Thermal-Physics Guided Infrared Image Super-Resolution with Dynamic High-Frequency Amplification

Mingxuan Zhou[1*], Yirui Shen[1*], Shuang Li[2†], Jing Geng[1†], Yutang Zhang[2], Shuigen Wang[3]

[1]School of Computer Science & Technology, Beijing Institute of Technology, Beijing, China
[2]School of Artificial Intelligence, Beihang University, Beijing, China
[3]Iray Technology Co., Ltd., Shandong, China
{mingxuanzhou, yrshen, janegeng}@bit.edu.cn, {shuangliai, yutangzhang}@buaa.edu.cn, shuigen.wang@iraytek.com

Abstract

The practical deployment of infrared imaging is hindered by its inherent output of low-resolution (LR) images. While the super-resolution (SR) technique is a promising remedy, we discover two major challenges concerning infrared image SR: preserving accurate thermal distributions, which are fundamental to infrared imaging, and addressing the ambiguity of high-frequency elements compared to visible images. To tackle these issues, we propose **ThesIS**, a tailored framework that utilizes **The**rmal-Physic**s** guidance and dynamic high-frequency amplification for **I**nfrared image **S**uper-resolution to produce high-resolution (HR) images with accurate physical properties and delicate visual details. Specifically, Thermal Regularization is introduced to reconstruct the accurate thermal radiation distribution via the introduced Infrared Radiation Intensity Alignment Loss, mitigating the adverse effects of complex degradations while conducting initial upscaling. Additionally, we design a guidance mechanism to counter the randomness of the diffusion model, further refining the preservation of physical information. The proposed Dynamic High-Frequency Amplification effectively strengthens the ambiguous high-frequency information present in infrared images, leading to improved texture details and superior visual quality. Extensive experiments demonstrate that ThesIS successfully recovers accurate thermal information while delivering visually satisfying results with state-of-the-art performance. Furthermore, we introduce the **InfraredSR** dataset, which comprises 39,833 images at a resolution of 512 $\times$ 512, hoping to advance research in this field.

Introduction

Infrared imaging has long been a versatile and widely adopted technology, rooted in capturing thermal radiation emitted in diverse environments (Bao et al. 2023; Huang et al. 2022). By analyzing this radiation, we extract valuable information about temperature changes and object shapes that are difficult to obtain in the visible domain, especially under poor lighting conditions (e.g., dark nights and foggy days). This unique capability facilitates applicability across multiple domains, including surveillance, medical imaging, and temperature monitoring (Yu et al. 2022, 2023; Bao et al. 2023). However, infrared images are frequently in low resolutions, due to constraints in sensor technology and adverse environmental conditions (Huang et al. 2021). Enhancing the quality of these images is critical for accurate analysis and interpretation. While improving the performance of infrared devices is one approach, it is costly and inevitably constrained by physical limitations (Huang et al. 2022). Alternatively, super-resolution algorithms offer a promising solution to enhance the infrared image quality, directly refining the outputs of infrared imaging systems.

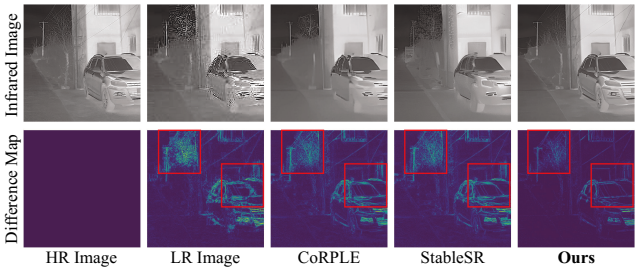

Figure 1: HR image, the corresponding LR image, and the SR results generated by CoRPLE (Li et al. 2024), StableSR (Wang et al. 2024), and our method (first row). **Spatial difference maps** (second row) demonstrate that our method produces results that closely match the infrared radiation intensity distribution of the HR Image.

Super-resolution is a fundamental task in computer vision that has gained significant attention. Some works (Wang et al. 2021c; Liang et al. 2021; Zhang et al. 2021) utilize CNN (LeCun et al. 1998) and Transformer (Vaswani et al. 2017) architectures to produce outputs with refined details and enhanced resolutions. Recent years have witnessed the notable progress of Diffusion Models (Ho et al. 2020), prompting exploration of their powerful generative prior in super-resolution tasks. Typical works such as (Lin et al. 2024; Wang et al. 2024; Saharia et al. 2023) demonstrate the diffusion model's ability to generate super-resolved results with realistic detail information. Nevertheless, these methods are proposed for visible data, which fundamentally differ from infrared data in their imaging principles.

*These authors contributed equally.
†Corresponding authors are Shuang Li and Jing Geng.

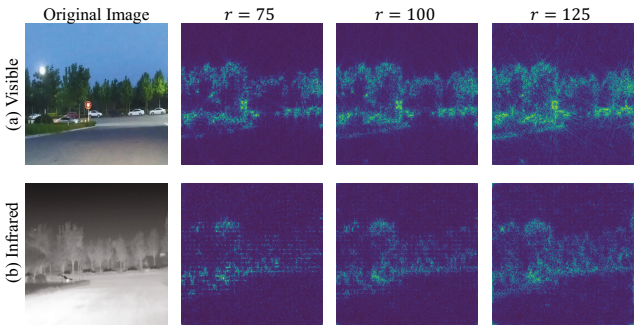

Figure 2: We conduct high-pass operations to visible and infrared images in the Fourier Spectrum domain and then reverse the FFT process. r stands for the radius of the filter mask. It is evident that the high-frequency information in the infrared images is **more ambiguous and weaker**.

On one hand, as previously discussed, infrared imaging captures the thermal radiation of objects. The intensity and spectral distribution of this radiation reflect an object's temperature, surface properties, and emissivity (Cui et al. 2020). As a result, thermal information is crucial for correctly interpreting infrared images and applying them in downstream tasks. Sadly, current super-resolution methods, whether designed for visible images or specifically tailored for the infrared domain, neglect this important aspect. As is shown in Figure 1, the outputs of these methods exhibit significant differences in infrared radiation intensity distributions, which in turn leads to suboptimal visual quality in our observation. On the other hand, as illustrated in Figure 2, the high-frequency information in infrared images is weaker and limited than that of visible images because of restrictions of infrared imaging devices (Huang et al. 2022), which leads to less clear and more ambiguous detail information. This creates more challenges to infrared image super-resolution tasks, as the desired method must struggle to reconstruct refined details utilizing modest high-frequency information.

Taking these two factors into account, we propose a diffusion model-based framework called **ThesIS** for infrared image super-resolution. Our practical findings indicate that relying solely on a pre-trained diffusion model is insufficient for reconstructing high-resolution infrared images while preserving correct thermal radiation distribution. Therefore, ThesIS involves two tailored operations, **Thermal Regularization** and **Dynamic High-Frequency Amplification**, to achieve a more feasible reconstruction of both visual representations and physical information. In Thermal Regularization, we first conduct *Infrared Radiation Intensity Restoration*, where a Thermal Restoration Network is utilized to restore the correct thermal radiation distribution, mitigating the damage caused by complex image degradation. This network is constrained by *Infrared Radiation Intensity Alignment Loss*, which is built on the principle of **one-to-one mapping** between grayscale values and radiation intensity. Given the network's limited capacity for visual reconstruction, its output is consequently exploited to fine-tune the pre-trained Diffusion model for refined visual details. Furthermore, we modify the sampling process by introducing *Thermal-Physics Guidance*, ensuring more accurate thermal radiation distribution during the denoising process. To address the ambiguous high-frequency information within infrared images, we propose *Dynamic High-Frequency Amplification*. Previous research (Si et al. 2024) has explored incorporating frequency operation in diffusion architectures to improve their output qualities. However, such approaches often rely on manually tuned parameters, which are difficult to generalize across diverse infrared imaging devices and degradation conditions. This motivates us to introduce *Dynamic Frequency Filter Block*, which involves learning a dynamic filter and adaptively enhancing high-frequency elements to acquire plausible and delicate image details.

It is also worth noticing the scarcity of publicly available infrared image super-resolution datasets, prompting us to introduce a novel dataset named **InfraredSR**. This dataset comprises 39,833 infrared images with a resolution of 512×512, significantly expanding upon the scale of the previous high-resolution infrared datasets (Shen et al. 2023).

Related Work

Visible Image Super-Resolution

Image super-resolution aims to enhance LR images into HR outputs. Traditional SR methods (Kim et al. 2016; Dong et al. 2016; Shi et al. 2016; Haris et al. 2018) utilize neural networks to learn a mapping from LR to HR using paired datasets, where LR images are directly down-sampled from the corresponding HR images. However, this simple pattern fails to reflect real-world image degradation. Alternative approaches (Bulat et al. 2018; Shocher et al. 2018; Wang et al. 2018b, 2021b,a) have been proposed to address this limitation by generating LR images adhering to realistic blurry distributions. GANs have also been explored for super-resolution tasks (Ledig et al. 2017; Wang et al. 2018a, 2021d; Liang 2021), achieving satisfactory results.

Recently, Diffusion Probability Models (Ho et al. 2020; Dhariwal et al. 2021) have gained significant attention. Researchers explore leveraging diffusion models to effectively enhance image resolution. SR3 (Saharia et al. 2023), a pioneering work in this area, adapts diffusion models and performs super-resolution through an iterative denoising process. Several subsequent studies (Lin et al. 2024; Wang et al. 2024; Yu et al. 2024) incorporate additional conditions to fine-tune the pre-trained Stable Diffusion. For instance, StableSR (Wang et al. 2024) introduces a time-aware encoder for model guidance and a controllable feature wrapping module to trade between quality and fidelity. While these methods have seen success in visible domains, they struggle to effectively address the unique challenges of infrared image super-resolution and produce suboptimal results.

Infrared Image Super-Resolution

Unlike visible one, infrared imaging faces challenges due to limitations in its imaging equipment, resulting in reduced contrast, diminished details and textures, which are reflected by high-frequency elements. To address these issues, specialized architectures have been developed.

Inspired by the advancements in the visible domain, researchers begin leveraging CNNs to super-resolve infrared images (Fan et al. 2021; Zou et al. 2021; He et al. 2018). Some approaches further incorporate information from the visible spectrum to aid in this task. For instance, PSRGAN (Huang et al. 2021) employs a GAN framework that integrates transfer learning to achieve high-quality results. Recently, Transformer-based methods have gained traction in this realm. CoRPLE (Li et al. 2024) involves prompt learning in the transformer architecture to improve output quality. The latest advancements include DifIISR (Li et al. 2025), which integrates gradient information from visual and perceptual priors into diffusion models (Ho et al. 2020), leading to superior visual fidelity and enhanced performance on downstream tasks. However, few methods take the important thermal properties of infrared images into account, leading to inaccurate thermal distributions and ultimately compromising result fidelity. Although (Zou et al. 2024) attempts to constrain thermal information in the spectral domain, it does not adequately address the degradation of thermal consistency under complex degradation scenarios.

Proposed Method

We propose an innovative architecture, **ThesIS**, for infrared image super-resolution towards real physical thermal properties and delicate visual representations. As illustrated in Figure 3, ThesIS comprises two core operations for robust training and progressive restoration. In *Thermal Regularization*, we propose the Infrared Radiation Intensity Alignment Loss to restore the accurate thermal radiation distribution disrupted by complex degradations and guide the denoising process of Stable Diffusion during inference. The *Dynamic High-frequency Amplification* leverages the Dynamic Frequency Filter Block (DFFB) to extract and enhance high-frequency components within the given feature maps, enabling the diffusion model to achieve super-resolution results with enhanced visual quality and refined details.

Thermal Regularization

Infrared Radiation Intensity Restoration. Based on the principles of infrared imaging, accurately restoring the infrared radiation intensity distribution requires alignment of the grayscale value distribution. In the infrared domain, grayscale histograms are commonly used to depict this distribution by illustrating the frequency of pixels at each intensity level. Consequently, we aim to leverage histogram-based tools to further regulate the grayscale distributions between super-resolution results and ground truth images. Inspired by (Avi-Aharon et al. 2023), we propose an *Infrared Radiation Intensity Alignment Loss* to serve as the thermal regularization based on differentiable histograms.

For a given infrared image $\boldsymbol{X}$, each pixel of it has a grayscale value $g(x_i) \in [0, 255]$, where x_i denotes the pixel-wise element of $\boldsymbol{X}$. The intensity of an image pixel is denoted as $I(x_i) \in [0, 1]$, where $I(x_i)$ is equals to $g(x_i)/255$. We then use the Kernel Density Estimation (KDE) (Davis et al. 2011; Parzen 1962) with a reflection boundary correction strategy to estimate the probability density function f_I of gray-level intensity:

$$\hat{f}_I(y) = \frac{1}{NW} \sum_{x_i \in \boldsymbol{X}} [\mathcal{K}(\frac{I(x_i) - y}{W}) + \mathcal{K}(\frac{I(x_i) + y - 2a}{W}) + \mathcal{K}(\frac{I(x_i) + y - 2b}{W})], \tag{1}$$

where $y \in [0, 1]$, N is the number of pixels and W represents the bandwidth. The kernel function $\mathcal{K}(\cdot)$ is defined as the derivative of a sigmoid function $\sigma(x)$, given by $\mathcal{K}(x) = \sigma'(x)$. Here, $a = 0$, and $b = 1$ represent the minimum and maximum values of the intensity interval.

We divide the pixel grayscale intensity interval $[0, 1]$ into 256 sub-intervals $\{B_k\}_{k=0}^{255}$. Each sub-interval has a length of $L = \frac{1}{256}$ and a midpoint $\mu_k = L(k + \frac{1}{2})$. Based on this setting, the probability of a pixel in the image falling within a particular grayscale interval can be derived as:

$$P_I(k) = \int_{y \in B_k} \hat{f}_I(y)\, dy. \tag{2}$$

The probability function $P_I(k)$ denotes the value of the k-th bin in our grayscale histogram. A differentiable histogram $\boldsymbol{H}$ for a given image then can be constructed:

$$\boldsymbol{H} = \{\mu_k, P_I(k)\}_{k=0}^{255}. \tag{3}$$

Let $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ represent the infrared images under comparison. The disparity in grayscale distribution can be quantified using the differentiable histograms. We employ the Wasserstein Distance (Rubner et al. 2000) to measure the similarity between these distributions, as it is particularly effective for this purpose. Denoting $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ as the histograms of images $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ respectively, Wasserstein Distance is equivalent to Mallows' distance for one-dimensional histograms with equal areas (Levina et al. 2001):

$$\mathrm{W}(\boldsymbol{H}_1, \boldsymbol{H}_2) = \left(\frac{1}{K}\right)^{\frac{1}{l}} \left\|\mathrm{CDF}(\boldsymbol{H}_1) - \mathrm{CDF}(\boldsymbol{H}_2)\right\|_l, \tag{4}$$

where $K = 256$, $\mathrm{CDF}(\cdot)$ is the cumulative density function. We then propose the **Infrared Radiation Intensity Alignment Loss** with the following formulation:

$$\mathcal{L}_{IRA}(\boldsymbol{I}_1, \boldsymbol{I}_2) = \sum_{i=0}^{255} (\mathrm{CDF}_i(\boldsymbol{H}_1) - \mathrm{CDF}_i(\boldsymbol{H}_2))^2, \tag{5}$$

where $\mathrm{CDF}_i(\boldsymbol{H})$ represents the i-th element of the cumulative density function of $\boldsymbol{H}$.

To address the distortion of true thermal radiation intensity, we propose a Thermal Restoration Network. Given the low-resolution image $\boldsymbol{I}_{LR}$, the proposed network generates the upsampled image $\boldsymbol{I}_{UP}$. Parameter optimization is achieved by minimizing the following loss function:

$$\mathcal{L} = \mathcal{L}_1(\boldsymbol{I}_{UP}, \boldsymbol{I}_{GT}) + \lambda \mathcal{L}_{IRA}(\boldsymbol{I}_{UP}, \boldsymbol{I}_{GT}), \tag{6}$$

where $\mathcal{L}_1(\boldsymbol{I}_{UP}, \boldsymbol{I}_{GT}) = ||\boldsymbol{I}_{UP} - \boldsymbol{I}_{GT}||_1$, and $\boldsymbol{I}_{GT}$ denotes the corresponding ground truth image.

This restoration module can effectively and reasonably restore the distribution of thermal radiation intensity as well

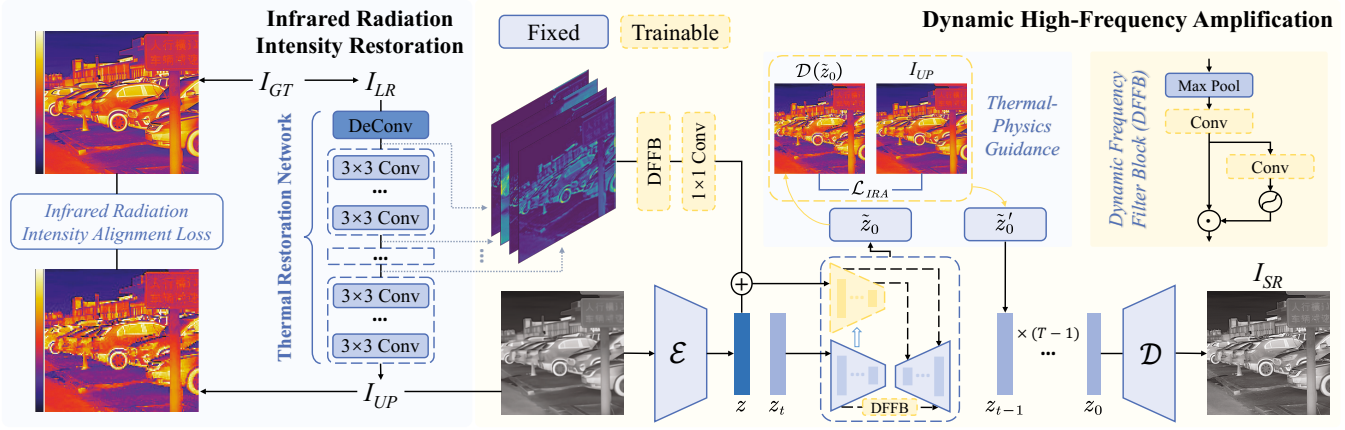

Figure 3: An overview of ThesIS. Infrared Radiation Intensity Restoration is proposed to recover the accurate thermal radiation distribution and guide the denoising process of Stable Diffusion. The Dynamic High-Frequency Amplification incorporates high-frequency prior into diffusion models, facilitating the generation of outputs with enhanced textures and finer details.

as perform preliminary upsampling. Furthermore, $\boldsymbol{I}_{UP}$ will serve as a guide for the finetuning of the generation module.

Thermal-Physics Guidance. Inspired by the remarkable generative capabilities of Stable Diffusion (Rombach et al. 2022), we incorporate it as the generative prior in our infrared super-resolution tasks. During the *diffusion process*, we use the encoder of Stable Diffusion to map $\boldsymbol{I}_{UP}$ into the latent space, obtaining the initial latent representation $\boldsymbol{z} = \mathcal{E}(\boldsymbol{I}_{UP})$, which is used both for the diffusion process and as the additional condition. We take $\boldsymbol{z} = \mathcal{E}(\boldsymbol{I}_{UP})$ as the starting point of the forward Markov process and gradually add Gaussian noise to generate the noisy features $\boldsymbol{z}_t$:

$$\boldsymbol{z}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{z} + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{7}$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$, $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

In the *denoising process*, a network ϵ_θ is learned by predicting the noise ϵ conditioned on c. At a time step t, we can obtain a cleaner latent $\boldsymbol{z}_{t-1}$:

$$\begin{aligned} \boldsymbol{z}_{t-1} &= \sqrt{\alpha_{t-1}}\tilde{\boldsymbol{z}}_0 + \sqrt{1-\alpha_{t-1}}\epsilon_\theta(\boldsymbol{z}_t|t,c), \\ \tilde{\boldsymbol{z}}_0 &= \frac{\boldsymbol{z}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\boldsymbol{z}_t|t,c)}{\sqrt{\bar{\alpha}_t}}. \end{aligned} \tag{8}$$

Therefore, the sampling process of $\boldsymbol{z}_{t-1}$ can be defined as $q(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t, \tilde{\boldsymbol{z}}_0)$. The iterative inference steps of the denoising process allow us to guide the sampling process with auxiliary information. Inspired by (Fei et al. 2023), we leverage the *Infrared Radiation Intensity Alignment Loss* to further steer the denoising process towards super-resolution results that maintain accurate physical properties. At each time step t, the denoiser of Stable Diffusion predicts the noise ϵ_t of the noisy latent $\boldsymbol{z}_t$. From this, we can derive $\tilde{\boldsymbol{z}}_0$ as outlined in Equation 8. To ensure that the distribution of thermal radiation remains accurate and undisturbed, we apply $\mathcal{L}_{IRA}$ between $\mathcal{D}(\tilde{\boldsymbol{z}}_0)$ and $\boldsymbol{I}_{UP}$ in the pixel space, and then update the latent $\tilde{\boldsymbol{z}}_0$ accordingly, as illustrated in Algorithm 1.

This guidance mechanism is essential for mitigating the intrinsic randomness of diffusion models and aiding the generation of infrared images with reasonable thermal radiation distribution during the denoising process.

Algorithm 1: Thermal-Physics Guidance

Input: Image $\boldsymbol{I}_{UP}$, control conditions c, guidance scale s
Output: Super-resolution result

Sample $\boldsymbol{z}_T$ from $\mathcal{N}(0, \mathbf{I})$
for t from T to 1 **do**
  $\tilde{\boldsymbol{z}}_0 = \dfrac{\boldsymbol{z}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\boldsymbol{z}_t|t,c)}{\sqrt{\bar{\alpha}_t}}$
  $\mathcal{L} = \mathcal{L}_{IRA}(\mathcal{D}(\tilde{\boldsymbol{z}}_0), \boldsymbol{I}_{UP})$
  Sample $\boldsymbol{z}_{t-1}$ from $q(z_{t-1}|z_t, \tilde{\boldsymbol{z}}_0 - s\nabla_{\tilde{z}_0}\mathcal{L})$
end for
return super-resolution result $\boldsymbol{I}_{SR} = \mathcal{D}(\boldsymbol{z}_0)$

Dynamic High-Frequency Amplification

Recent research has demonstrated that skip connections in U-Net architectures (Ronneberger et al. 2015; Zhou et al. 2018) introduce high-frequency features into the decoder module. Building on this insight, some approaches have proposed applying spectral modulation in the Fourier domain to selectively reduce low-frequency components in skip connections, thereby enhancing detail generation (Si et al. 2024). However, relying on a frequency-dependent scaling factor to retain high-frequency details in skip features often requires extensive manual tuning. This challenge is particularly pronounced in infrared images, where high-frequency information is less prominent, and diverse degradations can impact frequency content in complex ways. A pre-set scaling factor may therefore struggle to effectively address the variety of real-world degradation scenarios. To overcome this limitation, we propose a dynamic high-frequency feature amplification approach, implemented through Dynamic Frequency Filter Blocks (DFFB), as illustrated in Figure 3.

The DFFB aims to decompose the given features into separate frequency parts based on dynamically learned filters and retain the amplified high-frequency components. Specifically, given the input features $\boldsymbol{F}_s \in \mathbb{R}^{H\times W\times C}$, we propose

to learn dynamic frequency filters:

$$\begin{aligned} \boldsymbol{F}_s^{'} &= Conv(MaxPooling(\boldsymbol{F}_s)), \\ \boldsymbol{f}_{hp} &= \sigma(Conv(\boldsymbol{F}_s^{'})) \odot \boldsymbol{F}_s^{'}, \end{aligned} \tag{9}$$

where $Conv(\cdot)$ represents the convolutional layer, $\sigma(\cdot)$ is the sigmoid activation layer and $\odot$ denotes the element-wise product. The learned dynamic filters $\boldsymbol{f}_{hp}$ are used to extract high-frequency information within the given features.

We incorporate the proposed DFFB into the U-Net skip features to amplify high-frequency components for integration into the U-Net decoder. This frequency filter uses a dynamically learned approach to extract high-frequency information, aiding the diffusion model in generating super-resolution infrared images with enhanced texture fidelity.

In practical applications, we found that relying solely on a diffusion model fails to ensure consistency between the SR results and the original output, which is especially critical for high-frequency information in infrared images. This phenomenon prompted us to introduce conditional controls into the pre-trained diffusion models (Zhang et al. 2023). Multi-scale feature maps of the Thermal Restoration Network can provide relevant information about the image, especially the texture and detail information (Ding et al. 2021; Shang et al. 2024). These feature maps are concatenated along the channel dimension, and the proposed DFFB is also applied to extract and amplify high-frequency components. The enhanced high-frequency features are then processed through several convolutional layers to adjust their dimensionality, matching the shape of the latent features $\boldsymbol{z}$. The fusion of multi-scale features with the latent representation $\boldsymbol{z}$ offers refined guidance for the denoising process, enhancing the model's ability to recover fine details in infrared images.

InfraredSR Dataset

We propose a high-resolution infrared dataset, *InfraredSR*, containing 39,833 infrared images in total, intending to appraise the proficiency of approaches for infrared super-resolution. Specifically, the dataset includes high-resolution images captured using long-wavelength infrared (LWIR) cameras, with subsequent preprocessing performed to ensure uniformity across samples. All HR images are provided at a resolution of 512×512 for training purposes. For evaluation, we construct two test sets, *InfraredSR-Synthetic* and *InfraredSR-Real*, which are images with a resolution of 128×128 manually degraded using the multiple high-order degradation pipeline as (Wang et al. 2021c) and captured in real-world conditions, respectively.

Experiments

Settings

Implementation details. We perform $\times 4$ super-resolution tasks for both training and evaluation. In the *Thermal Regularization*, a Thermal Restoration Network is trained, comprising the first four blocks of VGG16 (with max-pooling layers removed) and three deconvolution layers. The Adam optimizer is adopted, and the learning rate is set as 1×10^{-4}.

Methods	PSNR↑	SSIM↑	LPIPS↓	FID↓	NIQE↓	MANIQA↑
RealSR	17.9849	0.5377	0.4699	62.6513	4.6909	0.2408
BSRGAN	20.5212	0.5760	0.4048	65.4193	4.9319	0.2840
RealESRGAN	21.6359	0.6022	0.4033	42.4908	4.9303	0.2640
SwinIR	21.2873	0.5989	0.4191	56.2623	6.4174	0.2809
HAT	18.8475	0.5984	0.5033	67.8188	6.2342	0.3002
StableSR	20.3682	0.6043	<u>0.3098</u>	38.7586	<u>4.4325</u>	0.3545
DiffBIR	20.5470	0.5993	0.3114	<u>38.5637</u>	4.4692	0.3312
DiffPIR	21.5780	0.6047	0.3596	39.3274	4.5347	<u>0.3624</u>
LKFormer	18.9746	0.6057	0.5031	63.1862	6.0996	0.2940
CoRPLE	21.0760	<u>0.6509</u>	0.5484	59.7039	6.9174	0.2703
DifIISR	<u>22.2931</u>	0.6469	0.3629	40.6326	4.5226	0.3604
Ours	**23.0170**	**0.6781**	**0.2921**	**37.1473**	**4.3803**	**0.3729**

Table 1: Comparison on the InfraredSR-Synthetic test set. We mark the best and second best performance in **bold** and <u>underline</u> texts respectively.

For the *Dynamic High-Frequency Amplification*, we construct our generation module based on Stable Diffusion with text prompts kept empty. The Adam optimizer is employed with a learning rate of 5×10^{-5}. As for inference, we adopt DDPM (Ho et al. 2020) sampling utilizing the proposed thermal-physics guidance with 50 time steps.

Datasets. We utilize the proposed InfraredSR dataset for both training and evaluation. Additionally, we evaluate our method on publicly available infrared image datasets, including IR700 (Datset 2021) and 101Thermal (Danaci et al. 2022), applying the degradation pipeline consistent with our InfraredSR-Synthetic to generate corresponding LR inputs.

Baseline Methods. We conduct a comprehensive comparison of our proposed method with several visible image super-resolution methods, including RealSR (Ji et al. 2020), BSRGAN (Zhang et al. 2021), RealESRGAN (Wang et al. 2021c), SwinIR (Liang et al. 2021), and HAT (Chen et al. 2023). Additionally, we compare our method with diffusion-based super-resolution models such as StableSR (Wang et al. 2024), DiffBIR (Lin et al. 2024), and DiffPIR (Zhu et al. 2023). We also include LKFormer (Qin et al. 2024), CoRPLE (Li et al. 2024), and DifIISR (Li et al. 2025), which are methods specifically designed for infrared images. All models are obtained from their official releases.

Evaluation Metrics. We utilize various image quality assessment metrics, including PSNR, SSIM, LPIPS (Zhang et al. 2018), FID (Heusel et al. 2017), NIQE (Mittal et al. 2012), and MANIQA (Yang et al. 2022), to evaluate the perceptual quality of super-resolved results on the InfraredSR-Synthetic test set. For the publicly available IR700 and 101Thermal datasets, we report metrics such as PSNR and SSIM. To further validate the effectiveness of our approach in real-world scenarios, we report NIQE and MANIQA scores on the InfraredSR-Real test set. In addition, we assess the alignment of infrared radiation intensity by computing the Correlation (COR) and Bhattacharyya distance (BHAD) (Kailath 1967) between the grayscale histograms of the ground-truth and super-resolved images.

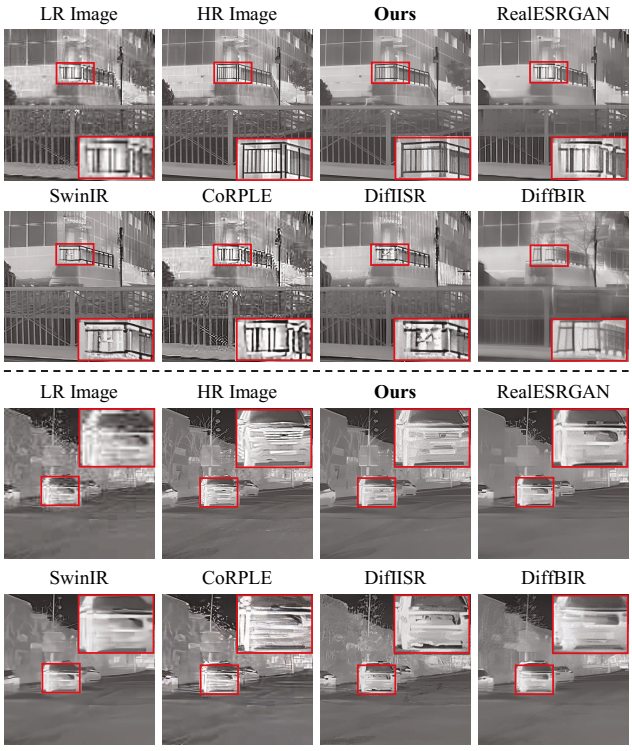

Figure 4: Qualitative comparisons with different methods on InfraredSR-Synthetic test set. Our method produces favorable super-resolution results close to the ground truth and with desirable visual details. Please zoom in for details.

Methods	InfraredSR		IR700		101Thermal	
	COR↑	BHAD↓	COR↑	BHAD↓	COR↑	BHAD↓
RealSR	0.9100	0.1533	0.9084	0.1612	0.9173	0.1509
BSRGAN	0.8984	0.1607	0.8674	0.1650	0.8982	0.1608
RealESRGAN	0.8421	0.1930	0.9151	0.1435	0.8729	0.1680
SwinIR	0.9214	0.1405	0.9250	0.1379	0.9034	0.1551
HAT	0.9133	0.1452	0.9210	0.1337	0.9231	0.1472
StableSR	0.8609	0.1629	0.8818	0.1636	0.8757	0.1692
DiffBIR	0.9276	0.1425	0.9281	0.1378	0.9196	0.1519
DiffPIR	0.8308	0.1929	0.8671	0.1927	0.8196	0.2019
LKFormer	0.9189	0.1448	0.9326	0.1306	0.9201	0.1489
CoRPLE	0.8917	0.1510	0.8981	0.1512	0.8929	0.1649
DifIISR	0.9238	0.1418	0.9223	0.1404	0.9180	0.1526
Ours	**0.9319**	**0.1387**	**0.9407**	**0.1294**	**0.9281**	**0.1454**

Table 2: Comparison based on grayscale histogram similarity metrics on the test set InfraredSR-Synthetic and public datasets IR700 and 101Thermal. Our method achieves SOTA performance to recover precise physical information.

Main Results

Quantitative Comparisons with Baseline Methods. Table 1 presents the evaluation metrics for the InfraredSR-Synthetic test set. Our method achieves state-of-the-art (SOTA) performance across all evaluation metrics. Notably, current SOTA methods proposed for visible images (Liang et al. 2021; Chen et al. 2023) struggle to achieve competitive scores, underscoring the unique challenges of infrared super-resolution and the effectiveness of our approach in the infrared domain. More importantly, Table 2 illustrates the similarity between the grayscale histograms of ground-truth images and super-resolution results. Our method generates outputs that most closely align with the thermal radiation intensity distribution of real-world scenes, demonstrating its ability to accurately restore infrared radiation information.

l_1 Loss for Thermal Restoration Network	IRA Loss for Thermal Restoration Network	PSNR↑	COR↑	BHAD↓
✗	✗	21.8976	0.9163	0.1508
✓	✗	22.5589	0.9274	0.1431
✓	✓	**23.0170**	**0.9319**	**0.1387**

Table 3: Ablation studies on components of our proposed method on the benchmark of InfraredSR-Synthetic dataset.

Qualitative Comparisons. We present comparisons for InfraredSR-Synthetic and InfraredSR-Real in Figure 4 and Figure 5, respectively. Our approach demonstrates superior visual quality, preserving delicate textures and fine-grained details. We further analyze the distribution of infrared radiation intensity and visualize the differences through difference histograms between the HR image and its counterparts in Figure 1. SR result of ThesIS closely matches the infrared radiation intensity distribution of the HR image, reinforcing its capacity to restore accurate physical information.

Ablation Studies

The Importance of Thermal Restoration Network. We ablate the Thermal Restoration Network by directly upsampling the LR image and feeding it into the diffusion model, while retaining the same feature extraction network for guidance. Evaluation is conducted on the InfraredSR-Synthetic test set, with metrics reported in Table 3. Removal of the Thermal Restoration Network leads to a noticeable decline in performance scores, highlighting its crucial role in alleviating real-world degradation effects and performing initial restoration of the thermal radiation information.

The Necessity of IRA Loss. We utilize the IRA Loss to regulate the alignment of infrared radiation distribution in Thermal Regularization. To evaluate its impact, we remove the IRA Loss and train the Thermal Restoration Network with only l_1 loss. As shown in Table 3, this modification results in the model failing to achieve satisfactory scores in terms of COR and BHAD, leading to unreliable infrared outputs. These results underscore the effectiveness of the IRA Loss in restoring the infrared radiation intensity distribution.

The Effectiveness of DFFB. We present the log amplitudes of Fourier-transformed U-Net features in Figure 6, affirming that DFFB amplifies high-frequency components. Additional studies are conducted by either removing the DFFB or replacing it with frequency scaling factors (Si et al. 2024). Quantitative results are provided in Table 4. The ob-

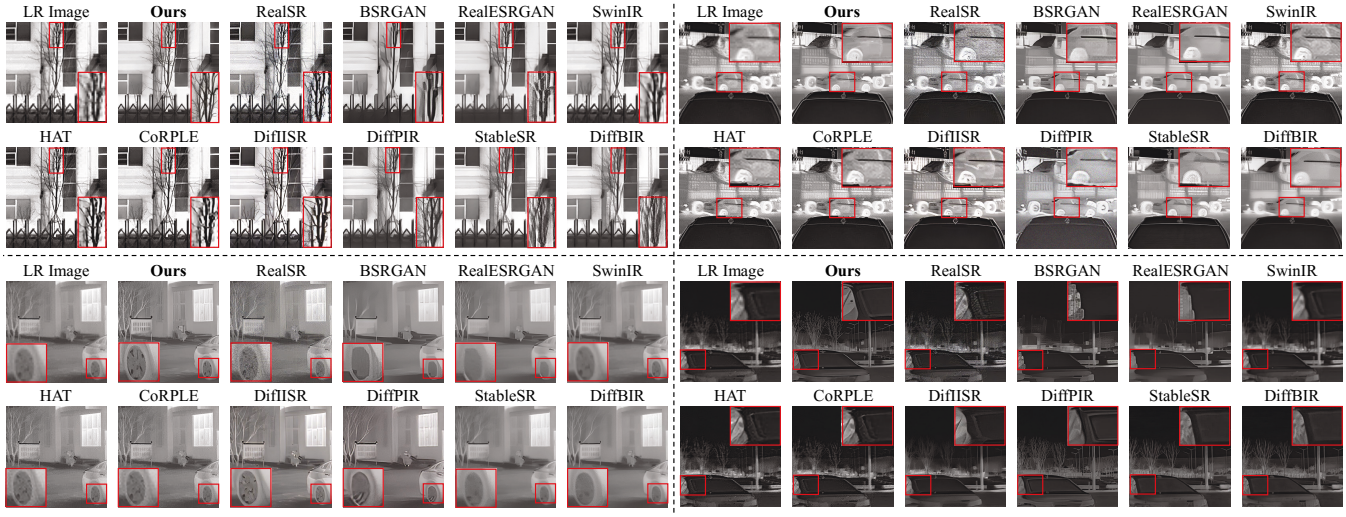

Figure 5: Qualitative comparisons with different methods on InfraredSR-Real test set. Our method yields favorable results with delicate details from a smaller resolution under complicated real-world settings. Please zoom in for details.

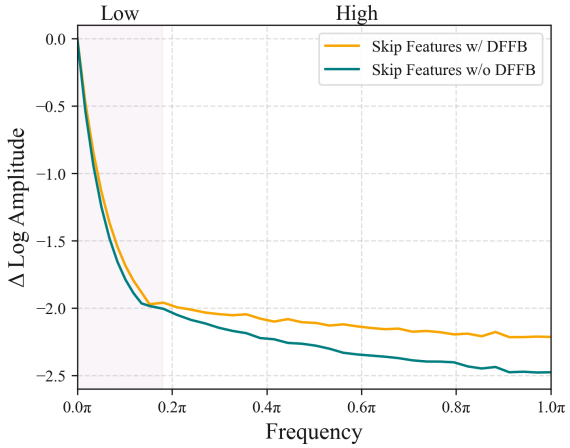

Figure 6: Log amplitudes in U-Net skip connections.

Methods	PSNR↑	SSIM↑	FID↓	MANIQA↑
Baseline	22.3626	0.6160	38.8312	0.3598
Frequency Scaling Factor	22.7863	0.6526	38.0522	0.3617
Ours	**23.0170**	**0.6781**	**37.1473**	**0.3729**

Table 4: Ablations on DFFB in U-Net.

served decline in performance scores when omitting the dynamic high-frequency operation or substituting it with fixed scaling factors confirms DFFB's role in retaining and amplifying high-frequency information in infrared images.

The Indispensability of Thermal-Physics Guidance and the Impact of Its Scale. We explore the necessity of the guidance mechanism for inference and investigate the impact of its scale s on super-resolution results, as shown in Table 5. Removing the guidance mechanism ($s = 0$) leads to unsatisfactory scores in terms of COR and BHAD. As the scale increases, both the fidelity of the images (measured by PSNR) and the preservation of accurate physical information improve and gradually converge, but the quality of outputs (represented by MANIQA) continues to decline. A larger scale of guidance may constrain the generated latent to be closer to the preliminary upsampled image, thus failing to produce refined textures and details. Therefore, we choose $s = 1$, as it doesn't significantly compromise image quality while effectively restoring accurate physical information.

Scale	PSNR↑	MANIQA↑	COR↑	BHAD↓
$s = 0$	22.3960	0.3877	0.9209	0.1408
$s = 0.5$	23.0325	0.3783	0.9287	0.1396
$s = 1$ **(Ours)**	23.0170	0.3729	0.9319	0.1387
$s = 2$	23.0179	0.3701	0.9321	0.1382
$s = 4$	23.0181	0.3694	0.9317	0.1385

Table 5: Ablations on different scales of the proposed thermal-physics guidance.

Conclusion

In this paper, we propose **ThesIS**, a novel framework specifically designed for infrared image super-resolution. Our approach employs **Thermal Regularization** and **Dynamic High-Frequency Amplification** operations to progressively restore physical properties and enhance the resolution of infrared images. We further propose **InfraredSR**, a new infrared super-resolution dataset comprising 39,833 images, to fill the current blank in this area. Experimental results demonstrate that our model achieves SOTA performance on both synthetic and real-world datasets, highlighting its effectiveness. We hope that our innovative focus on thermal radiation intensity for infrared super-resolution tasks, alongside the proposed dataset, will accelerate progress in both related research and practical applications.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (No. 62376026, 62388101, 42571343) and Beijing Nova Program (No. 20230484296).

References

Avi-Aharon, M.; et al. 2023. Differentiable histogram loss functions for intensity-based image-to-image translation. *TPAMI*, 45(10): 11642–11653.

Bao, F.; Wang, X.; Sureshbabu, S. H.; Sreekumar, G.; Yang, L.; Aggarwal, V.; Boddeti, V. N.; and Jacob, Z. 2023. Heat-assisted detection and ranging. *Nature*, 619(7971): 743–748.

Bulat, A.; et al. 2018. To Learn Image Super-Resolution, Use a GAN to Learn How to Do Image Degradation First. In *ECCV*, 187–202.

Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; and Dong, C. 2023. Activating More Pixels in Image Super-Resolution Transformer. In *CVPR*, 22367–22377.

Cui, S.-l.; et al. 2020. Universal formula of blackbody waveband radiation brightness response in the infrared temperature measurement technology. *Spectroscopy and Spectral Analysis*, 40(5): 1329–1333.

Danaci, K. I.; et al. 2022. A Survey on Infrared Image and Video Sets. *arXiv preprint arXiv:2203.08581*.

Datset, I. 2021. Public IR700 Dataset for Infrared Image Super-resolution.

Davis, R. A.; et al. 2011. Remarks on some nonparametric estimates of a density function. *Selected Works of Murray Rosenblatt*, 95–100.

Dhariwal, P.; et al. 2021. Diffusion Models Beat GANs on Image Synthesis. In *NeurIPS*, 8780–8794.

Ding, K.; Liu, Y.; Zou, X.; Wang, S.; and Ma, K. 2021. Locally adaptive structure and texture similarity for image quality assessment. In *ACM MM*, 2483–2491.

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2016. Image Super-Resolution Using Deep Convolutional Networks. *TPAMI*, 38(2): 295–307.

Fan, K.; et al. 2021. Infrared image super-resolution via progressive compact distillation network. *Electronics*, 10(24): 3107.

Fei, B.; Lyu, Z.; Pan, L.; Zhang, J.; Yang, W.; Luo, T.; Zhang, B.; and Dai, B. 2023. Generative diffusion prior for unified image restoration and enhancement. In *CVPR*, 9935–9946.

Haris, M.; et al. 2018. Deep Back-Projection Networks for Super-Resolution. In *CVPR*, 1664–1673.

He, Z.; Tang, S.; Yang, J.; Cao, Y.; Yang, M. Y.; and Cao, Y. 2018. Cascaded deep networks with multiple receptive fields for infrared image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(8): 2310–2322.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *NeurIPS*, 6626–6637.

Ho, J.; et al. 2020. Denoising diffusion probabilistic models. *NeurIPS*, 33: 6840–6851.

Huang, Y.; Jiang, Z.; Lan, R.; Zhang, S.; and Pi, K. 2021. Infrared image super-resolution via transfer learning and PSRGAN. *IEEE Signal Processing Letters*, 28: 982–986.

Huang, Y.; Miyazaki, T.; Liu, X.; and Omachi, S. 2022. Infrared Image Super-Resolution: Systematic Review, and Future Trends. *arXiv preprint arXiv:2212.12322*.

Ji, X.; Cao, Y.; Tai, Y.; Wang, C.; Li, J.; and Huang, F. 2020. Real-World Super-Resolution via Kernel Estimation and Noise Injection. In *CVPR Workshops*, 466–467.

Kailath, T. 1967. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1): 52–60.

Kim, J.; et al. 2016. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 1646–1654.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278–2324.

Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A. P.; Tejani, A.; Totz, J.; Wang, Z.; and Shi, W. 2017. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *CVPR*, 105–114.

Levina, E.; et al. 2001. The earth mover's distance is the mallows distance: Some insights from statistics. In *ICCV*, volume 2, 251–256.

Li, X.; Liu, J.; Chen, Z.; Zou, Y.; Ma, L.; Fan, X.; and Liu, R. 2024. Contourlet Residual for Prompt Learning Enhanced Infrared Image Super-Resolution. In *ECCV*, 270–288.

Li, X.; Wang, Z.; Zou, Y.; Chen, Z.; Ma, J.; Jiang, Z.; Ma, L.; and Liu, J. 2025. Difiisr: A diffusion model with gradient guidance for infrared image super-resolution. In *CVPR*, 7534–7544.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Gool, L. V.; and Timofte, R. 2021. SwinIR: Image Restoration Using Swin Transformer. In *ICCV Workshops*, 1833–1844.

Liang, Y. 2021. Unsupervised Super Resolution Reconstruction of Traffic Surveillance Vehicle Images. In *ICMLC*, 336–341.

Lin, X.; He, J.; Chen, Z.; Lyu, Z.; Dai, B.; Yu, F.; Qiao, Y.; Ouyang, W.; and Dong, C. 2024. Diffbir: Toward blind image restoration with generative diffusion prior. In *ECCV*, 430–448. Springer.

Mittal, A.; et al. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3): 209–212.

Parzen, E. 1962. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3): 1065–1076.

Qin, F.; Yan, K.; Wang, C.; Ge, R.; Peng, Y.; and Zhang, K. 2024. LKFormer: large kernel transformer for infrared image super-resolution. *Multimedia Tools and Applications*, 1–15.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *CVPR*, 10674–10685.

Ronneberger, O.; et al. 2015. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 234–241. Springer.

Rubner, Y.; et al. 2000. The earth mover's distance as a metric for image retrieval. *IJCV*, 40: 99–121.

Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, D. J.; and Norouzi, M. 2023. Image Super-Resolution via Iterative Refinement. *TPAMI*, 45(4): 4713–4726.

Shang, W.; Ren, D.; Zhang, W.; Fang, Y.; Zuo, W.; and Ma, K. 2024. Arbitrary-Scale Video Super-Resolution with Structural and Textural Priors. In *ECCV*, 73–90.

Shen, Y.; Kang, J.; Li, S.; Yu, Z.; and Wang, S. 2023. Style Transfer Meets Super-Resolution: Advancing Unpaired Infrared-to-Visible Image Translation with Detail Enhancement. In *ACM MM*, 4340–4348.

Shi, W.; Caballero, J.; Huszar, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; and Wang, Z. 2016. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In *CVPR*, 1874–1883.

Shocher, A.; et al. 2018. "Zero-Shot" Super-Resolution Using Deep Internal Learning. In *CVPR*, 3118–3126.

Si, C.; Huang, Z.; Jiang, Y.; and Liu, Z. 2024. Freeu: Free lunch in diffusion u-net. In *CVPR*, 4733–4743.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In *NeurIPS*, 5998–6008.

Wang, J.; Yue, Z.; Zhou, S.; Chan, K. C.; and Loy, C. C. 2024. Exploiting diffusion prior for real-world image super-resolution. *IJCV*, 1–21.

Wang, L.; Wang, Y.; Dong, X.; Xu, Q.; Yang, J.; An, W.; and Guo, Y. 2021a. Unsupervised Degradation Representation Learning for Blind Super-Resolution. In *CVPR*, 10581–10590.

Wang, W.; Zhang, H.; Yuan, Z.; and Wang, C. 2021b. Unsupervised Real-World Super-Resolution: A Domain Adaptation Perspective. In *ICCV*, 4298–4307.

Wang, X.; Xie, L.; Dong, C.; and Shan, Y. 2021c. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In *ICCV Workshops*, 1905–1914.

Wang, X.; Xie, L.; Dong, C.; and Shan, Y. 2021d. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In *ICCV Workshops*, 1905–1914.

Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; and Loy, C. C. 2018a. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In *ECCV Workshops*, 63–79.

Wang, Y.; Perazzi, F.; McWilliams, B.; Sorkine-Hornung, A.; Sorkine-Hornung, O.; and Schroers, C. 2018b. A Fully Progressive Approach to Single-Image Super-Resolution. In *CVPR Workshops*, 864–873.

Yang, S.; Wu, T.; Shi, S.; Lao, S.; Gong, Y.; Cao, M.; Wang, J.; and Yang, Y. 2022. MANIQA: Multi-dimension Attention Network for No-Reference Image Quality Assessment. In *CVPR Workshops*, 1190–1199.

Yu, F.; Gu, J.; Li, Z.; Hu, J.; Kong, X.; Wang, X.; He, J.; Qiao, Y.; and Dong, C. 2024. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *CVPR*, 25669–25680.

Yu, Z.; Chen, K.; Li, S.; Han, B.; Liu, C. H.; and Wang, S. 2022. ROMA: Cross-Domain Region Similarity Matching for Unpaired Nighttime Infrared to Daytime Visible Video Translation. In *ACM MM*, 5294–5302.

Yu, Z.; Li, S.; Shen, Y.; Liu, C. H.; and Wang, S. 2023. On the Difficulty of Unpaired Infrared-to-Visible Video Translation: Fine-Grained Content-Rich Patches Transfer. In *CVPR*, 1631–1640.

Zhang, K.; Liang, J.; Gool, L. V.; and Timofte, R. 2021. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution. In *ICCV*, 4771–4780.

Zhang, L.; et al. 2023. Adding conditional control to text-to-image diffusion models. In *ICCV*, 3836–3847.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*, 586–595.

Zhou, Z.; Rahman Siddiquee, M. M.; Tajbakhsh, N.; and Liang, J. 2018. Unet++: A nested u-net architecture for medical image segmentation. In *DLMIA*, 3–11. Springer.

Zhu, Y.; Zhang, K.; Liang, J.; Cao, J.; Wen, B.; Timofte, R.; and Van Gool, L. 2023. Denoising diffusion models for plug-and-play image restoration. In *CVPR*, 1219–1229.

Zou, Y.; Chen, Z.; Zhang, Z.; Li, X.; Ma, L.; Liu, J.; Wang, P.; and Zhang, Y. 2024. Contourlet Refinement Gate Framework for Thermal Spectrum Distribution Regularized Infrared Image Super-Resolution. *arXiv preprint arXiv:2411.12530*.

Zou, Y.; Zhang, L.; Liu, C.; Wang, B.; Hu, Y.; and Chen, Q. 2021. Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections. *Optics and Lasers in Engineering*, 146: 106717.