IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 October 2023, accepted 21 November 2023, date of publication 24 November 2023, date of current version 5 February 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3336684

RESEARCH ARTICLE

TNT++: A Spectral Super-Resolution Method Based on the Entropy of Pathological Images

HUIYUAN ZHANG, ZHAOHUA YANG, (Member, IEEE), ZEYUAN DONG, AND YIJING CHEN
School of Instrument Science and Optoelectronics, Beihang University, Beijing 100191, China

Corresponding author: Zhaohua Yang (yangzh@buaa.edu.cn)

This work was supported in part by the Haidian Transformation Special Fund of the Peking University Third Hospital under Grant HDCXZHKC2022220, and in part by the National Natural Science Foundation of China under Grant 61973018.

ABSTRACT Spectral super-resolution is critical in transforming multispectral images into hyperspectral variants. Its profound importance is evident, yet its adoption in medical imaging reveals a palpable gap. Historically, many networks rely on correlation for grouping spectral bands within the visible light spectrum. However, several medical case images are enriched with information in the near-infrared spectrum, mainly attributed to the near-infrared's ability to penetrate the cellular surface, thereby accessing deeper layers of information. Therefore, grouping from a new perspective is very important. To bridge this gap, we introduce a Spatial-attention Transformer In Spectral-probability Transformer Network (TNT++), explicitly designed to enhance the spectral super-resolution of medical imagery. This methodology is tailored uniquely, drawing upon the inherent pixel statistical properties typical of medical hyperspectral images. Notably, by calculating the entropy value based on the pixel distribution of individual spectral bands, we unveiled the inherent joint spectral entropy patterns in the dataset, introducing an entropy-based grouping and revealing the nuances in image disorder levels—subtleties previously neglected. Our revamped transformer exhibits superior adaptability, proficiently capturing both the spatial and spectral complexities while adeptly navigating the intricacies of image architectures. Rigorous evaluations on the open-source Multidimensional Liver Cancer pathology dataset validate our model's excellence. Outshining six contemporary state-of-the-art (SOTA) techniques across four established metrics, it achieves a PSNR of 31.95dB and an SSIM of 0.9065, marking a significant stride forward in this discipline.

INDEX TERMS Entropy, medical images, spectral super-resolution, transformer.

I. INTRODUCTION

Hyperspectral images find extensive applications across various sectors such as remote sensing, agriculture, mining, military, and more. In the medical domain, based on the electromagnetic properties of hyperspectral images and their interactions with different materials, distinct organs, and tissue molecules exhibit unique response curves to varying spectral wavelengths. Building on this foundational theory, a myriad of spectral imaging systems have been employed to probe biological organ tissues, aiming to glean enhanced medical insights. Such information propels medical advancements, emphasizing the growing importance of biomedical spectral imaging. Among the numerous promising prospects, microscopic cellular hyperspectral images stand out. Unlike conventional RGB images, Microscopic HyperSpectral Images (MHSI) amalgamate both spatial and spectral information and have been successfully deployed in identifying a range of diseases, including brain cancer, oral cancer, skin cancer, prostate cancer, tongue cancer, and cholangiocarcinoma, among others [1], [2], [3], [4], [5], [6]. Such data often furnishes more precise outcomes for segmentation and classification. Tissue cells and molecules reflect differently across the spectrum. However, from a hardware perspective in capturing hyperspectral images, various acquisition methods like Whiskbroom [7], Pushbroom [8], Snapshot [9], and Staring [10] have been proposed. Still, compromises are inherent concerning acquisition costs and

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

other factors. Instruments commonly used for microscopic hyperspectral acquisition are prohibitively expensive, coupled with a convoluted data-gathering process. Hardware constraints often make it challenging to achieve high spatial and spectral resolutions concurrently, limiting the application scope of hyperspectral images. Consequently, there's a growing interest in reconstructing hyperspectral images from conventional RGB images, a technique aptly termed Spectral Super-Resolution (SSR) [11].

SSR techniques primarily bifurcate into data-driven and prior-based approaches [12]. Within the prior-based category, prominent methods encompass Dictionary Learning, Manifold Learning, and Gaussian Process. The hallmark of these methods lies in their capability to reconstruct hyperspectral images even when confronted with limited datasets. Nguyen et al. proposed a method that fits color images through different camera response functions, which are then subjected to white balance adjustments. Subsequently, the reconstruction of hyperspectral images is achieved based on radial functions [13]. Robles-Kelly introduced a hand-crafted inversion algorithm, assimilating insights from changes in light sources and appearance [12]. Arad and Ben-Shahar capitalized on spectral priors to establish a super-complete dictionary for hyperspectral image reconstruction [11]. An enhancement to the Anchored Neighborhood Regression (ANR) combined the merits of ANR and Simple Functions (SF). A+ hinges on the features of ANR and anchored regressors. However, in contrast to learning regressors in the dictionary, it leverages the entirety of the training material, echoing the approach of SF [14]. Methods grounded in priors are often constrained by a handful of known handcrafted features priors, enabling them to restore only superficial features. Their efficacy tends to wane when dealing with contemporary datasets characterized by complexity and rich image structures.

In recent years, with the enrichment of hardware equipment and data, data-driven methods have gradually become the mainstream for hyperspectral reconstruction, for example in textile, agriculture, and remote sensing industries [15], [16], [17], [18]. Convolutional Neural Networks (CNN)-based methods like the HSCNN-D [19] approach integrate path expansion through a dense structure, concatenating the outputs of convolutional kernels across various scales, thereby surpassing the earlier version, HSCNN [20]. Multiscale CNN [21]harnesses spectral similarity to reconstruct hyperspectral images in a multi-granular manner. By synergistically combining the strengths of both 2-D and 3-D convolutional neural networks, it achieves a refined and comprehensive reconstruction, progressing from a coarse to a more detailed representation. The Dense-Unet [18] operates without relying on any prior knowledge, offering an end-to-end reconstruction of hyperspectral images. Attention-based methods like HRnet [22], AWAN [23], and DRCR [24] guide models to pay attention to spectral similarity within images. The HRnet method [22] utilizes Pixel-Shuffle for sampling to retain more information and incorporates the Squeeze and Excitation (SE) channel attention mechanism to feed multi-scale information into the network. The AWAN [23] method, drawing inspiration from the Nonlocal network, introduced the Patch-level Second-order Nonlocal (PSNL) and DRAB dual residual channel attention blocks. Tailored loss functions were devised based on the camera response function. Notably, in the 2020 NTIRE competition, it clinched the first prize in the clean track. DRCR method [24], when symmetrically embedded with the Channel Re-calibration Modules (CRM) module under the UNet structure, is used to learn spectral similarity. It achieved third place in the NTIRE2022 competition. MST++ [25] and MST [26] use transformers to capture long-distance information. However, methods based on CNN are insufficient in capturing long-range information and struggle with complex medical images. Attention-based and transformer-based networks rely on spectral self-similarity. Analyzing spectral images using spectral similarity methods has already been applied in various fields [27]. However, in the domain of SSR, some studies have noted the occurrence of interference between spectral bands, and [28] suggests that networks should be designed based on sensor characteristics. Hang et al. [29] harnessed the interrelation between spectral bands and the inherent association between hyperspectral and RGB images. They introduced a methodology that involves grouped reconstruction and back-mapping tailored for RGB training. A type of model for spectral group recovery based on physical prior design has been proposed. However, These methods often lack a more intuitive basis for categorization. Medical data information is predominantly concentrated in the red and near-infrared spectrum. Conventional grouping strategies calculate correlations and primarily rely on mapping hyperspectral images to RGB priors. However, this approach might be unsuitable for medical image data involving the near-infrared band. Thus, developing a reliable grouping strategy tailored for specific medical data is an urgent issue [30].

Park et al. [31] and Kim et al. [32] have utilized regression analysis for SSR of medical RGB images. The advantage is that less data and just a smartphone are sufficient for SSR. However, linear and polynomial regression can only reconstruct shallow features. Sharma and Hefeeda used residual networks for hyperspectral reconstruction of RGB images of veins [33]. However, due to the limitations of the method, The reconstruction results can still be further improved. Ma et al. [34] proposed an approach based on unsupervised learning to enhance the spectral and spatial resolution of microscopic cell imaging, which is of significant importance. However, there is still a need for spectral scanning equipment to obtain low spatial resolution spectral images. In practical situations, this is more costly compared to obtaining high-resolution RGB images. Ortega et al. [35] proposed a computationally efficient super-resolution method to enhance spatial resolution, which holds tremendous application value. However, this also necessitates the use of spectral imaging equipment to capture low-resolution hyperspectral images. In summary, existing deep learning-based SSR methods have paid less attention to the medical field, and most methods are deployed in daily life scenarios, as well as remote sensing and agricultural imaging datasets [36]. Only a few works have recognized the potential of SSR in the medical field, but these methods are based on expensive snapshot hyperspectral cameras [37].

In light of the advent of the Vision Transformer, a variety of transformer-based models have been proposed in recent years. Among them, Transformer in Transformer (TNT) [38] stands as a superior deep learning model, employing two layers of transformers for visual tasks. In this architecture, each image patch is treated as an individual sequence. The inner transformer performs a secondary segmentation on the image to extract more refined object features. The processed results are then added to the input and subsequently fed into the outer transformer for final output.

The Metaformer [39] posits that the impressive performance of transformers across various computer vision tasks is attributable to the robustness of the transformer architecture itself. To substantiate this claim, the Metaformer replaces the multi-head attention mechanism with simple pooling layers. Experimental results on multiple high-level tasks, such as object detection, semantic segmentation, and image classification, demonstrate that the network retains its competitiveness even with this modification. The presence of various transformer models, particularly TNT and Metaformer, indicate the versatility and robustness of the transformer architecture, thereby making a strong case for its applicability in diverse visual computing scenarios.

To address the aforementioned issues, this study proposes the Spectral-Possibility Transformer and TNT++, which utilize the degree of image disorder for spectral grouping and perform SSR from both spatial and spectral dimensions. Incorporating the advantages of both Metaformer and TNT and delving deeply into the laws of image entropy, this study aims to address the aforementioned issues. We introduce the Spectral-probability Transformer and TNT++, which leverage the degree of disorder in images for spectral grouping and calculating the spectral similarity matrix. These models perform SSR from both spatial and spectral dimensions.

1) This study established a theoretical model for microscopic hyperspectral reconstruction and analyzed the regularity of the level of disorder in hyperspectral images based on entropy values. Moreover, to the best of our knowledge, there is limited research on reconstructing microscopic hyperspectral images from microscopic RGB images.

2) This study customizes a new transformer for medical datasets. Thanks to prior analysis, it possesses enhanced generalization capabilities and reduced computational complexity.

3)This study introduces entropy-based grouping. Distinct from previous grouping methods, entropy-based grouping neither relies on hardware information nor on the prior mapping from hyperspectral to RGB. It can be achieved with straightforward computation. To the best of our knowledge, this marks the inaugural application of information entropy in this domain.

4) The TNT++ proposed in this study surpasses the 6 advanced methods of RGB reconstruction of microscopic hyperspectral images. It reconstructs the microscopic hyperspectral images of 30 bands in the 550nm-1000nm range from the RGB image.

The remainder of this paper primarily introduces the proposed Probability Multi-head Self-Attention (P-MSA) mechanism and the model structure, presenting the reconstruction results in three ways. Ablation experiments validate the effectiveness of the proposed module and grouping strategy

II. RELATED WORK

A. HISTOPATHOLOGY DATA

Histopathology holds tremendous potential value for disease diagnosis. However, the majority of existing histopathology datasets focus on RGB images. Such RGB images limit the capacity of deep learning models to fully harness their information utilization capabilities. According to the literature [5], this dataset is among the few hyperspectral microscopy datasets available. Furthermore, the dataset provides detailed annotations. Moreover, as provided in the literature, near-infrared can penetrate the cellular surface [40], accessing deeper layers of information, a feature frequently observed in case images. The microscopy cell imaging and dataset acquisition process is illustrated in Figure 1, as highlighted by the yellow box depicting the microscopy imaging system. All these aspects underscore the considerable potential and value of this dataset for spectral super-resolution research.

B. PROBLEM DESCRIPTION

The microscopic image, illuminated by a 550-1000nm light source, is transmitted to the color CCD. Its photosensitive elements collect the light intensity, and the Bayer filter records the blue, red, and green components. These color components are then read out as numerical values. $\mathbf{L}_\lambda$ represents the intensity distribution of the halogen lamp light source wavelength, $\mathbf{M}_\lambda$ represents the reflectance of the microcellular surface after staining with He dye, and $\mathbf{B}_c^\lambda$ as the SRF (Spectral Response Function).

$$\mathbf{I}_{(H\times W\times C)} = \int_\lambda \mathbf{L}_\lambda \mathbf{M}_\lambda \mathbf{B}_c^\lambda \tag{1}$$

In a general sense, $\mathbf{X}_{(H\times W\times \lambda)}$ can be considered as the real MHSI, $\mathbf{X}_{(H\times W\times \lambda)}$ is obtained by multiplying $\mathbf{L}_\lambda$ with $\mathbf{M}_\lambda$.

$$\mathbf{I}_{(H\times W\times C)} = \int_\lambda \mathbf{X}_{(H\times W\times \lambda)} \mathbf{B}_c^\lambda \tag{2}$$

In discrete cases, equation (1) degenerates to equation (2), where n is the index of the discrete representation of the wavelengths. Based on the spectral resolution of the dataset, the spectral band step size is obtained as 7.5nm.

$$\mathbf{I}_{(H\times W\times C)} = \sum\nolimits_n \mathbf{X}_{(H\times W\times \lambda_n)} \mathbf{B}_c^{\lambda_n} \tag{3}$$

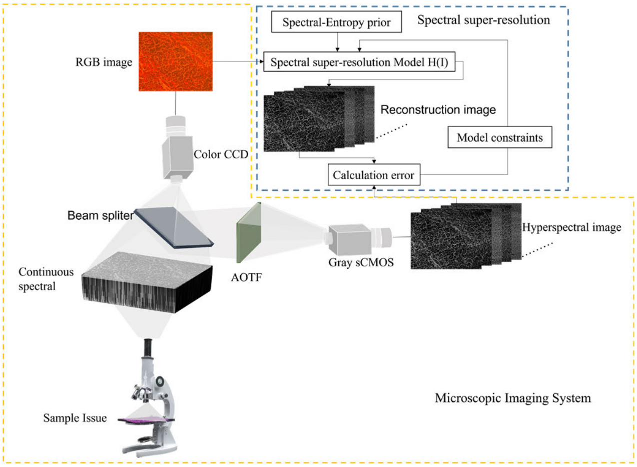

FIGURE 1. **Microscopic hyperspectral image acquisition process and spectral super-resolution.**

To simplify the expression in equation (3). where $\mathbf{I} \in \mathbb{R}^{HW\times C}$ denotes the vectorial representation of RGB image, $\mathbf{X} \in \mathbb{R}^{HW\times \lambda_n}$ represents the corresponding HSI, and $\mathbf{B} \in \mathbb{R}^{\lambda_n \times C}$ is the SRF. Lens noise $\mathbf{D} \in \mathbb{R}^{HW\times C}$ is also incorporated to simulate a real camera shooting scenario.

$$\mathbf{I} = \mathbf{XB} + \mathbf{D} \tag{4}$$

As can be seen, super-resolving hyperspectral images from RGB images are a severe ill-posed problem. According to the pioneering work of Arad and Ben-Shahar [11], the feasibility of reconstructing hyperspectral images from RGB can be categorized under the following two assumptions:

1) The spectral signal of the sensor is restricted to a low-dimensional form in high-dimensional space.

2) Within this low-dimensional form, the frequency of isochromatic heterogeneities is very low.

As illustrated in Fig.1, this reconstruction satisfies the aforementioned two conditions. Hyperspectral images and microscopic color images are obtained under the same light source. Through Acousto-Optic Tunable Filter(AOTF), the continuous spectral signal is reduced to discrete spectra. As the observed samples are all microscopic cells stained with H&E dye, there are relatively few instances of same-color-different-spectrum.

C. THE ENTROPY OF MHSI

Entropy is often used to measure the level of disorder in an image [41]. Based on this, the pixel distribution of a single-band image is computed. The index i indicates the statistical probability of pixels of size i, and the index j refers to the j th spectral segment. The notation $\mathbf{P}_{ij}^{\mathbf{X}} \in \mathbb{R}^{\lambda_n \times 256}$ denotes the matrix of pixel distributions across all spectral bands, and $\mathbf{v}^{\mathbf{X}} \in \mathbb{R}^{\lambda_n \times 1}$ denotes the vector of entropy values for each spectral band. count($\mathbf{X}_{(H\times W\times \lambda_n)}$) is used to calculate the pixel distribution of each spectral band.

$$\mathbf{P}_{ij}^{\mathbf{X}} = \frac{\mathrm{count}(X_{(H\times W\times \lambda_n)})}{HW} \tag{5}$$

$$\mathbf{v}^{X} = -\sum_{i=0}^{255} \mathbf{P}_{ij}^{\mathbf{X}} \log_2 \mathbf{P}_{ij}^{\mathbf{X}} \tag{6}$$

As illustrated in Fig.2. The middle bands 6 to 20, i.e., the red bands, are complex and relatively chaotic, while the entropy values of the early bands 21 to 30 and late bands 1 to 5 are lower, and the entropy distribution of the late bands is concentrated. This regularity can be used to solve the problem of the number of groups in group recovery.

Various molecules within cellular tissue (such as proteins, lipids, nucleic acids, etc.) possess corresponding absorption and emission spectra, resulting in significant differences in results across different spectral bands. The reconstruction model, denoted as H, is considered. The input is the RGB image I, and the reconstruction result is$\hat{\mathbf{X}}$.

$$\hat{\mathbf{X}} = \mathrm{H}(\mathbf{I}) \tag{7}$$

S(X) represents the prior constraint of the entropy range, serving as the prior knowledge of the model. When the entropy of different bands in $\mathbf{X}$ exceeds the reward range, it returns to penalize the loss of the network.

Compared to previous direct computations of the norms between the reconstructed result $\hat{\mathbf{X}}$ and the sample $\mathbf{X}$ the introduction of statistical measures $\mathbf{P}_{ij}^{\mathbf{X}}$ and $\mathbf{P}_{ij}^{\hat{\mathbf{X}}}$ can be viewed as an entropy convex optimization problem. The first term represents the distance between the reconstructed distribution and the sample distribution, calculating the spectral aggregation feature differences. The second term calculates the difference in entropy values between the two, indicating the disparity in their levels of disorder. This first term is used to design the new input for the transformer.

$$\mathrm{S}\left(\mathbf{X},\hat{\mathbf{X}}\right) = \sum_{i,j} \mathbf{P}_{ij}^{\hat{\mathbf{X}}} \log \frac{\mathbf{P}_{ij}^{\hat{\mathbf{X}}}}{\mathbf{P}_{ij}^{\mathbf{X}}} + \sum \left| \mathbf{v}^{\mathbf{X}} - \mathbf{v}^{\hat{\mathbf{X}}} \right| \tag{8}$$

The process of solving model H is a constrained optimization process. $\mathrm{L}\left(\mathbf{X},\hat{\mathbf{X}}\right)$ represents the prior knowledge concerning the number of spectral groups. $\mathbf{R}_g$ indicates the grouping of the matrix by channels, as shown by the entropy prior principle in Fig.2 where g is the grouping index, and G is the total number of groups. The notation $\|\|_F$ represents the computation of the discrepancy between the reconstruction result and the true result.

$$\mathrm{L}\left(\mathbf{X},\hat{\mathbf{X}}\right) = \sum_{g=1}^{G} \left\| \mathbf{R}_g(\mathbf{X}) - \mathbf{R}_g\left(\hat{\mathbf{X}}\right) \right\|_F \tag{9}$$

Meanwhile, α and λ serve as the influencing coefficients. Therefore, the problem is formulated as follows:

$$\mathrm{H} = \arg\min_{\mathrm{H}} \left\| \mathbf{X} - \hat{\mathbf{X}} \right\|_F + \alpha \mathrm{L}\left(\mathbf{X},\hat{\mathbf{X}}\right) + \lambda \mathrm{S}\left(\mathbf{X},\hat{\mathbf{X}}\right),$$
$$s.t.\ \hat{\mathbf{X}} = \mathrm{H}(\mathbf{I}) \tag{10}$$

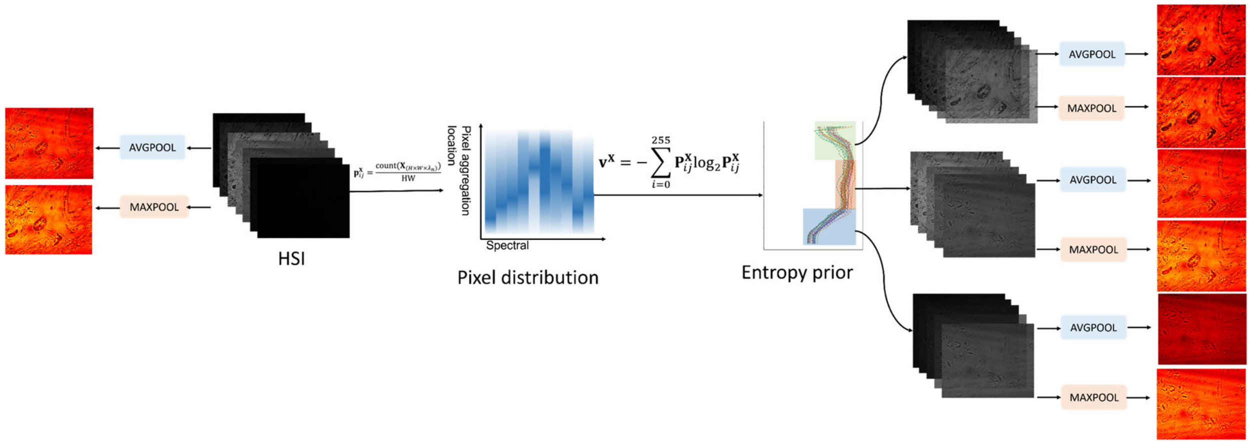
AVGPOOL
MAXPOOL
HSI
$\mathbf{p}_{ij}^{\mathbf{X}} = \frac{\mathrm{count}(\mathbf{X}_{(H\times W\times\lambda_n)})}{\mathrm{HW}}$
Pixel aggregation location
Spectral
Pixel distribution
$\mathbf{v}^{\mathbf{X}} = -\sum_{i=0}^{255} \mathbf{P}_{ij}^{\mathbf{X}} \log_2 \mathbf{P}_{ij}^{\mathbf{X}}$
Entropy prior
AVGPOOL
MAXPOOL
AVGPOOL
MAXPOOL
AVGPOOL
MAXPOOL

FIGURE 2. Comparison of entropy-based pooling versus global pooling. Global pooling can extract spatial information with higher entropy values but lacks representation for spatial information with lower entropy values. This limitation primarily arises because global pooling captures scenarios where information frequencies are high, overlooking representations when information frequencies are low. After grouped pooling, spatial information during low entropy values is preserved, providing a reliable prior for subsequent local information modeling.

D. ATTENTION MECHANISM AND TRANSFORMER

The flexible use of attention mechanisms in image segmentation, classification, super-resolution, and many other computer vision fields has gained great attention. The attention mechanism is combined with the design of the human cognitive process so that the model emphasizes the attention to different details of the image. Based on this, the transformer architecture further developed in recent years has a stronger ability to process sequences. According to different tasks, researchers develop different attention mechanism strategies and embed them in transformer. For example, swin-transformer based on Window Multi-head Self-Attention (W-MSA) uses different Windows as headers. Channel Multi-head Self-Attention (C-MSA) for channel attention mechanisms is a Restromer developed with a single channel as the header.

III. PROPOSED METHOD

Multi-level structures process feature maps of different scales, thus avoiding the spatial information loss caused by Unet-like methods [30]. The inspiration for this work comes from TNT [38] and Metaformer [39]. On one hand, we combine the internal and external reconstruction of images from TNT, and on the other hand, Metaformer proposes a general architecture for transformers. We leverage the idea of a general framework to enhance the performance.

capability of the spatial attention mechanism [38] The inner transformer adopts spatial attention, is capable of modeling local information [42], and uses residual connections [43] to avoid gradient vanishing. There exist two conflicting viewpoints in the SSR field [30], i.e., spectral similarity [25], [29], and interference among different spectral bands [21], [44]. To address this issue, the inner transformer performs group recovery based on the level of entropy disorder, with the number of groups set to 3. The outer transformer performs channel-weighted reconstruction, and the design of the P-MSA considers pixel distribution to generate $\mathbf{Q}_j$ and $\mathbf{K}_j$. By considering the degree of feature map disorder, channel dimension information is adjusted again.

Based on Fig 2 and equation (9), group pooling will be applied to the internal transformer reconstruction, and as analyzed from the visualization results in Fig 2, different group pooling can better retain multi-dimensional information. Furthermore, according to the entropy prior curve on the right of Fig 2, and equations (8) and (6), the pixel distribution possesses a degree of similarity, therefore, weighting scores should be assigned based on the pixel distribution, which can be computed according to equation (10). The model achieves the optimal solution under several prior constraint scenarios.

A. NETWORK ARCHITECTURE

In equation (9), the variable i represents the scale ratio at the current branch of the multi-scale network. The pixel unshuffle operation is defined as the channel concatenation of all possible shifted versions. As shown in Fig 3(a)

$$\mathbf{A}^{(\frac{H}{i}\times\frac{W}{i}\times 4^{i}*C)} = \mathrm{Pixelunshuffle}_{i}\left(\mathbf{A}_{in}^{(H\times W\times C)}\right) \tag{11}$$

we represent $\mathbf{A}^{(\frac{H}{i}\times\frac{W}{i}\times 4^{i}*C)}$ as $\mathbf{A}^{i\downarrow}$. The index i represents the network sampling scaling ratio. Feature maps from different sampling layers will pass through TNT++, and the results of different scales will be up-sampled and concatenated at different levels. The inputs from these different levels will then go through TNT++ again, facilitating cross-scale feature fusion. We denote TNT++ as T.

$$\mathbf{B}_{i} = \mathrm{T}\left(\mathbf{A}^{i\downarrow}\right) \tag{12}$$

Pixel-shuffle is an up-sampling method, and $\mathbf{B}_i$ represents the result after the i-th sampling feature map passes through the backbone.

$$\mathbf{C}_{2} = \mathrm{Concat}[\mathrm{Pixelshuffle}\left(\mathbf{B}_{3}\right), \mathbf{B}_{2}] \tag{13}$$

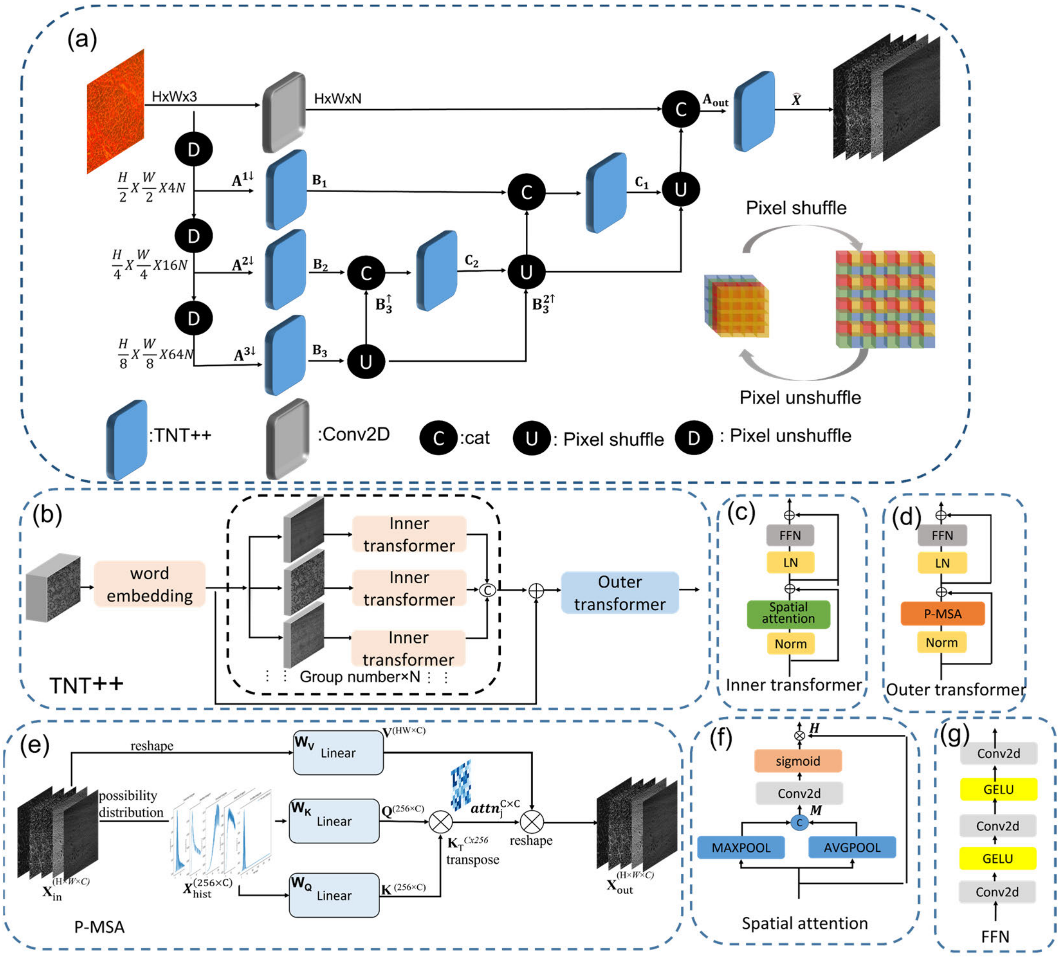

FIGURE 3. **(a) Overall architecture of our model (b) Composition of TNT++ (c) Inner transformer (e) Structure of a single-head P-MSA (f) Spatial structure for complex understanding of space similarity (g) Feedforward network The entropy grouping strategy is employed to compute spatial contextual information. The external transformer takes into account pixel aggregation characteristics and utilizes a linear layer to adjust the aggregation distribution features, aligning them closer to the actual distribution.**

we represent Pixelshuffle ($\mathbf{B}_3$) as $\mathbf{B}_3^{2\uparrow}$ and so on.

$$\mathbf{C}_1 = \text{Concat}\left(\mathbf{B}_3^{2\uparrow}, \mathbf{C}_2^{\uparrow}, \mathbf{B}_1\right) \tag{14}$$

Finally, the previous features are concatenated and combined with the input to obtain the result, and the multi-channel features are fused through the convolutional layer.

$$\mathbf{A}_{\text{Out}} = \text{Conv2d}\{\text{Concat}[\mathbf{B}_3^{3\uparrow}, \mathbf{C}_2^{2\uparrow}\mathbf{C}_1^{\uparrow}, \text{Conv2d}(\mathbf{A}_{\text{in}})\ ]\} \tag{15}$$

B. INNER TRANSFORMER

Based on Fig.2 and Equation (9), the internal transformer will conduct entropy grouping, completing the preliminary grouping reconstruction. The input is divided into groups to obtain $\mathbf{X}_{in}$, and spatial weights are generated by processing single-group information through average pooling and max pooling. As shown in Fig 3(c)

$$\mathbf{X}_{\text{in}} = \mathbf{R}_{\text{g}}\left(\text{Conv2d}(\mathbf{A})\right) \tag{16}$$

$$\mathbf{M} = \mathbf{Concat}\left[\text{Maxpool}\left(\mathbf{X}_{\text{in}}\right), \text{Avgpool}\left(\mathbf{X}_{\text{in}}\right)\right] \tag{17}$$

Subsequently, the spatial attention matrix is normalized. As shown in Fig 3(d)

$$\mathbf{Attn}^{\frac{H}{i}X\frac{W}{i}} = \text{sigmoid}\left[\text{Conv2d}\left(\mathbf{M}\right)\right] \tag{18}$$

The spatial attention matrix is subjected to a Hadamard product with the input, and the result is propagated forward, yielding the output of the inner transformer.

$$\mathbf{H} = \mathbf{X}_{in} \cdot \mathbf{Attn} + \mathbf{X}_{in} \tag{19}$$

$$\text{Inner}\left(\mathbf{A}_i\right) = \text{Gelu}\{\text{Conv2d}[\text{Layernorm}\left(\mathbf{H}\right)\ ]\}) + \mathbf{H} \tag{20}$$

C. OUTER TRANSFORMER

Based on equations (5) and (8), the computation of pixel distribution difference will be carried out through P-MSA.The result of the inner transformer is normalized first. Then, for the P-MSA, the distribution of Xin is initially evaluated through pixel statistics, and $\mathbf{Q}_j$ and $\mathbf{K}_j$ are calculated through

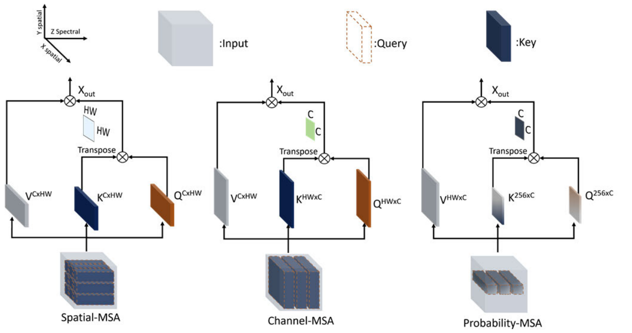

FIGURE 4. Comparison of three self-attention mechanisms, Compared to traditional approaches that directly use features as queries and keys, computing through pixel probability distribution offers lower computational complexity.

a linear layer. As shown in Fig 3(d) and Fig 3(e)

$$\mathbf{X}_{count}^{(256\times C)} = \frac{\text{count}\left(\mathbf{X}_{in}^{(H\times W\times C)}\right)}{HW} \tag{21}$$

$$\mathbf{Q}^{256\times C}, \mathbf{K}^{256\times C} = \text{Linear}\left[\mathbf{X}_{count}^{(256\times C)}\right] \tag{22}$$

$$\mathbf{V} = \mathbf{Linear}\left[\text{reshape}\left(\mathbf{X}_{in}^{(H\times W\times C)}\right)\right] \tag{23}$$

The attention matrix is calculated based on the $\mathbf{Q}_j$ and $\mathbf{K}_j$ values obtained through the pixel distribution computation.

$$\mathbf{Attn}_j^{C\times C} = \text{softmax}\left(\mathbf{K}_j^T \mathbf{Q}_j\right) \tag{24}$$

$$\mathbf{X}_{out} = \text{Concat}\left(\mathbf{V}_j \cdot \mathbf{Attn}_j^{C\times C}\right) \tag{25}$$

D. COMPARISON OF DIFFERENT SELF-ATTENTION

The following are simplified diagrams of three types of attention, excluding tricks such as position encoding and masking. The focus is on the generation of attention and the result output computation, which are the parts with the greatest differences in complexity.

As shown in Fig 4, Spatial-MSA (S-MSA) utilizes global information to generate Q, K, creating spatial information weights. Channel-MSA (C-MSA) employs channel information for generating **Q**, **K**, producing channel weights. On the other hand, P-MSA solely requires pixel distribution probability information to generate **Q**, **K**, also yielding channel weights.

The most significant difference in complexity calculation lies in two matrix multiplications. The complexity of the spatial multi-head attention mechanism can be obtained from the [25].

$$\text{O}\left(\text{S}-\text{MSA}\right) = 2C\left(HW\right)^2 \tag{26}$$

The computational complexity of the spectral multi-head attention mechanism can be obtained from [25] and [45]. N is considered as the number of heads in the multi-head attention mechanism.

$$\text{O}\left(\text{C}-\text{MSA}\right) = 2\frac{C^2 HW}{N} \tag{27}$$

TABLE 1. Computational complexity of different attention mechanisms.

Complexity	S-MSA	C-MSA	P-MSA
Params	68080	59980	**45720**
Giga FLOPs	1.89	0.06	**0.02**

The computational complexity of the P-MSA mechanism can be calculated as follows.

$$\mathbf{O}\left(\mathbf{P}-\mathbf{MSA}\right) = \frac{C^2 HW}{N} + \frac{C^2 256}{N} \tag{28}$$

It can be observed that, when the spatial resolution exceeds 256, the complexity of the channel multi-head attention mechanism and spatial multi-head attention mechanism will be higher than the complexity of the pixel distribution multi-head attention mechanism. Additionally, the P-MSA encapsulates spatial dimension information. Therefore, for images with different structures but the same pixel distribution, the calculated self-similarity matrix will remain consistent, showcasing better generalization capabilities. TABLE 1 computes the computational complexity of different multi-head attention mechanisms.

IV. EXPERIMENT

A. SETTING

The hyperspectral dataset used in this study is sourced from [5], with the original image dimensions being 1024 × 1280. The RGB image is synthesized through SRF and augmented with lens noise to simulate a real-world dataset. The data splitting strategy employed in this study ensures a fair distribution of instances across the training, testing, and validation sets to mitigate the risk of inter-patient biases. Specifically, a stratified splitting approach was adopted, where the data was divided in a manner that maintains the same distribution of patient characteristics (e.g., age, condition severity, etc.) across all three sets. This strategy ensures that each set—training, testing, and validation—is representative of the overall patient population, thus providing a reliable basis for evaluating the model's performance without overestimation. The partitioning was performed such that 80% of the data was allocated to the training set, and 10% each to the testing and validation sets, while adhering to the stratified splitting criteria to maintain the consistency in patient representation across the sets.

This study employs a patch-based training strategy. The image patch size is set at 256 × 256, the feature extraction size is 30, and the batch size is 12. The learning rate is configured at 0.0001 and employs a cosine annealing strategy for decay over 200 epochs to ensure the model's convergence on the test set. All experiments were conducted on an NVIDIA RTX 4090, with 128G RAM, and a CPU of i7-13700KF. The results are shown in Table 2, where the latency refers to the time required to compute for images with a resolution of 1024 × 1280.

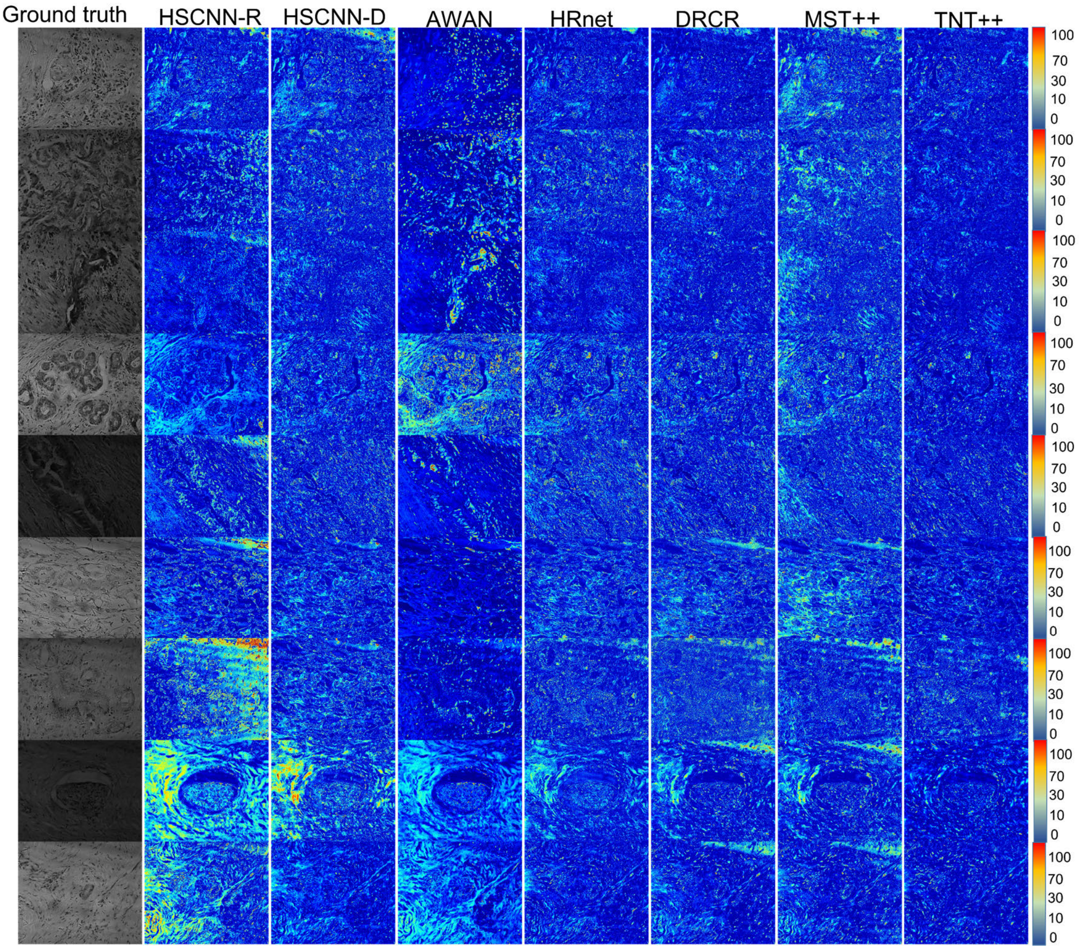

FIGURE 5. Compared with the reconstruction results of different SOTA methods for nine targets, the error heatmap calculated by MRAE is displayed, with the error bar chart on the far right. Our TNT++ algorithm exhibits superior global reconstruction performance.

B. LOSS FUNCTION

Unlike previous approaches, although we use Mean Relative Absolute Error (MRAE) for training, we add a small constant of 1e-6 to the denominator to prevent gradient explosion. MRAE [19] allows the model to converge faster and provides a better understanding of the discrepancy between the image and reality.

$$\mathrm{LOSS}\left(\mathbf{X},\hat{\mathbf{X}}\right)=\frac{1}{HW}\sum_{i=0}^{H}\sum_{j=0}^{W}\frac{\left|\mathbf{X}(i,j)-\hat{\mathbf{X}}(i,j)\right|}{\mathbf{X}\,(i,j)+\varepsilon} \tag{29}$$

C. RESULT AND PERFORMANCE METRICS

MRAE, is used to calculate the average pixel deviation between the true hyperspectral image and the reconstructed hyperspectral image. It is also commonly used to represent overall reconstruction effects and error heatmaps, as shown in Fig.5.

$$\mathrm{MRAE}\left(\mathbf{X},\hat{\mathbf{X}}\right)=\frac{1}{HW}\sum_{i=0}^{H}\sum_{j=0}^{W}\frac{\left|\mathbf{X}(i,j)-\hat{\mathbf{X}}(i,j)\right|}{\mathbf{X}\,(i,j)} \tag{30}$$

RMSE (Root Mean Square Error) serves as a metric to gauge the discrepancy within individual spectral bands. In this context, we compute the difference between the reconstructed single $\mathbf{X}$ and $\hat{\mathbf{X}}$ itself. Ultimately, the average is taken from the sum of the RMSE values across all hyperspectral bands. The smaller the RMSE and MRAE values, the better the reconstruction quality

$$\mathrm{RMSE}\left(\mathbf{X},\hat{\mathbf{X}}\right)=\sqrt{\frac{1}{HW}\sum_{i=0}^{H}\sum_{j=0}^{W}\left(\mathbf{X}(i,j)-\hat{\mathbf{X}}(i,j)\right)^{2}} \tag{31}$$

PSNR (Peak Signal-to-Noise Ratio) is derived from RMSE. The higher the PSNR value, the better the reconstruction quality.

$$\mathrm{PSNR}\left(\mathbf{X},\hat{\mathbf{X}}\right)=20\times\log_{10}\left(\frac{\max\,(\mathbf{X})}{\mathrm{RMSE}\left(\mathbf{X},\hat{\mathbf{X}}\right)}\right) \tag{32}$$

SSIM (Structural Similarity Index) measures the degree of structural similarity and ranges between 0 and 1. A value closer to 1 indicates better quality, while a value closer to

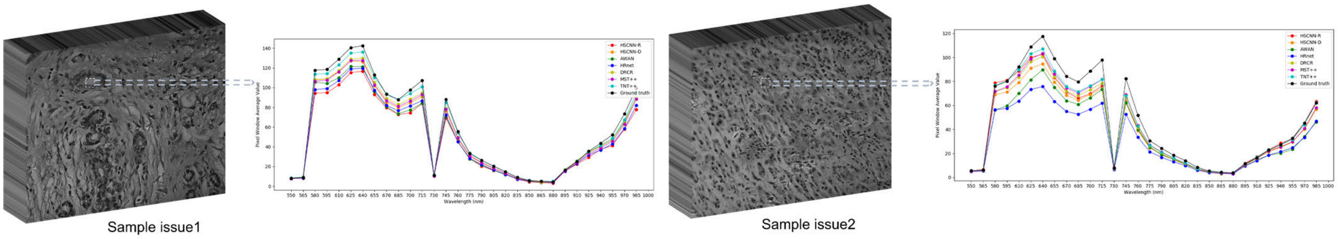

FIGURE 6. Wavelength-pixel size curve for a certain window in two samples, with the true value represented in black.

TABLE 2. Comparison with several SOTA methods.

Method	MRAE	RMSE	PSNR	SSIM	Params(Gflops)	**Traning** time(h)	Latency(s)
HSCNN-R	0.1808	0.04709	26.86	0.8358	2.87	25.12	13.37
HSCNN-D	0.1269	0.03112	30.33	0.8773	4.65	49.36	16.27
AWAN	0.1212	0.03123	30.26	0.8782	4.04	76.32	13.54
HRnet	0.1196	0.03080	30.37	0.8857	31.70	37.94	22.07
DRCR	0.1204	0.02903	30.53	0.8815	13.2	64.21	14.89
MST++	0.1155	0.02870	31.04	0.8867	**1.62**	31.68	11.22
TNT++	**0.1030**	**0.02568**	**31.95**	**0.9065**	3.82	**21.93**	**10.87**

TABLE 3. Different spectral group results.

Spectral Group	MRAE	RMSE	PSNR	SSIM
1-30	0.1118	0.03016	30.63	0.8867
1-15,16-30	0.1087	0.02796	31.22	0.8951
1-10,11-20,21-30	0.1051	0.02612	31.78	0.9014
1-8,9-15,16-23,23-30	0.1065	0.02720	31.44	0.8988
1-6,7-12,13-18,19-24,24-30	0.1079	0.02738	31.41	0.8971

0 signifies poorer quality.

$$\mathrm{SSIM}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{(\mu_x \mu_{\hat{x}} + C_1)\,(2\sigma_{x\hat{x}} + C_2)}{\left(\mu_x^2 + \mu_{\hat{x}}^2 + C_1\right)\left(\sigma_{\hat{x}}^2 + \sigma_x^2 + C_2\right)} \tag{33}$$

D. ABLATION STUDY

To demonstrate the reliability of our spectral group prior, we also conducted ablation experiments on different spectral groupings. The results are shown in Table 3. The outcomes from various grouping strategies are found to be inferior to those of entropy grouping, underscoring the reliable effectiveness of entropy-based partitioning. Furthermore, splitting evenly three times produced the most favorable results relative to other grouping approaches. When there are too many groups, the performance deteriorates because generalization decreases, making it more challenging to fit a uniform pattern.

To demonstrate the effectiveness of the internal and external transformers proposed in this study, we observed the experimental results after separately eliminating the internal and external components. The results are shown in Table 4. Even without inner and outer situations, this structure can still produce good results. It can be deduced that the joint TNT structure is meaningful for maintaining the reconstruction results, and reconstructing both inner and outer channels in TNT enables a more refined reconstruction outcome. The absence of either inner or outer transformers results in a decrease in performance, and the absence of the outer transformer has a more significant impact. It can be concluded that both outer transformers contribute significantly to the performance improvement.

TABLE 4. The impact of inner transformer and outer transformer on result reconstruction. 'without' is denoted as 'Wo'.

Strategy	MRAE	RMSE	PSNR	SSIM
Wo inner	0.1158	0.02883	30.92	0.8916
Wo outer	0.1167	0.02914	30.84	0.8912
Wo inner and outer	0.1326	0.03328	29.75	0.8732

We computed the variations in the results caused by adopting different pairings of multi-head attention mechanisms. The S-MSA and C-MSA structures maintain consistency with the attention mechanisms referenced in Section II-D. The results are shown in Table 5. It is evident from the two comparative experiments that incorporated P-MSA, the final performance improved. This suggests a superior generalization ability.

V. DISCUSSION

In instances where there are pronounced similarity differences between different wavelengths, the performance of reconstruction sees a decline. Also, the model's ability to generalize diminishes when the number of groups is excessive. Conversely, with fewer groups, interference occurs among the relationships between different wavelengths. Offering a fresh viewpoint on spectral grouping, this study is centered around entropy prior grouping. This approach is capable of

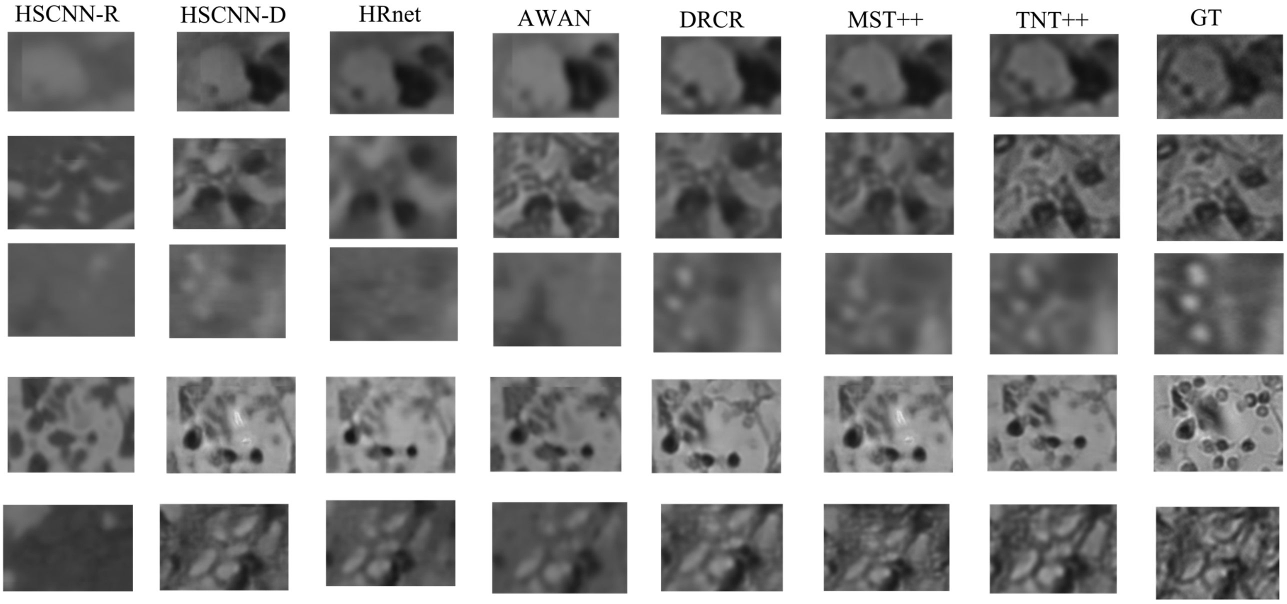

FIGURE 7. For detailed reconstruction results, it can be observed that TNT++ is closer to reality in terms of detail. In microscopic cell data, different spectral bands penetrate the cell surface, resulting in significant differences in inter-spectral detail.

TABLE 5. The influence of internally and externally nested multi-head attention mechanisms on the results.

C-MSA	S-MSA	P-MSA	MRAE	RMSE	PSNR	SSIM
×	√	√	0.1133	0.02835	31.11	0.8916
√	×	√	0.1163	0.02957	30.84	0.8857
√	√	×	0.1203	0.03035	30.48	0.8815

delivering a relatively reliable grouping prior without the need for specific hardware information. Embedding hyperspectral joint entropy prior to network design is worth trying in multiple domains. Especially for advanced visual tasks like classification, segmentation, etc., in hyperspectral images, extending the application of information entropy evidently enables further exploitation of information. The level of disorder provides an alternative perspective for the identification of different entities, given that each entity possesses distinct entropy values. Moreover, as pixel statistics are performed, image enhancement and transformation do not significantly impact it, which also contributes to the accelerated learning rate of this model. This also bodes well for better integration into several other domains.

In the case of hyperspectral classification, different features have distinct pixel distributions across different bands. Correspondingly, for the same feature with different spatial representations, can be classified as the same feature and further generalized. Due to the complex and intricate variations in medical images, a higher-dimensional statistical experience is required for reconstruction. P-MSA enables weight learning in the pixel statistical dimension, resulting in faster learning and an improved reconstruction process. RGB and hyperspectral images are obtained under the same light source. This is an ideal data acquisition environment for the field of SSR. Our algorithm holds positive significance for the advancement of spectroscopy and tissue optics.

VI. LIMITATION

Although integrating pixel distribution and entropy priors in the reconstruction process yields improved results, it relies heavily on the assimilation of additional prior information for segmentation. This leads to an increased necessity for pre-processing the data. A fully automated analysis that can reliably segment information and offer a dependable segmentation process can further enhance the accuracy. This direction of self-reliable segmentation will be pivotal for future research and holds significant implications for large multi-modal models. The development of real-time spectral super-resolution reconstruction algorithms is meaningful for practical applications. Current algorithms can achieve good reconstruction results, but they require extended reconstruction times, making them unsuitable for real-time observation processes.

VII. CONCLUSION

Acquiring MHSI is challenging and comes with complex conditions. SSR has the potential to further extend the applicability of hyperspectral analysis in medical imaging. Most existing SSR algorithms overlook their utility in the medical field. The medical domain demands higher accuracy for SSR, yet also offers more prior information. Our proposed TNT++ method, based on information entropy, not only achieves higher accuracy but also speeds up model convergence and reduces computational complexity through entropy grouping and P-MSA. This is primarily due to the consistent pixel distribution across different images, despite their varying spatial characteristics. Unlike previous research, we have reconstructed images from three similar perspectives: spectral similarity within a single image, spatial similarity, and similarity in complexity or entropy across different images. For medical data, the entropy patterns and pixel distribution enabled by this method allow the network to better understand and learn pathological cell data. Additionally, the introduction of priors for object categories in super-resolution ensures that different objects have distinct entropy priors, leading to improved model performance. In summary, the proposed method outperforms previous approaches due to the incorporation of P-MSA, which enables the model to recognize dataset characteristics, and the well-arranged combination of group reconstruction.

We have validated our method on multi-dimensional gallbladder and liver cancer medical datasets, an area often overlooked in previous studies. Our approach effectively reduces super-resolution errors and yields better image quality.

REFERENCES

[1] H. Fabelo, "Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations," *PLoS ONE*, vol. 13, no. 3, Mar. 2018, Art. no. e0193721.

[2] J.-R. Duann, C.-I. Jan, M. Ou-Yang, C.-Y. Lin, J.-F. Mo, Y.-J. Lin, M.-H. Tsai, and J.-C. Chiou, "Separating spectral mixtures in hyperspectral image data using independent component analysis: Validation with oral cancer tissue sections," *J. Biomed. Opt.*, vol. 18, no. 12, Dec. 2013, Art. no. 126005, doi: 10.1117/1.jbo.18.12.126005.

[3] H. Akbari, L. V. Halig, D. M. Schuster, A. Osunkoya, V. Master, P. T. Nieh, G. Z. Chen, and B. Fei, "Hyperspectral imaging and quantitative analysis for prostate cancer detection," *J. Biomed. Opt.*, vol. 17, no. 7, Jul. 2012, Art. no. 0760051, doi: 10.1117/1.jbo.17.7.076005.

[4] Z. Liu, H. Wang, and Q. Li, "Tongue tumor detection in medical hyperspectral images," *Sensors*, vol. 12, no. 1, pp. 162–174, Dec. 2011. [Online]. Available: https://www.mdpi.com/1424-8220/12/1/162

[5] Q. Zhang, Q. Li, G. Yu, L. Sun, M. Zhou, and J. Chu, "A multidimensional choledoch database and benchmarks for cholangiocarcinoma diagnosis," *IEEE Access*, vol. 7, pp. 149414–149421, 2019, doi: 10.1109/ACCESS.2019.2947470.

[6] Y. Ji, C. Jones, Y. Baek, G. K. Park, S. Kashiwagi, and H. S. Choi, "Near-infrared fluorescence imaging in immunotherapy," *Adv. Drug Del. Rev.*, vol. 167, pp. 121–134, Dec. 2020, doi: 10.1016/j.addr.2020.06.012.

[7] J. E. Fowler, "Compressive pushbroom and whiskbroom sensing for hyperspectral remote-sensing imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 684–688.

[8] K. C. Lawrence, B. Park, W. R. Windham, and C. Mao, "Calibration of a pushbroom hyperspectral imaging system for agricultural inspection," *Trans. ASAE*, vol. 46, no. 2, p. 513, 2003.

[9] R. T. Kester, N. Bedard, L. Gao, and T. S. Tkaczyk, "Real-time snapshot hyperspectral imaging endoscope," *J. Biomed. Opt.*, vol. 16, no. 5, 2011, Art. no. 056005.

[10] N. Gupta, "Development of staring hyperspectral imagers," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2011, pp. 1–8.

[11] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. Eur. Conf. Comput. Vis.*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 19–34.

[12] A. Robles-Kelly, "Single image spectral reconstruction for multimedia applications," in *Proc. 23rd ACM Int. Conf. Multimedia*, Oct. 2015, pp. 251–260.

[13] R. M. Nguyen, D. K. Prasad, and M. S. Brown, "Training-based spectral reconstruction from a single RGB image," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, Sep. 2014, pp. 186–201.

[14] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vis.* Singapore, Nov. 2014, pp. 111–126.

[15] A. Gkillas, D. Kosmopoulos, and K. Berberidis, "Cost-efficient coupled learning methods for recovering near-infrared information from RGB signals: Application in precision agriculture," *Comput. Electron. Agricult.*, vol. 209, Jun. 2023, Art. no. 107833, doi: 10.1016/j.compag.2023.107833.

[16] S. Hu, R. Hou, L. Ming, S. Meifang, and P. Chen, "A hyperspectral image reconstruction algorithm based on RGB image using multi-scale atrous residual convolution network," *Frontiers Mar. Sci.*, vol. 9, Jan. 2023, Art. no. 1006452, doi: 10.3389/fmars.2022.1006452.

[17] Y. Liu, J. Zhang, and Y. Zhang, "Hyperspectral reconstruction from a single textile RGB image based on the generative adversarial network," *Textile Res. J.*, vol. 93, nos. 1–2, pp. 307–316, Jan. 2023, doi: 10.1177/00405175221118105.

[18] S. Galliani, C. Lanaras, D. Marmanis, E. Baltsavias, and K. Schindler, "Learned spectral super-resolution," Tech. Rep., 2017.

[19] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 939–947, doi: 10.1109/CVPRW.2018.00139.

[20] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu, "HSCNN: CNN-based hyperspectral image recovery from spectrally undersampled projections," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 518–525, doi: 10.1109/ICCVW.2017.68.

[21] S. Mei, Y. Geng, J. Hou, and Q. Du, "Learning hyperspectral images from RGB images via a coarse-to-fine CNN," *Sci. China Inf. Sci.*, vol. 65, no. 5, pp. 1–14, May 2022.

[22] Y. Zhao, L.-M. Po, Q. Yan, W. Liu, and T. Lin, "Hierarchical regression network for spectral reconstruction from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1695–1704, doi: 10.1109/cvprw50498.2020.00219.

[23] J. Li, C. Wu, R. Song, Y. Li, and F. Liu, "Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1894–1903, doi: 10.1109/cvprw50498.2020.00239.

[24] J. Li, S. Du, C. Wu, Y. Leng, R. Song, and Y. Li, "DRCR Net: Dense residual channel re-calibration network with non-local purification for spectral super resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1258–1267.

[25] Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, R. Timofte, and L. V. Gool, "MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 744–754, doi: 10.1109/CVPRW56347.2022.00090.

[26] Y. Cai, J. Lin, X. Hu, H. Wang, X. Yuan, Y. Zhang, R. Timofte, and L. Van Gool, "Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17481–17490.

[27] P. Du, T. Fang, H. Tang, and P. Shi, "Similarity measure of spectral vectors based on set theory and its application in hyperspectral RS image retrieval," *Chin. Opt. Lett.*, vol. 1, no. 11, pp. 637–640, 2003. [Online]. Available: https://opg.optica.org/col/abstract.cfm?URI=col-1-11-637

[28] C. L. C. Liu, Z. H. Z. Han, and T. X. T. Xie, "Hyperspectral high-dynamic-range endoscopic mucosal imaging," *Chin. Opt. Lett.*, vol. 13, no. 7, pp. 71701–71705, 2015. [Online]. Available: https://opg.optica.org/col/abstract.cfm?URI=col-13-7-071701

[29] R. Hang, Q. Liu, and Z. Li, "Spectral super-resolution network guided by intrinsic properties of hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 30, pp. 7256–7265, 2021, doi: 10.1109/TIP.2021.3104177.

[30] J. He, Q. Yuan, J. Li, Y. Xiao, D. Liu, H. Shen, and L. Zhang, ‘‘Spectral super-resolution meets deep learning: Achievements and challenges,’’ *Inf. Fusion*, vol. 97, Sep. 2023, Art. no. 101812.

[31] S. M. Park, M. A. Visbal-Onufrak, M. M. Haque, M. C. Were, V. Naanyu, M. K. Hasan, and Y. L. Kim, ‘‘mHealth spectroscopy of blood hemoglobin with spectral super-resolution,’’ *Optica*, vol. 7, no. 6, p. 563, 2020.

[32] T. Kim, M. A. Visbal-Onufrak, R. L. Konger, and Y. L. Kim, ‘‘Data-driven imaging of tissue inflammation using RGB-based hyperspectral reconstruction toward personal monitoring of dermatologic health,’’ *Biomed. Opt. Exp.*, vol. 8, no. 11, p. 5282, 2017.

[33] N. Sharma and M. Hefeeda, ‘‘Hyperspectral reconstruction from RGB images for vein visualization,’’ in *Proc. 11th ACM Multimedia Syst. Conf.*, May 2020, pp. 77–87.

[34] L. Ma, A. Rathgeb, H. Mubarak, M. Tran, and B. Fei, ‘‘Unsupervised super-resolution reconstruction of hyperspectral histology images for whole-slide imaging,’’ *J. Biomed. Opt.*, vol. 27, no. 5, May 2022, Art. no. 056502.

[35] C. U. Ortega, E. Q. Gutiérrez, L. Quintana, S. Ortega, H. Fabelo, L. S. Falcón, and G. M. Callico, ‘‘Towards real-time hyperspectral multi-image super-resolution reconstruction applied to histological samples,’’ *Sensors*, vol. 23, no. 4, p. 1863, Feb. 2023.

[36] J. Zhang, R. Su, Q. Fu, W. Ren, F. Heide, and Y. Nie, ‘‘A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging,’’ *Sci. Rep.*, vol. 12, no. 1, p. 11905, Jul. 2022, doi: 10.1038/s41598-022-16223-1.

[37] Z. Meng, M. Qiao, J. Ma, Z. Yu, K. Xu, and X. Yuan, ‘‘Snapshot multispectral endomicroscopy,’’ *Opt. Lett.*, vol. 45, no. 14, p. 3897, 2020.

[38] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, ‘‘Transformer in transformer,’’ in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 15908–15919.

[39] W. Yu, M. Luo, P. Zhou, C. Si, Y. Zhou, X. Wang, J. Feng, and S. Yan, ‘‘MetaFormer is actually what you need for vision,’’ in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10809–10819.

[40] Y. Chen, S. Wang, and F. Zhang, ‘‘Near-infrared luminescence high-contrast in vivo biomedical imaging,’’ *Nature Rev. Bioeng.*, vol. 1, no. 1, pp. 60–78, Jan. 2023.

[41] W. Wu, ‘‘Information entropy-based strategy for the quantitative evaluation of extensive hyperspectral images to better unveil spatial heterogeneity in mass spectrometry imaging,’’ *Anal. Chem.*, vol. 94, no. 29, pp. 10355–10366, Jul. 2022, doi: 10.1021/acs.analchem.2c00370.

[42] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, ‘‘CBAM: Convolutional block attention module,’’ in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.

[43] K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[44] U. B. Gewali, S. T. Monteiro, and E. Saber, ‘‘Spectral super-resolution with optimized bands,’’ *Remote Sens.*, vol. 11, no. 14, p. 1648, Jul. 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/14/1648

[45] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. Yang, ‘‘Restormer: Efficient transformer for high-resolution image restoration,’’ in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5718–5729, doi: 10.1109/CVPR52688.2022.00564.

HUIYUAN ZHANG received the bachelor’s degree from the Nanchang University of Aeronautics, in 2022. He is currently pursuing the degree with the School of Instrument Science and Optoelectronics, Beihang University. His primary research interests include spectral super-resolution, medical image processing, and information entropy.

ZHAOHUA YANG (Member, IEEE) received the bachelor’s degree in instrument science and technology from the Harbin Institute of Technology, in 1998, and the master’s degree in control science and engineering from the Gansu University of Technology, in 2001. She further advanced her studies with the Harbin Institute of Technology, from 2001 to 2004, specializing in precision instruments and machinery. Professionally, she has a storied career with the School of Instrument Science and Optoelectronics, Beihang University. She began her journey there, from May 2004 to May 2006, followed by another tenure, from May 2006 to August 2007. Recently, she rejoined the Faculty, in September 2021, where she has been contributing. This domain incorporates technologies spanning machine vision, signal processing, and light field modulation and control. Spacecraft Attitude Measurement and Control: This encompasses the control of spacecraft actuators and the high-precision attitude control techniques of carriers. Micro-Nano Resonant Cavity Sensing Technology: Vital for high-precision measurements of microscopic scales of particles and angular velocity, this technology holds significant value in fields like biomedical engineering, autonomous navigation of spacecraft, and mobile phone attitude measurement. Her primary research interests include quantum imaging detection and its applications: quantum imaging, characterized by high sensitivity and super-resolution, has emerged as a pivotal branch in optoelectronic imaging, and computational imaging.

ZEYUAN DONG received the bachelor’s degree in mechanical engineering from Southwest Jiaotong University, in 2012, and the master’s degree from the School of Automation Science and Electrical Engineering, Beihang University, Beijing, in 2018, where she is currently pursuing the Ph.D. degree in precision instruments and machinery. She is a Researcher with Beihang University. Her primary area of research interests include medical hyperspectral imaging, where she aims to advance diagnostic techniques and treatment options.

YIJING CHEN received the bachelor’s degree from the Department of Measurement, Control Technology, and Instruments, Wuhan University of Technology, in 2022. He is currently pursuing the Ph.D. degree in precision instruments and machinery with Beihang University, Beijing. His research interest includes medical hyperspectral imaging.