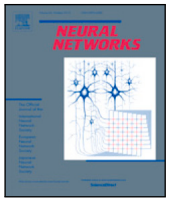
Neural Networks 178 (2024) 106489

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



Full Length Article

NFMPAtt-Unet: Neighborhood Fuzzy C-means Multi-scale Pyramid Hybrid Attention Unet for medical image segmentation

Xinpeng Zhao, Weihua Xu *

College of Artificial Intelligence, Southwest University, Chongqing, 400715, PR China

ARTICLE INFO

Keywords:
Fuzzy C-means
Medical image segmentation
Neighborhood rough set
U-net

ABSTRACT

Medical image segmentation is crucial for understanding anatomical or pathological changes, playing a key role in computer-aided diagnosis and advancing intelligent healthcare. Currently, important issues in medical image segmentation need to be addressed, particularly the problem of segmenting blurry edge regions and the generalizability of segmentation models. Therefore, this study focuses on different medical image segmentation tasks and the issue of blurriness. By addressing these tasks, the study significantly improves diagnostic efficiency and accuracy, contributing to the overall enhancement of healthcare outcomes. To optimize segmentation performance and leverage feature information, we propose a Neighborhood Fuzzy c-Means Multiscale Pyramid Hybrid Attention Unet (NFMPAtt-Unet) model. NFMPAtt-Unet comprises three core components: the Multiscale Dynamic Weight Feature Pyramid module (MDWFP), the Hybrid Weighted Attention mechanism (HWA), and the Neighborhood Rough Set-based Fuzzy c-Means Feature Extraction module (NFCMFE). The MDWFP dynamically adjusts weights across multiple scales, improving feature information capture. The HWA enhances the network's ability to capture and utilize crucial features, while the NFCMFE, grounded in neighborhood rough set concepts, aids in fuzzy C-means feature extraction, addressing complex structures and uncertainties in medical images, thereby enhancing adaptability. Experimental results demonstrate that NFMPAtt-Unet outperforms state-of-the-art models, highlighting its efficacy in medical image segmentation.

1. Introduction

Medical image segmentation stands as a pivotal technique within healthcare and diagnostics, playing a crucial role in extracting meaningful information from intricate medical images. As advancements in imaging technologies continue to burgeon, the demand for accurate and efficient methods to delineate and analyze anatomical structures becomes increasingly paramount. In this context, image segmentation serves as a linchpin, enabling the precise identification and isolation of specific regions of interest within medical images.

Traditional medical image segmentation methods typically include techniques such as threshold-based (Bhargavi & Jyothi, 2014), region-based (Lewis, O'Callaghan, Nikolov, Bull, & Canagarajah, 2007), and edge detection-based approaches (Huang, Tung, Chen, Wang, & Wu, 2005). While these methods can yield satisfactory segmentation results to some extent, they may encounter limitations, particularly when dealing with complex and diverse medical images. For instance, threshold-based methods may struggle with images exhibiting complex grayscale distributions, while region-based methods may fail to accurately segment structures with irregular shapes. Therefore, as the complexity and diversity of medical images increase, traditional approaches may prove to be less flexible and accurate in certain scenarios. In contrast, deep learning-based medical image segmentation methods often demonstrate superior performance when dealing with complex and diverse medical images.

In contrast, significant strides have been made in the field of medical image segmentation with the advancement of deep learning. Introducing models like U-Net (Ronneberger, Fischer, & Brox, 2015), U-Net++(Zhou, Rahman Siddiquee, Tajbakhsh, & Liang, 2018), ResUNet++(Jha et al., 2019), nnU-Net (Isensee, Jaeger, Kohl, Petersen, & Maier-Hein, 2021), and TransUNet (Chen et al., 2021) has greatly enhanced the accuracy and robustness of medical image segmentation tasks. These models leverage the potent feature learning capabilities of deep learning, particularly convolutional neural networks (CNNs), to improve segmentation effectiveness through techniques such as skip connections, dense feature fusion, multi-scale receptive fields, and attention mechanisms. However, medical image segmentation still encounters challenges such as noise, blurred boundaries, and inadequate low-level features. These issues may impact the performance of deep learning models, especially when dealing with medical images featuring complex structures and low contrast. Difficulties arise from relying

* Corresponding author.
E-mail addresses: zhaoxpengpeng@163.com (X. Zhao), chxuwh@gmail.com (W.H. Xu).

https://doi.org/10.1016/j.neunet.2024.106489
Received 7 February 2024; Received in revised form 27 May 2024; Accepted 20 June 2024
Available online 22 June 2024

solely on low-level features, as semantic features alone may fail to provide precise boundary information, resulting in inferior boundary quality of segmentation results.

Therefore, to address the issues of noise and blurry boundaries in medical images, we propose a feature extraction module based on neighborhood rough set and fuzzy C-means. Neighborhood rough set is an extension of rough set theory that focuses more on the local structure and neighborhood information within the data (Guo et al., 2024). It considers the relationships between elements in the dataset, allowing for a more accurate description of data features and local patterns. By introducing neighborhood rough set into fuzzy C-means, we aim to extract features from medical images, thus addressing problems such as blurry boundaries and uncertainty in the images.

In the field of image processing, particularly in the medical domain, rough set methods are widely utilized. Their advantage lies in their ability to effectively handle complex information within medical images and extract useful features and patterns from them. The scope of application of this method encompasses various aspects such as image denoising (Phophalia, Rajwade, & Mitra, 2014), image segmentation (Hirano & Tsumoto, 2002), and image classification (Jothi et al., 2016). Its uniqueness lies in its capability to deal with uncertainty and fuzziness in the data, thereby enabling more accurate and reliable results in medical image analysis.

Additionally, to enhance feature selectivity and multi-scale receptive fields while addressing the challenge of relying on low-level features, we propose a Multi-scale Dynamic Weight Feature Pyramid (MDWFP) module and a Hybrid Weighted Attention mechanism (HWA). The MDWFP module dynamically adjusts the weights of different scales to better capture feature information at various scales, enhancing the network's sensitivity to information. By adjusting the weights across multiple scales, it improves feature diversity and representation, thus aiding in enhancing model performance.

The Hybrid Weighted Attention mechanism combines spatial and channel attention by dynamically adjusting weights across different scales to adapt to various tasks and datasets. It evaluates scales through convolutional layers, then normalizes them using the softmax function to obtain dynamic weights. This mechanism dynamically adjusts weights in both spatial and channel domains, reducing redundancy in descriptions. By integrating features from different scales, it improves adaptability and the ability to capture relevant features.

The paper presents a network model called NFMPAtt-Unet, which serves as an enhancement of the U-Net architecture. It incorporates three key components: the multi-scale dynamic weight feature pyramid module (MDWFP), hybrid weighted attention mechanism (HWA), and the fuzzy C-means feature extraction module based on neighborhood rough set (NFCMFE). The primary contributions of this paper are as follows:

- A multi-scale dynamic weight feature pyramid module is proposed. By introducing this module, the network is allowed to dynamically adjust weights across multiple scales, facilitating better capturing of feature information at different scales. This helps enhance the network's sensitivity to information at various scales, thereby boosting feature diversity and expressive capability.
- A novel hybrid weighted attention mechanism is proposed. The incorporation of a hybrid weighted attention mechanism in the network enables a more targeted focus on specific regions, thereby enhancing the model's attention to critical information. This mechanism augments the network's capacity to capture and leverage important features effectively.
- A fuzzy C-means feature extraction module based on neighborhood rough set is proposed. This module leverages the concept of neighborhood rough sets for fuzzy C-means feature extraction, aiding in better handling complex structures and uncertainties in medical images. It enhances the adaptability of the network to image features.
- To validate the effectiveness and superiority of our model, we conducted experiments comparing it with eight different U-Net models on six medical image datasets. The experimental results demonstrate that our proposed model outperforms the others.

The paper's structure is outlined as follows: The second section focuses on related work, the third section presents a detailed description of our proposed network model and its core components, the fourth section conducts experiments comparing our model with other state-of-the-art models, and the concluding fifth section summarizes the findings and conclusions of the article.

2. Related work

This section offers an overview of methods utilized in medical image segmentation, emphasizing a summary of the frequently employed U-Net model and its associated approaches. Subsequently, we delve into a detailed summary of the widely used FCM (Fuzzy C-means) algorithm. Furthermore, we encapsulate the concepts of feature fusion, extraction fusion methods, and attention mechanisms in the context of medical image segmentation.

2.1. Medical image segmentation based on U-Net

In 2015 Ronneberger et al. (2015), introduced the U-Net network model tailored for biomedical image analysis. Renowned for its streamlined network architecture and enhanced generalization capabilities, the model garnered significant attention immediately upon its publication. Compared to other convolutional neural network models, U-Net stands out for its simpler structure, stronger generalization capabilities, and robust feature extraction abilities, which are particularly advantageous for medical image analysis. Subsequently, numerous scholars have made improvements upon the U-Net model, introducing various enhanced versions.

In 2018, OZan et al. Oktay et al. (2018) introduced the Attention U-Net model by incorporating an attention gate module into the standard U-Net architecture. This model automatically learns to focus on the shape and size of target structures, suppressing irrelevant regions and highlighting useful salient features. It demonstrated superior performance in tasks such as pancreatic segmentation. However, it did not fully exploit features from different scales and depths, as it focused solely on feature maps with the same scale. Subsequently, Zhou et al. (2018) enhanced the U-Net architecture by incorporating multiple skip connections, forming a densely connected network. Through a redesign of the skip connections and upsampling modules, this network facilitates improved fusion of feature maps at different depths, thereby enhancing segmentation accuracy and robustness. However, its training necessitates the use of a deep supervision loss function, potentially increasing the difficulty and instability of the training process. In 2020, Huang et al. (2020) introduced UNet3+, a neural network designed for medical image segmentation. UNet3+ leverages full-scale skip connections and deep supervision to fuse feature maps at various scales and depths, thereby enhancing segmentation accuracy and robustness. In 2021, Sha et al. integrated Transformer modules with the U-Net architecture, as described in their work (Sha, Zhang, Ji, & Hu, 2021). This integration successfully accomplished a mapping from low-quality images to high-quality images. Noteworthy for its ability to simultaneously process spatial and frequency-domain information, this approach enhances image contrast, clarity, and detail. In 2023, Iqbal et al. Iqbal and Sharif (2023) introduced PDF-UNet, a semi-supervised neural network designed for breast tumor image segmentation. The proposed approach incorporates a U-shaped pyramid dilated network to boost both the accuracy and robustness of the segmentation process. Building upon the aforementioned research and to fully leverage image features, we integrate a novel FCM feature selection module based on neighborhood rough sets into the U-shaped network. Leveraging the

capabilities of neighborhood rough sets in handling local neighborhoods and uncertainty, this module enhances the model's ability to extract fuzzy and local features effectively. Additionally, by introducing a multiscale dynamic weight feature pyramid module, the network dynamically adjusts the weights of different scales, enabling better capture of feature information across various scales and enhancing the network's sensitivity to information. The incorporation of a hybrid weighted attention mechanism allows the model to selectively focus on specific regions, thereby enhancing the model's attention to key information and improving feature utilization efficiency.

2.2. Image segmentation with Fuzzy C-Means method

FCM (Bezdek, Ehrlich, & Full, 1984) is a soft clustering method based on fuzzy set theory, capable of handling data uncertainty and fuzziness. Image segmentation based on FCM (Yu, Jiang, Fan, Xie, & Lan, 2024) involves using the Fuzzy C-Means clustering algorithm to categorize pixels in an image. It partitions the image into several regions based on pixel intensity or color values, achieving segmentation. The advantage of FCM-based image segmentation lies in its ability to handle uncertainty and fuzziness in images, particularly in cases where category boundaries are not well-defined. However, drawbacks include the need to specify the number of clusters beforehand, sensitivity to noise and outliers, and computational complexity.

Gong et al. introduced a balanced weighted fuzzy factor in their work (Gong, Liang, Shi, Ma, & Ma, 2012). This factor takes into account both spatial distance and grayscale differences of all neighboring pixels simultaneously. This factor accurately estimates the decay level of neighboring pixels. Furthermore, they incorporated a kernel distance metric to replace the Euclidean distance, thereby improving the algorithm's robustness against noise and outliers. Tang et al. presented a fuzzy c-means clustering algorithm based on image blocks and structural similarity for image segmentation in their work (Tang, Ren, & Pedrycz, 2020). Guo et al. introduced an improved fuzzy c-means clustering algorithm in their work (Guo, Shi, Chen, Chen & and Ding, 2023). This enhancement involved the introduction of a new affinity matrix to store and incorporate spatial information of the image as a prior, contributing to the regularization of the fuzzy clustering method and yielding superior segmentation results. They also introduced a new coefficient to control the update of the membership matrix, making it more accurately reflect the membership of image blocks.

Neighborhood rough set (Pan, Xu, & Ran, 2023) can leverage various domain functions to characterize the similarity or dissimilarity between objects, thereby obtaining diverse rough approximations (Xu et al., 2023). To comprehensively consider both grayscale and spatial information of pixels, we integrate neighborhood rough set with fuzzy c-means clustering. We propose a method based on the neighborhood FCM module, incorporating it as a component of the U-Net network to enhance U-Net's capability in capturing overall features.

2.3. Multi-scale feature extraction and fusion

The pyramid module, as introduced in the work by Zhao, Shi, Qi, Wang, and Jia (2017), is a network structure specifically designed for feature extraction. This module facilitates the generation of global scene prior information on the final layer feature map of deep neural networks. By doing so, it enhances the network's capability to capture global information, providing a more comprehensive understanding of the overall context within the input data (Guo, Xu, Qian, & Ding, 2023). The pyramid module is particularly valuable in tasks where global context is crucial, such as in the analysis of complex scenes or in the segmentation of medical images with diverse structures. The fundamental idea of the pyramid module is to obtain feature maps of multiple sizes by employing pooling operations at different scales. Following this process, the feature maps, which include the original feature map, are concatenated along the channel dimension. This results in a composite feature map that effectively integrates information from multiple scales, contributing to a more comprehensive representation of the input data. The concatenation along the channel dimension ensures that the model can leverage features from different scales in a unified manner, enhancing its ability to capture both local and global details during subsequent processing steps. The advantage of the pyramid module lies in its ability to simultaneously consider global semantic information and local details, thereby improving the robustness and flexibility of the network.

Hierarchical parsing net (Shi et al., 2018) introduces a context feature encoding and fusion mechanism. This mechanism enhances the local features of each object, considering both scene–object context and object–object context. In RAPNet (Zhang, Liu, Lei, Wang, & Lu, 2020), a residual atrous spatial pyramid (RASP) module is introduced. This module sequentially aggregates contextual information from global to local scales, enhancing label consistency in a residual-refinement manner. In the P2T model introduced by Wu, Liu, Zhan, and Cheng (2022), a pyramid pooling module is utilized to improve the multi-head self-attention mechanism. This module serves the dual purpose of reducing the sequence length of image labels while capturing powerful contextual features. By incorporating this pyramid pooling module into the attention mechanism, the model is better equipped to handle and process contextual information across different scales, ultimately improving its performance in tasks such as image labeling. P2T stands out for its capacity to balance global and local information, taking into account both spatial and channel dimensions of the feature map. This equilibrium enhances the expressive power and robustness of the features, reducing redundancy and contributing to more effective and efficient processing.

The feature pyramid's advantage lies in its capability to strike a balance between global and local information, taking into account both spatial and channel dimensions of the feature map. This approach enhances feature expression and robustness, providing a comprehensive representation of the input data. Building upon this, we propose a multi-scale dynamic weight feature pyramid module, assigning different weights to various scales to ensure more accurate fusion of the obtained feature maps.

2.4. Attention mechanism in deep learning

In the realm of deep learning, the attention mechanism, as described by Li, Jin, Zhou, Kubota, and Ju (2020), mimics the human visual and cognitive systems. This approach enables neural networks to focus their attention on relevant parts while processing input data. Through the incorporation of attention mechanisms, neural networks can autonomously learn and selectively attend to crucial information in the input, thereby enhancing the model's performance and generalization capabilities.

The most typical attention mechanisms include self-attention mechanism (Vaswani et al., 2017), spatial attention mechanism (Jaderberg, Simonyan, Zisserman, & kavukcuoglu, 2015), and temporal attention mechanism (Yao et al., 2015). The integration of attention mechanisms empowers the model to assign varying weights to different positions in the input sequence. This capability allows the model to focus on the most relevant parts during the processing of each sequence element, enhancing its ability to capture and prioritize important information. In their work, Hu, Li, Zhao, and Zhang (2020) introduced a parallel deep learning algorithm that incorporates a hybrid attention mechanism into DenseNet. This approach effectively extracts and integrates spatial and channel features from images. In the TransAttUnet model, as proposed by Chen, Liu, Zhang, Lu, and Kong (2023), the traditional U-Net structure is enhanced through the incorporation of a self-aware attention module. This SAA module consists of transformer self-attention and global spatial attention, aiming to augment non-local interactions between encoder features.

The channel attention mechanism assesses the importance of each channel by calculating its significance, while the spatial attention mechanism incorporates an attention module that enables the model to dynamically learn attention weights for distinct regions. This combination enhances the model's ability to focus on relevant features both within individual channels and across different spatial regions, minimizing redundancy and improving its overall adaptability. By combining them, the model can simultaneously capture information about both channels and regions. Hence, our proposal introduces a hybrid weighted attention mechanism that concurrently integrates the spatial attention mechanism and channel attention mechanism. This approach assigns distinct weights at various scales, ensuring varied weights for both channel and spatial attention mechanisms across different scales. The parallel combination of these mechanisms aims to enhance the model's adaptability and capture relevant features effectively across different spatial and channel dimensions.

3. The proposed approach

In this section, we initially outline the overall framework of our proposed NFMPAtt-Unet model. Subsequently, we provide a detailed introduction to the core components of the network, including the multi-scale dynamic weight feature pyramid module (MDWFP), the hybrid weighted attention mechanism (HWA), and the fuzzy c-means feature extraction module based on neighborhood rough set (NFCMFE).

3.1. NFMPAtt-Unet

As illustrated in Fig. 1, the framework diagram of the proposed NFMPAtt-Unet model is depicted. Clearly, this network model adopts a U-shaped structure. The left half of the network consists of a series of downsampling operations, including convolutional layers and pooling layers, while the right half performs the opposite upsampling operations. Additionally, the network model includes the proposed multi-scale dynamic weight feature pyramid module (MDWFP), hybrid weighted attention mechanism (HWA), and fuzzy c-means feature extraction module based on neighborhood rough set (NFCMFE).

During downsampling, different-scale features are input into MDWFP for feature fusion based on scale-specific weights. The fused feature map is then input into MWA, where different attention mechanisms are applied. This network model enables the acquisition of rich information at multiple scales, achieving a more comprehensive context through pyramid feature fusion. The attention mechanism allows fine-grained attention adjustments across the entire fused feature map to enhance focus on critical regions for the task. Simultaneously, the original image is input into NFCMFE to obtain processed feature maps. The combination of these two feature maps using a loss function yields the final output result.

To achieve superior results, we have devised a comprehensive loss function, denoted as L, specifically designed for this model. Its mathematical expression is as follows:

$$L = L_{net} + \lambda * L_{FCM} \tag{1}$$

where L_{net} represents the loss function for the lower part of the network, L_{FCM} is the loss function for the NFCMFE module, and λ is the parameter balancing between the network and NFCMFE.

3.2. Fuzzy C-means feature extraction module based on neighborhood rough set

Fig. 2 illustrates the basic framework of our proposed NFCMFE module. Firstly, we conduct image analysis using the theory of neighborhood rough sets. Neighborhood rough set theory is an extension of rough set theory that introduces the concept of neighborhood to better capture spatial correlations between data elements. In our approach, we employ a neighborhood function to compute the upper approximation, lower approximation, and boundary region for each pixel, thus constructing feature spaces with different neighborhood systems. This approach allows us to more accurately describe spatial relationships between pixels, providing a foundation for subsequent image segmentation and clustering.In this paper, we use the Euclidean distance function as the neighborhood function, with the formula as follows:

$$d(P,Q) = \sqrt{(R_p - R_q)^2 + (G_p - G_q)^2 + (B_p - B_q)^2} \tag{2}$$

where P and Q represent two pixels in the image, where $P = (R_p, G_p, B_p)$ and $Q = (R_q, G_q, B_q)$.

Next, we employ the fuzzy c-means (FCM) clustering method for image segmentation. FCM is a classic clustering algorithm designed to partition data points into multiple clusters so that similar data points belong to the same cluster center. In our approach, we integrate both the grayscale values of each pixel and the previously computed neighborhood system to calculate the membership degree of each pixel to each cluster center, thereby obtaining a fuzzy partition of pixels. This method not only considers the grayscale information of pixels but also utilizes the spatial relationships between pixels, enhancing the accuracy and robustness of image segmentation. The formula for calculating the membership degree involved is as follows:

$$d(P,Q) = \sqrt{\sum_{i=1}^{n} (P_i - Q_i)^2} \tag{3}$$

By calculating the neighborhood function between pixels, we can obtain the neighborhood relationship. Let s_{ij} represent the neighborhood relationship between the ith and jth pixel points, taking values of 0 or 1. If $s_{ij} = 1$, it signifies that the ith and jth pixel points belong to the same neighborhood; if $s_{ij} = 0$, it implies that the ith and jth pixel points do not belong to the same neighborhood.

Using the fuzzy c-means clustering method and taking into account the grayscale values of each pixel and the previously computed neighborhood system, compute the membership degree of each pixel to each cluster center. This process results in a fuzzy partition. This step integrates both grayscale and spatial information from the image, establishing the groundwork for subsequent consistency regularization. The traditional FCM distance metric only considers the grayscale or color information of pixel points, overlooking spatial and neighborhood information. To address this issue, we introduced previously computed neighborhood knowledge into the FCM clustering, obtaining the distance metric in FCM clustering. The formula is as follows:

$$d_{ij} = \|x_i - v_j\| + \lambda \sum_{k=1}^{n} s_{ik} \|u_{kj} - u_{ij}\| \tag{4}$$

where x_i denotes the feature vector of the ith pixel point, v_j represents the jth cluster center, u_{ij} denotes the membership degree of the ith pixel point to the jth cluster center, and λ is a regularization parameter regulating the impact of the neighborhood system on the distance metric.

After obtaining the clustering results, we introduce an additional consistency regularization term. The purpose of this term is to promote both local and global consistency in the image by maximizing the similarity of membership degrees between pixels and their neighboring pixels. The design of this regularization term takes into account both global and local consistency. By comparing distances between pixels and their adjacent pixels or region centers, it encourages consistency between different regions and pixels. The objective function for this consistency regularization term is as follows:

$$R(U) = \sum_{i=1}^{o} \sum_{j=1}^{c} u_{ij}^{m} (\alpha\, d(x_i, x_j) + \beta \sum_{k=1}^{c} d(v_j, v_k)) \tag{5}$$

where U is an $n \times c$ membership matrix, with u_{ij}? representing the membership degree of the ith pixel to the jth cluster center. The term $d(x_i, x_j?)$ denotes the distance between the ith and jth pixels, and

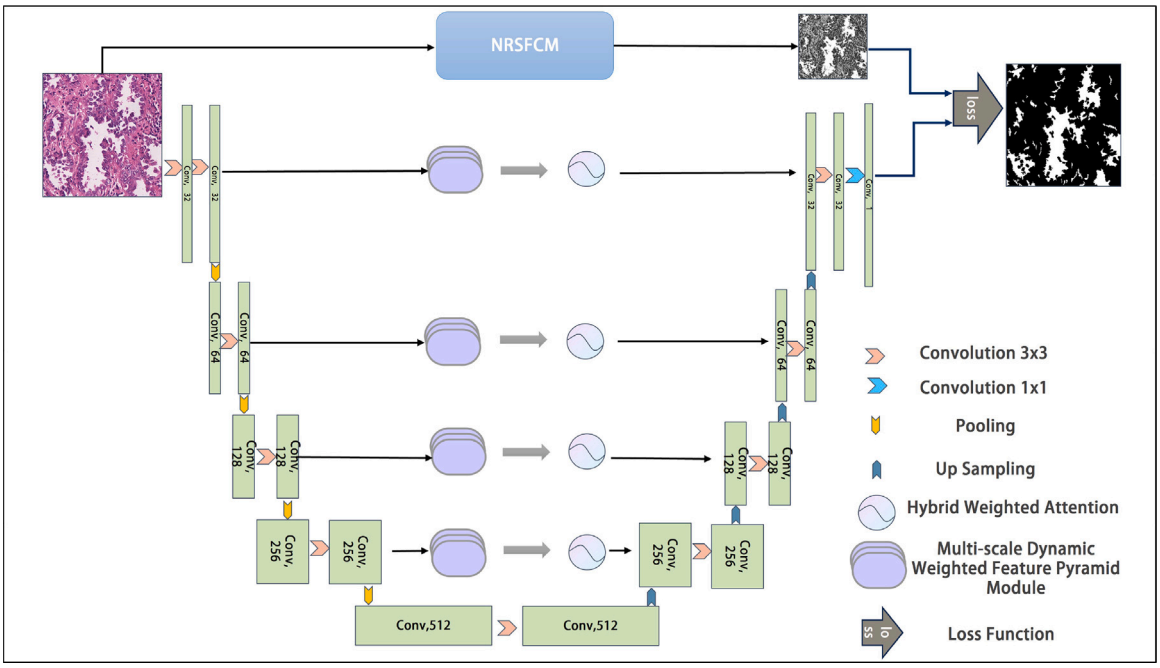


Fig. 1. The proposed NFMPAtt-Unet model framework.

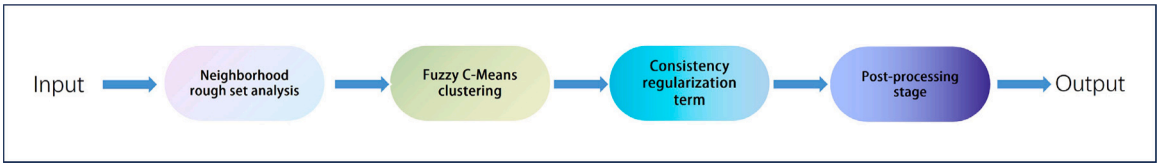


Fig. 2. The fundamental framework of the NFCMFE module begins with the application of neighborhood rough set theory for neighborhood partitioning. Subsequently, the FCM algorithm is employed for clustering and regularization. Finally, some post-processing steps are conducted before the output is generated.

$d(v_j, v_k)$ represents the distance between the jth and kth cluster centers. The parameter m is a fuzziness exponent greater than 1, while α and β are two regularization parameters utilized to control the weights of local and global consistency.

Based on the defined regularization term $R(U)$, we can derive a loss function $L(LCM)$ related to FCM, the formula of which is as follows:

$$L(FCM) = L(U,V) = J_m(U,V) + \lambda R(U) - \mu D(U) \tag{6}$$

$$J_m(U,V) = -\sum_{i=1}^{n}\sum_{j=1}^{c} \omega_{ij}[u_{ij}^m log(v_{ij})] \tag{7}$$

$$D(U) = \frac{2\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{gt} u_{ij}^{pred}}{\sum_{i=1}^{n}\sum_{j=1}^{c}(u_{ij}^{gt} + u_{ij}^{pred})} \tag{8}$$

we need to determine the degree of membership of each pixel to each cluster center, which is the purpose of the U matrix. The U matrix has a size of $n \times c$, where n represents the number of pixels in the image, and c represents the number of cluster centers. In the U matrix, each row represents a pixel, and each column represents a cluster center. Each element u_{ij} in the matrix represents the degree to which the ith pixel belongs to the jth cluster center. V is a $c \times p$ matrix representing the positions of the cluster centers, where p denotes the dimensionality of the feature space. Each row in the V matrix represents a cluster center, and each column represents a feature dimension, indicating the position of the cluster center in the feature space. $J_m(U,V)$ is the weighted cross-entropy loss function, used to measure the difference between the fuzzy clustering result and the ground truth labels, where w_{ij} is a weighting term typically used to adjust the contribution of different samples to the loss function, u_{ij} is the membership degree of pixel i to cluster center j, and v_{ij} is the feature value corresponding to cluster center j. $R(U)$ is the consistency regularization term, which promotes the local and global consistency of the image. It maximizes the similarity of membership degrees between adjacent pixels based on the degree of membership of the pixels. This helps ensure that the clustering results have continuity and smoothness in space. $D(U)$ is the Dice loss function, where u_{ij}^{gt} is the membership degree of the ith pixel's ground truth label to the jth cluster center, and u_{ij}^{pred} is the membership degree of the ith pixel's predicted label to the jth cluster center.

Finally, post-processing is applied to the obtained feature maps, including tasks such as removing small regions, filling holes, and smoothing edges, to further enhance the overall effectiveness.

3.3. Multi-scale dynamic weight feature pyramid module

To enhance the utilization of feature maps at different scales, we introduce a multi-scale dynamic weight feature pyramid module (MD-WFP), as depicted in the structural diagram shown in Fig. 3.

Firstly, the input includes feature maps from low to high levels, with each feature map corresponding to a different scale. The scale evaluation module, as shown in Fig. 3, employs convolution for scale evaluation, producing a scalar as the weight for each scale. For each scale i, denoting the scale weight as w_i, the output of the scale evaluation module is $w_1, w_2, \ldots, w_n$. The output is subsequently passed through a softmax function to ensure weight normalization, transforming it into an effective probability distribution. The dynamic weights d_i can be calculated using the following formula:

$$d_i = \frac{e^{w_i}}{\sum_{j=1}^{n} e^{w_j}} \tag{9}$$

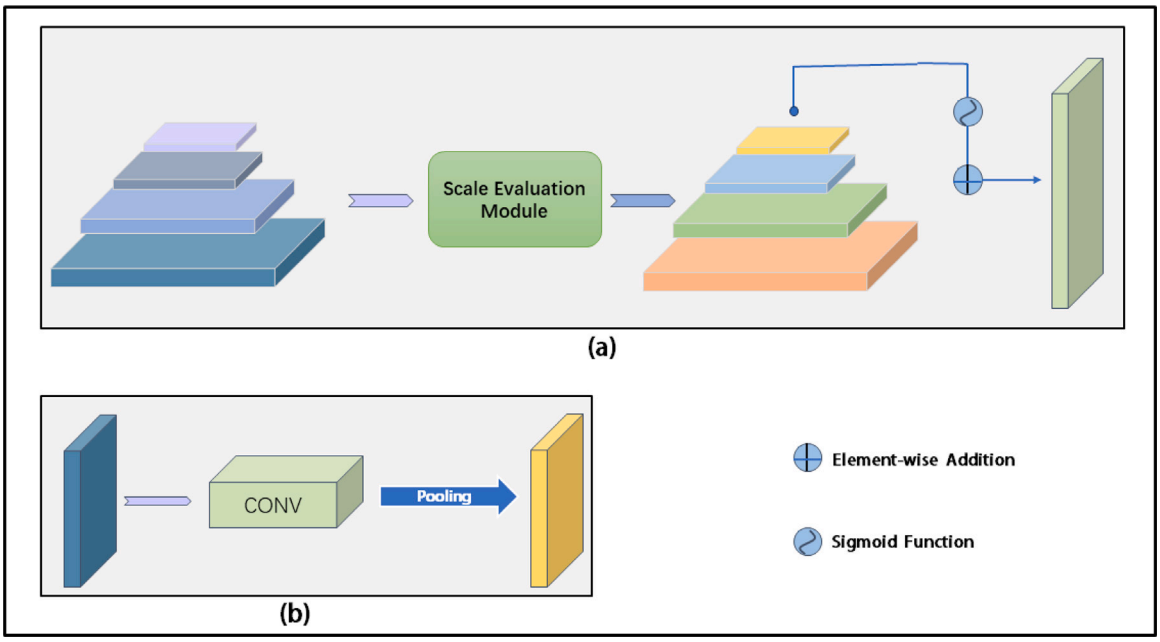


Fig. 3. (a) illustrates the framework of the MDWFP module, which is utilized for evaluating feature inputs of multiple scales and generating corresponding feature weights. These feature weights are then processed by an activation function to obtain the final weighted features. (b) represents the structural diagram of the scale evaluation module in (a).

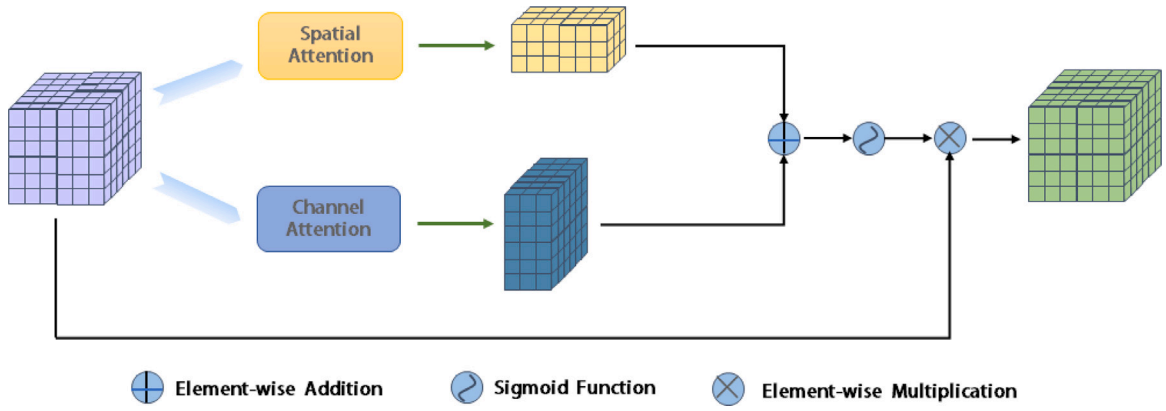


Fig. 4. The framework of the Hybrid Weighted Attention Mechanism involves obtaining different weighted features for channel attention and spatial attention through a weight matrix. Subsequently, these features undergo fusion and activation functions to produce the final feature map.

where e^{w_i} represents the exponent of w_i.

For each scale i, multiply the original feature map by the corresponding dynamic weight d_i. For each spatial position and channel, the fused feature map can be represented as:

$$F_{\text{fused}}(x, y, c) = \sum_{i=1}^{z} d_i \cdot F_i(x, y, c) \tag{10}$$

where F_i represents the feature map for the ith scale. The fused feature map serves as the final output of the module.

3.4. Hybrid weighted attention mechanism

The proposed hybrid weighted attention (MWA) mechanism, shown in Fig. 4, dynamically adjusts the weights in both spatial and channel domains across different scales to adapt to various tasks and datasets. It integrates spatial attention and channel attention to extract and fuse features across various scales, minimizing redundancy in the description. Convolutional layers are employed for scale evaluation, followed by normalization using the softmax function to obtain dynamic weights. Formulas (9), (10), and (11) represent the MWA mechanism, spatial attention scores, and channel context information, respectively.

$$H(Q, K, V) = S(Q, K, V) \odot C(Q, K, V) \tag{11}$$

where $\odot$ denotes element-wise multiplication. The symbols Q, K, V denote the query, key, and value matrices, respectively. These matrices are obtained through linear transformations of the input feature map.

$$S(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{12}$$

where d_k represents the dimension of K, and the Softmax operation normalizes along the last dimension.

$$C(Q, K, V) = \sigma\left(W_c \cdot ReLU\left(W_s \cdot Concat\left(Q, K, V\right)\right)\right) \tag{13}$$

where σ represents the sigmoid function. In this context, W_s and W_c represent two weight matrices, where W_s is used for spatial dimension processing, and W_c is used for channel dimension processing. Concatenate operation is applied to concatenate Q, K, and V along the channel dimension. Subsequently, ReLU and Sigmoid functions are used to calculate the channel attention weights.

Table 1
Information about datasets used.

Dataset	Number of training set	Number of test set	Image size
Eye	54	27	4288 × 2848
WSSS4LUAD	80	40	256 × 256
Lung	632	72	3000 × 2919
FootUlcer	810	200	512 × 512
Cell	1000	101	512 × 512
ISIC	2000	150	512 × 512

By introducing scale adaptability, the hybrid-weighted attention mechanism has made significant progress in flexibility, enabling neural networks to intelligently learn and adapt to diverse scale features in different tasks or scenarios. This advancement facilitates more effective capture and processing of various spatial scale differences present in input data. Through automatic weight adjustments, the network can handle information more intricately across different levels and resolutions, thereby enhancing performance on complex tasks.

4. Experimental setup and outcomes

In this section, we experimentally validate the efficacy and performance of our proposed model, NFMPAtt-Unet. Initially, we present the datasets and equipment employed in our experiments. Subsequently, we present the experimental results and conduct a detailed analysis.

4.1. Basic introduction

In the experiments, we implemented our method using PyTorch (version 2.0.0) and conducted the experiments on an NVIDIA GTX 3090 GPU. Different batch sizes (1, 8, 16) were employed for various datasets. We initialized the learning rate at 0.001 with a weight decay rate of 0.003. Furthermore, we employed active data augmentation techniques, incorporating random angle rotations within the range of −45 degrees to 45 degrees, along with random horizontal and vertical flips. For standardizing image processing, we uniformly resize images to a size of 512×512 pixels for experimental purposes.

We utilized a total of six different medical image segmentation datasets, and their basic information is presented in Table 1.

WSSS4LUAD (Han et al., 2022): This dataset was derived by scanning H&E stained slides from Guangdong Provincial People's Hospital (GDPH) and collecting Whole Slide Images (WSI) from the Cancer Genome Atlas (TCGA) at a resolution of 256×256. Pixel-level predictions were conducted for three prevalent and significant tissue types: tumor epithelial tissue, tumor-associated stromal tissue, and normal tissue.

Cell (Ma et al., 2023): This dataset encompasses the diversity of microscopic images across four dimensions: cell source, staining method, microscope type, and cell morphology. There is significant variation in the source of cells in microscopic images, differences in staining methods, the use of different types of microscopes, and noticeable variations in cell morphology across different cell types.

FootUlcer (Wang et al., 2020): This dataset comprises 1109 images of foot ulcers captured from 889 patients. The original images were captured using a Canon SX 620 HS digital camera and an iPad Pro, often under uncontrolled lighting conditions, leading to varied backgrounds. The dataset underwent uniform resizing, standardizing the pixel dimensions to 512×512.

Eye: This dataset comprises various fundus images with diverse labels. For this experiment, we focused solely on using the pupil as the target label. The dataset comprises a total of 54 training images and 27 testing images.

ISIC (Rotemberg et al., 2021): This dataset is a skin lesion image dataset used for segmenting melanomas from dermoscopic images. It comprises 2000 training images and 150 testing images, minimizing redundancy in the description. The images in the dataset have varying sizes, which we standardized to 512×512.

Lung: This dataset consists of X-ray chest images with segmented labels corresponding to lung regions. It includes 632 training images and 72 testing images.

4.2. Evaluation indicators

To evaluate the performance of various methods, we utilized a set of five evaluation metrics: accuracy (Acc), precision (Prec), F1 score (F1), intersection over union (IoU), Area Under the Curve-Receiver Operating Characteristic (AUC-ROC) , and Dice coefficient (Dice). Their definitions are as follows:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{14}$$

$$\text{Prec} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{15}$$

$$\text{F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{16}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{17}$$

$$\text{AUC-ROC} = \int_0^1 \text{TPR}\,(\text{FPR})\, d\,\text{FPR} \tag{18}$$

$$\text{Dice} = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{19}$$

In these formulas, TP (True Positive) denotes the number of samples predicted as positive and actually belonging to the positive class, while TN (True Negative) represents the number of samples predicted as negative and actually belonging to the negative class. FP (False Positive) represents the number of samples predicted as positive but actually belonging to the negative class, and FN represents the number of samples predicted as negative but actually belonging to the positive class. TPR (True Positive Rate) is calculated as $TP/(TP+FN)$, and FPR (False Positive Rate) is calculated as $FP/(FP+TN)$. X and Y are two sets, $|X \cap Y|$ is the number of elements in their intersection, and $|X|$ and $|Y|$ are the number of elements in sets X and Y, respectively.

Accuracy evaluates the proportion of correctly predicted samples out of the total samples, focusing on overall correctness. Precision measures the accuracy of positive predictions among samples predicted as positive, indicating the precision of the predictions. F1 score is the harmonic mean of precision and recall, offering a balanced consideration of both accuracy and completeness in predictions. Intersection over Union is the ratio of the intersection to the union of the predicted and true regions, representing the degree of overlap. A higher value indicates better segmentation results. The ROC curve plots the false positive rate (FPR) against the true positive rate (TPR) at various thresholds, reflecting sensitivity and specificity at different thresholds. AUC-ROC, the area under the ROC curve, ranges from 0.5 to 1, with a higher value indicating better classification performance. The Dice coefficient is a statistical tool used to measure the similarity between two samples. The value of the Dice coefficient ranges from 0 to 1, where 1 indicates complete similarity and 0 indicates no similarity.

4.3. Experimental results

In this section, we assessed the performance of the proposed NFMPAtt-Unet model and conducted a comparative analysis with eight other UNet-related models. These models include: U-Net(2015) (Ronneberger et al., 2015), U-Net++(2018) (Zhou et al., 2018), SA-UNet (2021) (Guo et al., 2021), AB-UNet(2021) (Saidu & Csató, 2021), DC-UNet(2021) (Lou, Guan, & Loew, 2021), DenseRes-UNet(2022) (Kiran, Raza, Ijaz, & Khan, 2022), PDF-UNet(2023) (Iqbal & Sharif, 2023) and META-UNet(2023) (Wu, Zhao, & Wang, 2023). The experimental results are detailed in Table 2, Table 3 and Table 4. Furthermore, visual

Table 2
Comparison with other models on mentioned datasets (WSSS4LUAD, Cell, FootUlcer).

Dataset	Model	Acc (%)	Prec (%)	F1 (%)	IoU (%)	AUC-ROC (%)
WSSS4LUAD	U-Net	91.37	90.33	87.70	78.09	90.36
	U-Net++	92.76	90.28	86.87	76.78	90.04
	SA-UNet	92.81	87.85	87.37	77.57	91.04
	AB-UNet	93.21	**96.93**	86.92	76.87	88.89
	DC-UNet	93.16	90.15	87.73	78.15	90.85
	DenseRes-UNet	93.42	92.10	88.00	78.57	90.67
	PDF-UNet	93.44	90.88	88.20	78.90	**91.12**
	META-UNet	90.17	86.99	81.80	69.20	86.28
	NFMPAtt-Unet	**93.62**	93.27	**88.25**	**78.98**	90.66
Cell	U-Net	86.32	73.12	74.48	59.34	82.97
	U-Net++	86.98	74.77	75.52	60.67	83.55
	SA-UNet	87.76	72.87	78.55	64.68	86.94
	AB-UNet	88.39	76.91	78.35	64.41	85.64
	DC-UNet	88.09	73.72	78.98	65.26	87.11
	DenseRes-UNet	88.29	76.82	78.14	64.13	85.47
	PDF-UNet	88.14	**79.16**	76.79	62.32	83.77
	META-UNet	88.71	77.39	79.00	65.28	86.13
	NFMPAtt-Unet	**89.94**	77.78	**80.74**	**67.71**	**87.95**
FootUlcer	U-Net	90.44	76.49	77.60	66.15	88.97
	U-Net++	90.54	82.83	78.64	64.79	89.03
	SA-UNet	92.51	82.99	82.11	69.64	90.50
	AB-UNet	91.44	76.71	80.99	68.05	92.71
	DC-UNet	92.70	73.20	77.37	63.10	90.82
	DenseRes-UNet	90.69	86.89	86.79	70.97	88.36
	PDF-UNet	92.19	87.82	85.00	73.91	91.09
	META-UNet	93.71	89.80	89.63	**81.21**	92.76
	NFMPAtt-Unet	**94.36**	**92.26**	**90.09**	77.55	**93.91**

Table 3
Comparison with other models on mentioned datasets (Eye, ISIC, Lung).

Dataset	Model	Acc (%)	Prec (%)	F1 (%)	IoU (%)	AUC-ROC (%)
Eye	U-Net	90.05	56.90	62.68	45.65	84.42
	U-Net++	90.49	61.01	65.52	44.79	85.13
	SA-UNet	92.89	71.20	71.74	55.93	85.88
	AB-UNet	93.85	60.55	68.52	34.71	87.57
	DC-UNet	92.12	20.24	32.05	18.76	84.71
	DenseRes-UNet	93.02	74.87	65.70	48.92	89.05
	PDF-UNet	95.98	83.35	**83.33**	55.24	90.20
	META-UNet	95.79	70.21	78.60	64.75	**94.30**
	NFMPAtt-Unet	**96.99**	**85.17**	79.51	**65.98**	91.94
ISIC	U-Net	92.52	87.86	80.81	67.80	81.76
	U-Net++	92.97	89.93	80.98	68.04	85.92
	SA-UNet	93.08	86.32	79.17	65.53	85.29
	AB-UNet	93.15	87.24	79.21	69.93	88.53
	DC-UNet	93.25	83.32	80.63	67.54	87.34
	DenseRes-UNet	89.40	72.37	69.28	53.00	80.44
	PDF-UNet	94.09	89.89	82.15	69.70	86.88
	META-UNet	94.91	**90.94**	84.91	**73.78**	88.95
	NFMPAtt-Unet	**95.64**	90.81	**86.38**	71.70	**89.53**
Lung	U-Net	97.33	95.77	95.11	90.75	93.48
	U-Net++	97.54	95.63	94.15	92.63	95.53
	SA-UNet	97.94	94.80	95.78	91.91	96.42
	AB-UNet	97.37	94.44	95.54	93.39	93.43
	DC-UNet	97.75	95.61	96.03	92.35	96.61
	DenseRes-UNet	97.42	97.33	95.34	91.09	95.59
	PDF-UNet	98.13	97.09	96.67	93.55	96.58
	META-UNet	97.17	97.02	96.19	93.02	95.37
	NFMPAtt-Unet	**98.21**	**97.51**	**96.80**	**93.81**	**97.57**

comparisons of NFMPAtt-Unet with other models on the dataset are presented in Fig. 6.

Analyzing the experimental results from Tables 2 and 3, it is evident that our proposed NFMPAtt-Unet model exhibits superior performance compared to other models. On the WSSS4LUAD dataset, NFMPAtt-Unet attains the highest performance, achieving the best values for Acc, F1, and IoU, with scores of 93.62%, 88.25%, and 78.98%, respectively. In the Cell dataset, the NFMPAtt-Unet model exhibits the best performance across the Acc, F1, IoU, and AUC-ROC metrics, with values of 89.94%, 80.74%, 67.71% and 87.95%, respectively. On the FootUlcer dataset, the NFMPAtt-Unet model demonstrates the best performance across the Acc, Prec, F1, and AUC-ROC metrics, with values of 94.36%, 92.26%, 90.09%, and 93.91%, respectively. In the Eye dataset, the NFMPAtt-Unet model achieves the best performance in terms of Acc, Prec, and IoU, with values of 96.99%, 85.17%, and 65.98%, respectively. On the ISIC dataset, the NFMPAtt-Unet model demonstrates the best performance in terms of Acc, F1, and AUC-ROC, with values of 95.64%, 86.38%, and 89.53%, respectively. On the Lung dataset, the

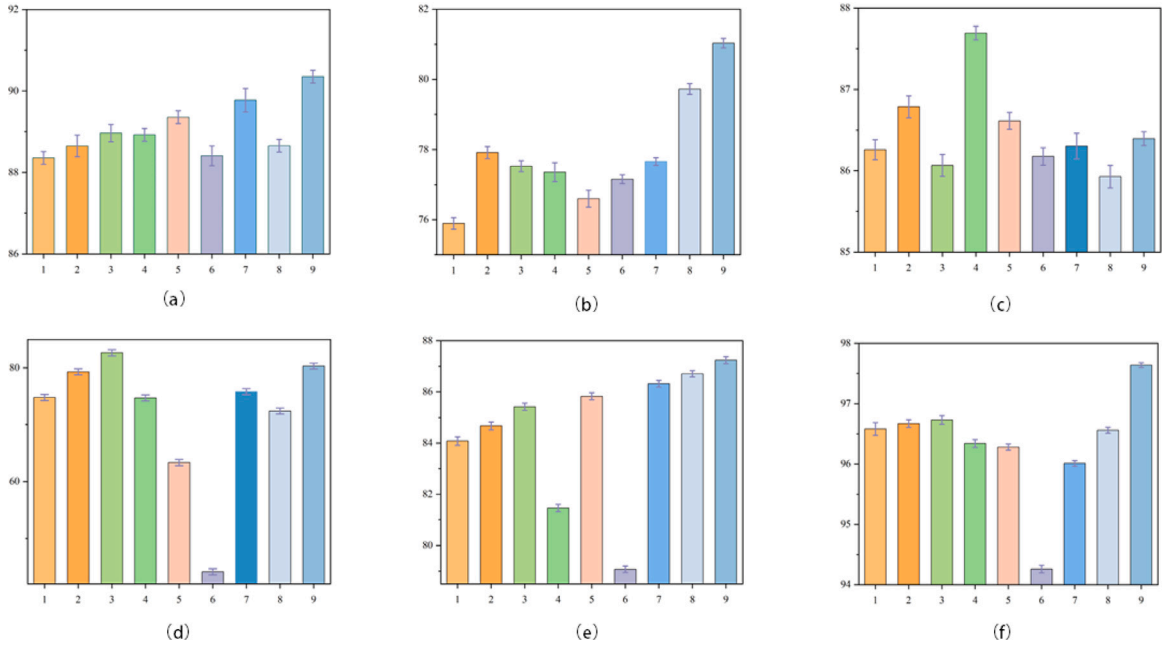


Fig. 5. The bar chart illustrates the DICE scores of different models on various datasets, with (a), (b), (c), (d), (e), and (f) corresponding to datasets WSSS4LUAD, Cell, FootUlcer, Eye, ISIC, and Lung, respectively. Models 1 through 9 represent U-Net, U-Net++, SA-UNet, AB-UNet, DC-UNet, DenseRes-UNet, PDF-UNet, META-UNet, and NFMPAtt-UNet, respectively.

Table 4
The experimental results of the model's DICE scores (%) on different datasets are presented as follows: (average value ± standard deviation).

Models \ Datasets	WSSS4LUAD	Cell	FootUlcer	Eye	ISIC	Lung
U-Net	88.35±0.157	75.89±0.167	86.25±0.123	74.79±0.054	84.08±0.065	96.58±0.105
U-Net++	88.65±0.266	77.91±0.174	86.78±0.135	79.30±0.042	84.67±0.051	96.67±0.064
SA-UNet	88.96±0.215	77.52±0.154	86.06±0.134	**82.65±0.054**	85.42±0.038	96.73±0.072
AB-UNet	88.92±0.157	77.35±0.268	**87.69±0.083**	74.70±0.035	81.46±0.042	96.34±0.065
DC-UNet	89.35±0.157	76.59±0.241	86.61±0.104	63.32±0.033	85.83±0.035	96.28±0.052
DenseRes-UNet	88.40±0.245	77.15±0.124	86.17±0.107	44.10±0.025	79.07±0.025	94.26±0.062
PDF-UNet	89.77±0.287	77.67±0.111	86.30±0.157	75.81±0.021	86.32±0.031	96.01±0.047
META-UNet	88.65±0.154	79.72±0.152	85.92±0.139	72.40±0.024	86.71±0.023	96.56±0.049
NFMPAtt-UNet	**90.35±0.187**	**81.03±0.136**	86.39±0.085	80.32±0.022	**87.24±0.034**	**97.64±0.038**

NFMPAtt-Unet model excels in all five metrics — Acc, Prec, F1, IoU, and AUC-ROC, with values of 98.21%, 97.51%, 96.80%, 93.81%, and 97.57%, respectively. From the results in Table 4 and Fig. 5, we can observe that the NFMPAtt-UNet model outperforms other models on most datasets and exhibits greater stability.

Consequently, considering the experimental results, it can be concluded that our proposed NFMPAtt-Unet model surpasses other models, showcasing superior performance across different datasets and evaluation metrics.

As depicted in Fig. 6, the first row showcases the segmentation results for breast cancer glands. In this task, both NFMPAtt-Unet and other models successfully capture the main contour areas. However, compared to other models, NFMPAtt-Unet exhibits more accurate segmentation, particularly in smaller regions. For cell segmentation (second row), NFMPAtt-Unet demonstrates superior accuracy in capturing cell contours and local areas compared to other models. In wound segmentation (third row), NFMPAtt-Unet showcases a more precise ability to segment the entire wound contour and small wounds. When dealing with segmentation tasks involving small and detailed structures like pupils and melanomas (fourth and fifth rows), NFMPAtt-Unet also demonstrates more precise segmentation. The last row displays lung segmentation results, highlighting NFMPAtt-Unet's superior accuracy in capturing lung contours. These observations emphasize the outstanding performance of the NFMPAtt-Unet model across various segmentation tasks.

4.4. Ablation study

As depicted in Fig. 1, as described in the third section, our proposed NFMPAtt-Unet incorporates three additional core components: MD-WFP, HWA, and NFCMFE, in comparison to the Unet network. To assess the effectiveness of each component, we performed an ablation study by selectively removing certain components from NFMPAtt-Unet and evaluated the model on the FootUlcer and Eye datasets (see Table 5).

The experimental results of the ablation study are presented in Table 4. Upon comparing the data in Table 4, it is apparent that the incorporation of the three proposed core components substantially improves the experimental performance in comparison to the standard U-Net. Each core component has been validated for its effectiveness on the two datasets used, providing support for the superior performance of the model.

4.5. Computational complexity

The model complexity is shown in Table 6, with the input image format being (1, 3, 128, 128). From the experimental results, it is evident that although our model's parameter count and computational complexity are not the lowest, our segmentation performance is superior. Additionally, the computational complexity of SA-UNet, AB-UNet, and DenseRes-UNet is significantly higher than that of NFMPAtt-UNet. Therefore, NFMPAtt-UNet balances the model's computational cost while improving segmentation accuracy.

Table 5
Ablation results from the FootUlcer and Eye datasets.

MDWFP	HWA	NFCMFE	FootUlcer					Eye				
			Acc	Prec	F1	IoU	AUC-ROC	Acc	Prec	F1	IoU	AUC-ROC
×	×	×	90.44	76.49	77.60	66.15	88.97	90.05	56.90	62.68	45.65	84.42
×	✓	✓	92.65	88.77	87.84	75.08	90.92	95.83	83.69	77.80	62.89	90.36
✓	×	✓	91.26	86.75	83.48	69.72	90.40	95.96	82.40	77.06	63.26	90.59
✓	✓	×	92.79	88.75	87.49	74.23	91.49	95.44	83.91	78.07	63.55	91.05
✓	✓	✓	**94.36**	**92.26**	**90.09**	**77.55**	**93.91**	**96.99**	**85.17**	**79.51**	**65.98**	**91.94**

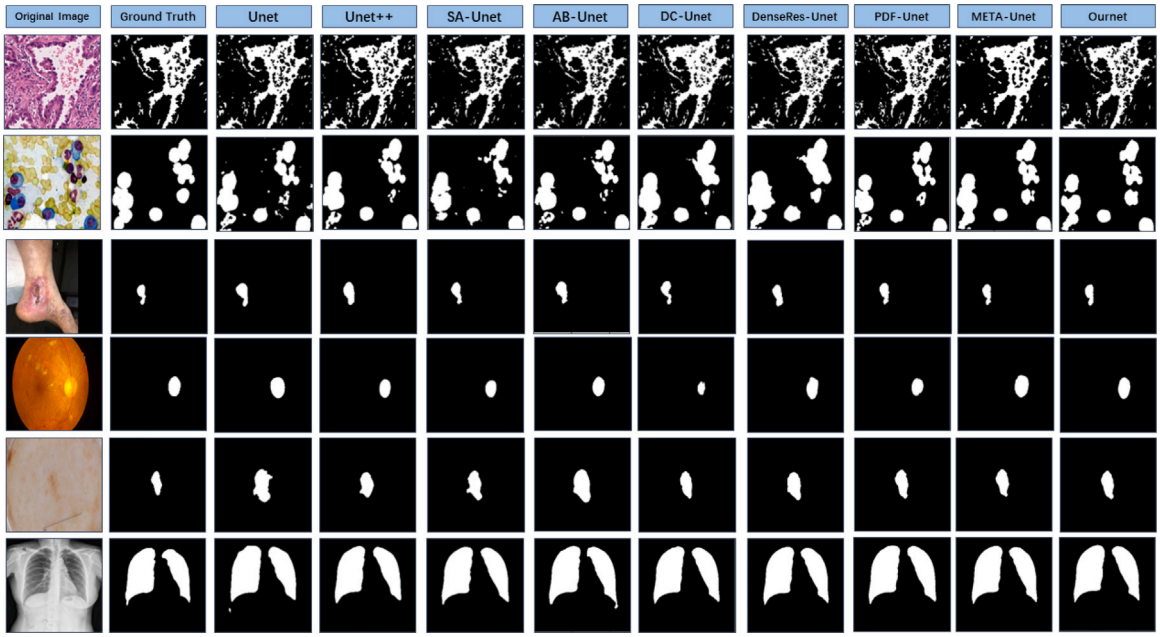


Fig. 6. Visual comparison of different algorithm models on various datasets. From the first row to the last row, the datasets are as follows: WSSS4LUAD, Cell, FootUlcer, Eye, ISIC, Lung.

Table 6
Different models' parameter count, FLOPs, and FPS comparison, measured at a resolution of 128 × 128.

Models	Params (M)	FLOPs (G)	FPS
U-Net	**7.850**	3.525	**251.369**
U-Net++	9.160	8.725	173.369
SA-UNet	60.339	19.221	58.663
AB-UNet	31.038	13.686	100.809
DC-UNet	10.811	5.958	34.950
DenseRes-UNet	11.796	90.234	26.766
PDF-UNet	10.523	2.796	120.281
META-UNet	11.628	**1.268**	78.322
NFMPAtt – UNet	9.628	6.760	125.850

4.6. Limitations and future work

Although our proposed NFMPAtt-Unet model has shown significant performance improvement in various medical image segmentation tasks compared to many advanced methods, there are still limitations in segmentation tasks. As shown in Table 6, the main issue lies in the high number of parameters. While our model has fewer parameters compared to some models and achieves other performance improvements, it still increases the hardware burden to some extent. This is because the neighborhood rough set proposed in our network model is based on pixel extraction, and higher-resolution images increase computational complexity. Additionally, the introduced hybrid weighted attention mechanism adds two additional weight matrices during weight generation. While these operations enhance the segmentation performance of the model, they also increase the number of parameters in the model. In future work, we will focus on reducing model parameters to make the proposed deep learning model more suitable for deployment in clinical applications.

5. Conclusion

In this study, we propose a novel U-Net network model designed to address the inherent challenges posed by the fuzzy boundaries in medical image segmentation. This enhanced U-Net architecture integrates three key components: the Multi-scale Dynamic Weight Feature Pyramid Module (MDWFP), the Hybrid Weighted Attention Mechanism (HWA), and the Fuzzy C-means Feature Extraction Module based on Neighborhood Rough Set (NFCMFE). The MDWFP module improves feature fusion by dynamically assigning weights to features across different scales, thereby enhancing the network's ability to capture multi-scale information effectively. The HWA mechanism enhances the network's feature selection process by integrating both channel-wise and spatial attention mechanisms, enabling more efficient utilization of crucial features for segmentation tasks. Finally, the NFCMFE module leverages the concept of neighborhood rough sets to extract features using fuzzy C-means clustering, enabling the model to handle complex structures and uncertainties inherent in medical images more effectively.

Furthermore, we conducted comparative experiments involving our proposed model and other state-of-the-art models on multiple datasets. The experimental results substantiate the superior advantages and performance of our proposed model.

CRediT authorship contribution statement

Xinpeng Zhao: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation. **Weihua Xu:** Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

Data availability

No data was used for the research described in the article.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (NO. 62376229) and Natural Science Foundation of Chongqing (NO. CSTB2023NSCQ-LZX0027).

References

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2–3), 191–203.

Bhargavi, K., & Jyothi, S. (2014). A survey on threshold based segmentation technique in image processing. *International Journal of Innovative Research and Development*, *3*(12), 234–239.

Chen, B., Liu, Y., Zhang, Z., Lu, G., & Kong, A. W. K. (2023). Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *8*(1), 55–68.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., et al. (2021). Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.

Gong, M., Liang, Y., Shi, J., Ma, W., & Ma, J. (2012). Fuzzy c-means clustering with local information and kernel metric for image segmentation. *IEEE Transactions on Image Processing*, *22*(2), 573–584.

Guo, L., Shi, P., Chen, L., Chen, C., & Ding, W. (2023). Pixel and region level information fusion in membership regularized fuzzy clustering for image segmentation. *Information Fusion*, *92*, 479–497.

Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., & Fan, C. (2021). Sa-unet: Spatial attention u-net for retinal vessel segmentation. In *2020 IEEE international conference on pattern recognition* (pp. 1236–1242). IEEE.

Guo, D., Xu, W., Ding, W., Yao, Y., Wang, X., Pedrycz, W., et al. (2024). Concept-cognitive learning survey: Mining and fusing knowledge from data. *Information Fusion*, *109*, 102426.

Guo, D., Xu, W., Qian, Y., & Ding, W. (2023). M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion. *Information Fusion*, *100*, 101962.

Han, C., Pan, X., Yan, L., Lin, H., Li, B., Yao, S., et al. (2022). Wsss4luad: Grand challenge on weakly-supervised tissue semantic segmentation for lung adenocarcinoma. arXiv preprint arXiv:2204.06455.

Hirano, S., & Tsumoto, S. (2002). Segmentation of medical images based on approximations in rough set theory. In *2002 international conference on rough sets and current trends in computing* (pp. 554–563). Springer.

Hu, H., Li, Q., Zhao, Y., & Zhang, Y. (2020). Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors. *IEEE Transactions on Industrial Informatics*, *17*(4), 2880–2889.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., et al. (2020). Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing* (pp. 1055–1059). IEEE.

Huang, Y.-C., Tung, Y.-S., Chen, J.-C., Wang, S.-W., & Wu, J.-L. (2005). An adaptive edge detection based colorization algorithm and its applications. In *2005 ACM international conference on multimedia* (pp. 351–354). ACM.

Iqbal, A., & Sharif, M. (2023). UNet: A semi-supervised method for segmentation of breast tumor images using a U-shaped pyramid-dilated network. *Expert Systems with Applications*, *221*, Article 119718.

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, *18*(2), 203–211.

Jaderberg, M., Simonyan, K., Zisserman, A., & kavukcuoglu, k. (2015). Spatial transformer networks. In *2015 advances in neural information processing systems*. Curran Associates, Inc.

Jha, D., Smedsrud, P. H., Riegler, M. A., Johansen, D., De Lange, T., Halvorsen, P., et al. (2019). Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE international symposium on multimedia* (pp. 225–2255). IEEE.

Jothi, G., et al. (2016). Hybrid tolerance rough set–firefly based supervised feature selection for MRI brain tumor image classification. *Applied Soft Computing*, *46*, 639–651.

Kiran, I., Raza, B., Ijaz, A., & Khan, M. A. (2022). DenseRes-Unet: Segmentation of overlapped/clustered nuclei from multi organ histopathology images. *Computers in Biology and Medicine*, *143*, 105267.

Lewis, J. J., O'Callaghan, R. J., Nikolov, S. G., Bull, D. R., & Canagarajah, N. (2007). Pixel-and region-based image fusion with complex wavelets. *Information Fusion*, *8*(2), 119–130.

Li, J., Jin, K., Zhou, D., Kubota, N., & Ju, Z. (2020). Attention mechanism-based CNN for facial expression recognition. *Neurocomputing*, *411*, 340–350.

Lou, A., Guan, S., & Loew, M. (2021). DC-UNet: rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation. In *Medical imaging 2021: Image processing* (pp. 758–768). SPIE.

Ma, J., Xie, R., Ayyadhury, S., Ge, C., Gupta, A., Gupta, R., et al. (2023). The multi-modality cell segmentation challenge: Towards universal solutions. arXiv: 2308.05864.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., et al. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

Pan, Y., Xu, W., & Ran, Q. (2023). An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets. *International Journal of Machine Learning and Cybernetics*, *14*(4), 1217–1233.

Phophalia, A., Rajwade, A., & Mitra, S. K. (2014). Rough set based image denoising for brain MR images. *Signal Processing*, *103*, 24–35.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *2015 medical image computing and computer-assisted intervention* (pp. 234–241). Springer.

Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., et al. (2021). A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific Data*, *8*(1), 34.

Saidu, I. C., & Csató, L. (2021). Active learning with bayesian UNet for efficient semantic image segmentation. *Journal of Imaging*, *7*(2), 37.

Sha, Y., Zhang, Y., Ji, X., & Hu, L. (2021). Transformer-unet: Raw image processing with unet. arXiv preprint arXiv:2109.08417.

Shi, H., Li, H., Meng, F., Wu, Q., Xu, L., & Ngan, K. N. (2018). Hierarchical parsing net: Semantic scene parsing from global scene to objects. *IEEE Transactions on Multimedia*, *20*(10), 2670–2682.

Tang, Y., Ren, F., & Pedrycz, W. (2020). Fuzzy C-means clustering through SSIM and patch for image segmentation. *Applied Soft Computing*, *87*, Article 105928.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *2017 advances in neural information processing systems*. Curran Associates, Inc.

Wang, C., Anisuzzaman, D., Williamson, V., Dhar, M. K., Rostami, B., Niezgoda, J., et al. (2020). Fully automatic wound segmentation with deep convolutional neural networks. *Scientific Reports*, *10*(1), 21897.

Wu, Y.-H., Liu, Y., Zhan, X., & Cheng, M.-M. (2022). P2T: Pyramid pooling transformer for scene understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(11), 12760–12771.

Wu, H., Zhao, Z., & Wang, Z. (2023). META-Unet: Multi-scale efficient transformer attention Unet for fast and high-accuracy polyp segmentation. *IEEE Transactions on Automation Science and Engineering*, 1–12.. http://dx.doi.org/10.1109/TASE.2023.3292373.

Xu, W., Guo, D., Mi, J., Qian, Y., Zheng, K., & Ding, W. (2023). Two-way concept-cognitive learning via concept movement viewpoint. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(10), 6798–6812.

Yao, L., Torabi, A., Cho, K., Ballas, N., Pal, C., Larochelle, H., et al. (2015). Describing videos by exploiting temporal structure. In *2015 IEEE international conference on computer vision* (pp. 4507–4515). IEEE.

Yu, H., Jiang, L., Fan, J., Xie, S., & Lan, R. (2024). A feature-weighted suppressed possibilistic fuzzy c-means clustering algorithm and its application on color image segmentation. *Expert Systems with Applications*, *241*, Article 122270.

Zhang, P., Liu, W., Lei, Y., Wang, H., & Lu, H. (2020). RAPNet: Residual atrous pyramid network for importance-aware street scene parsing. *IEEE Transactions on Image Processing*, *29*, 5010–5021.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *2017 IEEE conference on computer vision and pattern recognition* (pp. 2881–2890). IEEE.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *2018 deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3–11). Springer.