

Article

A Tumor MRI Image Segmentation Framework Based on Class-Correlation Pattern Aggregation in Medical Decision-Making System

Hui Wei ¹, Baolong Lv ¹, Feng Liu ^{2,3}, Haojun Tang ^{4,*}, Fangfang Gou ⁴  and Jia Wu ^{4,5} ¹ School of Modern Service Management, Shandong Youth University of Political Science, Jinan 250102, China² School of Information Engineering, Shandong Youth University of Political Science, Jinan 250102, China³ New Technology Research and Development Center of Intelligent Information Controlling in Universities of Shandong, Jinan 250103, China⁴ School of Computer Science and Engineering, Central South University, Changsha 410083, China⁵ Research Center for Artificial Intelligence, Monash University, Melbourne, Clayton, VIC 3800, Australia

* Correspondence: live_0.0@csu.edu.cn

Abstract: Medical image analysis methods have been applied to clinical scenarios of tumor diagnosis and treatment. Many studies have attempted to optimize the effectiveness of tumor MRI image segmentation by deep learning, but they do not consider the optimization of local details and the interaction of global semantic information. Second, although medical image pattern recognition can learn representative semantic features, it is challenging to ignore useless features in order to learn generalizable embeddings. Thus, a tumor-assisted segmentation method is proposed to detect tumor lesion regions and boundaries with complex shapes. Specifically, we introduce a denoising convolutional autoencoder (DCAE) for MRI image noise reduction. Furthermore, we design a novel tumor MRI image segmentation framework (NFSR-U-Net) based on class-correlation pattern aggregation, which first aggregates class-correlation patterns in MRI images to form a class-correlational representation. Then the relationship of similar class features is identified to closely correlate the dense representations of local features for classification, which is conducive to identifying image data with high heterogeneity. Meanwhile, the model uses a spatial attention mechanism and residual structure to extract effective information of the spatial dimension and enhance statistical information in MRI images, which bridges the semantic gap in skip connections. In the study, over 4000 MRI images from the Monash University Research Center for Artificial Intelligence are analyzed. The results show that the method achieves segmentation accuracy of up to 96% for tumor MRI images with low resource consumption.

Keywords: MRI tumor segmentation; intelligent assisted diagnosis; classification; class-correlational representation

MSC: 68T01



Citation: 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. <https://doi.org/10.3390/math11051187>

Academic Editor: Jakub Nalepa

Received: 1 February 2023

Revised: 24 February 2023

Accepted: 24 February 2023

Published: 28 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Automatic tumor image recognition and analysis methods are of vast potential value for improving the diagnosis and treatment planning of individual patients. During clinical diagnosis, tumor detection methods include symptomatic examination, imaging examination, and pathological biopsy [1]. Among them, magnetic resonance imaging (MRI) is an effective tool for diagnosing tumors. MRI causes little damage to biological tissues, has multi-planarity with high tissue contrast resolution, and allows to demonstrate the location of the lesion area and distinguish benign from malignant tumors. Nevertheless, in the process of tumor diagnosis, massive redundant MRI image data is generated for each tumor patient, and it is laborious for clinicians to identify tumor images.

Due to their high clinical relevance, medical image processing techniques play a crucial role in computer-aided diagnosis systems, giving rise to automatic, semi-automatic, and interactive MRI image recognition methods for various tumor structures [2]. In the context of a small image sample size and a lack of annotation data, transfer learning has become a popular method to alleviate the problem of the scarcity of medical image annotation data. Researchers have attempted to bypass data annotation and use semi-supervised and unsupervised learning to solve the problem of label defects. Among them, domain-adaptive algorithms are widely used in the field of medical image segmentation, which trains classifiers by learning an image-to-image transformation model between the target and source modalities to extend the data [3]. Moreover, few-shot classification methods are also dedicated to learning new visual concepts through a small number of sample images that provide transferrable information across categories and show good performance in the low-data modality [4].

Medical image processing techniques provide a feasible solution for vision tasks with low data modalities; however, medical images also suffer from problems such as low resolution and noise redundancy. For image denoising, traditional image processing techniques such as Gaussian smoothing, anisotropic diffusion, and wavelet denoising have performed better in digital image signal processing [5], which facilitates the resolution of image quality problems caused by external distortion and noise. Recently, researchers have also proposed deep learning strategies for the image enhancement of medical images. In the case of tumor MRI images, Mehta et al. [6] constructed an architecture consisting of encoder-decoder pairs to denoise tumor MRI images and achieved a higher peak signal-to-noise ratio (PSNR) at multiple fundamental noise levels. Yang et al. [7] proposed a joint network applied to brain tumor denoising and classification that consists of a CNN baseline for noise reduction of MRI scans and has an excellent performance in identifying noisy medical images.

In conditions where the annotated data are sufficient, supervised learning can predict the exact results more easily based on prior experience. Deep learning-based methods for MRI tumor image analysis have several popular frameworks, among which, a transformer with excellent global representation capability has been successfully applied in the field of computer vision, and it possesses good performance for medical image segmentation in scenarios with large datasets. However, the transformer still has certain shortcomings, with problems such as complex calculations, high risk of overfitting, and large training data required, which requires the high performance of the hardware equipment equipped in the primary hospital [8]. In contrast, the U-shaped CNN architecture and its variants with powerful local information extraction have achieved excellent performance in medical image segmentation [9]. Compared with 3D structures, 2D structures require fewer parameters and have lower computational complexity. The existing 2D U-Net focuses on the optimization of local details and the interaction of global semantic information in MRI tumor images by rationally improving the model structure, thus enabling the network to exhibit significant segmentation performance [10].

Medical image processing technology has alleviated the challenge of low efficiency in tumor diagnosis to some extent. However, due to the specificity of tumors, MRI images of different tumors have diverse characteristics. Taking the MRI images of osteosarcoma as an example, the current supervised learning approach still has the following difficulties:

- (1) The process of osteosarcoma MRI detection relies mainly on manual identification by professionals. Each patient with osteosarcoma generates 600–700 MRI images in a single diagnosis, but few of them are valuable [11–13]. Redundant data aggravate the workload of identification and consume a lot of time and energy from doctors, resulting in low diagnostic efficiency and being prone to misdiagnosis [14,15].
- (2) Osteosarcoma is very costly to diagnose [16,17]. Developing countries are economically backward and lack a well-developed medical system [18], facing difficulties in acquiring high-priced MRI equipment and a scarcity of clinicians. Patients are prone to delay the best treatment time due to economic and geographical reasons [19,20].

- (3) Osteosarcoma is very difficult to diagnose. Osteosarcoma has high variability in shape and location [21,22], and the MRI images often contain redundant noise information from outside the target background, which makes it difficult for doctors to distinguish tumor tissue from surrounding normal tissue [23,24]. Most hospitals lack a complete osteosarcoma-assisted segmentation system to detect potential features of osteosarcoma images that cannot be identified with the naked eye by quantitative analysis [25].
- (4) Weakness of osteosarcoma MRI image detection methods. In order to enhance the tumor segmentation effect, many studies have learned the mapping relationships of different features by machine learning [26–28], but they do not consider the implicit features to obtain valid information. Although training the classifier by computing numerous features can improve the segmentation accuracy, its overly complex structure can lead to a dramatic increase in parameters, making the training of the model inefficient [29,30].

In this study, an osteosarcoma-assisted segmentation method based on an MRI image segmentation framework (NFSR-U-Net) is proposed. In the method, we use the Mean Teacher semi-supervised learning method to divide the original image dataset and input it sequentially into the preprocessing process to enhance the training efficiency of the model. Meanwhile, a denoising convolutional autoencoder (DCAE) with a lightweight structure is introduced to perform noise reduction on the image, which facilitates the accuracy of segmentation. In model design, we propose an MRI image segmentation framework (NFSR-U-Net) based on class-correlation pattern aggregation to achieve detailed boundary segmentation of osteosarcoma lesion tissues. The suggested osteosarcoma-assisted segmentation method has high accuracy and few parameters, which have an indispensable impact on the auxiliary diagnosis and treatment of the tumor.

The main contributions of the study are presented below:

- (1) We employ the Mean Teacher method to partition the dataset and input it sequentially into the preprocessing process, which is conducive to improving the training efficiency of the model. At the same time, we introduce a denoising convolutional autoencoder (DCAE) to eliminate unwanted noise, which improves the feasibility of osteosarcoma image segmentation.
- (2) We propose the NFSR-U-Net, which aggregates local correlation patterns in MRI images to form class-correlational representations and identifies similar semantic features in the discrete feature space for local matching to closely correlate dense representations of local features. The model enables pixel-level embeddings of similar classes to achieve a high fit for classification by learning intra-class similarity in MRI images, which shows excellent performance in tumor tissue segmentation with large shape differences.
- (3) The NFSR-U-Net learns highly representative and hierarchical semantic features by rescaling high-level features in the middle and late stages using the spatial attention mechanism. The effectiveness of extracting spatial features of various depths by using the residual structure is also used to enhance the statistical information of textures and boundaries in MRI images. It bridges the semantic gap of skip connections in U-Net.
- (4) In this paper, over 4000 sample data acquired from the Monash University Research Center for Artificial Intelligence were used for analysis. Compared with other methods, the tumor MRI image segmentation method has a better segmentation effect and possesses fewer parameters, which facilitates model training.

2. Related Works

With the deepening of research on the assisted diagnosis of osteosarcoma and other diseases, more and more medical image segmentation technologies have been applied to the automatic segmentation of osteosarcoma, effectively improving the performance and accuracy of tumor segmentation models.

Osteosarcoma cells have multiple morphologies with large spatial and structural variability. Learning effective feature information from images and making reasonable feature selections is beneficial for achieving higher accuracy in image segmentation. Zhang et al. [31] put several supervised output blocks into the residual network, which learned shape features from images and effectively segmented osteosarcoma. Huang et al. [32] guided multi-scale feature learning by introducing supervised layers into a convolutional network to depict osteosarcoma boundaries. Pan et al. [33] used convolutional autoencoders with feature cross-fusion learning methods to generate fine fusion features and combined them with residual neural networks for label prediction. Nabid [34] introduced a network composed of convolution and recurrent unit blocks to classify osteosarcoma images precisely. Shuai et al. [35] designed an osteosarcoma segmentation model consisting of two U-shaped networks and dense skip connections by combining adaptive monitoring methods.

Magnetic resonance imaging possesses great soft tissue recognition and high spatial resolution, but it has different and unique intensity and texture, so the automatic segmentation of osteosarcoma MRI images is a challenging task. Obaid [36] introduced a method that can effectively segment MRI images of osteosarcoma with different textures, positions, or intensities by using k-means clustering and iterative Gaussian filtering strategies. Baidya [37] adopted diffusion-weighted imaging to classify the MRI images of osteosarcoma, and this method can quantitatively analyze the changes in tumor cells. Chen et al. [38] established a set of CE FS T1WI features by comparing the radionics features of MRI, which facilitates preoperative knowledge of the pathological characteristics of patients with osteosarcoma in response to neoadjuvant chemotherapy.

Apart from the segmentation of osteosarcoma, researchers have also conducted research related to image segmentation techniques for other tumors, such as brain tumors, lung tumors, etc. Guan et al. [39] proposed an AGSE-VNet segmentation model that combined channel relationships to strengthen the salient information in channels and used attention mechanisms to weaken edge features. Zhang et al. [40] proposed a hybrid clustering algorithm to segment brain tumors that fused K-means clustering with the C-means algorithm and combined morphological operations, which effectively reduced the image sensitivity to noise. Dutande et al. [41] combined the Maximum Intensity Projection method with DRS-CNN to accomplish automated segmentation of lung tumor CT images. Zhang [42] introduced the scale attention composite mechanism into U-Net, which completed the global spatial information modeling.

Medical image analysis techniques, with their powerful feature extraction capabilities, have alleviated the challenge of tumor diagnosis to some extent. In response to the struggle of models to detect potential features in images, researchers have attempted to analyze similar classes of correlation patterns in images and use them as feature transformations in deep learning to reveal the structural layout of images [43]. Kim et al. [44] learned sampling patterns and correlation measures of local structures in images based on his proposed FCSS descriptor, and finally matched points between different instances of the same object class. Zheng et al. [45] applied a spatial correlation loss method to capture the spatial relationships for the transformation of the amount of information in the image. Inspired by the above work, we introduce the class-correlational representation which is an approach more suitable for addressing the heterogeneity problem of image data. This method uses convolutional structures in the feature representation process to analyze and aggregate class-correlation patterns within the image representation, making the embeddings of similar classes as close as possible. Furthermore, on many vision tasks, P. Ramachandran et al. [46] proposed stand-alone self-attention to compute class correlation coefficients as attention weights for aggregation. Wang et al. [47] introduced non-local neural networks to apply non-local operations to capture remote dependencies. Nevertheless, these methods lack the learning of representations in a way that utilizes the class correlation tensor, and the models are prone to cause image semantic loss during image detection. Therefore, we suggest packing properties in the form of Hadamard transformations to

provide the model with the necessary semantic information in the feature representation and to learn generalizable relational embeddings.

In multimodal biomedical image segmentation research, researchers have improved the effectiveness of image segmentation by optimizing the model structure and combining it with advanced CV techniques. Among them, U-Net [48] is suitable for the segmentation of medical images. Drozdal et al. [49] verified the utility of skip connections during image segmentation, which allows models to pass spatial information and recover lost spatial information during the stitching process. The bottleneck feature of U-Net is composed of high-level semantic features collected by the contraction path. Its spatial features are associated with the location information of the segmented object, and the features between channels are concentrated on the semantic category. However, the bottleneck features come from deeper layers with redundant features, which can interfere with the segmentation results. In addition, the features coming from skip connections are calculated in the early stage, and there is a semantic gap between the features of the bottleneck feature and the features of the expansive path [50]. The attention mechanism based on a heuristic search approach enables feature selection. Woo et al. [51] introduced the convolutional block attention module, which improves the representational power by emphasizing salient features in the channel and space dimensions. Networks such as SENet [52] and ECA [53] correct the channels by learning the importance of channels. However, both of them adopt the global average pooling (GAP) operation when compressing the feature map, which directly calculates the average value of the feature map, resulting in the loss of spatial information. These models ignore the spatial characteristics of each channel.

The above shows that medical image segmentation techniques have a major impact on tumor diagnosis. However, the osteosarcoma MRI images have much noise, the bottleneck features extracted by convolutional networks have redundant features, and there is a semantic gap of skip connections, which affects the performance of the model. Therefore, we introduce an MRI image segmentation framework (NFSR-U-Net) based on class-correlation pattern aggregation in the decision-making system, which aggregates local correlation patterns in MRI images to locally match similar semantic features in the discrete feature space of tumor images, resulting in a high fit of pixel-level embeddings of similar classes, which is conducive to addressing the heterogeneity problem of tumor image data. Furthermore, NFSR-U-Net learns highly representative and hierarchical semantic features by rescaling high-level features in the middle and late stages using the spatial attention mechanism. The effectiveness of extracting spatial features of various depths by using the residual structure is also used to enhance the statistical information of textures and boundaries in MRI images. These initiatives bridge the semantic gap of skip connections in the network. Through the strategies of data set segmentation, preprocessing, and osteosarcoma MRI image segmentation, this medical decision-making system can detect the location and edges of lesion regions and realize the automated segmentation of osteosarcoma images, which effectively enhances the accuracy and reliability of medical diagnosis.

3. System Model Design

Osteosarcoma is a malignant neoplastic disease located in bone tissue. During MRI detection of osteosarcoma, patients generate a lot of complex image data. The manual screening of images and tumor diagnosis by doctors alone can be time and energy-consuming, and the tremendous workload can lead to an increased rate of misdiagnosis. Additionally, in developing countries, medical resources are tighter, and the doctor–patient ratio is severely imbalanced, making it difficult for patients to receive effective services. Moreover, the long and costly diagnostic cycle of osteosarcoma requires many families to bear high costs. AI-based image processing technology can automatically screen medical images to achieve the diagnosis of patients. The diagnostic results provide an auxiliary basis for doctors, thus improving their efficiency and reducing the cost of diagnosing. In this study, an osteosarcoma-assisted medical system based on NFSR-U-Net is proposed to assist doctors in classifying. Figure 1 shows the architecture of it.

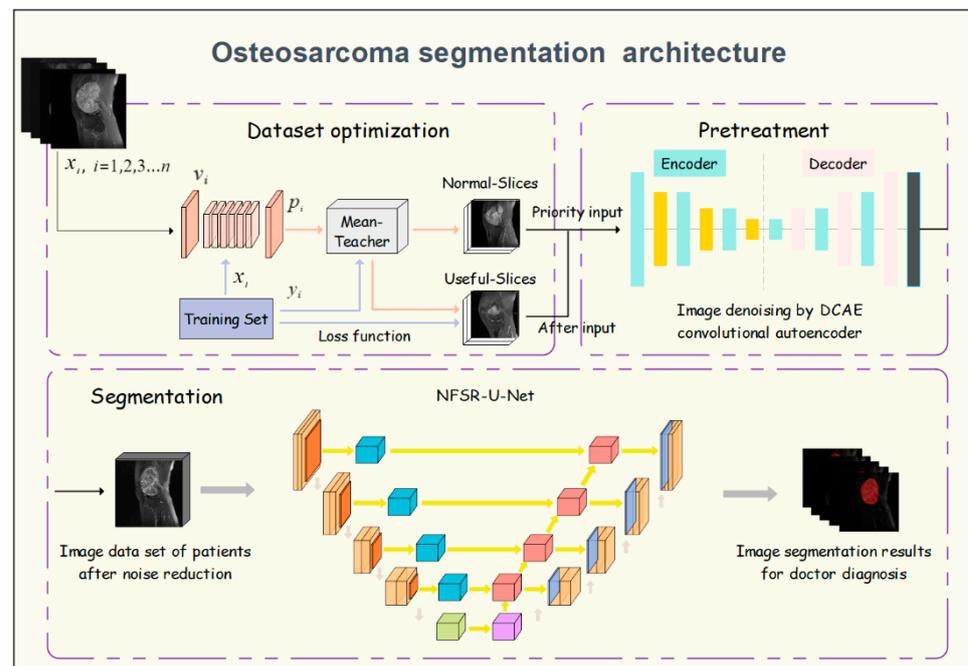


Figure 1. Overview of osteosarcoma segmentation system architecture.

This system is organized into three sections: dataset optimization, preprocessing, and image segmentation. In dataset optimization, the dataset in this thesis was collected from the Monash University Research Center for Artificial Intelligence [54]. We divide the original osteosarcoma MRI image dataset into useful slices and normal slices by the Mean Teacher algorithm and input each into the preprocessing process in turn. In preprocessing, the MRI images of patients are input to the denoising convolutional autoencoder (DCAE), and the image data are noise reduced, followed by feeding the dataset into the network. Finally, NFSR-U-Net is used for image segmentation, which helps doctors complete the subsequent diagnosis of osteosarcoma and discern the extent of tumor tissue invasion.

To further elaborate on the osteosarcoma segmentation system, we divide this chapter into three subsections. In Section 3.1, we introduce the steps for optimizing the MRI image dataset. Section 3.2 introduces the principle and structure of DCAE. Section 3.3 elaborates on the NFSR-U-Net segmentation model.

Three strategies were set up to enhance the detection effect:

- (1) Dataset optimization. We use the Mean Teacher semi-supervised learning method to optimize the original dataset by dividing the osteosarcoma image dataset into US and NS, which facilitates the training of the model.
- (2) Preprocessing. We introduce an unsupervised denoising convolutional autoencoder (DCAE), which eliminates unnecessary noise from osteosarcoma MRI images.
- (3) NFSR-U-Net. We design several modules to improve the bottleneck features and skip connections in U-Net for precise classification of tumor MRI images.

The relevant symbols involved in this paper are explained, as displayed in Table 1.

3.1. Dataset Optimization

For the original dataset of osteosarcoma MRI images, there are often images that are difficult to train. These images contain noise, and the tumor area in them may be extremely small, which makes it hard to be clearly visualized. Directly using these images as the first input dataset may cause the model to become slow and inefficient during training. Therefore, it is necessary to partition the original dataset. We used the ResNet-7 model to partition the original osteosarcoma image dataset into useful slices (US) and normal slices (NS), where US denotes the image slices that are easy to train and NS denotes the image slices that are time-consuming during the training process. As shown in Figure 2 (below),

the model is composed of 6 layers of residual modules and 1 layer of fully connected layers. Between each connected layer, we add a 3×3 max-pooling layer to minimize the scale of the feature map.

Table 1. Description of some symbols in this chapter.

Symbol	Paraphrase
D_1, D_2	Original osteosarcoma MRI image data set
T_1	MRI image labels of osteosarcoma in dataset D_1
P_{S1}, P_{S2}	Predicted probability of Student Model output
P_{T2}	Predicted probability of Teacher Model output
γ_S, γ_T	Parameter set of Student Model and Teacher Model
$E(r, q)$	Class-correlation computation
$F(\cdot)$	Local class-correlation pattern aggregation function
Z	Class-correlational representation
$P(\cdot)$	Global feature pooling operation
$ \cdot _{\text{odd}}$	The nearest odd number calculation function
$Q(\cdot)$	Neighbor feature selection function
Y_{NFS}	The output of neighbor feature selection
σ	Sigmoid activation function
ϕ_S	Spatial dimension extension function
$S(\cdot)$	Spatial attention operation
ϕ_C	Channel dimension extension function
Q_{up}	2 times upsampling bilinear interpolation operation

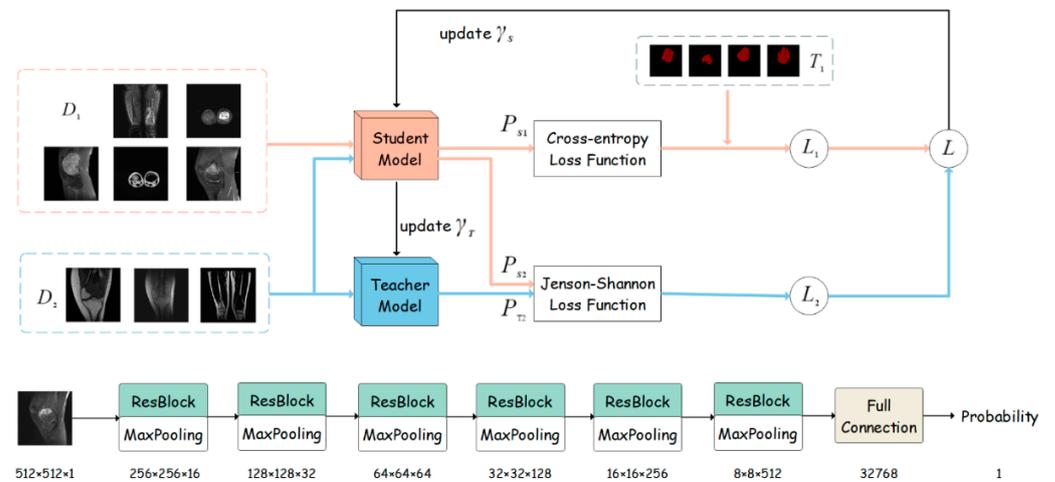


Figure 2. Schematic diagram of Mean Teacher dataset optimization method (upper) and ResNet-7 model (below). Where the Student Model and Teacher Model are both based on the ResNet-7.

In order to better adapt to the additional osteosarcoma image dataset and to enhance the robustness of the model, the Mean Teacher semi-supervised learning method [55] was used to optimize the dataset. As shown in Figure 2 (upper), the general framework of the method is divided into the Student Model and Teacher Model, which both use the ResNet-7, and the parameters of the two parts are γ_S and γ_T . We randomly split the data sets into D_1 (70%) and D_2 (30%) at the patient level, containing 2756 MRI images and 1244 MRI images, respectively, where the images in dataset D_1 contain label T_1 and dataset D_2 is unlabeled. The training procedure of the Mean Teacher is described below:

Input data sets D_1 and D_2 in the Student Model, and the prediction probability output by the module is recorded as P_{S1} and P_{S2} , respectively; then, input the data set D_2 in the Teacher Model, and the prediction probability value obtained by the module is recorded as P_{T2} ;

The loss value l_1 is calculated from the label T_1 in the data set D_1 and the predicted probability value P_{S1} is obtained in the previous step. The loss function formula L_1 is given as follows:

$$L_1 = -\frac{1}{M} \sum_{i=0}^M t_i \cdot \log(p_i) + (1 - t_i) \cdot \log(1 - p_i), \quad t_i \in T_1, p_i \in P_{S1} \tag{1}$$

The loss value l_2 is calculated from P_{S2} , P_{T2} and loss function L_2 (as illustrated in Equation (4)).

Student Model generates the loss which is recorded as $l_{total} = l_1 + l_2$. According to the calculated loss value l_{total} , the gradient descent is performed to update the parameter γ'_S of the model. Teacher Model updates parameter γ'_T through Formula (2) moving average, where α denotes the exponential moving average (EMA) decay:

$$\gamma'_T = \alpha \gamma_T + (1 - \alpha) \gamma'_S \tag{2}$$

Among them, L_1 is the cross-entropy loss function. The loss function L_2 uses the Kullback–Leibler divergence (KL) calculation formula [56], which can be applied to estimate the level of overlap between two distributions as a measure of the difference. A lower KL divergence represents a greater degree of overlap, and the KL is calculated as follows:

$$KL(N||M) = \sum m(x) \cdot \log\left(\frac{m(x)}{n(x)}\right) \tag{3}$$

However, KL divergence suffers from the disadvantage of asymmetry. Although the experiments try to align the predicted distributions of the two models, it is difficult to know which side has more accurate prediction results. Hence, the Jensen–Shannon (JS) method is used to solve the asymmetry of KL divergence. The loss function is calculated as follows.

$$L_2 = \frac{1}{2} KL(P_{S2}||P_{T2}) + \frac{1}{2} KL(P_{T2}||P_{S2}) \tag{4}$$

After the above steps, the original dataset was segmented into US (52.7%) and NS (47.3%), where US included 2108 osteosarcoma MRI images and NS included 1892. Finally, we input the relatively simple dataset sample US into the segmentation model first, and the NS second. Relevant studies show that it is more beneficial to the training of the model. Moreover, the datasets are automatically divided by computer, which reduces the workload of manual image screening by doctors and improves the efficiency of osteosarcoma detection.

3.2. Pretreatment

MRI medical imaging usually contains a lot of useless noise information from the background of the task target, which will significantly affect the effect of osteosarcoma segmentation. Therefore, in this study, we introduced a denoising convolutional autoencoder (DCAE) to denoise the MRI image data of osteosarcoma. Compared with the traditional automatic encoder, the convolution automatic encoder [57] is more suitable for image processing because its convolution ability can make full use of the image structure. Additionally, its weight is shared among various inputs, thus maintaining a high local spatiality, which also outperforms the NL mean and median filter according to the validation in the literature [58].

DCAE uses a relatively simple architecture, which is illustrated in Figure 3. It is composed of an encoder and a decoder. The encoder is made up of a convolution layer and a max-pooling layer. The convolution layer extracts features from the osteosarcoma image through convolution operation. After each convolution layer, there will be a ReLU activation and batch normalization layer. The max-pooling layer performs 2 times downsampling to filter useless information, thus eliminating unnecessary noise and other artifacts in the tumor MRI image. Then, the decoder expands the feature map and enlarges the important feature information through deconvolution and 2 times upsampling, so as to restore the image. Therefore, DCAE can effectively reduce the noise of the input image \tilde{x} with noise.

Compared with the original MRI image of osteosarcoma, the noise of the output image x is greatly reduced and has more useful feature information, which is helpful for more accurate segmentation of osteosarcoma in the following steps. At the same time, the MRI images of osteosarcoma after noise reduction can be served as a reference for clinical diagnosis by doctors and provide the necessary conditions for doctors to carry out further diagnosis and treatment for patients.

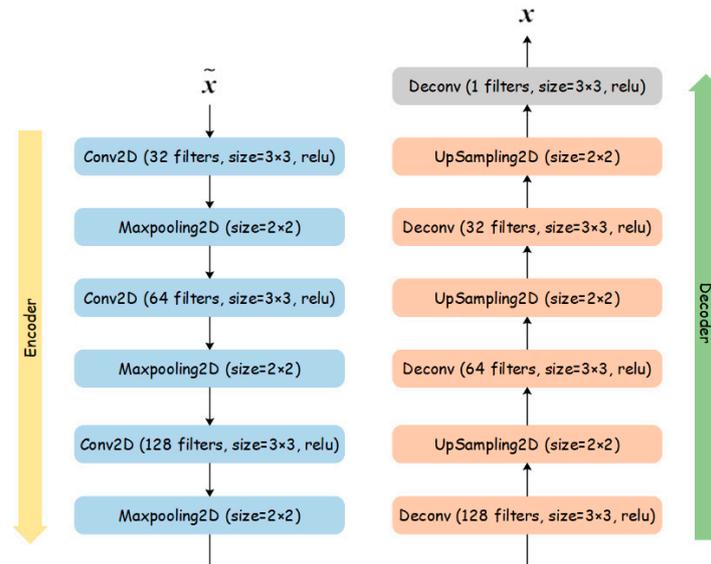


Figure 3. Architecture of the proposed DCAE.

3.3. Osteosarcoma MRI Image Segmentation

Segmenting the diseased tissue from the surrounding normal tissue has been a major challenge in medical image segmentation. In order to identify the osteosarcoma lesion region more accurately, we used an MRI image segmentation framework (NFSR-U-Net) based on class-correlation pattern aggregation, as shown in Figure 4. Compared with the traditional U-Net structure, NFSR-U-Net achieves advanced feature filtering and effective fusion of spatial features in osteosarcoma MRI images by adding the neighbor feature selection module to optimize the representation of bottleneck features between contracting and expansive paths and by adding skip connection multi-level semantic feature residual fusion, which helps the network perform fine-grained osteosarcoma MRI image segmentation. The NFSR-U-Net model comprises four main structures:

Encoder: The contracting path is composed of four contracting blocks, each containing two convolutional layers, where the convolutional layer comprises a 3×3 convolution, a batch normalization (BN), and a rectified linear unit (ReLU). Each contracting block is followed by a downsampling via 2×2 max-pooling with a stride of 2.

Decoder: The expansive path includes four expansive blocks, each of which uses 2 times upsampling to produce the extended feature map. The extended feature map is spliced with the high-resolution feature map of the same layer of the contracting path.

Bottleneck layer: The neighbor feature selection module (NFS) is designed to generate the class-correlational representation by analyzing the local correlation patterns in MRI images and then identifying the relationship between similar class features to correlate the dense representation of image local features. NFS enables the embedding of similar classes to achieve a high fit by learning the intra-class similarity in osteosarcoma images, which exhibit high accuracy in tumor tissue segmentation with large shape differences.

Skip connections: The spatial attention module (SAM) and spatial feature residual connection module (SFRC) are designed to optimize skip connections. SAM uses 1×1 convolution to compress the channel dimension to 1, followed by a sigmoid activation function. SFRC uses the residual connection method to upsample the feature map after

compressing the channel with 1×1 convolution for the lower-level features, and the result is summed with the SAM-processed feature matrix.

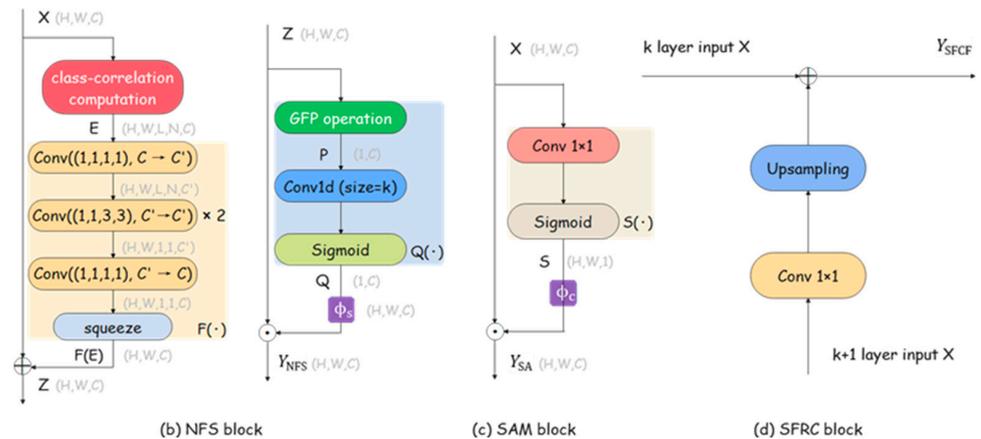
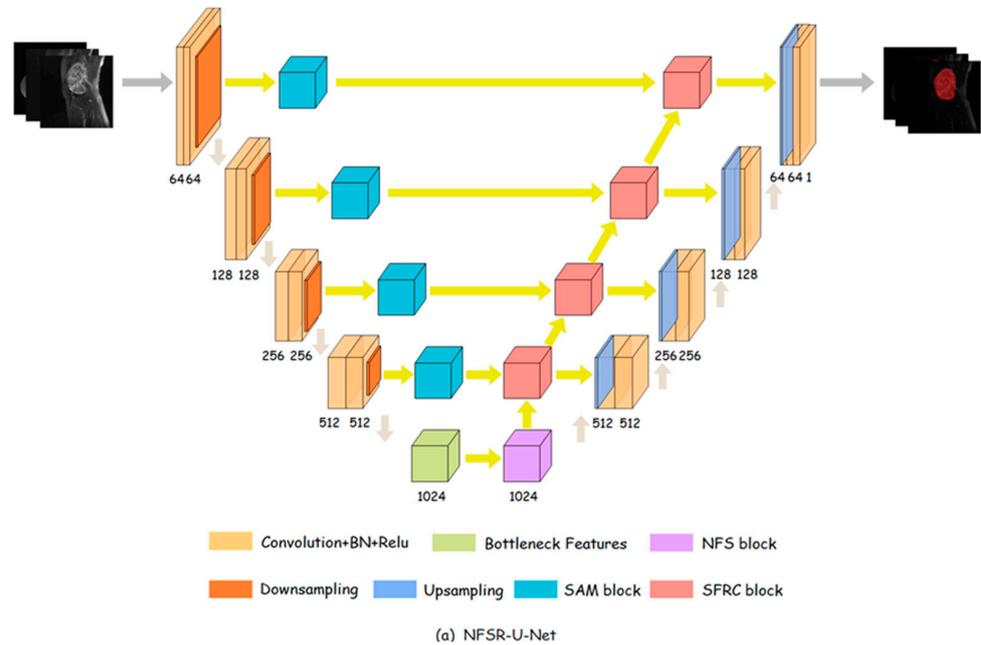


Figure 4. The general framework of NFSR-U-Net and its key components ((a) NFSR-U-Net architecture; (b) NPS block; (c) SAM block; (d) SFRC block). The number below each layer in (a) indicates the number of output channels.

3.3.1. Neighbor Feature Selection Module

Medical image pattern recognition and analysis can learn data-driven, hierarchical semantic features from sufficient image data, however, the distinctive differences and variations of features within different classes in MRI images can easily make embeddings of similar classes far apart. In contrast, class-correlational representation can provide semantic information about similar classes and learn generalizable relational embeddings. In this paper, for osteosarcoma MRI images with high heterogeneity, we introduce the neighborhood feature selection module, as shown in Figure 4b. This module first aggregates class-correlation patterns by measuring the similarity of nearby patches to form the class-correlational representation. Then, it analyzes the correlational representations of features between similar classes so that the embeddings of the same class are as close as possible.

Class-correlation computation. For a basic representation of the input $X \in R^{H \times W \times C}$, this part obtains the class-correlation tensor $E \in R^{H \times W \times L \times N \times C}$ by computing the Hadamard product of each position $r \in [1, H] \times [1, W]$.

$$E(r, q) = \frac{X(r)}{\|X(r)\|} \cdot \frac{X(r+q)}{\|X(r+q)\|} \tag{5}$$

Class-correlational representation. Then, we design a convolutional structure $F(\cdot)$ for the analysis and learning of class-correlation patterns. As shown in Figure 4b, $F(\cdot)$ consists of several 4D convolutions. One convolution reduces the number of channels of the class-correlation tensor E to C' , two 3×3 convolutions transform the tensor, and one convolution changes its channels to C and reduces the spatial dimension to 1×1 . Between convolutions, there are ReLU and BN layers. Finally, use a squeeze to generate $F(E) \in R^{H \times W \times C}$. By gradually aggregating the local class-correlation patterns using $F(\cdot)$, we calculate the class-correlational representation $Z \in R^{H \times W \times C}$ as follows.

$$Z = F(E) + X \tag{6}$$

Global feature pooling operation. We improve the global average pooling (GAP) operation in SENet [53] and propose the global feature pooling (GFP) operation to compress the feature map. GFP takes a mixture of local features in class-correlational representation Z to identify the relationship of dense features between similar classes, enabling a higher fit of pixel-level embeddings of similar classes. Meanwhile, GFP improves the expressiveness of the network by incorporating the spatial features of the osteosarcoma MRI images while compressing feature maps.

$$P(Z) = Z \cdot \left(\text{softmax} \left(Z^T W_N^T \right) \right) \tag{7}$$

where $P(Z) \in R^{C \times 1}$ represents the global feature pooling operation with input $Z \in R^{C \times HW}$. $W_N^T \in R^{C \times 1}$ indicates the transpose of the 1×1 convolution matrix.

Neighbor feature selection (NFS). As shown in $Q(\cdot)$, NFS uses GFP operation in compressing the feature maps of osteosarcoma MRI images. It identifies semantic features in the discrete feature space of the tumor image by associating n neighboring feature channels in the feature map and searches for similar class features for local matching to more closely associate the dense representation of local features of the image for the joint classification task. NFS facilitates more subtle feature representation while capturing cross-channel interactions, which is suitable for processing osteosarcoma MRI images with heterogeneous distribution of tumor lesion tissue. The NFS is implemented by 1D convolution with a kernel of size k , which represents the coverage of channel interactions in this group. To avoid manual adjustment of n , the formulation given by ECA-Net [53] is used in this paper as follows:

$$n = k = \psi(C) = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \tag{8}$$

where $\lfloor \cdot \rfloor_{\text{odd}}$ indicates the nearest odd number, C denotes the number of feature channels. The values of γ and b are 2 and 1, respectively. NFS is calculated as:

$$\begin{aligned} Q(Z, k) &= \sigma(C_k^{1D}(P(Z))) \\ Y_{\text{NFS}}(Z, k) &= \phi_S(Q(Z, k)) \cdot Z \end{aligned} \tag{9}$$

where $Q(Z, k) \in R^{C \times 1}$ indicates the neighbor feature selection operation with input $Z \in R^{C \times HW}$ and convolution kernel size k . C_k^{1D} denotes the 1D convolution with kernel size k . ϕ_S is a spatial dimension expansion function to expand the spatial dimension of $Q(Z, k)$ consistent with the input Z . Output $Y_{\text{NFS}}(Z, k) \in R^{C \times HW}$.

3.3.2. Spatial Attention Module

Medical image processing and analysis can learn highly representative and hierarchical semantic features in MRI images, but it remains a potential challenge to ignore useless features to learn generalizable embeddings. The spatial attention mechanism aims to recalibrate the spatial location importance of feature maps and ignore relatively irrelevant locations, which has a positive impact on fine-grained tumor image segmentation. To extract highly representative features in the spatial axis dimension of MRI images and improve the characterization ability of the model, we introduced the spatial attention module (SAM). As shown in Figure 4c, SAM uses 1×1 convolution to compress the channel dimension to 1 to ensure the spatial dimensional feature consistency. SAM is calculated as:

$$\begin{aligned} S(X) &= \sigma(W_S X) \\ Y_{SA}(X) &= \phi_C(S(X)) \cdot X \end{aligned} \tag{10}$$

where $S(X) \in R^{1 \times HW}$ denotes spatial attention operation. $W_S \in R^{1 \times C}$ is a 1×1 convolution matrix for compressing the channel dimension to 1. The output of SAM is $Y_{SA}(X) \in R^{C \times HW}$, and ϕ_C is the channel dimension extension function by which the channel dimension can be extended to C .

3.3.3. Spatial Feature Residual Connection Module

To retrieve the spatial information of osteosarcoma MRI images lost by pooling and reduce the semantic difference between contracting path and expansive path features, the spatial feature residual connection module (SFRC) is introduced. SFRC effectively enhances the statistical information of texture and boundary in osteosarcoma MRI images by extracting effective information in different spatial dimensions and introducing advanced features in the middle and late stages of early skip connections to learn the positional relationships between pixels. As shown in Figure 4d, SFRC receives the upper feature (layer k) and lower feature (layer $k+1$) of the model as input, and k is an integer in [1,4]. The calculation formula for SFRC is:

$$Y_{SFRC}(X_k, X_{k+1}) = X_k + Q_{up}(W_F X_{k+1}) \tag{11}$$

where $X_k \in R^{C \times HW}$ is the output feature from the k -layer contracting path, $X_{k+1} \in R^{2C \times \overline{HW}}$ is the output feature from SFRC of layer $k+1$, and X_{k+1} is the output of NFS only when $k = 4$. The \overline{HW} indicates that the H and W sizes are halved respectively. $W_F \in R^{C \times 2C}$ is a 1×1 convolution operation for the purpose of compressing the channel number from $2C$ to C . Q_{up} is a 2 times upsampling bilinear interpolation operation, by which the feature map of the image can be expanded so that the output Y_{SFRC} is consistent with X_k in the feature dimension.

3.3.4. Loss Function

To address the positive and negative sample imbalance in the osteosarcoma MRI image dataset, we use a combination of binary cross-entropy (L_{BCE}) and dice loss (L_{Dice}) weighting in this paper after analyzing the stability and accuracy of training. Since L_{Dice} may cause drastic gradient changes and thus affect the back-propagation, leading to difficulties in training the osteosarcoma segmentation model, we will reduce the weight of L_{Dice} appropriately. The loss function is calculated as:

$$\begin{aligned} L(y, \hat{y}) &= \omega L_{Dice}(y, \hat{y}) + (1 - \omega)L_{BCE}(y, \hat{y}) \\ L_{BCE}(y, \hat{y}) &= -(y \log \hat{y} + (1 - y) \log(1 - \hat{y})) \\ L_{Dice}(y, \hat{y}) &= 1 - \frac{2|y \cap \hat{y}| + \varepsilon}{|y| + |\hat{y}| + \varepsilon} \end{aligned} \tag{12}$$

where y denotes the value of the true osteosarcoma segmentation map, \hat{y} is the value of the osteosarcoma segmentation map predicted by the model, ω is the weight of the two losses, and ε is the smoothing term set to avoid a denominator of 0. In the course of this experiment, the loss function L is tested in general, and the smoothing term ε of L_{Dice} is

modified in this paper. We first set the weight parameter ω to 0.05, then progressively increase the value of ω to increase the loss reward for fewer samples by adjusting the contribution of the class weights to loss, and make the overall value of the two types of loss functions reach an equilibrium minimum in each epoch. In the end, we adjusted the smoothing term ε to achieve a relatively smaller value of loss at each epoch while refraining from a denominator of 0. After the experiments, we set the loss weight parameter ω to 0.3 and the smoothing term ε to 1.0.

The above is the general architecture of NFSR-U-Net. Before image segmentation, we preprocess the dataset with strategies such as dataset optimization and noise reduction. By means of automatic computer screening and processing of MRI images, the efficiency of osteosarcoma detection by doctors is improved. During segmentation, we designed the NFS, SAM, and SFRC modules in NFSR-U-Net to optimize the bottleneck layer and the skip connections of the network, realizing a finer-grained tumor lesion tissue segmentation of osteosarcoma MRI images. After segmentation, the diagnostic results given by the system provide an auxiliary basis for doctors to help treat patients. While ensuring segmentation accuracy, the network has fewer parameters, which facilitates model training.

4. Results

4.1. Data Set

The datasets in the paper were acquired from the Monash University Research Center for Artificial Intelligence [54]. We acquired over 4000 osteosarcoma MRI images and other metric data from 204 patients with osteosarcoma disease. To prevent leakage of patient information between training and testing and to avoid the risk of matching the model to patients, the dataset is divided at the patient level in this paper. Since the volume of MRI images collected from a single patient was not equal, we divided the 204 patients according to the number 156:48, which is roughly an 8:2 ratio. The delineated image dataset was used as the training set and the test set, containing 3108 MRI images versus 892 MRI images, respectively. This trial studied 204 patients involved, and specific information on their data is listed in Table 2. In the process of data splitting, this thesis conducts a 5-fold cross-validation on the original image dataset. We randomly slice the dataset into 5 disjoint subsets of the same size, pick four subsets each time as the training set, and the remaining one as the test set, and repeat the process 5 times. Finally, the test errors generated were averaged for cross-validation to estimate the robustness and repeatability of the model.

Table 2. Baseline of patient characteristics.

Characteristics		Total	Training Set	Test Set
Age	<15	48 (23.5%)	38 (23.2%)	10 (25%)
	15~25	131 (64.2%)	107 (65.2%)	24 (60.0%)
	>25	25 (12.3%)	19 (11.6%)	6 (15.0%)
Sex	Female	92 (45.1%)	69 (42.1%)	23 (57.5%)
	Male	112 (54.9%)	95 (57.9%)	17 (42.5%)
Marital status	Married	32 (15.7%)	19 (11.6%)	13 (32.5%)
	Unmarried	172 (84.3%)	145 (88.4%)	27 (67.5%)
Surgery	Yes	181 (88.8%)	146 (89.0%)	35 (87.5%)
	No	23 (11.2%)	18 (11.0%)	5 (12.5%)
SES	Low SES	78 (38.2%)	66 (40.2%)	12 (30.0%)
	High SES	126 (61.8%)	98 (59.8%)	28 (70.0%)
Grade	Low grade	41 (20.1%)	15 (9.1%)	26 (65%)
	High grade	163 (79.9%)	149 (90.9%)	14 (35%)
Location	Axial	29 (14.2%)	21 (12.8%)	8 (20%)
	Extremity	138 (67.7%)	109 (66.5%)	29 (72.5%)
	Other	37 (18.1%)	34 (20.7%)	3 (7.5%)

4.2. Evaluation Index

For the purposes of evaluating the segmentation effect of the model, we chose the widely used evaluation metrics of accuracy (Acc), precision (Pre), recall (Re), F1-score (F1), intersection of union (IOU), and dice similarity coefficient (DSC) for model performance evaluation.

In this paper, the Params metric represents the number of parameters of the segmentation model, which indicates the size of the storage space occupied by the model. The floating-point operations (FLOPs) metric reflects the computational complexity of the segmentation network. In this osteosarcoma MRI segmentation experiment, we circumvent the occurrence of missed diagnoses by improving the recall evaluation index.

4.3. Comparison Algorithm

For the sake of evaluating the performance and complexity of the NFSR-U-Net, several MRI image segmentation models were selected for analysis and these models were compared with the NFSR-U-Net for the segmentation effect of osteosarcoma MRI images. These segmentation models consist of U-Net [48], FCN [59], FPN [60], PSPNet [61], MSRN [62], MSFCN [32], and AIMSost [63].

4.4. Training Strategy

Before formally training the NFSR-U-Net tumor MRI image segmentation model, we performed noise reduction on the original dataset images to avoid the model from over-focusing on invalid features and effectively enhance the robustness of the model. At the same time, we expanded the osteosarcoma MRI image dataset by rotating and expanding (or reducing) the images, and the dataset was increased by nearly 20%. During the training of the NFSR-U-Net network, we set the learning rate to 0.01 and trained a total of 800 epochs. The segmentation model set Adam as the optimizer and used PolynomialDecay to dynamically adjust the learning rate of each epoch.

4.5. Segmentation Effect Evaluation

Before training the segmentation model, we divided the dataset into US (52.7%) and NS (47.3%), where US included 2108 osteosarcoma MRI images and NS included 1892. As shown in Figure 5 (left), it is a useful-slice dataset, in which the tumor location in US is very obvious, and the boundary between the tumor region and the surrounding tissues is also clearer. These advantages speed up the convergence of the model during training, improve the image segmentation accuracy of the model, and reduce the training burden significantly, so it is suitable as the priority input training set. In Figure 5 (right), this image is a normal-slices dataset, and the boundary between the tumor region and the surrounding tissues is blurred. If it is used as the first input dataset, the model will become inefficient during training, so it should be used as the later input dataset.

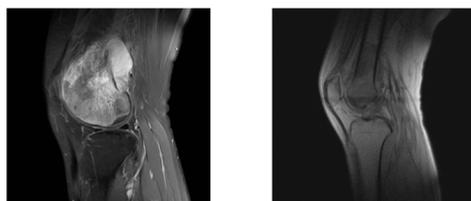


Figure 5. Useful-Slices (US) (left) and Normal-Slices (NS) (right).

The convergence and performance of the Mean Teacher model depend on the parameter settings of the model. Therefore, in this experiment, we selected two parameter variables, train data size and α , where α is the decay of EMA, as is displayed in Table 3. We estimate the performance of the mean teacher model by adjusting the values of these two parameters under the conditions of a fixed epoch of 500 and a batch size of 32. The results show that the mean teacher algorithm achieves the best enhancement of the segmentation

model with the parameter setting of 70% of the train data size and α of 0.99, and the dice coefficient reaches the highest value. This may be attributed to the fact that the EMA strategy can effectively prevent the model from overfitting, especially when the larger-scale network parameters are learned from the limited training data.

Table 3. The influence of the parameter settings of the Mean Teacher on the segmentation results estimated by dice coefficient.

Method	Parameters		Metrics
	Train Data Size	α	Dice (%)
Mean Teacher	2756	0.98	91.89
	2756	0.99	92.08
	3108	0.98	91.72
	3108	0.99	91.74

In order to study the influence of the data set optimization process on the results, this paper compares the segmentation effect of the data optimization process after the Mean Teacher algorithm with the unoptimized process in the training phase. Observing Figure 6, the image shows the process of change in Acc metrics formed by the two cases of data optimization and non-optimization, which are noted as the standard curve and the under-optimization curve, respectively (800 training epochs were completed, and 50 epochs were randomly selected). It can be concluded that the input of US and NS optimized by the Mean Teacher algorithm as the combined order effectively enhances the training efficiency of the model, which enables the convergence speed of NFSR-U-Net during training to be accelerated, and the peak of Acc metrics is reached at the 10th epoch. Then, this paper further analyzes the differences in the robustness of the model caused by the heterogeneity of image data. Based on the inherent heterogeneity of osteosarcoma, patients may produce MRI images that are difficult to identify during detection, such as containing edematous areas, cartilage in the interstitial space, and may lack classical tumor markers. In this study, we added 8 such images every 1 epoch to the standard curve, for a total of 400 images. The course of this curve during the training phase of the model is shown in Figure 6 and is noted as a heterogeneity curve. The results show that under the condition that the training time is long and does not affect the practical use, although the Acc of the curve decreases slightly during the training, the fluctuations of the metric are smoother, and it is closer to the standard curve with enough iterations, which indicates that the robustness and fairness of the model are enhanced to some extent by the strategy of increasing the heterogeneity of the data.

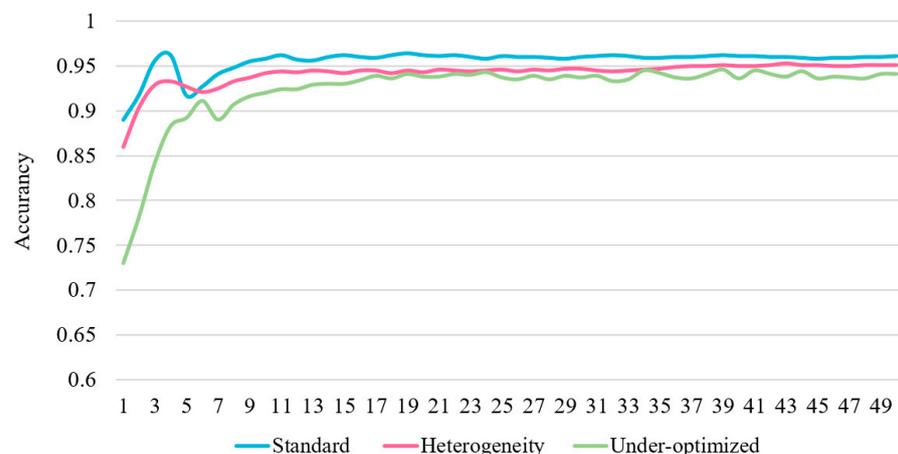


Figure 6. Comparison of Acc metrics for standard, under-optimized, and heterogeneous curves in the model training phase.

Immediately after that, we input the MRI image datasets US and NS of osteosarcoma, which were divided by the Mean Teacher algorithm, as a combination, into the NFSR-U-Net model sequentially, while at the same time, the initial image dataset NS was input into the model as another separate group for comparison of test results. Observing Figure 7, the image demonstrates the difference in the evaluation metric scores formed by the different combinations of US+NS and NS only in the testing phase of the model. It can be concluded that the NFSR-U-Net model shows better segmentation performance in the testing phase after the sequential input of the US+NS combination optimized by the Mean Teacher approach to the segmentation network, in which the scores of all metrics are improved. On the contrary, with the under-optimized NS image set as the input to the segmentation model, the NFSR-U-Net showed a weakness in the relevant metrics in the test phase, and the scores of each metric decreased by 0.10 ± 0.04 points.

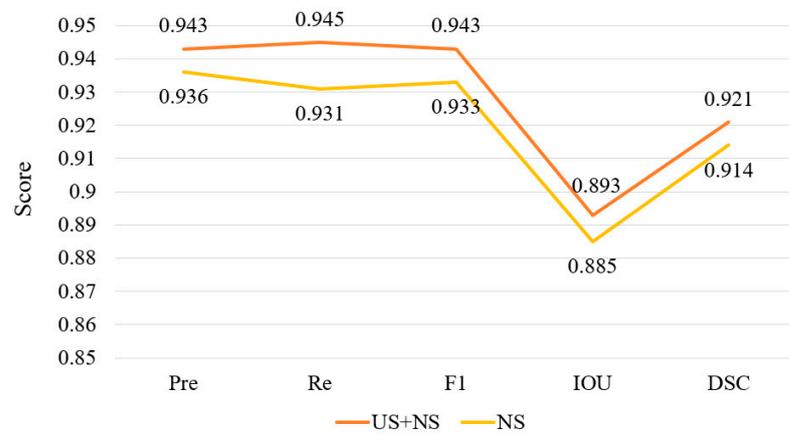


Figure 7. The scores of the concerned evaluation metrics of the NFSR-U-Net imposed by the input of a combination of US and NS, and the input of NS only.

After dataset optimization, we performed a data preprocessing process for image noise reduction using DCAE on MRI image data. Figure 8 shows the results of model segmentation with and without dataset preprocessing. Among them, Figure 8a indicates the original label of the osteosarcoma image, and Figure 8b shows the MRI image segmentation result of the network without data preprocessing. Figure 8c shows the MRI image segmentation result with dataset preprocessing. It is observed that the tumor segmentation region in the middle column is not complete. In contrast, the tumor region obtained from the segmentation after dataset optimization is more accurate, which indicates that preprocessing effectively enhances the quality of MRI images and the segmentation effect is substantially improved.

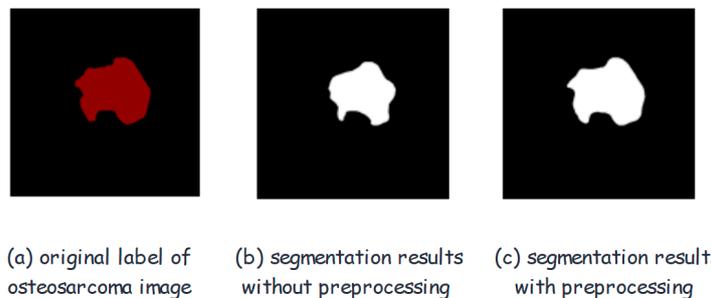


Figure 8. Comparison of MRI image segmentation results of NFSR-U-Net with and without dataset preprocessing process. (a) original label of osteosarcoma image; (b) segmentation results without preprocessing; (c) segmentation results with preprocessing.

With regard to the denoising process of osteosarcoma MRI images, a comparative analysis of DCAE with two architectures [6,7] based on different deep learning algorithms was performed. Combined with the cross-entropy and MSE loss functions, we evaluated the performance of different architectures for the denoising task of osteosarcoma MRI scans, as presented in Equation (13).

$$\begin{aligned}
 MSE &= \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \\
 L_{CE}(p, q) &= - \sum_{i=1}^n p(x_i) \log[q(y_i)] \\
 L_{to} &= wL_{CE} + (1 - w)MSE
 \end{aligned}
 \tag{13}$$

The mean accuracy values for the two-parameter settings are displayed in Table 4. The table reveals that the DCAE architecture still exhibits the best denoising performance with three various sets of weights w . DCAE filters the useless information in the noisy osteosarcoma MRI scans, thus eliminating unnecessary noise and other artifacts in the image and enhancing the image’s peak signal-to-noise ratio (PSNR).

Table 4. The mean accuracy exhibited by each architecture with various weighting parameters.

Weight Setting	Architecture		
	U-Net with 2 Encoder-Decoder Pairs	ACDN	DCAE
(0.4, 0.6)	0.954	0.955	0.962
(0.5, 0.5)	0.961	0.957	0.967
(0.7, 0.3)	0.958	0.962	0.965

In this experiment, we conducted a comparative analysis of the segmentation effect of each model for osteosarcoma MRI images, as shown in Figure 9, in which the first and second columns are the labels of the original image, label, respectively, while the later columns are the effect maps of each model for osteosarcoma image segmentation. It can be clearly seen that for some tumor images with complex shapes, the MSFCN, MSRN, and FCN16 models show the problem of over-dividing the tumor regions, and among them, the MSFCN model shows the problem that the tumor regions are more blurred and not well differentiated from the surrounding tissues. In contrast, our proposed NFSR-U-Net model delineates the detailed part of the osteosarcoma region well and effectively distinguishes the tumor region from surrounding tissues, which means it achieves accurate segmentation of tumor images with complex shapes.

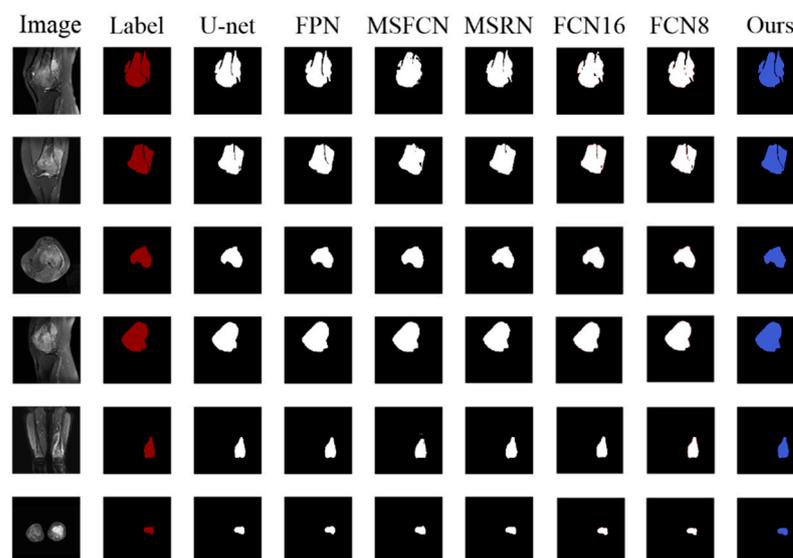


Figure 9. Comparison of image segmentation effects of different models.

For the sake of evaluating the performance of segmentation models more clearly, we used various evaluation indexes to quantitatively compare the segmentation results derived from each model. Table 5 shows a detailed comparison of the evaluation index values of different segmentation models. From this table, it is obvious that our proposed NFSR-U-Net has higher values of Pre, Re, IOU, and other evaluation indexes than those of other models, which indicates its good performance in MRI image segmentation. To further analyze the properties of the NFSR-U-Net model on MRI image segmentation of osteosarcoma, we compared it with the existing work of other researchers on this dataset. Compared with the AIMSost model, NFSR-U-Net improved 0.014 ± 0.003 points in Pre, Re, F1, and IOU metrics, but its Params increased by only 9.5 M. This indicates that although the model has a high accuracy of osteosarcoma segmentation, its parameters and computational complexity do not increase much, which is conducive to cost saving by ensuring that doctors obtain highly accurate segmentation results while the performance requirements of primary hospitals equipped with hardware equipment are low.

Table 5. Comparison of evaluation indexes of different MRI image segmentation models for osteosarcoma.

Model	Pre	Re	F1	IOU	DSC	Params	FLOPs
FCN-16s	0.922	0.882	0.9	0.824	0.859	134.3 M	190.35 G
FCN-8s	0.941	0.873	0.901	0.83	0.876	134.3 M	190.08 G
PSPNet	0.856	0.888	0.872	0.772	0.87	46.70 M	101.55 G
MSFCN	0.881	0.936	0.906	0.841	0.874	23.38 M	1524.34 G
MSRN	0.893	0.945	0.918	0.853	0.887	14.27 M	1461.23 G
FPN	0.914	0.924	0.919	0.852	0.888	48.20 M	141.45 G
U-net	0.922	0.924	0.923	0.867	0.892	17.26 M	160.16 G
AIMSost	0.928	0.931	0.926	0.882	0.912	15.91 M	171.72 G
Ours	0.943	0.945	0.943	0.893	0.921	25.41 M	204.01 G

We compared the values of Params metrics and DSC metrics of different segmentation models, which are displayed in Figure 10. It is observed that the NFSR-U-Net has a great improvement in accuracy compared with other segmentation models, and the value of DSC reaches 0.921, which is nearly 3 percentage points higher than the second-place U-Net model. Meanwhile, the model still maintains a low parameter size with only 25.41 M Params, which is much lower than the number of parameters of models such as FCN-16s and FCN-8s, indicating that our proposed MRI image segmentation system for osteosarcoma occupies few resources and the model is trained faster, which is beneficial to the practical application of the system.

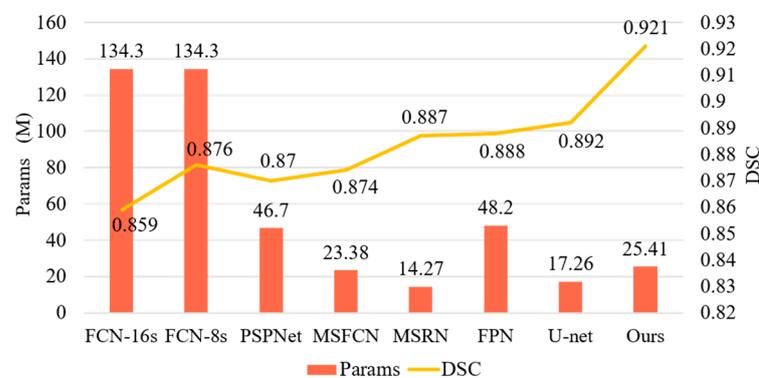


Figure 10. Params and DSC of each model.

Figure 11 shows the values of FLOPs and DSC metrics for different segmentation models. It is observed that the NFSR-U-Net has the highest DSC values, and the segmentation results are all better than the segmentation models such as U-Net, FCN-16s, MSRN, and PSPNet. Moreover, it ensures high segmentation accuracy of osteosarcoma

MRI images while maintaining the FLOPs at a low value of 204.01 G, which reflects the low computational complexity of it. Compared with the MSFCN and MSRN, NFSR-U-Net has a lower computational cost and is more conducive to the training and application of the model. Observing the AIMSost model in the figure, it has a lower FLOPs index of 171.72 G. It is attributed to its reconstruction of the component attention condenser in AttendSeg, and the utilization of such a component reduces the computational complexity of the model to a large extent. However, the NFSR-U-Net model performs better on the DSC metric, almost 0.01 points higher than the AIMSost model, but the computational complexity remains at a low level, which is of importance for hospitals that are required to deliver highly accurate osteosarcoma diagnoses.

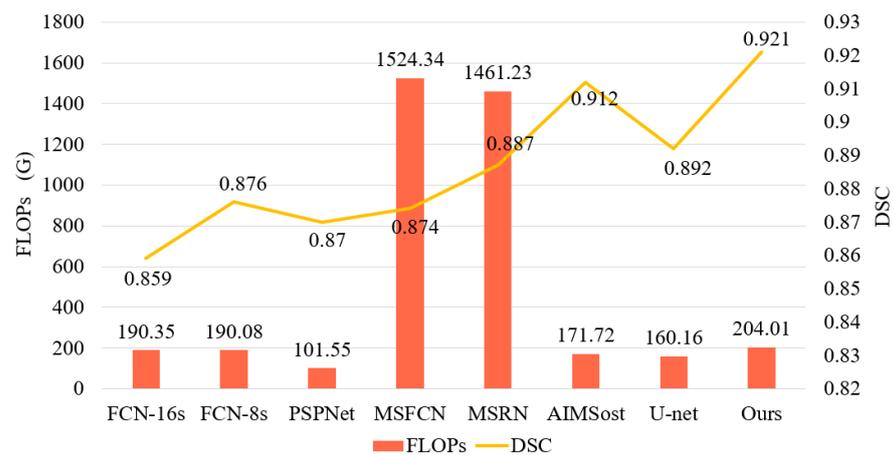


Figure 11. FLOPs and DSC of each model.

Figure 12 shows how the image segmentation accuracy of each model changes as the number of training rounds increases. A total of 800 training epochs were completed, and 50 epochs (randomly select 1 epoch out of every 16 epochs) were randomly selected for the comparative analysis of the segmentation accuracy values of different models. From the line graph, it is observed that after training nearly 160 epochs, the training effect of each segmentation model reaches a more stable state, in which the segmentation accuracy value of NFSR-U-Net is close to 96%, higher than that of FPN, MSRN, and other models, and its accuracy value is very stable. The accuracy ranking of each segmentation model is: NFSR-U-Net (ours) > U-Net > FPN > MSRN > MSFCN > FCN-8s > PSPNet.

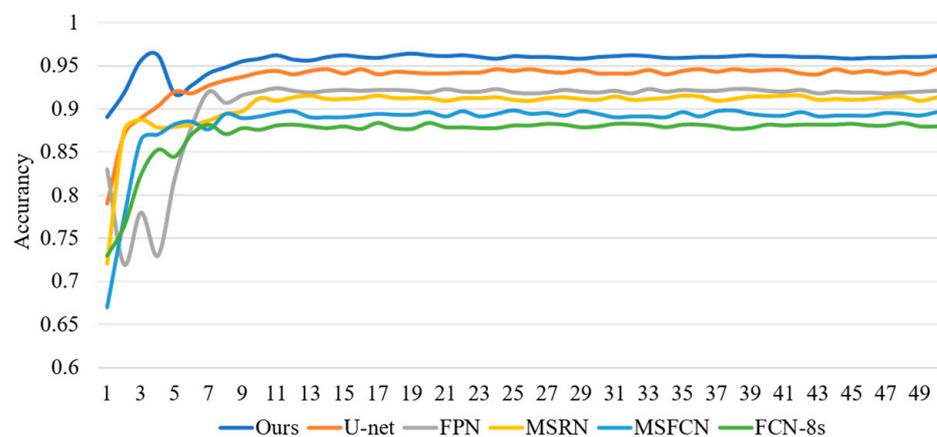


Figure 12. Accuracy changes of different models.

Additionally, recall is also a very important evaluation index when evaluating the image segmentation effect, which reflects the possibility of the occurrence of missed diagnoses. Figure 13 shows the variation in recall for each model. It can be seen that the recall values

of U-Net, FPN, and MSFCN change sharply in the early training period but stabilize in the late training period, while the recall of the MSRN model fluctuates throughout the whole process. In contrast, the Recall of NFSR-U-Net has less fluctuation and higher values after stabilization, which reflects that it is capable of effectively circumventing the occurrence of missed diagnoses and has better performance. Then, we compared the F1-score of NFSR-U-Net with other partial models. Figure 14 displays the variation of the F1-score for each model. It is observed that the F1-score of the NFSR-U-Net always maintains a high value, and the final value remains stable, which indicates that the NFSR-U-Net is very effective for osteosarcoma MRI image segmentation and has high robustness.

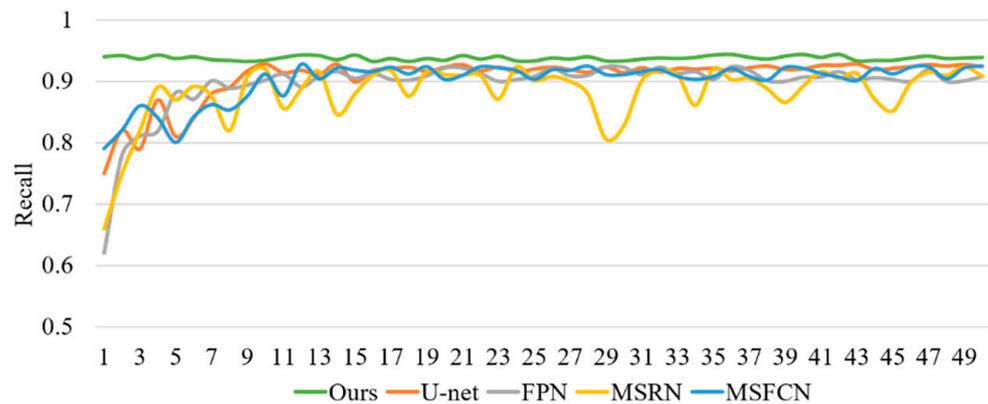


Figure 13. Recall changes of different models.

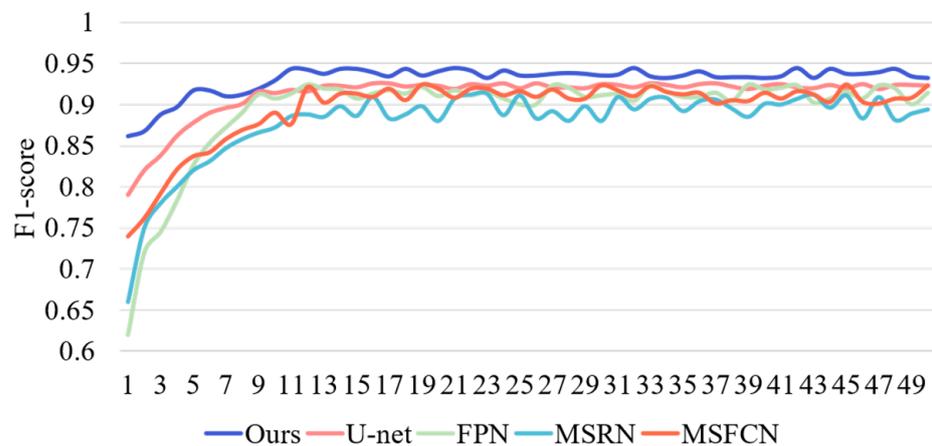


Figure 14. F1-score changes of different models.

Figure 15 plots the graph of the loss function values for individual models in epochs based on the experimental data. In this experiment, we used L2 regularization for the loss function, set the weight decay to 50, and used data enhancement techniques such as image flipping and scaling to reduce the possibility of model overfitting. It can be observed that the loss values of individual models decreased numerically in general and eventually reached a smooth lower loss value, which indicates that the NFSR-U-Net and the other models underwent normal convergence. The initial loss values of the loss functions of MSRN and MSFCN are relatively large because the initial parameters of the models are set separately, and the initial loss values might also differ. In addition, the MSRN model has more dramatic fluctuations in the loss values compared to the other models, which may be a result of differences in the generalization ability, complexity, learning rate, or probability problems of individual models, leading to the models falling into local optima.

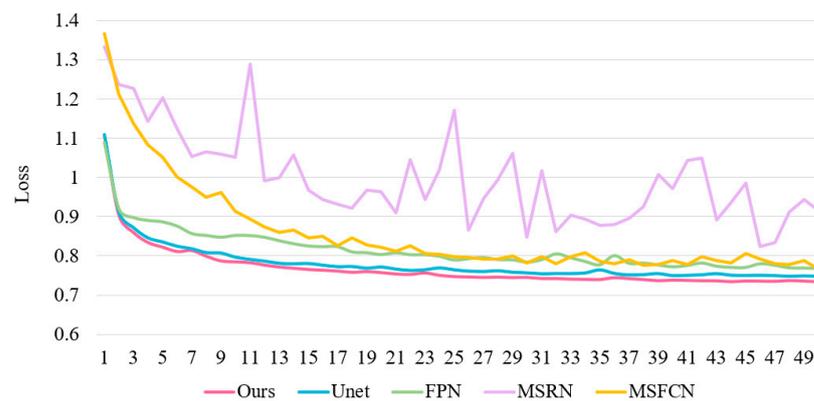


Figure 15. The variation of loss function values for individual models in epochs.

5. Discussion

In tumor MRI image segmentation studies, low contrast and edge blurring often tend to affect the segmentation performance of the network. It is an attractive research direction for the current tumor segmentation domain to precisely segment tumor tissues with variable shapes and locations.

From the experimental results, we observe that the data optimization and preprocessing process on the original tumor MRI image dataset is an effective way to improve the model training efficiency and tumor MRI image segmentation accuracy. In this way, the convergence speed of the model during training was accelerated by dividing the MRI image dataset into US and NS using the Mean Teacher algorithm. In addition, the process of preprocessing using DCAE removes artifacts and noise from MRI images, which effectively enhances the accuracy of tumor MRI image segmentation.

Meanwhile, the analysis of evaluation indexes such as DSC, IOU, and Pre indicates that the NFSR-U-Net proposed in this paper has an excellent segmentation effect on tumor tissues with variable shapes and locations. From Figure 9 and Table 5, it can be observed that NFSR-U-Net obtained the highest values for the Pre, IOU, and DSC metrics. Among them, the Pre of NFSR-U-Net reached 0.943, which exceeded 0.941 of the FCN-8s and 0.922 of the FCN-16s. Moreover, its DSC reached 0.921, which surpassed 0.892 of the U-Net and 0.888 of the FPN. Specifically, FCN-8s and FCN-16s are insensitive to MRI image details and lack spatial coherence by not fully considering pixel-to-pixel relationships, resulting in less refined tumor region segmentation results. The interpretable reasons for the high segmentation accuracy achieved by NFSR-U-Net are divided into two aspects. On the one hand, NFSR-U-Net enables pixel-level embeddings of similar classes to achieve a high fit for classification by learning intra-class similarity in MRI images. On the other hand, this network learns highly representative semantic features by rescaling high-level features using the spatial attention mechanism, and the effectiveness of extracting spatial features by using residual structures is also used to enhance the statistical information of textures and boundaries in tumor MRI images.

In terms of computational complexity, the Params and FLOPs metrics are analyzed in this experiment. Compared with other models, the NFSR-U-Net has a low parameter size of 25.41 M Params, which is lower than 134.3 M of the FCN-8s and 46.70 M of the PSPNet, and its FLOPs are maintained at a low value of 204.01 G, much lower than 1524.34 G of the MSFCN and 1461.23 G of the MSRN. Although the MSRB and MSFCN merge hierarchical image features and reconstruct them at the end of the network, the feature maps contain cumbersome information, and simple concatenation of features at various scales will lead to underutilization of local features, and thus have relatively high computational complexity. In the NFSR-U-Net, we optimize the bottleneck section to ignore useless features and select more relevant ones. Furthermore, the combination of the attention mechanism with the residual structure allows NFSR-U-Net to migrate attention to refine the details of the tumor

tissue portion of the MRI image by targeting particular regions of the MRI image that are of interest rather than the entire image. As a result, there is low computational complexity.

From the analysis of Figures 13 and 14, it can be concluded that NFSR-U-Net achieves the highest values in the Re and F1 metrics, and the final values remain stable. Among them, its Re reaches 0.945, higher than 0.867 of the U-net, and F1 is 0.943, higher than 0.919 of the FPN, which reflects the better tumor MRI segmentation performance of the network. This is attributed to the fact that we focus on the optimization of local details and the interaction of global semantic information by improving the model structure rationally, thus enabling the network to exhibit significant segmentation performance. Concerning the limitations, the network is not optimized sufficiently for the traditional convolutional blocks, so we will consider further improvements for it.

6. Conclusions

In the article, a tumor MRI image segmentation framework, NFSR-U-Net, based on class-correlation pattern aggregation in medical decision-making systems is used with over 4000 MRI images from the Monash University Centre for Artificial Intelligence Research. In this method, we achieve detailed segmentation of the boundaries of tumor lesion tissues by optimizing the dataset, preprocessing, and image segmentation steps. Compared with other models, the results show that this model achieves better segmentation effects for tumor MRI images. Furthermore, the model is more lightweight and consumes fewer resources.

In the future, the framework will be improved to optimize the aggregation of feature patterns so that it can better handle the interaction of local details and global information in tumor MRI images. Meanwhile, we will enhance the statistical information in MRI images to refine the segmentation boundaries of lesion regions and identify texture features of tumor images. We aim to improve the generalization capability of the model so that it can be applied to MRI images of other tumor types.

Author Contributions: Conceptualization, H.W. and H.T.; Methodology, B.L. and J.W.; Software, F.L.; Validation, F.L.; Formal analysis, J.W.; Data curation, J.W.; Writing—original draft, H.T.; Writing—review & editing, H.T. and F.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in the Shandong Humanities and Social Sciences Project, 2022-YYGL-01.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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. Rathore, R.; Van Tine, B.A. Pathogenesis and current treatment of osteosarcoma: Perspectives for future therapies. *J. Clin. Med.* **2021**, *10*, 1182. [[CrossRef](#)] [[PubMed](#)]
2. Wen, Y.; Zhang, L.; Meng, X.; Ye, X. Rethinking the transfer learning for FCN based polyp segmentation in colonoscopy. *arXiv* **2022**, arXiv:2211.02416. [[CrossRef](#)]
3. Zhao, Z.; Zhou, F.; Xu, K.; Zeng, Z.; Guan, C.; Zhou, S.K. LE-UDA: Label-efficient unsupervised domain adaptation for medical image segmentation. *IEEE Trans. Med. Imaging* **2022**, *8*, 102–118. [[CrossRef](#)]
4. Gama, P.H.T.; Oliveira, H.N.; Marcato, J.; Santos, J.D. Weakly Supervised Few-Shot Segmentation Via Meta-Learning. *IEEE Trans. Multimed.* **2022**, *14*, 52–66. [[CrossRef](#)]
5. Hamache, A.; Boudaren, M.E.Y.; Pieczynski, W. Kernel smoothing classification of multiattribute data in the belief function framework: Application to multichannel image segmentation. *Multimed. Tools Appl.* **2022**, *81*, 29587–29608. [[CrossRef](#)]
6. Mehta, D.; Padalia, D.; Vora, K.; Mehendale, N. MRI image denoising using U-Net and Image Processing Techniques. In Proceedings of the 2022 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2–3 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 306–313.

7. Yang, H.; Chen, C.; Lin, W.; Yi, Y. A New CNN-based Joint Network for Brain Tumor Denoising and Classification. In Proceedings of the 2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Nanjing, China, 23–25 September 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 506–510.
8. Cheng, J.; Liu, J.; Kuang, H.; Wang, J. A fully automated multimodal MRI-based multi-task learning for glioma segmentation and IDH genotyping. *IEEE Trans. Med. Imaging* **2022**, *41*, 1520–1532. [[CrossRef](#)]
9. Luu, H.M.; Park, S.-H. Extending nn-UNet for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*; Springer: Cham, Switzerland, 2022; pp. 173–186.
10. Futrega, M.; Milesi, A.; Marcinkiewicz, M.; Ribalta, P. Optimized U-Net for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*; Springer: Cham, Switzerland, 2022; pp. 15–29.
11. Shen, Y.; Gou, F.; Dai, Z. Osteosarcoma MRI image-assisted segmentation system base on guided aggregated bilateral network. *Mathematics* **2022**, *10*, 1090. [[CrossRef](#)]
12. 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. Available online: <https://www.mdpi.com/2227-9032/10/11/2189> (accessed on 25 August 2022). [[CrossRef](#)]
13. 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](#)]
14. 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](#)]
15. Chang, L.; Wu, J.; Moustafa, N.; Bashir, A.K.; Yu, K. AI-driven synthetic biology for non-small cell lung cancer drug effectiveness-cost analysis in intelligent assisted medical systems. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 5055–5066. [[CrossRef](#)]
16. 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](#)]
17. Cai, Z.; Wu, J. Efficient content transmission algorithm based on multi-community and edge-caching in ICN-SIoT. *Peer-to-Peer Netw. Appl.* **2022**, *12*, 1–18. [[CrossRef](#)]
18. Cui, R.; Chen, Z.; Wu, J.; Tan, Y.; Yu, G. A multiprocessing scheme for PET image pre-screening, noise reduction, segmentation and lesion partitioning. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 1699–1711. [[CrossRef](#)]
19. Yu, G.; Chen, Z.; Wu, J.; Tan, Y. Medical decision support system for cancer treatment in precision medicine in developing countries. *Expert Syst. Appl.* **2021**, *186*, 115725. [[CrossRef](#)]
20. Wu, J.; Tan, Y.; Chen, Z.; Zhao, M. Data decision and drug therapy based on non-small cell lung cancer in a big data medical system in developing countries. *Symmetry* **2018**, *10*, 152. [[CrossRef](#)]
21. Arunachalam, H.B.; Mishra, R.; Daescu, O.; Cederberg, K.; Rakheja, D.; Sengupta, A.; Leonard, D.; Hallac, R.; Leavey, P. Viable and necrotic tumor assessment from whole slide images of osteosarcoma using machine-learning and deep-learning models. *PLoS ONE* **2019**, *14*, e0210706. [[CrossRef](#)] [[PubMed](#)]
22. 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](#)] [[PubMed](#)]
23. Bansal, P.; Gehlot, K.; Singhal, A. Automatic Detection of Osteosarcoma Based on Integrated Features and Feature Selection Using Binary Arithmetic Optimization Algorithm. *Multimed. Tools Appl.* **2022**, *81*, 8807–8834. [[CrossRef](#)]
24. 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](#)]
25. Yang, W.; Luo, J.; Wu, J. Application of information transmission control strategy based on incremental community division in IoT platform. *IEEE Sens. J.* **2021**, *21*, 21968–21978. [[CrossRef](#)]
26. Chen, H.; Liu, J.; Cheng, Z.; Lu, X.; Wang, X.; Lu, M.; Zhao, Y. Development and external validation of an MRI-based radiomics nomogram for pretreatment prediction for early relapse in osteosarcoma: A retrospective multicenter study. *Eur. J. Radiol.* **2020**, *129*, 109066. [[CrossRef](#)]
27. Gou, F.; Wu, J. Novel data transmission technology based on complex IoT system in opportunistic social networks. *Peer-to-Peer Netw. Appl.* **2022**, *13*, 1–18. [[CrossRef](#)]
28. 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](#)]
29. 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](#)]
30. Tian, X.; Jia, W. Optimal matching method based on rare plants in opportunistic social networks. *J. Comput. Sci.* **2022**, *64*, 101875. [[CrossRef](#)]
31. 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](#)] [[PubMed](#)]
32. Huang, L.; Xia, W.; Zhang, B.; Qiu, B.; Gao, X. MSFCN-multiple supervised fully convolutional networks for the osteosarcoma segmentation of CT images. *Comput. Methods Programs Biomed.* **2017**, *143*, 67–74. [[CrossRef](#)]
33. Pan, L.; Wang, H.; Wang, L.; Ji, B.; Liu, M.; Chongcheawchamnan, M.; Yuan, J.; Peng, S. Noise-reducing attention cross fusion learning transformer for histological image classification of osteosarcoma. *Biomed. Signal Process. Control.* **2022**, *77*, 103824. [[CrossRef](#)]

34. Nabid, R.A.; Rahman, M.L.; Hossain, M.F. Classification of Osteosarcoma Tumor from Histological Image Using Sequential RCNN. In Proceedings of the 2020 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 17–19 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 363–366.
35. 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: Piscataway, NJ, USA, 2021; pp. 93–99.
36. Nador, 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](#)]
37. Kayal, E.B.; Kandasamy, D.; Sharma, R.; Bakhshi, S.; Mehndiratta, A. Segmentation of osteosarcoma tumor using diffusion weighted MRI: A comparative study using nine segmentation algorithms. *Signal Image Video Process.* **2020**, *14*, 727–735. [[CrossRef](#)]
38. Chen, H.; Zhang, X.; Wang, X.; Quan, X.; Deng, Y.; Lu, M.; Wei, Q.; Ye, Q.; Zhou, Q.; Xiang, Z.; et al. MRI-based radiomics signature for pretreatment prediction of pathological response to neoadjuvant chemotherapy in osteosarcoma: A multicenter study. *Eur. Radiol.* **2021**, *31*, 7913–7924. [[CrossRef](#)] [[PubMed](#)]
39. Guan, X.; Yang, G.; Ye, J.; Yang, W.; Xu, X.; Jiang, W.; Lai, X. 3D AGSE-VNet: An automatic brain tumor MRI data segmentation framework. *BMC Med. Imaging* **2022**, *22*, 6. [[CrossRef](#)] [[PubMed](#)]
40. Zhang, C.; Shen, X.; Cheng, H.; Qian, Q. Brain tumor segmentation based on hybrid clustering and morphological operations. *Int. J. Biomed. Imaging* **2019**, *2019*, 7305832. [[CrossRef](#)]
41. Dutande, P.; Baid, U.; Talbar, S. Deep Residual Separable Convolutional Neural Network for lung tumor segmentation. *Comput. Biol. Med.* **2022**, *141*, 105161. [[CrossRef](#)]
42. Zhang, C.; Lu, J.; Hua, Q.; Li, C.; Wang, P. SAA-Net: U-shaped network with Scale-Axis-Attention for liver tumor segmentation. *Biomed. Signal Process. Control.* **2022**, *73*, 103460. [[CrossRef](#)]
43. Torabi, A.; Bilodeau, G.-A. Local self-similarity-based registration of human ROIs in pairs of stereo thermal-visible videos. *Pattern Recognit.* **2013**, *46*, 578–589. [[CrossRef](#)]
44. Kim, S.; Min, D.; Ham, B.; Jeon, S.; Lin, S.; Sohn, K. Fcsc: Fully convolutional self-similarity for dense semantic correspondence. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6560–6569.
45. Zheng, C.; Cham, T.-J.; Cai, J. The spatially-correlative loss for various image translation tasks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 16407–16417.
46. Ramachandran, P.; Parmar, N.; Vaswani, A.; Bello, I.; Levskaya, A.; Shlens, J. Stand-alone self-attention in vision models. *Adv. Neural Inf. Process. Syst.* **2019**, *14*, 32. [[CrossRef](#)]
47. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
48. 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.
49. Drozdal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; Pal, C. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 179–187.
50. Ibtehaz, N.; Rahman, M.S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw.* **2020**, *121*, 74–87. [[CrossRef](#)] [[PubMed](#)]
51. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
52. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
53. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542. [[CrossRef](#)]
54. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 25–38. [[CrossRef](#)]
55. Ponti, M.; Kittler, J.; Riva, M.; de Campos, T.; Zor, C. A decision cognizant Kullback–Leibler divergence. *Pattern Recognit.* **2017**, *61*, 470–478. [[CrossRef](#)]
56. Masci, J.; Meier, U.; Cireşan, D.; Schmidhuber, J. Stacked convolutional auto-encoders for hierarchical feature extraction. In *International Conference on Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 52–59.
57. Gondara, L. Medical image denoising using convolutional denoising autoencoders. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 12–15 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 241–246.
58. 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](#)]
59. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

60. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
61. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
62. Li, J.; Fang, F.; Mei, K.; Zhang, G. Multi-scale residual network for image super-resolution. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 517–532.
63. 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; IEEE: Piscataway, NJ, USA, 2022; pp. 1539–1543.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.