mathematics

MDPI

Article

A Novel Medical Decision-Making System Based on Multi-Scale Feature Enhancement for Small Samples

Keke He [1], Yue Qin [2,*], Fangfang Gou [2,*] and Jia Wu [2,3,*]

[1] School of Computer Science and Engineering, Changsha University, Changsha 410003, China; z20131046@ccsu.edu.cn
[2] School of Computer Science and Engineering, Central South University, Changsha 410083, China
[3] Research Center for Artificial Intelligence, Monash University, Clayton, Melbourne, VIC 3800, Australia
* Correspondence: 214711088@csu.edu.cn (Y.Q.); gff8221@csu.edu.cn (F.G.); jiawu0510@csu.edu.cn (J.W.)

Abstract: The medical decision-making system is an advanced system for patients that can assist doctors in their medical work. Osteosarcoma is a primary malignant tumor of the bone, due to its specificity, such as its blurred borders, diverse tumor morphology, and inconsistent scales. Diagnosis is quite difficult, especially for developing countries, where medical resources are inadequate per capita and there is a lack of professionals, and the time spent in the diagnosis process may lead to a gradual deterioration of the disease. To address these, we discuss an osteosarcoma-assisted diagnosis system (OSADS) based on small samples with multi-scale feature enhancement that can assist doctors in performing preliminary automatic segmentation of osteosarcoma and reduce the workload. We proposed a multi-scale feature enhancement network (MFENet) based on few-shot learning in OSADS. Global and local feature information is extracted to effectively segment the boundaries of osteosarcoma by feeding the images into MFENet. Simultaneously, a prior mask is introduced into the network to help it maintain a certain accuracy range when segmenting different shapes and sizes, saving computational costs. In the experiments, we used 5000 osteosarcoma MRI images provided by Monash University for testing. The experiments show that our proposed method achieves 93.1% accuracy and has the highest comprehensive evaluation index compared with other methods.

Keywords: decision-making system; imaging segmentation; multi-scale feature enhancement; few-shot

MSC: 68T07

check for updates

Citation: He, K.; Qin, Y.; Gou, F.; Wu, J. A Novel Medical Decision-Making System Based on Multi-Scale Feature Enhancement for Small Samples. *Mathematics* **2023**, *11*, 2116. https://doi.org/10.3390/math11092116

Academic Editor: Vasile Preda

Received: 23 March 2023
Revised: 24 April 2023
Accepted: 27 April 2023
Published: 29 April 2023

1. Introduction

A medical decision system is an advanced information system for patients that can assist doctors in diagnosis. It is a scientific decision-making system that assists physicians in diagnosis by synthesizing clinical data and organically combining numerous models. It has been well used in some common cancers such as breast cancer, lung cancer, and brain tumors. As a rare bone malignancy [1], osteosarcoma is a rare malignant tumor which is currently poorly treated due to the lack of experience of physicians [2]. Especially in developing countries, large-scale clinical trials cannot be conducted due to resource, environmental and demographic constraints, and up to 75% of patients have a poor curative effect. Most patients miss the best period of tumor resection and develop blood metastases in the later period, which leads to the exacerbation of the disease and an unsatisfactory prognosis [3,4]. Therefore, for developing countries, building a medical decision system is one of the best ideas to improve the diagnostic environment and increase the survival rate of patients with osteosarcoma.

Optimal segmentation of medical images is important for facilitating automated methods in medical decision-making and disease diagnosis. Imaging is an important tool in the detection of osteosarcoma. The radiologists identify the area of the lesion by

manually tracing the contour lines of the target structure on medical images and delineating the target area from the surrounding tissues and organs. Magnetic resonance imaging (MRI) is often used in the preoperative diagnosis of osteosarcoma because of its very high tissue resolution. MRI image analysis before treatment can not only be used to predict osteosarcoma, but is also a major step in estimating clinical parameters and assisting disease diagnosis and surgical planning [5,6].

Unfortunately, due to their largely underdeveloped healthcare infrastructures, the majority of developing nations experience the same problem with osteosarcoma diagnosis [7–11]. It requires a large number of professionals, particularly in the field of medical imaging. In China, for example, the growth rate of medical imaging data is very fast, with an annual growth rate of 30%. However, the number of radiologists is growing slowly, with an annual growth rate of only about 4%. The gap between the two growth rates is huge. Additionally, there is a massive shortage of pathologists in China (an average of 1 pathologist every 70,000 people in China compared to 1 pathologist every 2000 people in the United States). The shortage of medical professionals in China and their heavy workload are the reasons for the high rate of misdiagnosis and underdiagnosis. According to a misdiagnosis data report from the Chinese Medical Association, China has a clinical misdiagnosis rate of 27.8% overall, with malignant tumors having an average misdiagnosed rate of 40% [12]. The most important thing is that the radiologist's reading speed is limited. They need to mark the position of the organ and tumor on each image while outlining the images. This process usually takes a lot of time [13,14]. Moreover, after finding the location of the tumor, doctors also need to design specific irradiation plans or surgical plans of radiation according to the tumor's size and shape. Therefore, the speed of medical imaging diagnosis is extremely limited in developing countries, and there is an urgent clinical need for a fast, accurate, and reproducible automated osteosarcoma-assisted segmentation method.

With the rise of computer-aided diagnosis systems, image segmentation is one of the key steps. Researchers can automatically mine and analyze the deep features of input images by using the good ability of deep learning to express complex nonlinear relationships. The identification of functional tissue and the diagnosis of tumors provides interpretive information. In a practical sense, the application of deep learning technology in the diagnosis of osteosarcoma can not only assist doctors to complete the initial automatic segmentation of osteosarcoma, but also can allow doctors to conduct further processing on the automatically segmented images, which can improve the diagnostic accuracy and doctors' work efficiency and reduce the pressure on doctors [15]. Additionally, it can successfully address the issues associated with the limited and unequal distribution of medical resources in developing nations, while simultaneously lowering the high cost of the growing number of high-quality medical resources and raising the standard of primary care [16].

Currently, MRI image segmentation based on deep learning methods is mainly applied to the heart, brain, liver, and other regions [17–20]. Most of them are CNN-based methods for image segmentation, which have a small field of view due to the limitations of CNN, resulting in a lack of accuracy in segmentation. Especially, it is not applicable to rare diseases such as osteosarcoma, because there are several factors affecting osteosarcoma segmentation: (1) The limited quantity of information available; the model is prone to overfitting despite the enormous amount of medical picture data that is accessible. (2) Individual variation; there are many types of osteosarcoma lesions, and the form of the lesion varies with the individual. Tumors may vary greatly in size and location from patient to patient. (3) Tumor heterogeneity: osteosarcoma originates from the bone and surrounding soft tissues. The variation in the composition and morphology of local tissues makes the borders blurred and difficult to identify. All of these reasons make MRI image segmentation of osteosarcoma a challenging task.

Therefore, to reduce the pressure on hospitals in developing countries and the time-consuming problem of traditional manual segmentation, this paper proposes a novel medical decision-making system based on small samples with multi-scale feature enhancement (OSADS) for osteosarcoma. The core of the system is a multi-scale feature enhancement

segmentation network (MFENet) based on the few-shot method. The network can obtain the global and local feature information through MRI and accurately segment the boundary with a small number of samples. Additionally, the computational cost can be saved by introducing a prior mask in the model. Furthermore, the network can adapt to lesion regions of different sizes and locations by generating multi-scale features. It can effectively assist doctors in diagnosis in clinical practice. It greatly improves the efficiency of doctors and reduces their workload.

The following is a list of the contributions to this paper:

(1) This paper discusses an osteosarcoma-assisted diagnosis system (OSADS) in conjunction with artificial intelligence techniques. Physicians can use an OSADS for automatic image segmentation and the results can be used as a second opinion to assist doctors in their diagnosis. This not only addresses the high rate of misdiagnosis caused by doctors' enormous workloads in underdeveloped nations, but also allows patients to obtain timely and effective treatment.
(2) We design a multi-scale feature enhancement network (MFENet). In this network, we use several transformer blocks to extract global information and convolution to extract local information. The feature information of different scales is generated at the same time. This enables the model to maintain a range of accuracy when segmenting osteosarcomas of different shapes and sizes.
(3) We add a prior mask to the image segmentation model. It can replace the reconstruction process from features to segment MRI images. This enables a rapid response to provide auxiliary results to doctors.
(4) The experiments are conducted using real data of osteosarcoma images provided by Monash University. The results show that our proposed method of osteosarcoma segmentation is better than other methods. The system is extremely important for osteosarcoma diagnosis, the course of treatment, and outlook.

The rest of the paper is structured as follows. In Section 2, this paper presents the work related to the preliminary research and the latest research results. In Section 3, this paper details the osteosarcoma-assisted diagnosis system (OSADS), a framework that includes four main parts: classification, preprocessing, segmentation, and diagnosis. In the fourth chapter, we analyze and discuss the experimental results. In the fifth chapter, we conclude the paper and describe future research directions.

2. Related Work

Early non-invasive inspection and quantitative analysis are needed to separate and identify lesions from the healthy surrounding tissue with complex features. Additionally, it is important for neoadjuvant chemotherapy efficacy and evaluation. In the early years, there was some research on osteosarcoma segmentation techniques [21–24]. However, most of them were based on traditional segmentation methods, usually on low-level features for osteosarcoma. Due to the specificity of osteosarcoma, it originates in the bone as well as the surrounding soft tissue. The grayscale and textural features within the tumor are inconsistent. These methods have limited ability to express features. To achieve higher accuracy in image segmentation, researchers have started to experiment with new image segmentation methods. The most common one is the method based on CNN architecture [2,25,26].

These methods have greatly improved performance compared to traditional segmentation methods. However, CNN-based methods rely on convolutional operations, which only collect information from the neighborhood with limited sensory fields and lack the ability to explicitly capture long-range dependencies. They cannot effectively identify the boundary of osteosarcoma, carrying this out with low robustness.

Transformers have recently attracted a lot of interest from the computer vision community and were first used to express sequence-to-sequence prediction in NLP tasks [27]. Because the multiple self-attention mechanisms can effectively establish global connections between sequence tags, their long-range correlation modeling capabilities work effectively

in dense prediction tasks such as image segmentation. A self-aware model with a moving window is called a Swin Transformer [28]. Concatenated window self-attention operation and sliding window self-attention operation are used to obtain global attention, and feature fusion is used to extract multi-scale features from the input image. In SETR [29], the input image is treated as a sequence of image patches represented by learnt patch embeddings, and this sequence is transformed for the purpose of learning discriminative feature representations. In medical image segmentation, the DS-TransUNet [30] is a two-scale encoder sub-network that introduces a hierarchical Swim Transformer into the encoder and decoder of a U-shaped structure. The network can extract features of different scales. TransFuse [31] is a new parallel branching architecture that combines the Swin Transformer and CNN to efficiently capture global dependencies and low-level spatial details. PMTrans [32] utilizes a pyramid network structure to combine multi-scale attention and CNN feature extraction to accommodate multi-scale and multi-resolution patches instead of fixed-size patches, enabling robust adaptation to objects of various sizes and shapes.

Although the above-mentioned literature shows that the Swin Transformer shas great performance in image segmentation, it still faces two bottlenecks in clinical practice: (1) in clinical practice, high computational costs burden hardware requirements and virtually increase patient waiting times. (2) Medical image datasets are usually much smaller than everyday images. Traditional Swin Transformer pre-training relies on large datasets; the model is prone to overfitting and has poor generalization ability.

The burden of expensive model training data annotation plagues the majority of supervised deep learning algorithms for medical picture segmentation. To ease this pressure, a few-shot segmentation method was recently presented. MSHNet [33] can establish a strong semantic relationship between the support and query images together with cosine similarity. It is used to reduce the overfitting problem caused by a small amount of data. CRNet [34] is a cross-reference network for few-shot segmentation. The model can make predictions on both support images and query images. It can better discover co-occurring objects in two images through a cross-reference mechanism, helping to complete the task of few-shot segmentation. MFNet [35] is a novel multiplexed (class) encoding and decoding architecture. It successfully combines multi-scale query and multi-class support data into a single query support embedment. As was already mentioned, current approaches only take into account local data when implementing query support capabilities, and they ignore global data. It is well known that global relationship modeling is crucial for scene understanding in computer vision [36–38].

The medical decision-making system is a comprehensive application of a large amount of data, through the combination of models for human–computer interaction, to assist doctors in making decisions. Currently, there are many studies in this area in China and abroad. For example, Xue et al. [39] proposed a data-driven decision weight and reliability fusion method to provide a solution for thyroid cancer. Its idea is to model and describe the evaluation of each criterion for thyroid cancer diagnosis by using three types of language scaling functions. Thus, the weight and reliability of each radiologist's assessment of each criterion are determined. Vaiyapuri et al. [40] proposed an intelligent decision system (IDLDMS-PTC) for guiding pancreatic tumor segmentation, which combines multi-level-threshold (EPO-MLT) technology and optimizes the EPO algorithm for threshold selection to guide tumor segmentation. The method has shown superior performance through experiments. Zhou et al. [41] proposed a new intelligent decision method for GC screening (ID-GCS). It is a data-driven decision system based on multimodal semantic fusion. ID-GCS uses a hybrid attention mechanism to extract text semantics from multimodal gastroscopy reports, thus improving the interpretability of gastroscopy results. Intelligent decision-making systems have demonstrated their superior performance in the medical field.

In summary, this paper proposes a medical decision-making system for osteosarcoma which is a segmentation scheme based on the few-shot method, named the multi-scale feature enhancement network (MFENet). The network uses the Swin Transformer to guide global information on the merged query support features to make up for the deficiency of

the few-shot method, improving segmentation efficiency. This method not only addresses the insufficiency of osteosarcoma data, but can effectively segment the boundaries of osteosarcoma by capturing global contextual information and local details.

3. System Model Design

MRI is the current routine imaging modality used to detect osteosarcoma. Accurate segmentation of osteosarcoma areas from MRI images can provide accurate tumor quantification for clinical preoperative planning for radiotherapy and assessment of postoperative treatment efficacy. This is important for neoadjuvant chemotherapy efficacy and evaluation. Because there are not enough medical resources per person in developing nations, radiologists must manually outline tumor areas from vast amounts of picture data every day. It is labor-intensive, complex, and time-consuming work. It is difficult for patients to receive timely and effective treatment. Therefore, we hope to build an automatic segmentation system for osteosarcoma MRI images to assist doctors in segmenting osteosarcoma regions, reduce the workload of doctors and improve the efficiency of diagnosis and treatment. The most important thing is to allow patients with osteosarcoma to receive timely treatment. The system architecture is shown in Figure 1. The system is mainly divided into four stages as follows.

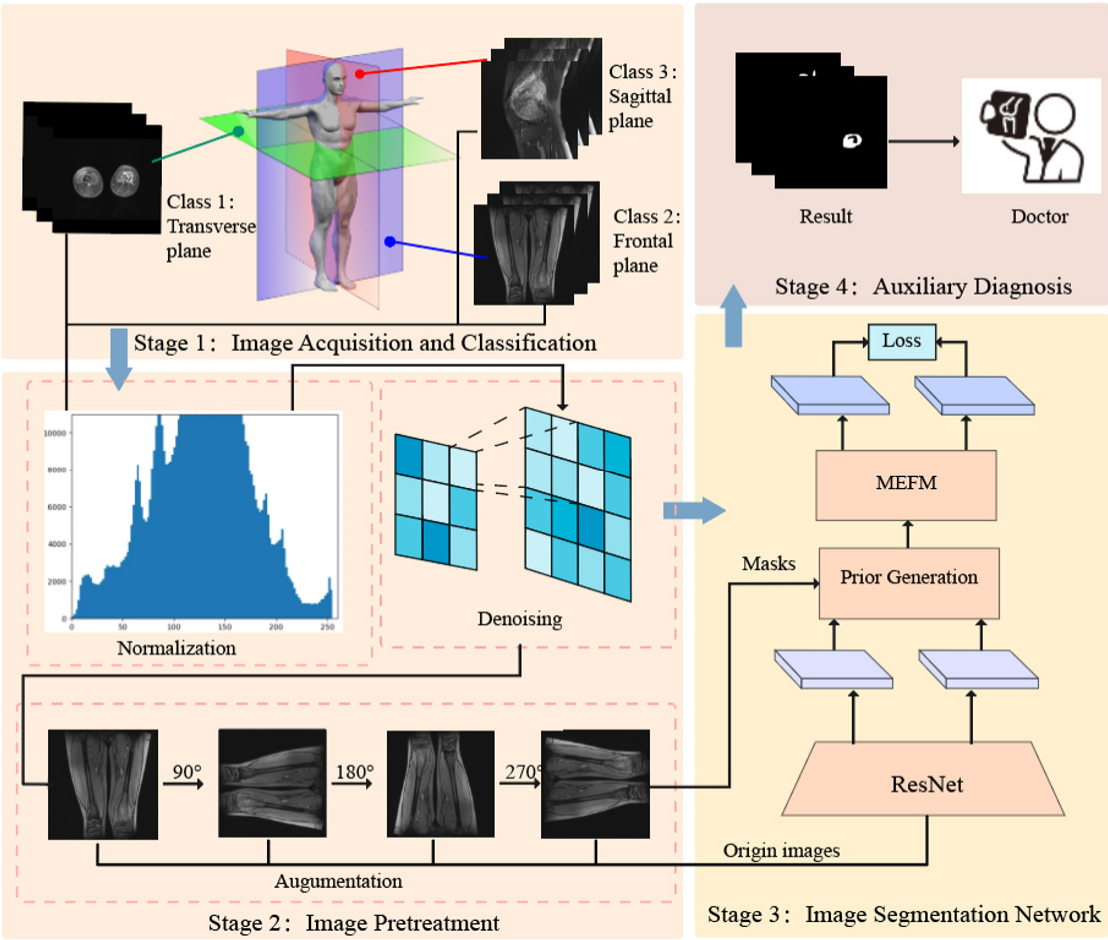

Figure 1. The framework of OSADS (including four stages: classification, pretreatment, segmentation and diagnosis).

First, in this experiment we collect real data on osteosarcoma and divide them into three categories: sagittal plane, transverse plane, and frontal plane. Next, the acquired data images are preprocessed, including image denoising, data normalization, and data augmentation operations to improve image quality. Then, we use MEFNet to segment the

image and adjust the parameters to make the model achieve optimal results. Finally, the segmentation results of the system can be used as a "second opinion" to assist doctors in diagnosis.

3.1. Problem Definition

Similar to the work of [42–44], we aim to build the support and query sets in the osteosarcoma segmentation task to accommodate few-shot segmentation. As mentioned before, due to the specificity and individual differences of osteosarcoma, the shape characteristics of osteosarcoma such as size and location are not exactly the same for each patient. We treat osteosarcoma images from different people as separate classes in this paper, and the tumors that correspond to each class are considered its members. Then, as the support set, we sample the members of various classes, while the remaining members of the relevant class serve as the query set.

We follow the formula for few-shot segmentation in [44], Given a training set $D_{train} = \{(S_i, Q_i)\}_{i=1}^{N_{train}}$, each osteosarcoma image (X_i) in it is considered a unique class. This training set is cropped into patches to build the support set $S_i = \{\{x_i^{S_k}, y_i^{S_k}\}_{k=1}^{K}\}_{i=1}^{N_{train}}$ and the query set $Q_i = \{x_i, y_i\}_{i=1}^{N_{train}}$. That is, each query set (Q_i) is associated with a small (K-shot) support set (S_i). x_i and y_i are the i-th query image and the corresponding ground truth mask, respectively. Each training episode includes a support set and a query set. To make the experimental data distribution more reasonable, we let $D_{test} = \{(S_i, Q_i)\}_{i=1}^{N_{test}}$ represent the test episode. It is worth noting that $Q_i = \{x_i\}_{i=1}^{N_{test}}$, $S_i = \{\{x_i^{S_k}, y_i^{S_k}\}_{k=1}^{K}\}_{i=1}^{N_{test}}$. The model only needs to segment x_i based on the information provided by the support set $\{x_i^{S_k}, y_i^{S_k}\}_{k=1}^{K}$, i.e., it only needs to segment the class with the same ground truth mask in the support set. The main symbols in this chapter are shown in Table 1.

Table 1. Symbol Description.

Symbol	Meaning
D_{train}	train set (including support set and query set)
D_{test}	test set
$S_i = \{\{x_i^{S_k}, y_i^{S_k}\}_{k=1}^{K}\}_{i=1}^{N_{train}}$	support set
$Q_i = \{x_i, y_i\}_{i=1}^{N_{train}}$	query set
x_i	i-th image
y_i	i-th image with label
M_Q	query feature map
M_S	support feature map
c_S	maximum similarity index
$O = \{O^1, O^2, \ldots, O^n\}$	the spatial resolution of average pooling
$I = \{I_1, I_2, \ldots, I_n\}$	feature pyramid

3.2. Data Preprocessing

During the training process, all data input into the neural network are required to have the same dimensionality, i.e., each image input to the neural network needs to maintain the same dimensions. To speed up the program, we normalize the input image by converting the value of each pixel (W) in the image so that it is within the range of 0–1. Additionally, the sizes of MRI images generated by different medical devices are inconsistent, so normalization can solve the above problems.

$$W_{norm} = \frac{W - \min(W)}{\max(W) - \min(W)} \tag{1}$$

It is inevitable that some noise will be doped when medical equipment generates MRI images. The most important effect of noise on an image is that it can override and reduce the visibility of certain features in the image. Due to the low contrast inside osteosarcoma, the loss of visibility has a significant impact on the image. It has a hindering effect on

the subsequent segmentation. Therefore, image denoising is an integral part of image preprocessing. We use the median filter method to denoise the image, and the formula is as follows:

$$F(x,y) = med\{f(x-m, y-n), (m,n) \in K\} \tag{2}$$

where *med* is the median filter function and K is a two-dimensional template, usually taken as 3×3 or 5×5. Finally, we extend the dataset by scaling up (scaling down) the images, rotating and flipping them to prevent overfitting during subsequent model training.

3.3. Segmentation Network

In this section, we will introduce the specific steps of the MFENet proposed in this paper to achieve the task of detailed automatic osteosarcoma segmentation. To reduce the computational time of the system, we first map the osteosarcoma MRI images from high-dimensional features to low-dimensional features. The backbone of the feature extractor is ResNet-50 [45]. ImageNet [46] is used to train the backbone model. ResNet layers are separated into four blocks based on spatial resolution, which corresponds to four different representation levels. To produce features, we select blocks 2 and 3 and drop the layers following block 3. After block 2, all feature maps have a fixed size of 1/8 of the input osteosarcoma MRI image. After block 2 and block 3, the features are concatenated and encoded into 256 dimensions using 3×3 convolution. During training, we always keep the weights of ResNet unchanged. Simply, our formulas are all based on (S, Q) in a single episode. Supposing G, $M_Q \in R^{H \times W \times C}$, $M_{S_k} \in R^{H \times W \times C}$ represent the backbone function, the query feature map and the k-th support feature map, respectively, H, W, and C represent the height, width and number of feature channels of the MRI images, respectively. The following formula can be obtained:

$$M_Q = G(x), M_{S_k} = G(x^{S_k}) \odot y^{S_k} \tag{3}$$

From [47], it is clear that the high-level features of the image lead to lower recognition performance and the mid-level features have better performance because they form part of the object shared by the invisible class, especially in the few-shot semantic segmentation task, although higher-level features can directly provide training-class semantic information, reducing the training loss. The model is not good for evaluating unseen test classes. Hence, we add a prior mask to the model and make the osteosarcoma support images generate a "prior" in it. This prior knowledge helps the model identify targets in query images of osteosarcoma. In other words, it can use high-level features to pre-estimate the likelihood of pixels in osteosarcoma images belonging to the target class. The flow of the module is shown in Figure 2, in three steps.

In the first step, the cosine similarity is calculated to obtain the similarity matrix (W). $\cos(m_q, m_{s_k})$ calculates the cosine similarity between pixels m_q and m_{s_k}, $m_q \in M_Q$, $m_{s_k} \in M_{S_k}$, as shown in Equation (4):

$$\cos(m_q, m_{s_k}) = \frac{m_q^T m_{s_k}}{\|m_q\| \|m_{s_k}\|} q, s_k \in \{1, 2, 3, \ldots, hw\} \tag{4}$$

In the second step, the similarity matrix is used to find the similarity between a pixel in M_Q and each pixel in m_q, $m_q \in M_Q$. We compute the similarity of this point with each pixel of M_{S_k} and obtain the maximum similarity index denoted as c_{S_k}, as shown in Equation (5):

$$c_{S_k} = \underset{s_k \in \{1,2,3,\ldots,hw\}}{\operatorname{argmax}} (\cos(m_q, m_{s_k})) \tag{5}$$

In the third step, to enhance the relevant features of osteosarcoma images and improve the accuracy of osteosarcoma segmentation, we further obtain the support prototype (M_s) for guiding query image segmentation as follows:

$$M_S = \frac{\sum_{k=1}^{K} GAP(M_{S_k}[y^{S_k}, :])}{K} \tag{6}$$

where GAP represents the global average pooling of the spatial dimension, and we resize the ground truth mask ($y^{S_k} \in R^{H \times W}$) to the feature resolution. Intuitively, the support prototype ($M_s \in R^C$) is a feature vector obtained by averaging the foreground features from the support set, which encodes representative information in the osteosarcoma images.

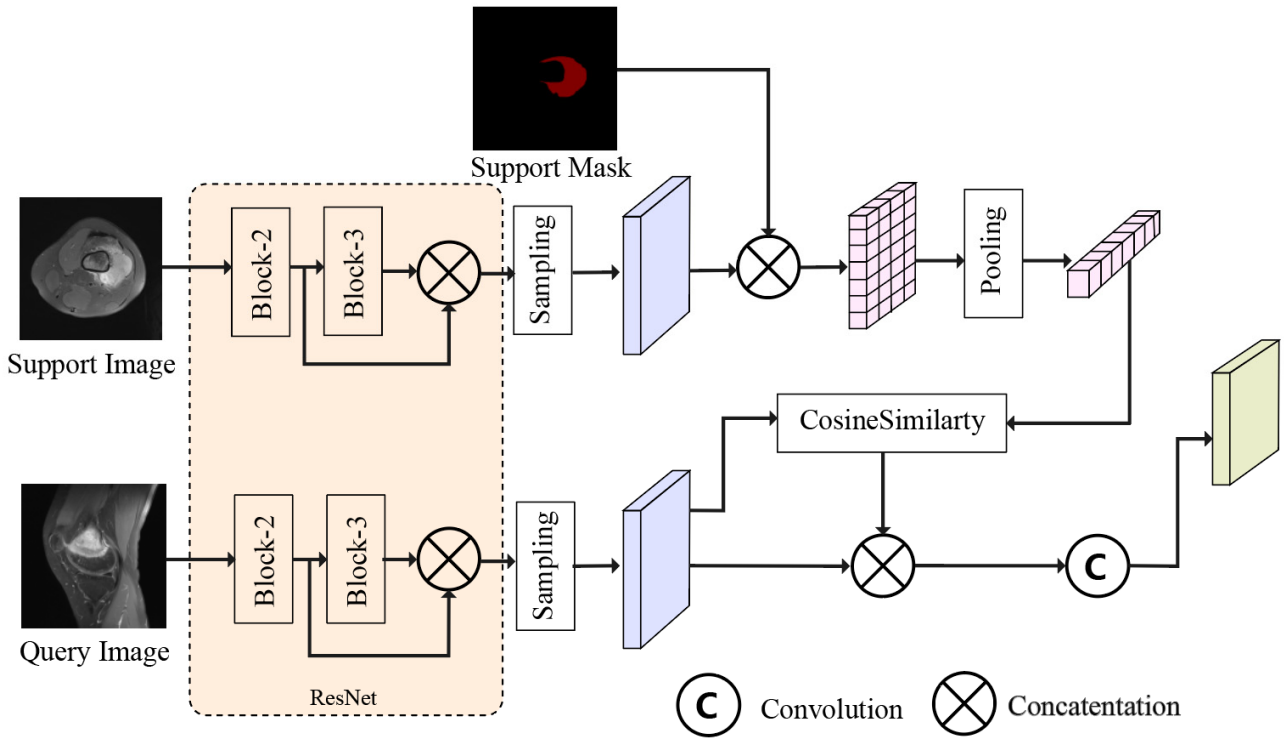

Figure 2. Prior generation module. The red markers in the image indicate the tumor area.

Since the size of the objects in the osteosarcoma support image and the query image may vary considerably, we design a multi-scale feature enhancement module (MFEM) with input feature map I so that the information on different scales of the osteosarcoma image can be utilized. It mainly consists of two parts: global enhancement block and local enhancement block. As shown in Figure 3, with the global and local enhancement blocks, not only can the entire tumor area be effectively segmented, but they are also important for improving the border segmentation of osteosarcoma. To obtain features at different scales, we use an adaptive averaging pool. Letting $O = \{O^1, O^2, \ldots, O^n\}$ denote the spatial resolution of average pooling, we assume $O^1 > O^2 > \ldots > O^n$. The feature I_i with space size P^i can be expressed as

$$I_i = GAP_{I_i}(I) \tag{7}$$

In the wa, the size of the output feature is P^i. At the same time, we obtain a feature pyramid of $\{I_1, I_2, \ldots, I_n\}$. Each feature pyramid will be handled by a global enhancement block and a local enhancement block. It is worth noting that $I_i \in R^{R_i \times R_i \times (2C+1)}$.

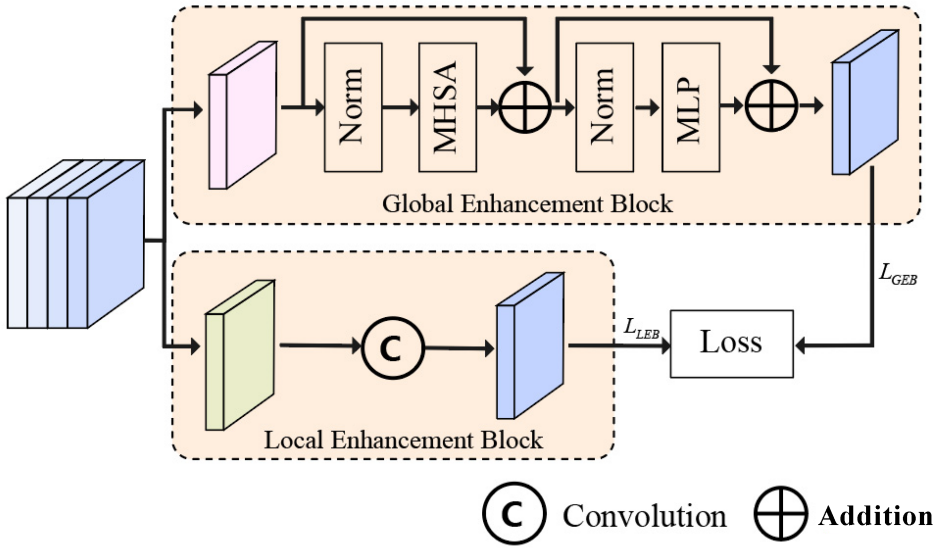

Figure 3. Muti-scale feature enhancement module.

Global Enhancement Block (GEB). In order to effectively learn the texture and shape features of osteosarcoma in MRI images. We need to obtain global features of osteosarcoma images. Different from previous work [48] that used convolutional layers to refine the combined features, due to the limitation of the convolution operator, the receptive field is limited. We utilize the Swin Transformer to enhance features so that global information can be exploited. First, we reduce the number of channels of I_i through the fully connected layer to obtain $I_i' \in R^{R_i \times R_i \times C}$. I_i' merges the output features from the branch refinement (I_{i-1}) by feature merging. If $i = 1$, merging is not performed and output directly. Therefore, the following equation can be obtained.

$$O_i = \begin{cases} Conv_{1\times1}(Concat(I_i', T_{i-1}^j)) + I_i', & if\ i > 1 \\ I_i' & if\ i = 1 \end{cases} \tag{8}$$

where $Concat(\cdot)$ represents feature connections across channels, and $Conv_{1\times1}$ represents 1×1 convolution with output channel C. We reshape O_i into $R^{{R_i}^2 \times C}$, The vector sequence is obtained by using the J Transformer blocks to explore the global information as follows.

$$\begin{aligned} &T_i^0 = O_i, \\ &\hat{T}_i^j = MHSA(T_i^{j-1}) + T_i^{j-1},\ j = 1, \ldots, J, \\ &T_i^j = MLP(\hat{T}_i^j) + \hat{T}_i^j,\ j = 1, \ldots, J \end{aligned} \tag{9}$$

In the formula, $MHSA(\cdot)$ represents the standard multi-head self-attention in the Swin Transformer, and the specific structure is shown in Figure 3. $MLP(\cdot)$ is a two-layer multilayer perceptron. After using the Swin Transformer, we obtain $T_i^J \in R^{{R_i}^2 \times C}$, and reshape it into $R^{R_i \times R_i \times C}$. In our experiments, we use $J = 3$. After dealing with different scales, we obtain $\{T_1^J, T_2^J, \ldots, T_n^J\}$. The final output features of the global enhancement blocks are formed by a series of n-enhanced feature maps (T_i^J), denoted as

$$\begin{aligned} &T_i^0 = Y_i, \\ &\hat{T}_i^k - MHSA(T_i^{k-1}) + T_i^{k-1}, k = 1, \ldots, K, \\ &T = Concat(T_1^K, T_2^K, \ldots, T_n^K) \end{aligned} \tag{10}$$

The final obtained T is used to predict the target mask (M).

Local Enhancement Block (LEB). Local enhancement follows the same process as GEB. I_i is processed by a fully connected layer and FMU to generate O_i. Unlike using Swin Transformer blocks to handle O_i, LEB utilizes traditional convolution to refine Y_i, thereby encoding local information. Global and local information can complement each other. After LEB, let $\{Z_1, Z_2, \ldots, Z_n\}$ denote output features from different scales. Similarly, the final output feature of LEB is formed by the interpolation and concatenation of Z_i, denoted as

$$Z = Concat(Z_1, T_2, \ldots, T_n) \tag{11}$$

The output (Z) is used to predict the target mask (M).

3.4. Loss Function

The features from both GEB and LEB are used to predict the target mask of the query image with the losses L_{GEB} and L_{LEB}. The final loss of the whole network is defined as

$$L_{seg} = L_{GEB} + L_{LEB} \tag{12}$$

Here, both L_{GEB} and L_{LEB} are common cross-entropy losses in semantic segmentation [49]. During testing, the final prediction for the query image is the average of the output of the prediction by the global augmentation module and the local augmentation module.

The osteosarcoma segmentation method proposed in this study not only overcomes the difficulty of working with less labeled data for osteosarcoma, but also maintains a certain range of accuracy when segmenting osteosarcomas of different shapes and sizes, effectively segmenting the boundaries of osteosarcoma. More importantly, this method can effectively assist doctors in diagnosing osteosarcoma and solve common medical problems in developing countries.

4. Experiment

4.1. Experimental Settings

4.1.1. Dataset

In order to better measure the accuracy, robustness and effectiveness of the algorithm proposed in this paper, we utilized more than 5000 MRI images of osteosarcoma provided by the Monash University. These osteosarcoma images were from 204 patients (92 men and 112 women). Among these patients, 181 were diagnosed with osteosarcoma and the remaining 23 were undiagnosed. We selected 80% of the data as the training set (D_{train}) and 20% of the data as the test set (D_{test}). The details of the patient information items are shown in Table 2.

Table 2. Patient information items.

Characteristic	**Type**	***D* = 204**	
		D_{train} **= 164 (80.4%)**	D_{test} **= 40 (16.9%)**
Sex	Female	69 (42.1%)	23 (57.5%)
	Male	95 (57.9%)	17 (42.5%)
Marital status	Married	19 (11.6%)	13 (32.5%)
	Unmarried	145 (88.4%)	27 (67.5%)
SES	Low	66 (40.2%)	12 (30.0%)
	High	98 (59.8%)	28 (70.0%)
Surgery	Yes	146 (89.0%)	35 (87.5%)
	No	18 (11.0%)	5 (12.5%)
Grade	Low	15 (9.1%)	26 (65.0%)
	High	149 (90.9%)	14 (35%)
Location	Axial	21 (12.8%)	8 (20.0%)
	Extremity	109 (66.5%)	29 (72.5%)

4.1.2. Evaluation Metrics

We selected accuracy, precision, recall, and F1-score as metrics to evaluate the quality of the model. The first three indicators are quantified by the four parameters, TP, TN, FP, and FN, in the confusion matrix, TP (true positive) indicating that the actual osteosarcoma area is consistent with the segmentation result, TN (true negative) indicating that the actual normal area is consistent with the segmentation result, FP (false positive) indicating that it is actually a normal area, but it is determined to be an osteosarcoma area, and FN (false negative) indicating that it is actually an osteosarcoma area, but iis judged as a normal area. Therefore, we define the relevant indicators as follows:

Accuracy (*Acc*) is the proportion of the correct number of samples (*TP* and *TN*) to the total number of samples.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

Precision (*Pre*): the proportion of correctly predicted positive samples (*TP*) to all predicted positives (*TP* and *FP*).

$$Pre = \frac{TP}{TP + FP} \tag{14}$$

Recall (*Re*): the proportion of correctly predicted positive samples (*TP*) to all actual positives (*TP* and *FN*).

$$Re = \frac{TP}{TP + FN} \tag{15}$$

*F*1 score (*F*1): due to accuracy and recall affecting each other, we find a balance between them in order to combine their performance. It is defined as follows.

$$F1 = \frac{2 * Pre * Re}{Pre + Re} \tag{16}$$

In addition to the above-mentioned common model evaluation indicators, there are usually two special indicators in the segmentation task, the dice similarity coefficient (*DSC*) and the intersection of union (*IoU*), where *M* is the mask and *Z* is the result of the segmentation [50].

$$DSC = \frac{2 * |Z \cap M|}{|Z| + |M|} \tag{17}$$

$$IoU = \frac{Z \cap M}{Z \cup M} \tag{18}$$

4.2. Brief Introduction of Comparison Algorithms

We use the FPN [51], MSRN [25], MSFCN [2], PFENet [43], PSPNet [52], UNet [53], FCN [54] and our proposed MFENet for comparative experimental analysis. Below is a brief introduction to them:

1. The feature pyramid network (FPN) [51] aims to use the hierarchical semantic features inherent in convolutional networks to build feature pyramids. It substantially improves the performance of small object detection by a simple change in network connectivity, with essentially no increase in the computational effort of the original model.
2. The multiple supervised residual network (MSRN) [25] incorporates three supervised edge output modules in the residual network to guide the learning of low-level shape features and high-level semantic features. This helps to segment low-contrast tumor regions.
3. The multiple supervised fully convolutional network (MSFCN) [2] introduces convolution kernels of different scales based on fully convolutional neural networks to guide multi-scale feature learning as a contraction structure, so that local and global image features can be captured.

4. The prior-guided feature enrichment network (PFENet) [44] is a few-shot segmentation network that can quickly adapt to new classes with few labeled support samples. Simultaneously, a prior mask and feature enrichment module is developed, which uses supporting features and prior masks to adaptively enrich query features, improving the model's segmentation performance.
5. The pyramid scene parsing network (PSPNet) [52] provides a pyramid scene parsing network that may combine contextual information from several regions, boosting the ability to obtain global knowledge.
6. The UNet [53] is an encoder–decoder U-shaped network structure. The encoder part completes feature extraction, and the decoder part completes upsampling by introducing skip connections and information connections from different convolutional layers. It is a classic network model in the field of medical image segmentation.
7. The fully convolutional network (FCN) [54] is an end-to-end network structure that replaces the fully connected layer in the classification network with convolutional layers and pooling layers, and introduces a skip connection structure, so that the network structure can adapt to pixel-level dense prediction tasks.
8. The multi-scale feature enhancement network (MFENet) is a new method we propose in this paper. It uses transformer blocks to extract global information and convolution to extract local information and generate feature information at different scales, effectively combining the advantages of both [50,55,56].

4.3. Evaluation of Segmentation Effect

We examine the effects before and after applying the MFEM in the segmentation of osteosarcoma MRI images in this paper to quantify its effectiveness. The findings are displayed in Figure 4. The ground truth mask is on the left, the segmentation effect without the MFEM is in the middle, and the segmentation effect with the MFEM is on the right. We can observe that the segmentation result has an incorrect segmentation region before the prior mask is introduced. In contrast, the segmentation results are closer to the true labels after the prior mask is added. The accuracy of segmentation has been improved.

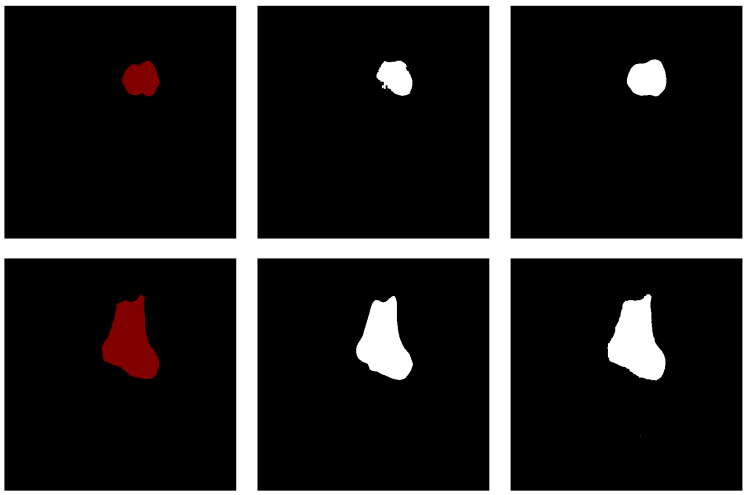

Figure 4. Comparison of segmentation effects before and after adding MFEM. The red part of the figure is the ground truth marked by the doctor. The white area is the tumor area predicted by the model.

When improving global features, we additionally assess the effects of various transformer block counts on DSC and IoU metrics in the MFEM. The outcomes demonstrate the robustness of our method in terms of the various L selections and the comparability of our method's output. However, Layer = 3 is when the best performance is shown. This is most likely a result of the tiny Layer = 2 not properly utilizing the global information. The net-

work may over-adapt to the base class and perform somewhat worse if more transformer blocks (Layer = 4) are employed. Table 3 displays the specific outcomes.

Table 3. The effect of different numbers of transformer blocks.

Block	Layer	DSC	IOU
GEB	3	0.883	0.865
LEB	3	0.881	0.876
GEB + LEB	2	0.916	0.925
GEB + LEB	3	0.942	0.952
GEB + LEB	4	0.937	0.939

Figure 5 shows the comparison of the segmentation effect of osteosarcoma under different models. We list four groups of osteosarcoma images of different types in total. The initial image of the osteosarcoma is shown in the first column, the lesion area is shown in the second column, and the segmentation results for each model are shown in the following columns. By visually comparing the comparative photos of osteosarcoma segmentation in these four groups of images, we discovered that the suggested osteosarcoma segmentation model segmented the best. In particular, the best fit to the real lesion area was achieved when segmenting relatively small size lesion areas. On the contrary, other models, especially the FCN, perform poorly in segmentation performance, mainly because most of these models are based on convolution operations with limited receptive fields and cannot obtain global features of images. Additionally, the model proposed in this paper can effectively enhance the global features and improve the accuracy of segmentation by adding Swin Transformer blocks.

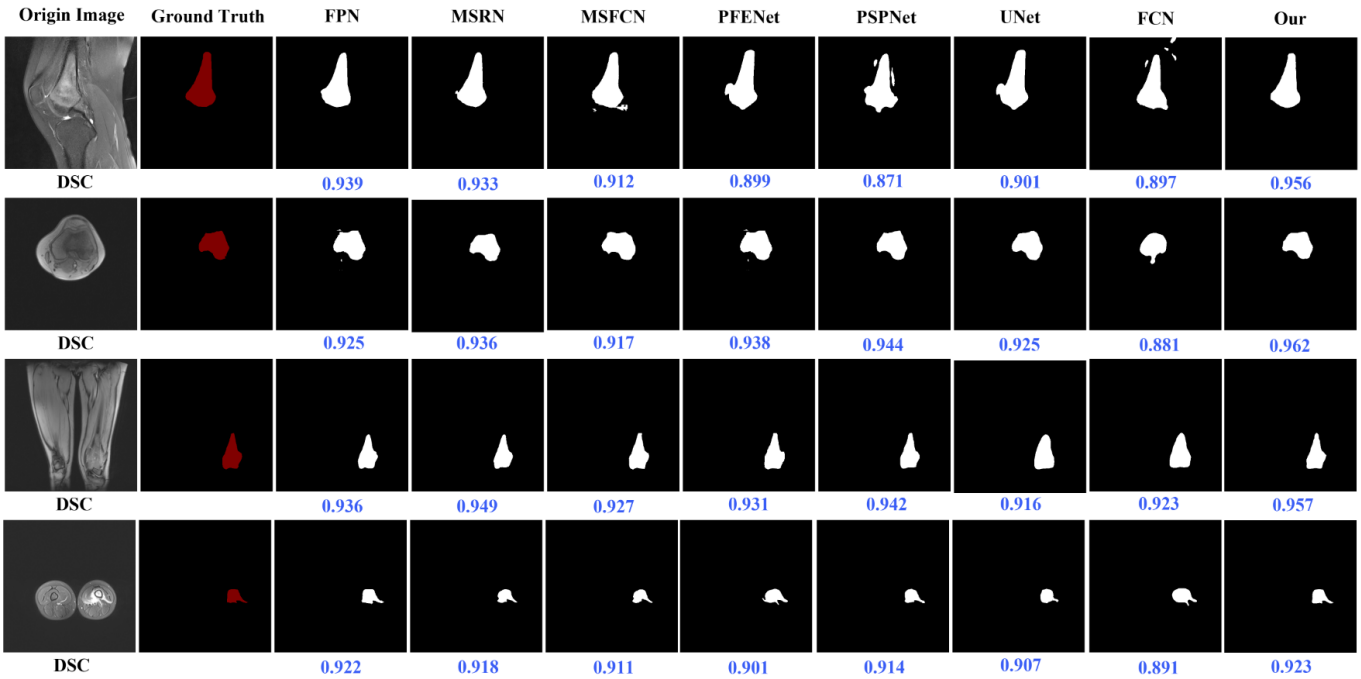

Figure 5. Comparison of segmentation effects of different models. The red part of the figure is the ground truth marked by the doctor. the white area is the tumor area predicted by different models.

In order to evaluate the performance of different methods more clearly. We trained a total of 300 epochs, and the results were averaged and then compared and analyzed. Additionally, four common model evaluation metrics were used to evaluate the various model models, and the results are shown in Figures 6–9.

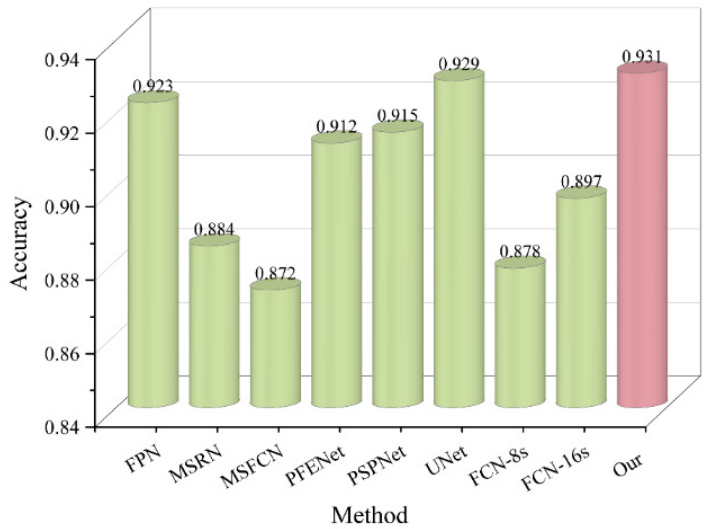

Figure 6. Comparison of the accuracy.

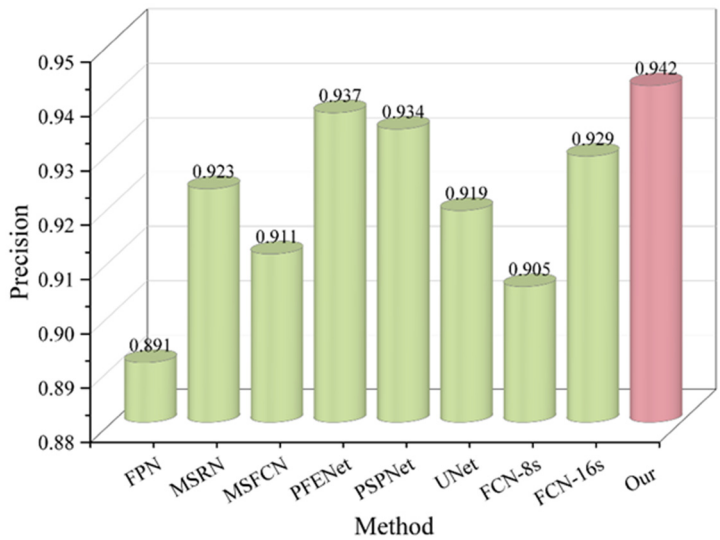

Figure 7. Comparison of the precision.

Figure 6 shows the comparison of the accuracy of each model. We can see that the accuracy of the three models except the MSRN, MSFCN and FCN is below 90%. Because the FCN architecture is based on the VGG top-level feature segmentation as the target, the receptive field is limited. However, its own structure does not take into account the relationship between pixels and lacks spatial consistency. This will result in the loss of feature information in several small-scale osteosarcoma pictures, lowering segmentation accuracy. The PSPNet gathers features using a feature pyramid module, and its accuracy is 91.5%, which is second only to the UNet and our technique. Compared with other traditional methods, the PSPNet adds a pyramid structure that can aggregate contextual information from different regions, thereby improving the ability to obtain global information. demonstrating that multi-scale features can help improve segmentation performance. However, overall, the study's suggested model performs the best, with a 93.1% accuracy rate, demonstrating that by incorporating Swin Transformer blocks to strengthen the global characteristics, segmentation accuracy may be increased and clinicians can obtain accurate diagnosis findings.

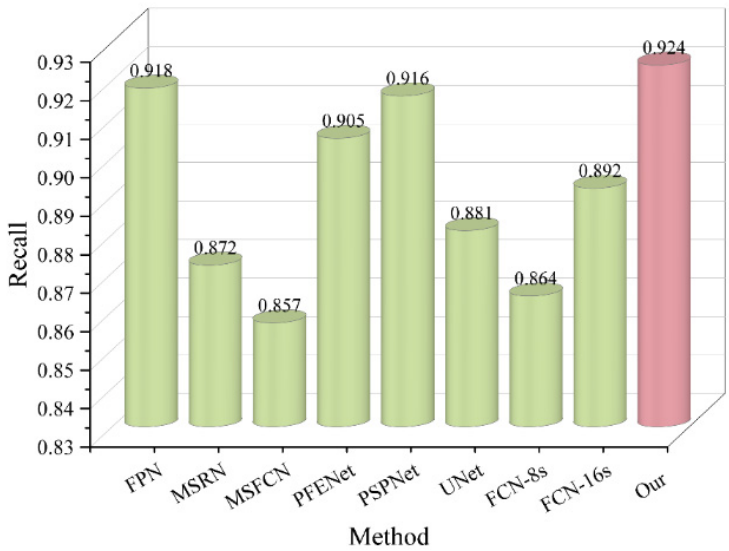

Figure 8. Comparison of the recall.

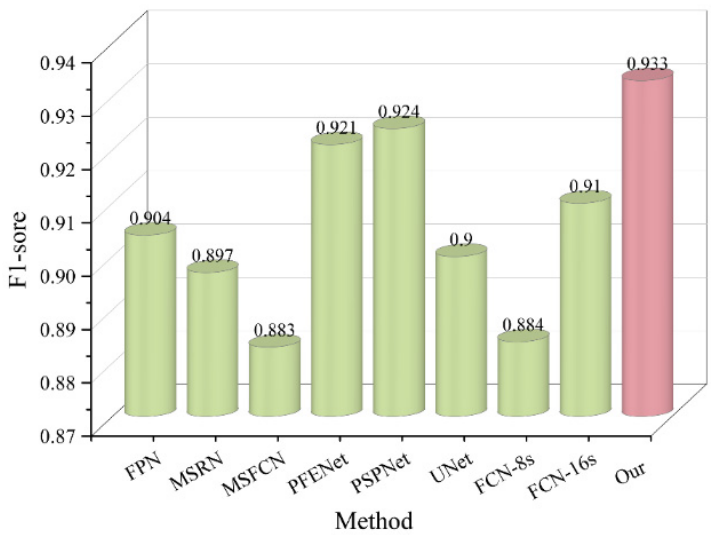

Figure 9. Comparison of the F1 score.

The precision and recall of each model are compared in Figures 7 and 8. According to the data in the figure, the model suggested in this study has the best accuracy, at 94.2%. Compared with the PFENet and PSPNet, the improvements are 0.5% and 0.8%, respectively. It is shown that using transformer in the few-shot segmentation method to guide global information on the merged query support features while capturing multi-scale features can effectively improve segmentation precision. Comparing with the recall rate, our proposed method far surpasses that of other methods, reaching 93.4%, indicating that the method can better predict the osteosarcoma samples of those small targets without missing a lot of relatively unlikely but relatively small samples. In clinical diagnosis, the occurrence of misdiagnosis and missed diagnosis can be better avoided.

Since the accuracy and recall rates affect each other, we would like to achieve both ideally, but they are mutually constrained in practice. As a result, the model developed in this study was thoroughly examined in order to balance the effects of precision and recall. We added an evaluation metric, the F1 score. As shown in Figure 9 above, although our proposed method is slightly inferior in accuracy, its F1 value is always the highest, reaching 93.3%. The other models performed averagely except for the PSPNet, which reached 92.4%

accuracy, and the PFENet, which reached 92.1% accuracy. The results show that the model we proposed has good robustness.

DSC and IoU are two types of special metrics for image segmentation. As shown in Figure 10, the point model closer to the upper right corner shows a higher IoU-DSC value and better segmentation efficiency. The closer to the point in the lower left corner, the lower the IoU-DSC value exhibited by the model, and the less ideal the segmentation effect. We can see by comparing the positions of other models in the figure that the DSC value of the method proposed in this paper is 91.5% while the IoU value reaches 88.1%, which far exceeds that of the other methods. The results demonstrate that the method presented in this research has the best segmentation impact and can better automatically define the lesion region without over-segmentation and under-segmentation. The method can become a good reference for doctors, better assist doctors in diagnosis and reduce the workload of doctors.

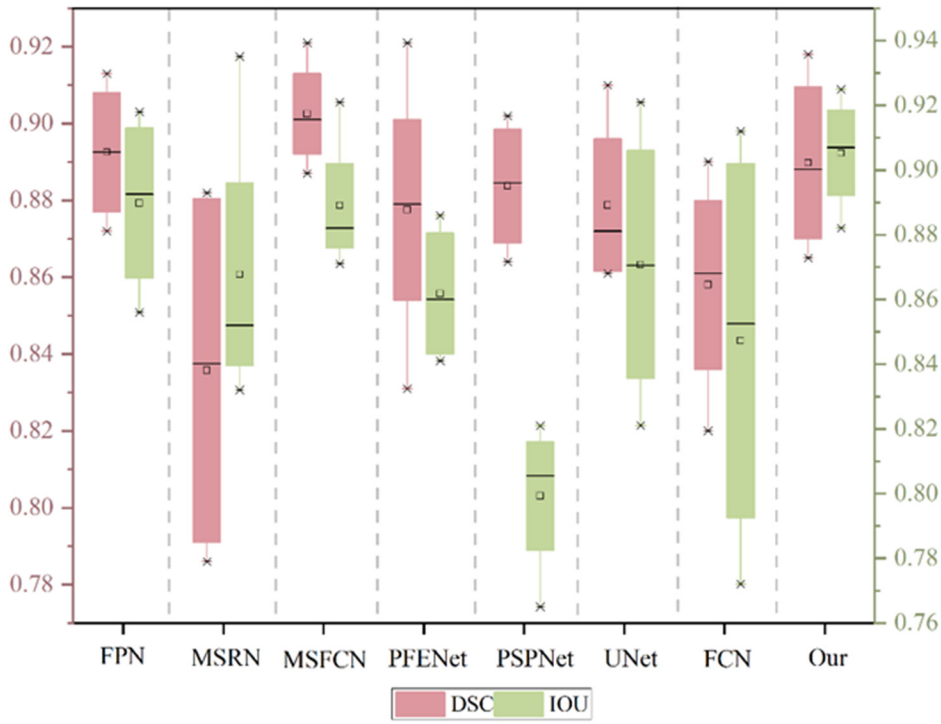

Figure 10. Comparison of the IoU and DSC. The X-axis is to represent the different models.

To demonstrate that adding a prior mask to the model can speed up the computation of the model, we compared the MFENet model which has a prior mask with the model without it, as shown in Table 4. We can see that the run rate with a prior mask is higher than the rate of the model without it, where FPS represents the number of frames per second of transmitted images.

Table 4. The effect of using a prior block.

Methods	DSC	IOU	Speed
Ours (without prior)	0.892	0.865	15.8 FPS
Ours	0.905	0.891	17.1 FPS

5. Conclusions

In our study, we discuss a medical decision-making system, the core of which is an automatic segmentation network for osteosarcoma named MFENet. It extracts local and global multi-scale features by introducing traditional convolutional neural networks and

transformer blocks. It may not only assist physicians in segmenting osteosarcomas of different shapes and sizes with a range of accuracy, but also greatly save computational costs by introducing a prior mask in the model to replace the reconstruction process from features when segmenting MRI images. However, there are some limitations in this paper, which include the fact our proposed method can only outline the general area of the tumor. In general, the interior of a tumor includes necrotic areas, edematous areas, and the actual tumor area. Single-mode MRI images cannot distinguish tumors from muscle tissue and joint fluid because they all show low signals in the image. In the future, we should introduce more modal data to improve the experimental content based on this experiment.

The two most critical problems in AI-based medical image analysis are fine-grained lesion segmentation and disease classification. In order to help clinicians diagnose osteosarcoma more thoroughly and address issues with osteosarcoma diagnosis in underdeveloped nations, in this study, we performed a segmentation of osteosarcoma. In the future, we will use the segmentation results of this paper as samples for the next step of disease grading, and effectively grade the tumor mass according to the relevant parameters of osteosarcoma.

Author Contributions: Writing—original draft preparation, K.H. and Y.Q.; writing—review and editing, F.G. and J.W.; visualization, K.H., Y.Q.; funding acquisition, K.H. All authors have read and agreed to the published version of the manuscript.

Funding: The general project of Changsha Technology Bureau (grant no. KC1705026) and Natural Science Foundation of Hunan Province (grant no. 2020JJ4647 and grant no. 2020JJ6064) supported this work.

Data Availability Statement: Data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 12 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tahmasbi-Arashlow, M.; Barnts, K.L.; Nair, M.K.; Cheng, Y.-S.L.; Reddy, L.V. Radiographic manifestations of fibroblastic osteosarcoma: A diagnostic challenge. *Imaging Sci. Dent.* **2019**, *49*, 235–240. [CrossRef] [PubMed]
2. Liu, F.; Zhu, J.; Lv, B.; Yang, L.; Sun, W.; Dai, Z.; Gou, F.; Wu, J. Auxiliary Segmentation Method of Osteosarcoma MRI Image Based on Transformer and U-Net. *Comput. Intell. Neurosci.* **2022**, *2022*, 9990092. [CrossRef] [PubMed]
3. Harrison, D.J.; Geller, D.S.; Gill, J.D.; Lewis, V.O.; Gorlick, R. Current and future therapeutic approaches for osteosarcoma. *Expert Rev. Anticancer Ther.* **2018**, *18*, 39–50. [CrossRef]
4. Ouyang, T.; Yang, S.; Gou, F.; Dai, Z.; Wu, J. Rethinking U-Net from an Attention Perspective with Transformers for Osteosarcoma MRI Image Segmentation. *Comput. Intell. Neurosci.* **2022**, *2022*, 7973404. [CrossRef] [PubMed]
5. Burkett, B.J.; Fagan, A.J.; Felmlee, J.P.; Black, D.F.; Lane, J.I.; Port, J.D.; Rydberg, C.H.; Welker, K.M. Clinical 7-T MRI for neuroradiology: Strengths, weaknesses, and ongoing challenges. *Neuroradiology* **2021**, *63*, 167–177. [CrossRef] [PubMed]
6. Liu, F.; Gou, F.; Wu, J. An Attention-Preserving Network-Based Method for Assisted Segmentation of Osteosarcoma MRI Images. *Mathematics* **2022**, *10*, 1665. [CrossRef]
7. Ling, Z.; Yang, S.; Gou, F.; Dai, Z.; Wu, J. Intelligent Assistant Diagnosis System of Osteosarcoma MRI Image Based on Transformer and Convolution in Developing Countries. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5563–5574. [CrossRef]
8. Zhou, Z.; Gou, F.; Tan, Y.; Wu, J. A Cascaded Multi-Stage Framework for Automatic Detection and Segmentation of Pulmonary Nodules in Developing Countries. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5619–5630. [CrossRef]
9. Wu, J.; Guo, Y.; Gou, F.; Dai, Z. A medical assistant segmentation method for MRI images of osteosarcoma based on DecoupleSegNet. *Int. J. Intell. Syst.* **2022**, *37*, 8436–8461. [CrossRef]
10. Tang, H.; Huang, H.; Liu, J.; Zhu, J.; Gou, F.; Wu, J. AI-Assisted Diagnosis and Decision-Making Method in Developing Countries for Osteosarcoma. *Healthcare* **2022**, *10*, 2313. [CrossRef]
11. Wu, J.; Gou, F.; Tan, Y. A Staging Auxiliary Diagnosis Model for Nonsmall Cell Lung Cancer Based on the Intelligent Medical System. *Comput. Math. Methods Med.* **2021**, *2021*, 6654946. [CrossRef]
12. Qin, Y.; Li, X.; Wu, J.; Yu, K. A management method of chronic diseases in the elderly based on IoT security environment. *Comput. Electr. Eng.* **2022**, *102*, 108188. [CrossRef]
13. Wu, J.; Yang, S.; Gou, F.; Zhou, Z.; Xie, P.; Xu, N.; Dai, Z. Intelligent Segmentation Medical Assistance System for MRI Images of Osteosarcoma in Developing Countries. *Comput. Math. Methods Med.* **2022**, *2022*, 7703583. [CrossRef]
14. Shen, Y.; Gou, F.; Dai, Z. Osteosarcoma MRI Image-Assisted Segmentation System Base on Guided Aggregated Bilateral Network. *Mathematics* **2022**, *10*, 1090. [CrossRef]

15. Gou, F.; Wu, J. Novel data transmission technology based on complex IoT system in opportunistic social networks. *Peer-to-Peer Netw. Appl.* **2022**, *11*, 23. [CrossRef]
16. Zhan, X.; Long, H.; Gou, F.; Duan, X.; Kong, G.; Wu, J. A Convolutional Neural Network-Based Intelligent Medical System with Sensors for Assistive Diagnosis and Decision-Making in Non-Small Cell Lung Cancer. *Sensors* **2021**, *21*, 7996. [CrossRef]
17. Guo, F.; Ng, M.; Wright, G. Cardiac Cine MRI Left Ventricle Segmentation Combining Deep Learning and Graphical Models. In Proceedings of the Medical Imaging Conference—Image Processing, Houston, TX, USA, 15–20 February 2020; Volume 11313, p. 2021.
18. Wang, L.; Yu, L.; Zhu, J.; Tang, H.; Gou, F.; Wu, J. Auxiliary Segmentation Method of Osteosarcoma in MRI Images Based on Denoising and Local Enhancement. *Healthcare* **2022**, *10*, 1468. [CrossRef]
19. Mo, S.; Cai, M.; Lin, L.; Tong, R.; Chen, Q.; Wang, F.; Hu, H.; Iwamoto, Y.; Han, X.-H.; Chen, Y.-W. Mutual Information-Based Graph Co-Attention Networks for Multimodal Prior-Guided Magnetic Resonance Imaging Segmentation. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 2512–2526. [CrossRef]
20. Wu, J.; Zhou, L.; Gou, F.; Tan, Y. A Residual Fusion Network for Osteosarcoma MRI Image Segmentation in Developing Countries. *Comput. Intell. Neurosci.* **2022**, *2022*, 7285600. [CrossRef]
21. Gou, F.; Liu, J.; Zhu, J.; Wu, J. A Multimodal Auxiliary Classification System for Osteosarcoma Histopathological Images Based on Deep Active Learning. *Healthcare* **2022**, *10*, 2189. [CrossRef] [PubMed]
22. Zhuang, Q.; Dai, Z.; Wu, J. Deep Active Learning Framework for Lymph Node Metastasis Prediction in Medical Support System. *Comput. Intell. Neurosci.* **2022**, *2022*, 4601696. [CrossRef] [PubMed]
23. Wu, J.; Xiao, P.; Huang, H.; Gou, F.; Zhou, Z.; Dai, Z. An Artificial Intelligence Multiprocessing Scheme for the Diagnosis of Osteosarcoma MRI Images. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4656–4667. [CrossRef]
24. Nasor, M.; Obaid, W. Segmentation of osteosarcoma in MRI images by K-means clustering, Chan-Vese segmentation, and iterative Gaussian filtering. *IET Image Process.* **2021**, *15*, 1310–1318. [CrossRef]
25. Zhang, R.; Huang, L.; Xia, W.; Zhang, B.; Qiu, B.; Gao, X. Multiple supervised residual network for osteosarcoma segmentation in CT images. *Comput. Med. Imaging Graph.* **2018**, *63*, 1–8. [CrossRef]
26. Shuai, L.; Gao, X.; Wang, J. Wnet++: A Nested W-shaped Network with Multiscale Input and Adaptive Deep Supervision for Osteosarcoma Segmentation. In *Proceedings of the 2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT), Xi'an, China, 18–20 August 2021*; IEEE: Washington, DC, USA, 2021; pp. 93–99.
27. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
28. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021.
29. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Zhang, L. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
30. Lin, A.; Chen, B.; Xu, J.; Zhang, Z.; Lu, G. DS-TransUNet: Dual Swin Transformer U-Net for Medical Image Segmentation. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 4005615. [CrossRef]
31. Zhang, Y.; Liu, H.; Hu, Q. TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021; pp. 14–24.
32. Zhang, Z.; Sun, B.; Zhang, W. Pyramid Medical Transformer for Medical Image Segmentation. *arXiv* **2021**, arXiv:2104.14702.
33. Shi, X.; Cui, Z.; Zhang, S.; Cheng, M.; He, L.; Tang, X. Multi-similarity based Hyperrelation Network for few-shot segmentation. *IET Image Process.* **2022**, *17*, 204–214. [CrossRef]
34. Liu, W.; Zhang, C.; Lin, G.; Liu, F. CRNet: Cross-Reference Networks for Few-Shot Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 4164–4172.
35. Zhang, M.; Shi, M.; Li, L. MFNet: Multi-class Few-shot Segmentation Network with Pixel-wise Metric Learning. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 8586–8598. [CrossRef]
36. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 $\times$ 16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
37. Liu, Y.; Sun, G.; Qiu, Y.; Zhang, L.; Chhatkuli, A.; Van Gool, L. Transformer in Convolutional Neural Networks. *arXiv* **2021**, arXiv:2106.03180.
38. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
39. Xue, M.; Cao, P.; Hou, B.; Liu, W. Data-driven decision-making with weights and reliabilities for diagnosis of thyroid cancer. *Int. J. Mach. Learn. Cybern.* **2022**, *13*, 2257–2271. [CrossRef]
40. Vaiyapuri, T.; Dutta, A.K.; Punithavathi, I.S.H.; Duraipandy, P.; Alotaibi, S.S.; Alsolai, H.; Mohamed, A.; Mahgoub, H. Intelligent Deep-Learning-Enabled Decision-Making Medical System for Pancreatic Tumor Classification on CT Images. *Healthcare* **2022**, *10*, 677. [CrossRef] [PubMed]
41. Ding, S.; Hu, S.; Li, X.; Zhang, Y.; Wu, D.D. Leveraging Multimodal Semantic Fusion for Gastric Cancer Screening via Hierarchical Attention Mechanism. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *52*, 4286–4299. [CrossRef]

42. Wu, J.; Liu, Z.; Gou, F.; Zhu, J.; Tang, H.; Zhou, X.; Xiong, W. BA-GCA Net: Boundary-Aware Grid Contextual Attention Net in Osteosarcoma MRI Image Segmentation. *Comput. Intell. Neurosci.* **2022**, *2022*, 3881833. [CrossRef]
43. Gou, F.; Wu, J. Triad link prediction method based on the evolutionary analysis with IoT in opportunistic social networks. *Comput. Commun.* **2022**, *181*, 143–155. [CrossRef]
44. Lv, B.; Liu, F.; Gou, F.; Wu, J. Multi-Scale Tumor Localization Based on Priori Guidance-Based Segmentation Method for Osteosarcoma MRI Images. *Mathematics* **2022**, *10*, 2099. [CrossRef]
45. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
46. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
47. Zhang, C.; Lin, G.; Liu, F.; Yao, R.; Shen, C.; Soc, I.C. CANet: Class-Agnostic Segmentation Networks with Iterative Refinement and Attentive Few-Shot Learning. In Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–21 June 2019; pp. 5212–5221.
48. Hu, T.; Yang, P.; Zhang, C.; Yu, G.; Mu, Y.; Snoek, C.G. Attention-Based Multi-Context Guiding for Few-Shot Semantic Segmentation. In Proceedings of the 33rd AAAI Conference on Artificial Intelligence/31st Innovative Applications of Artificial Intelligence Conference/9th AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 8441–8448.
49. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
50. Gou, F.; Wu, J. An Attention-based AI-assisted Segmentation System for Osteosarcoma MRI Images. In Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA, 6–8 December 2022; pp. 1539–1543. [CrossRef]
51. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
52. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
53. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
54. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
55. Zhan, X.; Liu, J.; Long, H.; Zhu, J.; Tang, H.; Gou, F.; Wu, J. An Intelligent Auxiliary Framework for Bone Malignant Tumor Lesion Segmentation in Medical Image Analysis. *Diagnostics* **2023**, *13*, 223. [CrossRef]
56. Wei, H.; Lv, B.; Liu, F.; Tang, H.; Gou, F.; Wu, J. A Tumor MRI Image Segmentation Framework Based on Class-Correlation Pattern Aggregation in Medical Decision-Making System. *Mathematics* **2023**, *11*, 1187. [CrossRef]