Computer Methods and Programs in Biomedicine 269 (2025) 108905

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: https://www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

DC-MSSFF Net: Dule-channel multi-scale spatial-spectral feature fusion network for cholangiocarcinoma pathology high-resolution hyperspectral image segmentation

Meiyan Liang [a,b,1,*], Zelin Xi [c,1], Bo Li [d], Lin Wang [e,**]

[a] Shanxi Key Laboratory of Wireless Communication and Detection, School of Physics and Electronic Engineering, Shanxi University, Taiyuan, 030006, Shanxi, China
[b] Shanxi Beike Biotechnology, Taiyuan, 030006, Shanxi, China
[c] Shanxi Institute of Mechanical and Electrical Engineering, Changzhi, 046000, Shanxi, China
[d] Department of Rehabilitation Treatment, Shanxi Rongjun Hospital, Taiyuan, 030000, Shanxi, China
[e] Department of pathology, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University, Taiyuan, 030032, Shanxi, China

ARTICLE INFO

Dataset link: https://www.kaggle.com/datasets/ethelzq/multiDimension-choledoch-database

Keywords:
Cholangiocarcinoma
Ensemble learning
Graph convolutional network
Graph-within-graph
Hyperspectral image

ABSTRACT

Background and Objective: High-precision segmentation of pathological images is a challenging task in the field of medical image processing. Hyperspectral microscopic imaging offers a distinct advantage in histopathological image segmentation due to its abundance of spectral and spatial data.
Methods: Here, a Dule-Channel Multi-Scale Spatial-Spectral Feature Fusion Network (DC-MSSFF Net) is proposed for semantic segmentation of cholangiocarcinoma hyperspectral images (HSI). The DC-MSSFF Net is composed of two parallel channels, graph-within-graph (GwG) and multi-scale CNN. The GwG can greatly reduce the computational burden while establishing the spatial context relationship of the HSI image. The multi-scale CNN channel is able to fine-tune the segmented edges of the HSI images at the pixel-level based on hyperspectral information in the depth dimension. Afterwards, the segmentation results are achieved by fusing the features from the two channels. Furthermore, an ensemble-based framework is applied to further improve the performance of the model.
Results: The image segmentation evaluation indexes such as dice similarity coefficient (Dice) of the Cholangiocarcinoma HSI data can reach 70.47, which is much higher than the SOTA method and RGB-based image segmentation methods.
Conclusion: The superior performance of the DC-MSSFF network pioneers the inductive learning task of deep frameworks for semantic segmentation of high-resolution hyperspectral image (HR-HSI).

1. Introduction

Cholangiocarcinoma is the second most common primary hepatic malignancy after hepatocellular carcinoma worldwide, accounting for approximately 15 of all primary liver tumors [1,2]. The incidence and mortality rates of cholangiocarcinoma have increased steadily over the past decades. To date, the histopathological examination is still considered to be the ‘‘gold standard’’ for the diagnosis and treatment of primary tumors, including cholangiocarcinoma. However, the diagnosis of cholangiocarcinoma is suffered from its insidious clinical features, low specificity of most diagnostic modalities, and lack of absolute diagnostic criteria. This is mainly because traditional pathology images can only provide two-dimensional spatial information, which limits the accuracy improvement of image processing algorithms.

Hyperspectral microscopic imaging (HMI) is an advanced 3D imaging technology that combines optical microscopy and hyperspectral imaging (HSI) to obtain both 2D image and 1D spectral information in a hyperspectral data cube. This method not only records the spatial feature of a sample, but also reveals its chemical properties through discrete/continuous spectrum of the reflected light at each pixel of the image. Compared with the traditional 2D pathology images,

* Corresponding author at: Shanxi Key Laboratory of Wireless Communication and Detection, School of Physics and Electronic Engineering, Shanxi University, Taiyuan, 030006, Shanxi, China.
** Corresponding author.
E-mail addresses: meiyanliang@sxu.edu.cn (M. Liang), 202122616117@email.sxu.edu.cn (Z. Xi), wanglin@sxbqeh.com.cn (L. Wang).
[1] The two authors contribute equally to this work.

https://doi.org/10.1016/j.cmpb.2025.108905
Received 17 November 2024; Received in revised form 4 June 2025; Accepted 8 June 2025
Available online 21 June 2025

hyperspectral data cube provides more comprehensive lesion information of tissues across a wide spectrum of light rather than just assigning primary colors (red, green and blue) to each pixel. Therefore, it has shown great potential for applications in computational pathology, such as pathological image classification and segmentation. Meanwhile, Silvia Seidlitz and Jan Sellner [3] have also demonstrated that image segmentation results are more accurate based on HSI modality rather than RGB images through the U-Net. Typically, it also provides more possibilities for the diagnosis of intractable lesions such as cholangiocarcinoma image segmentation.

2. Related works

2.1. Machine learning-based methods

For hyperspectral image segmentation tasks, the mainstream research includes methods based on machine learning [4] and approaches related to deep learning [4,5]. Over the past decade, hyperspectral image classification and segmentation have been primarily based on machine learning and its related methods. In these studies, support vector machine (SVM), random forests (RF), K-Nearest Neighbors (KNN) [6], linear discriminant analysis (LDA) [7] and other classic image processing techniques [8] are employed for hyperspectral pathological image segmentation. Here, methods such as principal component analysis (PCA) [9,10], independent component analysis (ICA) [10] and non-negative matrix factorization (NMF) [8] are used for dimensionality reduction in the preprocessing stage to avoid the curse of dimensionality.

Specifically, Nathan et al. [11] combined hyperspectral imaging technology and SVM classifier together to distinguish the different types of cancer, where PCA was used as a feature extraction and down-scaling method in hyperspectral image preprocessing. In 2019, Neeraj Kumar [8] et al. proposed a semi-supervised approach for processing hyperspectral pathology images, in which they utilized NMF for dimension reduction and preprocessing. Subsequently, hierarchical clustering was harnessed as a classifier to perform pixel-level image segmentation. The feasibility of the method was rigorously evaluated and validated on a mid-infrared colon pathological image dataset. It also provides deep insights into the importance of using k-means for pre-clustering, illustrating the merits of hierarchical clustering compared with SVM and RF, especially in the presence of substantial intra-class variation and limited labeled samples. Lu [12] et al. also applied hyperspectral imaging technology to the field of digital pathology, by establishing the correlation between selected histopathological features and spectral signatures with Spearman's rank correlation coefficient, which attempts to distinguish tongue neoplasia from normal tissues in early stage. However, these machine learning approaches have limited performance on various types of tasks due to the specific feature representations obtained manually.

2.2. Deep learning-based methods

2.2.1. CNN-based approaches

Deep learning has demonstrated strong capabilities in extracting hierarchical semantic features by integrating both spatial and spectral dimensions of hyperspectral data, significantly improving segmentation and classification accuracy. Although initially developed for RGB imagery, its application to hyperspectral data processing has shown considerable promise and has attracted growing attention in fields such as remote sensing [13–19] and biomedicine [20,21]. In early works, Li et al. [22] explored the use of 2D CNNs to derive deep spatial representations from hyperspectral remote sensing data, subsequently employing support vector machines (SVMs) for pixel-level classification. However, this approach faced inherent limitations in capturing spectral information — a critical aspect of hyperspectral data — which notably affected classification performance. To address this shortcoming, Chen et al. [23] advanced the field by introducing 3D CNNs that jointly extract spatial and spectral features. While effective, this method introduced significant computational challenges due to the high dimensionality of HSI. Afterwards, Swalpa Kumar Roy et al. [24] proposed HybridSN, a modular architecture that combines spectral-spatial 3D convolutions with subsequent 2D spatial processing. This hybrid design achieves higher classification accuracy than standalone 3D CNN models while reducing computational costs, offering a more efficient balance between feature complexity and performance. It is important to note that in these approaches, hyperspectral images are generally treated as static 3D data cubes, ignoring the continuous spectral information contained within each pixel. Each pixel, in fact, comprises a reflectance sequence that forms a unique spectral signature. Recognizing this, Wang et al. [25] proposed the ASSMN framework, which strategically integrates LSTM module to simulate multi-scale spectral and spatial structures. This approach emphasizes the scientific importance of capturing spectral sequences, enabling a deeper exploration of the intrinsic structure of hyperspectral data.

However, these CNN-based predictive methods are all based on transductive learning due to the inherent image properties and specific tasks in remote sensing. While in the medical field, transductive learning approaches are relatively meaningless and have poor generalization ability, when directly applied to medical image classification and segmentation tasks. Contrastly, inductive learning approaches are obviously more applicable and significant for processing high-resolution hyperspectral medical images.

2.2.2. UNet-based approaches

In the field of medical image segmentation, UNet and its variants [26–29] have achieved great success on images in RGB modality by fusing multi-scale spatial features from up-sampling layers to corresponding down-sampling layers. These methods have been applied to the segmentation of hyperspectral medical images and can be viewed as an inductive learning task. For instance, Stojan Trajanovski et al. [30] introduced HSI-UNet, a dual-stream architecture that partitions hyperspectral cubes into visible and near-infrared band to exploit information from specific wavelength ranges independently. Their findings demonstrated that segmentation based solely on the visible spectrum was insufficient for tumor identification; however, the integration of near-infrared spectra enabled accurate detection of positive regions. This highlights the critical role of multi-band spectral fusion in overcoming the limitations of single-modality imaging for medical segmentation tasks. Similarly, The FDSS Encoder [31] utilizes frequency disentanglement to decompose hyperspectral data into high-frequency semantic and low-frequency style components across both spatial and spectral domains. This decomposition facilitates more comprehensive utilization of multi-scale information. Building on this foundation, Zhan et al. [32] developed HyperUNet, incorporating a multi-scale supervised loss mechanism to guide spectral-spatial feature learning for cholangiocarcinoma segmentation. By employing a linear transformation module to enhance discriminative spectral components and suppress noise before feeding the data into the network, HyperUNet achieved a Dice score of 68.32 on the cholangiocarcinoma dataset. This innovation demonstrated that spectral preprocessing within UNet architectures can significantly enhance segmentation accuracy. To better exploit the rich spectral features within 3D hyperspectral data, Wang et al. [33] developed the Hyper-Net architecture, which enhances the deep fusion of spectral and spatial information. Hyper-Net, a variant of UNet, utilizes 3D convolutions and dual-path dilated convolutions to capture fine-grained information in melanoma datasets. Although HyperUnet outperforms 2D counterparts by effectively utilizing spectral information, its high computational complexity remains a significant challenge for high-resolution hyperspectral images, highlighting the challenge of balancing feature expression ability and computational complexity in clinical hyperspectral analysis.

In summary, these studies illustrate a clear evolutionary trajectory: Although deep learning has promoted the development of hyperspectral image segmentation through multi-stream architecture, spectral feature engineering and 3D CNN, the classic UNet-based models still face two persistent challenges: insufficient spectral feature utilization and high computational cost. Therefore, the search for architectures that effectively balance spectral-spatial representation with computational efficiency remains a critical frontier for translating hyperspectral imaging into practical clinical applications.

2.2.3. GCN-based transductive learning

Inspired by the field of remote sensing, Graph Convolutional Network (GCN) has received increasing attention in the medical field due to their ability to handle irregularly structured data. Unlike traditional CNN based networks that use shared parameters and fixed-size kernels, GCN can adaptively construct edges according to the local features of nodes in a hyperspectral image. This allows GCN to perform convolution flexibly on arbitrary irregular regions.

In 2019, Qin et al. [34] proposed a semi-supervised learning framework spectral-spatial graph convolutional networks (S2GCN) for hyperspectral image classification, which considers each pixel in a hyperspectral image as a node and uses the neighborhood relationship between pixels to build edges. It is apparent that this pixel-level graph construction approach is not applicable to HR-HSI since the consumption complexity grows exponentially with the number of nodes and edges. To reduce the amount of calculation when applying GCN, in 2021, Hong et al. [35] proposed the miniGCN network, which allows to construct and train large-scale GCNs in a minibatch fashion. Note that, this network has certain inductive properties and can also reduce the amount of calculation. Afterwards, Wan et al. [36] presented a multiscale dynamic GCN (MDGCN), which introduces superpixels concept into GCN framework and greatly improves the computational efficiency compared with pixel-based GCN. Notably, the superpixel-based GCN achieved a performance improvement of 3.48% on the same dataset compared to the miniGCN. In 2022, Ding et al. [37] proposed a superpixel-based Multi-Feature Fusion Network (MFGCN) that combines multi-scale GCN and multi-scale CNN to extract the spatial and spectral information of HSI, respectively. By specific spatial and spectral processing channels, it achieves an accuracy improvement of 3.81% compared to the miniGCN on transduction task. Unlike the aforementioned methods, Feng et al. [38] proposed discriminative information refinement (DIR) module tackles spectral redundancy by unifying spectral dimensions via adaptive mapping and modeling band relationships with k-NN graphs. This spectral-centric approach fundamentally deviates from spatial superpixel-based GCN methods, directly resolving the spectral complexity inherent in HSI data. However, GCN-based approaches continue to exhibit several fundamental limitations: (1) in hyperspectral image (HSI) processing, traditional GCN models inherently lack inductive capabilities, thereby restricting their effectiveness in pathological hyperspectral image segmentation; (2) although superpixel-based GCN frameworks can alleviate computational burdens to some extent, achieving an optimal balance between contextual feature representation and computational efficiency remains a significant challenge; and (3) heterogeneous pixels are inevitably introduced during the HSI pre-processing stage, primarily due to the inherent noise and spectral variability of HSI data, particularly near tissue phenotype boundaries.

To address this issue, a novel DC-MSSFF Net is proposed for pathological HR-HSI segmentation. It adopts an inductive learning method [39–41], which has more generalization ability and application prospects compared to transductive learning. Note that, it is a combination of GCN-based and CNN-based framework, which fully utilize the spatial information and spectral information of hyperspectral image in parallel channels to achieve fine-grained pathological image segmentation. In DC-MSSFF Net, GCN-based channel uses Graph-within-Graph (GwG) architecture, which can greatly reduce the computational burden while establishing the spatial context relationship of the high-resolution hyperspectral. The multi-scale CNN channel is able to fine-tune the segmentation boundaries of HSI image at the pixel-level according to the spatial–spectral information. Ultimately, the segmentation results of the microscopic hyperspectral choledoch dataset [42] are achieved by fusing the obtained features of the two channels. Moreover, an ensemble-based framework [43] is applied to further improve the performance of the model.

As the first inductive learning task built upon a joint multi-scale GCN and CNN dual-channel architecture, this work achieves state-of-the-art performance in cholangiocarcinoma high-resolution hyperspectral image (HR-HSI) semantic segmentation. The main contributions of this paper are summarized as follows:

- **Inductive GCN for HR-HSI Segmentation:** Drawing inspiration from GraphSAGE, we successfully integrate neighbor sampling and feature aggregation into our Graph-within-Graph backbone. This integration not only significantly enhances the inductive capability of the DC-MSSFF Net but also improves computational efficiency.
- **Hierarchical Contextual Feature Representation:** We introduce a hierarchical Graph-within-Graph dual-topology representation that embeds subgraphs into a global graph structure, effectively integrating local and global HSI features. This approach balances the need for context-aware feature representation with the computational burden, offering a robust framework for efficient processing.
- **Fine-Tuning and Optimization:** The multi-scale CNN branch leverages multi-scale spectral-spatial features to further refine segmentation accuracy. It fine-tunes the model to mitigate the impact of heterogeneous pixels introduced during the preprocessing stage and captures subtle feature variations, thereby generating more accurate superpixel boundaries.
- **Ensemble Learning:** By employing ensembles of three base learners configured with different hyperparameters, we further enhance the accuracy and reliability of HSI segmentation results through a majority voting mechanism. This ensemble strategy not only improves segmentation performance but also strengthens the model's generalization capability, which is crucial for practical applications in pathology.

3. Method

3.1. The framework of DC-MSSFF Net

The block diagram of Dual-Channel multi-scale spatial–spectral feature fusion network (DC-MSSFF Net) is illustrated in Fig. 1. As shown, DC-MSSFF Net consists of two sequential stages: image preprocessing and modeling. In the preprocessing stage, dimensionality reduction and pixel-to-superpixel mapping are performed. During modeling, the image cube is processed through two parallel channels: a multi-scale GCN spatial feature extraction channel (MSFE-GCN channel) and multi-scale CNN spectral feature extraction channel (MSFE-CNN channel). The MSFE-GCN channel performs efficient segmentation using superpixel representations, while the MSFE-CNN channel refines segmentation edges at the pixel level. Features from both channels are subsequently fused to obtain enhanced segmentation results. Finally, ensemble learning is employed to mitigate performance sensitivity to hyperparameter configurations.

3.2. Data pre-processing

Given a cholangiocarcinoma HSI dataset $D = \{X, Y\} = \{X_i, Y_i\}_{i=1}^{N}$, where X_i and Y_i are the ith input HSI and the corresponding pixel-level label. For each input HSI image, it can be expressed as $X_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,H\times W}\} \in R^{H\times W\times B}$. Where $H \times W$ and B denote the number of spatial pixels and the number of bands in the input image, respectively. The pixel-level label $Y_i \in R^{H\times W}$ is provided in the training process, which is a binary value in our image segmentation task.

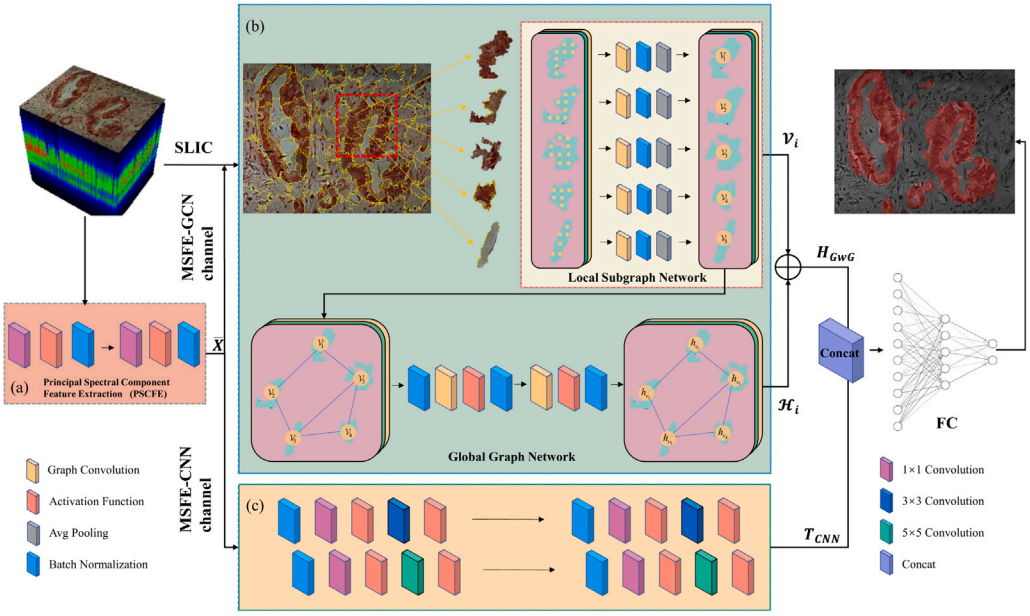

Fig. 1. DC-MSSFF Net architecture for hyperspectral cholangiocarcinoma segmentation. (a) PSCFE module using 1 × 1 convolution to compress 60 band to 32. (b) MSFE-GCN channel is used to capture the context-aware information of the image. (c) MSFE-CNN channel is applied for refining the segmentation boundaries of the HSI.

3.2.1. Dimension reduction

For the input HSI, dimension reduction is first performed through principal spectral component feature extraction (PSCFE) module to avoid the curse of dimensionality. Let $\bar{X}$ be the output feature of the PSCFE, and it can be expressed as

$$\bar{X} = \sigma(conv1(BN(\tilde{X}))) \tag{1}$$

where $\tilde{X} = SpectralNorm(X)$ denotes the feature of the input image after the spectral normalization, $conv1(\cdot)$ indicates 1 × 1 convolution operation, BN stands for the batch normalization operation and $\sigma(\cdot)$ denotes the activation function. Here, $\bar{X}$ is also considered as an input for the two downstream channels.

3.2.2. Pixel to superpixel assignment in MSFE-GCN channel

For MSFE-GCN channel, the pathology microscopic hyperspectral image is adaptively partitioned into superpixels based on SLIC algorithm according to the homogeneity of the HSI image (Fig. 2). Here, superpixel mapping matrix is obtained. As shown in Fig. 2, the hyperspectral image is partitioned into j superpixel regions, and each pixel is assigned a superpixel label. For instance, the superpixel region in the upper right corner is assigned a pre-segmentation label 22 according to the superpixel mapping matrix label. For ith HSI, the superpixel mapping matrix can be denoted as $S_i = \{S_{i,1}, \ldots, S_{i,j}\} \in R^{H \times W}$. Where each element $S_{i,j}$ represent jth superpixel of the ith image. For the jth superpixel, it contains a total of j' pixels and can be represented as $S_{i,j} = \{S_{(i,j),1}, \ldots, S_{(i,j),j'}\}$. Here, the relationship between the number of superpixels j in a HSI and hyperparameter j' is given by

$$j = \frac{H \times W}{j'} \tag{2}$$

where hyperparameter j' is the number of pixels within each superpixel. In our case, candidate value of $j' = 300$, 400 or 500. It adaptively adjusted around the candidate values with a variance of 200 pixels according to the boundary of tissue types within each HSI.

3.3. Multi-scale GCN spatial feature extraction channel

Graph Convolutional Networks (GCNs) extend convolutional operations to non-Euclidean data by leveraging graph structural information. They effectively capture non-linear node relationships and demonstrate strong generalization capabilities. However, pixel-based graph construction imposes significant computational burdens due to complex node connectivity. While superpixel-based methods alleviate this by treating statistically aggregated superpixels as nodes, they critically neglect inter-superpixel details, such as fine textures and boundary information. For HSI, an ideal model should exploit relationships both within and between superpixels. To address this, we propose a hierarchical Graph-within-Graph (GwG) structure to hierarchically establish the spatial multiscale feature of pathological HSI. Specifically, the overall framework of GwG is composed of a global graph network architecture and a certain number of local subgraph networks. Here, the subgraphs are obtained via the superpixel mapping matrix S in image preprocessing. And the global graph is constructed based on these subgraphs in a hierarchical manner. This means that each local subgraph is constructed based on pixels, while the global graph is build based on the superpixels.

For the ith input HSI image feature $\bar{X}_i$ after PSCFE module, the global graph $\mathcal{G}_i^{glob} = \{\mathcal{G}_i^S, \mathcal{E}_i^{glob}\}$ is constructed using the obtained superpixel mapping matrix S_i, in which $\mathcal{G}_i{}^S = \{\mathcal{G}_{i,1}^S, \ldots, \mathcal{G}_{i,j}^S\}$ is a set of local subgraphs and a node set of $\mathcal{G}_i^{glob}$. $\mathcal{E}_i^{glob}$ represent the edges between the superpixels. Similarly, each subgraph can be denoted as $\mathcal{G}_{i,j}^S = \{\mathcal{V}_{i,j}, \mathcal{E}_{i,j}\}$, where $\mathcal{V}_{i,j}$ and $\mathcal{E}_{i,j}$ represent the vertex set and edge set of the subgraph $\mathcal{G}_{i,j}^S$, respectively. Here, each vertex $\mathcal{V}_{i,j}$ is a pixel in the subgraph, which can be expressed as $\mathcal{V}_{i,j} = \bar{X}_{i,j} = [\bar{X}_{(i,j),1}, \ldots, \bar{X}_{(i,j),a}, \bar{X}_{i,j,b}..., \bar{X}_{(i,j),j'}]^T$. The element $\bar{X}_{(i,j),j'}$ stands for the j'th node of the jth subgraph in the input $\bar{X}_i$. $\mathcal{E}_{i,j}$ denotes the edge set of subgraphs. where $A_{i,j}$ is the adjacency matrix of subgraph $\mathcal{G}_{i,j}^S$ for the input $\bar{X}_i$, which can be expressed as

$$A_{i,j}^{a,b} = \begin{cases} 1 & if\,(X_{(i,j),a}, X_{(i,j),b}) \in \mathcal{E}_{i,j} \\ 0 & others \end{cases} \tag{3}$$

where $A_{i,j}^{a,b}$ denotes the connection between node a and node b in the jth superpixel.

3.3.1. Local subgraph network

For the cholangiocarcinoma hyperspectral image dataset, each HSI image contains 60 spectral bands. Although superpixel segmentation can partition similar pixels into homogeneous regions according to the weighted Euclidean distance of spectral and spatial features, there still

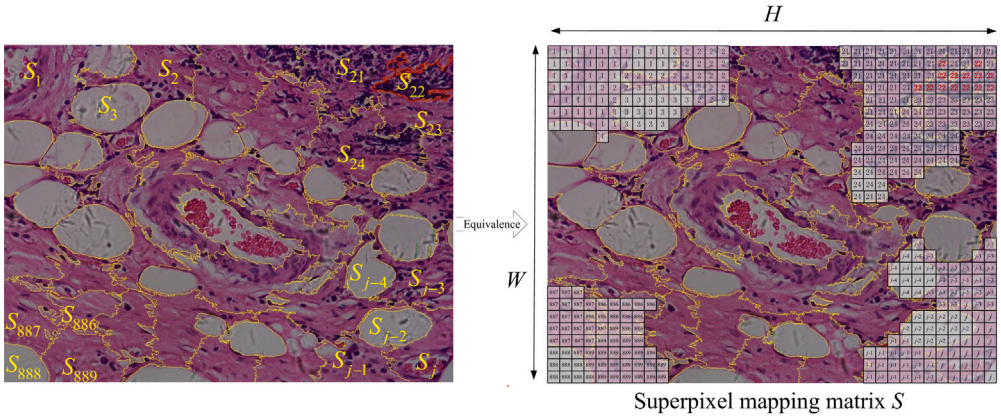

Fig. 2. The schematic diagram of superpixel mapping matrix. All pixels within a superpixel share identical index values.

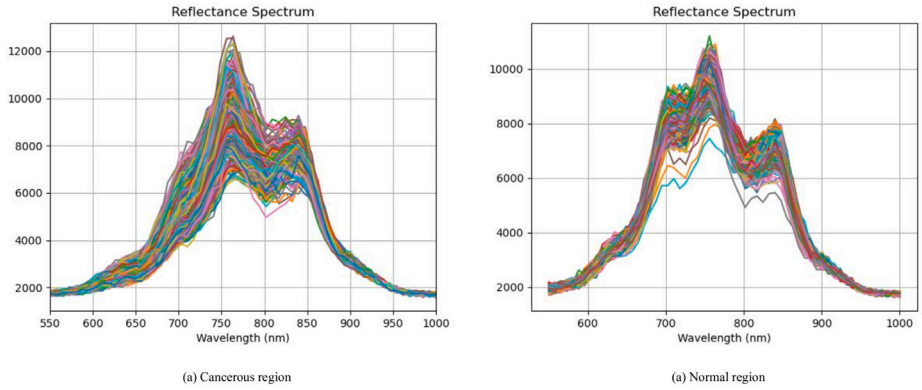

Fig. 3. Spectral reflectance variation in two randomly selected superpixels (550–1000 nm wavelength range).

exists significant spectral differences within superpixels, as illustrated in Fig. 3. It can be clearly observed from Fig. 3 that there is a significant spectral variability within each superpixel, for both cancerous regions and normal regions of the tissue. Specifically, although the conventional statistical method such as maximum value can maintain the most salient spectral features of superpixels, it may be over exaggerated due to spectral variation, as shown in Fig. 3(b). It is obviously that maximum value on some bands are significantly different from the majority, which may lead to biased decision based on such an inappropriate feature representation. While for the classical average value statistical method, it may be a good alternative option for suppressing noise. However, this method vulnerable to external conditions. Besides, these simple non-trainable statistical methods cannot provide accurate feature representation of superpixels, thus affecting the accuracy of final segmentation results. Therefore, it is not an optimal option to represent each superpixel base on pixel-wise uniform statistical features.

Compared with CNN-based methods, the GCN framework can handle arbitrary irregular-shaped images, which makes it a priority for superpixel processing. Theoretically, pixels within a superpixel should share similar spatial features and final latent representations. Consequently, we convert each superpixel into a pixel-level graph named local subgraph network for fine-grained information mining. Here, each pixel is treated as a node of the local subgraph network. As shown in Fig. 4, a superpixel is converted into a local subgraph network, in which each pixel (red dot) is connected by its four neighbors (green dot). The feature vector of each node is the pixel-wise spectral feature along the depth dimension.

After pixel-wise graph convolution operation, the output of the local subgraph network is given by

$$Y_{local}^{(l)} = \sigma(\tilde{D}_{local}^{(-1/2)} \tilde{A}_{local} \tilde{D}_{local}^{(-1/2)} Y_{local}^{(l-1)} W^{(l)}) \tag{4}$$

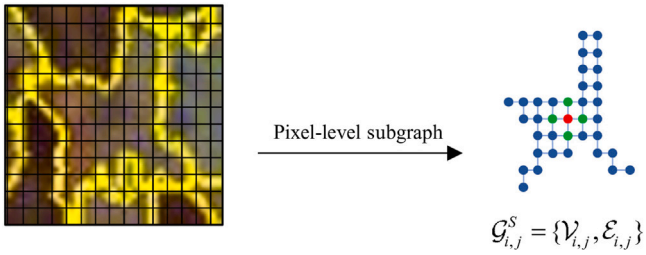

Fig. 4. The local subgraph network. The red dot indicates the target node, and the green dot indicates the four neighbors of the target node.

where $Y_{local}^{(l-1)}$ is the output feature of the $(l-1)$th GCN layer. Here $Y_{local}^{(0)} = \tilde{X}_{i,j}$, when $l = 0$. Where $W^{(l)}$ is the connection weight matrix learned by the graph convolution layer. Here, $\tilde{A}_{local} = A_{local} + I$ denotes a new adjacency matrix of the local subgraph network by adding an identity matrix I. $\tilde{D}_{local}$ denotes the degree of the adjacency matrix $\tilde{A}_{local}$. $\sigma(\cdot)$ denotes the activation function like ReLU.

After message passing between nodes in the subgraph, a global average pooling (GAP) layer is added to obtain the overall features of the superpixels. The feature of jth superpixel is obtained by

$$v_{i,j} = GAP(Y_{i,j}) \tag{5}$$

where Y_{local} is the output of the GCN layer in each subgraph, $GAP(\cdot)$ denote the global average pooling operation. $v_{i,j}$ is the aggregated feature vector of each subgraph, which is also regarded as a node in the global graph. Therefore, the feature representation of the node set in global graph can be expressed as $\mathcal{V}_i = \{v_{i,1}, v_{i,2}, \ldots, v_{i,j}\}$.

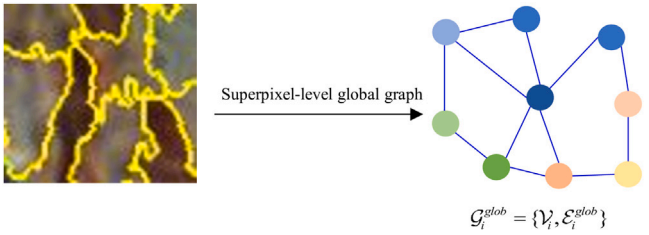

Fig. 5. The global graph network $\mathcal{G}_i^{glob}$ is constructed based on the superpixels. The colored nodes in the right represent the superpixels.

3.3.2. Global graph network

For pathological images, superpixel partition would over-segment HSI to ensure the homogeneity of superpixels, while integral tissue phenotype may be scattered in several adjacent superpixels. Furthermore, spatial relationships between various tissue phenotypes can also aid in HSI segmentation. Therefore, classifying hyperspectral pixels in HSI images only considering the output features $\mathcal{V}_i$ of subgraph may lead to overall performance degradation due to the inherent limitation of considering local features. Here, a global graph network is built hierarchically based on the adjacent relationships between superpixels and the spatial patterns of tissue phenotypes. Each node in the global graph represents an embedding of local subgraph features. The hierarchical structure of GwG allows for the utilization of both discriminative global contextual information and local deep features from HSI, resulting in improved image segmentation. Specifically, the global graph can be expressed in a transformed form $\mathcal{G}_i^{glob} = \{\mathcal{V}_i, \mathcal{E}_i^{glob}\}$ based on the node feature embeddings, as shown in Fig. 5. Here, $\mathcal{V}_i$ and $\mathcal{E}_i^{glob}$ are the node set and the edge set in global graph $\mathcal{G}_i^{glob}$, respectively. The architecture of GwG based on hierarchical feature embedding is shown in Fig. 1.

In the literature, there are two main types of learning problems on graphs: transductive learning and inductive learning. Transductive learning is inherently a single graph node classification or regression problem. This means that that a portion of the nodes are labeled for training, and the task is to predict the labels of unannotated nodes. It is worth noting that such methods were originally developed to classify hyperspectral remote sensing images. In contrast, inductive learning can be regarded as an entire graph classification or regression problem. It is the learning process in which the learners discover general rules by observing the training data, and then apply the rules to the unseen test data. The main difference is that during transductive learning, the model has already encountered both the training and testing data in the training process. However, inductive learning encounters only the training data in the learning process and applies the learned mapping function to a previously unseen dataset. In our case, the inductive capability is essential for high-throughput computer-assisted medical image processing systems. Thus, inductive learning is more applicable for processing pathological HSI rather than transductive learning does.

Here, GraphSAGE [44] is used as a backbone of our global graph network, as it is an inductive algorithm that exploits node feature information to generate node embeddings for unseen nodes or graphs. Besides, the GraphSAGE is also extremely efficient in generating node embeddings of large graphs. The key insight behind it is to decompose the large graph into multiple mini-graphs according to the adjacency relationship of the superpixels, and then aggregate the information of the local neighbors on each mini-graph to obtain the node embeddings.

Fig. 6 is the schematic diagram of the mini graph. In Fig. 6, the green nodes represent the superpixels on the global graph, which are also considered as the target nodes for mini graph decomposition. The red nodes represent the 1-hop neighbors of each target node. The blue nodes are the 2-hop neighbors, and so on to obtain k-hop neighbors for each target node. In our case, all k-hop neighbors of the target node are sampled for each mini graph, since the neighbors of the target node do not contain redundant information for it. Meanwhile, the computational burden is also acceptable in this scenario. Here, k is also known as 'searching depth' in the global graph network. As shown in Fig. 6, each target node and its k-hop neighbors are treated as a mini-batch, and their features are then aggregated on the mini graph to form the node embeddings.

The k-hop feature aggregation of the target node can be obtained recursively from its $(k-1)$-hop neighbors according to the following equation

$$h^k_{Ne(v_{i,j})} = AGGREGATE_k(\{h^{k-1}_{i,u}, \forall u \in Ne(v_{i,j})\}) \tag{6}$$

where $Ne(v_{i,j})$ denotes the neighbors of the node $v_{i,j}$. For each node $v_{i,j} \in \mathcal{V}_i$, Eq. (6) aggregates its $(k-1)$-hop neighbors to obtain the feature vector $h^k_{Ne(v_{i,j})}$. This aggregation process relies on the representations generated in the previous iteration, where the representations at $k = 0$ correspond to the input node features. For the node $v_{i,j}$, the obtained feature vector $h^k_{Ne(v_{i,j})}$ and its node feature $h^{k-1}_{i,v_{i,j}}$ from the previous aggregator are fused via concatenation operation. Afterwards, the contextual feature representation of the node $v_{i,j}$ can be transformed and expressed as

$$h^k_{i,v_{i,j}} = \sigma(W^k concat(h^{k-1}_{i,v_{i,j}}, h^k_{Ne(v_{i,j})})) \tag{7}$$

where σ is a non-linearity activation function like ReLU. Then, the L2 norm is used to normalize the feature embeddings of each node in the global graph network:

$$\tilde{h}_{i,v_{i,j}} = h^k_{i,v_{i,j}} / \left\| h^k_{i,v_{i,j}} \right\|_2 \tag{8}$$

Therefore, the output feature of the global graph network can be denoted as $\mathcal{H}_i = [\tilde{h}_{i,v_{i,1}}, \ldots, \tilde{h}_{i,v_{i,j}}]^T$. Here, skip connection is applied in MSFE-GCN channel to perform feature fusion on the superpixels (node), and the two features of each node are obtained through local subgraph and mini graph heterogeneous frameworks, respectively. That is,

$$H_{GwG} = \mathcal{V}_i + \mathcal{H}_i \tag{9}$$

This approach alleviates the issue of a significant computational complexity surge when applying graphs to HR-HSI. However, the segmentation results obtained by using the MSFE-GCN channel alone inevitably introduce heterogeneous pixels in the superpixels based on the SLIC algorithm. Therefore, it is necessary to fine-tune the partition boundaries of the superpixels by introducing a specific channel.

3.4. Multi-scale CNN spectral feature extraction channel

The Multi-Scale GCN Spatial Feature Extraction Channel (MSFE-GCN) performs HSI segmentation using a hierarchical Graph-within-Graph (GwG) topology operating at the superpixel level. To further enhance segmentation accuracy, we introduce a parallel Multi-Scale CNN Spectral Feature Extraction Channel (MSFE-CNN). As illustrated in Fig. 1, this channel refines superpixel boundaries at the pixel level, compensating for lost intra-superpixel detail. Since high-precision segmentation requires multi-scale local features, the MSFE-CNN channel explicitly incorporates dual receptive fields to capture complementary spectral characteristics.

In each subchannel, we split the 3D convolution kernel into 1D spectral convolution kernel and 2D spatial convolution kernel to avoid a significant increase in parameters and computational complexity. Here, 1D convolution kernel in each subchannel is used to further extract and refine the spectral features of HSI. While the following 2D convolution kernels are applied to extract spatial features of HSI in different scales based on 3×3 and 5×5 convolution kernels, respectively. As shown in Fig. 1, the sequential stacking of multiple convolutional layers allows the MSFE-CNN channel to capture more complex and abstract spectral features for pixel-level fine-tuning of hyperspectral image segmentation results of the GwG channel. Therefore, the output of the lth layer is given by

$$T^l_r = \sigma(W'^l_r \cdot \sigma(W^l_r \tilde{T}^{l-1}) + b^l_r) \tag{10}$$

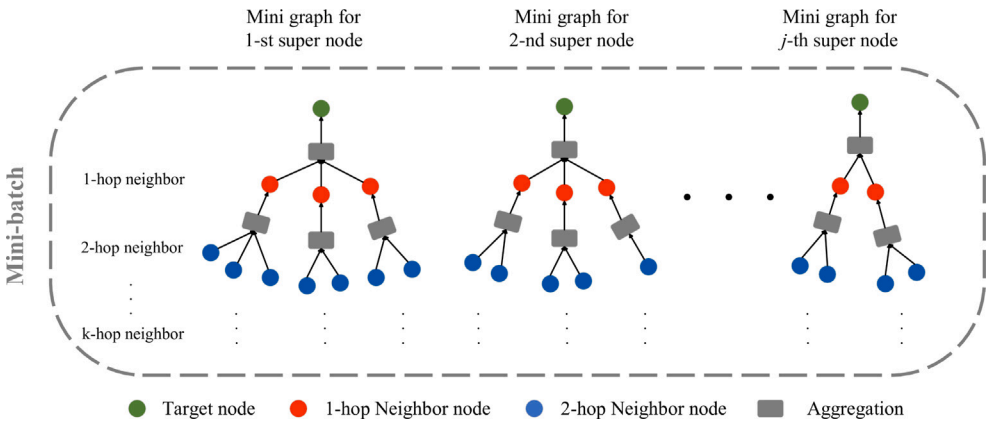

Fig. 6. Schematic diagram of the mini graph. The feature of the target node is progressively aggregated from k-hop to itself, thereby refining the embedding of the node.

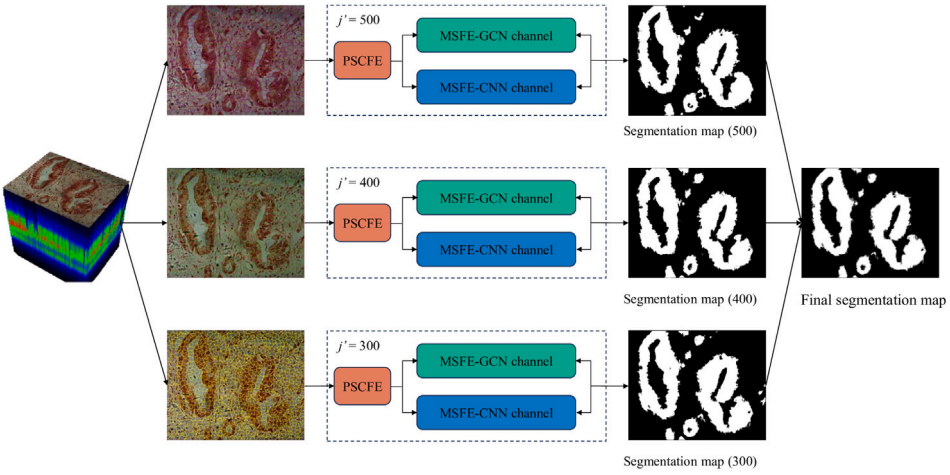

Fig. 7. Ensemble learning architecture.

where W_r^l and b_r^l denote the 1D connection weights and bias, respectively. While $W_r'^l$ denotes 2D convolution kernels with different receptive fields after the corresponding 1D convolution kernels. Specifically, it is a 3×3 or 5×5 convolution kernels convolutional kernel in each sub-channel, respectively. Here, $\tilde{T}^{l-1} = BN(T^{l-1})$ denotes the normalized output of $(l-1)$th layer, which can also be regarded as the input of the l layer. As shown in Eqs. (1) and (10), the input of MSFE-CNN channel can be expressed as $T^0 = \tilde{X}$. Finally, the features obtained from the MSFE-CNN channel are denoted as T_{CNN}.

Afterwards, the obtained features of the two channels are fused through concatenate operation

$$F = concat(H_{GwG} \parallel T_{CNN}) \tag{11}$$

where F is the final output after feature fusion. $concat(\cdot \parallel \cdot)$ denotes the concatenate operation. Finally, the segmentation result is obtained after the fully connected layer, which is also illustrated in Fig. 1. Here, the DC-MSSFF Net is optimized in an end-to-end manner by a minimizing the cross-entropy of predicted labels and ground truth.

3.5. Ensemble learning

In data pre-processing stage, the number of superpixels j in a HSI is a significant hyperparameter for representing different tissue phenotype, which severely affects the final image segmentation quality and accuracy. Specifically, small j would introduce heterogeneous pixels in the superpixels which will lead to the performance degradation. Whereas large j would increase the running time and computational burden for the hardware. To tackle this issue, an ensemble learning framework is used with DC-MSSFF Net as the base learner. In our case, there are three base learners, and the only difference between them is the setting of the hyperparameter j', which takes values of 300, 400 and 500, respectively. The final segmentation result is obtained by the maximum voting of the three base learners. Fig. 7 shows the ensemble learning architecture for processing the HSI images.

4. Experiments

4.1. Dataset and implementation details

Hyperspectral choledochal database. It is one of the largest public pathological hyperspectral image datasets obtained from the Changhai hospital, Shanghai, China. The image size of the microscopy hyperspectral date cube is $1024 \times 1280 \times 60$. Each hyperspectral image contains 60 bands, which are obtained by sampling uniformly over the wavelength range of 550 nm to 1000 nm. The HSI data cube are stored in the BSQ (band sequential) format. While the labels are marked with polygons and stored as .xml format.

In the experiments, a total of 338 cholangiocarcinoma hyperspectral images were divided into training, validation, and test sets following an approximate 8:1:1 ratio, comprising 258, 40, and 40 images respectively. The model was trained on an NVIDIA RTX 3090 (24 GB) GPU for 75 epochs with a batch size of 1, resulting in a total training time of approximately 120 h. Performance was evaluated using the Sorensen-Dice coefficient (Dice) and Intersection over Union (IoU). Detailed hyperparameter configurations are summarized in Table 1.

Table 1
Hyperparameter settings.

Category	Parameter	Value	Remarks
Data Partition	Training set	258	8:1:1 split ratio
	Validation set	40	
	Test set	40	
Training Setup	Optimizer	Adam	
	Initial learning rate	0.001	
	Batch size	1	Limited by GPU memory
	Training epochs	75	
Model Architecture	PSCFE output channels	32	Principal spectral component extraction
	GCN search depth (k)	3	For global graph aggregation
	Ensemble learners	3	$j' = 300, 400, 500$

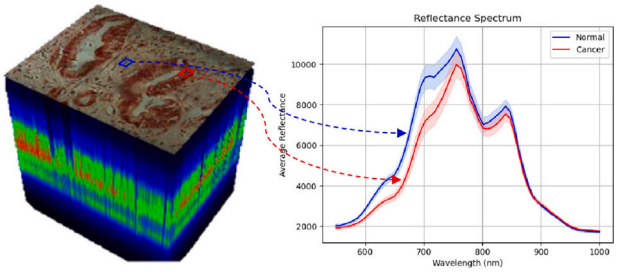

Fig. 8. The average reflectance intensity values versus wavelength in negative region (blue) and positive region (red).

Table 2
Dice (%) and IoU (%) indices of pathological images based on HSI and RGB Modalities.

Modality	Dice↑	IoU↑
RGB	64.76	48.57
HSI	**70.47**	**55.33**

4.2. Results

HSI vs. RGB: In field of pathological image processing, HSI cubes with high spatial and spectral resolution may potentially contain more diagnostic information than H&E-stained RGB images.

Fig. 8 presents the choledochal reflectance intensity values versus wavelength for cancerous and normal regions (the size of region are 20×20) in the testing set. In this figure, the horizontal axis represents the wavelength range from 550 nm to 1000 nm, while the vertical axis depicts the tissue reflectance intensity values corresponding to each wavelength. The darker lines indicate the average reflectance intensity values, whereas the shaded areas represent the standard error, reflecting the variance across the curves. As illustrated in Fig. 8, it is evident that the reflectance intensity values for the normal regions (blue line) are significantly higher than those for the cancerous regions (red line) within the spectral range of 600 nm to 750 nm, a distinction that could potentially benefit image segmentation tasks.

To further validate the effectiveness of hyperspectral imaging (HSI), segmentation experiments were conducted on pathological HSI images and their corresponding RGB images using the DC-MSSFF Net. The results of the cholangiocarcinoma HSI segmentation are summarized in Table 2. It is apparent that the Dice and Intersection over Union (IoU) indices for the segmentation results based on the hyperspectral cube are substantially higher than those obtained from the RGB modality. This improvement can be attributed to the rich spectral information embedded in each pixel of the pathological HSI cubes, which is extremely valuable for achieving high-precision image segmentation, particularly for lesions that lack distinct features in RGB images.

The result of DC-MSSFF Net: Fig. 9 and Table 3 show the qualitative and quantitative results of HSI segmentation, respectively. Since single-band grayscale images from the mid-spectral range (band 25th–band 30th) are relatively optimal for human interpretation. Here, band 30th is selected for visual analyzation. Fig. 9(a_7)–(e_7) are the image segmentation results of cholangiocarcinoma HSI using DC-MSSFF Net. While Fig. 9(a_8)–(e_8) represent the corresponding ground truth. To clearly illustrate the segmentation results, Fig. 9(a_1)–(e_1)show the original gray-scale images of cholangiocarcinoma in band 30th as a reference. The positive regions of each image are highlighted with a red mask.

As illustrated in Fig. 9, the segmentation results demonstrate that the pathological HSI image segmentation based on DC-MSSFF Net has high concordance with the ground truth, with sharply defined boundaries between positive and negative regions. The average Dice score and IOU can reach 70.47 and 55.33, further demonstrating its effectiveness in pathological HR-HSI segmentation tasks. Specifically, the DC-MSSFF Net accurately segmented the malignant epithelial tissue of the gland as positive, while predicting the large white empty space within the gland corresponding to the lumen of cholangiocarcinoma cells as negative. This observation underscores the model's capability for fine-grained discrimination in pathological HSI segmentation. It can be primarily attributed to the spectral information contained in hyperspectral images. While our method effectively leverages these spectral variations to perform fine-grained image segmentation. Besides, small glands merge with the surrounding tissues and are prone to being easily ignored, but DC-MSSFF Net effectively identifies the epithelial structures and accurately distinguishes them from adjacent luminal regions, which is marked with yellow square in Fig. 9(b_7). It is worth mentioning that DC-MSSFF Net, as the first inductive learning task based on joint multiscale GCN and CNN dual-channel architecture, which achieves state-of-the-art performance on pathological HR-HSI semantic segmentation.

4.3. Discussion

DC-MSSFF Net vs. Other image segmentation networks: To validate the effectiveness of the proposed DC-MSSFF Net, we conducted comprehensive comparisons qualitatively and quantitatively with other state-of-the-art models, which are also presented in Fig. 9 and Table 3.

(i) Classical Image Segmentation Frameworks: These methods, which include UNet and UNet++, are classical medical image segmentation networks. For the Unet, the green squares in Fig. 9(a_2) and (d_2) show that the malignant glands in each image are clearly segmented as a whole, and the large white empty regions in the gland corresponding to the lumen of the are also segmented as positive. This fact illustrates that the utilization of multi-scale spatial information in the Unet can ensure the segmentation of the most distinctive malignant glands. However, some smaller malignant gland are prone to be ignored according to the ground truth, which is highlighted with yellow squares in Fig. 9(b_2), (c_2) and (e_2). This is mainly because the model cannot capture the subtle spectral variation in cholangiocarcinoma HSI to achieve fine-grained image segmentation. In contrast, the segmentation performance of UNet++ is improved compared with Unet due to the properties of multi-scale spatial feature extraction and fusion operation.

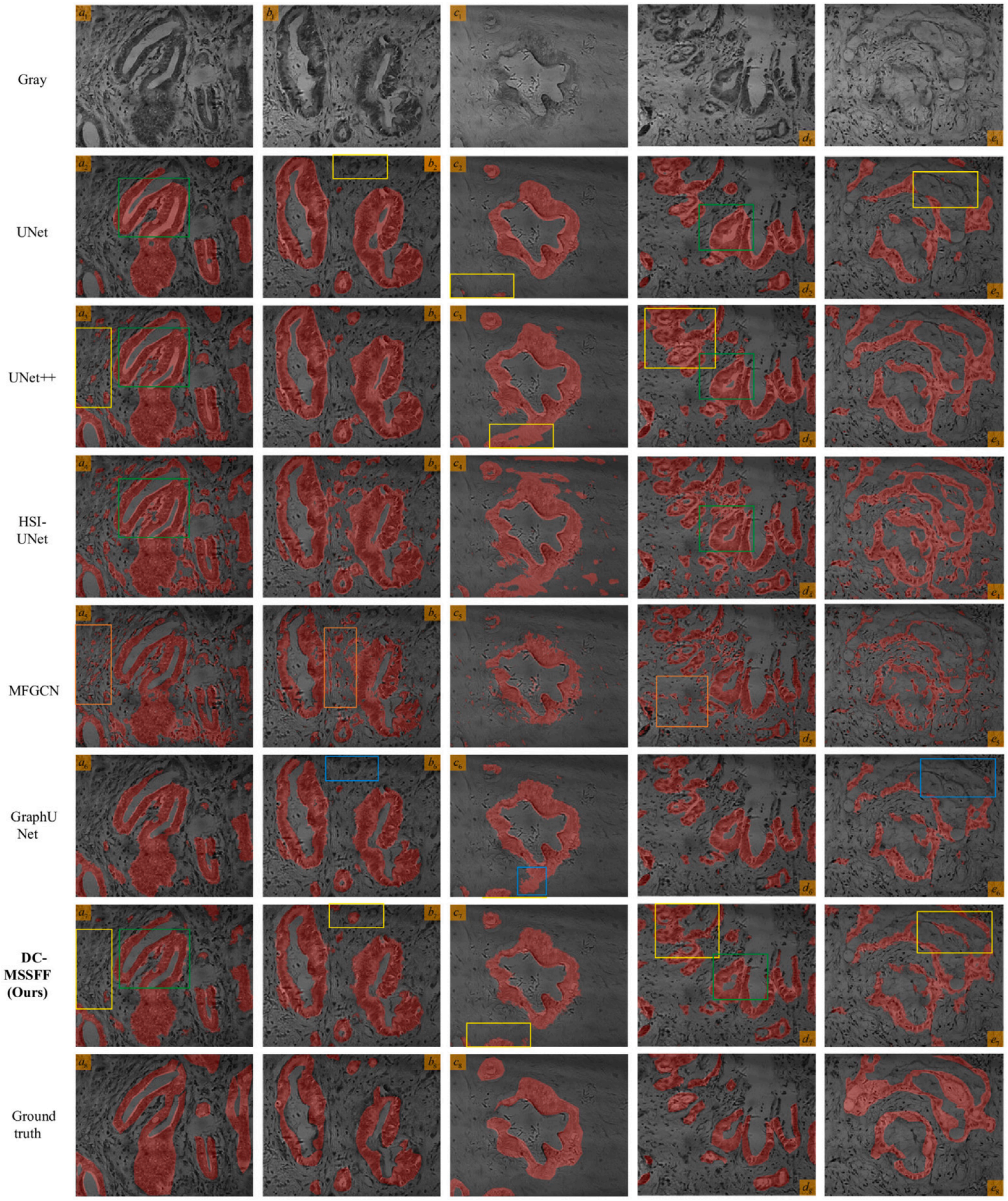

Fig. 9. Qualitative comparison of HSI segmentation using the proposed DC-MSSFF Net. Here, we selected grayscale images from the 30th spectral band and overlaid the predicted segmentation masks (highlighted in red) onto the images using alpha blending ($\alpha = 0.3$).

Table 3
Quantitative comparison of **DC-MSSFF Net vs. SOTA image segmentation networks** based on cholangiocarcinoma HSI dataset.

Method	Mean dice↑ (Median, Max, Min)	IoU↑	Parameters
HSI UNet [30]	62.92(66.71, 99.93, 12.59)	49.48	/
MFGCN [37]	63.21(66.43, 89.22, 16.71)	49.66	/
UNet [26]	65.80(66.52, 93.87, 17.97)	50.91	34.53 M
UNet++[27]	66.32(66.91, 94.66, 17.43)	51.43	47.20 M
HyperUNet [32]	68.32	/	/
SSTE-Former [45]	67.09(67.30, 88.50, 19.16)	52.26	/
GraphUNet [46]	68.57(69.50, 84.63, 29.55)	53.50	/
DC-MSSFF Net ($j' = 300$)	68.68(69.50, 88.64, 21.10)	53.29	0.033 M
DC-MSSFF Net ($j' = 400$)	69.75(71.47, 87.06, 24.99)	54.48	0.033 M
DC-MSSFF Net ($j' = 500$)	69.12(69.81, 86.50, 24.56)	53.74	0.033 M
DC-MSSFF Net (Ensamble)	**70.47(71.47, 87.43, 24.37)**	**55.33**	**0.099 M**

Nevertheless, the white empty regions in the gland are still partially separated from the positive area, which highlighted with green squares in Fig. 9(a_3) and (d_3). Meanwhile, it also exists the issue of over-segmentation, as shown by the yellow squares in Fig. 9(a_3), (c_3) and (d_3). These are mainly beacuse the classic UNet-like achitectures failed to effectively capture the necessary spectral information in HR-HSI, which seriously affects the segmentation performance. Furthermore, Applying Unet and its variants directly to proccess hyperspectral images

has drawback that they cannot handle images with large size, espacially HR-HSI.

(ii) U-net Variants: Methods such as HSI UNet and Hyper UNet represent UNet variants that incorporate specialized spectral information processing blocks in hyperspectral image analysis. For HSI UNet, the perfomance of the model lags behind all the comparison models. This is primarily due to the architectural limitations of the model in capturing and representing multi-scale spatial features. Nonetheless, it also demonstrated the significance of spectral information in fine-grained HSI image segmentation, typical cases are highlighted with green square in Fig. 9(a_4) and (d_4). While for Hyper UNet, it uses multi-scale supervision at each layer to regulate low-level detailed features and high-level semantic features at different scales, so that the output of the latter scale could benefit from the former scale. The Dice score of the model improved to 68.32 by utilizing the spectral information of HSI [32]. Therefore, it can be concluded that fully extraction of spatial and spectral information in HR-HSI images is an important factor for image segmentation.

(iii) Transformer-based Frameworks: Among the compared models, SSTE-Former employs a transformer-based architecture to model spectral-spatial dependencies, demonstrating moderate performance in hyperspectral feature integration. However, in hyperspectral imaging, not all pixel instances are mutually related; consequently, computing attention intensities for irrelevant instances not only results in marginal performance gains but also imposes an undesirable computational burden. Furthermore, the transformer-based architecture is limited in its ability to fully exploit the rich spectral information embedded within the hyperspectral data cube. As a result, the proposed DC-MSSFF Net surpasses SSTE-Former by 3.38% in Dice coefficient and 3.07% in Intersection over Union (IoU), confirming its superior performance in pathological HR-HSI segmentation tasks.

(iv) GCN-based Methods: Task-specific GCN-based methods, including MFGCN and GraphUNet, utilize GCN to construct contextual information of the input image. However, They all suffer from the insufficient feature representation issue of contextual fine-grained information. MFGCN is only based on superpixels and does not represent the internal features of superpixels in detail. Therefore, it cannot capture the significant spectral differences within superpixels and is not suitable for high-precision segmentation of pathological HR-HSI. Thus, there are many segmentation errors based on MFGCN, which are marked with orange squares in Fig. 9(a_5)–(e_5). For the GraphUNet, context feature representation and reconstruction are achieved through node aggregation and projection. However, only the top ranked nodes are retained during downsampling, resulting in the loss of fine-grained information in the reconstruction stage, as shown by the blue squares in Fig. 9(b_6), (c_6) and (e_6).

In summary, the DC-MSSFF network demonstrates notable advantages in HR-HSI segmentation by jointly employing the MSFE-GCN and MSFE-CNN channels. The innovative dual-channel architecture effectively integrates tissue topology with spatial–spectral detail features, thereby enabling more refined segmentation of diseased tissues and their boundaries. By hierarchically capturing spatial dependencies and simultaneously mitigating intra-superpixel heterogeneity, the proposed framework successfully addresses the limitations of traditional methods in multi-scale feature fusion and computational efficiency. Experimental results validate its superior performance, achieving a Dice score of 70.47 and an IoU of 55.33 on hyperspectral cholangiocarcinoma images. A representative case is highlighted with a yellow/green square in Fig. 9(a_7)–(e_7).

4.4. Ablation studies

To demonstrate the effectiveness of the proposed model, ablation of the local subgraph network within GwG, ablation of the MSFE-CNN channel, and ablation of the PSCFE module are conducted on DC-MSSFF Net. The results of these ablation studies are summarized in Table 4.

Table 4
Ablation studies.

Method	Dice↑	IoU↑
w/o Local subgraph network	64.51	48.72
w/o MSFE-CNN channel	68.27	52.63
DC-MSSFF Net	**69.12**	**53.74**

Table 5
The ablation studies of PSCFE module and the PSCFE module with different output channels.

DC-MSSFF	w/o 1DCNN	128 output	32 output	16 output
Dice	68.03	67.02	**69.12**	64.12
IoU	52.47	51.43	**53.74**	48.33
FLOPs	**31.93G**	76.57G	43.35G	37.81G

As shown in the Table, ablation of the local subgraph network in the MSFE-GCN channel causes a significant performance degradation, with the Dice score dropping from 69.12 to 64.51. This degradation is primarily caused by the failure to fully utilize the fine-grained spectral information of pixel within each superpixel. Without local subgraph network, the superpixel features are simply based on statistical pooling of pixel features, which is insufficient to represent the underlying spatial–spectral heterogeneity within a superpixel, such as the continuous spectral variations shown in Fig. 3. This ablation operation demonstrates the significance of hierarchical fine-grained contextual feature representation in HSI segmentation, especially in complex histological environments with less heterogeneity, as shown in Fig. 10(a_2), (b_2) and (d_2).

Moreover, although the MSFE-GCN channel provides hierarchical multi-scale spatial–spectral information, it still leads to blurred boundaries and over-segmentation of glands when acting alone, as shown in Fig. 10(a_3)–(f_3). This underscores the critical role of the MSFE-CNN channel in rectifying segmentation errors from the superpixel preprocessing stage. The performance degradation further demonstrates that pixel-level feature refinement via the CNN channel is essential to compensate for the inherent limitations of superpixel segmentation methods like SLIC.

Table 5 shows the ablation experiments of the PSCFE module, which demonstrates the impact of dimensionality reduction on model efficiency and accuracy. As shown in Table 5, the model without 1D-CNN will result in overall performance degradation. This can be attributed to the curse of dimensionality, as excessively high feature dimensions will lead to overfitting and reduced generalization capability. Furthermore, quantitative evaluation across varying channel reduction rates demonstrated that the configuration with 32 output channels provides the trade-off between segmentation accuracy and computational burden.

5. Conclusions

The diagnosis of cholangiocarcinoma is a major challenge in the field of medical image processing. For its atypical features, it is extremely difficult to make a high-precision diagnosis even based on H&E-stained pathological images. In this study, DC-MSSFF Net is proposed for semantic segmentation of cholangiocarcinoma HR-HSI. This approach fully utilizes the spatial information and spectral information of HR-HSI in parallel channels to achieve pathological image segmentation through a GwG convolution network and a multi-scale CNN architecture, respectively. To further improve the performance and reliability of the model, ensemble learning framework is also applied, where each DC-MSSFF Net acts as a base learner. The result demonstrated that the Dice score of the Cholangiocarcinoma HR-HSI data can reach 70.47 after ensemble learning, which outperforms the SOTA method and RGB-based image segmentation methods despite the limited amount of the microscopic hyperspectral choledoch pathology

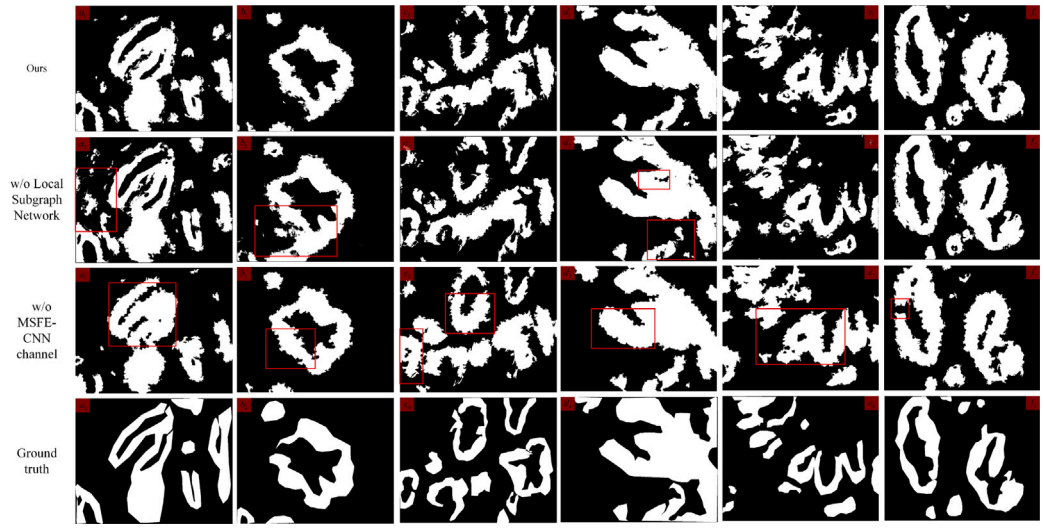

Fig. 10. The generated masks of the HR-HSI in ablation studies. (a_1)–(f_1) show the result of our DC-MSSFF Net. (a_2)–(f_2) are the model w/o local subgraph network in MSFE-GCN channel. (a_3)–(f_3) show the model w/o MSFE-CNN channel. (a_4)–(f_4) are the corresponding ground truth (White areas represent the positive regions, while black areas denote the negative regions).

data. Qualitative and quantitative comparisons is performed with other image segmentation methods, which illustrate the superior performance of the proposed algorithm in pathological HR-HSI segmentation. Notably, this framework pioneers the inductive learning task of deep frameworks based on the joint of GCN and CNN for semantic segmentation of HR-HSI data.

However, the DC-MSSFF Net still faces certain limitations in representing heterogeneous features across superpixels within each hyperspectral data cube and is subject to considerable computational burden. In each hyperspectral cube, there exists significant diversity among negative samples associated with individual superpixels; the model, however, does not currently account for the influence of these heterogeneous negative samples, which can adversely impact segmentation performance to some extent. As part of future work, we plan to develop a diversity-based negative sample selection algorithm to enhance the model's robustness and overall segmentation accuracy. Moreover, given the substantial computational demands associated with processing large-scale hyperspectral datasets, balancing computational complexity with segmentation performance emerges as a critical challenge that must be urgently addressed.

CRediT authorship contribution statement

Meiyan Liang: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Conceptualization. **Zelin Xi:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Bo Li:** Writing – review & editing. **Lin Wang:** Visualization, Investigation.

Declaration of competing interest

The authors declared that they have no conflicts of interest in this work.

Acknowledgments

This work is supported in part by Natural Science Foundation of Shanxi Province, China under Grant 202303021211014, National Natural Science Foundation of China under Grant 11804209. It is also supported in part by Shanxi Scholarship Council of China 2023-010, China Postdoctoral Science Foundation, China 2023M742577, Postgraduate Education Innovation Program of Shanxi Province, China 2024SJ016 and Continued Funding Project for High-level Research Achievements of Shanxi Bethune Hospital (Shanxi Academy of Medical Sciences), China (2024GSPYJ08).

Data availability

The Hyperspectral choledochal database is available from the kaggle platform data sharing (https://www.kaggle.com/datasets/ethelzq/multiDimension-choledoch-database).

References

[1] P.J. Brindley, M. Bachini, S.I. Ilyas, S.A. Khan, A. Loukas, A.E. Sirica, B.T. Teh, S. Wongkham, G.J. Gores, Cholangiocarcinoma, Nat. Rev. Dis. Prim. 7 (1) (2021) 65, http://dx.doi.org/10.1038/s41572-021-00300-2.

[2] K.N. Lazaridis, G.J. Gores, Cholangiocarcinoma, Gastroenterology 128 (6) (2005) 1655–1667, http://dx.doi.org/10.1053/j.gastro.2005.03.040.

[3] S. Seidlitz, J. Sellner, J. Odenthal, B. Özdemir, A. Studier-Fischer, S. Knödler, L. Ayala, T.J. Adler, H.G. Kenngott, M. Tizabi, et al., Robust deep learning-based semantic organ segmentation in hyperspectral images, Med. Image Anal. 80 (2022) 102488, http://dx.doi.org/10.1016/j.media.2022.102488.

[4] Y. Tarabalka, J. Chanussot, J.A. Benediktsson, Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers, IEEE Trans. Syst. Man Cybern. B 40 (5) (2010) 1267–1279, http://dx.doi.org/10.1109/TSMCB.2009.2037132.

[5] M. Paoletti, J. Haut, J. Plaza, A. Plaza, Deep learning classifiers for hyperspectral imaging: A review, ISPRS J. Photogramm. Remote Sens. 158 (2019) 279–317, http://dx.doi.org/10.1016/j.isprsjprs.2019.09.006, URL: https://www.sciencedirect.com/science/article/pii/S0924271619302187.

[6] G. Lu, X. Qin, D. Wang, S. Muller, H. Zhang, A. Chen, Z.G. Chen, B. Fei, Hyperspectral imaging of neoplastic progression in a mouse model of oral carcinogenesis, in: Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 9788, SPIE, 2016, pp. 252–259, http://dx.doi.org/10.1117/12.2216553.

[7] A. Madooei, R.M. Abdlaty, L. Doerwald-Munoz, J. Hayward, M.S. Drew, Q. Fang, J. Zerubia, Hyperspectral image processing for detection and grading of skin erythema, in: Medical Imaging 2017: Image Processing, vol. 10133, SPIE, 2017, pp. 577–583, http://dx.doi.org/10.1117/12.2254132.

[8] N. Kumar, P. Uppala, K. Duddu, H. Sreedhar, V. Varma, G. Guzman, M. Walsh, A. Sethi, Hyperspectral tissue image segmentation using semi-supervised NMF and hierarchical clustering, IEEE Trans. Med. Imaging 38 (5) (2019) 1304–1313, http://dx.doi.org/10.1109/TMI.2018.2883301.

[9] C. Rodarmel, J. Shan, Principal component analysis for hyperspectral image classification, Surv. Land Inf. Sci. 62 (2) (2002) 115–122.

[10] E. Aloupogianni, M. Ishikawa, T. Ichimura, M. Hamada, T. Murakami, A. Sasaki, K. Nakamura, N. Kobayashi, T. Obi, Effects of dimension reduction of hyperspectral images in skin gross pathology, Ski. Res. Technol. 29 (2) (2023) e13270, http://dx.doi.org/10.1111/srt.13270.

[11] M. Nathan, A.S. Kabatznik, A. Mahmood, Hyperspectral imaging for cancer detection and classification, in: 2018 3rd Biennial South African Biomedical Engineering Conference, SAIBMEC, Stellenbosch, South Africa, 2018, pp. 1–4, http://dx.doi.org/10.1109/SAIBMEC.2018.8363180.

[12] G. Lu, D. Wang, X. Qin, S. Muller, J.V. Little, X. Wang, A.Y. Chen, G. Chen, B. Fei, Histopathology feature mining and association with hyperspectral imaging for the detection of squamous neoplasia, Sci. Rep. 9 (1) (2019) 17863, http://dx.doi.org/10.1038/s41598-019-54139-5.

[13] A.F. Goetz, Three decades of hyperspectral remote sensing of the Earth: A personal view, Remote Sens. Environ. 113 (2009) S5–S16, http://dx.doi.org/10.1016/j.rse.2007.12.014.

[14] A. Sellami, M. Farah, I. Riadh Farah, B. Solaiman, Hyperspectral imagery classification based on semi-supervised 3-D deep neural network and adaptive band selection, Expert Syst. Appl. 129 (2019) 246–259, http://dx.doi.org/10.1016/j.eswa.2019.04.006, URL: https://www.sciencedirect.com/science/article/pii/S0957417419302374.

[15] F. Zhao, J. Zhang, Z. Meng, H. Liu, Z. Chang, J. Fan, Multiple vision architectures-based hybrid network for hyperspectral image classification, Expert Syst. Appl. 234 (2023) 121032, http://dx.doi.org/10.1016/j.eswa.2023.121032, URL: https://www.sciencedirect.com/science/article/pii/S0957417423015348.

[16] W. Ye, W. Zhang, W. Lei, W. Zhang, X. Chen, Y. Wang, Remote sensing image instance segmentation network with transformer and multi-scale feature representation, Expert Syst. Appl. 234 (2023) 121007, http://dx.doi.org/10.1016/j.eswa.2023.121007, URL: https://www.sciencedirect.com/science/article/pii/S0957417423015099.

[17] R. Dian, T. Shan, W. He, H. Liu, Spectral super-resolution via model-guided cross-fusion network, IEEE Trans. Neural Networks Learn. Syst. 35 (7) (2024) 10059–10070, http://dx.doi.org/10.1109/TNNLS.2023.3238506.

[18] R. Dian, Y. Liu, S. Li, Low-rank transformer for high-resolution hyperspectral computational imaging, Int. J. Comput. Vis. (2025).

[19] B. Qin, S. Feng, C. Zhao, W. Li, R. Tao, J. Zhou, Language-enhanced dual-level contrastive learning network for open-set hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 63 (2025) 1–14, http://dx.doi.org/10.1109/TGRS.2025.3549049.

[20] G. Lu, B. Fei, Medical hyperspectral imaging: A review, J. Biomed. Opt. 19 (1) (2014) 010901, http://dx.doi.org/10.1117/1.JBO.19.1.010901, URL: https://doi.org/10.1117/1.JBO.19.1.010901.

[21] V. Dremin, Z. Marcinkevics, E. Zherebtsov, A. Popov, A. Grabovskis, H. Kronberga, K. Geldnere, A. Doronin, I. Meglinski, A. Bykov, Skin complications of diabetes mellitus revealed by polarized hyperspectral imaging and machine learning, IEEE Trans. Med. Imaging 40 (4) (2021) 1207–1216, http://dx.doi.org/10.1109/TMI.2021.3049591.

[22] G. Cheng, Z. Li, J. Han, X. Yao, L. Guo, Exploring hierarchical convolutional features for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 56 (11) (2018) 6712–6722, http://dx.doi.org/10.1109/TGRS.2018.2841823.

[23] Y. Chen, H. Jiang, C. Li, X. Jia, P. Ghamisi, Deep feature extraction and classification of hyperspectral images based on convolutional neural networks, IEEE Trans. Geosci. Remote Sens. 54 (10) (2016) 6232–6251, http://dx.doi.org/10.1109/TGRS.2016.2584107.

[24] S.K. Roy, G. Krishna, S.R. Dubey, B.B. Chaudhuri, Hybridsn: Exploring 3-d–2-d CNN feature hierarchy for hyperspectral image classification, IEEE Geosci. Remote. Sens. Lett. 17 (2) (2020) 277–281, http://dx.doi.org/10.1109/LGRS.2019.2918719.

[25] D. Wang, B. Du, L. Zhang, Y. Xu, Adaptive spectral–spatial multiscale contextual feature extraction for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 59 (3) (2021) 2461–2477, http://dx.doi.org/10.1109/TGRS.2020.2999957.

[26] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241, http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[27] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: Redesigning skip connections to exploit multiscale features in image segmentation, IEEE Trans. Med. Imaging 39 (6) (2020) 1856–1867, http://dx.doi.org/10.1109/TMI.2019.2959609.

[28] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, 2021, arXiv preprint arXiv:2102.04306.

[29] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention U-net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.

[30] S. Trajanovski, C. Shan, P.J.C. Weijtmans, S.G.B. de Koning, T.J.M. Ruers, Tongue tumor detection in hyperspectral images using deep learning semantic segmentation, IEEE Trans. Biomed. Eng. 68 (4) (2021) 1330–1340, http://dx.doi.org/10.1109/TBME.2020.3026683.

[31] B. Qin, S. Feng, C. Zhao, B. Xi, W. Li, R. Tao, FDGNet: Frequency disentanglement and data geometry for domain generalization in cross-scene hyperspectral image classification, IEEE Trans. Neural Networks Learn. Syst. (2024) 1–14, http://dx.doi.org/10.1109/TNNLS.2024.3445136.

[32] G. Zhan, Y. Uwamoto, Y.-W. Chen, HyperUNet for medical hyperspectral image segmentation on a choledochal database, in: 2022 IEEE International Conference on Consumer Electronics, ICCE, Las Vegas, NV, USA, 2022, pp. 1–5, http://dx.doi.org/10.1109/ICCE53296.2022.9730171.

[33] Q. Wang, L. Sun, Y. Wang, M. Zhou, M. Hu, J. Chen, Y. Wen, Q. Li, Identification of melanoma from hyperspectral pathology image using 3D convolutional networks, IEEE Trans. Med. Imaging 40 (1) (2021) 218–227, http://dx.doi.org/10.1109/TMI.2020.3024923.

[34] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, Y.Y. Tang, Spectral–spatial graph convolutional networks for semisupervised hyperspectral image classification, IEEE Geosci. Remote. Sens. Lett. 16 (2) (2019) 241–245, http://dx.doi.org/10.1109/LGRS.2018.2869563.

[35] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, J. Chanussot, Graph convolutional networks for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 59 (7) (2021) 5966–5978, http://dx.doi.org/10.1109/TGRS.2020.3015157.

[36] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, J. Yang, Multiscale dynamic graph convolutional network for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 58 (5) (2020) 3162–3177, http://dx.doi.org/10.1109/TGRS.2019.2949180.

[37] Y. Ding, Z. Zhang, X. Zhao, D. Hong, W. Cai, C. Yu, N. Yang, W. Cai, Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification, Neurocomputing 501 (2022) 246–257, http://dx.doi.org/10.1016/j.neucom.2022.06.031, URL: https://www.sciencedirect.com/science/article/pii/S0925231222007329.

[38] S. Feng, H. Zhang, B. Xi, C. Zhao, Y. Li, J. Chanussot, Cross-domain few-shot learning based on decoupled knowledge distillation for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–14, http://dx.doi.org/10.1109/TGRS.2024.3476116.

[39] R.S. Michalski, 4 — A theory and methodology of inductive learning, in: R.S. Michalski, J.G. Carbonell, T.M. Mitchell (Eds.), Machine Learning, Morgan Kaufmann, San Francisco (CA), 1983, pp. 83–134, http://dx.doi.org/10.1016/B978-0-08-051054-5.50008-X, URL: https://www.sciencedirect.com/science/article/pii/B978008051054550008X.

[40] M. Bianchini, A. Belahcen, F. Scarselli, A comparative study of inductive and transductive learning with feedforward neural networks, in: G. Adorni, S. Cagnoni, M. Gori, M. Maratea (Eds.), AI*IA 2016 Advances in Artificial Intelligence, Springer International Publishing, Cham, 2016, pp. 283–293, http://dx.doi.org/10.1007/978-3-319-49130-1_21.

[41] A. Rossi, M. Tiezzi, G.M. Dimitri, M. Bianchini, M. Maggini, F. Scarselli, Inductive–transductive learning with graph neural networks, in: L. Pancioni, F. Schwenker, E. Trentin (Eds.), Artificial Neural Networks in Pattern Recognition, Springer International Publishing, Cham, 2018, pp. 201–212, http://dx.doi.org/10.1007/978-3-319-99978-4_16.

[42] Q. Zhang, Q. Li, G. Yu, L. Sun, M. Zhou, J. Chu, A multidimensional choledoch database and benchmarks for cholangiocarcinoma diagnosis, IEEE Access 7 (2019) 149414–149421, http://dx.doi.org/10.1109/ACCESS.2019.2947470.

[43] M. Ganaie, M. Hu, A. Malik, M. Tanveer, P. Suganthan, Ensemble deep learning: A review, Eng. Appl. Artif. Intell. 115 (2022) 105151, http://dx.doi.org/10.1016/j.engappai.2022.105151, URL: https://www.sciencedirect.com/science/article/pii/S095219762200269X.

[44] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, Adv. Neural Inf. Process. Syst. 30 (2017).

[45] K. Wu, J. Fan, P. Ye, M. Zhu, Hyperspectral image classification using spectral–spatial token enhanced transformer with hash-based positional embedding, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–16, http://dx.doi.org/10.1109/TGRS.2023.3258488.

[46] H. Gao, S. Ji, Graph U-nets, IEEE Trans. Pattern Anal. Mach. Intell. 44 (9) (2022) 4948–4960, http://dx.doi.org/10.1109/TPAMI.2021.3081010.