications, methods, and challenges in IRSR, targeting both academic researchers and industrial practitioners. This survey aims to bridge the gap between conventional SR techniques and the unique characteristics of IR imaging, offering insights into the development of

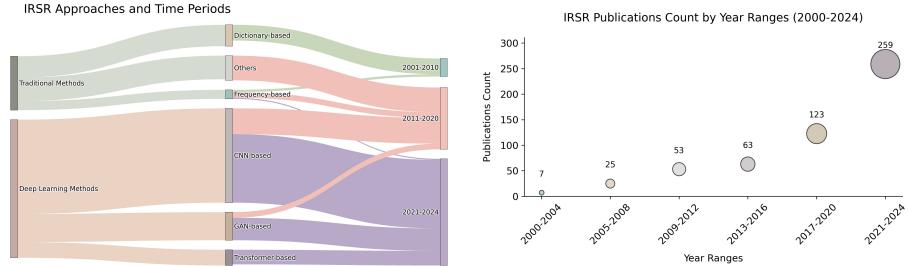


Figure 2: [Left] Evolution of IRSR methodologies over time. [Right] Number of IRSR-related publications indexed in Web of Science from 2000 to 2024.

domain-specific algorithms. For researchers, we provide a comprehensive analysis of existing methodologies and unresolved challenges. For industrial practitioners, we highlight practical considerations such as hardware constraints and deployment challenges, with guidance on optimizing IRSR techniques for applications in autonomous driving, medical imaging, and remote sensing. By analyzing existing approaches and identifying open issues, this survey serves as a reference for advancing research and applications in IR image super-resolution. The hierarchical and structural taxonomy of this survey is shown in Fig.1.

Unlike previous surveys, our work places a stronger emphasis on deep-learning-based IRSR techniques and the unique statistical properties of IR images that distinguish them from visible-light images. This distinction is critical, as directly applying visible-light SR models to IR images often results in poor performance. Additionally, we offer a detailed discussion on why hardware-based enhancements alone are insufficient for improving IR image quality. Unlike previous SR surveys [117, 80, 16, 110, 20], this paper focuses on how the unique degradation patterns, spectral dependencies, and sensor characteristics of IR images shape IRSR methodologies.

The left panel of Fig. 2 categorizes IRSR methodologies into traditional techniques, CNN-based models, GAN-based approaches, and Transformer architectures. Traditional frequency-based and dictionary-based methods dominated the field from 2000 to 2010, focusing on sparse representation and frequency-domain modeling. The introduction of CNNs in 2011 significantly improved SR performance, leading to nearly a decade of dominance in the field. More recently, GAN-based approaches have demonstrated superior perceptual quality, particularly in texture synthesis for low-contrast IR images, but remain sensitive to domain shifts. Transformer-based models, on the other hand, provide improved global contextual modeling, making them particularly effective for preserving structural consistency in IRSR.

The right panel of Fig. 2 presents the increasing number of IRSR publications from 2000 to 2024, highlighting the field's rapid growth. The significant rise in publications from 2017 to 2024 corresponds to the adoption of deep-learning-based methods and their expanding applications in real-world scenarios. This trend underscores the growing importance of IRSR research, making this survey

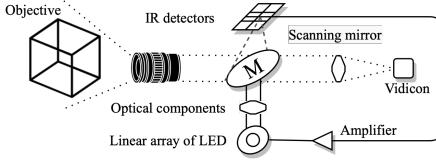


Figure 3: Explanation of the operation mechanism in a FLIR system.

a timely and valuable resource.

The remainder of this paper is structured as follows: Section 2 introduces the fundamentals of IR imaging and its applications. Section 3 reviews IRSR methodologies. Section 4 presents datasets and evaluation metrics. Section 5 discusses challenges and future directions. Finally, Section 6 concludes the survey.

2. Fundamentals of Infrared Images

In this section, we briefly introduce the fundamentals of infrared (IR) images, focusing on key components of IR imaging systems and an overview of traditional IR image synthesis models. These discussions highlight the inherent challenges—both from hardware limitations and the simplified physical modeling—that motivate the need for advanced processing techniques.

2.1. Imaging Systems

Traditional IR imaging systems rely on a few basic components, including lenses, sensors, and processors [80, 46]. As shown in Fig. 3, a classic forward-looking infrared (FLIR) system converts a scene into a digital image through the following steps: the objective forms an image on a linear array of IR detectors; each detector’s amplified signal drives a corresponding LED; the LED array is reflected by a scanning mirror onto an image tube, and finally the resulting signal is displayed on a cathode ray tube. Modern systems commonly employ solid-state detectors to enhance performance.

Despite these advancements, improving IR image quality via hardware enhancements faces substantial challenges: **Lenses:** Increasing the amount of light captured requires larger lenses, which not only raises costs and complicates design but also enlarges detector size, hindering system integration—especially in dynamic environments [46]. **Sensors:** IR sensors inherently produce fixed pattern noise (FPN) and other artifacts that are difficult to fully eliminate due to fundamental physical limitations [43, 112], [46]. **Processors:** Many existing systems use traditional microprocessors that struggle to run sophisticated image processing algorithms in real time, further limiting performance [132, 103].

These hardware constraints emphasize that direct improvements in imaging quality via physical enhancements are both costly and limited, thereby motivating advanced computational approaches.

Table 1: IR image super-resolution applications.

Medical biochemical engineering	Vision tasks	Other engineering tasks	
Pharmaceutical industry[77, 78]	Image conversion[48]	Automated vehicle[32, 95]	Food quality control[71]
Medical science[106, 12]	Multispectral matching[11]	Remote sensing[124, 76, 56]	Agricultural management[14, 60]
Cellular observations[111, 10]	Targets detection[51]	Terrain models[56]	Water resource management[62, 82]
Fluorescence microscopy[63, 72]	Face recognition[105, 86]	Land surface[18, 3]	Star formation[125]

2.2. IR Image Synthesis Model

Beyond hardware, the quality of IR images is also determined by the underlying synthesis models. Traditional IR image synthesis attempts to simulate the imaging process using physical models—often based on radiance estimation, heat transfer, and atmospheric effects [134]. Early models approximate Planck’s Law and employ simplified integrations to estimate surface radiance. However, such models usually fall short because they: Simplify complex heat transfer processes by neglecting interactions between different surface facets. Rely on basic coloring and degradation assumptions that do not capture the limited color range and overlapping frequency information characteristic of IR images. Thus, while these models provide a theoretical basis for IR image generation, they cannot fully account for the unique degradation and limited information in IR images. To summarize, the enhancement of IR images faces significant challenges, primarily due to hardware constraints arising from physical limitations and the complexity of image synthesis models. Further, we can summarize the unique patterns in IR images from the model used to generated these images:

- **Blur kernel.** The blur kernel employed in the reconstruction of visible images is comparatively straightforward, typically encompassing Gaussian noise and JPEG compression [27, 114, 64, 136, 84]. However, these blur kernels lack the robustness required for IR images to accurately depict complex real-world scenarios or heat transfer model.
- **Degradation.** The deployment of IR cameras in uncontrolled, natural environments further exacerbates the complexity of IR image degradation.
- **Space information limited.** Because of the coloring model in IR image generating, the IR image presents simple color[27, 115], insignificant gradients[107, 130, 33], and information overlap between high and low frequencies[70]. Increased focus on these unique patterns will benefit to models applicable to IR images being proposed.

Based on the above concerns, we would resort to the subsequent image processing technology to improve the IR image quality. In the next sections, we will present IR image super-resolution methods, more economical, for applications in different fields.

2.3. Applications

IR image super-resolution techniques have critical applications across various fields including medical biochemical engineering, vision tasks, and other

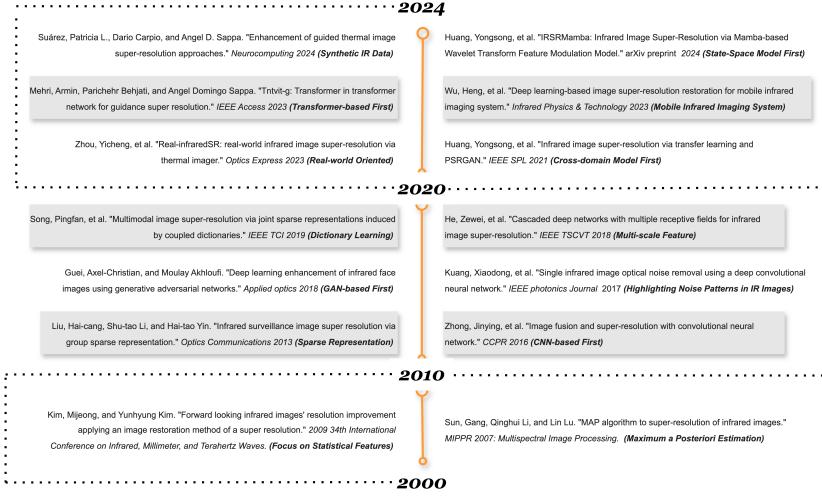


Figure 4: The timeline of representative IRSR algorithms (from year 2000 to 2024).

engineering domains (see Tab. 1). The practical impact of these techniques ranges from enhancing diagnostic imaging and assisting in disaster response to improving remote sensing and automated vehicle performance.

3. Related Methods

In this section, we will introduce IR image super-resolution methods, including traditional methods and deep learning-based methods. For the traditional algorithms, there will be three parts: frequency domain-based, dictionary-based, and other methods. Then, the deep learning methods will be presented.

Fig.4 provides a chronological overview of representative IRSR algorithms, illustrating the key advancements in this field from 2000 to 2024. The timeline highlights how traditional methods, such as frequency domain-based approaches and dictionary learning algorithms, laid the foundation for early IRSR research. These methods, prevalent from 2000 to 2010, primarily focused on mathematical models and handcrafted features to enhance IR image quality[42, 102, 58]. However, they often faced limitations in handling complex noise patterns[44] and adapting to diverse real-world scenarios. This paved the way for the adoption of deep learning approaches, starting in 2011, particularly CNN-based methods[142]. By learning hierarchical feature representations, these methods achieved significant improvements in reconstruction quality and robustness, marking a turning point in the development of IRSR techniques.

In recent years, GANs and Transformer-based models have pushed the boundaries of IRSR methodologies even further. GANs, introduced around 2018, enabled realistic texture generation, effectively addressing the unique noise characteristics of IR images[29]. Meanwhile, Transformer-based models, exemplified

Table 2: Performance comparison [in terms of PSNR (dB) and SSIM] over selected test dataset for 4 or 3 upscaling factors. Some of these works use visible images[1] as training datasets, such as 7 and 9. The test datasets for these representative methods depend on the task, check the references for more details.

No.	Methods	Year	Ref. \times 4 (PSNR/SSIM)	Keywords
1	Wu, Wenhao, et al. [120]	2022	31.69/0.7877	Meta-learning; Lightweight network
2	Huang, Yongsong, et al.[36]	2021	33.13/0.8282	Transfer learning; Small sample
3	Liu, Qing-Ming, et al. [59]	2021	32.23/0.8720	Attention mechanism; GAN
4	Prajapati, Kalpesh, et al.[84]	2021	34.90/0.9134	Attention mechanism; Channel splitting
5	Marivani, Iman, et al. [69]	2020	35.19/0.9888	Sparse coding; Image fusion
6	Marivani, Iman, et al. [70]	2020	34.49/0.9853	Multimodal image; Residual learning
7	Yao, Tingting, et al.[131]	2020	34.54/0.8807	Dictionary learning; CNN
8	Song, Pingfan, et al.[98]	2020	36.36/0.9796	Dictionary learning; Sparse representations
9	Batchulun, Ganbayar, et al. [9]	2020	22.99/0.9760	Image deblurring; GAN
10	Rivadeneira, Rafael E, et.al.[87]	2019	37.85/ \varnothing	Thermal images; CNN
11	Suryanarayana, Gunnam, et al.[104]	2019	31.40/0.9513	Discrete wavelet transform; CNN
12	Marivani, Iman, et al. [68]	2019	33.19/ \varnothing	Sparse coding; CNN
13	He, Zewei, et al.[34]	2018	36.02/0.9230	Multiple receptive fields; CNN
14	Sun, Chao, et al.[101]	2018	32.71/ \varnothing	Zoom mechanism; Transfer learning
15	Han, Tao Young, et al.[33]	2018	39.69/0.9582 ($\times 3$)	Frequency components; CNN
16	Zhao, Yao, et al. [141]	2016	30.69/0.9031 ($\times 3$)	Compressed sensing; Dictionary learning;

fied by recent works such as [73], have enhanced global contextual modeling capabilities, resulting in superior performance in preserving structural consistency and fine details. This evolution reflects a broader trend toward leveraging advanced deep learning architectures to overcome the inherent challenges of IRSR. Additionally, recent studies have extended the scope of IRSR by exploring cross-domain learning[36], mobile infrared imaging systems[119], and state-space modeling[38], demonstrating the growing practical relevance and applicability of these techniques.

Based on the advancements in IR imaging systems and deep learning techniques, researchers are increasingly focusing on developing more efficient image processing methods to enhance IR image quality further. As such, this survey categorizes image processing approaches into two primary types: traditional methods and deep learning-based techniques. To provide a clear comparison, Table 2 lists several representative works, summarizing their performance and contributions. This comparison not only highlights the progress made in the field but also underscores the evolving role of deep learning in addressing the specific challenges of IR image super-resolution.

3.1. Traditional Methods

Traditional IRSR methods largely follow the paradigm of visible image super-resolution and mainly include three types: frequency domain-based, dictionary-based, and other methods.

3.1.1. Frequency Domain-based Methods

These methods decompose the IR image into distinct frequency components (e.g., high and low frequencies) and enhance the high-frequency details via engineered priors. For instance, *Choi et al.* propose to first distinguish edge pixels

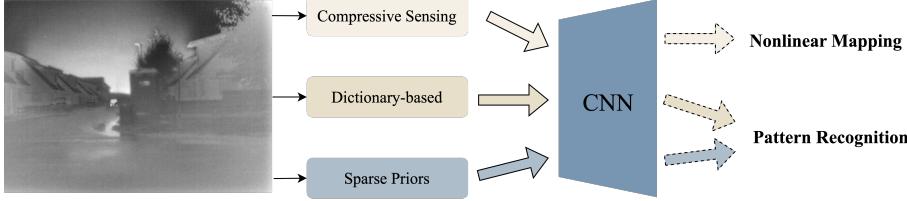


Figure 5: CNNs contribute to other traditional methods used to learn nonlinear mappings and pattern recognition.

and then independently enhance high-frequency information using a formulation such as:

$$e(\hat{k}, l) = \left(1 - e^{-\frac{(\lambda_1 + \lambda_2)}{\sigma_{\text{sum}}}}\right) \left(1 - e^{-\frac{(\lambda_1 + \varepsilon)/(\lambda_2 + \varepsilon)}{\sigma_{\text{ratio}}}}\right) \quad (4)$$

Here, λ_1 and λ_2 are eigenvalues from the structure tensor at pixel (\hat{k}, l) , and σ_{sum} , σ_{ratio} , and ε are preset parameters. Although effective in highlighting edges, such methods require carefully designed priors and face challenges (e.g., edge registration and acceleration) [80, 50, 17, 67].

3.1.2. Dictionary-based Methods

Dictionary-based approaches bridge the gap between low-resolution (LR) and high-resolution (HR) images via patch-based representations. These methods exploit inter-patch relationships—such as group similarity, multi-view, and multi-scale information—to construct dictionaries that robustly capture image structures [15, 19, 128, 129].

3.1.3. Other Traditional Algorithms

Additional approaches include iterative methods (e.g., for iris recognition [7, 2]), sparsity-based representations, projection techniques, and regularization methods [61, 31, 80, 133]. Although these methods offer good interpretability, their reliance on handcrafted features and simplistic priors limits their performance when addressing the complex, nonlinear degradations in IR images.

Table 3: CNN transfers on traditional methods.

Category	Method	Highlight
Compressive Sensing	Zhang, Xudong, et al. (2018)[138]	Nonlinear Mapping
Dictionary-based	Yao, Tingting, et al. (2020)[131]	
Sparse Priors	Marivani, Iman, et al. (2020)[69]	Pattern Recognition

Figure 5 and Table 3 illustrate how CNNs have been applied to enhance traditional methods by learning robust nonlinear mappings, thereby bridging the gap between conventional priors and data-driven approaches.

3.2. Deep Learning Methods

With the development of deep learning, neural networks have shown remarkable performance in image super-resolution with powerful fitting ability[22, 23, 41, 55, 113, 47]. However, few previously published studies have been able to draw on any systematic research into IR image-specific patterns. These unique patterns include gradient information, high or low-frequency information, extra information, etc[36, 35, 25, 139, 57]. In this section, we will also focus on the differences between IR images and visible images in the deep learning field of super-resolution. In summary, the challenges of IR image super-resolution in deep learning include specific patterns, difficulty in representing patterns, and poor image quality. More details will be shown in the next section.

3.2.1. CNN-based

There are two major trends for CNN-based models in IR image super-resolution: **1)** First, researchers introduced CNNs as contributors to improving traditional algorithms. **2)** Further, it became popular to use CNNs to build end-to-end models used in IR image super-resolution.

CNN & Traditional algorithms: As mentioned in Sec .3.1, traditional methods require enough prior knowledge to achieve image super-resolution. However, weak mathematical analysis limits the proposed algorithms. CNN can learn nonlinear mappings without indication because of its powerful fitting ability[22, 23, 41]. It allows people to free themselves from the task to seek for priors. Initially, neural networks were not used to reconstruct IR images directly. Rather, it was carefully used to help traditional algorithms improve performance through **nonlinear mapping and pattern recognition**.

1) Nonlinear mapping: In infrared camera systems, there will be little high-frequency information and low-frequency information in the captured infrared image compared to the visible image due to the poor imaging environment (see Sec.2.1). It becomes a challenge to better represent these two kinds of information in IR images super-resolution. Then, the researchers proposed that CNN can be used to represent the nonlinear mapping of low and high frequency information in the latent space. In *Zhang's work*[138], the SR image is reconstructed using compressed sensing first. Then the SR image and HR image are reduced to receive high-frequency noise information. Finally, the information is fed into CNN to learn nonlinear mapping.

$$\mathcal{L} = \frac{1}{2N} \sum_{i=1}^N \|\tilde{R}(\Theta, i) - R(\Theta, i)\|^2 \quad (5)$$

Eq.5 denotes the loss function to learn the trainable parameters Θ in CNN. Corresponding to i th training image, $\tilde{R}(\Theta, i)$ represents the estimated residual image produced by CNN, while $R(\Theta, i)$ represents the true residual image used for training[138]. Experiments show that this approach using CNNs to represent nonlinear mappings for reconstructing IR images can enhance detailed information.

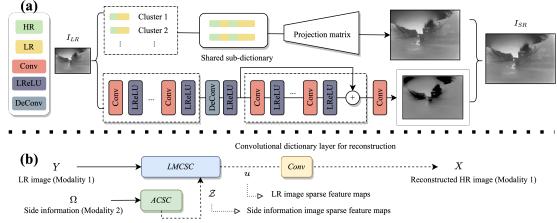


Figure 6: (a) shows the CNN used as a feature extractor in a dictionary-based approach; (b) is an illustration for the CNN used to represent prior.

2) Pattern recognition: On the other hand, *Yao, Tingtin, et al.* [131] proposed to use CNN as a feature extractor in the dictionary-based approach. This method also focuses on the problem that the detailed information in IR images is important, and CNN can build the dictionary with more details added to the representation. Novel interpretable operators are proposed to construct the basic module as the sparse representation for prior extraction, another module to represent the edge information of visible images, and finally fused in a residual network using skip connection[69].

More details are shown in Fig.6. For (b), \$X\$ and \$Y\$ denote the LR image, and reconstructed SR image, respectively. The convolutional sparse codes \$\mathcal{Z}\$ of the guidance HR image \$\Omega\$ are similar to \$u\$ by means of the \$\ell_1\$-norm according to the literature[69]. Furthermore, residual networks with multiple branches are used to reconstruct high and low-frequency information, respectively[130].

Beyond introducing CNNs, these existing methods started to introduce visible images to guide the reconstruction of IR images purposefully, which represents an entirely new direction. Further, attention has been focused on image characteristics, specifically the difference between high and low-frequency information in IR images. Details will be described in the next sections.

CNN & End-to-end models: In this section, the end-to-end models used in IR image super-resolution will be presented. The high-frequency information in IR images represents fine details, while the low-frequency information primarily captures the overall structure and contours of the image. Moreover, the IR image has less edge information compared to the visible image because of the imaging system (see Sec.2.1). First, the researchers propose that the independent extraction for high-frequency and low-frequency information in IR images is achieved by CNN. In [144], *Zou et al.* used residual networks to build a model similar to U-Net. The multi-receptive field module in this work is supposed to help represent the features of high and low-frequency information. Similar work is [25], and the residual network is also used. However, the information distillation approach is used in the backbone network, which was first proposed in [36] to benefit IR image super-resolution. This Information distillation-based model is also considered to be helpful for different information extraction. In addition, *Kwasniewska et al.* used a wide receptive field residual network constructed by dense connection[45]. According to the experimental results, this work con-

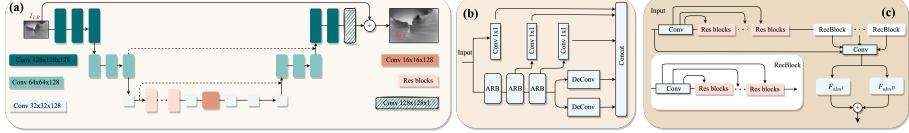


Figure 7: In (a), the researchers used a network similar to U-Net to extract high and low-frequency information independently[144]. For the same purpose, (b) demonstrates that the information distillation method can also be used[45]. Finally, the module in (c) shows that wide receptive fields are beneficial for low-contrast images[27].

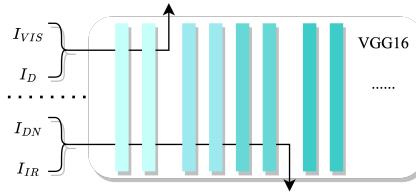


Figure 8: I_{DN} denotes the denoised output, while I_D is the output from the author's proposed network[130].

firmed that wide receptive fields are effectively used for low-contrast images. *Prajapati, Kalpesh, et al.* suggest that the common residual design allows for retaining too many redundant features in the network. A new module for integration with multiple attention mechanisms is proposed in their work. According to the experimental results, this novel attention mechanism module will help to enhance the high-frequency information representation[84]. Finally, not only the residual network is used, but also the importance of edges is observed[27]. The method that enhances high-frequency details and contacts information-tail modules to improve details is the highlight. More details are shown in Fig.7. And F_{nlm^D} is the output of the previous (*i.e.* D) recursive block.

The approaches mentioned above show that the two variables in the latent space (low-frequency and high-frequency information) are treated independently. Then novel methods are used to represent the nonlinear mapping between these two independent variables respectively. In the next section, we will describe another category of methods that introduce extra information: visible image information.

For introducing visible image information, the researchers focused on the following components: dataset, network structure, and loss function. It is natural to associate using visible images to train neural networks because they are cheap and easy to use, with rich detailed information. However, it is difficult to fit the data distribution for IR images with the model trained on the visible image dataset due to the domain transfer challenge[36, 87]. In other words, the SR images have poor quality. Then, adaptations to the loss function are proposed. As we all know, the loss function is important for optimizing neural networks. *Patel, Heena M., et al.* proposed that a combined loss function, with

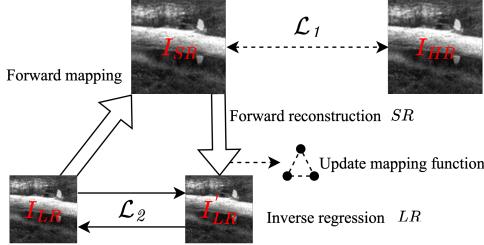


Figure 9: Regression network structure.

two terms for L1 and SSIM loss, could be used[81]. According to the ablation experiments, this combined loss function would be beneficial for their proposed improved dense blocks. On the other hand, it also inspired that appropriate loss functions would be useful for introducing information from visible images In [130], the authors propose that the perceptual similarity between IR images I_{IR} and paired visible images I_{VIS} can be captured by CNNs. And there are features present in the middle layer of the neural network and they can be used as loss functions (see Fig.8).

Next, we will introduce novel network structures for using visible image patterns. *Zou, Yan, et al.* proposed a dual-path residual network that directly fuses the features from visible and IR images in the channel[143]. Further, *Oz, Navot, et al.* consider that dimensionality compression will influence the performance in feature fusion. Because the correlation between adjacent pixels is ignored. Then, the dimensional change network for increasing channels was proposed[79].

Finally, we will also present models related to super-resolution in multi-vision tasks. In many vision tasks[108], such as segmentation and fusion, super-resolution is applied to help other subjects as a pre-processing task. In particular, the regression network proposed in [108] is dedicated to prevent the irrelevant function mapping space on the reconstructed images by using forward generation and backward regression. This double mapping constraint is used for IR image super-resolution, and then the output that has better image quality is used in the fusion task[26, 92]. More details can be found in Fig.9.

In summary, the approaches in IR images and CNNs have first tried to combine the pattern recognition capabilities of CNNs into traditional reconstruction methods, such as dictionary reconstruction and sparse coding. Then started to reconstruct IR images directly using end-to-end models. Many module-based improvement methods were proposed, such as multi-scale information extraction. However, the high and low-frequency information in IR images have a significant gap. To use this information, the model structure has been improved further, for example, by presenting the information separately in different modules or using more complex modules. Compared to IR images, visible images have information with rich details. For this reason, researchers have started to try to use visible images to help reconstruct IR images. It includes designing

new network structures and using strategies such as transfer learning. The purpose is to help neural networks can introduce more patterns and information from visible images.

3.2.2. GAN-based

After GAN[28] was proposed, research based on adversarial training model is emerging in the super-resolution field. SRGAN[47], ESRGAN[113], and various other types of GAN models[66] moved the field forward together through model improvement and mathematical analysis (WGAN[30]) modifications. The same attention has been focused on the application and research of GAN models in the IR image super-resolution.

Initially, GAN models using normal images have been directly used in IR image reconstruction. Researchers[94] used SRGAN straightly to reconstruct IR images and received reconstructed images that were quite acceptable. However, the blurred edges and unclear details are still a challenge. [29] used a modified DCGAN to reconstruct IR images, but the experimental results were not compared between the same category GAN methods. For this reason, it is difficult to describe the actual effectiveness. On the other hand, the work from [135] compared the existing SR algorithms cross-sectionally on IR images. The experimental results show that the SRFBN[52] model has the best universalization ability. But, for the GAN model, there are always unpleasant artifacts due to the model collapse. All these works illustrate the domain transfer difficulties possibly suffered by algorithms using visible images that are used in the IR images super-resolution task. Then, people started to consider the possibility of designing GAN models specifically for IR images. These studies are divided into the following categories: **module improvement, introduce extra information.**

1) *Module improvement:* It is natural first to consider that the improvement method is to explore new modules to improve performance. *Rivadeneira et al./89/* uses the CycleGAN structure and employs ResNet as the generator's encoder. The self-attentive module is applied in the encoder and a new loss function is also proposed (see Eq. 6). According to the discussion in this work, the Sobel edge detector can capture the contour consistency between the input image and the cyclic-generated image by calculating the mean squared difference between the images.

$$\mathcal{L}_{\text{Sobel}} = \frac{1}{N} \sum_i \| \text{Sobel}(G_{H2L}(G_{L2H}(I_L))) - \text{Sobel}(I_L) \| \quad (6)$$

Another GAN model based on the attention mechanism has also been proposed, which differs in the loss function and discriminator used[59]. As shown in Fig. 10. Liu, Qing-Ming, et al. use a module based on the channel attention mechanism in the generator at their work. Moreover, they also replaced the original discriminator with a new one designed by the gradient penalty approach from WGAN. It will benefit network convergence. For the loss function, Wasserstein distance is used to evaluate the target image I_{HR} and the reconstructed image \hat{I} (see Eq.7).

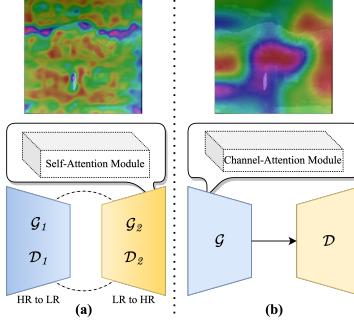


Figure 10: Attentional mechanisms and GAN-based models for IR image super-resolution. Cam Heatmap is also presented.

$$\begin{aligned} \mathcal{L}_{\text{WGAN}} = & E [D_\theta(I_{HR}) - D_\theta(G_\theta(I_{LR}))] + \\ & \lambda E \left[\left\| \nabla_I D(\hat{I})_2 - 1 \right\|^2 \right] \end{aligned} \quad (7)$$

where the second term on the right side of the equal sign is the gradient penalty term. Furthermore, WGAN is used to optimize the training process is also mentioned in this work[35].

2) Introduce extra information: This approach is focused on the issue that IR images have fewer patterns and further proposes to use visible images to guide the reconstruction. The main options include the following: **hybrid and split**. For the hybrid model, the features from visible and IR images are not purposely distinguished but mixed directly in the module. In [4], the information is initially extracted by residual blocks and then fused using 1x1 convolution. It is called multimodal, visual-thermal fusion model[123, 21]. Experiments show that visible images are beneficial for improving high-frequency details in IR images.

For the split model, *Huang et al.*[36] proposed using transfer learning to help introduce the patterns in visible images to reconstruct IR images. In this approach, PSRGAN, two components are proposed: the main path, which is responsible for extracting the features from visible images, and the branch path, which is used to represent the patterns from IR images. Finally, the IR image is reconstructed by a transfer learning strategy that maps both of them to a common feature space. Such an approach can better capture and utilize the features in the model and show better results in the experiments.

Finally, there are also approaches that focus on the connection between SR methods and other tasks, such as denoising, and then models that combine cross-tasks are proposed. *Batchuluun, Ganbayar, et al.*[8] proposed methods that can be used for super-resolution and detection, where the super-resolution component focuses on both denoising and SR. In this work, the number of modules in the generator has been increased. In [75], researchers proposed a GAN-based

Table 4: Available dataset in IR image super-resolution fields. From left to right, the columns depict the name and the reference, pixel resolution of images, and a brief description, respectively

Name	Classes	pxl	Description
NATO SET-140[118]	Universal	320x256 640x480 384x288x2	The imagery was recorded for different target / background settings, camera and/or object movements and temperature contrasts. The paired samples are from the real world.
CASIA Interval v3[7]	Iris images	280x320	This database has 2,655 NIR images of 280×320 pixels from 249 contributors captured in 2 sessions with a close-up iris camera, totaling 396 different eyes. HR-LR image pairs could be available from bicubic downampling processing.
ULB17-VT[4]	Universal	240 × 320	Thermal and RGB images were manually extracted and annotated with a total of 570 pairs. The framework is divided into 512 training and validation samples and 58 test samples.
IR-COLOR2000[24]	Universal	128x128	The training datasets (IR-COLOR2000), which contain 2000 pairs of images. Paired RGB images are available.
IR100[36]	Universal	640x480	100 samples from FIR Sequence Pedestrian Dataset, the paired LR images come from bicubic downampling.
CVC-09-1K[35]	Universal	640x480	1000 randomly selected images from the FIR Dataset, the paired LR images come from bicubic downsampling.
New dataset[87]	Universal	640x512	This dataset was acquired using a TAU2 thermal camera with a 13mm lens (45° HFOV) in a resolution of 640x512, with a depth of 8 bits and save it in PNG format
CDN-MRF dataset[34]	Universal	640x480	They build up a HR infrared image dataset covering a wide range of scenarios (e.g., vehicle, machine, pedestrian and building). The paired samples are from the real world.

framework for joint cross-modal and super-resolution aerial image vehicle detection. The first sub-network of this approach uses the GAN architecture to generate two different SR images by cross-domain transforms.

In summary, there is no doubt that the GAN model has great advantages in the field of generation, leading to a fairly good performance in SR algorithms. The results achieved in the visible field have attracted the application to the IR image super-resolution. However, according to the experimental results, the direct unmodified introduction of the normal image super-resolution reconstruction method encounters the difficulty of domain shift[94, 29, 135]. The edges in the reconstructed images are not clear. Therefore, new models have been proposed by considering the characteristics of IR images. These models are divided into improved modules[89, 59, 84], introduce extra information[4, 88, 122], and multitask[61, 9, 8, 75]. These algorithms have achieved some success, but missing standard benchmarks and datasets that are uniform leads to an unfavorable cross-sectional comparison between different types of methods. In the next section, we present available datasets and evaluation metrics in the IR image super-resolution field.

3.2.3. Transformer-based

Transformer-based methods have transformed IRSR by addressing challenges unique to IR imaging, such as low contrast and limited gradients. Leveraging self-attention mechanisms[40], these models excel in capturing long-range dependencies, making them ideal for reconstructing high-resolution IR images with superior structural consistency.

The architectural innovations within Transformer-based IR SR models demonstrate a strong focus on achieving efficient global feature representation. Swin Transformer-based models, such as SwinIBSR[96] and SwinIPISR[121], extend the Swin architecture with residual connections and tailored degradation models, improving generalization and robustness in real-world scenarios. Similarly,

LKFormer replaces conventional self-attention with large kernel convolutional attention blocks, striking a balance between computational efficiency and non-local feature modeling[85]. These advancements reflect a shared objective to optimize performance while maintaining scalability.

A defining characteristic of these methods is their adaptability to specific tasks and modalities. Techniques like Cross-modal Texture Transformers (CMTT) effectively transfer information from visible-spectrum images to improve IR image quality, while edge-focused models such as TESR incorporate auxiliary networks to enhance texture and detail recovery[39]. Meanwhile, DASR utilizes dual-attention mechanisms to simultaneously capture local and global information, underscoring the versatility of Transformer architectures in addressing diverse challenges[54].

Compared to GANs, which excel in texture synthesis but are prone to artifacts under domain shifts, Transformer-based methods demonstrate superior robustness through explicit global contextual modeling. Similarly, while CNNs are computationally efficient and effective in learning localized features, Transformers outperform in scenarios requiring long-range dependency modeling and structural consistency. These distinctions make Transformers particularly suitable for dynamic environments where IR imaging data is sparse or noisy. For instance, TnTViT-G[73] demonstrates how visible-spectrum guidance can compensate for missing IR image details, exemplifying the potential of cross-spectral applications.

In summary, Transformer-based methods represent a significant advancement in IRSR by combining efficient feature extraction with innovative attention mechanisms. However, their computational complexity remains a challenge, particularly for high-resolution IR image reconstruction, where scaling attention mechanisms can be resource-intensive. Addressing these limitations could further enhance their applicability in real-time systems. These approaches share a unified goal of overcoming the limitations of IR imaging systems while optimizing for scalability and practical deployment. Future research will likely explore deeper integration with emerging imaging modalities and further enhancements for real-time applications.

4. Datasets & Metrics

In this section, we introduce the datasets and image quality assessment metrics used in IR image super-resolution.

4.1. Datasets

For IRSR, two types of datasets are commonly employed: training datasets and test datasets. The training dataset should contain a large number of samples covering diverse natural and realistic conditions, which is crucial for enhancing model generalization. In contrast, test datasets are typically task-specific—for example, a face dataset for facial reconstruction or an iris dataset for iris reconstruction.

Although several datasets such as IR100 and NATO SET-140 have been proposed, most available datasets lack paired samples that encompass varying environmental conditions and multi-modal information. Future efforts should focus on developing standardized datasets that reflect real-world challenges (e.g., adverse weather, varying illumination) and that include relevant imaging parameters. Representative datasets include the publicly available collection by Weiss et al. [118], a benchmark of 1,872 NIR iris images by Alonso-Fernandez et al. [6], and the ULB17-VT dataset for thermal and visual image pairs [5]. Additional datasets, such as those proposed by Du et al. [24] and Huang et al. [36, 35], are summarized in Table 4. Note that many datasets do not provide complete metadata (e.g., equipment parameters) and differ in sample sizes, which hinders direct comparisons.

4.2. Metrics

Evaluating the quality of reconstructed IR images is critical. In practice, three types of metrics are used: reference metrics, non-reference metrics, and human evaluations.

4.2.1. Reference Metrics

The most common reference metrics are the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM). PSNR is calculated as:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{L^2}{\frac{1}{N} \sum_{i=1}^N (I(i) - \hat{I}(i))^2} \right) \quad (8)$$

where L (typically 255 for 8-bit images) is the maximum pixel value, I is the high-resolution ground truth image, and \hat{I} is the reconstructed image. However, because PSNR is based solely on pixel-wise MSE, it often does not correlate well with human perception.

To better capture perceptual quality, SSIM is used:

$$\text{SSIM}(I, \hat{I}) = [\mathcal{C}_l(I, \hat{I})]^\alpha [\mathcal{C}_c(I, \hat{I})]^\beta [\mathcal{C}_s(I, \hat{I})]^\gamma, \quad (9)$$

where \mathcal{C}_l , \mathcal{C}_c , and \mathcal{C}_s denote comparisons of luminance, contrast, and structure respectively, and α , β , and γ are weighting factors. Both PSNR and SSIM require ground-truth images, which may not be available in real-world scenarios.

4.2.2. No Reference Metrics

No reference metrics, such as the Natural Image Quality Evaluator (NIQE) [74] and the Learned Perceptual Image Patch Similarity (LPIPS) [137], are used to assess image quality without relying on ground-truth images. These metrics leverage statistical properties or deep feature representations to evaluate perceptual quality.

4.2.3. Human Evaluation

Human subjective evaluation remains the most reliable measure of visual quality. However, it requires significant time and resources to collect and analyze scores from multiple observers, making it less practical for routine comparisons.

In summary, while reference metrics like PSNR and SSIM are widely used for benchmarking, their limitations in reflecting human perception have motivated the development of no reference metrics and the use of human evaluation in specific scenarios.

5. Future Trends

In this section, we outline promising research directions that can further advance infrared image super-resolution (IRSR).

5.1. New Models & New Paradigms

Advanced model design can enhance reconstructed image quality by learning more effective nonlinear mappings in high-dimensional spaces. Inspired by breakthroughs in natural language processing with Large Language Models (LLMs), the emergence of Large Vision Models (LVMs) is expected to have a profound impact on IRSR. In particular, future research may focus on understanding and exploiting the correlation between low- and high-frequency information in IR images to build more robust representations.

Generative Models. With rapid advances in deep learning, state-space and diffusion models [91, 93, 65, 13] are attracting increasing attention. These models, capable of efficiently handling long-range dependencies, offer a promising alternative to traditional Transformer architectures for low-level vision tasks. Their integration into IRSR frameworks is anticipated to significantly improve detail synthesis and overall reconstruction quality.

Data Augmentation. Innovative data augmentation approaches—such as generating synthetic thermal-like images from visible inputs using CycleGAN [100] can effectively bridge the gap between limited IR datasets and real-world variability. Future work should explore domain-specific synthetic data generation and cross-modality transformations to enhance model robustness and alleviate data scarcity.

Large Vision Models. Just as LLMs have revolutionized text processing, LVMs can potentially transform computer vision tasks like IRSR. Their ability to extract and upscale intricate details from low-resolution IR images, combined with transfer learning from vast visual datasets, could mitigate the need for extensive infrared-specific training data and lead to significant performance improvements.

5.2. New Applications

Although many super-resolution algorithms have been proposed, their performance in real-world scenarios often falls short because they typically assume simple degradation models such as bicubic downsampling. In reality, IR images

are affected by complex degradations—such as noise, blur, and compression artifacts—that are not fully captured by these assumptions [49, 37, 136, 114, 97]. Blind super-resolution, which aims to recover images degraded by unknown factors, is therefore an attractive direction. Future research should focus on accurately modeling real-world IR degradations and developing reconstruction algorithms that are robust to such complexities.

5.3. Datasets and Evaluation Metrics

Robust datasets are critical for training and benchmarking IRSR algorithms. However, many current datasets are limited and non-standardized, lacking essential capture details (e.g., optical parameters, resolution, environmental conditions). Establishing standardized benchmarks—with paired visible images under similar conditions, as seen in initiatives like the TISR challenge under PBVS—will enable fair comparisons across methods [90, 109].

Similarly, traditional evaluation metrics (e.g., PSNR, SSIM) and non-reference (e.g., NIQE, LPIPS)—may not fully capture the unique perceptual qualities of IR images. Developing specialized metrics that reflect IR-specific characteristics and human perception is a key direction for future research.

6. Conclusion

In this paper, we provide a comprehensive survey of IR image super-resolution research from the past two decades. We discuss its fundamental role in engineering applications and analyze key factors limiting IR imaging quality, such as noise and hardware constraints, highlighting the high costs associated with hardware redesign. We systematically classify and summarize both traditional algorithms and deep learning-based methods, alongside essential datasets and image quality assessment metrics. Our review bridges the gap between traditional super-resolution approaches and the unique challenges of IR imaging, offering actionable insights for developing domain-specific algorithms. Future research could explore integrating Large Vision Models and advanced data augmentation techniques to address current limitations and expand the applicability of IR image super-resolution in diverse real-world scenarios.

Declaration of competing interest.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability.

Data will be made available on request.

Acknowledgments.

This work was supported by JSPS KAKENHI Grant Number JP23KJ0118.

References

- [1] Agustsson, E., Timofte, R., 2017. Ntire 2017 challenge on single image super-resolution: Dataset and study, in: CVPR workshops, pp. 126–135.
- [2] Ahmadi, S., et al., 2020. Super resolution laser line scanning thermography. Optics and Lasers in Engineering 134, 106279.
- [3] Allred, B.a., 2021. Time of day impact on mapping agricultural subsurface drainage systems with uav thermal infrared imagery. Agricultural Water Management 256, 107071.
- [4] Almasri, F., Debeir, O., 2018a. Multimodal sensor fusion in single thermal image super-resolution, in: ACCV, Springer. pp. 418–433.
- [5] Almasri, F., Debeir, O., 2018b. Multimodal sensor fusion in single thermal image super-resolution, in: ACCV Workshops.
- [6] Alonso-Fernandez, F., et al., 2015. Eigen-patch iris super-resolution for iris recognition improvement, in: 2015 23rd European Signal Processing Conference (EUSIPCO), IEEE. pp. 76–80.
- [7] Alonso-Fernandez, F., et al., 2017. Iris super-resolution using iterative neighbor embedding, in: CVPR Workshops, pp. 153–161.
- [8] Batchuluun, G., et al., 2020a. Deep learning-based thermal image reconstruction and object detection. IEEE Access 9, 5951–5971.
- [9] Batchuluun, G., et al., 2020b. Thermal image reconstruction using deep learning. IEEE Access 8, 126839–126858.
- [10] Beenard, M., et al., 2021. Optimization of advanced live-cell imaging through red/near-infrared dye labeling and fluorescence lifetime-based strategies. International Journal of Molecular Sciences 22.
- [11] Bodensteiner, C., et al., 2018. Multispectral matching using conditional generative appearance modeling. 2018 15th IEEE AVSS , 1–6.
- [12] Canales-Fiscal, M.R., et al., 2021. Covid-19 classification using thermal images: thermal images capability for identifying covid-19 using traditional machine learning classifiers. ACM CBCAHI .
- [13] Cao, H., et al., 2022. A survey on generative diffusion model. arXiv preprint arXiv:2209.02646 .

- [14] Cao, Y., et al., 2020. Monitoring of sugar beet growth indicators using wide-dynamic-range vegetation index (wdrvi) derived from uav multispectral images. *Computers and Electronics in Agriculture* 171, 105331.
- [15] Chang, H., et al., 2004. Super-resolution through neighbor embedding, in: CVPR, IEEE. pp. I–I.
- [16] Chen, H., et al., 2022. Real-world single image super-resolution: A brief review. *Information Fusion* 79, 124–145.
- [17] Chen, Q., Shao, Q., 2024. Single image super-resolution based on trainable feature matching attention network. *Pattern Recognition* 149, 110289.
- [18] Chen, S., et al., 2021. Geometry and adjacency effects in urban land surface temperature retrieval from high-spatial-resolution thermal infrared images. *Remote Sensing of Environment* 262, 112518.
- [19] Cheng-Zhi, D., et al., 2014. Infrared image super-resolution via locality-constrained group sparse model. *Acta Physica Sinica* 63.
- [20] Danaci, K.I., Akagunduz, E., 2022. A survey on infrared image and video sets. arXiv preprint arXiv:2203.08581 .
- [21] Dong, A., et al., 2024. Mfifusion: An infrared and visible image enhanced fusion network based on multi-level feature injection. *Pattern Recognition* 152, 110445.
- [22] Dong, C., et al., 2015. Image super-resolution using deep convolutional networks. *IEEE T-PAMI* 38, 295–307.
- [23] Dong, C.a., 2016. Accelerating the super-resolution convolutional neural network, in: ECCV, Springer. pp. 391–407.
- [24] Du, J., et al., 2020. Rgb-ir cross input and sub-pixel upsampling network for infrared image super-resolution. *Sensors* 20, 281.
- [25] Fan, K., et al., 2021. Infrared image super-resolution via progressive compact distillation network. *Electronics* 10, 3107.
- [26] Fang, A., Li, Y., 2024. Dynamic and static fusion mechanisms of infrared and visible images. *Pattern Recognition* , 110689.
- [27] Gao, Z., Chen, J., 2022. Maritime infrared image super-resolution using cascaded residual network and novel evaluation metric. *IEEE Access* 10, 17760–17767.
- [28] Goodfellow, I., et al., 2020. Generative adversarial networks. *Communications of the ACM* 63, 139–144.
- [29] Guei, A.C., Akhloufi, M., 2018. Deep learning enhancement of infrared face images using generative adversarial networks. *Applied optics* 57, D98–D107.

- [30] Gulrajani, I., et al., 2017. Improved training of wasserstein gans. NeurIPS 30.
- [31] Güngör, A., Kar, O.F., 2019. A transform learning based deconvolution technique with super-resolution and microscanning applications, in: ICIP, IEEE. pp. 2159–2163.
- [32] Gupta, H., Mitra, K., 2022. Toward unaligned guided thermal super-resolution. IEEE TIP 31, 433–445.
- [33] Han, T.Y., et al., 2018. Infrared image super-resolution using auxiliary convolutional neural network and visible image under low-light conditions. Journal of Visual Communication and Image Representation 51, 191–200.
- [34] He, Z., et al., 2018. Cascaded deep networks with multiple receptive fields for infrared image super-resolution. IEEE TCSVT 29, 2310–2322.
- [35] Huang, Y., et al., 2021a. Infrared image super-resolution via heterogeneous convolutional wgan, in: PRICAI, Springer. pp. 461–472.
- [36] Huang, Y., et al., 2021b. Infrared image super-resolution via transfer learning and psrgan. IEEE Signal Processing Letters 28, 982–986.
- [37] Huang, Y., et al., 2022. Rethinking degradation: Radiograph super-resolution via aid-srgan. arXiv preprint arXiv:2208.03008 .
- [38] Huang, Y., et al., 2024. Irsrmamba: Infrared image super-resolution via mamba-based wavelet transform feature modulation model. arXiv preprint arXiv:2405.09873 .
- [39] Jiang, Y., et al., 2024. Cross-modal texture transformer for thermal infrared reference-based super-resolution reconstruction. Optics & Laser Technology 176, 110914.
- [40] Khan, S., et al., 2022. Transformers in vision: A survey. ACM computing surveys (CSUR) 54, 1–41.
- [41] Kim, J., Lee, J.K., Lee, K.M., 2016. Accurate image super-resolution using very deep convolutional networks, in: CVPR, pp. 1646–1654.
- [42] Kim, M., Kim, Y., 2009. Forward looking infrared images' resolution improvement applying an image restoration method of a super resolution, in: 2009 34th International Conference on Infrared, Millimeter, and Terahertz Waves, IEEE. pp. 1–2.
- [43] Kim, S.H., et al., 2018. Averaging current adjustment technique for reducing pixel resistance variation in a bolometer-type uncooled infrared image sensor. Journal of Sensor Science and Technology 27, 357–361.
- [44] Kuang, X., et al., 2017. Single infrared image optical noise removal using a deep convolutional neural network. IEEE photonics Journal 10, 1–15.

- [45] Kwasniewska, A., et al., 2020. Super-resolved thermal imagery for high-accuracy facial areas detection and analysis. *Engineering Applications of Artificial Intelligence* 87, 103263.
- [46] Laikin, M., 2018. *Lens design*. CRC Press.
- [47] Ledig, C., et al., 2017. Photo-realistic single image super-resolution using a generative adversarial network, in: CVPR, pp. 4681–4690.
- [48] Lee, I.H., et al., 2022a. Super-resolution thermal generative adversarial networks for infrared image enhancement. *Journal of Institute of Control, Robotics and Systems* .
- [49] Lee, S., et al., 2022b. Learning multiple probabilistic degradation generators for unsupervised real world image super resolution. arXiv preprint arXiv:2201.10747 .
- [50] Li, L., et al., 2025. Sanet: Face super-resolution based on self-similarity prior and attention integration. *Pattern Recognition* 157, 110854.
- [51] Li, X., et al., 2022. Improved yolov4 network using infrared images for personnel detection in coal mines. JEI 31, 013017 – 013017.
- [52] Li, Z., et al., 2019. Feedback network for image super-resolution, in: CVPR, pp. 3867–3876.
- [53] Liang, J., et al., 2021. Swinir: Image restoration using swin transformer, in: CVPR, pp. 1833–1844.
- [54] Liang, S., et al., 2023. Dasr: Dual-attention transformer for infrared image super-resolution. *Infrared Physics & Technology* 133, 104837.
- [55] Lim, B., et al., 2017. Enhanced deep residual networks for single image super-resolution, in: CVPR workshops, pp. 136–144.
- [56] Lin, C.H., Wang, T.Y., 2021. A novel convolutional neural network architecture of multispectral remote sensing images for automatic material classification. *Signal Processing: Image Communication* 97, 116329.
- [57] Liu, H., et al., 2022. An infrared image denoising model with unidirectional gradient and sparsity constraint on biomedical images. *Infrared Physics & Technology* 126, 104348.
- [58] Liu, H.c., et al., 2013. Infrared surveillance image super resolution via group sparse representation. *Optics Communications* 289, 45–52.
- [59] Liu, Q.M., et al., 2021a. Infrared image super-resolution reconstruction by using generative adversarial network with an attention mechanism. *Applied Intelligence* 51, 2018–2030.

- [60] Liu, S., et al., 2021b. Simulating the leaf area index of rice from multi-spectral images. *Remote Sensing* 13, 3663.
- [61] Liu, X., et al., 2019. Infrared image super-resolution reconstruction based on quaternion and high-order overlapping group sparse total variation. *Sensors* 19, 5139.
- [62] Lloyd, D.T., et al., 2021. Optically enhanced super-resolution of sea surface temperature using deep learning. *IEEE TGRS* 60, 1–14.
- [63] Lukose, J., et al., 2021. Optical technologies for the detection of viruses like covid-19: Progress and prospects. *Biosensors and Bioelectronics* 178, 113004.
- [64] Luo, Z., et al., 2022. Learning the degradation distribution for blind image super-resolution, in: CVPR, pp. 6063–6072.
- [65] Lyu, Z., et al., 2022. Accelerating diffusion models via early stop of the diffusion process. arXiv preprint arXiv:2205.12524 .
- [66] Ma, C., et al., 2020. Structure-preserving super resolution with gradient guidance, in: CVPR, pp. 7769–7778.
- [67] Mao, Y., et al., 2016. An infrared image super-resolution reconstruction method based on compressive sensing. *Infrared Physics & Technology* 76, 735–739.
- [68] Marivani, I., et al., 2019. Multimodal image super-resolution via deep unfolding with side information, in: 2019 27th European Signal Processing Conference (EUSIPCO), IEEE. pp. 1–5.
- [69] Marivani, I., et al., 2020a. Joint image super-resolution via recurrent convolutional neural networks with coupled sparse priors, in: ICIP, IEEE. pp. 868–872.
- [70] Marivani, I., et al., 2020b. Multimodal deep unfolding for guided image super-resolution. *IEEE TIP* 29, 8443–8456.
- [71] Martínez Gila, D.M., et al., 2022. The advantage of multispectral images in fruit quality control for extra virgin olive oil production. *Food Analytical Methods* 15, 75–84.
- [72] Mazaheri, L., et al., 2022. Investigating the performances of wide-field raman microscopy with stochastic optical reconstruction post-processing. *Applied spectroscopy* 76, 340–351.
- [73] Mehri, A., et al., 2023. Tntvit-g: Transformer in transformer network for guidance super resolution. *IEEE Access* 11, 11529–11540.
- [74] Mittal, A., Soundararajan, R., Bovik, A.C., 2012. Making a “completely blind” image quality analyzer. *IEEE SPL* 20, 209–212.

- [75] Mostafa, M., et al., 2020. A joint cross-modal super-resolution approach for vehicle detection in aerial imagery, in: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II, SPIE. pp. 184–194.
- [76] Nie, J., et al., 2020. Land surface temperature and emissivity retrieval from nighttime middle-infrared and thermal-infrared sentinel-3 images. *IEEE Geoscience and Remote Sensing Letters* 18, 915–919.
- [77] Offroy, M., Roggo, Y., Duponchel, L., 2012. Increasing the spatial resolution of near infrared chemical images (nir-ci): The super-resolution paradigm applied to pharmaceutical products. *Chemometrics and Intelligent Laboratory Systems* 117, 183–188.
- [78] Offroy, M., et al., 2010. Infrared chemical imaging: spatial resolution evaluation and super-resolution concept. *Analytica chimica acta* 674 2, 220–6.
- [79] Oz, N., et al., 2020. Rapid super resolution for infrared imagery. *Optics Express* 28, 27196–27209.
- [80] Park, S.C., et al., 2003. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine* 20, 21–36.
- [81] Patel, H.M., et al., 2021. Thermisrnet: an efficient thermal image super-resolution network. *Optical Engineering* 60, 073101.
- [82] Ping, B., et al., 2021. Can the structure similarity of training patches affect the sea surface temperature deep learning super-resolution? *Remote Sensing* 13, 3568.
- [83] Poyser, M., Breckon, T.P., 2024. Neural architecture search: A contemporary literature review for computer vision applications. *Pattern Recognition* 147, 110052.
- [84] Prajapati, K., et al., 2021. Channel split convolutional neural network (chasnet) for thermal image super-resolution, in: CVPR, pp. 4368–4377.
- [85] Qin, F., et al., 2024. Lkformer: large kernel transformer for infrared image super-resolution. *Multimedia Tools and Applications* , 1–15.
- [86] Redlich, R., et al., 2016. An embedded hardware architecture for real-time super-resolution in infrared cameras, in: 2016 Euromicro conference on digital system design (DSD), IEEE. pp. 184–191.
- [87] Rivadeneira, R.E., Suárez, P.L., et al., 2019. Thermal image superresolution through deep convolutional neural network, in: International conference on image analysis and recognition, Springer. pp. 417–426.
- [88] Rivadeneira, R.E., et al., 2020. Thermal image super-resolution: A novel architecture and dataset., in: VISIGRAPP (4: VISAPP), pp. 111–119.

- [89] Rivadeneira, R.E., et al., 2022a. A novel domain transfer-based approach for unsupervised thermal image super-resolution. *Sensors* 22, 2254.
- [90] Rivadeneira, R.E., et al., 2022b. Thermal image super-resolution challenge results-pbvs 2022, in: CVPR, pp. 418–426.
- [91] Rombach, R., et al., 2022. High-resolution image synthesis with latent diffusion models, in: CVPR, pp. 10684–10695.
- [92] Ruan, Z., et al., 2024. Semantic attention-based heterogeneous feature aggregation network for image fusion. *Pattern Recognition* 155, 110728.
- [93] Salimans, T., Ho, J., 2022. Progressive distillation for fast sampling of diffusion models. arXiv preprint arXiv:2202.00512 .
- [94] SHAO, B., et al., 2018. Single frame infrared image super-resolution algorithm based on generative adversarial nets. *Journal of Infrared and Millimeter Wave* 37, 427–432.
- [95] Shen, J., et al., 2019. A depth estimation framework based on unsupervised learning and cross-modal translation, in: Target and Background Signatures V, SPIE. pp. 21–31.
- [96] Shi, Y., Chen, N., Pu, Y., Zhang, J., Yao, L., 2024. Swinibsr: Towards real-world infrared image super-resolution. *Infrared Physics & Technology* 139, 105279.
- [97] Son, S., et al., 2021. Toward real-world super-resolution via adaptive downsampling models. *IEEE T-PAMI* .
- [98] Song, P., et al., 2019. Multimodal image super-resolution via joint sparse representations induced by coupled dictionaries. *IEEE Transactions on Computational Imaging* 6, 57–72.
- [99] Su, H., et al., 2024. A review of deep-learning-based super-resolution: From methods to applications. *Pattern Recognition* , 110935.
- [100] Suárez, P.L., Carpio, D., Sappa, A.D., 2024. Enhancement of guided thermal image super-resolution approaches. *Neurocomputing* 573, 127197.
- [101] Sun, C., et al., 2018. A rapid and accurate infrared image super-resolution method based on zoom mechanism. *Infrared Physics & Technology* 88, 228–238.
- [102] Sun, G., et al., 2007. Map algorithm to super-resolution of infrared images, in: MIPPR 2007, SPIE. pp. 148–156.
- [103] Sun, M., Yu, K., 2013. A sur-pixel scan method for super-resolution reconstruction. *Optik* 124, 6905–6909.

- [104] Suryanarayana, G., Tu, E., Yang, J., 2019. Infrared super-resolution imaging using multi-scale saliency and deep wavelet residuals. *Infrared Physics & Technology* 97, 177–186.
- [105] Szankin, M., et al., 2019. Influence of thermal imagery resolution on accuracy of deep learning based face recognition, in: 2019 12th HSI, IEEE. pp. 1–6.
- [106] Torra, J., et al., 2022. Versatile near-infrared super-resolution imaging of amyloid fibrils with the fluorogenic probe cranad-2. *Chemistry* .
- [107] Wang, B., et al., 2021a. Super resolution reconstruction of low light level image based on the feature extraction convolution neural network, in: Computational Imaging VI, SPIE. pp. 74–79.
- [108] Wang, B., et al., 2022a. Multimodal super-resolution reconstruction of infrared and visible images via deep learning. *Optics and Lasers in Engineering* 156, 107078.
- [109] Wang, K., et al., 2022b. Cippsrnet: A camera internal parameters perception network based contrastive learning for thermal image super-resolution, in: CVPR, pp. 342–349.
- [110] Wang, L., Yoon, K.J., 2021. Deep learning for hdr imaging: State-of-the-art and future trends. *IEEE T-PAMI* .
- [111] Wang, S., Ding, M., et al., 2018a. Live cell visualization of multiple protein-protein interactions with bifc rainbow. *ACS chemical biology* 13 5, 1180–1188.
- [112] Wang, S.P., 2016. Stripe noise removal for infrared image by minimizing difference between columns. *Infrared Physics & Technology* 77, 58–64.
- [113] Wang, X., et al., 2018b. Esrgan: Enhanced super-resolution generative adversarial networks, in: ECCV workshops, pp. 0–0.
- [114] Wang, X., et al., 2021b. Real-esrgan: Training real-world blind super-resolution with pure synthetic data, in: CVPR, pp. 1905–1914.
- [115] Wang, Y., et al., 2022c. Compressed sensing super-resolution method for improving the accuracy of infrared diagnosis of power equipment. *Applied Sciences* 12, 4046.
- [116] Wang, Y., et al., 2025. Lightweight remote sensing super-resolution with multi-scale graph attention network. *Pattern Recognition* 160, 111178.
- [117] Wang, Z.a.o., 2020. Deep learning for image super-resolution: A survey. *IEEE T-PAMI* 43, 3365–3387.

- [118] Weiß, A.R., et al., 2012. A standard data set for performance analysis of advanced ir image processing techniques, in: Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIII, SPIE. pp. 354–363.
- [119] Wu, H., et al., 2023. Deep learning-based image super-resolution restoration for mobile infrared imaging system. *Infrared Physics & Technology* 132, 104762.
- [120] Wu, W., et al., 2022a. Meta transfer learning-based super-resolution infrared imaging. *Digital Signal Processing* 131, 103730.
- [121] Wu, X., et al., . Swinipisr: A super-resolution method for infrared polarization imaging sensors via swin transformer. *IEEE Sensors J.* .
- [122] Wu, Y., et al., 2022b. Infrared and visible light dual-camera super-resolution imaging with texture transfer network. *Signal Processing: Image Communication* 108, 116825.
- [123] Wu, Z., et al., 2025. Freprompter: Frequency self-prompt for all-in-one image restoration. *Pattern Recognition* 161, 111223.
- [124] Xiong, Q., et al., 2021. Deriving non-cloud contaminated sentinel-2 images with rgb and near-infrared bands from sentinel-1 images based on a conditional generative adversarial network. *Remote Sensing* 13, 1512.
- [125] Yamaguchi, M., et al., 2020. Super-resolution imaging of the protoplanetary disk hd 142527 using sparse modeling. *The Astrophysical Journal* 895, 84.
- [126] Yang, J., et al., 2008. Image super-resolution as sparse representation of raw image patches, in: 2008 CVPR, IEEE. pp. 1–8.
- [127] Yang, J.o., 2010. Image super-resolution via sparse representation. *IEEE TIP* 19, 2861–2873.
- [128] Yang, X., et al., 2015. Infrared image recovery from visible image by using multi-scale and multi-view sparse representation, in: 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), IEEE. pp. 554–559.
- [129] Yang, X., et al., 2017. Multi-sensor image super-resolution with fuzzy cluster by using multi-scale and multi-view sparse coding for infrared image. *Multimedia Tools and Applications* 76, 24871–24902.
- [130] Yang, Y., et al., 2020. Deep networks with detail enhancement for infrared image super-resolution. *IEEE Access* 8, 158690–158701.
- [131] Yao, T., et al., 2020. Infrared image super-resolution via discriminative dictionary and deep residual network. *Infrared Physics & Technology* 107, 103314.

- [132] Young, S.S., Driggers, R.G., 2006. Superresolution image reconstruction from a sequence of aliased imagery. *Applied Optics* 45, 5073–5085.
- [133] Yu, H., et al., 2013. Single infrared image super-resolution combining non-local means with kernel regression. *Infrared Physics & Technology* 61, 50–59.
- [134] Yu, W., et al., 1998. An infrared image synthesis model based on infrared physics and heat transfer. *International journal of infrared and millimeter waves* 19, 1661–1669.
- [135] Zhang, C., et al., 2020. Single infrared remote sensing image super-resolution via supervised deep learning, in: *Image and Signal Processing for Remote Sensing XXVI*, SPIE. pp. 241–249.
- [136] Zhang, K., et al., 2021. Designing a practical degradation model for deep blind image super-resolution, in: *CVPR*, pp. 4791–4800.
- [137] Zhang, R., et al., 2018a. The unreasonable effectiveness of deep features as a perceptual metric, in: *CVPR*, pp. 586–595.
- [138] Zhang, X., et al., 2018b. Infrared image super resolution by combining compressive sensing and deep learning. *Sensors* 18, 2587.
- [139] Zhang, X., et al., 2019. Super-resolution imaging for infrared micro-scanning optical system. *Optics express* 27, 7719–7737.
- [140] Zhao, L., et al., 2024. Ssir: Spatial shuffle multi-head self-attention for single image super-resolution. *Pattern Recognition* 148, 110195.
- [141] Zhao, Y., et al., 2016. Learning-based compressed sensing for infrared image super resolution. *Infrared Physics & Technology* 76, 139–147.
- [142] Zhong, J., et al., 2016. Image fusion and super-resolution with convolutional neural network, in: *Chinese Conference on Pattern Recognition*, Springer. pp. 78–88.
- [143] Zou, Y., et al., 2020. An infrared image super-resolution imaging algorithm based on auxiliary convolution neural network, in: *Optics Frontier Online 2020: Optics Imaging and Display*, SPIE. pp. 335–340.
- [144] Zou, Y., et al., 2021. Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections. *Optics and Lasers in Engineering* 146, 106717.