OTIAS: OcTree Implicit Adaptive Sampling for Multispectral and Hyperspectral Image Fusion

Shangqi Deng[1,2], **Jun Ma**[3], **Liang-Jian Deng**[3,†], **Ping Wei**[1,2,†]

[1]National Key Laboratory of Human-Machine Hybrid Augmented Intelligence
[2]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
[3]University of Electronic Science and Technology of China

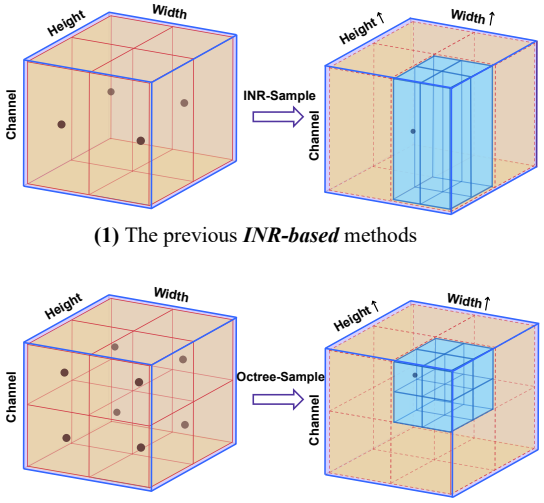
Abstract

Implicit Neural Representation (INR) methods have demonstrated great potential in arbitrary-scale super-resolution tasks. This success is primarily due to their ability to continuously represent images using coordinates. In the task of remote sensing image fusion, INR methods have also shown promising applications. However, the previous INR methods neglect channel-wise modeling, while sharing a single kernel across all channels at each position, resulting in a lack of sensitivity to data specificity. To address these issues, we propose the OcTree Implicit Adaptive Sampling (OTIAS) method, which innovatively applies the octree structure to restore data from both horizontal and vertical directions, effectively incorporating spatial and spectral information from hyperspectral data. Additionally, we introduce a novel method to adaptively generate interpolation kernels based on coordinates. This approach efficiently produces customized interpolation kernel parameters for octree nodes, tailored to different spectral information. Overall, our method achieves state-of-the-art performance on the CAVE and Harvard datasets with $4\times$ and $8\times$ scaling factors, outperforming existing approaches.

Code — https://github.com/shangqideng/OTIAS

Introduction

Hyperspectral imaging is an advanced imaging technique used to acquire image data in different wavelength ranges. Unlike common color images, hyperspectral imaging captures information not only in the three primary colors of R, G, and B but in dozens or even hundreds of very narrow wavelength ranges, providing detailed spectral information. This means that each pixel contains data not only about color but also about the spectral characteristics of the materials. Therefore, hyperspectral imaging finds wide applications in fields (Bedini 2017; Adam, Mutanga, and Rugege 2010; Wang et al. 2023, 2024) such as material identification, vegetation studies, mineral exploration, and more. However, in hyperspectral imaging, physical constraints impose a trade-off between the number of spectral bands and spatial resolution. Typically, increasing spectral resolution results in reduced spatial resolution, and enhancing spatial resolution leads to a decrease in spectral resolution. To address this issue, practical imaging involves capturing the same scene with different sensors, respectively yielding a multispectral image with a high spatial resolution (HR-MSI) and a hyperspectral image with a low spatial resolution (LR-HSI). Recently, multispectral and hyperspectral image fusion (MHIF) (Li et al. 2024, 2023b,a, 2022; Xu et al. 2024; Ran et al. 2023; Cao et al. 2024) has become a crucial area, since it could fuse HR-MSI and LR-HSI to obtain a hyperspectral image with high spatial resolution (HR-HSI).



(1) The previous ***INR-based*** methods

(2) The proposed ***octree-based*** INR method

Figure 1: (1) The previous INR-based methods (Tang, Chen, and Zeng 2021; Zhu et al. 2023) can be viewed as utilizing a quadtree structure for spatial sampling to enhance resolution. (2) The proposed OTIAS considers both spatial and channel directions to subdivide the grid of the original image cube, thereby improving the quality of the fused product, seeing in discussion section.

Implicit Neural Representation (INR) (Mildenhall et al. 2021) is a neural network-based approach for learning continuous representations of objects without the need for explicit geometric information. It has shown remarkable success in various 2D tasks such as image super-resolution (Tang, Chen, and Zeng 2021), thanks to its multi-

†Corresponding authors: liangjian.deng@uestc.edu.cn, pingwei@xjtu.edu.cn
Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

modal (Li and Fan 2022; Li et al. 2023e,d,f,c; Zhu et al. 2024) information fusion capability. INR utilizes Multi-Layer Perceptron (MLP) networks to map input spatial coordinates into continuous signal transformations. However, INR has certain limitations, such as relying on a global MLP network that may struggle to adapt to variations across different coordinates and channels. Consequently, it may lack sufficient sensitivity to local information in some scenarios, particularly in fusion tasks where adapting to detailed variations in objects or scenes is crucial. Furthermore, methods based on INR are primarily focused on spatial sampling and do not consider channel-wise information. Directly applying INR to remote sensing data with a large number of spectral bands may result in poor reconstruction performance.

The octree structure allows for the hierarchical modeling of image data, enabling efficient organization and management of image information. Given the success of this concept, we specifically design a spatial-spectral sampling method for the MHIF task to enhance the efficiency and quality of sampling. In detail, we propose OcTree Implicit Adaptive Sampling (OTIAS), which combines octree structure with the INR interpolation process, achieving feature fusion through hierarchical sampling. Furthermore, an adaptive sampling kernel method is designed, allowing the network to better adjust to variations across different coordinates and channels, thereby improving INR performance in the MHIF task.

Our contributions are listed as follows:

1. We leverage the OcTree Hierarchy (OTH) structure for fusion super-resolution by partitioning leaf nodes in both spatial and spectral domains to store low-resolution data, while root nodes store the feature maps after fusion. Specifically, OTH uses the octree's three-dimensional structure to model the relationships between spatial and spectral domains.
2. We propose an Adaptive Synthesis Kernel (ASK) module that allows the network to capture spatial and spectral variations at the leaf nodes of a single-layer octree. To reduce the kernel generation network's parameter count, we design a lightweight kernel that efficiently captures terrain information in remote sensing images while minimizing overhead.
3. We propose OcTree Implicit Adaptive Sampling (OTIAS), a novel octree-based structure that integrates spatial and spectral adaptability for MHIF. Our method achieves state-of-the-art performance on two public datasets, and ablation studies demonstrate the effectiveness of each OTIAS module.

Related Work

Neural Network Methods for MHIF

In the context of MHIF (Multispectral and Hyperspectral Image Fusion) task, several CNN-based methods have been introduced. Among these methods, DBIN (Wang et al. 2019), an iterative fusion framework, is designed for blind hyperspectral image fusion, seamlessly fusing LR-HSI with HR-MSI without prior knowledge about the observation model and preserving spectral and spatial information simultaneously. SSR-Net (Zhang et al. 2020) employs two consecutive CNN modules to reconstruct the spectral and spatial information in HR-HSI after cross-overlapping LR-HSI and HR-MSI in the spectral domain. Similarly, ResTFNet (Liu, Liu, and Wang 2020) utilizes a residual network (ResNet (He et al. 2016)) and a two-stream architecture to extract spatial and spectral information separately, followed by a residual CNN for their fusion. Meanwhile, MHF-Net (Xie et al. 2022) establishes a linear relationship from the HR-HSI image to the HR-MSI and LR-HSI images, offering clear interpretability. MoG-DCN (Dong et al. 2021) introduces a novel model-guided deep convolutional network (MoG-DCN), which takes the observation matrix of HSI into account during end-to-end optimization. For the concurrent retrieval of both spatial and spectral data, HSR-Net (Hu et al. 2021) employs adaptive structures in both spatial and channel domains. Additionally, BDT (Deng et al. 2023b) introduces a specialized Transformer-based architecture, leveraging spatial and spectral multi-head self-attention mechanisms along with a bidirectional hierarchy strategy, resulting in superior performance in MHIF task. QIS (Zhu et al. 2023) demonstrates the potential of INR techniques in handling remote sensing data.

Implicit Neural Representation

Recently, Implicit Neural Representation (INR) has garnered significant attention in the field of 2D image processing (Chen, Liu, and Wang 2021). Drawing inspiration from INR's applications in 3D scenes (Mildenhall et al. 2021), the Local Implicit Image Function (LIIF) (Chen, Liu, and Wang 2021) employs 2D coordinates and features from the image for implicit representation. Due to the continuity of input coordinates, LIIF's output results can be presented at arbitrary resolutions. Derived from LIIF, UltraSR (Xu, Wang, and Shi 2021) incorporates various architectural designs, such as deep coordinate fusion and residual MLP, combined with spatial encoding, achieving remarkable super-resolution results. Additionally, JIIF (Tang, Chen, and Zeng 2021) suggests using INR to reconstruct depth images in the Low-Resolution (LR) domain by guiding them with High-Resolution (HR) RGB image to address noise-related issues. Furthermore, LTE (Lee and Jin 2022) represents image texture in 2D Fourier space and is jointly trained with a deep super-resolution architecture, facilitating the learning of high-frequency components.

Motivation

To enhance the fidelity of multispectral and hyperspectral fusion images, we consider employing the Implicit Neural Representation (INR) method. However, the previous INR-based methods (Tang, Chen, and Zeng 2021; Zhu et al. 2023) for remote sensing image fusion tasks, to some extent, neglect spectral information (see Fig. 1). Inspired by PixelShuffle (Shi et al. 2016), we establish an OcTree Hierarchy (OTH) sampling model in the spatial-spectral dimensions, encompassing information at different scales, coordinates, and channels. To enhance the model with data specificity, we design an Adaptive Synthesis Kernel (ASK) module for

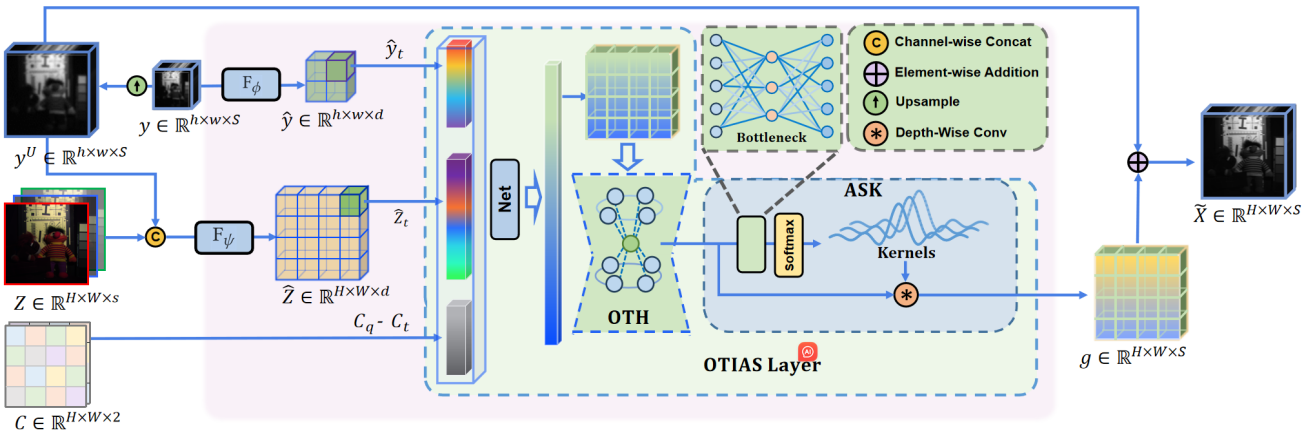


Figure 2: The proposed OTIAS integrates LR-HSI and HR-MSI images to reconstruct HR-HSI, employing encoders for spectral and spatial features. The OTH and ASK are our key components, where OTH partitions leaf nodes from spatial and spectral domains, and ASK generates adaptive kernels for octree leaf nodes.

each leaf node of the OTH, understanding differences in spectral information at different locations. Simultaneously, we consider reducing the network parameters for generating customized kernels, incorporating a lightweight design for improved efficiency.

Methodology

Generally, the problem of multispectral and hyperspectral image fusion (MHIF) can be defined as follows:

$$\mathbf{Y} = \mathbf{XBS}, \quad \mathbf{Z} = \mathbf{RX}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{S\times hw}$, $\mathbf{Z} \in \mathbb{R}^{s\times HW}$, and $\mathbf{X} \in \mathbb{R}^{S\times HW}$ represent the matrix forms of LR-HSI, HR-MSI, and HR-HSI, respectively. Additionally, h, w, and S represent the height, width, and number of bands for LR-HSI, while H, W, and s denote the corresponding dimensions for HR-MSI. Furthermore, $\mathbf{B} \in \mathbb{R}^{HW\times HW}$ is the blur matrix, $\mathbf{S} \in \mathbb{R}^{HW\times hw}$ denotes the downsampling pattern, and $\mathbf{R} \in \mathbb{R}^{s\times S}$ represents the spectral response matrix. In short, our task aims to recover $\mathbf{X}$ from the information contained in $\mathbf{Y}$ and $\mathbf{Z}$.

Preliminary

We drew inspiration from INR methods (Sitzmann et al. 2020; Wu et al. 2024) and re-framed the MHIF problem in a similar fashion, as outlined below:

$$\Phi(C) - \mathcal{X} = 0, \quad \Phi : C \to \Phi(C, \mathcal{Y}, \mathcal{Z}), \tag{2}$$

where $\Phi(\cdot)$ signifies an implicit neural network, $\mathcal{Y} \in \mathbb{R}^{h\times w\times S}$ denotes LR-HSI, $\mathcal{Z} \in \mathbb{R}^{H\times W\times s}$ pertains to HR-LSI, $\mathcal{X} \in \mathbb{R}^{H\times W\times S}$ denotes HR-HSI, and $C \in \mathbb{R}^{H\times W\times 2}$ denotes the normalization of coordinates within the high-resolution (HR) domain. Specifically, we represent pixels using their center positions and map the coordinates of $H\times W$ to a square grid of size $[-1, 1]\times[-1, 1]$ for the convenience of sharing coordinates in both the high-resolution and low-resolution domains. The normalization process (Chen, Liu, and Wang 2021) in the high-resolution domain can be formulated as follows:

$$C(i, j) = \left[-1 + \frac{2i+1}{H}, -1 + \frac{2j+1}{W}\right], \tag{3}$$

where $i \in [0, H-1], j \in [0, W-1]$. Furthermore, to ensure network stability and achieve better fusion results, we encode $\mathcal{Y}$ and $\mathcal{Z}$ through neural networks, mapping them into the latent space:

$$\begin{cases} \hat{\mathcal{Y}} = F_\phi(\mathcal{Y}), \\ \hat{\mathcal{Z}} = F_\psi(Cat(\mathcal{Y}^U, \mathcal{Z})), \end{cases} \tag{4}$$

where $\hat{\mathcal{Y}} \in \mathbb{R}^{h\times w\times d}$ represents the feature map of the spectral modality, $\hat{\mathcal{Z}} \in \mathbb{R}^{H\times W\times d}$ represents the feature map of the spatial modality, and $\mathcal{Y}^U \in \mathbb{R}^{H\times W\times S}$ denotes the bicubic interpolation upsampled LR-HSI. d corresponds to the channel numbers of feature maps. The symbols ϕ and ψ denote learnable parameters associated with the spectral and spatial encoder functions. Subsequently, we define the INR sampling process at location C_q in the form of an quadtree, where the four leaf nodes are defined as:

$$\mathcal{N}_q = \{f_\theta(\hat{\mathcal{Y}}_t, \hat{\mathcal{Z}}_t, C_q - C_t) | C_t \in \mathcal{C}\}, \quad \mathcal{N}_q \in \mathbb{R}^{4\times d}. \tag{5}$$

$C_q \in \mathbb{R}^2$ is the normalized coordinate of the query pixel in the HR domain, and $\mathcal{C} \in \mathbb{R}^4$ denotes the coordinate of neighboring pixels at position C_q in the LR domain. Additionally, $\hat{\mathcal{Y}}_t \in \mathbb{R}^d$ and $\hat{\mathcal{Z}}_t \in \mathbb{R}^d$ denote the spectral and spatial latent code at the position C_t, respectively. The term $C_q - C_t$ represents relative distance, f_θ denotes an MLP layer, and θ is the learnable parameters. Then, the output of $\Phi(\cdot)$ at position C_q is stored in the root node of the quadtree, which is expressed as follows:

$$\Phi(C_q) = \mathcal{N}_q * \mathcal{K}_q, \tag{6}$$

where $\mathcal{K}_q \in \mathbb{R}^{4\times 1}$ can be regarded as a kernel function used to associate the leaf nodes with the corresponding root

node. It is not difficult to observe that this kernel function is channel-shared. We use the LIIF (Chen, Liu, and Wang 2021) method to specify $\mathcal{K}_q$. LIIF utilizes a straightforward area-weighted interpolation formula (Chen, Liu, and Wang 2021):

$$\mathcal{K}_{q,t} = \frac{A_{q,t}}{A}, \tag{7}$$

where $\mathcal{K}_{q,t} \in \mathbb{R}$, $A_{q,t}$ represents the area formed by the pixels diagonally opposite to the t corner pixel with respect to C_q, and $A = \sum_{t \in \mathcal{C}} A_{q,t}$ represents the total area serving as the denominator. The four computed corner values of $\mathcal{K}_{q,t}$ are used to form the kernel function $\mathcal{K}_q$.

The Overall Architecture

The overall architecture of our model is shown in Fig. 2. We demonstrate a novel OcTree Implicit Adaptive Sampling (OTIAS) approach that leverages the OcTree Hierarchy (OTH) structure to model the sampling process. Additionally, we devise an Adaptive Synthesis Kernel (ASK) module and successfully apply it to the OTH framework. The proposed OTIAS employs an approach based on INR to fuse the LR-HSI image $\mathcal{Y} \in \mathbb{R}^{h \times w \times S}$ and the HR-MSI image $\mathcal{Z} \in \mathbb{R}^{H \times W \times s}$, aiming to reconstruct the HR-HSI image $\mathcal{X} \in \mathbb{R}^{H \times W \times S}$. First, we yield a spectral feature map $\hat{\mathcal{Y}} \in \mathbb{R}^{h \times w \times d}$ and a spatial feature map $\hat{\mathcal{Z}} \in \mathbb{R}^{H \times W \times d}$ via dual networks. Subsequently, we take the normalized coordinates $C \in \mathbb{R}^{H \times W \times 2}$ as input, which is represented as:

$$\mathcal{G} = \Psi(C), \tag{8}$$

where $\Psi(\cdot)$ denotes the OTIAS layer. $\mathcal{G}$ represents the features learned by OTH, sharing the same shape as HR-HSI.

The OTIAS layer consists of the OcTree Hierarchy $\Omega(\cdot)$ structure and the Adaptive Synthesis Kernel $\psi(\cdot)$. Therefore, Eq. (8) can be expressed in the form:

$$\Psi(C) = \psi(\Omega(\hat{\mathcal{Y}}, \hat{\mathcal{Z}}, C)), \tag{9}$$

where $\Omega(\cdot)$ involves partitioning leaf nodes both in spatial and spectral domains. It leverages the three-dimensional structure of an octree to better capture inter-band correlations. Furthermore, $\psi(\cdot)$ dynamically generates kernels for the octree leaf nodes based on the input information, allowing for improved adaptation to local characteristics.

Consequently, based on prior research that suggests long skip connections in local implicit representations can enrich high-frequency components in residuals and stabilize convergence (Kim, Lee, and Lee 2016), we incorporate the bicubic interpolated LR-HSI $\mathcal{Y}^U$ as a long skip connection. The final fusion result takes the following form:

$$\tilde{\mathcal{X}} = \mathcal{G} + \mathcal{Y}^U. \tag{10}$$

Eq. (9) allows us to stack OTIAS layers to extend the network depth. Specifically, the OTIAS layer at the C_q position can be represented as:

$$\Psi(C_q) = \psi(\Omega(f_\theta(\hat{\mathcal{Y}}_t, \hat{\mathcal{Z}}_t, C_q - C_t), \mathcal{C}), C_q), \tag{11}$$

where $\Omega(\cdot)$ denotes the OTH, $\psi(\cdot)$ means the Adaptive Synthesis Kernel, and $\hat{\mathcal{Y}}_t \in \mathbb{R}^{1 \times 1 \times d}$. The detailed process of the OTIAS layer is optimized for batch parallelism and is provided in Algorithm. 1.

Algorithm 1: Pseudo code of OTIAS layer in a PyTorch-like style.

Input: LR-domian feature $\hat{\mathcal{Y}}$: `(B, h, w, d)`, HR-domain grid c: `(B, HW, 2)`, LR-domain grid $c\downarrow$: `(B, hw, 2)`, HR-domain feature $\hat{\mathcal{Z}}$: `(B, H, W, d)`.
▷ Here, B means batch-size.
Output: OTIAS layer output $recon$: `(B, H, W, d)`.
▷ Init Δx, Δy, c_.
$\Delta x = 1/h$, $\Delta y = 1/w$, $c\_ =$ `c.clone()`;
▷ Obtain query coordinate c_.
$c\_[:, :, 0]$, $[:, :, 1] \pm = [-1, 1] * \Delta x, [-1, 1] * \Delta y$;
▷ Init Net(·), Reduce(·) and Span(·);
`Net`$(\cdot)$ = `MLP`$(2d+2, 2d)$;
`Reduce`$(\cdot)$ = `MLP`$(d, d//r)$;
`Span`$(\cdot)$ = `MLP`$(d//r, d)$;
def `forwardFunction`$(\hat{\mathcal{Y}}, c, c\downarrow, \hat{\mathcal{Z}})$**:**
  ▷ Sample the LR-domian feature f based on grid c.
  $\hat{\mathcal{Y}}\_s$: `(B, HW, 4, d)` $\leftarrow$ `F.grid_Sample`$(\hat{\mathcal{Y}}, c\_)$;
  ▷ Sample the LR-domain grid $c\downarrow$ based on grid c.
  c_s: `(B, HW, 4, 2)` $\leftarrow$ `F.grid_Sample`$(c\downarrow, c\_)$;
  ▷ Relative sampling distance.
  Δc: `(B, HW, 4, 2)` $\leftarrow$ `Minus`(c, c_s);
  ▷ Construct the octree hierarchy.
  o: `(B, HW, 8, d)` $\leftarrow$
   `Net(cat(`$\hat{\mathcal{Y}}\_s, \hat{\mathcal{Z}}, \Delta c$`)).view()`;
  ▷ Obtain adaptive synthesis kernel.
  k: `(B, HW, 8, d)` $\leftarrow$
   `F.softmax(Span(Reduce(`o`)), dim=-2)`;
  ▷ Obtain the fusion reconstruction.
  $recon$ `(B, H, W, d)` $\leftarrow$
   $(o \circledast k)$`.sum(-2).view()`;
  return $recon$;
return `forwardFunction(`$recon$`)`;

OcTree Hierarchy (OTH)

For previous INR super-resolution methods, they sample Low-Resolution (LR) domain pixels in the spatial domain by calculating offsets Δx and Δy in the up, down, left, and right directions to obtain the nearest neighboring $\mathcal{C} \in \mathbb{R}^4$ four pixels. This process can be viewed as one layer in the form of quadtree sampling, where four leaf nodes store information about the nearest neighboring pixels, and a newly generated root node stores information about the target pixels after super-resolution. The sampling process is expressed as:

$$\mathcal{N}_q = \text{QuadTree}(\hat{\mathcal{Y}}_t, \hat{\mathcal{Z}}_t, C_q, \mathcal{C}), \tag{12}$$

where QuadTree$(\cdot)$ represents the quadtree sampling describing the previous INR methods. $\mathcal{N}_q \in \mathbb{R}^{4 \times d}$ represents the information of the leaf nodes at position C_q.

Due to $\mathcal{C} \in \mathbb{R}^4$ having only spatial direction offsets nearby the query position C_q, the previous INR methods exclusively

consider the spatial domain, neglecting the spectral information. Inspired by Pixel-Shuffle (Shi et al. 2016), we propose the OcTree Hierarchy (OTH) structure denoted as $\Omega(\cdot)$, which expands channel information to the spatial domain resulted in extending the sampling space. Specifically, we further split the channels and rearrange the information on the channels into space. The structure of OTH can be viewed as an upgraded version of a quadtree, characterized by an extension along the channel dimension, which can be described as:

$$\hat{\mathcal{N}}_q = \Omega(\hat{\mathcal{Y}}_t, \hat{\mathcal{Z}}_t, C_q, \mathcal{C}). \tag{13}$$

In detail, based on $\mathcal{C} \in \mathbb{R}^4$, we query four LR-domain feature vectors, HR-domain feature vectors, and relative sampling distance at position C_q, obtaining four leaf nodes $\{\mathcal{N}_{q0}, \mathcal{N}_{q1}, \mathcal{N}_{q2}, \mathcal{N}_{q3}\} \in \mathbb{R}^{4\times d}$. After that, through an MLP, we increase the channel dimension to $2d$. Subsequently, we group and rearrange them along the channel direction, obtaining the new leaf nodes $\hat{\mathcal{N}}_q = \{\mathcal{N}_{q0}, \mathcal{N}_{q1}, \cdots, \mathcal{N}_{q7}\} \in \mathbb{R}^{8\times d}$, as illustrated in the Algorithm 1.

Adaptive Synthesis Kernel (ASK)

In order to fully account for the specificity of data located at different positions C_q, we design a strategy called Adaptive Synthesis Kernel (ASK). Taking into account a set of leaf nodes represented as $\hat{\mathcal{N}}_q \in \mathbb{R}^{8\times d}$, our objective is to create an adaptive kernel based on C_q with the shape $\mathbb{R}^{d\times d\times 8}$. Next, we apply the adaptive kernel based on C_q to the leaf nodes $\hat{\mathcal{N}}_q = \{\mathcal{N}_{q0}, \mathcal{N}_{q1}, \cdots, \mathcal{N}_{q7}\} \in \mathbb{R}^{8\times d}$ through matrix multiplication to generate the root node $\Psi(C_q) \in \mathbb{R}^{1\times d}$. The above process can be represented as:

$$\Psi(C_q) = \psi(\hat{\mathcal{N}}_q, C_q). \tag{14}$$

However, we find that the cost of generating such a shaped kernel for each C_q is enormous. Assuming the target resolution size is 64×64, with a channel size of $d = 128$, we would need 2 Gigabytes to store the parameters of the generated kernel for each layer, which is unacceptable for the MHIF task. To alleviate this issue, we simplify the structure of the generation kernel by designing it in the form of a depthwise convolutional kernel. Transforming it into a depthwise convolutional kernel reduces the number of parameters to 1% of the original, significantly easing the computational burden. After obtaining eight leaf nodes $\hat{\mathcal{N}}_q \in \mathbb{R}^{8\times d}$ through Eq. (13), the leaf nodes are then fed into a multi-layer perceptron with a bottleneck shape. Subsequently, a softmax operation is performed on the first dimension. The entire process is represented as:

$$\hat{\mathcal{K}}_q = \text{Softmax}(\text{Span}(\text{Reduce}(\hat{\mathcal{N}}_q))), \tag{15}$$

where $\hat{\mathcal{K}}_q \in \mathbb{R}^{8\times d}$ denotes the customized kernel. $\text{Span}(\cdot)$ and $\text{Reduce}(\cdot)$ refers to Algorithm 1.

Then, we utilize the customized kernel $\hat{\mathcal{K}}_q$ for executing depthwise convolution on the leaf nodes $\hat{\mathcal{N}}_q$ as follows:

$$\psi(\hat{\mathcal{N}}_q, C_q) = \hat{\mathcal{N}}_q \circledast \hat{\mathcal{K}}_q, \tag{16}$$

where $\circledast$ means the matrix product operation. Finally, we parallelize the computation of the adaptive synthesis kernel and the depth-wise convolution operation to obtain the restored result, $recon \in \mathbb{R}^{H\times W\times d}$, as demonstrated in Algorithm 1.

The learnable parameters in the entire network denoted as θ, are optimized through data-driven iterations. The learning problem is defined as follows:

$$\hat{\theta} = \arg\min_{\theta} \sum_{m}^{M} \parallel \Psi(C) - \mathcal{X} \parallel, \tag{17}$$

where M represents the number of training pairs.

Experiment

Experiment Setting

Datasets. We conduct experiments to assess the performance of our model on the CAVE[1] and Harvard[2]. The CAVE dataset comprises 32 hyperspectral images (HSIs) with 31 spectral bands ranging in wavelengths from 400 nm to 700 nm, and a spectral resolution of 10 nm. The pervious works (Hu et al. 2021, 2022) select 20 images for training, leaving the remaining 11 images for the test dataset. The Harvard dataset consists of 77 HSIs of both indoor and outdoor scenes, each with a size of $1392 \times 1040 \times 31$ and spanning the spectral range from 420 nm to 720 nm. Twenty of these images are selected, with the upper-left portion cropped to (1000×1000), and 10 images are used for testing while the remaining 10 are used for training.

Data Simulation. We input a pair of LR-HSI and HR-MSI images, denoted as $(\mathcal{Y}, \mathcal{Z})$, into the proposed OTIAS, utilizing the related HR-HSI $\bar{\mathcal{X}}$, for training. As the ground-truth (GT), $\mathcal{X}$, is not available, a simulation process is required. For the CAVE dataset, we crop the 20 selected training images, generating 3920 overlapping patches with a size of $64\times 64\times 31$, and these patches are used as GT, $\mathcal{X}$. To generate the corresponding LR-HSIs, we apply a 3×3 Gaussian kernel with a standard deviation of 0.5 to blur the HR-HSIs, followed by downsampling the blurred patches by a factor of 4. Moreover, we use the common spectral response function of the Nikon D700[3] camera to create the HR-MSI patches from the corresponding HR-HSIs. Thus, we generate 3920 LR-HSIs with a size of $16 \times 16 \times 31$ and 3920 HR-MSIs with a size of $64 \times 64 \times 3$. The simulated pairs with the associated GTs are randomly divided into training (80%) and testing (20%) sets. For the Harvard dataset, we simulate images following the same procedure.

Implementation Details and Quality Indexes. The proposed network is implemented in PyTorch 2.4.0 and Python 3.11. Additionally, the AdamW optimizer (P and Ba 2014) is used during training with a learning rate of 0.0001 to minimize the sum of the absolute difference (ℓ_1). The training

[1]https://www.cs.columbia.edu/CAVE/databases/multispectral/

[2]http://vision.seas.harvard.edu/hyperspec/index.html

[3]https://www.maxmax.com/nikon_d700_study.htm

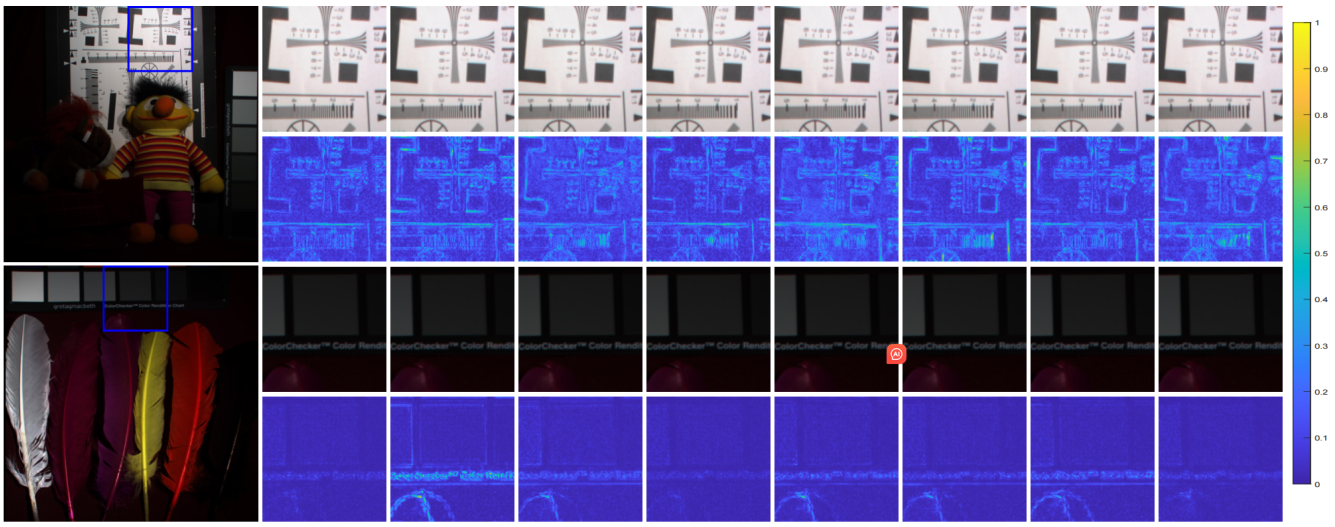


Figure 3: The first and third rows show the results using the pseudo-color (R-31, G-20, B-10) representation on "*chart and stuffed toy*" and "*feathers*", respectively, from the CAVE dataset. Some close-ups are depicted in the blue rectangles. The second and fourth rows show the residuals between the GT and the fused products. (a) GT. (b) Ours. (c) MIMO (Fang et al. 2024). (d) DCT (Ma et al. 2024). (e) QIS (Zhu et al. 2023). (f) DSPNet (Sun et al. 2023). (g) 3DT-Net (Ma et al. 2023). (h) PSRT (Deng et al. 2023a). (i) DHIF (Huang et al. 2022).

epochs is fixed into 1000 on a Linux operating system with an NVIDIA RTX4090 GPU (24G).

We adopt four widely-used quality metrics (Vivone 2023; Vivone et al. 2020; Wang et al. 2004): the peak signal-to-noise ratio (PSNR), the spectral angle mapper (SAM), the relative dimensionless global error synthesis (ERGAS), and the structural similarity index measure (SSIM).

Comparison with other Methods

To validate the superiority of the proposed OTIAS, we compare it with various state-of-the-art methods, including CSTF-FUS (Li et al. 2018), LTTR (Dian, Li, and Fang 2019), LTMR (Dian and Li 2019), IR-TenSR (Xu et al. 2022), SSRNet (Zhang et al. 2020), ResTFNet (Liu, Liu, and Wang 2020), HSRNet (Hu et al. 2021), MogDCN (Dong et al. 2021), Fusformer (Hu et al. 2022), DHIF (Huang et al. 2022), PSRT (Deng et al. 2023a), 3DT-Net (Ma et al. 2023), DSPNet (Sun et al. 2023), QIS (Zhu et al. 2023), DCT (Ma et al. 2024) and MIMO (Fang et al. 2024).

Results on CAVE Dataset. We assess our OTIAS method on CAVE dataset with a scaling factor of 4, conducting a comparative analysis with existing MHIF methods. As shown in the left part of Tab. 2, our OTIAS significantly surpasses other state-of-the-art models. For instance, our OTIAS improves PSNR by 1.57 dB, 2.46 dB, 0.21 dB, and 1.25 dB compared with MIMO (Fang et al. 2024), DCT (Ma et al. 2024), QIS (Zhu et al. 2023), and DSPNet (Sun et al. 2023), respectively. Concerning the absolute error maps in Fig. 3, as the reconstruction impact approaches the original picture more closely, the color of the error map tends to be more blue. It is clear that OTIAS excels in restoring texture details compared to the other techniques under consideration, aligning with the findings presented in Tab. 2.

To ensure a fair comparison, we further evaluate the proposed OTIAS against two other INR methods (Tang, Chen, and Zeng 2021; Zhu et al. 2023) on the CAVE dataset, with the results summarized in Tab. 1. The results indicate that, with comparable parameter counts and FLOPs, our method delivers significantly better performance. Additionally, the absolute error maps in Fig. 4 further highlight the advantages of our approach in restoring fine details and textures, demonstrating its superior capability in preserving structural information and achieving high-quality reconstructions.

Method	PSNR (↑)	SAM (↓)	ERGAS (↓)	#Params	#Flops
JIIF	50.93	2.13	1.12	2.832M	8.041G
QIS	52.22	1.98	1.02	2.914M	7.658G
Ours	52.43	1.94	0.99	2.99M	8.722G

Table 1: The average three QIs, the parameters and flops on the CAVE dataset simulating a scaling factor of 4.

Results on Harvard Dataset. The right part of Tab. 2 provides the comparison on Harvard dataset, with a scaling factor of 4. Clearly, the proposed OTIAS exhibits an average PSNR value that surpasses the second-best and third-best methods by 0.2 dB and 0.79 dB, respectively. While our model exhibits a slight inferiority compared to the third-best DSPNet (Sun et al. 2023) in terms of SAM and SSIM, it is noteworthy that we achieve a reduction of 50.7% in parameters compared with DSPNet. Furthermore, our model attains the best results in terms of ERGAS.

Ablation Study

We examine the effect of Adaptive Synthesis Kernel (ASK) module and the OcTree Hierarchy (OTH).

Methods	CAVE ×4				Harvard ×4				#Params	#FLOPs
	PSNR (↑)	SAM (↓)	ERGAS (↓)	SSIM (↑)	PSNR (↑)	SAM (↓)	ERGAS (↓)	SSIM (↑)		
Bicubic	34.33±3.88	4.45±1.62	7.21±4.90	0.944±0.0291	38.71±4.33	2.53±0.67	4.45±1.81	0.948±0.0268	–	–
CSTF-FUS (Li et al. 2018)	34.46±4.28	14.37±5.30	8.29±5.29	0.866±0.0747	39.15±3.45	6.93±2.69	4.66±1.81	0.914±0.0489	–	–
LTTR (Dian, Li, and Fang 2019)	35.85±3.49	6.99±2.55	5.99±2.92	0.956±0.0288	40.88±3.94	4.01±1.27	4.03±2.18	0.957±0.0350	–	–
LTMR (Dian and Li 2019)	36.54±3.30	6.71±2.19	5.39±2.53	0.963±0.0208	42.06±3.56	3.51±0.99	3.59±2.03	0.970±0.0201	–	–
IR-TenSR (Xu et al. 2022)	35.61±3.45	12.30±4.68	5.90±3.05	0.945±0.0267	40.47±3.04	4.36±1.52	5.57±1.57	0.963±0.0137	–	–
ResTFNet (Liu, Liu, and Wang 2020)	45.58±5.47	2.82±0.70	2.36±2.59	0.993±0.0056	45.94±4.35	2.61±0.69	2.56±1.32	0.985±0.0082	2.387M	1.75G
SSRNet (Zhang et al. 2020)	48.62±3.92	2.54±0.84	1.63±1.21	0.995±0.0023	48.00±3.36	2.31±0.60	2.30±1.42	0.987±0.0070	0.027M	0.11G
HSRNet (Hu et al. 2021)	50.38±3.38	2.23±0.66	1.20±0.75	0.996±0.0014	48.29±3.03	2.26±0.56	1.87±0.81	0.988±0.0073	1.09M	2.00G
MogDCN (Dong et al. 2021)	51.63±4.10	2.03±0.62	1.11±0.82	0.997±0.0018	47.89±4.09	2.11±0.52	1.89±0.82	0.988±0.0073	6.840M	47.534G
Fusformer (Hu et al. 2022)	49.98±8.10	2.20±0.85	2.50±5.21	0.994±0.0111	47.87±5.13	2.84±2.07	2.04±0.99	0.986±0.0101	0.504M	10.124G
DHIF (Huang et al. 2022)	51.07±4.17	2.01±0.63	1.22±0.97	0.997±0.0016	47.68±3.85	2.32±0.53	1.95±0.92	0.988±0.0074	22.462M	54.273G
PSRT (Deng et al. 2023a)	50.47±6.19	2.19±0.64	2.06±3.71	0.996±0.0031	47.96±3.21	2.18±0.55	1.89±0.86	0.989±0.0059	0.247M	1.138G
3DT-Net (Ma et al. 2023)	51.38±4.18	2.16±0.70	1.14±1.00	0.996±0.0026	47.78±4.42	2.04±0.51	1.98±0.86	0.989±0.0059	3.464M	68.19G
DSPNet (Sun et al. 2023)	51.18±3.92	2.15±0.64	1.13±0.82	0.997±0.0014	48.29±3.16	2.30±0.55	1.92±0.91	0.988±0.0061	6.065M	6.855G
QIS (Zhu et al. 2023)	52.22±4.21	1.98±0.60	1.02±0.81	0.997±0.0014	48.88±3.33	2.14±0.52	1.77±0.81	0.989±0.0064	2.914M	7.658G
DCT (Ma et al. 2024)	49.97±3.58	2.51±0.80	1.25±0.88	0.996±0.0017	46.42±5.30	2.19±0.55	2.30±0.90	0.988±0.0066	5.262M	53.362G
MIMO (Fang et al. 2024)	50.86±3.45	2.28±0.71	1.18±0.72	0.996±0.0014	47.03±5.57	2.08±0.52	2.02±0.79	0.988±0.0069	4.583M	1.583G
OTIAS (Ours)	52.43±4.07	1.94±0.58	0.99±0.76	0.997±0.0013	49.08±3.28	2.09±0.52	1.72±0.76	0.989±0.0066	2.99M	8.722G

Table 2: The average and standard deviation calculated for all the compared approaches on 11 CAVE examples and 10 Harvard examples simulating a scaling factor of 4. The best results are in red and the second best results are in blue.

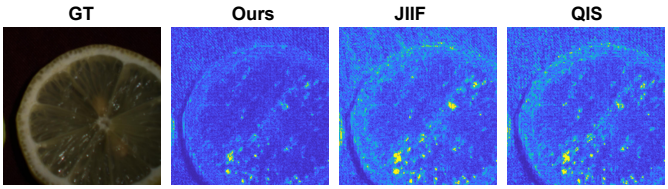


Figure 4: The residuals effect diagram on the "*fake and real lemon slices*" in the CAVE dataset.

Method	PSNR (↑)	SAM (↓)	ERGAS (↓)	#Params	#Flops
JIIF	50.93	2.13	1.12	2.832M	8.041G
JIIF(ASK)	51.22	1.98	1.02	2.839M	8.07G
QIS	52.22	1.98	1.02	3.187M	8.811G
QIS(ASK)	52.32	1.97	0.99	3.617M	10.089G

Table 3: The presence of 'ASK' in parentheses indicates the application of ASK technology. The average three QIs, the corresponding parameters and flops on the CAVE dataset simulating a scaling factor of 4.

ASK with previous INR methods. To further illustrate the effectiveness of the Adaptive Synthesis Kernel (ASK) strategy, we integrate it into two representative INR methods, namely JIIF (Tang, Chen, and Zeng 2021) and QIS (Zhu et al. 2023). The results, presented in Tab. 3, clearly demonstrate that the incorporation of the proposed ASK design enhances the performance of the original models while introducing only a negligible increase in computational cost and parameter count. This significant improvement, achieved with minimal overhead, highlights the efficiency of the ASK strategy. These experiments not only validate the robustness and generalizability of the proposed ASK design but also emphasize its potential to serve as a universal enhancement for INR frameworks.

Structure of OTH. We investigate the impact of the OcTree Hierarchy (OTH) structure on spectral fidelity. We modify our network model, where in Tab. 4, the number 1 indicates a single-layer octree structure, and the number 2 indicates a two-layer octree structure. We apply these models to an $8\times$ scaling factor task. As shown in Tab. 4, the experimental results reveal that as the number of layers increases, the spectral fidelity of fusion results improves, with only a marginal increase in the number of parameters.

Layer	SAM (↓)	ERGAS (↓)	SSIM (↑)	#Params	#Flops
1	2.59	0.74	0.996	2.825M	29.419G
2	2.58	0.76	0.996	2.924M	31.116G
3	2.51	0.70	0.996	2.99M	33.842G

Table 4: The average three QIs, the parameters and flops on the CAVE dataset simulating a scaling factor of 8.

Conclusion

In this article, we introduce a novel spatial-spectral adaptive octree sampling structure, OcTree Implicit Adaptive Sampling (OTIAS). Compared with previous methods, we innovatively integrate OcTree Hierarchy (OTH) with INR for fusion tasks, enhancing both feature extraction and representation capabilities. Additionally, the design of the Adaptive Synthesis Kernel (ASK) module effectively improves the performance of octree sampling in the MHIF task. Experimental results demonstrate that our method achieves state-of-the-art performance in the MHIF task.

Acknowledgements

This research was partially supported by National Natural Science Foundation of China (No.U23B2060, No.62088102), Youth Innovation Team of Shaanxi Universities, and partially supported by National Natural Science Foundation of China (No.12271083). We sincerely thank Yu-Jie Liang and Zihan Cao for their valuable support in this work.

References

Adam, E.; Mutanga, O.; and Rugege, D. 2010. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. *Wetlands ecology and management*, 18: 281–296.

Bedini, E. 2017. The use of hyperspectral remote sensing for mineral exploration: A review. *Journal of Hyperspectral Remote Sensing*, 7(4): 189–211.

Cao, Z.; Cao, S.; Deng, L.-J.; Wu, X.; Hou, J.; and Vivone, G. 2024. Diffusion model with disentangled modulations for sharpening multispectral and hyperspectral images. *Information Fusion*, 104: 102158.

Chen, Y.; Liu, S.; and Wang, X. 2021. Learning continuous image representation with local implicit image function. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8628–8638.

Deng, S.; Deng, L.-J.; Wu, X.; Ran, R.; Hong, D.; and Vivone, G. 2023a. PSRT: Pyramid shuffle-and-reshuffle transformer for multispectral and hyperspectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–15.

Deng, S.; Deng, L.-J.; Wu, X.; Ran, R.; and Wen, R. 2023b. Bidirectional Dilation Transformer for Multispectral and Hyperspectral Image Fusion. *International Joint Conference on Artificial Intelligence (IJCAI)*, 3633–3641.

Dian, R.; and Li, S. 2019. Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. *IEEE Transactions on Image Processing*, 28(10): 5135–5146.

Dian, R.; Li, S.; and Fang, L. 2019. Learning a low tensor-train rank representation for hyperspectral image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9): 2672–2683.

Dong, W.; Zhou, C.; Wu, F.; Wu, J.; Shi, G.; and Li, X. 2021. Model-guided deep hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 30: 5754–5768.

Fang, J.; Yang, J.; Khader, A.; and Xiao, L. 2024. MIMO-SST: Multi-Input Multi-Output Spatial-Spectral Transformer for Hyperspectral and Multispectral Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–20.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Conference on Computer Vision and Pattern RecognitionC*, 770–778.

Hu, J.; Huang, T.; Deng, L.; Dou, H.; Hong, D.; and Vivone, G. 2022. Fusformer: A Transformer-Based Fusion Network for Hyperspectral Image Super-Resolution. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Hu, J.; Huang, T.; Deng, L.; Jiang, T.; Vivone, G.; and Chanussot, J. 2021. Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33: 7251–7265.

Huang, T.; Dong, W.; Wu, J.; Li, L.; Li, X.; and Shi, G. 2022. Deep Hyperspectral Image Fusion Network With Iterative Spatio-Spectral Regularization. *IEEE Transactions on Computational Imaging*, 8: 201–214.

Kim, J.; Lee, J. K.; and Lee, K. M. 2016. Accurate image super-resolution using very deep convolutional networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1646–1654.

Lee, J.; and Jin, K. H. 2022. Local texture estimator for implicit representation function. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1929–1938.

Li, J.; Zheng, K.; Gao, L.; Ni, L.; Huang, M.; and Chanussot, J. 2024. Model-Informed Multistage Unsupervised Network for Hyperspectral Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–17.

Li, J.; Zheng, K.; Li, Z.; Gao, L.; and Jia, X. 2023a. X-Shaped Interactive Autoencoders With Cross-Modality Mutual Learning for Unsupervised Hyperspectral Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–17.

Li, J.; Zheng, K.; Liu, W.; Li, Z.; Yu, H.; and Ni, L. 2023b. Model-Guided Coarse-to-Fine Fusion Network for Unsupervised Hyperspectral Image Super-Resolution. *IEEE Geoscience and Remote Sensing Letters*, 20: 1–5.

Li, J.; Zheng, K.; Yao, J.; Gao, L.; and Hong, D. 2022. Deep unsupervised blind hyperspectral and multispectral data fusion. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Li, S.; Dian, R.; Fang, L.; and Bioucas-Dias, J. M. 2018. Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. *IEEE Transactions on Image Processing*, 27(8): 4118–4130.

Li, W.; and Fan, X. 2022. Image-Text Alignment and Retrieval Using Light-Weight Transformer. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 4758–4762.

Li, W.; Ma, Z.; Deng, L.-J.; Man, H.; and Fan, X. 2023c. Modality-Fusion Spiking Transformer Network for Audio-Visual Zero-Shot Learning. In *IEEE International Conference on Multimedia and Expo (ICME)*, 426–431.

Li, W.; Ma, Z.; Deng, L.-J.; Wang, P.; Shi, J.; and Fan, X. 2023d. Reservoir Computing Transformer for Image-Text Retrieval. In *ACM International Conference on Multimedia (ACM MM)*, 5605–5613.

Li, W.; Ma, Z.; Shi, J.; and Fan, X. 2023e. The Style Transformer With Common Knowledge Optimization for Image-Text Retrieval. *IEEE Signal Processing Letters*, 30: 1197–1201.

Li, W.; Zhao, X.-L.; Ma, Z.; Wang, X.; Fan, X.; and Tian, Y. 2023f. Motion-Decoupled Spiking Transformer for Audio-Visual Zero-Shot Learning. In *ACM International Conference on Multimedia (ACM MM)*, 3994–4002.

Liu, X.; Liu, Q.; and Wang, Y. 2020. Remote sensing image fusion based on two-stream fusion network. *Information Fusion*, 55: 1–15.

Ma, Q.; Jiang, J.; Liu, X.; and Ma, J. 2023. Learning a 3D-CNN and transformer prior for hyperspectral image super-resolution. *Information Fusion*, 101907.

Ma, Q.; Jiang, J.; Liu, X.; and Ma, J. 2024. Reciprocal transformer for hyperspectral and multispectral image fusion. *Information Fusion*, 104: 102148.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

P, K. D.; and Ba, J. 2014. Adam: A method for stochastic optimization. *Adam: A Method for Stochastic Optimization*.

Ran, R.; Deng, L.-J.; Jiang, T.-X.; Hu, J.-F.; Chanussot, J.; and Vivone, G. 2023. GuidedNet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution. *IEEE Transactions on Cybernetics*, 53(7): 4148–4161.

Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; and Wang, Z. 2016. Real-time single image and video super-resolution using an efficient subpixel convolutional neural network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1874–1883.

Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit neural representations with periodic activation functions. *Neural Information Processing Systems (NeurIPS)*, 33: 7462–7473.

Sun, Y.; Xu, H.; Ma, Y.; Wu, M.; Mei, X.; Huang, J.; and Ma, J. 2023. Dual spatial-spectral pyramid network with transformer for hyperspectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–16.

Tang, J.; Chen, X.; and Zeng, G. 2021. Joint implicit image function for guided depth super-resolution. In *ACM International Conference on Multimedia (ACM MM)*, 4390–4399.

Vivone, G. 2023. Multispectral and hyperspectral image fusion in remote sensing: A survey. *Information Fusion*, 89: 405–417.

Vivone, G.; Dalla Mura, M.; Garzelli, A.; Restaino, R.; Scarpa, G.; Ulfarsson, M. O.; Alparone, L.; and Chanussot, J. 2020. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geoscience and Remote Sensing Magazine*, 9(1): 53–81.

Wang, W.; Zeng, W.; Huang, Y.; Ding, X.; and Paisley, J. 2019. Deep Blind Hyperspectral Image Fusion. In *IEEE/CVF International Conference on Computer Vision (ICCV)*.

Wang, Y.; He, X.; Dong, Y.; Lin, Y.; Huang, Y.; and Ding, X. 2024. Cross-Modality Interaction Network for Pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–16.

Wang, Y.; Lin, Y.; Meng, G.; Fu, Z.; Dong, Y.; Fan, L.; Yu, H.; Ding, X.; and Huang, Y. 2023. Learning high-frequency feature enhancement and alignment for pan-sharpening. In *ACM International Conference on Multimedia (ACM MM)*, 358–367.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612.

Wu, R.-C.; Deng, S.; Ran, R.; Dou, H.-X.; and Deng, L.-J. 2024. INF3: Implicit Neural Feature Fusion Function for Multispectral and Hyperspectral Image Fusion. *IEEE Transactions on Computational Imaging*, 10: 1547–1558.

Xie, Q.; Zhou, M.; Zhao, Q.; Xu, Z.; and Meng, D. 2022. MHF-net: An interpretable deep network for multispectral and hyperspectral image fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44: 1457–1473.

Xu, T.; Huang, T.; Deng, L.; and Yokoya, N. 2022. An Iterative Regularization Method based on Tensor Subspace Representation for Hyperspectral Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–16.

Xu, T.; Huang, T.-Z.; Deng, L.-J.; Xiao, J.-L.; Broni-Bediako, C.; Xia, J.; and Yokoya, N. 2024. A Coupled Tensor Double-Factor Method for Hyperspectral and Multispectral Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–17.

Xu, X.; Wang, Z.; and Shi, H. 2021. Ultrasr: Spatial encoding is a missing key for implicit image function-based arbitrary-scale super-resolution. *arXiv preprint arXiv:2103.12716*.

Zhang, X.; Huang, W.; Wang, Q.; and Li, X. 2020. SSR-NET: Spatial–spectral reconstruction network for hyperspectral and multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7): 5953–5965.

Zhu, C.; Deng, S.; Zhou, Y.; Deng, L.-J.; and Wu, Q. 2023. QIS-GAN: A lightweight adversarial network with quadtree implicit sampling for multispectral and hyperspectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–15.

Zhu, C.; Zhang, T.; Wu, Q.; Li, Y.; and Zhong, Q. 2024. An Implicit Transformer-based Fusion Method for Hyperspectral and Multispectral Remote Sensing Image. *International Journal of Applied Earth Observation and Geoinformation*, 131: 103955.