JPEG IMAGES RESTORATION WITH THUMBNAIL GUIDANCE AND SUPER-RESOLUTION

QIONGYANG HU

MPhil

The Hong Kong Polytechnic University

2026

The Hong Kong Polytechnic University
Department of Electrical and Electronic Engineering

JPEG Images Restoration with Thumbnail Guidance and Super-Resolution

Qiongyang Hu

A thesis submitted in partial fulfilment of the requirements
for the degree of
Master of Philosophy

August 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

Qiongyang HU (Name of student)

Abstract

Images are essential for communication, documentation, and artistic expression. However, during storage and transmission, bit errors may occur due to storage or transmission issues, leading to pixel distortion and affecting overall visual quality. Image restoration is crucial for applications in photography, medical imaging, and security. High-quality images enhance the perception of visual content and improve decision-making processes in various fields. As technology advances, the demand for robust restoration methods grows, making this research timely and significant. The primary goal of this thesis is to develop an optimized deep learning framework for image restoration, focusing on addressing challenges posed by bit errors and low resolution images. We first explore the challenges associated with restoring images affected by bitstream damage, highlighting prevalent issues such as loss of detail and the introduction of artifacts. For those type of bit errors, we utilize a three-step Mamba-based thumbnail-guided network designed to restore damaged images effectively. The proposed network features two parallel branches: one for direct image restoration and another utilizing thumbnails as auxiliary information to guide the restoration process. The methodology demonstrates significant improvements in restoration quality, showcasing the effectiveness of combining thumbnail data with deep learning techniques. Second, we address the issue of insufficient image resolution, aiming to enhance the image size while maintaining clarity and detail upon enlargement. We propose a graph-based framework to accomplish the single image super resolution task. The results reveal notable advancements in super-resolution performance, validating the proposed graph-based network's effectiveness in overcoming existing challenges. In summary, this thesis presents a novel dual-branch network to address bit error issues, along with a graph-based approach to tackle critical challenges related to low-resolution images, thereby making significant contributions to the field of image restoration. The findings underscore the potential of advanced deep learning techniques in improving image quality, paving the way for future research and applications in various domains of image processing technology.

Keywords: Deep learning, image restoration, bitstream corruption, image super-resolution, graph neural network

Publications Arising from the Thesis

Conference Paper

1. **Qiongyang Hu**, Yi Wang and Lap-Pui Chau. Restoration of Bitstream-Corrupted Images: A Mamba-based Thumbnail-guided Network. Published in IEEE International Symposium on Circuits and Systems (ISCAS), 2025.

Submitted Paper

1. **Qiongyang Hu**, Wenyang Liu, Wenbin Zou, Yuejiao Su, Lap-Pui Chau and Yi Wang. Beyond Fixed Neighborhoods: Single Image Super-resolution through Heterogeneous Subgraph Modeling. Submitted in IEEE Transactions on Multimedia.

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Dr WANG Yi, and co-supervisor Prof. CHAU Lap Pui. Their exceptional guidance, unwavering support, and generosity have profoundly influenced my academic journey and personal development. It has been a true honor to be part of their research group, and my two years at The Hong Kong Polytechnic University have been enriched by the company of brilliant, innovative, and compassionate individuals. Throughout this journey, I have not only enhanced my research skills but also learned the value of collaboration with people from diverse backgrounds to solve complex problems. The rigorous academic standards and passion for knowledge demonstrated by my mentors have inspired me to strive for excellence. I will forever cherish their teachings and endeavor to apply what I've learned to future challenges. Once again, I extend my heartfelt thanks to my supervisors and peers; your support and companionship have made my university experience truly unforgettable.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Image restoration [6, 16, 25, 73, 104, 119] is a critical aspect of image processing, aimed at recovering an original image from a degraded version. One of the significant challenges in this field is the impact of bit errors [76, 81], which can arise during image transmission or storage. In Fig. 1.1, the left side illustrates a bit error, which occurs during data transmission or storage when certain bits become corrupted. Bit errors, unlike blur and haze, introduce specific pixel-level distortions due to data corruption during storage or transmission, resulting in artifacts like color shifts and mosaic effects. While blur softens edges and reduces detail from motion or focus issues, and haze creates a washed-out appearance due to atmospheric conditions, bit errors can render parts of an image unrecognizable, as illustrated in Fig. 1.1 on the right.

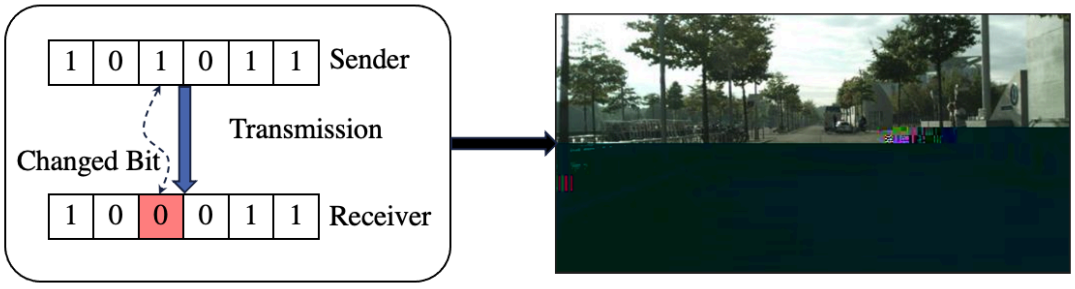


Figure 1.1: Bit error in an image.

Understanding these impacts is crucial for developing effective image restoration techniques processes. Restoring images affected by bit errors requires specialized algorithms that can correct these precise inaccuracies, highlighting the need for targeted

approaches in image processing. Despite advancements in image restoration techniques, several limitations remain. Traditional methods often struggle with complex degradations, while deep learning approaches [2, 56, 98], although powerful, can be sensitive to noise and artifacts introduced by bit errors. This gap highlights the need for more robust restoration methods that can effectively handle bit errors while improving image quality. Early restoration techniques relied on linear filters and statistical approaches. Common methods include: Median Filtering [4, 78]: Effective for removing salt-and-pepper noise, but can blur edges. Wiener Filtering [21, 29]: Based on statistical estimation, it adapts to local image variance but is not robust to all types of noise. Inverse Filtering [64]: Attempts to reverse the degradation process; however, it amplifies noise, especially when the degradation model is inaccurate. Deep learning has revolutionized image restoration but comes with its own set of challenges [12, 107]: Data Dependency: These methods require vast amounts of high-quality training data, which may not always be available. Bit Error Sensitivity: Deep learning models can be sensitive to bit errors, leading to significant degradation in output quality.

Thumbnails provide a way to offer valuable auxiliary information in image restoration, particularly in the context of bit errors. As low-resolution representations of full images stored in file headers, they can help guide the restoration process by providing a reference point for color and structural integrity. Besides, they supply contextual information that can aid in filling gaps left by bit errors, improving the overall recovery of the full-resolution image. To tackle the challenges posed by bitstream damage, such as detail loss and the emergence of artifacts, we propose a novel approach using a three-step Mamba-based thumbnail-guided network. This network comprises two parallel branches: one dedicated to direct image restoration and the other leveraging thumbnails as auxiliary data to enhance the restoration process.

The first approach hinges on the availability of the pristine reference image, yet such ground-truth data is not always retrievable in practical scenarios. Specifically, when decoding processes break down, and only low-resolution thumbnails remain accessible, super-resolution emerges as the sole viable remedy. Furthermore, in cases of high bitstream error rates where conventional restoration techniques prove ineffective, thumbnail data becomes the only usable input for subsequent processing. Given

that thumbnail super-resolution constitutes a foundational task within the broader domain of image restoration, our subsequent research focus shifts to this problem, which forms the core of our second contribution. Super Resolution (SR) [3, 20, 65, 67, 88] has been a significant topic in the field for many years. Super resolution technology seeks to enhance the resolution and detail of images, primarily focusing on the generation of high-resolution (HR) images from low-resolution (LR) inputs. However, the cost and technological constraints associated with acquiring high-resolution images render the processing of low-resolution images a critical challenge. Consequently, super resolution has become a pivotal research area aimed at improving image quality without incurring additional hardware expenses. Among the various approaches to super resolution, Single Image Super Resolution (SISR) [42, 96] stands out as one of the most prevalent methodologies. SISR specifically concentrates on leveraging the features of a single low-resolution image to facilitate high-resolution image reconstruction. This approach typically employs machine learning and deep learning techniques to extract salient features from low-resolution images, thereby producing clearer and more detailed high-resolution outputs. We propose a Heterogeneous Subgraph Network (HSNet) that integrates multiple subgraphs to enhance single image super resolution by capturing both local topologies and contextual relationships, achieving notable performance improvements.

1.1 Contributions and Thesis Organization

The entire thesis consists of five parts: Introduction, Literature Review, Restoration of Bitstream-corrupted Images, Single Image Super Resolution and Conclusion and Future Work, as shown in the Fig.1.2. The content of the following chapters is as follows.

Chapter 1: Introduction

An overview of the research topic, objectives, and significance.

Chapter 2: Literature Review

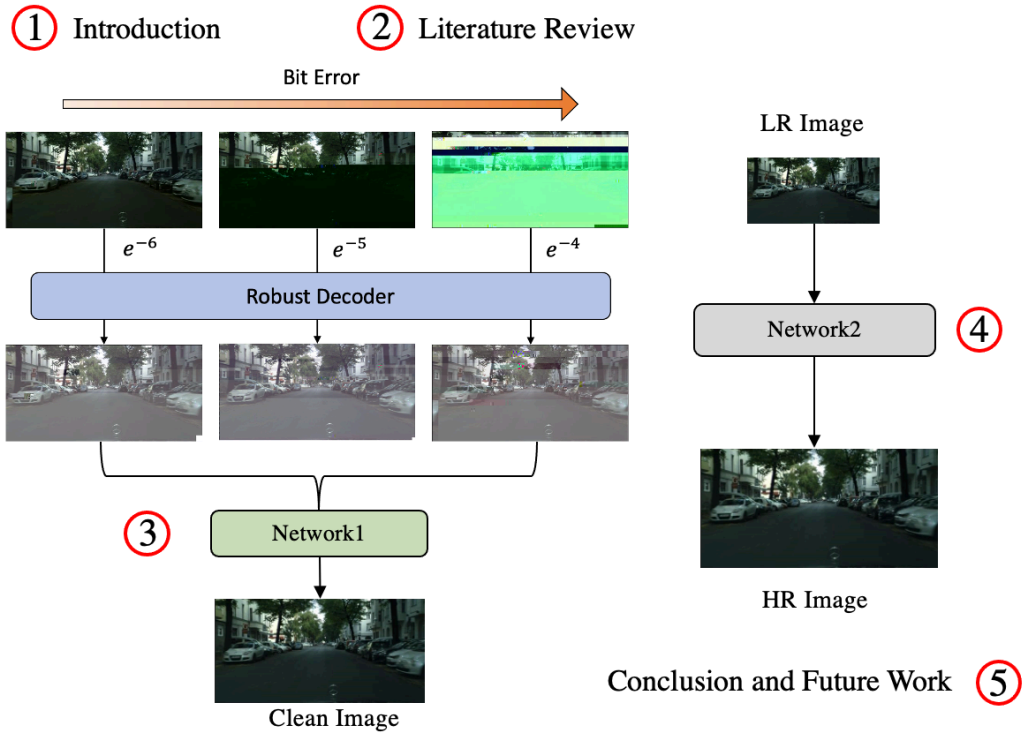


Figure 1.2: Thesis organization.

A review of existing literature related to image restoration methods.

Chapter 3: Restoration of Bitstream-corrupted Images

• Objective: To restore damaged images to closely resemble the original quality as much as possible.

• Key Contributions:

1. Introduces a Mamba-based thumbnail-guided network, designed to reconstruct corrupted images through a guided, multi-step process.

2. Incorporates Point-to-Point Restoration (PPR) Block to progressively merge information from the thumbnail at various depths.

3. Demonstrates superior performance on various image restoration benchmark datasets compared to existing methods.

Chapter 4: Single Image Super Resolution

• Objective: To extract and enhance details from low-resolution images, resulting in clearer and more realistic high-resolution images.

• Key Contributions:

1. Proposes HSNet, which leverages subgraph structure to enhance feature representation.

2. Enhances the richness and diversity of the feature set by combining information from multiple subgraphs, which is crucial for effective representation.

3. Achieves cleaner restoration results with fine textures compared to other SISR restoration algorithms.

Chapter 5: Conclusion and Future Work

Reviews the contributions of the thesis and discusses potential avenues for future research.

Chapter 2

Literature Review

In this chapter, we review in details the previous researches on corrupted JPEG images restoration and single image super-resolution, systematically analyze their intrinsic connections, existing limitations, and unaddressed research gaps. This review aims to justify the necessity of our proposed research by highlighting the urgent need for more robust and integrated solutions to tackle the challenges of bitstream-corrupted JPEG image recovery, which cannot be fully resolved by current fragmented techniques.

2.1 Corrupted JPEG Images Restoration

Traditional JPEG image restoration [6, 38, 52] focuses on addressing the artifacts and quality degradation that arise from lossy compression. The JPEG format employs discrete cosine transform (DCT) to compress images, which can lead to visual distortions such as blocking artifacts, ringing, and loss of detail. Common causes of JPEG corruption include transmission errors, storage issues, and repeated editing. Different types of damage, such as compression artifacts, color distortion, and block effects, significantly affect visual quality and subsequent processing tasks. Various methods have been developed to mitigate these effects, including spatial domain techniques, frequency domain approaches, and hybrid models. Spatial domain methods often involve filtering techniques [37, 72] that aim to smooth out artifacts, while frequency

domain methods [89, 95] attempt to reconstruct lost details by manipulating the DCT coefficients directly. However, these traditional approaches suffer from a fundamental trade-off: over-smoothing to eliminate artifacts inevitably erodes fine image details, especially in highly compressed images where the boundary between artifacts and valid details is blurred. This limitation makes them incompetent for handling severe bitstream corruption, which introduces irregular and pixel-level distortions beyond mere compression artifacts.

Despite advancements, these traditional approaches often struggle with balancing artifact reduction and detail preservation, particularly in highly compressed images. Recent studies [46, 52, 77, 106] have made notable progress in this field. For instance, U-Net [27] has gained prominence for its ability to capture multi-scale features through its encoder-decoder architecture, making it particularly effective in tasks like denoising and artifact removal. Similarly, Transformer-based approaches [13, 46, 103] have emerged, leveraging self-attention mechanisms to model long-range dependencies within images, which enhances the restoration process by enabling the model to focus on relevant contextual information. Nevertheless, most of these modern methods are designed for general compression artifact removal or mild noise reduction, rather than targeted recovery from bitstream corruption. They lack explicit mechanisms to model the irregular and random nature of bit errors, leading to suboptimal performance when faced with severe data loss, color shifts, or mosaic effects caused by bitstream damage.

Building on the foundation of traditional restoration techniques, bitstream-corrupted JPEG image restoration addresses a more specific challenge: the recovery of images that have been damaged during transmission or storage. This type of restoration is crucial in scenarios where JPEG images become corrupted due to data loss, noise, or interference. Notably, this task is inherently linked to single image super-resolution (SISR), as bitstream corruption often renders high-resolution (HR) images unreadable, leaving only low-resolution (LR) thumbnails (stored in JPEG file headers) as the sole available data. In such cases, restoring the corrupted image first requires enhancing the thumbnail resolution via SISR to obtain a usable reference, forming a sequential dependency between the two tasks.

Techniques in this area typically involve analyzing the structure of the JPEG bitstream to identify and correct errors, often utilizing redundancy in the encoding process. Recent advancements [50] have incorporated machine learning and deep learning methods to enhance restoration quality, allowing for more effective recovery of both structural integrity and visual fidelity. These modern techniques outperform traditional methods by better understanding the underlying patterns in image data, thereby providing a robust solution to the challenges posed by bitstream corruption. However, a critical gap remains: existing bitstream restoration methods either rely on the availability of intact HR image components (which is unavailable in extreme corruption scenarios) or fail to integrate auxiliary LR information (e.g., thumbnails) to guide the restoration process. This disconnect leaves a void in handling scenarios where only LR thumbnails are accessible, highlighting the need for integrated models that combine bitstream restoration with SISR.

2.2 Single Image Super-resolution

Single image super-resolution (SISR) refers to techniques that enhance the resolution of images, producing high-resolution outputs from low-resolution inputs. This field has gained significant attention due to its applications in various domains, including medical imaging, satellite imagery, and video enhancement.

For bitstream-corrupted JPEG image restoration, SISR is not merely a standalone task but a prerequisite: when the HR image bitstream is severely damaged, thumbnails (LR) become the only viable input, and high-quality SISR is essential to generate a reliable reference for subsequent restoration. This interdependency underscores the need for SISR methods that are robust to the artifacts often present in thumbnails derived from corrupted files. Early methods for super-resolution relied on interpolation techniques, such as bilinear and bicubic interpolation. While these methods are simple and computationally efficient, they often fail to restore fine details and can introduce artifacts. Key techniques in this area include: Interpolation Methods: Bilinear, Bicubic.

Reconstruction-Based Methods: These involve multiple low-resolution images of the same scene, aligning them to reconstruct a high-resolution image.

Single image super-resolution (SISR) has witnessed remarkable progress in recent years, with state-of-the-art methods integrating innovative architectural designs and learning strategies to further elevate the fidelity of super-resolved images. A large body of research has focused on refining convolutional neural network (CNN) architectures [17, 47, 50, 83, 84, 92] for enhanced SISR performance. While CNNs exhibit superior performance in modeling local spatial patterns, their inherent limitation of finite receptive fields renders them ineffective at capturing long-range contextual dependencies in images. To mitigate this drawback, researchers have proposed a suite of improvement strategies, among which dilated convolutions [32, 49, 55, 115] have been widely adopted to expand the effective kernel size, thereby enlarging the receptive field and boosting the modeling capacity for long-range information in SISR tasks.

Another prominent paradigm for SISR is the Transformer-based framework [8, 40, 41, 60, 61], which leverages self-attention mechanisms to model long-range dependencies in a flexible manner. Nevertheless, this strong representational capacity is accompanied by substantial computational overhead, which hinders its practical deployment. To alleviate this computational burden, researchers have explored content-aware optimization strategies [30, 93] that adopt adaptive information extraction schemes for different levels of image details. For example, several studies [35, 70] have attempted to enhance inter-window interactions by establishing semantic connections between non-overlapping local windows via shifted window mechanisms. Despite these optimizations, Transformer-based SISR methods suffer from high sensitivity to artifacts in low-resolution (LR) inputs (e.g., noise induced by bit errors), which tend to be amplified during the super-resolution process and thus degrade the quality of the final high-resolution (HR) outputs.

To further break the constraints of local window-based modeling, category-aware attention mechanisms have been introduced [110] to establish long-range correlations between semantically similar structural patterns within images. Even so, the majority of existing models still remain constrained by the inherent limitations of local

window partitioning. A critical research gap in current SISR research is the lack of methods that can simultaneously capture flexible local-topological correlations and global-contextual dependencies, while maintaining robust performance against artifacts caused by bitstream corruption. Existing approaches tend to prioritize either local detail modeling (CNNs) or long-range dependency capture (Transformers), failing to achieve an effective integration of both properties; moreover, none of these methods are explicitly designed to address the irregular artifacts inherent in thumbnail images derived from corrupted JPEG bitstreams.

In contrast, our proposed method abandons the restrictive local window paradigm and instead leverages subgraphs to model long-range spatial relationships, which enables the extraction of flexible and adaptive image representations that integrate both local and global information.

2.3 Graph Neural Network

In recent years, research has begun to shift focus toward irregular data structures (e.g., point clouds and social networks), where graph-based methods have been widely adopted to enable more flexible data modeling and processing. Nevertheless, the direct application of graph models to regular grid-structured data such as images remains a non-trivial challenge that demands targeted solutions. As a universal data structure, graphs offer a flexible paradigm for visual perception tasks: modeling an image as a graph facilitates the exploration of complex pixel-wise interdependencies, enables more robust handling of structural irregularities in visual data, and enhances the capacity to model intricate spatial patterns within images. For single image super-resolution (SISR), graph neural networks (GNNs) exhibit a unique inherent advantage: they can effectively model irregular pixel dependencies induced by bitstream corruption and capture global structural correlations that conventional CNNs and Transformer-based methods fail to represent adequately. This renders GNNs a highly promising framework for addressing the inherent limitations of existing SISR approaches.

Graph-based methods have been demonstrated to outperform traditional techniques in modeling both local and global contextual dependencies for visual tasks [57, 82, 94]. To balance modeling performance and computational efficiency, Han et al. [28] treat image patches rather than individual pixels as graph nodes, which effectively reduces model complexity while preserving the ability to capture fine-grained spatial relationships and patterns. Ren et al. [71] also adopt patch features as graph nodes and design a Key-Graph Constructor to generate a sparse and representative key graph by selectively establishing connections between salient nodes. In contrast, Tian et al. [85] construct image graphs with individual pixels as basic nodes instead of patches. Despite these advancements, existing GNN-based image processing methods suffer from two critical limitations that are particularly pronounced in the context of SISR with bitstream corruption: (1) Full-graph construction strategies are often adopted, which incurs exorbitant computational overhead; (2) The lack of adaptive mechanisms for selecting critical nodes and edges leads to either prohibitive computational costs or incomplete modeling of intrinsic spatial relationships in images.

To address these issues, our proposed method leverages the graph construction paradigm introduced in [117] to construct subgraphs, which significantly reduces the computational cost associated with full-graph construction. We then generate a robust heterogeneous graph from the constructed subgraph set. This heterogeneous subgraph design effectively mitigates the limitations of existing GNN-based methods through two key innovations: multi-scale node representations are integrated via a dedicated node sampling strategy (elaborated in the subsequent section), and computational overhead is substantially reduced through the modular construction of multiple subgraphs instead of a single full graph.

2.4 Feature Scanning in Mamba

Unlike conventional CNNs and Transformers, the emergence of Mamba [53, 109] has unlocked novel research avenues in the field of image restoration. As a hardware-aware algorithm, Mamba leverages recursive sequential scanning and dynamic adaptability to fully exploit GPU computational capabilities, while effectively mitigating inherent limitations of traditional convolutional methods—such as restricted receptive fields and inadequate modeling of long-range dependencies. Its sequential feature scanning mechanism and flexible directionality render it particularly well-suited for handling images with irregular artifacts, as it can adaptively focus on intact image regions to guide high-fidelity restoration. Furthermore, Mamba's inherent efficiency addresses the critical computational bottleneck of Transformer-based approaches, making it a promising and viable component for lightweight, high-performance image restoration models.

Owing to its intrinsic design, Mamba processes image patches in a sequential manner, supporting diverse scanning directions that substantially enhance image semantic understanding. In contrast, Vision Transformers [58] rely on multi-head self-attention mechanisms to model inter-patch relationships, whereas Mamba's sequential processing paradigm enables flexible exploration of multiple scanning trajectories across image patches. To date, extensive research efforts have been devoted to investigating novel scanning directions and their combinations, with the core objective of boosting Mamba's capability to capture fine-grained textures and global contextual information for image understanding tasks.

Notably, various selective scan techniques have been proposed under the umbrella of Selective Scanning Methods (SSM) for image and video processing tasks. A multitude of recent works [10, 54, 100] have identified and validated several common and effective scanning patterns, including Z-shaped, diagonal, and S-shaped trajectories. These diverse scanning strategies not only accelerate the research progress in Mamba-based models but also contribute to notable performance improvements in image processing tasks. Additionally, Pei et al. [68] proposed an atrous-based selective scan approach, which employs efficient skip sampling to simultaneously capture global contextual

information and fine-grained local features. Inspired by the efficacy of such selective scan mechanisms, we introduce a Node Sampling Strategy (NSS) in our work, designed to strategically sample critical nodes that are indispensable for constructing discriminative and efficient graphs in the context of super-resolution.

Chapter 3

Restoration of Bitstream-corrupted Images

We investigate the real-world JPEG image restoration problem with bit errors on the compressed bitstream. To mimic the effect of bit errors encountered in real images, we automatically inject various bit errors to generate damaged images, thereby simulating the bitstream-corrupted JPEG images in real situations.As illustrated in Fig. 3.1 and Fig. 3.2, the process involves the generation of corrupted JPEG images through encoding and decoding.

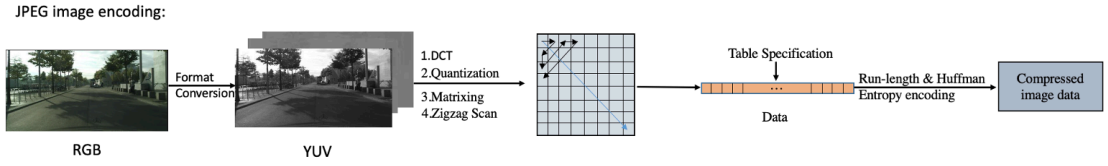


Figure 3.1: Data generation of compressed image data.

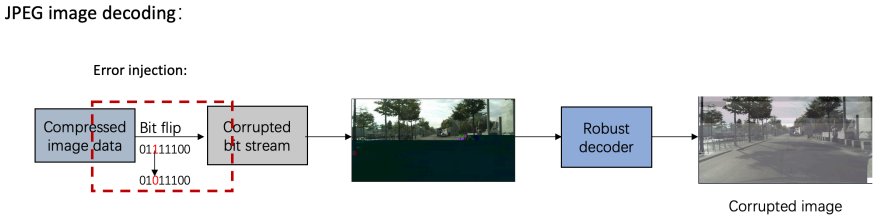


Figure 3.2: Process of bitstream corruption in a compressed JPEG image.

The image restoration problem is proposed to recover these images caused by bit errors that conventional decoders cannot perfectly decode. Typically, when a bit stream containing bit errors is decoded by the robust decoder, the resulting image exhibits two distinct characteristics: color casts and block shifts. To solve those problems, we propose a Mamba-based thumbnail-guided network to address the impact of color casts and block shifts on the image. The proposed framework is structurally divided into three blocks. Firstly, we use a Feature Aggregation (FA) block to reassemble the information from the corrupted image and the thumbnail image into a more acceptable input format for the subsequent networks, allowing for self-adjustment of the input format. Then, we design a Point-to-Point Restoration (PPR) block as a decoder to parse the features of inputs and thumbnails to generate coarse images. Finally, with the guidance of coarse images, a Pyramid Fusion (PF) block is used to generate the refined images. Extensive experimental results demonstrate our model outperforms state-of-the-art methods. Ablation studies and comparisons with super-resolution methods illustrate the effectiveness of our approach.

3.1 Background

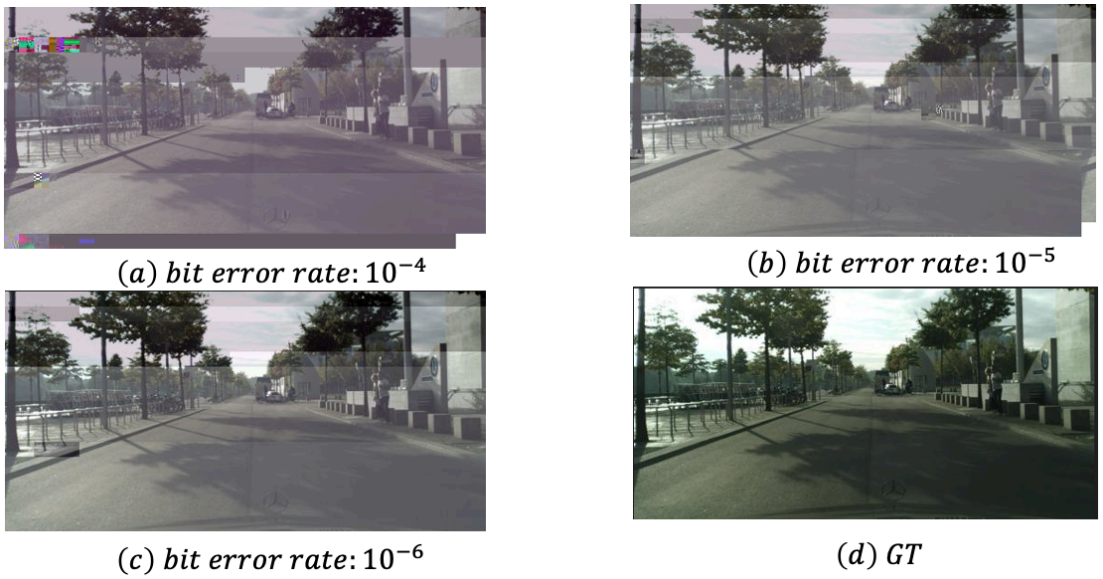
The image restoration task aims to recover accurate, precise, and complete original images from damaged or distorted ones. Restoration tasks can be categorized based on the type of damage, including image denoising [34], dehazing [43, 69], inpainting [9, 91], and super-resolution [79], all of which fall under pixel-level image restoration. Those approaches focus on individual pixel values to address quality degradation caused by noise or blur. In contrast, bit-level image restoration targets the underlying data (bitstreams), restoring damage from storage or transmission errors, often manifested as distortion, mosaic effects, and color deviations.

Bit-level errors commonly arise from aging data storage media, such as hard drives and memory cards [99], or during image uploads and downloads due to packet loss or noise [108]. These errors pose significant challenges, particularly in fields like medical image

analysis [66], where high image quality is critical. Therefore, rectifying bit errors is essential for maintaining data integrity.

Existing bit-level restoration methods can be divided into traditional and deep-learning approaches. Traditional techniques [23, 80] rely on error detection and correction to recover pixel-level images from corrupted bitstreams. In contrast, deep learning methods often employ a two-step strategy: first, mapping the erroneous bitstream to a pixel-level image, then applying advanced restoration techniques. This approach leverages neural networks' capabilities to recover high-quality images more effectively than traditional methods.

We adopt this two-step strategy by first employing a robust decoder [50] to transform bitstreams with errors into pixel-level images. However, adopting this approach leads to some adverse effects, including color casts and block shifts. Fig.3.3 shows the impact of various error rates on the decoded images.



(a) *bit error rate*: 10^{-4} (b) *bit error rate*: 10^{-5}

(c) *bit error rate*: 10^{-6} (d) *GT*

Figure 3.3: Examples of generated images with different levels of bit error rate.

Faced with those issues, Liu et al. [50] utilized a GAN-based method [86] for coarse restoration. However, this approach is less effective than more advanced methods. Transformer models, known for capturing long-range dependencies, face challenges with large image sizes due to computational complexity. In contrast, Mamba employs a hardware-aware algorithm that enhances training efficiency and effectively

addresses long-distance dependency issues. Besides, Mamba-based models exhibit enhanced training stability, which helps to mitigate issues such as mode collapse that are often encountered in GANs.

We propose a Mamba-based thumbnail-guided network to fully utilize the thumbnail information to guide full-resolution restoration. A single neural network may struggle with effective restoration, given the diverse degradation patterns. Thus, we introduce a Feature Aggregation (FA) block using a Mamba-based method to reconstruct favorable inputs for the network. Additionally, we integrate thumbnail information into a U-net-like Point-to-Point Restoration (PPR) block, leveraging multi-scale feature extraction to enhance the model's capability. This fusion of thumbnail features and transformed input features allows for a comprehensive representation. Unlike standard skip connections in U-Net [5, 19], we employ a mixed block to merge low-level and high-level features effectively. Finally, a Pyramid Fusion (PF) block generates refined images by progressively enhancing the thumbnail through multi-scale feature fusion, resulting in high-quality output.

The main contributions of this paper are as follows:

- Aiming to solve the image restoration problem caused by bit errors, we propose a three-step Mamba-based thumbnail-guided network, which achieves state-of-the-art restoration performance by capturing more comprehensive representations.

- To construct a learnable representation by the network, we incorporate a Feature Aggregation (FA) block, thereby converting the inputs into transformed inputs for the subsequent network.

- To fully excavate and utilize the relationship of thumbnails and transformed inputs, we propose a Point-to-Point Restoration (PPR) block to generate the coarse image.

- To integrate the information contained in the coarse image into the generation process of the final refined image, we use the pyramid fusion (PF) block, which

employs a coarse-to-fine manner, thereby yielding a notable enhancement in the fidelity of the ultimate refined image.

3.2 Method

As shown in Fig.3.4, our method consists of three components: the Feature Aggregation (FA) block, the Point-to-Point Restoration (PPR) block, and the Pyramid Fusion (PF) block.

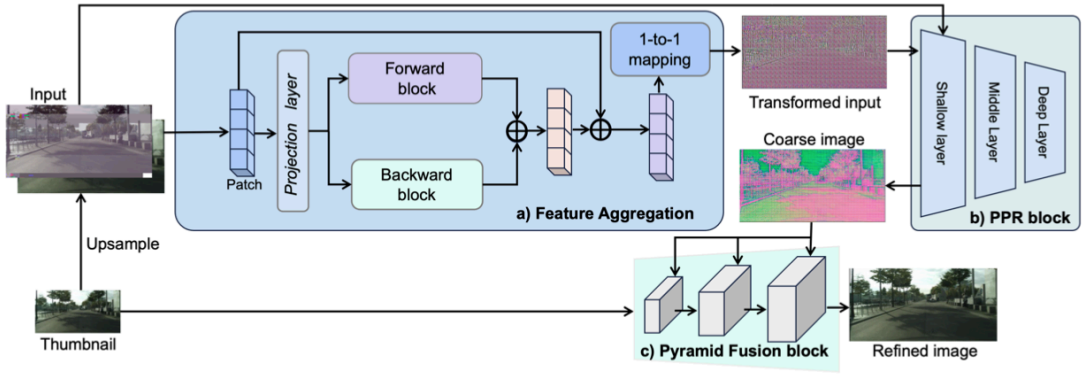


Figure 3.4: The architecture of our method consists of three components: Feature Aggregation (FA) block, Point-to-Point Restoration (PPR) block, and Pyramid Fusion (PF) block.

Specifically, we first use the encoder-decoder structure to perform coarse-grained image restoration. Next, building upon the thumbnail as the base, the information from the generated coarse image can be leveraged as auxiliary data to guide the final image restoration. The fine-grained restoration can be completed by integrating the coarse image and refining the thumbnail-based reconstruction with greater granularity.

3.2.1 Bit Error Generation in Bitstreams

The corrupted dataset is generated by injecting bit errors into bitstreams. Different from the previous work proposed by Liu et al. [50], we skip the process of decryption and encryption. The images that need to be restored can be simplified into the following formula:

$$Y = \mathrm{De}(\mathrm{Bitflip}(\mathrm{En}(x))), \tag{3.1}$$

where x denotes the original JPEG image, and De and En denote JPEG decoding and encoding, respectively. Bit flips denote the occurrence of random bit errors in the encoded data.

3.2.2 Feature Aggregation Block

For effective image reconstruction, it is essential to capture every detail across all channels. Hence, we propose a feature aggregation (FA) block, aiming to transform inputs into a learnable representation by the network. As shown in Fig.3.5, FA block consists of two main components: a feature extraction module based on the Mamba method and a 1-to-1 mapping block responsible for reconstructing the input.

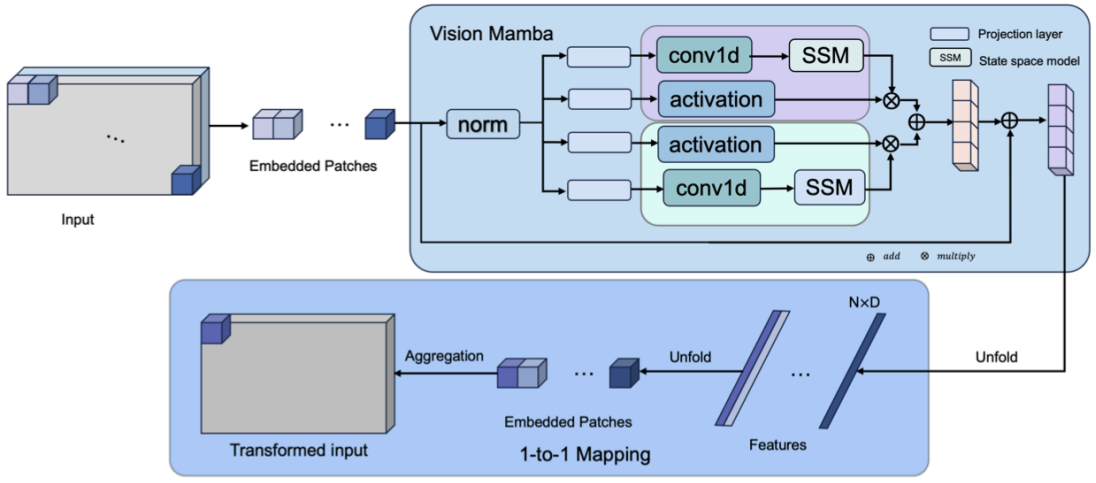


Figure 3.5: The Feature Aggregation (FA) block uses a Mamba-based feature extraction module and a 1-to-1 mapping block to transform inputs into a learnable representation for image reconstruction.

We concatenate damaged images with their corresponding thumbnails as inputs for the FA block. Thumbnails are low-resolution images stored in file headers, providing valuable clues for restoring full-resolution images. We utilize Vision-Mamba [120] as a robust feature extraction tool and then unfold the extracted features into the format of $[H, W, 3]$. The core rationale for selecting Vision Mamba as the backbone

architecture of our framework lies in its favorable computational complexity characteristics. Specifically, Vision Mamba operates with linear time and space complexity with respect to input image size, whereas Transformer-based models incur substantially higher computational and memory overheads, which become increasingly prohibitive when processing high-resolution visual data. Furthermore, in the context of low-quality image restoration tasks, Vision Mamba's proprietary selective state mechanism confers unique advantages: it can effectively suppress redundant noise components, amplify valid signal features, and mitigate the accumulation of restoration errors during iterative processing. Complementing this mechanism, its adaptive selective scanning capability enables dynamic allocation of computational resources, directing model focus toward image-damaged regions and structurally critical components (e.g., edges, textures, and object contours). This design circumvents a fundamental limitation of Transformer architectures, which adopt a uniform attention mechanism that treats all pixel positions equally. Such indiscriminate feature processing often leads to unintended noise amplification and the blurring of fine-grained details, ultimately compromising the quality of restored images.

Firstly, we divide inputs into patches. For the Vision-Mamba network, the input format is $[N, D]$, where N represents that the image has been divided into $(H/p \times W/p)$ blocks, D represents a patch that can be represented using D-dimensional features, p is the width and height of the patch. We specify D to $p \times p \times 3$ to ensure the size of the extracted features matches the reconstructed input in a one-to-one relationship, where 3 corresponds to the RGB channels. The 1-to-1 mapping block converts the extracted features into the transformed input for the subsequent network, which is a learnable representation in the shape of $[H, W, 3]$.

3.2.3 Point-to-Point Restoration Block

As shown in Fig.3.6, the PPR block has three layers, which process information from shallow to deep. The function of each layer can be split into three components: feature extraction, feature fusion, and feature selection.

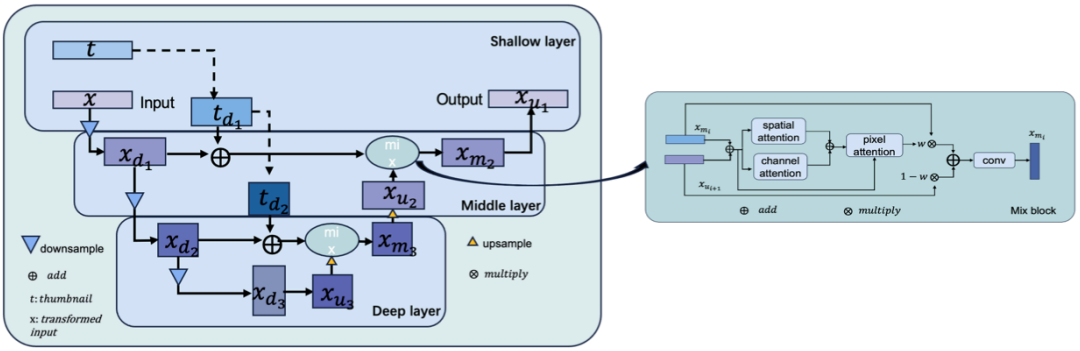


Figure 3.6: The Point-to-Point Restoration (PPR) block generates coarse images from transformed inputs and thumbnails, providing a preliminary rough restoration of the images.

We extract features from the reconstructed input and thumbnails. Then, we utilize different convolution operations in each block to capture more detailed information. The use of diverse convolution operations per layer solves the limitation of single-type convolutions that can only capture fixed receptive field information. By combining convolution types (e.g., standard, dilated, and depth-wise convolutions), the network extracts multi-scale details from both inputs, adapting to the varied feature distributions across shallow, middle, and deep layers. This multi-convolution extraction, paired with cross-layer additive fusion, ensures that shallow edge/texture details are preserved as the network deepens, avoiding the over-smoothing of restored images caused by shallow feature loss. Unlike U-net networks [59], the corresponding layers are directly added together for the feature fusion part. The fusion module integrates features from both shallow and deep layers, allowing for the perception of both high-level and low-level features. As the network deepens, the receptive field changes, and integrating information at different levels helps preserve features and prevents the network from losing shallow-level information. The three-layer shallow-to-deep progression of the PPR block further ensures a hierarchical refinement of features, where each stage builds on the previous one to balance structural consistency and detail richness—making it highly effective for generating the coarse image that underpins subsequent high-quality restoration.

3.2.4 Pyramid Fusion Block

The pyramid fusion (PF) block aims to gradually refine the low-resolution thumbnail by integrating the high-frequency details from the coarse image at different scales, effectively recovering a high-quality image. Unlike image super-resolution, the restoration process relies on the guidance of a coarse image recovered from the high-resolution corrupted input.

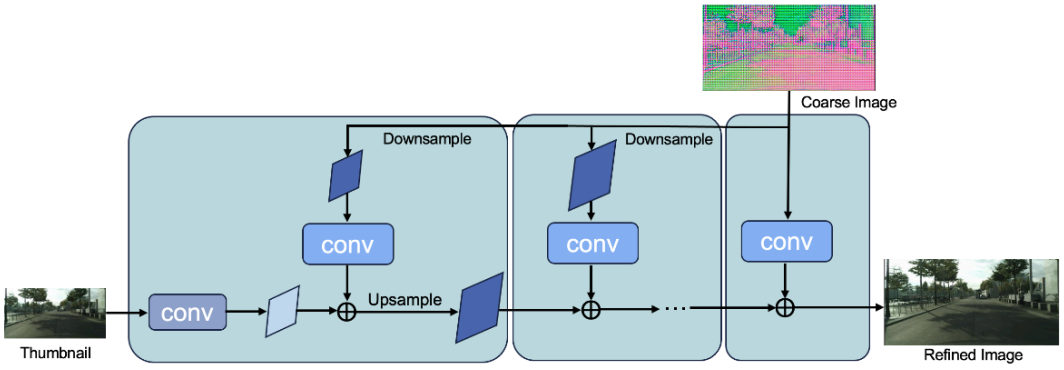


Figure 3.7: Pyramid fusion (PF) block.

By leveraging multi-scale feature fusion, the PF block is able to preserve rich details in the final output image, as illustrated in Fig. 3.7. Specifically, the upsampled thumbnail is gradually fused with the multi-scale high-frequency details extracted from the coarse image. Through this bidirectional approach, the information from thumbnail and coarse images is fused at different scales.

The PF block is selected for balancing detail preservation and structural consistency via its multi-scale bidirectional fusion mechanism. First, by decomposing the coarse image into multi-scale high-frequency detail maps, it can capture both local fine details (small scale) and global structural details (large scale), matching the hierarchical feature requirements of the upsampled thumbnail. Second, the gradual fusion strategy avoids the abrupt feature conflict caused by one-step fusion: the low-resolution thumbnail is upsampled progressively, and at each stage, it fuses with the corresponding scale of high-frequency details from the coarse image. This ensures that the supplemented details are consistent with the current resolution of the thumbnail, achieving precise detail filling. Third, unlike image super-resolution, which generates high-resolution

images from scratch based on low-resolution inputs, the PF block takes the coarse image (recovered from corrupted high-resolution data) as guidance, which provides a more reliable structural reference for detail restoration and significantly reduces the risk of generating unrealistic artifacts.

3.2.5 Loss Funciton

We employ contrast loss [90] in this paper. Inspired by the dehazing network, we use contrastive regularization (CR) to constrain the generated images. The overall loss function is formulated as:

$$L = \lambda_1(\min\|GT - R\|) + \lambda_2(\sum_{i=1}^{n} w_i \cdot \frac{D(F_i(GT), F_i(R))}{D(F_i(C), F_i(R))}), \tag{3.2}$$

where λ_1 and λ_2 are hyper-parameters. This loss has three inputs: refined image (R), ground truth (GT), and corrupted image (C). $F_i, i = 1,2,..n$ extracts the i_{th} hidden features from the pre-trained model [75]. $D(x,y)$ is the $L1$ distance between x and y. w_i is the weight coefficient.

3.3 Experiment

3.3.1 Implementation Details

Datasets. We conduct experiments on the Cityscapes [14] dataset. Its image resolution is 2048 $\times$ 1024, and we obtained image resolutions of 512 $\times$ 256 through downsampling. We generate damaged images through the SCA method in Two-ACIR [50]. We can generate images with different degrees of damage by controlling the bit error rates. For the thumbnail, we fix its size to 256 $\times$ 128, which is obtained through bicubic downsampling by ground truth.

Training details. We conducted our experiment on an NVIDIA GPU GeForce RTX 4090. We adopt the Adam optimizer with a start learning rate of 0.0001, a batch size of 5, and the exponential decay rates of $\beta 1$,$\beta 2$ = (0.9,0.999). The total epochs are set to 200. The hyper-parameters of the loss function are configured as (λ_1,λ_2) = (1, 0.1).

Evaluation metrics. Peak signal noise ratio (PSNR) and structural similarity index (SSIM) are commonly used indicators in restoration tasks to measure the performance of models. We use them to test the quality of restored images. The indicators can reflect the restoration ability of the model under different error conditions.

3.3.2 Comparisons of Image Restoration Methods

To assess the performance of our method, we conducted comparative studies against existing image restoration techniques (ACIR [50], DEA [11], UVM [116]) and super-resolution methods (SAFMN [79], Seemore [102]). These comparisons were performed on the Cityscapes dataset with varying bit error rates to evaluate the robustness of our approach across different levels of image degradation. Our evaluation results highlight the strengths of the proposed approach.

Table 3.1: Quantitative comparison of image restoration methods on the Cityscapes dataset with 512 × 256 resolution.

Methods \ Error rate	10^{-4}	10^{-5}	10^{-6}
2x Upsampled thumbnail	PSNR: 30.49 SSIM: 0.86		
UVM scaled [116]	PSNR: 25.71 SSIM: 0.85	PSNR: 27.95 SSIM: 0.94	PSNR: 28.04 SSIM: 0.95
DEA [11]	PSNR: 30.46 SSIM: 0.91	PSNR: 34.91 SSIM: 0.97	PSNR: 36.26 SSIM: 0.98
ACIR [50]	PSNR: 41.65 SSIM: 0.98	PSNR: 43.68 SSIM: 0.98	PSNR: 43.32 SSIM: 0.98
Our	**PSNR: 41.91 SSIM: 0.98**	**PSNR: 46.15 SSIM: 0.99**	**PSNR: 46.70 SSIM: 0.99**

We quantitatively compare our method with different methods in Tab. 3.1. Compared with ACIR, our method performs better under different errors. Our method outperforms baseline methods with an error rate of 10^{-6}, achieving a 3 dB PSNR enhancement in 512 $\times$ 256 resolution. In the case of a low error rate, such as 10^{-6} or 10^{-5}, our method is more capable of uncovering the relationship between the corrupted image and the thumbnail, thereby leveraging the detailed information contained in the generated coarse image to complete the image restoration. The similarity in performance between ACIR and the proposed approach at the high error rate of 10^{-4} can be attributed to the substantial degradation in the quality of the corrupted input data, which limits the ability to recover detailed information in the final output, irrespective of the specific method employed.

Due to the limited methods available for this specialized task, we compared the methods (DEA [11] and UVM scaled [116]) used for the image dehazing task. Because UVM has a large number of model parameters, which can lead to overfitting during the training process. We have adopted a reduced version that matches the number of layers in UVM with the number of layers mentioned in the paper, both being three layers. Due to the structure of DEA [11] and UVM scaled [116], the inputs for DEA [11] and UVM scaled [116] are both corrupted images, which limits their performance. Compared to our method, we introduce an additional information source (the thumbnail), which helps to improve the robustness and accuracy of the restoration. The experimental results demonstrate the superiority of our approach. Image dehazing methods that use only corrupted images as inputs, without leveraging the information from thumbnails, are far inferior to our approach.

Table 3.2: Quantitative comparison in PSNR of Bicubic, efficient SR networks, and our method at scale factor 2 on Cityscapes datasets with 512 $\times$ 256 resolution under an error rate of 10^{-4}.

Methods	scale factor: $\times$ 2
Bicubic	PSNR: 30.49
SAFMN [79]	PSNR: 30.56
Seemore [102]	PSNR: 32.20
Ours	**PSNR: 41.91**

Tab.3.2 reveals that employing thumbnail images as the input for state-of-the-art super-resolution methods, such as SAFMN [79] and Seemore [102], does not yield as effective image restoration performance as our proposed approach. We assess the performance at a bit error rate of 10^{-4}. Specifically, when the super-resolution methods use a scale factor of $\times$ 2, taking 128 $\times$ 256 thumbnail inputs to produce 256 $\times$ 512 outputs, they attain PSNR values of only 30.56 dB and 32.20 dB, respectively. In contrast, our proposed method achieves a significantly higher PSNR of 41.98 dB using a thumbnail size of 512 $\times$ 256 on the test dataset, outperforming the super-resolution-based approaches by a considerable margin.

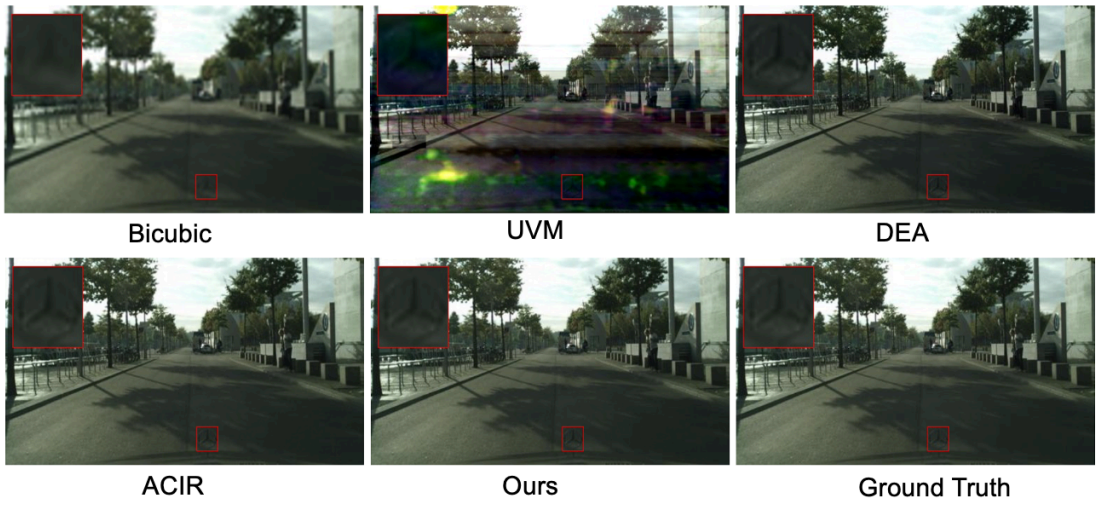


Figure 3.8: Visual comparison of different methods on Cityscapes dataset with 512 × 256 resolution under an error rate of 10^{-6}.

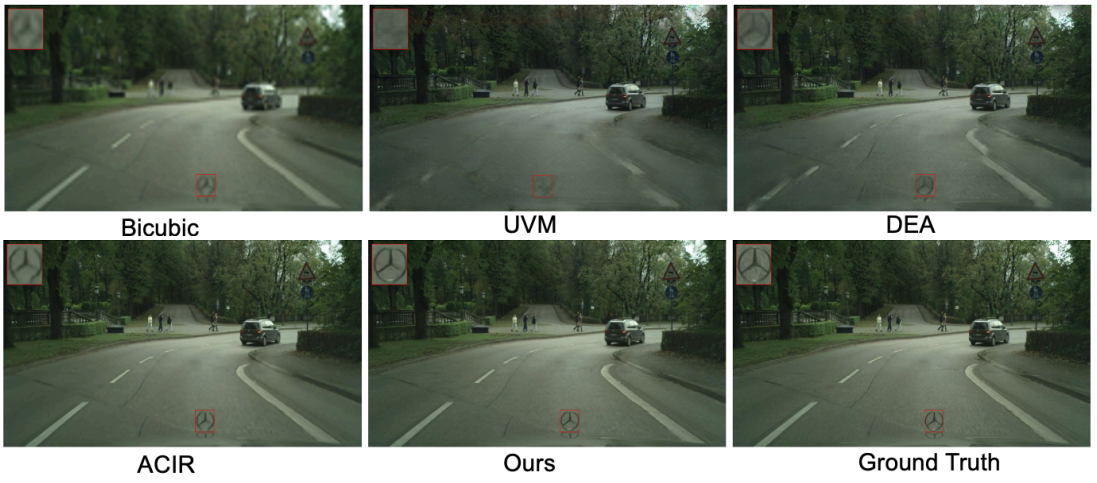


Figure 3.9: Visual comparison of different methods on Cityscapes dataset with 1024 × 512 resolution under an error rate of 10^{-4}.

The visual comparison of different methods on the Cityscapes dataset highlights their performance at various resolutions and error rates. Fig.3.8 presents a comparative analysis at a resolution of 512×256 pixels, maintaining an error rate of 10^{-6}. This figure effectively highlights the capability of each method to preserve critical details and clarity under minor error conditions, thereby offering insights into their efficacy in capturing essential features. Fig.3.9 extends this analysis to a higher resolution of 1024×512 pixels, with a more lenient error rate of 10^{-4}. The increased resolution permits a more nuanced examination of the methodologies, allowing for the discernment of intricate details. The visual results reveal significant variations in fidelity, illustrating how each approach adapts to elevated resolutions while managing the specified error constraints. Collectively, these figures underscore the influence of resolution and error rate on the performance of distinct image processing techniques.

3.3.3 Ablation Study

The effectiveness of each component: we removed several proposed components and tested their performance on the Cityscapes dataset. We evaluate the performance under a bit error rate of 10^{-4}.

Table 3.3: Ablation Study.

Methods \ Error rate			e^{-4}		e^{-5}		e^{-6}	
FA	PPR	PF	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
✓	✗	✗	36.55	0.94	38.66	0.96	38.94	0.96
✓	✓	✗	40.75	0.98	45.24	0.99	46.60	0.99
✓	✗	✓	40.74	0.98	44.62	0.99	45.46	0.99
✓	✓	✓	**41.91**	**0.98**	**46.15**	**0.99**	**46.70**	**0.99**

The FA block serves as a base component responsible for reconstructing the input form. The results presented in the first row of Table 3.3 demonstrate the effectiveness of the FA block as a core component of the proposed restoration framework. 1) Point-to-point restoration (PPR) block: Without the PPR, the restoration performance drops

by 1.2 PSNR on average compared to the full model. 2) Pyramid fusion (PF) block: Removing the PF block results in a 1.2 PSNR decrease in restoration quality.

Compared to the FA block, the addition of the PPR block results in an average PSNR enhancement of 4.2 dB. Furthermore, incorporating the PF block into the FA-based architecture leads to a 4.2 dB average PSNR increase. Ultimately, the employment of all three blocks - FA, PPR, and PF - attains an impressive 5.4 dB average PSNR improvement, demonstrating the highly effective restoration capabilities of the proposed framework for bitstream-corrupted JPEG images.

3.4 Summary

This chapter proposes a Mamba-based thumbnail-guided network that uses a three-step image restoration approach. Firstly, the Feature Aggregation (FA) block is to reconstruct the input form. Secondly, the Point-to-Point Restoration (PPR) block is leveraged to generate coarse images. It has a shallow-to-deep structure that can progressively merge the information from the thumbnail. Finally, the Pyramid Fusion (PF) block is used to refine the image. The proposed three-step method combines complementary techniques to address the intricate task of image restoration, enabling the network to produce high-fidelity restored images effectively. Our proposed method demonstrates significant effectiveness, as evidenced by extensive experimental results and ablation studies. These findings highlight the potential of our approach to address this challenging problem. We believe further exploration of this problem and our solution holds promising avenues for future research.

Chapter 4

HSNet Framework Introduction

Motivated by the thumbnail-related findings presented in the preceding chapter, this work aims to conduct an in-depth investigation into the impacts of low-resolution (LR) degradation on image restoration tasks. Specifically, our research focus is directed toward single image super-resolution (SISR), a fundamental yet challenging task in computer vision. Existing SISR methods based on convolutional neural networks (CNNs) and attention-based mechanisms are inherently constrained by their rigid architectural designs, which limit their flexibility in modeling complex image dependencies. In contrast, graph-based approaches, which offer superior flexibility in capturing intricate spatial correlations, are often plagued by prohibitive computational overhead, hindering their practical applicability and scalability. To address these inherent limitations of existing SISR paradigms, this chapter introduces the Heterogeneous Subgraph Network (HSNet), a novel and computationally efficient framework designed to fully leverage the advantages of graph-based modeling while mitigating its computational drawbacks. The core insight underlying HSNet is the decomposition of the computationally expensive global graph construction problem into a series of modular, manageable sub-tasks. To this end, we first introduce the Construct Subgraph Set Block (CSSB), a dedicated module engineered to generate a diverse ensemble of complementary subgraphs. Rather than constructing a single resource-heavy global graph, the CSSB encapsulates the multi-faceted intrinsic characteristics of visual data by explicitly modeling distinct relational patterns and feature modalities. In doing so,

it yields a rich set of graph structures spanning both local and global scales, which collaboratively encode comprehensive semantic and structural information of the input image.

Subsequently, we propose the Subgraph Aggregation Block (SAB) to effectively fuse the discriminative features embedded within each individual subgraph. By dynamically assigning attention weights to different subgraphs and adaptively integrating information from multiple graph perspectives, the SAB constructs a holistic and robust feature representation that accurately captures subtle inter-node dependencies and global structural correlations. Furthermore, to boost model efficiency while preserving the discriminability of learned features, we devise a dedicated Node Sampling Strategy (NSS) that adaptively selects and extracts the most prominent and informative node features from the input, thus minimizing computational overhead without degrading super-resolution performance.

Extensive experimental evaluations are conducted on standard benchmark datasets to validate the effectiveness and efficiency of the proposed HSNet. Extensive experimental results validate that our proposed framework attains state-of-the-art (SOTA) performance against contemporary methods, while striking an optimal trade-off between super-resolution accuracy and computational efficiency.

4.1 Background

For digital image processing, the demand for image upscaling is pervasive across various practical scenarios. Nevertheless, traditional interpolation-based upscaling algorithms frequently introduce undesirable visual artifacts, such as blurriness and block effects, which essentially undermine the image fidelity required for high-quality visual perception. To mitigate these inherent limitations, single image super-resolution (SISR) [39, 88, 101] has evolved as a pivotal technology, which inherently enhances image clarity and fine-grained details by reconstructing high-resolution (HR) images from their corresponding low-resolution (LR) counterparts. As a result, this versatile

technology has been widely adopted in a diverse range of fields, including but not limited to medical imaging [24], remote sensing and satellite imagery analysis [74], and high-reliability video surveillance systems [97].

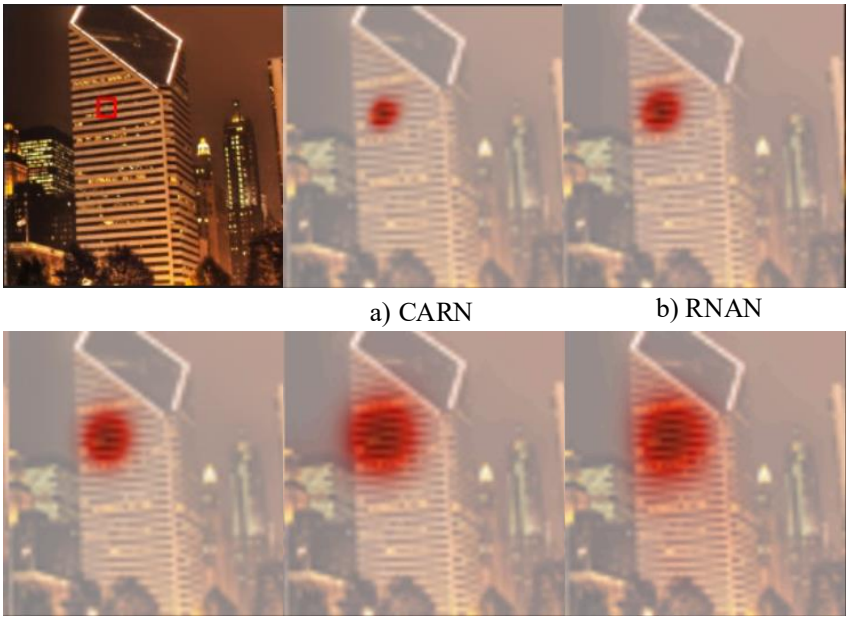


Figure 4.1: The compared neural networks with different core architectures. a) CARN [1]: Convolutional neural network. b) RNAN [114]: Non-local attentive network. c) RCAN [113]: Channel-wise attentive network. d) SAN [15]: Second-order attention-based network. e) Ours: Graph-based neural network.

To further enhance the quality of reconstructed high-resolution (HR) images, researchers have explored a diverse range of advanced learning-based techniques. Among these approaches, Convolutional Neural Networks (CNNs) [18, 36, 48, 87] and window-based attention mechanisms [110, 112] have become the two predominant paradigms in modern super-resolution (SR) methodologies. As depicted in Fig.4.1, CARN [1] serves as a typical representative of CNN-based SR networks. A key limitation of such CNN-based models is their inherently restricted focus areas, which inherently deprives them of the capability to capture extensive global contextual information. In contrast, the emergence of attention mechanisms is specifically aimed at addressing the intrinsic limitations of CNNs, as demonstrated by representative SR networks including RNAN [114], RCAN [113], and SAN [15]. By adaptively and selectively focusing on salient regions within the input low-resolution (LR) image, these attention-based models are able to more effectively model long-range spatial dependencies, which constitutes a

significant advantage over traditional CNN-based approaches. Yet this pivotal advantage is almost invariably coupled with a considerable surge in computational overhead. Notwithstanding their distinct merits, both CNN-based and attention-based paradigms share a fundamental inherent limitation: their information aggregation mechanisms are bounded to pre-defined local neighborhoods, thus impairing their adaptive capacity in addressing intricate image reconstruction tasks.

To tackle these inherent limitations, we design a novel Heterogeneous Subgraph Network (HSNet), which capitalizes on the structural characteristics of graphs to flexibly model image features without being constrained by pre-defined local neighborhoods. A critical bottleneck in applying full-graph construction to image processing tasks stems from the prohibitive computational overhead incurred by computing pairwise pixel relationships across the entire image domain. To alleviate this computational burden while preserving the representational advantages of graph-based modeling, we propose the Construct Subgraph Set Block (CSSB). Rather than constructing a single holistic graph for the entire image, the CSSB is designed to generate a diverse set of subgraphs. Conceptually, a subgraph refers to a subset of the global graph, comprising a curated set of its vertices and their associated edges. By selectively constructing and processing these subgraphs, our approach effectively retains the core structural properties of the global graph—such as topological connectivity and salient semantic paths—while enabling each subgraph to focus on and extract task-specific localized information. This strategy not only drastically reduces the computational complexity of graph construction but also markedly enhances the flexibility of feature extraction within the HSNet framework. Furthermore, building on the diverse set of subgraphs generated by the CSSB, we introduce the Subgraph Aggregation Block (SAB) to construct a robust and information-rich global heterogeneous graph representation. The SAB is specifically designed to distill and fuse the discriminative features embedded within individual subgraphs, yielding an aggregated graph where each node is endowed with a highly diverse feature representation. This enables the model to develop a comprehensive understanding of the underlying image data structure and inter-node relational patterns. By integrating complementary features from distinct subgraphs, the SAB augments the model's capacity to capture intricate spatial patterns and subtle

inter-dependencies, thereby substantially boosting the super-resolution performance of HSNet. As shown in Fig.4.1, our approach demonstrates a significantly broader receptive field in feature capture compared to state-of-the-art alternative methods.

Furthermore, to optimize this critical node selection process, we devise a strategic sampling mechanism that concurrently captures fine-grained local details and global structural features from various complementary views. Specifically, inspired by the mamba scanning strategy [121], we propose a dedicated Node Sampling Strategy (NSS), which adaptively selects salient node features that yield the most significant contributions to the learning process while effectively minimizing feature redundancy. By strategically discarding less informative node samples, the proposed NSS not only preserves critical structural and semantic features but also drastically reduces computational overhead; this in turn enhances the overall efficiency and precision of feature capture, thereby yielding substantial performance improvements for HSNet on the SISR task. The main contributions of this work are summarized as follows:

- We propose the Heterogeneous Subgraph Network (HSNet), a novel graph-based architecture that transcends the rigid local neighborhood constraints of CNNs and Transformer-based models by modeling image features via an ensemble of complementary subgraphs with rich structural expressiveness.
- We introduce the Construct Subgraph Set Block (CSSB), which generates an ensemble of complementary subgraphs—each dedicated to modeling distinct feature relationships—and propose the Node Sampling Strategy (NSS) to capture the most salient node features while substantially reducing computational overhead.
- We develop the Subgraph Aggregation Block (SAB), which intelligently fuses the distinct perspectives embedded within each subgraph to construct a holistic and discriminative feature representation, thereby enabling superior performance in image reconstruction tasks.
- The proposed HSNet achieves state-of-the-art (SOTA) performance across multiple standard super-resolution benchmarks. To validate the effectiveness and

generalizability of our approach, we conduct extensive ablation experiments and qualitative visualization analyses.

4.2 Our Method

For the effective application of graph structures to regular grid-structured data (e.g., images), it is imperative to convert image data into a graph representation, wherein individual pixels or image regions are mapped to graph nodes, and inter-node relationships—including spatial adjacency and feature similarity—are encoded as edges. Each node is equipped with discriminative features that capture task-relevant image information (e.g., color, texture, or spatial coordinates), thereby enhancing the graph's capacity to represent image content accurately. By capitalizing on the intrinsic flexibility of graph structures, we can explore the underlying relationships among irregularities in the input data, a key advantage that distinguishes graph-based methods from conventional window-based approaches. Graph structures inherently possess superior flexibility and can transcend the inherent constraints of fixed window partitioning to capture long-range contextual information. However, the construction of full-scale graphs typically incurs prohibitive computational overhead, which restricts their practical deployment in high-resolution image processing tasks. A straightforward yet efficient solution to this bottleneck is to reduce the number of nodes involved in graph computation, thereby mitigating the associated computational burden. To address this critical challenge, we propose a subgraph-based Heterogeneous Subgraph Network (HSNet).

In the subsequent sections, we first provide a comprehensive overview of the overall HSNet architecture. We then elaborate on the Construct Subgraph Set Block (CSSB), which consists of two core components: the Node Sampling Strategy (NSS) and the Subgraph Generation Block (SGB). Following this, we detail the design principles and operational mechanisms of the Subgraph Aggregation Block (SAB), and subsequently present a detailed exposition of the Graph Aggregation (GA) component embedded within the SAB.

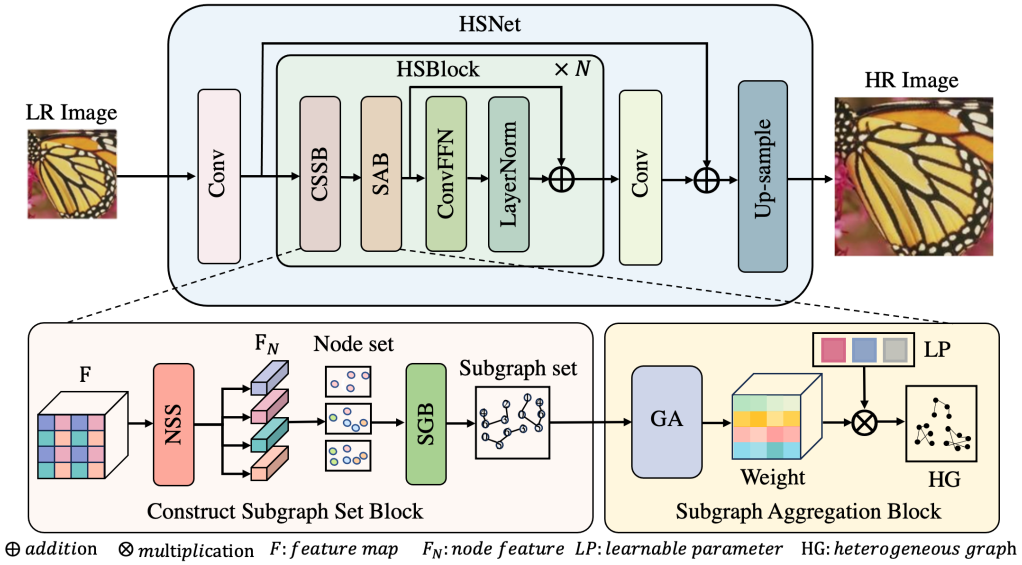


Figure 4.2: The architectural framework of HSNet is presented as follows: the Heterogeneous Subgraph Block (HSBlock) serves as its core module, which is composed of two essential submodules, namely the Construct Subgraph Set Block (CSSB) and the Subgraph Aggregation Block (SAB). In detail, the CSSB incorporates two functional components: the Node Sampling Strategy (NSS) is dedicated to node sampling and node set construction, and the generated node set is subsequently transmitted to the Subgraph Generation Block (SGB) to yield the subgraph set. For the SAB, it is responsible for the generation of heterogeneous graphs, and this module is constituted by the Graph Aggregation (GA) block and Learnable Parameters (LP).

4.2.1 Model Architecture

The overall architecture of HSNet adheres to the standard pipeline of super-resolution (SR) tasks and is decomposed into three sequential stages, as illustrated in Fig.4.2. First, a Shallow Feature Extraction stage—typically implemented via a single convolutional layer—transforms the input image into a low-dimensional feature map, laying a foundation for subsequent deep feature learning. Next is the Deep Feature Extraction phase, which is instantiated by our proposed Heterogeneous Subgraph Block (HSBlock). The final stage is Image Reconstruction, which maps the learned deep features back to the high-resolution (HR) image space to achieve resolution enhancement. As the core building block of the proposed HSNet, HSBlock is specifically designed to handle heterogeneous node and edge types, enabling the seamless integration of diverse information sources for comprehensive feature modeling. This block comprises

two pivotal modules: the Construct Subgraph Set Block (CSSB) and the Subgraph Aggregation Block (SAB). Specifically, the CSSB is tasked with generating a diverse ensemble of subgraphs by extracting task-relevant information according to adaptive criteria, which are dynamically determined by the Node Sampling Strategy (NSS) that employs distinct node selection mechanisms. Subsequently, the Subgraph Generation Block (SGB) constructs the set of subgraphs based on the sampled nodes and extracted information. Finally, the SAB aggregates this subgraph ensemble into a unified heterogeneous graph (HG), which encapsulates both local topological details and global contextual dependencies.

4.2.2 Construct Subgraph Set Block

CSSB initiates with the extraction of subgraphs from the large-scale graph via dedicated sampling strategies, where we leverage the Neighborhood Sampling Strategy (NSS) and Subgraph Generation Block (SGB) to construct diverse subgraph instances. Each subgraph not only encodes distinct graph subspaces but also adaptively selects target-node-relevant neighboring nodes, enabling efficient contextual information aggregation. By integrating heterogeneous sampling paradigms (i.e., local and global sampling), our model captures multi-scale structural and semantic information of the graph: local sampling prioritizes fine-grained relational patterns among proximal nodes, whereas global sampling excavates the holistic structural characteristics of the entire graph. This dual-scale sampling mechanism effectively enhances the model's capacity to model and interpret complex graph data with intricate topological structures.

4.2.2.1 Node Sampling Strategy

We propose the NSS, a selective sampling paradigm that identifies and preserves the most discriminative feature representations, thus enabling adaptive and flexible feature selection. Given a feature map $F \in R^{H \times W \times C}$, we perform feature selection with a

stride of 2, which partitions the entire feature space into four disjoint subspaces. Concretely, the spatial dimensions of height (H) and width (W) are each segmented into four distinct regions, which are formally defined as follows:

$$\begin{aligned} F_1 &= \left\{ F_{i,j,k,l} \in F \mid k \equiv 0, l \equiv 0 \pmod 2 \right\}, \\ F_2 &= \left\{ F_{j,i,k,l} \in F \mid k \equiv 0, l \equiv 1 \pmod 2 \right\}, \\ F_3 &= \left\{ F_{i,j,k,l} \in F \mid k \equiv 1, l \equiv 1 \pmod 2 \right\}, \\ F_4 &= \left\{ F_{j,i,k,l} \in F \mid k \equiv 1, l \equiv 1 \pmod 2 \right\}. \end{aligned} \tag{4.1}$$

In this equation, F_1 denotes the first subset, comprising elements $F_{i,j,k,l} \in F$ where $i, j \in \mathbb{Z}$ (i.e., i and j are integers), and both indices k and l are even, which is mathematically expressed as $k \equiv 0 \,(\mathrm{mod}\, 2)$ and $l \equiv 0 \,(\mathrm{mod}\, 2)$. The subset F_2 consists of the transposed elements $F_{j,i,k,l} \in F$, with k being even and l being odd (denoted by $l \equiv 1 \,(\mathrm{mod}\, 2)$).

For F_3, it contains elements $F_{i,j,k,l} \in F$ where k is even and l is odd. Meanwhile, F_4 includes the transposed elements $F_{j,i,k,l}$ in which both k and l are odd, represented as $k \equiv 1 \,(\mathrm{mod}\, 2)$ and $l \equiv 1 \,(\mathrm{mod}\, 2)$.

Following processing of the selected features, we yield $F_i \in R^{H/p,W/p,C}$, where p denotes the number of subsets. This operation enables four distinct sampling strategies for F, facilitating the capture of data features from diverse perspectives and thus enriching the feature representation. We then recombine these subsets to reconstruct the feature map $F_{concat} \in R^{H,W,C}$, as formulated below:

$$F_{concat} = [F_1, F_2, F_3, F_4]. \tag{4.2}$$

This reconstruction empowers the model to capture contextual features in a more comprehensive manner, thereby enhancing its semantic understanding of the input.

4.2.2.2 Subgraph Generation Block

The SGB's role is to create subgraphs, as illustrated in the Fig.4.3.

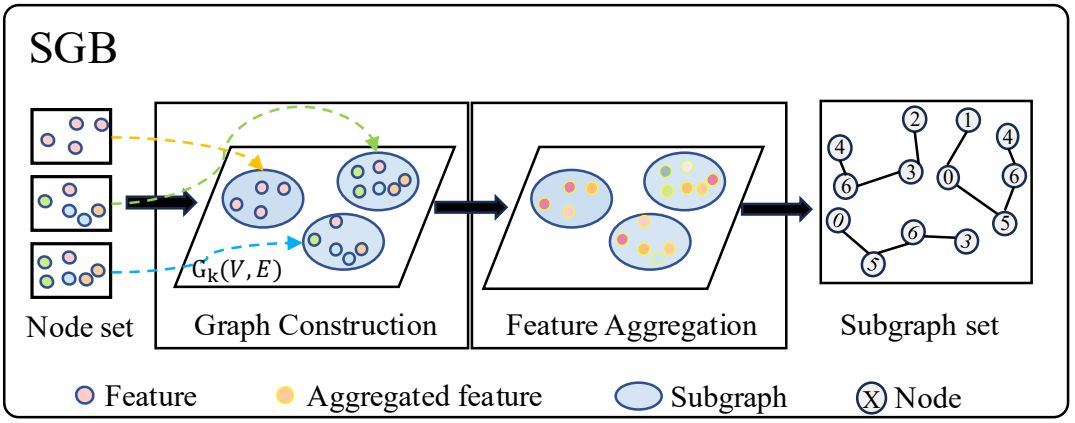


Figure 4.3: Architecture of the Subgraph Generation Block (SGB). It consists of two core components: Graph Construction and Feature Aggregation. It first takes as input a node set containing multiple groups of nodes with their respective features. In the Graph Construction component, nodes are organized into an initial different subgraphs based on feature correlation. Then, in the Feature Aggregation component, node features within each subgraph are fused and computed to generate representative aggregated features. Finally, a set of structured subgraphs is output as the subgraph set, providing feature-enhanced basic units for subsequent Subgraph Aggregation Block (SAB).

The SGB instantiates each subgraph through the deliberate selection of target nodes and associated edges. A growing body of SR literature has further capitalized on the inherent self-similarity of images as prior knowledge to enhance reconstruction performance. For instance, Zhou et al. [117] presented the cross-scale internal graph neural network (IGNN), a GNN-based framework designed to capture and model textural feature correlations across disparate scale levels. This work yields a critical insight from the aggregation of self-similar cross-scale feature patches: integrating self-similar feature information across the multi-scale space facilitates a holistic understanding of image feature distributions and enables effective extraction of deep, hierarchical image features.

Graph Construction. To capture cross-scale feature representations, we first adopt two distinct sampling schemes to extract feature patches across different scales. Specifically, grid sampling is applied to generate patches at fixed spatial intervals for each scale, which guarantees uniform spatial coverage of the feature space at every scale and thus facilitates the capture of fine-grained structural details. We then leverage the Structural Similarity Index Metric (SSIM) to quantify the similarity between image

patches; additionally, for each individual patch, we calculate its Euclidean distance to all other patches and retain only the K edges with the minimum distance values. This entire process yields a graph $G(V,E)$, where the vertex set V corresponds to all extracted feature patches and the edge set E denotes the pairwise connections between these patches. The edge weights are defined as the quantitative similarity measures between connected nodes, which enables a fine-grained characterization of the interrelationships among diverse feature patches. Accordingly, semantically similar feature patches and their intrinsic connections are modeled as graph nodes and edges, respectively. This graph-based structural modeling not only enables effective aggregation of correlated feature patches across scales but also preserves the inherent spatial topological relationships among all graph nodes.

Feature Aggregation. By aggregating semantically similar features across the graph, we integrate the characteristic information of adjacent nodes, thereby enabling a holistic contextual understanding of the feature space. This aggregation operation is formally formulated as follows:

$$y_i = 1/C(x) \sum f(x_i, x_j) g(x_j), \tag{4.3}$$

where x_i and x_j denote the image patches located at positions i and j, respectively. Here, $g(x_j)$ is designed to encapsulate the feature representation of patch x_j, whereas $f(x_i, x_j)$ serves to compute the aggregation weights assigned to $g(x_j)$ during the feature fusion process. To ensure the reliability and effectiveness of the aggregation mechanism, we introduce a normalization factor $C(x)$, which is defined as the sum of aggregation weights over all relevant patches in the image. This normalization step is of paramount importance, as it ensures that the aggregation weights are properly scaled and proportional to one another—thereby facilitating the accurate and effective integration of feature information from semantically similar patches.

4.2.3 Subgraph Aggregation Block

To address the need for comprehensive feature fusion across heterogeneous subgraph encodings, we introduce the Subgraph Aggregation Block (SAB), which adaptively fuses multi-perspective graph information via weighted aggregation to learn holistic feature representations and explicitly model the intricate inter-dependencies within the data. Concretely, upon constructing the subgraph set, we instantiate a Graph Aggregation (GA) module built on a multi-head attention mechanism to capture the intrinsic structural and semantic correlations across subgraphs and refine their feature encodings. Each attention head operates independently to learn task-adaptive attention weights and discriminative feature representations for distinct subgraphs, thereby enabling the simultaneous modeling of diverse relational patterns and feature modalities. This design inherently boosts the model's expressive capacity: individual heads specialize in encoding local fine-grained features (e.g., topological correlations among adjacent nodes), while others capture global structural features (e.g., the holistic graph topology and long-range node dependencies). Such hierarchical and multi-modal feature learning effectively enhances the model's overall representational power and task performance.

As illustrated in Fig.4.4, the attention module is devised to thoroughly mine the intrinsic interrelations across subgraphs and learn discriminative feature encodings for them. Realized via parallelized multi-head attention, this module empowers the model to conduct multi-perspective data analysis. Each attention head independently computes attention weights and derives tailored feature representations, with different heads specializing in distinct relational facets—for instance, local topological correlations among adjacent nodes or the global structural characteristics of the entire graph. This inherent diversity enables the capture of a rich spectrum of relational patterns and feature modalities, thereby substantially boosting the expressive capacity of the model.

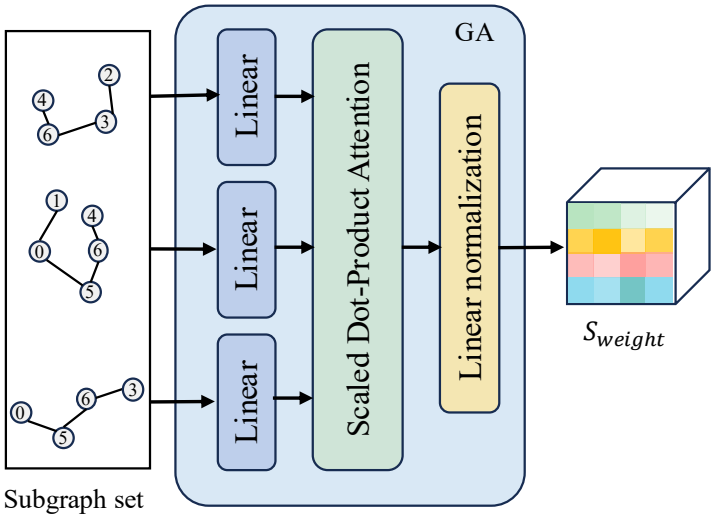


Figure 4.4: The workflow of the Graph Aggregation (GA) block: It takes a subgraph set as input, first passes each subgraph's features through separate linear layers for dimension mapping and encoding, then employs scaled dot-product attention to compute the correlation weights between subgraphs for adaptive aggregation, and finally converts the aggregated result into a structured weight tensor S_{weight} via linear normalization, thereby providing a feature representation that integrates inter-subgraph correlation information for subsequent heterogeneous graph processing.

4.2.3.1 Graph Aggregation

The input to the module comprises a collection of feature representation sets, which are typically instantiated as dense feature vectors encoding diverse data attributes (e.g., nodal characteristics in graph-structured data). We denote by $S^{(k)}$ the feature vector set corresponding to the k^{th} subgraph, formulated as:

$$S^{(k)} = \{s_1^{(k)}, s_2^{(k)}, \ldots, s_n^{(k)}\}, \tag{4.4}$$

where n is the number of nodes in the subgraph.

Each input feature set is then projected via linear layers to yield three distinct latent representations, namely queries (Q), keys (K), and values (V), respectively.

$$Q^{(k)} = W_Q^{(k)} S^{(k)}, \quad K^{(k)} = W_K^{(k)} S^{(k)}, \quad V^{(k)} = W_V^{(k)} S^{(k)}, \tag{4.5}$$

where $W_Q^{(k)}$, $W_K^{(k)}$,$W_V^{(k)}$ are learnable weight matrices. Then, we perform attention operations on different vectors, denoting them as $A^{(k)} = Attention(W_Q^{(k)}, W_K^{(k)}, W_V^{(k)})$.

The feature representations are refined via attention-weighted aggregation of feature values, as formulated in the following equation:

$$S'^{(k)} = A^{(k)} V^{(k)}. \tag{4.6}$$

Weights derived from the subset graph set following GA application are formulated as:

$$S_{weight} = \{S'^{(1)}, S'^{(2)}, \ldots, S'^{(J)}\}, \tag{4.7}$$

where J denotes the total number of subgraphs in the subset.

4.2.3.2 Combination

We further integrate representations of distinct subgraphs in a principled manner via learnable dynamic parameters, which are iteratively refined according to training feedback to endow the model with strong adaptive ability. Guided by these parameters, the model executes weighted fusion of subgraphs in an importance-aware fashion that aligns with their inherent relational characteristics, guaranteeing that more critical relational patterns impose a more pronounced effect on the ultimate feature learning outcome. This adaptive fusion strategy not only elevates the model's predictive performance but also enhances its interpretability by quantifying the contribution of each subgraph to the final representation. With the above architectural components, we effectively strengthen the node-wise representational power of the heterogeneous graph. The outputs from GA are combined using dynamic Learnable Parameter (LP) α_n.

$$HG = \alpha_1 S'^{(1)} + \alpha_2 S'^{(2)} + \ldots + \alpha_J S'^{(J)} \tag{4.8}$$

where J represents the total number of learnable parameters.

4.3 Experiment

4.3.1 Experimental details

Datasets. To guarantee a fair and rigorous evaluation, we adopt the standard training framework prevalent in recent super-resolution (SR) research. For the training of our lightweight SR model, we utilize the DIV2K dataset as the primary training source. We conduct an extensive performance comparison between our proposed model and a suite of representative SR baseline methods, with all evaluations performed on the widely adopted benchmark datasets: Set5 [7], Set14 [105], BSDS100 [62], Urban100 [31], and Manga109 [63].

Training details. For model training, input data is processed as 64×64 **cropped patches**; to enrich the training corpus and improve generalization, we employ standard data augmentation strategies including random horizontal/vertical flipping and discrete rotational augmentation at $0°, 90°, 180°$ and $270°$. The model is trained for a total of $500,000$ iterations with the **Adam optimizer** configured as $\beta_1 = 0.9$ and $\beta_2 = 0.99$, and an initial learning rate of 2^{-4}. A **multi-step learning rate decay schedule** is adopted, where the learning rate is halved at the iteration checkpoints of $250,000$, $400,000$, $450,000$ and $475,000$. All training is conducted on 4 NVIDIA RTX 4090 GPUs with a per-GPU batch size of 8 (a global batch size of 32), with the input patch size fixed at 64×64 throughout the training process.

Training Loss. The standard L_1 loss between the super-resolution prediction I_{SR} and the ground truth high-resolution image I_{HR} is utilized for training our models.

Evaluation metrics. We evaluate the performance of our proposed model against a comprehensive set of lightweight super-resolution (SR) baselines, with detailed comparative results presented in Tab. 4.1. The selected baselines cover four representative categories of lightweight SR methods: (1) patch-graph-based methods, including IPG-tiny [85] and GMN [22]; (2) CNN-based models, such as CRAN [1], IMDN [33], LAPAR-A [45], SAFMN [79], and SeemoRe [102]; (3) Transformer-based approaches, namely SwinIR-light [46], SRFormer-light [118], FIWHN [44], and CATANet [51]; and

(4) Mamba-framework-based methods, specifically MambaIR-light [26]. All SR models are evaluated at the common scaling factors of $\times 2$, $\times 3$, and $\times 4$, using two standard quantitative metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

4.3.2 Comparisons of Super-resolution Methods

4.3.2.1 Quantitative Comparison

We perform a comprehensive quantitative comparison between our proposed method and a diverse array of established baselines, with detailed results presented in Tab. 4.1. Notably, our method demonstrates a consistent and significant performance superiority over all competing models, which attests to its robustness and effectiveness in addressing super-resolution tasks. Specifically, on the Set14 benchmark at the ×4 scaling factor, HSNet achieves a PSNR of 28.93 dB—outperforming the current state-of-the-art (SOTA) by more than 0.03 dB under standard training protocols—while also delivering a notable improvement in SSIM with a score of 0.7886, representing a 0.0006 gain compared to the previous top-performing model. This measurable performance enhancement validates the efficacy of our proposed approach in boosting both the objective and perceptual quality of super-resolved images.

As illustrated in Tab. 4.1, our proposed method consistently achieves remarkable PSNR performance across diverse benchmark datasets. Furthermore, we compare the model parameters and computational complexity (measured in FLOPs) of state-of-the-art lightweight SR models on the Urban100 (×4) and Manga109 (×4) benchmarks. As shown in Tab. 4.2, our method outperforms previous Transformer-based and graph-based lightweight approaches by not only reducing the number of model parameters but also lowering the computational overhead significantly.

Table 4.1: Performance comparison of HSNet against recent super-resolution (SR) approaches, with the top and second-top performing results highlighted in red and blue.

Method	Scale	Params	SET5		SET14		B100		Urban100		Manga109	
			PSNR	*SSIM*	*PSNR*	*SSIM*	*PSNR*	*SSIM*	*PSNR*	*SSIM*	*PSNR*	*SSIM*
CARN (2018) [1]	×2	1592K	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256	38.36	0.9765
IMDN (2019) [33]	×2	694K	38.00	0.9605	33.63	0.9177	32.19	0.8996	32.17	0.9283	38.88	0.9774
LAPAR-A(2020) [45]	×2	548K	38.01	0.9605	33.62	0.9183	32.19	0.8999	32.10	0.9283	38.67	0.9772
SwinIR-light(2021) [46]	×2	878K	38.14	0.9611	33.86	0.9206	32.31	0.9012	32.76	0.9340	39.12	0.9783
ELAN-light (2022) [111]	×2	582K	38.17	0.9611	33.94	0.9207	32.30	0.9012	32.76	0.9340	39.11	0.9782
SRFormer-light (2023) [118]	×2	853K	38.23	0.9613	33.94	0.9209	32.36	0.9019	32.91	0.9353	39.28	0.9785
SAFMN (2023) [79]	×2	228K	38.00	0.9605	33.54	0.9177	32.16	0.8995	31.84	0.9256	38.71	0.9771
MambaIR-light (2024) [26]	×2	530K	38.15	0.9610	33.84	0.9207	32.31	0.9013	32.86	0.9343	39.35	0.9786
FIWHN (2024) [44]	×2	705K	38.16	0.9613	33.73	0.9194	32.27	0.9007	32.75	0.9337	39.07	0.9782
SeemoRe-T(2024) [102]	×2	220K	38.06	0.9608	33.65	0.9186	32.23	0.9004	32.22	0.9286	39.01	0.9777
IPG-tiny (2024) [85]	×2	872K	38.27	0.9616	34.24	0.9236	32.35	0.9018	33.04	0.9359	39.31	0.9786
CATANet (2025) [51]	×2	477K	38.28	0.9617	33.99	0.9217	32.37	0.9023	33.09	0.9372	39.37	0.9784
GMN(2025) [22]	×2	1110k	38.16	0.9611	33.88	0.9201	32.30	0.9012	32.55	0.9325	39.17	0.9781
HSNet	×2	870K	38.29	0.9620	34.22	0.9241	32.38	0.9031	33.27	0.9391	39.38	0.9802
CARN (2018) [1]	×3	1592K	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493	33.43	0.9427
IMDN (2019) [33]	×3	703K	34.36	0.9270	30.32	0.8417	29.09	0.8046	28.17	0.8519	33.61	0.9445
LAPAR-A(2020) [45]	×3	594K	34.36	0.9267	30.34	0.8421	29.11	0.8054	28.15	0.8523	33.51	0.9441
SwinIR-light(2021) [46]	×3	886K	34.62	0.9289	30.54	0.8463	29.20	0.8082	28.66	0.8624	33.98	0.9478
ELAN-light (2022) [111]	×3	590K	34.64	0.9288	30.55	0.8463	29.21	0.8081	28.69	0.8624	34.00	0.9478
SwinIR-light(2021) [46]	×3	886K	34.62	0.9289	30.54	0.8463	29.20	0.8082	28.66	0.8624	33.98	0.9478
SRFormer-light (2023) [118]	×3	861K	34.67	0.9296	30.57	0.8469	29.26	0.8099	28.81	0.8655	34.19	0.9489
SAFMN (2023) [79]	×3	233K	34.34	0.9267	30.33	0.8418	29.08	0.8048	27.95	0.8474	33.52	0.9437
MambaIR-light (2024) [26]	×3	538K	34.62	0.9286	30.54	0.8459	29.23	0.8083	28.73	0.8635	34.26	0.9482
FIWHN (2024) [44]	×3	713K	34.50	0.9283	30.50	0.8451	29.19	0.8077	28.62	0.8607	33.97	0.9472
SeemoRe-T(2024) [102]	×3	225K	34.46	0.9276	30.44	0.8445	29.15	0.8063	28.27	0.8538	33.92	0.9460
IPG-tiny (2024) [85]	×3	878K	34.64	0.9292	30.61	0.8470	29.26	0.8097	28.93	0.8666	34.30	0.9493
CATANet (2025) [51]	×3	550K	34.75	0.9300	30.67	0.8481	29.28	0.8101	29.04	0.8689	34.40	0.9499
GMN(2025) [22]	×3	1120K	34.60	0.9291	30.37	0.8454	29.23	0.8091	28.56	0.8609	34.14	0.9482
HSNet	×3	875K	34.76	0.9301	30.69	0.8490	29.30	0.8102	29.07	0.8691	34.42	0.9512
CARN (2018) [1]	×4	1592K	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837	30.42	0.9070
IMDN (2019) [33]	×4	715K	32.21	0.8948	28.58	0.7811	27.56	0.7353	26.04	0.7838	30.45	0.9075
LAPAR-A(2020) [45]	×4	659K	32.15	0.8944	28.61	0.7818	27.61	0.7366	26.14	0.7871	30.42	0.9074
SwinIR-light(2021) [46]	×4	897K	32.44	0.8976	28.77	0.7858	27.69	0.7406	26.47	0.7980	30.92	0.9151
ELAN-light (2022) [111]	×4	601K	32.43	0.8975	28.78	0.7858	27.69	0.7406	26.54	0.7982	30.92	0.9150
SwinIR-light(2021) [46]	×4	897K	32.44	0.8976	28.77	0.7858	27.69	0.7406	26.47	0.7980	30.92	0.9151
SRFormer-light (2023) [118]	×4	873K	32.51	0.8988	28.82	0.7872	27.73	0.7422	26.67	0.8032	31.17	0.9165
SAFMN (2023) [79]	×4	240K	32.18	0.8948	28.60	0.7813	27.58	0.7359	25.97	0.7809	30.43	0.9063
MambaIR-light (2024) [26]	×4	549K	32.41	0.8974	28.76	0.7849	27.70	0.7400	26.53	0.7983	31.05	0.9144
FIWHN (2024) [44]	×4	725K	32.30	0.8967	28.76	0.7849	27.68	0.7400	26.57	0.7989	30.93	0.9131
SeemoRe-T(2024) [102]	×4	232K	32.31	0.8965	28.72	0.7840	27.65	0.7384	26.23	0.7883	30.82	0.9107
IPG-tiny (2024) [85]	×4	887K	32.51	0.8987	28.85	0.7873	27.73	0.7418	26.78	0.8050	31.22	0.9176
CATANet (2025) [51]	×4	535K	32.58	0.8998	28.90	0.7880	27.75	0.7427	26.87	0.8081	31.31	0.9183
GMN(2025) [22]	×4	1134K	32.40	0.8979	28.66	0.7868	27.70	0.7416	26.40	0.7963	30.98	0.9143
HSNet	×4	882K	32.60	0.9012	28.93	0.7886	27.78	0.7437	26.89	0.8089	31.33	0.9181

4.3.2.2 Qualitative Comparison

In Fig.4.5, we showcase a series of visual examples that highlight the performance of various super-resolution methods.

By effectively aggregating multi-subgraph feature representations, HSNet is capable

Table 4.2: Comparison of Parameters and FLOPs for lightweight super-resolution models on Urban100 (×4) and Manga109 (×4) datasets: the optimal results are highlighted in **bold**, the suboptimal ones are marked with __; 'C' stands for CNN-based methods, 'T' for Transformer-based methods, and 'G' for graph-based methods.

Method	Type	Scale	Param	Flops	Urban100	Manga109
CARN [1]	C	×4	1592K	90.9G	26.07	30.47
IMDN [33]	C	×4	**715K**	**40.9G**	26.04	30.45
SwinIR-light [46]	T	×4	930K	63.6G	26.47	30.92
SRFormer-light [118]	T	×4	873K	62.8G	26.67	31.17
IPG-tiny [85]	G	×4	887K	61.3G	26.78	31.22
Ours	G	×4	882K	60.1G	**26.89**	**31.33**

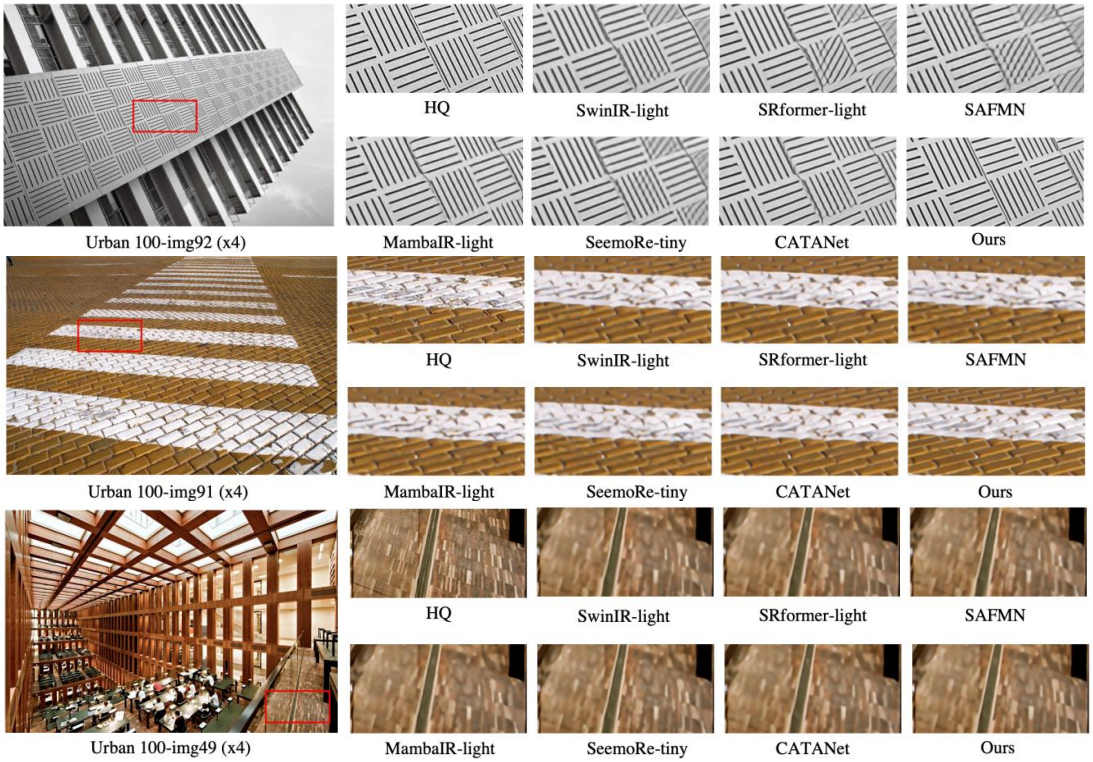


Figure 4.5: Visualized ×4 super-resolution results of different methods on the Urban100 dataset samples img_92, img_91 and img_49.

of capturing rich multi-scale contextual features in images. As visualized in Fig.4.5, our model achieves high-fidelity restoration of simple stripe structures for Urban100-img92, and this multi-scale feature modeling paradigm enables the simultaneous modeling of fine-grained local textures and global structural information, thus realizing holistic and high-quality image restoration. Moreover, for Urban100-img91 with repetitive texture patterns, our model also delivers robust and accurate restoration performance. In contrast, existing lightweight SR methods exhibit notable limitations in

recovering complex fine-grained details: taking Urban100-img49 as an example, these approaches often suffer from the loss of delicate textures and intricate structural details in their restoration results. This deficiency stems from a key limitation of current methods, which tend to over-rely on global feature encoding while neglecting the preservation and reconstruction of local fine details—an issue that merits further improvement in future research, such as the integration of advanced local feature extraction and refinement modules to enhance fine-detail modeling capabilities. Overall, our proposed method demonstrates superior performance in SR tasks, exhibiting an outstanding ability to restore sharp, clean edge structures and suppress visual artifacts significantly. This compelling performance gain is largely attributed to the innovative architectural design of HSNet, which is specifically tailored to capture and reconstruct intricate texture details with high precision.

4.3.3 Ablation Study

In this section, we conduct extensive ablation studies to comprehensively evaluate the efficacy of each key component in the proposed HSNet and elucidate its respective contribution to the overall performance. For the sake of fair and consistent comparison, all ablation experiments are performed on the ×4 version of HSNet under identical training protocols and experimental settings.

Ablation Study on the Subgraph Set Construction Block and Subgraph Aggregation Block. We investigate the performance impact of the Heterogeneous Subgraph Block (HSBlock) on the network, with a specific focus on its two core components: the Subgraph Set Construction Block (CSSB) and Subgraph Aggregation Block (SAB).

Table 4.3: Ablation experiments on the CSSB and SAB; PSNR is computed for the ×4 super-resolution task.

CSSB	SAB	Se5	Set14	B100	Urban100	Manga109
✓	✗	30.87	26.72	24.98	24.99	30.01
✗	✓	31.26	27.64	26.83	25.64	30.59
✓	✓	**32.60**	**28.93**	**27.78**	**26.89**	**31.33**

We quantify the performance of each component by evaluating the PSNR metric at the ×4 scaling factor across four benchmark datasets (Set14, B100, Urban100, Manga109), with detailed results presented in Tab. 4.3. **Impact of CSSB**: Experimental results verify that the incorporation of CSSB yields consistent and notable PSNR gains across all evaluated datasets, which corroborates that the CSSB's capability to adaptively extract discriminative subgraph features plays a pivotal role in elevating the quality of super-resolved images. **Impact of SAB**: When ablated individually, the SAB also brings about moderate PSNR improvements relative to the baseline model, validating its effectiveness in adaptive aggregation of multi-subgraph feature information. Nevertheless, its performance gain is less pronounced compared with that of the CSSB. **Synergistic Effect of Combined Components**: The model integrating both CSSB and SAB (dubbed HSBlock) achieves consistent and superior performance over either component alone across all datasets, which explicitly demonstrates the strong synergistic effect between the subgraph extraction mechanism and multi-subgraph aggregation strategy. This performance boost underscores the necessity of a holistic architectural design that jointly models and fuses diverse multi-scale subgraph information sources for super-resolution tasks.

Effects of Node Sampling Strategy. We evaluate the impact of the Node Selection Strategy (NSS) component on the performance of our proposed HSNet. As a core efficiency-enhancing module, NSS is specially designed to strengthen the network's capacity for capturing discriminative salient features while reducing computational overhead. To rigorously validate its efficacy, we conduct comparative experiments on HSNet with and without the NSS component across four standard benchmark datasets: Set14, B100, Urban100 and Manga109. The PSNR metrics, computed at the ×4 scaling factor, are systematically summarized in Tab. 4.4, enabling a quantitative assessment of NSS's contribution to both performance and efficiency.

Table 4.4: Ablation experiments on the Node Sampling Strategy (NSS); PSNR is computed for the ×4 super-resolution task.

Method	Se5	Set14	B100	Urban100	Manga109
wo NSS	32.42	28.71	27.61	26.72	31.24
w NSS	**32.60**	**28.93**	**27.78**	**26.89**	**31.33**

The experimental results demonstrate that integrating the Node Sampling Strategy (NSS) yields consistent and significant PSNR improvements across all evaluated benchmark datasets. For instance, on the Set14 dataset, the PSNR value increases by 0.22 dB—rising from 28.71 dB in the HSNet variant without NSS to 28.93 dB in the model equipped with NSS. These measurable performance gains corroborate that NSS plays a pivotal role in enhancing the network's capability for discriminative feature extraction and effective feature representation, which in turn contributes to substantial improvements in the objective quality of super-resolved image restoration.

Effects of Subgraph Generation Block. We investigate the impact of the Subgraph Generation Block (SGB) on the performance of our proposed HSNet architecture. To rigorously assess the efficacy of the SGB component, we conduct comparative experiments between HSNet variants with and without the SGB, evaluated across a suite of standard benchmark datasets.

Table 4.5: We perform an ablation study on the Subgraph Generation Block (SGB), where PSNR is evaluated with a ×4 upscaling factor.

Method	Se5	Set14	B100	Urban100	Manga109
wo SGB	31.74	27.23	26.32	25.64	30.81
w SGB	**32.60**	**28.93**	**27.78**	**26.89**	**31.33**

As presented in Tab. 4.5, the experimental results clearly demonstrate that integrating the Subgraph Generation Block (SGB) yields substantial and consistent PSNR improvements across all evaluated benchmark datasets. For instance, on the Set14 dataset, the PSNR metric increases by 1.70 dB—rising from 27.23 dB in the HSNet variant without SGB to 28.93 dB in the model incorporating SGB. This ablation study underscores the pivotal role of the SGB component in enhancing the overall performance of HSNet, confirming its critical contribution to advancing the model's capability in handling super-resolution tasks.

Effects of Graph Aggregation block. We investigate the impact of diverse graph aggregation techniques on the performance of our proposed HSNet, with a specific focus on their efficacy in fusing multi-subgraph feature representations. To this end, we

conduct comparative experiments across three distinct aggregation operations commonly adopted in graph-based learning: additive aggregation (add), concatenation-based aggregation (concat), and a custom-designed aggregation strategy (denoted as "aggregation" for brevity).

Table 4.6: Ablation experiments on the Graph Aggregation (GA) block; PSNR is computed for the ×4 super-resolution task.

Method	Se5	Set14	B100	Urban100	Manga109
add	31.23	27.81	26.59	25.71	30.81
concat	32.03	28.28	27.14	26.20	30.72
aggregation	**32.60**	**28.93**	**27.78**	**26.89**	**31.33**

As presented in Tab. 4.6, the experimental results clearly demonstrate that our proposed graph aggregation (GA) method outperforms both additive (add) and concatenation (concat) aggregation techniques consistently across all benchmark datasets. Quantitatively, the GA method achieves PSNR values ranging from 26.89 dB to 32.60 dB across the evaluated datasets, whereas the additive aggregation method yields PSNR scores in the range of 25.71 dB to 31.23 dB, and the concatenation approach achieves PSNR values between 26.30 dB and 32.03 dB. Collectively, these results confirm that the GA technique delivers measurable performance gains over the other two aggregation strategies. This ablation study underscores the critical importance of selecting an optimal graph aggregation mechanism for HSNet, with the proposed GA operation exhibiting superior capability in effectively fusing multi-subgraph feature representations to enhance super-resolution performance.

4.4 Summary

We present HSNet, a novel Heterogeneous Subgraph Network that amalgamates complementary feature information across multiple subgraphs, enabling concurrent capture of fine-grained local topological details and global contextual relationships in images. At the core of HSNet resides the Heterogeneous Subgraph Block (HSBlock), which encompasses two pivotal components: the Subgraph Set Construction Block (CSSB), which elicits diverse discriminative subgraph views from the input feature

space, and the Subgraph Aggregation Block (SAB), which adaptively aggregates these multi-subgraph views to yield enhanced, holistic feature representations. To strike an optimal balance between the model's representational capacity and inference efficiency, we further introduce a Node Sampling Strategy (NSS) that strategically prunes non-essential sampling nodes while retaining all salient structural and texture information. Finally, a dedicated Graph Aggregation (GA) block refines and fuses the enhanced multi-subgraph features, culminating in substantial performance gains for super-resolution tasks. For future research directions, advancing node sampling techniques holds great potential to further improve salient feature retention while reducing computational complexity. In particular, exploring adaptive sampling methods that dynamically adjust the sampling strategy based on the inherent features of graph structures (e.g., node density, topological complexity) could yield notable performance and efficiency gains for graph-based super-resolution models.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

This thesis investigates the application of Mamba and graph networks for image restoration tasks. It highlights the strengths of these methods in modeling complex mappings from degraded images to clean outputs while addressing persistent challenges in the field.

The proposed employs a three-step image restoration approach that significantly enhances the restoration process. First, the Feature Aggregation (FA) Block integrates initial features to establish a robust foundation for reconstructing the input form. Following this, the Point-to-Point Restoration (PPR) Block utilizes a shallow-to-deep structure to progressively merge information from the thumbnail, generating coarse images that improve restoration quality. Finally, the Pyramid Fusion (PF) Block refines the image, leveraging complementary techniques to deliver high-fidelity results. This innovative three-step method not only combines these complementary techniques to tackle the complexities of image restoration but also demonstrates its effectiveness through extensive experimental results and ablation studies, paving the way for promising future research in this challenging domain.

The Heterogeneous Subgraph Network (HSNet) is designed to effectively integrate information across diverse subgraphs, enhancing the overall feature representation. At

its core, the HSNet features the Heterogeneous Sub- graph Block (HSBlock), which aims to generate complementary graph structures. The HSBlock consists of two main components: the Construct Subgraph Set Block (CSSB), responsible for producing a set of subgraphs, and the Subgraph Aggregation Block (SAB), which integrates these subgraphs to create richer feature representations. This architecture allows HSNet to capture multiple features, ensuring a comprehensive understanding of both local topological details and global contextual relationships. The collaborative operation of its components significantly improves performance, paving the way for future exploration in complex data structures.

5.2 Future Work

The proposed methods in this thesis significantly enhance the performance of image restoration. Future work can expand upon these advancements from several key perspectives:

1. A Generative Compensation Framework for Thumbnail-Driven Processing Under Corrupted or Missing Thumbnail Conditions

Problem: A key limitation of the proposed thumbnail-guided framework stems from the absence of standardized specifications for thumbnails in the JPEG standard, where no unified criteria are defined for thumbnail quality, spatial resolution, or even mandatory inclusion. Consequently, the performance of our method is prone to significant degradation when confronted with extremely low-quality, heavily compressed thumbnails or in scenarios where thumbnails are entirely unavailable.

Opportunity: State-of-the-art generative models exhibit substantial potential to reconstruct useful semantic or structural information from corrupted, incomplete, or low-fidelity image data.

Focus: Leveraging the powerful content generation and restoration capabilities of generative models (e.g., diffusion models, GANs), we can potentially compensate for the

absence or defects of thumbnails, thereby enhancing the robustness and applicability of the thumbnail-guided framework in practical scenarios.

2. Effective Network for Severe Bit Error Images

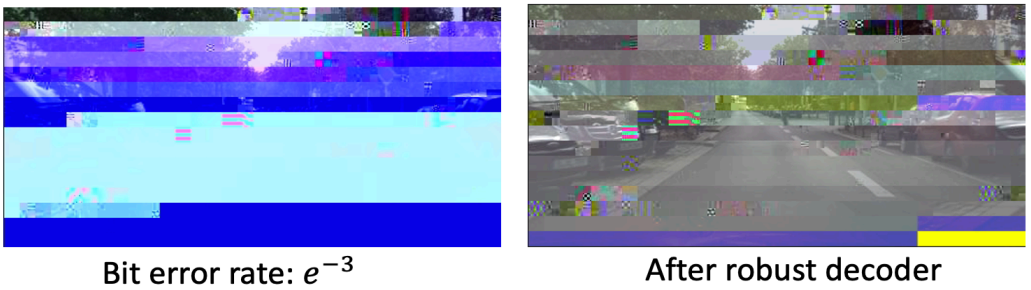


Figure 5.1: Under an error rate of 10^{-3}, robust decoder can not decode all pixel.

Problem: We observed that as the bit error rate increases, as shown in Fig.5.1, the images exhibit significant mosaic artifacts, severely obscuring or damaging the content. In this scenario, the spatial relationships between pixels and the RGB information are compromised, making it ineffective to use Anet for image restoration.

Opportunity: The thumbnail information remains intact during transmission, allowing us to leverage it directly for image reconstruction, achieving a super-resolution effect.

Focus: Further investigation is needed to extend our Mamba-based Network for different degrees of damage restoration.

3. Efficient SISR Network for High-resolution Images

Current Approach: Processing high-resolution images typically involves dividing them into smaller patches, which are restored individually and then stitched together. This patch-based method often ignores inter-patch dependencies, leading to suboptimal results.

Opportunity: The HSNet's ability to efficiently model long-range dependencies presents a promising avenue for directly addressing the challenges associated with 4K image restoration. The resolution of 4K is shown in Fig.5.2.

Focus: Further investigation is needed to extend HSNet for seamless 4K restoration, ensuring that interdependencies between patches are effectively utilized.

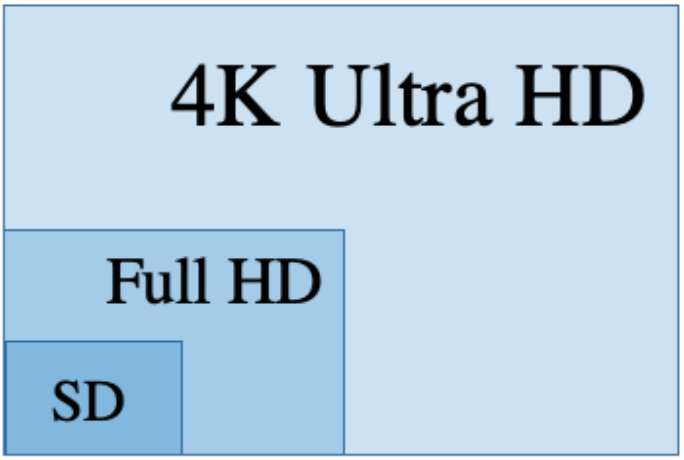


Figure 5.2: Standard Definition (SD): This is the lowest resolution, typically around 720x480 pixels. High Definition (HD): HD encompasses both 720p and 1080p resolutions, with 1080p also referred to as Full HD. Ultra High Definition (4K/UHD): 4K offers a resolution of 3840x2160 pixels, providing four times the pixel count of Full HD.

The proposed future research directions aim to build on the foundation established in this thesis, addressing practical challenges and further enhancing the capabilities of image restoration methods. Each direction offers valuable opportunities for exploration and innovation, contributing to advancements in the field.

Reference

[1] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. 2018. Fast, accurate, and lightweight super-resolution with cascading residual network. In Proceedings of the European conference on computer vision (ECCV). 252–268.

[2] Harry C Andrews. 2012. Digital image restoration: A survey. Computer 7, 5 (2012), 36–45.

[3] Saeed Anwar, Salman Khan, and Nick Barnes. 2020. A deep journey into super-resolution: A survey. ACM computing surveys (CSUR) 53, 3 (2020), 1–34.

[4] Kaoru Arakawa. 1996. Median filter based on fuzzy rules and its application to image restoration. Fuzzy sets and systems 77, 1 (1996), 3–13.

[5] Haoran Bai, Jinshan Pan, Xinguang Xiang, and Jinhui Tang. 2022. Self-guided image dehazing using progressive feature fusion. IEEE Transactions on Image Processing 31 (2022), 1217–1229.

[6] Mark R Banham and Aggelos K Katsaggelos. 2002. Digital image restoration. IEEE signal processing magazine 14, 2 (2002), 24–41.

[7] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. 2012. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. (2012).

[8] Qing Cai, Yiming Qian, Jinxing Li, Jun Lyu, Yee-Hong Yang, Feng Wu, and David Zhang. 2023. HIPA: Hierarchical patch transformer for single image super resolution. IEEE Transactions on Image Processing 32 (2023), 3226–3237.

[9] Chenjie Cao, Qiaole Dong, and Yanwei Fu. 2022. Learning prior feature and attention enhanced image inpainting. In European conference on computer vision. Springer, 306–322.

[10] Keyan Chen, Bowen Chen, Chenyang Liu, Wenyuan Li, Zhengxia Zou, and Zhenwei Shi. 2024. Rsmamba: Remote sensing image classification with state space model. IEEE Geoscience and Remote Sensing Letters (2024).

[11] Zixuan Chen, Zewei He, and Zhe-Ming Lu. 2024. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. IEEE Transactions on Image Processing (2024).

[12] Marcos Conde, Radu Timofte, Zihao Lu, Xiangyu Kong, Xiaoxia Xing, Fan Wang, Suejin Han, MinKyu Park, Tianyu Hao, Yuhong He, et al. 2025. NTIRE 2025 challenge on raw image restoration and super-resolution. In Proceedings of the Computer Vision and Pattern Recognition Conference. 1148–1171.

[13] Marcos V Conde, Ui-Jin Choi, Maxime Burchi, and Radu Timofte. 2022. Swin2sr: Swinv2 transformer for compressed image super-resolution and restoration. In European Conference on Computer Vision. Springer, 669–687.

[14] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition. 3213–3223.

[15] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. 2019. Second-order attention network for single image super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 11065–11074.

[16] Yehuda Dar, Michael Elad, and Alfred M Bruckstein. 2018. Restoration by compression. IEEE Transactions on Signal Processing 66, 22 (2018), 5833–5847.

[17] Monika Dixit and Ram Narayan Yadav. 2024. A review of single image super resolution techniques using convolutional neural networks. Multimedia Tools and Applications 83, 10 (2024), 29741–29775.

[18] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2014. Learning a deep convolutional network for image super-resolution. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13. Springer, 184–199.

[19] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. 2020. Multi-scale boosted dehazing network with dense feature fusion. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2157–2167.

[20] Sina Farsiu, Dirk Robinson, Michael Elad, and Peyman Milanfar. 2004. Advances and challenges in super-resolution. International Journal of Imaging Systems and Technology 14, 2 (2004), 47–57.

[21] Hiroko Furuya, Shintaro Eda, and Testuya Shimamura. 2009. Image restoration via Wiener filtering in the frequency domain. WSEAS transactions on signal processing 5, 2 (2009), 63–73.

[22] Garas Gendy, Jingchao Hou, Nabil Sabor, and Guanghui He. 2025. Lightweight image super-resolution network based on dynamic graph message passing and convolution mixer. Expert Systems with Applications 263 (2025), 125683.

[23] Firouzeh Golaghazadeh, Stéphane Coulombe, François-Xavier Coudoux, and Patrick Corlay. 2017. Checksum-filtered list decoding applied to H. 264 and H. 265 video error correction. IEEE Transactions on Circuits and Systems for Video Technology 28, 8 (2017), 1993–2006.

[24] Hayit Greenspan. 2009. Super-Resolution in Medical Imaging. Comput. J. 52, 1 (2009), 43–63. https://doi.org/10.1093/COMJNL/BXM075

[25] Bahadir Gunturk and Xin Li. 2018. Image restoration. CRC Press.

[26] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. 2024. Mambair: A simple baseline for image restoration with state-space model. arXiv preprint arXiv:2402.15648 (2024).

[27] Javier Gurrola-Ramos, Oscar Dalmau, and Teresa E Alarcon. 2021. A residual dense u-net neural network for image denoising. IEEE Access 9 (2021), 31742–31754.

[28] Yan Han, Peihao Wang, Souvik Kundu, Ying Ding, and Zhangyang Wang. 2023. Vision hgnn: An image is more than a graph of nodes. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 19878–19888.

[29] AD Hiller and Roland T Chin. 1990. Iterative Wiener filters for image restoration. In International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1901–1904.

[30] Jin-Fan Hu, Ting-Zhu Huang, Liang-Jian Deng, Hong-Xia Dou, Danfeng Hong, and Gemine Vivone. 2022. Fusformer: A transformer-based fusion network for hyperspectral image super-resolution. IEEE Geoscience and Remote Sensing Letters 19 (2022), 1–5.

[31] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. 2015. Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE conference on computer vision and pattern recognition. 5197–5206.

[32] Zehao Huang, Lingfeng Wang, Gaofeng Meng, and Chunhong Pan. 2017. Image super-resolution via deep dilated convolutional networks. In 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 953–957.

[33] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. 2019. Lightweight image super-resolution with information multi-distillation network. In Proceedings of the 27th acm international conference on multimedia. 2024–2032.

[34] Geonwoon Jang, Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. 2021. C2n: Practical generative noise modeling for real-world denoising. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 2350–2359.

[35] Hongcheng Jiang and ZhiQiang Chen. 2024. Flexible Window-based Self-attention Transformer in Thermal Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 3076–3085.

[36] Xiangtao Kong, Hengyuan Zhao, Yu Qiao, and Chao Dong. 2021. Classsr: A general framework to accelerate super-resolution networks by data characteristic. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 12016–12025.

[37] Reginald L Lagendijk and Jan Biemond. 2009. Basic methods for image restoration and identification. In The essential guide to image processing. Elsevier, 323–348.

[38] Edmund Y Lam. 2003. Image restoration in digital photography. IEEE Transactions on Consumer Electronics 49, 2 (2003), 269–274.

[39] Dawa Chyophel Lepcha, Bhawna Goyal, Ayush Dogra, and Vishal Goyal. 2023. Image super-resolution: A comprehensive review, recent trends, challenges and applications. Information Fusion 91 (2023), 230–260.

[40] Ao Li, Le Zhang, Yun Liu, and Ce Zhu. 2025. Exploring frequency-inspired optimization in transformer for efficient single image super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025).

[41] Feng Li, Runmin Cong, Jingjing Wu, Huihui Bai, Meng Wang, and Yao Zhao. 2025. Srconvnet: A transformer-style convnet for lightweight image super-resolution. International Journal of Computer Vision 133, 1 (2025), 173–189.

[42] Juncheng Li, Zehua Pei, Wenjie Li, Guangwei Gao, Longguang Wang, Yingqian Wang, and Tieyong Zeng. 2024. A systematic survey of deep learning-based single-image super-resolution. Comput. Surveys 56, 10 (2024), 1–40.

[43] Jie Li, Katherine A Skinner, Ryan M Eustice, and Matthew Johnson-Roberson. 2017. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. IEEE Robotics and Automation letters 3, 1 (2017), 387–394.

[44] Wenjie Li, Juncheng Li, Guangwei Gao, Weihong Deng, Jian Yang, Guo-Jun Qi, and Chia-Wen Lin. 2024. Efficient image super-resolution with feature interaction weighted hybrid network. IEEE Transactions on Multimedia (2024).

[45] Wenbo Li, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, and Jiaya Jia. 2020. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. Advances in Neural Information Processing Systems 33 (2020), 20343–20355.

[46] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision. 1833–1844.

[47] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced Deep Residual Networks for Single Image Super-Resolution. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017. IEEE Computer Society, 1132–1140. https://doi.org/10.1109/CVPRW.2017.151

[48] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 136–144.

[49] Guimin Lin, Qingxiang Wu, Lida Qiu, and Xixian Huang. 2018. Image super-resolution using a dilated convolutional neural network. Neurocomputing 275 (2018), 1219–1230.

[50] Wenyang Liu, Yi Wang, Kim-Hui Yap, and Lap-Pui Chau. 2023. Bitstream-corrupted jpeg images are restorable: Two-stage compensation and alignment

framework for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 9979–9988.

[51] Xin Liu, Jie Liu, Jie Tang, and Gangshan Wu. 2025. CATANet: Efficient Content-Aware Token Aggregation for Lightweight Image Super-Resolution. In Proceedings of the Computer Vision and Pattern Recognition Conference. 17902–17912.

[52] Xianming Liu, Xiaolin Wu, Jiantao Zhou, and Debin Zhao. 2015. Data-driven sparsity-based restoration of JPEG-compressed images in dual transform-pixel domain. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 5171–5178.

[53] Xiao Liu, Chenxu Zhang, and Lei Zhang. 2024. Vision mamba: A comprehensive survey and taxonomy. arXiv preprint arXiv:2405.04404 (2024).

[54] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. 2024. Vmamba: Visual state space model. Advances in neural information processing systems 37 (2024), 103031–103063.

[55] Yuan Liu, Ming Zhu, Jing Wang, Xiangji Guo, Yifan Yang, and Jiarong Wang. 2022. Multi-scale deep neural network based on dilated convolution for spacecraft image segmentation. Sensors 22, 11 (2022), 4222.

[56] Zihan Liu. 2022. Literature review on image restoration. In Journal of Physics: Conference Series, Vol. 2386. IOP Publishing, 012041.

[57] Ziyu Liu, Ruyi Feng, Lizhe Wang, Wei Han, and Tieyong Zeng. 2022. Dual learning-based graph neural network for remote sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing 60 (2022), 1–14.

[58] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using

shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision. 10012–10022.

[59] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. 3431–3440.

[60] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. 2022. Transformer for single image super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 457–466.

[61] Zhisheng Lu, Hong Liu, Juncheng Li, and Linlin Zhang. 2021. Efficient transformer for single image super-resolution. arXiv preprint arXiv:2108.11084 2 (2021).

[62] FowlkesC MartinD et al. 2001. Adatabaseof humansegmentednaturalimagesanditsapplicationto evaluating segmentation algorithms and measuring ecologicalstatistics. IEEEInternationalconference oncomputerVision. Vancouver 416 (2001), 423.

[63] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2017. Sketch-based manga retrieval using manga109 dataset. Multimedia tools and applications 76 (2017), 21811–21838.

[64] H Meertens, M Van Herk, and J Weeda. 1988. An inverse filter for digital restoration of portal images. Physics in Medicine & Biology 33, 6 (1988), 687.

[65] Peyman Milanfar. 2017. Super-resolution imaging. CRC press.

[66] Caner Ozer, Arda Guler, Aysel Turkvatan Cansever, and Ilkay Oksuz. 2023. Explainable Image Quality Assessment for Medical Imaging. arXiv:2303.14479 [eess.IV]

[67] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. 2003. Super-resolution image reconstruction: a technical overview. IEEE signal processing magazine 20, 3 (2003), 21–36.

[68] Xiaohuan Pei, Tao Huang, and Chang Xu. 2025. Efficientvmamba: Atrous selective scan for light weight visual mamba. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 39. 6443–6451.

[69] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. 2019. Enhanced pix2pix dehazing network. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 8160–8168.

[70] Abhisek Ray, Gaurav Kumar, and Maheshkumar H Kolekar. 2024. Cfat: Unleashing triangular windows for image super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 26120–26129.

[71] Bin Ren, Yawei Li, Jingyun Liang, Rakesh Ranjan, Mengyuan Liu, Rita Cucchiara, Luc Van Gool, and Nicu Sebe. 2024. Key-Graph Transformer for Image Restoration. arXiv preprint arXiv:2402.02634 (2024).

[72] Nadeem Salamat, Malik Muhammad Saad Missen, Nadeem Akhtar, Muhammad Mustahsan, and VB Surya Prasath. 2024. Color image restoration by filtering methods: a review. Soft Computing 28, 13 (2024), 7755–7782.

[73] M Ibrahim Sezan and A Murat Tekalp. 1990. Survey of recent developments in digital image restoration. Optical Engineering 29, 5 (1990), 393–404.

[74] Jacob Shermeyer and Adam Van Etten. 2019. The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019, Long Beach, CA, USA, June 16-20, 2019. Computer Vision Foundation / IEEE, 1432–1441. https://doi.org/10.1109/CVPRW.2019.00184

[75] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).

[76] Prabhishek Singh and Raj Shree. 2016. A comparative study to noise models and image restoration techniques. International Journal of Computer Applications 149, 1 (2016), 18–27.

[77] Jae Woong Soh and Nam Ik Cho. 2022. Variational deep image restoration. IEEE Transactions on Image Processing 31 (2022), 4363–4376.

[78] Hetvi Soni and Darshana Sankhe. 2019. Image restoration using adaptive median filtering. Image 6, 10 (2019).

[79] Long Sun, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. 2023. Spatially-adaptive feature modulation for efficient image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 13190–13199.

[80] Luca Superiori, Olivia Nemethova, and Markus Rupp. 2007. Performance of a H. 264/AVC error detection algorithm based on syntax analysis. Journal of mobile multimedia (2007), 314–330.

[81] Michael A Sutton, Stephen R McNeill, Jinseng Jang, and Majid Babai. 1988. Effects of subpixel image restoration on digital correlation error estimates. Optical Engineering 27, 10 (1988), 870–877.

[82] Shenggui Tang, Kaixuan Yao, Jianqing Liang, Zhiqiang Wang, and Jiye Liang. 2023. Graph neural networks with interlayer feature representation for image super-resolution. In Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining. 652–660.

[83] Chunwei Tian, Yong Xu, Wangmeng Zuo, Bob Zhang, Lunke Fei, and Chia-Wen Lin. 2020. Coarse-to-fine CNN for image super-resolution. IEEE Transactions on Multimedia 23 (2020), 1489–1502.

[84] Chunwei Tian, Ruibin Zhuge, Zhihao Wu, Yong Xu, Wangmeng Zuo, Chen Chen, and Chia-Wen Lin. 2020. Lightweight image super-resolution with enhanced CNN. Knowledge-Based Systems 205 (2020), 106235.

[85] Yuchuan Tian, Hanting Chen, Chao Xu, and Yunhe Wang. 2024. Image processing gnn: Breaking rigidity in super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 24108–24117.

[86] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 2018. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE conference on computer vision and pattern recognition. 8798–8807.

[87] Yi Wang, Hui Liu, and Lap-Pui Chau. 2017. Single underwater image restoration using adaptive attenuation-curve prior. IEEE Transactions on Circuits and Systems I: Regular Papers 65, 3 (2017), 992–1002.

[88] Zhihao Wang, Jian Chen, and Steven CH Hoi. 2020. Deep learning for image super-resolution: A survey. IEEE transactions on pattern analysis and machine intelligence 43, 10 (2020), 3365–3387.

[89] Heping Wen, Linchao Ma, Linhao Liu, Yiming Huang, Zefeng Chen, Rui Li, Zhen Liu, Wenxing Lin, Jiahao Wu, Yunqi Li, et al. 2022. High-quality restoration image encryption using DCT frequency-domain compression coding and chaos. Scientific Reports 12, 1 (2022), 16523.

[90] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. 2021. Contrastive learning for compact single image dehazing. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 10551–10560.

[91] Shaoan Xie, Zhifei Zhang, Zhe Lin, Tobias Hinz, and Kun Zhang. 2023. Smartbrush: Text and shape guided object inpainting with diffusion model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 22428–22437.

[92] Jin Yamanaka, Shigesumi Kuwashima, and Takio Kurita. 2017. Fast and accurate image super resolution by deep CNN with skip connection and network in network. In Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24. Springer, 217–225.

[93] Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. 2020. Learning texture transformer network for image super-resolution. In Proceedings of

the IEEE/CVF conference on computer vision and pattern recognition. 5791–5800.

[94] Yue Yang and Yong Qi. 2021. Image super-resolution via channel attention and spatial graph convolutional network. Pattern Recognition 112 (2021), 107798.

[95] Leonid P Yaroslavsky, Karen O Egiazarian, and Jaakko T Astola. 2001. Transform domain image restoration methods: review, comparison, and interpretation. Nonlinear Image Processing and Pattern Analysis XII 4304 (2001), 155–169.

[96] Shutong Ye, Shengyu Zhao, Yaocong Hu, and Chao Xie. 2023. Single-image super-resolution challenges: a brief review. Electronics 12, 13 (2023), 2975.

[97] Tomonari Yoshida, Tomokazu Takahashi, Daisuke Deguchi, Ichiro Ide, and Hiroshi Murase. 2012. Robust Face Super-Resolution Using Free-Form Deformations for Low-Quality Surveillance Video. In Proceedings of the 2012 IEEE International Conference on Multimedia and Expo, ICME 2012, Melbourne, Australia, July 9-13, 2012. IEEE Computer Society, 368–373. https://doi.org/10.1109/ICME.2012.162

[98] Guoshen Yu and Guillermo Sapiro. 2021. Image enhancement and restoration: Traditional approaches. In Computer Vision: A Reference Guide. Springer, 615–619.

[99] Qiao Yu, Wengui Zhang, Jorge Cardoso, and Odej Kao. 2023. Exploring Error Bits for Memory Failure Prediction: An In-Depth Correlative Study. In 2023 IEEE/ACM International Conference on Computer Aided Design (ICCAD). IEEE, 01–09.

[100] Weihao Yu and Xinchao Wang. 2025. Mambaout: Do we really need mamba for vision?. In Proceedings of the Computer Vision and Pattern Recognition Conference. 4484–4496.

[101] Linwei Yue, Huanfeng Shen, Jie Li, Qiangqiang Yuan, Hongyan Zhang, and Liangpei Zhang. 2016. Image super-resolution: The techniques, applications, and future. Signal processing 128 (2016), 389–408.

[102] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yulun Zhang, and Radu Timofte. 2024. See more details: Efficient image super-resolution by experts mining. In Forty-first International Conference on Machine Learning.

[103] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2022. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 5728–5739.

[104] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2021. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 14821–14831.

[105] Roman Zeyde, Michael Elad, and Matan Protter. 2010. On single image scale-up using sparse-representations. In International conference on curves and surfaces. Springer, 711–730.

[106] Zhiyuan Zha, Xin Yuan, Bihan Wen, Jiachao Zhang, Jiantao Zhou, and Ce Zhu. 2020. Image restoration using joint patch-group-based sparse representation. IEEE Transactions on Image Processing 29 (2020), 7735–7750.

[107] Lujun Zhai, Yonghui Wang, Suxia Cui, and Yu Zhou. 2023. A comprehensive review of deep learning-based real-world image restoration. IEEE Access 11 (2023), 21049–21067.

[108] Dong Zhang, Yi Guo, Houqiang Li, and Chang Wen Chen. 2010. Error resilient scalability for video bit-stream over heterogeneous packet loss networks. In Proceedings of 2010 IEEE International Symposium on Circuits and Systems. IEEE, 4201–4204.

[109] Hanwei Zhang, Ying Zhu, Dan Wang, Lijun Zhang, Tianxiang Chen, Ziyang Wang, and Zi Ye. 2024. A survey on visual mamba. Applied Sciences 14, 13 (2024), 5683.

[110] Leheng Zhang, Yawei Li, Xingyu Zhou, Xiaorui Zhao, and Shuhang Gu. 2024. Transcending the limit of local window: Advanced super-resolution transformer with adaptive token dictionary. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2856–2865.

[111] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang. 2022. Efficient long-range attention network for image super-resolution. In European conference on computer vision. Springer, 649–667.

[112] Xiang Zhang, Yulun Zhang, and Fisher Yu. 2024. HiT-SR: Hierarchical transformer for efficient image super-resolution. In European Conference on Computer Vision. Springer, 483–500.

[113] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. 2018. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV). 286–301.

[114] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. 2019. Residual non-local attention networks for image restoration. arXiv preprint arXiv:1903.10082 (2019).

[115] Zhendong Zhang, Xinran Wang, and Cheolkon Jung. 2018. DCSR: Dilated convolutions for single image super-resolution. IEEE Transactions on Image Processing 28, 4 (2018), 1625–1635.

[116] Zhuoran Zheng and Chen Wu. 2024. U-shaped vision mamba for single image dehazing. arXiv preprint arXiv:2402.04139 (2024).

[117] Shangchen Zhou, Jiawei Zhang, Wangmeng Zuo, and Chen Change Loy. 2020. Cross-scale internal graph neural network for image super-resolution. Advances in neural information processing systems 33 (2020), 3499–3509.

[118] Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. 2023. Srformer: Permuted self-attention for single image super-resolution.

In Proceedings of the IEEE/CVF International Conference on Computer Vision. 12780–12791.

[119] Y-T Zhou, Rama Chellappa, Aseem Vaid, and B Keith Jenkins. 1988. Image restoration using a neural network. IEEE transactions on acoustics, speech, and signal processing 36, 7 (1988), 1141–1151.

[120] Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. 2024. Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417 (2024).

[121] Qinfeng Zhu, Yuan Fang, Yuanzhi Cai, Cheng Chen, and Lei Fan. 2024. Rethinking scanning strategies with vision mamba in semantic segmentation of remote sensing imagery: An experimental study. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2024).