

Article

Artificial Intelligence-Aided Diagnosis Solution by Enhancing the Edge Features of Medical Images

Baolong Lv ¹, Feng Liu ^{2,3,*}, Yulin Li ¹ , Jianhua Nie ⁴, Fangfang Gou ⁵ and Jia Wu ^{5,6,*} 

¹ School of Modern Service Management, Shandong Youth University of Political Science, Jinan 250102, China; lvbaolong2010@sina.com (B.L.); lyfxs@126.com (Y.L.)

² School of Information Engineering, Shandong Youth University of Political Science, Jinan 250102, China

³ New Technology Research and Development Center of Intelligent Information Controlling in Universities of Shandong, Jinan 250103, China

⁴ Shandong Provincial People's Government Administration Guarantee Center, Jinan 250011, China; niejianhuasd@126.com

⁵ School of Computer Science and Engineering, Central South University, Changsha 410017, China; gff8221@163.com

⁶ Research Center for Artificial Intelligence, Monash University, Melbourne, VIC 3800, Australia

* Correspondence: liusdyu@sina.com (F.L.); jiawu0510@csu.edu.cn (J.W.)

Abstract: Bone malignant tumors are metastatic and aggressive. The manual screening of medical images is time-consuming and laborious, and computer technology is now being introduced to aid in diagnosis. Due to a large amount of noise and blurred lesion edges in osteosarcoma MRI images, high-precision segmentation methods require large computational resources and are difficult to use in developing countries with limited conditions. Therefore, this study proposes an artificial intelligence-aided diagnosis scheme by enhancing image edge features. First, a threshold screening filter (TSF) was used to pre-screen the MRI images to filter redundant data. Then, a fast NLM algorithm was introduced for denoising. Finally, a segmentation method with edge enhancement (TBNNet) was designed to segment the pre-processed images by fusing Transformer based on the UNet network. TBNNet is based on skip-free connected U-Net and includes a channel-edge cross-fusion transformer and a segmentation method with a combined loss function. This solution optimizes diagnostic efficiency and solves the segmentation problem of blurred edges, providing more help and reference for doctors to diagnose osteosarcoma. The results based on more than 4000 osteosarcoma MRI images show that our proposed method has a good segmentation effect and performance, with Dice Similarity Coefficient (DSC) reaching 0.949, and show that other evaluation indexes such as Intersection of Union (IOU) and recall are better than other methods.

Keywords: osteosarcoma; artificial intelligence; magnetic resonance imaging (MRI); pre-screening; denoising; edge enhancement

MSC: 68T01



Citation: Lv, B.; Liu, F.; Li, Y.; Nie, J.; Gou, F.; Wu, J. Artificial Intelligence-Aided Diagnosis Solution by Enhancing the Edge Features of Medical Images. *Diagnostics* **2023**, *13*, 1063. <https://doi.org/10.3390/diagnostics13061063>

Academic Editor: Dechang Chen

Received: 1 December 2022

Revised: 18 February 2023

Accepted: 9 March 2023

Published: 10 March 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

Osteosarcoma is the most common solid tumor of bone origin, accounting for approximately 20% of primary sarcomas of bone [1]. Osteosarcomas frequently occur in adolescents or children under 20 years of age, with more than 75% of patients having an age of onset younger than 25 years [2]. Although some osteosarcomas can be cured by surgical means, some cases remain that are highly fatal even with the most aggressive treatment measures. With the application of comprehensive therapeutic approaches, a cure rate of 65–70% can be achieved in some patients. However, osteosarcoma is prone to lesions with long treatment cycles and poor prognosis, leading to high mortality rates of malignant tumors [3]. Patients with advanced osteosarcoma had a 5-year survival rate of

only 20% [4]. The early detection of malignancies can significantly enhance disease cure rates and minimize patient death [5].

In the current application of evaluation of suspected osteosarcoma, radiographs are inaccurate in determining tumor borders, often resulting in results smaller than the actual size of the tumor [6]. Although CT shows the extent of bone destruction well, MRI performs better than CT in showing the extent of focal lesions [7]. MRI has the advantage of providing multidimensional images and allowing for more sensitive quantification of the extent of marrow cavity involvement [8]. Therefore, MRI images are very important for physicians to clinically examine patients with osteosarcoma.

The artificial intelligence-aided detection of medical images is important for predicting patient outcomes and monitoring disease progression with the development of treatment strategies [9–11]. Due to the general underdevelopment of health care systems, most countries, especially developing countries, still suffer from a strain and uneven distribution of health care resources [12]. Many hospitals have difficulty in meeting the hardware and staffing requirements for osteosarcoma treatment. Most diagnoses of osteosarcoma rely on the manual recognition of pictures [13]. However, the volume of image data from patients with osteosarcoma is enormous, but few images are of value. Only about 20 of the approximately 700 MRI images generated per patient with osteosarcoma may be useful to the physician for diagnosis [14,15]. The manual screening and processing of validated osteosarcoma images by physicians is a time-consuming and laborious process [16]. Considering the lack of uniformity in the diagnosis of histological features of osteosarcoma, it requires a high level of expertise and knowledge base in pathology on the part of the diagnosing physician. Diagnosis by inexperienced physicians is highly subjective, which can lead to an increased rate of misdiagnosis.

At present, image processing technology is developing rapidly, especially in image segmentation, and new methods are constantly proposed [17]. As medical image processing technology progress, more and more imaging modalities are being employed to diagnose osteosarcoma [18]. Analysis of the extent of tumor infiltration and the boundaries of tumor infiltration by MRI helps clinicians to localize the tumor. MRI images are susceptible to various noise sources (including thermal noise and physiological noise [19] caused by mechanical defects and external signals), which significantly reduces the image quality. The denoising operation is necessary to improve diagnostic efficiency. In addition, osteosarcoma itself has complex local tissue formation and morphological changes, difficult-to-maintain marginal features, and blurred tumor boundaries. Due to this, some medical image processing algorithms are less effective at identifying osteosarcomas, and it is challenging to obtain global multidirectional features and implicit features [20], as well as both accuracy and performance. It is crucial to improve the efficiency of osteosarcoma diagnosis by effectively extracting global features and solving the edge ambiguity segmentation problem without consuming too many computational resources and time costs. Edge feature-based methods are processing models employed in medical images, such as Transformer which has achieved wide application in the field of medical images by taking advantage of global modeling [20–22]. Such methods have achieved good results when dealing with simple tasks. When faced with the MRI segmentation task of osteosarcoma where there are a large number of complex boundaries and blurred lesion edges, it is difficult to acquire edge features globally and in multiple directions and to mine the implicit features, making the model accuracy and robustness poor [22]. Therefore, their segmentation results did not meet the expectations.

To improve the recognition accuracy of tumors, this study proposes an artificial intelligence-assisted diagnostic scheme for the MRI images of osteosarcoma with edge-enhanced features. It first processes the raw MRI images using threshold screening filtering (TSF) to filter out the redundant data in lesion-free regions. Then, the fast NLM algorithm and Fourier transform are fused to reduce the noise of the images. Finally, a segmentation network (TBNNet) with edge-enhancement features is designed to improve recognition accuracy by enhancing the tumor edge features. It effectively alleviates the problem of

the imprecise identification of fuzzy tumor boundaries by existing methods. The network introduces a channeled edge-crossing transformer instead of Unet's skip connection, and combines loss functions to globally segment tumor regions of different sizes at multiple scales, optimizing the effect of osteosarcoma segmentation.

2. Related Work

The segmentation of medical images by computer technology (including CT, MRI, X-ray, etc.), which in turn helps doctors to diagnose diseases, is gradually becoming a popular research topic. Listed and described below are some of the mainstream algorithms in the field.

Recently, visual transformers have been increasingly used for the segmentation of medical pictures. Jienengchen et al. proposed TransUNet [21], the first transformer-based segmentation method in medical imaging. The first transformer-based pure U-shaped design was Swin-Unet [22] proposed by Liu et al. MedT network was proposed by Jeya Maira Jose Valanarasu et al. [23]. To improve medical picture segmentation, Yuhe Gao et al. presented the UTNet [24], a simple and powerful hybrid transformer design that combines self-attention into convolutional neural networks. Yuan Fengji et al. proposed the unified transformer network Multi-Composite Transformer (MCTrans) for cross-scale dependency and semantic consistency learning problems [25]. Olivier Petit et al. combined the complementary capabilities of the U-shaped network and transformer and proposed the UNet-capable of combining the self-attention and cross-attention of both transformers [26]. TransFuse network [27] was suggested by Yundong Zhang et al., which mixes transformers and CNNs in tandem, proposing a fusion of the two techniques, in contrast to most earlier efforts, which replaced the convolutional layers in U-Net networks with the transformer or cascaded the two. HiFormer, a hierarchical multiscale representation transformer was recently proposed by Moein Heidari et al. [28]. Chen Wei et al. presented that HRSTNet [29] replaces the convolutional layer with a transformer module that creates varied resolution feature mapping information. Ruina Sun et al. [30] introduced an effective image classification segmentation technique based on an enhanced Swin transformer, which was designed specifically for lung cancer classification and segmentation. UNeXt model [31] was proposed by Vishal M. Patel et al. as a deep network architecture. CapsNet [32], a capsule network architecture built, was developed by Minh Tran et al. for medical picture segmentation.

In the disease diagnosis of osteosarcoma, image processing by computer technology as an aid to diagnosis has gradually become a research hotspot. The accurate classification of the morphology of different stages of osteosarcoma can enable the timely control of the spread and treatment of osteosarcoma patients in early diagnosis. In the article [33], a CNN has been used to pre-train a publicly available dataset of osteosarcoma tissue images to avoid extensive metastasis. Yu Fu [34] developed a DS-Net network capable of automatically classifying histological images. Hosein Barzekar et al. presented C-Net [35], a new convolutional neural network structure with multiple tandem CNNs. Rahad Arman Nabid [36] proposed a convolutional network to evaluate the grading of patients with osteosarcoma, but the model suffers from the problem of simple overfitting. To identify osteosarcoma cells from osteoblastic cells (MSC), the method proposed by Mario D'Acunto et al. [37] has an accuracy close to one and allows for the study of single cells but requires a huge amount of cellular data. Parlak et al. [38] employed diffusion-weighted imaging to accurately show Ewing's sarcoma and osteogenic sarcoma by measuring the expression diffusion coefficient (ADC) values for borderline case segmentation of Ewing's sarcoma and osteogenic sarcoma.

Many scholars in the literature have proposed the application of many image-processing techniques for the early identification and prediction of treatment options for patients with osteosarcoma. Su Young Jeong et al. [39] proposed the use of machine learning methods incorporating baseline 18-FDG positron emission tomography to predict textural features in scanned images. To this end, Hyung-Jun Im et al. [40] proposed various segmentation methods for pseudo-myelinating lesions, including the relative background threshold

method, the gradient-based method (PETedge), and Bluse Otsu (MO-PET). Shuai Limei et al. proposed W-net+ [41], a network structure based on a dense jump connection structure and cascaded dual U-Net. WB Huang et al. [42] proposed a fully automated MRI method for osteosarcoma detection. It is used to identify tumors with irregular structure and shape by using conditional random fields.

The above analysis demonstrates that picture segmentation techniques are becoming increasingly significant in illness diagnosis and prognosis assessment. However, because the pictures are vulnerable to noise, edge features in image segmentation of osteosarcoma are still challenging to retain and segmentation accuracy needs to be improved. In this study, an artificial intelligence-aided diagnosis method for osteosarcoma with edge-enhancement features is proposed, which improves the accuracy of tumor recognition by enhancing edge information.

3. Methodology

With the development of computer image processing technology, artificial intelligence solutions are widely used in the medical field. Its importance is to solve the problems of high consumption of medical resources and low efficiency of disease diagnosis in developing countries [43–46]. However, such methods still present major challenges in the recognition of tumors. Taking osteosarcoma MRI images as an example, existing methods have difficulty in handling images with complex lesions and blurred edges [10,47]. This paper proposes an artificial intelligence-aided diagnosis scheme to provide more options for osteosarcoma-aided diagnosis in developing countries. The overall design of the scheme is in Figure 1. First, we perform MRI image pre-screening by Threshold Screening Filter (TSF) to filter redundant data and extract useful images for submission to the next step; then, we introduce a combined fast Fourier transform NLM algorithm for noise reduction in MRI images to improve the diagnostic quality. Finally, to improve the recognition accuracy of the model for tumor edges, a U-shaped network with a fused transformer featuring edge enhancement is used to segment the pre-processed images. The channeled transformer bridges the semantic gap that exists in the UNet network. The method also introduces the edge-enhancement module (BAB) with a combined loss function to optimize the effect of edge segmentation.

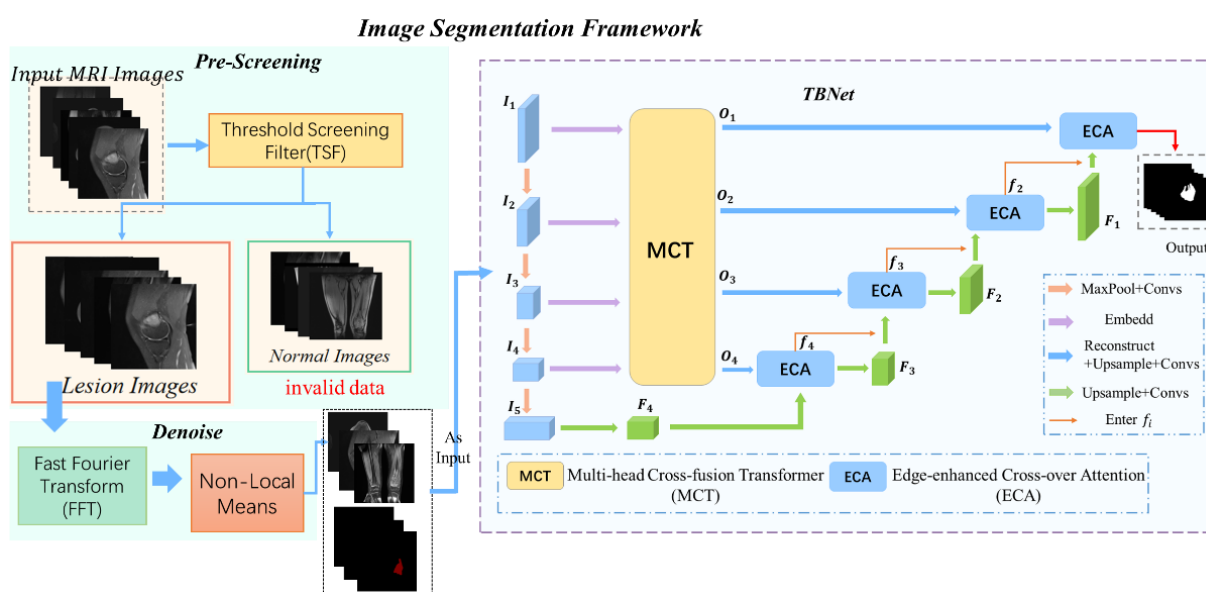


Figure 1. Image Segmentation Framework.

The artificial intelligence-aided diagnosis scheme of the osteosarcoma MRI images we constructed is divided into two main sections: Section 3.1 is data preprocessing; Section 3.2 is the image segmentation model of osteosarcoma MRI. In Section 3.1, we are mainly divided

into two parts, which are the pre-screening operation based on Threshold Screening Filter (TSF) and the noise-reduction processing based on a fast NLM. Table 1 shows the main mathematical symbols of this chapter and their annotations.

Table 1. Main mathematical notations and their annotations.

Notation	Meaning
$\mathbb{Z}^2$	The osteosarcoma MRI image domain
B_a	A block of image windows of size $n_1 \times n_1$
$\bar{B}_a$	The result of B_a initialized noise reduction
λ_T	The hyperparameter of the TSF
p	The Euclidean distance
λ_{FFT2}	The fixed threshold parameter
X	MR noise data
S	Original MR images and Phases
$k(a)$	Noise component intensity of the i -th pixel
I_0	The modified Bessel function
η_1, η_2	Real and virtual channel impact
$k(a)$	The activation function of TSF
$\omega(N_i, N_j)$	The function of weighted similarity
$\mathcal{F}_{2D}, \mathcal{F}_{2D}^{-1}$	The DCT transform operator (FFT) and its inverse
$Z(i)$	The normalization constant
$\phi(\cdot)$	The boundary level set
C_S	Search window side half-length
$M_i, \hat{M}_i$	The similarity matrix and Mask edge diagram
$\psi(\cdot)$	The instance normalization
$\delta(\cdot)$	The ReLU operator
f_{i-1}	Supplementary layer feature map for layer $i - 1$
$g(\cdot)$	The probabilistic output of the network

3.1. Data Pre-Processing

3.1.1. MRI Image Pre-Screening Based on Threshold Screening Filter (TSF)

The purpose of pre-screening is to obtain some valuable images from the original MRI images. For MRI image filtering, we used a threshold filter (TSF).

$$T(a) = k \left(\left| \frac{\sum_{1 \leq i \leq N_1} \sum_{1 \leq j \leq N_1} (\bar{B}_a(i, j) - 0.5\lambda_T)}{n_1^2} \right| - \lambda_T \right) \quad (1)$$

where λ_T is a hyperparameter of the TSF controlling the screening threshold of the lesion.

Noise significantly reduces the accuracy and there is no greater need for edge details during pre-screening for the time being. Therefore, we introduced the discrete Fourier transform (DFT) [48] for coarse initial denoising of noise at high frequencies first. $\bar{B}_a$ is the coarse initial noise reduction method for B_a .

$$\bar{B}_a = \mathcal{F}_{2D}^{-1}(s(\mathcal{F}_{2D}(B_a), \lambda_{FFT2} \sigma \sqrt{2 \log(n_1^2)})) \quad (2)$$

where $\mathcal{F}_{2D}$ represents the DCT transform operator, $\mathcal{F}_{2D}^{-1}$ represents the inverse of $\mathcal{F}_{2D}$, and λ_{FFT2} is the fixed threshold parameter. Additionally, s is defined as:

$$s(a, \lambda) = \begin{cases} 0, & -\lambda < a < \lambda \\ \lambda, & a > \lambda, a < -\lambda \end{cases} \quad (3)$$

In Equation (1), k is:

$$k(a) = \begin{cases} 0, & a \geq 0 \\ 1, & a < 0 \end{cases} \quad (4)$$

If the output of $k(a)$ is 1, the location of the window block B_a contains the lesion area, and the image is kept to mark the image as “lesion image”. When it is not 1, the image of window block B_a is considered invalid and the sliding traversal of the entire image continues. After the original data set is pre-screened, the redundant data without lesions are eliminated, and the useful and valuable image data are retained for further operations.

3.1.2. Noise Reduction Based on Fast NLM Algorithm

The degree of tumor invasion and the invasion boundary is different, which affects the recognition effect of the model to a certain extent. The source of noise influence in MRI images is different from the usual images. The usual images are perturbed by Gaussian white noise. MRI images, on the other hand, are influenced by the bias induced by the Rician distribution noise associated with the MR acquisition system signal, as well as Gaussian white noise. Real and imaginary channels are influenced by and collect complex MR data X .

$$X = S \cos \theta + \eta_1 + j(S \sin \theta + \eta_2) \quad (5)$$

where S is the original MR image and θ is the phase. Changing the noise distribution to Rician can be expressed as:

$$P\left(\frac{X}{S}, \sigma\right) = \frac{X}{\sigma^2} e^{-\left(\frac{X^2 + S^2}{2\sigma^2}\right)} I_0\left(\frac{XS}{\sigma^2}\right) \quad (6)$$

By first squaring the magnitude of the data X and then considering the expectation of both sides, it is possible to determine the bias caused by the Rician distribution.

The non-local mean (NLM) takes advantage of the redundancy of the image, and the pixels in the image filtered by non-local averaging are a weighted average of all other pixels [7]. NLM can successfully remove the aforementioned image noise. This number is determined by examining the picture block and determining the sliding window's similarity to the selected window. The NLM method's core equation is Equation (7).

$$NLM(I(i)) = \sum_{j \in N_i} \omega(N_i, N_j) I(j) \left(\omega(N_i, N_j) = \frac{1}{Z(i)} e^{\frac{(-p)}{h^2}} \right) \quad (7)$$

where $0 \leq \omega(N_i, N_j) \leq 1$, $\sum_{j \in N_i} \omega(N_i, N_j) = 1$, $\omega(N_i, N_j)$ is the weight to calculate the similarity between two patches, h represents the filtering parameter, and p represents the Euclidean distance between two patches.

The original NLM algorithm takes a lot of time to compute the similarity weights and has a high distance computation complexity [49]. To keep the NLM accurate while reducing the consumption of computational resources, we introduce the Fast Fourier Transform (FFT) strategy to implement a fast algorithm for computing pixel-level NLM. v is the input noisy image, V is the patch, c_s is the patch edge half-length, c is the patch side length as $c = 2 \times c_s + 1$, and c^2 is the number of pixels in the patch procedure. First, we rearrange the loop to consider all pixels i of all translation vectors $t \in [[-C_s, +C_s]]^2$. In the context of patches distance, the Euclidean distance can be calculated as follows:

$$s_t(i) = \|[v(i) - v(i+t)]\|_2^2, i = (i_1, i_2) \in \Omega \quad (8)$$

Then, in this case, the weighted norm of the patch difference is a discrete convolution:

$$\|V(i) - V(i+t)\|_{2,K}^2 = \sum_{\{b \in \mathbb{Z}^2: \|b\|_\infty \leq d_s\}} K(b) \|v(i+b) - v(i+t+b)\|_2^2 = (\tilde{K} * s_t)(i) \quad (9)$$

where $*$ denotes discrete convolution, $\tilde{K}(b) = K(-b)$. It is calculated by $\mathcal{F}$ and its inverse $\mathcal{F}^{-1}$.

$$\|V(i) - V(i+t)\|_{2,K}^2 = \mathcal{F}^{-1} \left(\mathcal{F}(\tilde{K}) \mathcal{F}(s_t) \right)(i) \quad (10)$$

The time complexity of FFT is $O(ND^2 \log(N))$, subject to the computation of any translation vector $t \in [[-C_s, +C_s]]^2$, and the computation of NLM weights is independent of the patch size. Assuming that the total number of image pixels is N , the introduction of FFT enables the NLM algorithm to obtain a good speedup. Specifically, the time complexity is reduced from the original $O(ND^2 c^2)$ to $O(ND^2 \log(N))$. As shown in Figure 1, after the image data are processed by noise reduction, they are then put into the subsequent segmentation operation.

3.2. Tumor Localization

The original data are input to the segmentation model TBNNet for segmentation after a series of data preprocessing, such as image pre-screening and noise reduction. TBNNet is a U-Net network with edge-enhanced features of a fusion transformer designed by us, which can effectively solve the edge blur segmentation problem caused by blurred lesion edges. TBNNet is based on the UNet network without skip connection, introducing a multi-head cross-fusion transformer (MCT), edge-enhanced cross-attention module (ECA) with combined loss function; this optimizes the segmentation effect and solves the edge blur segmentation problem. The model mainly consists of U-Net without a skip-connection mechanism [50], channel edge cross-fusion transformer (including multi-head cross-fusion transformer module MCT and edge-enhanced cross-attention module ECA), and combined loss function. Figure 2 shows the general design of the model.

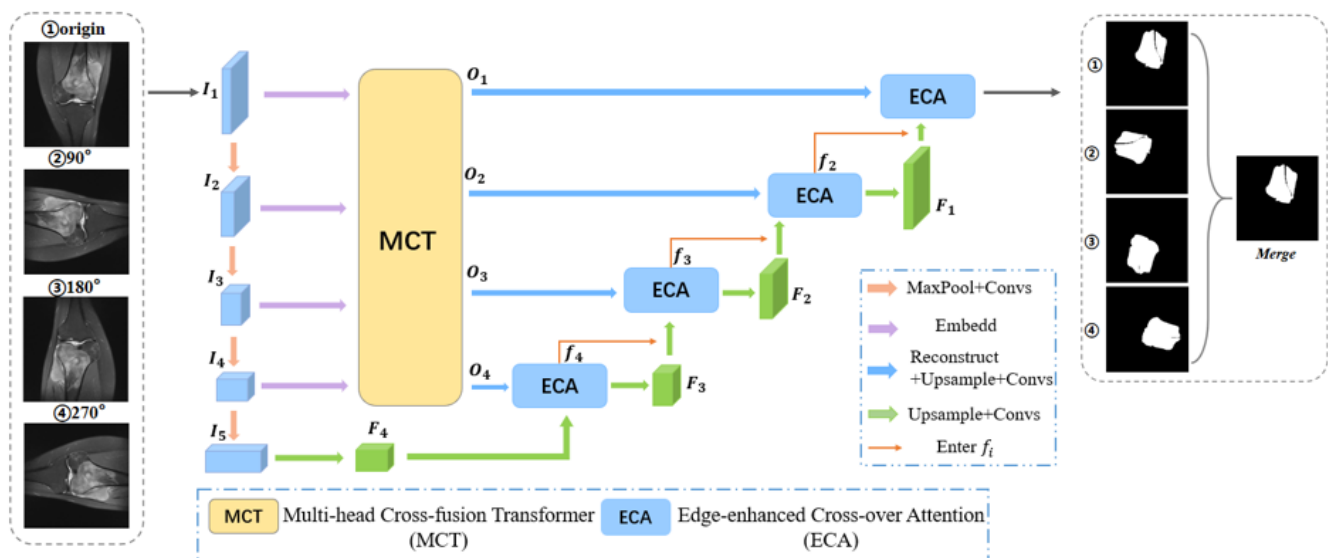


Figure 2. TBNNet Segmentation Model.

The original U-net does not work well in the osteosarcoma diagnosis task. Osteosarcoma itself has complex morphological changes, difficult-to-maintain edge features, and blurred tumor boundaries, coupled with the fact that MRI images, the tools relied on for diagnosis, are vulnerable to noise. This makes the skip connection mechanism have a certain semantic gap in the osteosarcoma segmentation recognition task, makes the stage features incompatible, and makes it difficult to obtain edge features and mine the implicit features globally and in multiple directions, thus making the model accuracy and robustness poor with a certain impact on the segmentation. We introduce a cross-fusion channel transformer with edge-enhancement features to replace the jump connection while taking advantage of the transformer and Unet for the cross-fusion of multi-scale channel information, solving the problem of semantic hierarchical inconsistency, and enhancing edge information to improve the edge segmentation ambiguity problem.

The channelized edge cross-fusion transformer consists of MCT for encoder feature transformation, MCT for encoder feature conversion, and the edge-enhancement cross-attention module (ECA) for decoder feature fusion.

(1) Multi-head cross-fusion transformer (MCT) for encoder feature transformation

MCT is a multi-scale global feature investigation of osteosarcoma MRI images using transformer long-dependent modeling to fuse multi-scale encoder features. It is divided into three stages: Embedding Multi-scale Features, Multi-Crossing Attention, and Multi-Layer Perceptron. Figure 3 shows the design of the MCT.

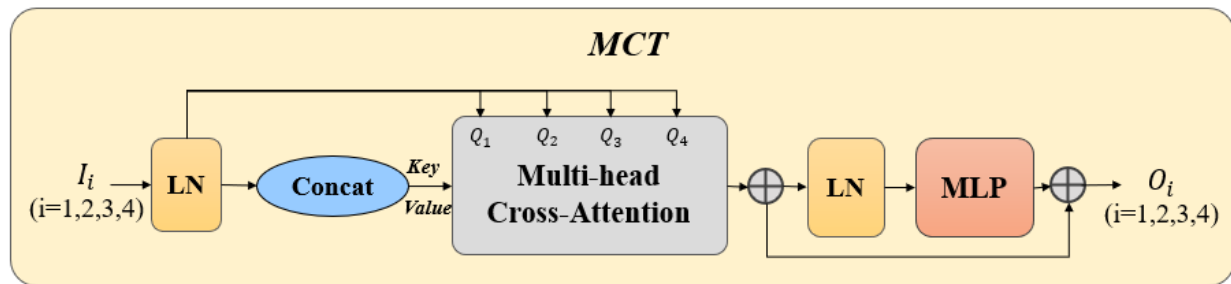


Figure 3. Multi-Head Cross-Fusion Transformer (MCT).

Embedding Multi-Scale Features. The embedding of multi-scale features starts by labeling the features with the I_i of the multi-scale feature of the given osteosarcoma MRI image, and reshaping the 2D patch of the flat feature sequence so that the patch size is P , $P/2$, $P/4$, and $P/8$. In this process, the channel size remains the same. We then connect the four layers as keys and values $L_\Sigma = \text{Concat}(L_1, L_2, L_3, L_4)$ and feed them into a multi-headed cross-attention module to encode the channels and dependencies to refine the features at the encoder level of each osteosarcoma MRI image using multiscale features.

Multi-Crossing Attention. As shown in Figure 4, there are five inputs required for this module. Let $T_i \in R^{C_i \times d}$, $T_{Key} \in R^{C_\Sigma \times d}$, $T_{Value} \in R^{C_\Sigma \times d}$, and $W_i \in R^{C_i \times d}$, $W_{Key} \in R^{C_\Sigma \times d}$, $W_{Value} \in R^{C_\Sigma \times d}$ be the weights of different inputs; d is the sequence length, C_i ($i = 1, 2, 3, 4$) is the channel size to skip the concatenated layer, and the four sizes we use are $C_1 = 64$, $C_2 = 128$, $C_3 = 256$, $C_4 = 512$. Through the cross-notice mechanism, a similarity matrix M_i is generated and T_{Value} is weighted.

$$T_i = L_i W_i, T_{Key} = L_\Sigma W_{Key}, T_{Value} = L_\Sigma W_{Value} \quad (11)$$

$$CA_i = M_i T_{Value}^T = \sigma \left[\psi \left(\frac{T_i^T T_{Key}}{\sqrt{C_\Sigma}} \right) \right] T_{Value}^T = \sigma \left[\psi \left(\frac{W_i^T L_i^T L_\Sigma W_{Key}}{\sqrt{C_\Sigma}} \right) \right] W_{Value}^T L_\Sigma^T \quad (12)$$

where $\psi(\cdot)$ and $\sigma(\cdot)$ denote the instance normalization and the softmax function, respectively. We change the original method of the attention operation along the patch-axis in the self-attention mechanism and change the direction to along the channel direction. Instance normalization is also used to normalize each instance's similarity matrix on the similar mapping, allowing for the gradient to flow smoothly. In the case of N-head focus:

$$MCA_i = \frac{\sum_{0 < j < N, j \in Z} CA_i^j}{N} \quad (13)$$

Multi-Layer Perceptron (MLP). After the multiple cross-notice mechanisms, the results are fed into the MLP with the residual structure and residual operator to obtain the output. An L-layer transformer is constructed by repeating the operation equation (14) L times. The four outputs O_1, O_2, O_3, O_4 of the L-th layer are reconstructed by up-sampling

the operations after convolutional layers. They are connected to the decoder features F_1, F_2, F_3, F_4 , respectively, and fed to the Edge-Enhanced Cross-Attention (ECA).

$$O_i = MCA_i + MLP(T_i + MCA_i) \quad (14)$$

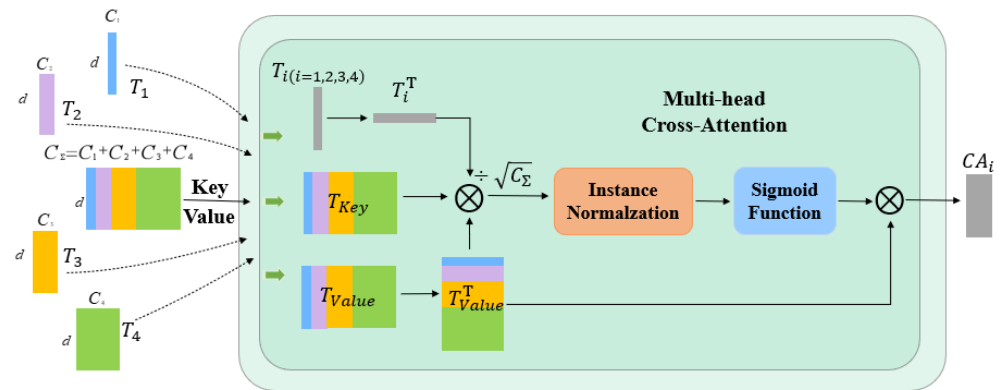


Figure 4. Multi-Head Cross-Attention Module.

The segmentation of the focus area in the MRI image is a binary classification task. The introduction of too many layers of transformers will lead to the explosive growth of computational complexity and model over-fitting problems. After balancing resource cost and segmentation accuracy, N and L are set to 4.

(2) Edge-Enhanced Cross Attention (ECA) module for decoder feature fusion

As shown in Figure 5, we propose a channel-based edge-enhancement cross-attention module to better fuse the semantic inconsistency characteristics between the channel transformer and the U-Net decoder, as well as to correct for fuzzy edge segmentation and the lack of some regions in osteosarcoma MRI images. The module is split into two sections: Decoder Channel Crossing Attention Module and Edge-Enhancement Module. It can not only direct the channel and information filtering of interceptor features to minimize ambiguity in decoder features, but it can also increase edge information to replenish missing regions.

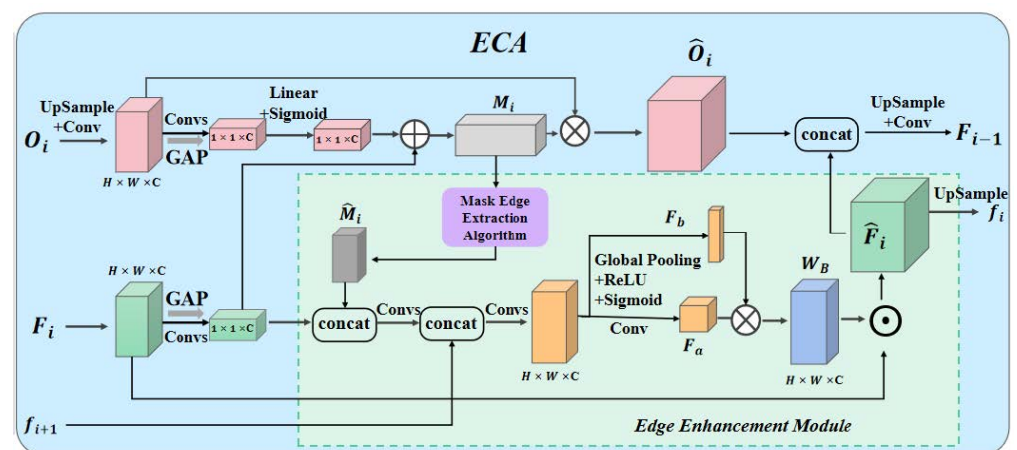


Figure 5. Edge-Enhanced Cross-Attention Module (ECA).

Decoder Channel Crossing Attention Module. We take level i transformer output $O_i \in R^{C \times H \times W}$ and $F_i \in R^{C \times H \times W}$ as the input for cross-attention. Spatial squeezing is performed through the GAP layer to construct attention masks.

$$M_i = X_1 \cdot \zeta(O_i) + X_2 \cdot \zeta(F_i) \quad (15)$$

where the resulting vector $\xi(x) \in R^{C \times 1 \times 1}$, the k -th channel $\xi(x) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X^k(i, j)$, $x_1 \in R^{C \times C}$ and $x_2 \in R^{C \times C}$. We recalibrate and excite O_i .

$$\widehat{O}_i = \sigma(M_i) \cdot O_i \quad (16)$$

Edge-Enhancement Module. The blurred focus edge in MRI images often leads to inaccurate tumor segmentation. To solve the problem of edge fuzzy segmentation, the edge-enhancement module is introduced into the cross-attention module of the feature fusion part of the encoder.

$$F_{ab} = d_3 \left[c \left(d_3 \left[c \left(d_1[F_i], \widehat{M}_i \right), f_{i+1} \right] \right) \right] \quad (17)$$

For the edge attention mechanism, we are inspired by the spatial channel compression and stimulus attention module and designed the attention module that focuses more on the edge feature information with the same simplicity and without increasing the parameters. The specific process can be expressed as follows: first, the input feature map $F_{ab} \in R^{C \times H \times W}$ is compressed in channel and space, and the compressed feature map $F_a \in R^{H \times W \times 1}$ obtained after the compression is multiplied with the vector $F_b \in R^{1 \times 1 \times C}$ to obtain a weight $W_B \in R^{C \times H \times W}$ of the same size as the input. This not only provides a corresponding weight for each pixel of the input feature map but can also highlight more important location information at the edges and suppress location information of a small value.

$$\widehat{F}_i = (F_a \times F_b) \odot F_{ab} \quad (18)$$

where $\times$ represents the expansion to direct multiplication and $\odot$ represents the pixel-by-pixel multiplication. Finally, the output $\widehat{O}_i$ of the decoder channel cross-attention module and the output $\widehat{F}_i$ of the edge-enhancement module will converge to obtain the final output F_i . F_i will continue to work as an input to the $i - 1$ layer.

In addition, the following loss function is proposed.

$$L = \alpha L_{Dice} + \beta L_{BD} \quad (19)$$

It combines dice loss and boundary loss:

$$L_{Dice} = 1 - \frac{2PT}{P^2 + T^2} \quad (20)$$

$$L_{BD} = \int_{\Omega} \phi_G(i) g(i) di \quad (21)$$

where P is the predicted value, T denotes the true label value, G is Ground Truth, and $g(\cdot)$ is the softmax probability output of the network. If $i \in G$, then $\phi_G(i)$ is the negative value of the distance between the point and G ; conversely, it has a positive value. We introduce parameters α and β as balancing coefficients in the defining Equation (19) of the integrated loss function L .

By focusing on both the tumor region and edge information, our method effectively alleviates the edge-blurring problem that has not been addressed in previous research methods.

4. Experimental Analysis

4.1. Dataset

The data for this article are from the Research Center for Artificial Intelligence of Monash University [51]. The original dataset included more than 4000 MRI image samples from 204 patients with osteosarcoma. All patients are diagnosed by specialists at the hospital based on clinical presentation and pathological images. All images were rotated by 90° , 180° , and 270° , respectively, to improve the generalization ability of the model. Finally, the dataset was partitioned into a training set, a validation set, and a test set in a

ratio of 7:1:2. In addition, the ground-truth label of our MRI images was carried out by a collaboration of three doctors at the hospital. As shown in Figure 1, in this mission, only the focal area needs to be segmented. Therefore, the ground-truth label consists of two parts—the tumor area is the foreground and the other part is the background.

The experiments in this paper were conducted under the Ubuntu 18.04.5 operating system with Python as the main programming language, Python version 3.8, and PyTorch version 1.10.0. The GPU used in the experiments was a GTX 3060, and the CPU used was an Intel(R) Xeon(R) E5-2630L v3 with 30 GB of RAM.

4.2. Evaluation Metrics

Based on an earlier study, classification accuracy (CA) expresses the ratio between the screened accurate samples and the expected samples [52]. We used CA to assess the validity of the MRI image pre-screening algorithm. Let C_R be the correctly classified image, i.e., the lesioned image that is correctly screened out by the normal image, C_F be the incorrect classification result, and $C_R + C_F$ be denoted as all the images of the input. The overall error of this process can be denoted by C_F .

$$CA = \frac{C_R}{C_R + C_F}$$

The peak signal-to-noise ratio (PSNR) can objectively quantify the denoising effect. NMISE represents the normalized mean squared error of integration between the denoised and noise-free data [53].

$$PSNR = 10 \log_{10} \left(\frac{\max(p_i)^2}{NMISE} \right)$$

Intersection of Union (IOU) is a metric that is commonly used to measure the similarity between the predicted tumor region and the actual tumor region [54]. Specifically, IOU is the ratio of the intersection region between the judged tumor region I_1 and the real tumor region I_2 . The larger this index is (the closer it is to 1), the greater the impact of the segmentation.

$$IOU = \frac{I_1 \cap I_2}{I_1 \cup I_2}$$

DSC is commonly used to indicate segmentation effects on a similarity measure from 0 to 1 [55]. The DSC is calculated as the ratio of twice the area of the junction of I_1 and I_2 to the entire area, as expressed in the following equation. The best segmentation effect is obtained when the DSC is 1.

$$DSC = \frac{2 * |I_1 \cap I_2|}{|I_1| + |I_2|}$$

The segmentation network's performance is explained using a confusion matrix. Among them, TP indicates that the area is predicted and is actually a focus area. TN indicates that both predicted and actual tissues are normal. FP represents the predicted tumor area, which is actually normal tissue. On the contrary, FN indicates that it is predicted to be normal, but it is actually a tumor. In this experiment, accuracy (ACC), precision (Pre), recall (Re), and F1-score (F1) are calculated by using the confusion matrix to measure the performance of each network [56].

4.3. Algorithm Comparison

We perform a comparative experimental analysis with our proposed TBNet using the following method. These methods are briefly described below.

- (1) A fully convolutional network (FCN) is a pixel-level classification of images, using skip structures to achieve fine segmentation [57]. In this paper, 2 networks with 8 and 16 up-sampling are used FCN-8s and FCN-16s.

- (2) The PSPNet focuses on the pyramid pool module as a technique of extracting global contextual information, collecting and fusing contextual information at various scales, thus being particularly effective in acquiring global information [58].
- (3) The MSFCN is a fully convolutional network with many supervised lateral output layers for automatic volume segmentation [59]. To enable the effective learning of local and global visual features, a supervised lateral layer is added to the three layers of the convolutional network to provide a system-level structure to guide multidimensional feature learning.
- (4) Multi-scale residual networks (MSRN) [60] can adaptively identify image features at different scales and make them interact to produce effective high-resolution picture information. The generation of structural multiscale residual block MSRBs is achieved by combining convolution kernels of different sizes on the basis of the creation of residual blocks.
- (5) U-Net is a symmetric U-shaped structure that enables image-semantic-level segmentation [61]. The left side is a convolutional layer (systolic path) responsible for feature extraction and the right side is an up-sampling layer (extended path) responsible for feature reduction. It can use very little data to obtain the best results.
- (6) Feature pyramid networks (FPN) [62] use low-level feature semantic information with accurate target locations and information-rich high-level feature semantic information to make predictions independently at different feature layers using multi-scale feature fusion.
- (7) The UTNet network [24] hybrid transformer design incorporates self-attention into convolutional neural networks. The self-attention module is used in both the encoder and decoder in UTNet, and it is combined with relative position coding to considerably minimize the complexity of the self-attention process. In addition, a self-attentive decoder is presented to recover fine-grained features from the encoder's skipped connections, addressing the problem that the transformer requires a considerable quantity of data to acquire the visual sensing bias.
- (8) The TransUNet network [21] combines the advantages of a transformer and U-Net in a hybrid CNN transformer design. The global context extraction encodes the tokenized picture blocks in the CNN feature map as input sequences by the transformer. The decoder performs up-sampling to achieve accurate localization. This operation is performed before combining the encoded features with the high-resolution CNN feature map.

4.4. Influence of Super Parameters

In the preparation before training, to improve the efficiency of training and model performance, we first performed the pre-screening and noise-reduction operation of the data to avoid meaningless redundant and invalid data, which affect the improvement in segmentation efficiency. For the fixed parameters of the TSF in the pre-screening operation, referring to the general settings of the study, we set n_1 to 7 and the fixed threshold parameter λ_{FFT2} to 0.82.

In addition, for the setting of parameter λ_T in TSF, we found that the reasonable setting interval of λ_T should be 90–160 by repeating the experiment and statistically analyzing the lesion areas on the MRI images of different osteosarcomas. If λ_T is set too small or too large, the accuracy of screening is reduced. The former is because it causes the TSF to incorrectly classify lesioned images as normal, reducing the accuracy of the screening process. The latter is because the TSF will incorrectly label healthy images as lesions. The average classification accuracy is maximum when $120 < \lambda_T < 140$, which may reach 95.82%, and the optimal interval of λ_T can be defined as 120–140. The trend of classification accuracy (CA) can be shown in Figure 6 for a suitable range of parameters.

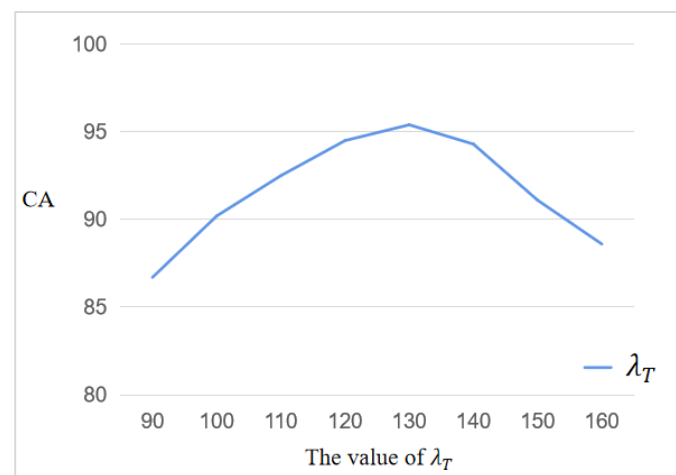


Figure 6. Trend diagram of CA with λ_T .

4.5. Results

From the MRI images screened by the TSF algorithm (the lesion images used for the experiment and the rejected normal images), 100 images were selected separately and the accuracy of the model was judged by the confusion matrix, as shown in Table 2. The predicted values indicate the results judged by the TSF algorithm. The classification results judged by the physician after examining the MRI images are represented by the actual values. Among them, the model correctly classified 93 normal images and lesion images, judged 3 lesion images as normal, and judged 4 normal images as lesion images. According to the analysis in Section 4.4, the best CA value of the TSF algorithm reached above 0.95. The CA value of the randomly selected results in Table 2 reached 0.93. It proves that the TSF algorithm is as expected and can effectively eliminate redundant data from lesion-free regions and filter out valuable training images.

Table 2. Analytical constant images of TSF algorithm screening results.

Predicted \ True	Lesion Image	Normal Image
	Lesion image	Normal image
Lesion image	49	4
Normal image	3	44

To verify the effectiveness of the denoising process, we used the evaluation metric PSNR to quantify the validity of this operation. Figure 7 shows the reconstruction results of the images before and after noise reduction with the corresponding PSNR values. By comparison, it can be observed that the post-noise-reduction image (Figure 7b) is clearer than the image before the noise-reduction treatment (Figure 7a) and is very close to the noise-free image (Figure 7c). Meanwhile, we can know that the PSNR value will be higher after the noise-reduction processing.

Specifically, Figure 8 shows the effect comparison of some randomly selected images before and after the noise-reduction process to prove that noise reduction has a great impact on the recognition accuracy of tumors. Comparing the segmentation effect before noise reduction (Figure 8B) and after noise reduction (Figure 8C), the segmentation effect is similar in the global view. However, we can visually see from the details of the local area magnification that the segmentation effect after noise reduction (Figure 8C) compensates for the detailed contour of the edge area of the lesion compared with the segmentation effect without noise reduction (Figure 8B), which is closer to the real image. The results demonstrate that noise-reduction processing increases accuracy.

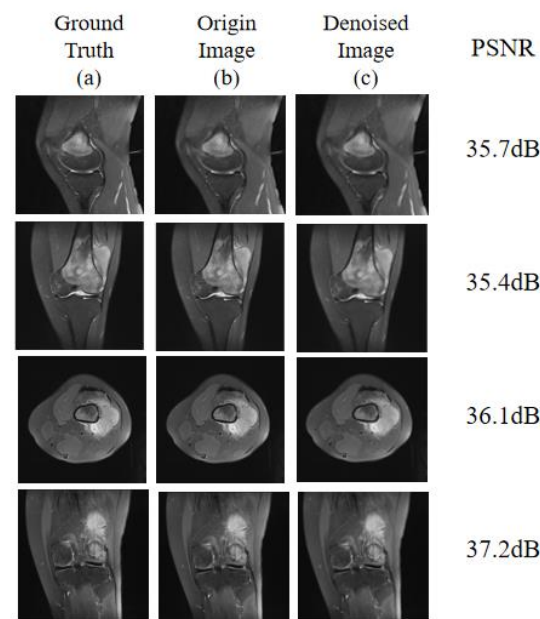


Figure 7. (a) Noise-free image (ground truth), (b) origin image before noise reduction, (c) reconstructed image after noise reduction.

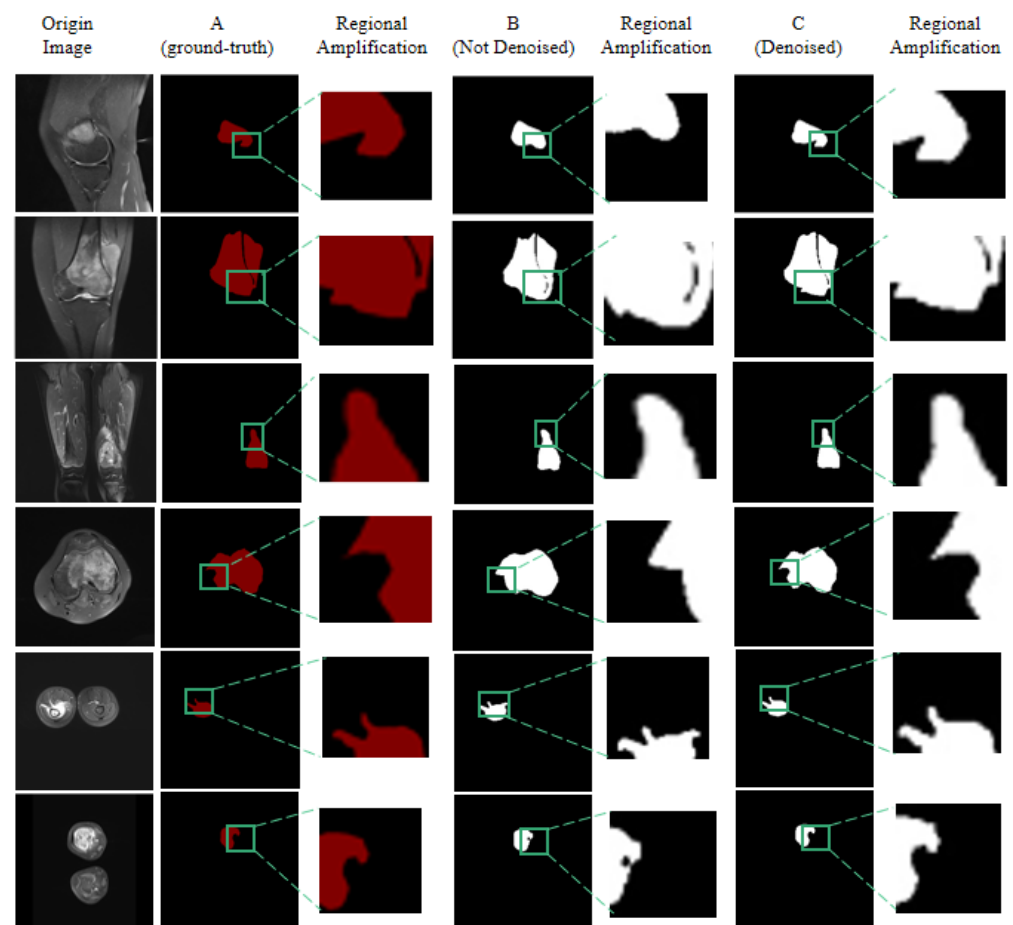


Figure 8. The segmentation effect before and after noise reduction. The sub-figure (A) represents the ground truth of the image, the sub-figure (B) represents the result of image segmentation without Denoised, and the sub-figure (C) represents the result of image segmentation with Denoised.

In Table 3, we further visualize the effect of the TSF filtering algorithm ($\lambda_T = 133$) and the NLM noise-reduction algorithm on the performance of the model. When the original MRI images were processed directly by the TBNNet network, all the other indexes performed poorly, although its Re value reached 0.974. When images without tumor regions (or tumor regions barely visible) were removed using the TSF method, the performance of our model improved more significantly. Precision increased by around 0.009 after image pre-screening, F1 increased by about 0.003, recall increased by about 0.001, and DSC increased by about 0.006. The most important DSC index improved by roughly 0.012 after the noise-reduction treatment, whereas precision, F1, and IOU improved by 0.005, 0.002, and 0.011, respectively.

Table 3. Comparison of TBNNet under different conditions.

Model	ACC	IOU	DSC	Pre	Re	F1
TBNNet	0.991	0.904	0.931	0.927	0.974	0.949
TBNNet+Denoise	0.993	0.915	0.943	0.932	0.968	0.951
TBNNet+Denoise+Pre-Screening	0.997	0.915	0.949	0.941	0.969	0.954

After the screening of the dataset by Threshold Screening Filter (TSF), more than 3000 MRI images remained. All comparison methods, including UNet, were experimented with using the pre-processed images to ensure fairness. The results of each model for the segmentation of osteosarcoma MRI images are shown in Figure 9 below. We can visually examine the model's segmentation performance using ground-truth photographs. Meanwhile, we have selected the DSC metrics, and based on the following six segmentation examples of osteosarcoma, we can find that the TBNNet network can obtain better segmentation results. When dealing with simple segmentation jobs, our technique can obtain the same segmentation outcomes as previous models. It performs better when dealing with tumor segmentation jobs with complex boundaries (cases 3 and 4). Our method, particularly in the specifics of the border contour of the osteosarcoma lesion, can obtain a more precise and comprehensive segmentation of the tumor borders with higher accuracy when compared to other methods that can handle it appropriately.

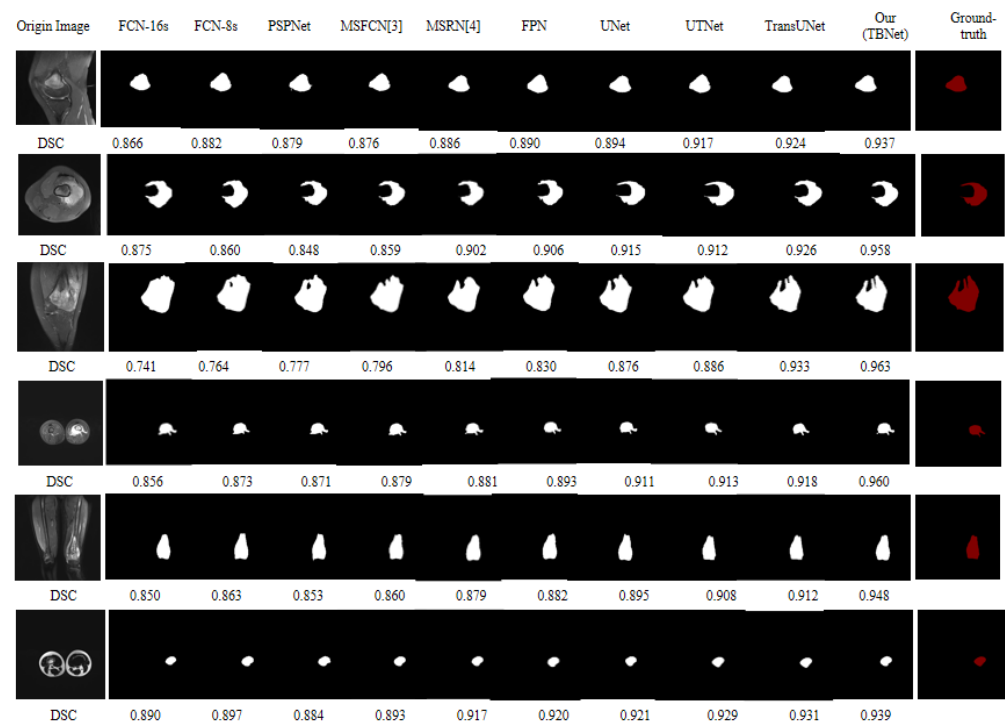


Figure 9. Comparison of different models for the segmentation of MRI images of osteosarcoma.

Similarly, for the same osteosarcoma images with complex edge blurring characteristics as in cases 3 and 4 above, we selected the following six examples of osteosarcoma images with blurred edges and used TransUNet, and our proposed TBNNet for image segmentation, respectively. As shown in Figure 10, our method can segment the boundary of the lesion region more effectively and precisely.

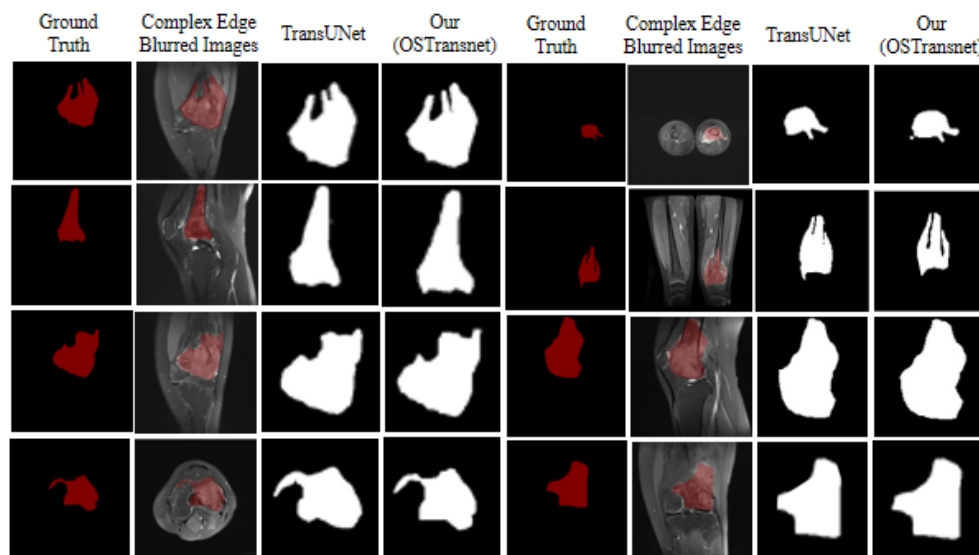


Figure 10. Segmentation effect with blurred edges.

We measured segmented data to quantitatively compare the effectiveness of the various methods by comparing several classical comparative metrics. The osteosarcoma task is shown in Table 4. Our proposed model TBNNet performs well overall in the task. Several metrics of our scheme are the highest. For example, ACC, IOU, DSC, and Re reach 0.997, 0.915, 0.949, and 0.969, respectively. The performance of the UNet network is also relatively good with an ACC value of 0.99. The MSRN network has the least params among all methods with only 12.42 M, and it also has a recall of 0.945. Comparatively, although the MSRN network also requires only 24.53 M parameters, it has significantly lower DSC values. For the PSPNet algorithm, it has the lowest FLOPS value but it has the worst IOU and DSC values. It indicates that the performance of this model is poor. Although our method has more parameters than UNet, MSFCN, and MSRN models, it has fewer parameters than some other convolution-based models. It has higher segmentation accuracy and faster computation speed with only a small number of additional parameters.

Table 4. Performance comparison of different models on the osteosarcoma task.

Model	ACC	IOU	DSC	Pre	Re	F1	FLOPS	Params
FCN-16s	0.989	0.824	0.859	0.922	0.882	0.900	187.35 G	122.4 M
FCN-8s	0.993	0.830	0.876	0.941	0.873	0.901	187.18 G	122.4 M
PSPNet	0.975	0.772	0.870	0.856	0.888	0.872	103.55 G	47.70 M
MSFCN	0.991	0.841	0.874	0.881	0.936	0.906	1642.43 G	24.53 M
MSRN	0.988	0.853	0.887	0.893	0.945	0.918	1346.12 G	12.42 M
FPN	0.989	0.852	0.888	0.914	0.924	0.919	134.14 G	47.82 M
UNet	0.990	0.867	0.892	0.922	0.924	0.923	160.16 G	17.26 M
UTNet	0.990	0.879	0.919	0.924	0.934	0.936	264.27 G	52.53 M
TransUNet	0.993	0.898	0.923	0.919	0.959	0.948	240.26 G	40.38 M
Our(TBNNet)+Denoise+Pre-Screening	0.997	0.915	0.949	0.941	0.969	0.954	235.26 G	36.51 M

Each model was trained for a total of 200 rounds, and 1 round was selected every 4 rounds for visualization. As can be seen from Figure 11, our method is the fastest to reach the best stability in terms of accuracy, recall, and F1 value. In Figure 11a, the accuracy

ranking among the models is ours > TransUNet > Unet > FPN > MSRN > MSFCN. In Figure 11b, overall, the recall rate of our suggested method has been kept as high as possible, ensuring that missed diagnoses are avoided to the greatest extent possible. Finally, the F1-score of each model in the training process is compared with our method. From Figure 11c, we can see that our F1 value fluctuates to some extent, but its average value remains the highest. It can be seen that the model is more robust and accurate in processing MRI images of osteosarcoma with different morphological characteristics.

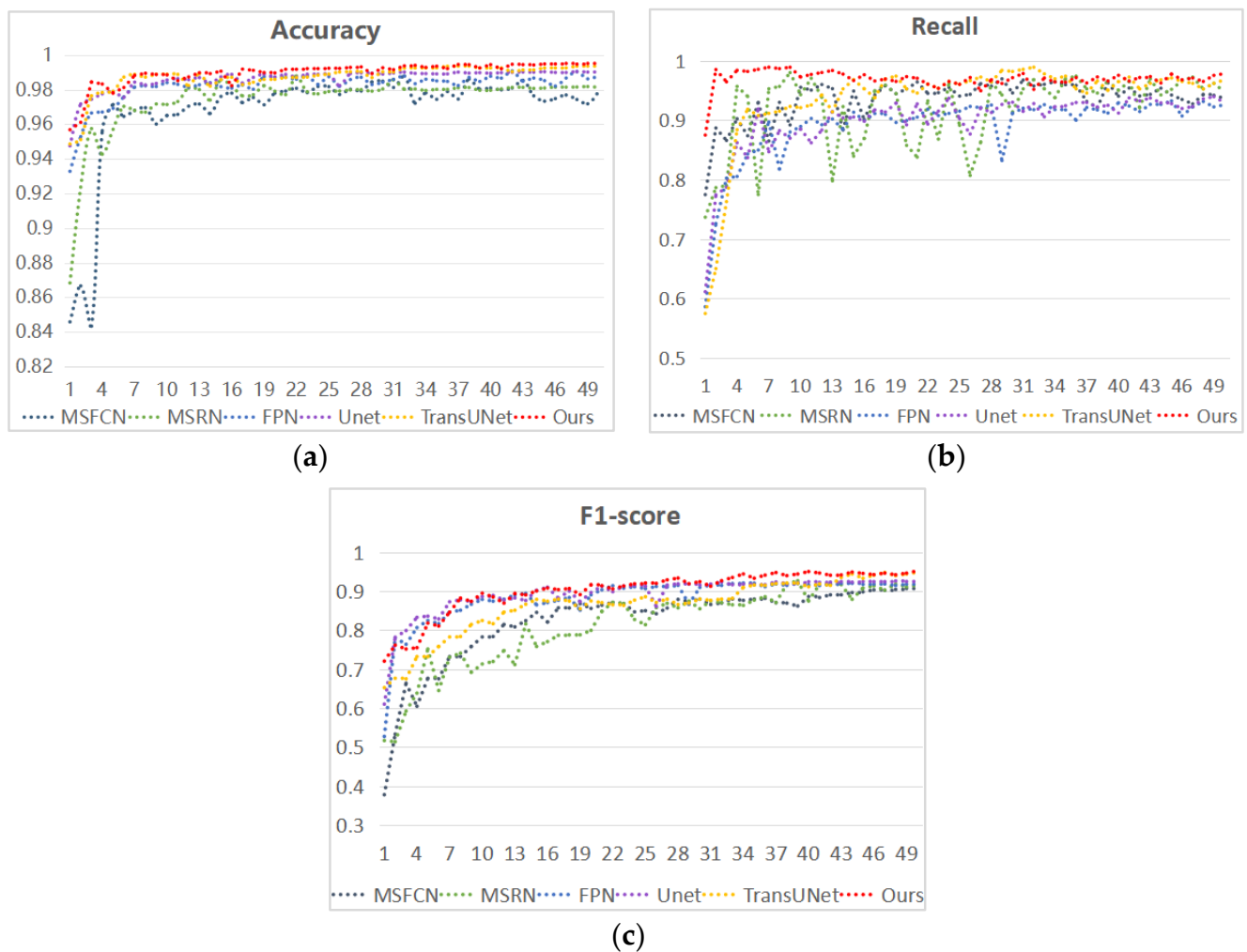


Figure 11. The trend of different metrics of each network during the training process. (a) is the trend of model accuracy, (b) is the trend of model recall, and (c) is the trend of model F1-score values.

4.6. Discussion

It is shown that the pre-screened TSF shows a robust and accurate classification capability, effectively filters redundant data, and improves segmentation performance while saving the running time of subsequent processing operations. The fast FFT-based NLM algorithm effectively improves image segmentation and largely reduces the impact of noise in osteosarcoma diagnosis.

We designed TBNNet, a U-Net network with a fusion transformer and edge enhancement, for the osteosarcoma MRI Image segmentation problem. Several comparison studies reveal that our model outperforms others in terms of assessment indices such as DSC and IOU. It offers significant benefits over other classical convolutional models for the proper processing of contour features in the complicated tumor border problem. In addition, comparative tests in terms of accuracy, recall, and F1 by systematic sampling of random samples show that our model has greater stability. This ensures the stability and

accuracy of segmentation results for practical diagnosis. This is mainly due to the introduction of a cross-fusion channel transformer with edge-enhancement features to replace the original U-Net jump connection in this paper. This approach takes advantage of both a transformer and Unet to perform the cross-fusion of multi-scale channel information, bridge the semantic gap, and solve the problem of semantic hierarchy inconsistency. In addition, the algorithm improves the problem of fuzzy edge segmentation by enhancing the edge information. Therefore, compared with the classical UNet model, the recognition accuracy of our method is higher. To apply to developing countries, our model reduces the computational cost as much as possible while ensuring accuracy and stability.

5. Conclusions

In this study, an artificial intelligence-aided diagnosis system for osteosarcoma MRI pictures with edge-enhancement characteristics is suggested. The method employs strategies such as image pre-screening, noise-reduction processing, and a segmentation model using edge-enhancement characteristics. It is used mainly to improve the recognition accuracy of osteosarcoma and assist doctors to detect the lesion location of patients more quickly and accurately. Our model is more robust in processing the MRI images of osteosarcoma with different shapes.

However, the recognition accuracy of the TBNNet model is relatively low for images without noise reduction and pre-screening. With the development of image processing technology, we will strive to improve the recognition accuracy of tumors in original images. In addition, improving the function of the auxiliary scheme in combination with the actual clinical performance is also the focus of research.

Author Contributions: Conceptualization, B.L., F.L. and J.W.; Data curation, B.L., Y.L. and J.N.; Formal analysis, F.G. and J.W. 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, 2020-NDGL-19.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data analyzed during the current study are included in the submission. 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 the publication of this article, will be considered by the corresponding author.

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

References

1. Corre, I.; Verrecchia, F.; Crenn, V.; Redini, F.; Trichet, V. The Osteosarcoma Microenvironment: A Complex but Targetable Ecosystem. *Cells* **2020**, *9*, 976. [[CrossRef](#)] [[PubMed](#)]
2. 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](#)] [[PubMed](#)]
3. Gao, S.-S.; Wang, Y.-J.; Zhang, G.-X.; Zhang, W.-T. Potential diagnostic value of miRNAs in peripheral blood for osteosarcoma: A meta-analysis. *J. Bone Oncol.* **2020**, *23*, 100307. [[CrossRef](#)]
4. Wu, J.; Guo, Y.; Gou, F.; Dai, Z. A medical assistant segmentation method for MRI images of osteosarcoma based on DecoupleSeg-Net. *Int. J. Intell. Syst.* **2022**, *37*, 8436–8461. [[CrossRef](#)]
5. 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](#)] [[PubMed](#)]
6. Chouhan, S.S.; Kaul, A.; Singh, U.P. Image Segmentation Using Computational Intelligence Techniques: Review. *Arch. Comput. Methods Eng.* **2019**, *26*, 533–596. [[CrossRef](#)]
7. Heo, Y.-C.; Kim, K.; Lee, Y. Image Denoising Using Non-Local Means (NLM) Approach in Magnetic Resonance (MR) Imaging: A Systematic Review. *Appl. Sci.* **2020**, *10*, 7028. [[CrossRef](#)]
8. 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](#)]

9. 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. [\[CrossRef\]](#)
10. Gou, F.; Wu, J. Data Transmission Strategy Based on Node Motion Prediction IoT System in Opportunistic Social Networks. *Wirel. Pers. Commun.* **2022**, *126*, 1751–1768. [\[CrossRef\]](#)
11. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
12. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Shen, Y.; Gou, F.; Dai, Z. Osteosarcoma MRI Image-Assisted Segmentation System Base on Guided Aggregated Bilateral Network. *Mathematics* **2022**, *10*, 1090. [\[CrossRef\]](#)
14. 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. [\[CrossRef\]](#)
15. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
16. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Jiao, Y.; Qi, H.; Wu, J. Capsule network assisted electrocardiogram classification model for smart healthcare. *Biocybern. Biomed. Eng.* **2022**, *42*, 543–555. [\[CrossRef\]](#)
18. 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. [\[CrossRef\]](#)
19. Lv, B.; Liu, F.; Gou, F.; Wu, J. Multi-Scale Tumor Localization Based on Prior Guidance-Based Segmentation Method for Osteosarcoma MRI Images. *Mathematics* **2022**, *10*, 2099. [\[CrossRef\]](#)
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. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
22. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *arXiv* **2021**, arXiv:2103.14030.
23. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical transformer: Gated axial-attention for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021, Proceedings, Part I 24*; Springer International Publishing: New York, NY, USA, 2021; pp. 36–46.
24. Gao, Y.; Zhou, M.; Metaxas, D. Utnet: A Hybrid Transformer Architecture for Medical Image Segmentation. *arXiv* **2021**, arXiv:2107.00781.
25. Ji, Y.; Zhang, R.; Wang, H.; Li, Z.; Wu, L.; Zhang, S.; Luo, P. Multi-compound transformer for accurate biomedical image segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1 2021, Proceedings, Part I 24*; Springer International Publishing: New York, NY, USA, 2021; pp. 326–336.
26. Petit, O.; Thome, N.; Rambour, C.; Soler, L. U-net transformer: Self and cross attention for medical image segmentation. *arXiv* **2021**, arXiv:2103.06104.
27. Zhang, Y.; Liu, H.; Hu, Q. TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021, Proceedings, Part I 24*; Springer International Publishing: New York, NY, USA, 2021; pp. 14–24. [\[CrossRef\]](#)
28. Heidari, M.; Kazerouni, A.; Soltany, M.; Azad, R.; Aghdam, E.K.; Cohen-Adad, J.; Merhof, D. HiFormer: Hierarchical Multi-scale Representations Using Transformers for Medical Image Segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 6191–6201. [\[CrossRef\]](#)
29. Wei, C.; Ren, S.; Guo, K.; Hu, H.; Liang, J. High-Resolution Swin Transformer for Automatic Medical Image Segmentation. *arXiv* **2022**, arXiv:2207.11553.
30. Sun, R.; Pang, Y.; Li, W. Efficient Lung Cancer Image Classification and Segmentation Algorithm Based on an Improved Swin Transformer. *Electronics* **2023**, *12*, 1024. [\[CrossRef\]](#)
31. Valanarasu, J.M.J.; Patel, V.M. UNeXt: MLP-Based Rapid Medical Image Segmentation Network. *arXiv* **2022**, arXiv:2203.04967.
32. Tran, M.; Vo-Ho, V.; Quinn, K.; Nguyen, H.V.; Luu, K.; Le, N.T. CapsNet for Medical Image Segmentation. *arXiv* **2022**, arXiv:2203.08948.
33. Anisuzzaman, D.; Barzekar, H.; Tong, L.; Luo, J.; Yu, Z. A deep learning study on osteosarcoma detection from histological images. *Biomed. Signal Process. Control.* **2021**, *69*, 102931. [\[CrossRef\]](#)

34. Fu, Y. Deep Model with Siamese Network for Viability and Necrosis Tumor Assessment in Osteosarcoma. *Med. Phys.* **2019**, *47*, 4895–4905. [\[CrossRef\]](#)
35. Barzakar, H.; Yu, Z. C-Net: A Reliable Convolutional Neural Network for Biomedical Image Classification. *arXiv* **2020**. [\[CrossRef\]](#)
36. Nabid, R.; Rahman, M.; 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.
37. D’Acunto, M.; Martinelli, M.; Moroni, D. From human mesenchymal stromal cells to osteosarcoma cells classification by deep learning. *J. Intell. Fuzzy Syst.* **2019**, *37*, 7199–7206. [\[CrossRef\]](#)
38. Parlak, Ş.; Ergen, F.B.; Yüksel, G.Y.; Karakaya, J.; Aydın, G.B.; Kösemehmetoğlu, K.; Aydingöz, Ü. Diffusion-weighted imaging for the differentiation of Ewing sarcoma from osteosarcoma. *Skelet. Radiol.* **2021**, *50*, 2023–2030. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Ho, D.J.; Agaram, N.P.; Schüffler, P.J.; Vanderbilt, C.M.; Jean, M.-H.; Hameed, M.R.; Fuchs, T.J. Deep Interactive Learning: An Efficient Labeling Approach for Deep Learning-Based Osteosarcoma Treatment Response Assessment. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2020; pp. 540–549.
40. Im, H.-J.; Solaiyappan, M.; Lee, I.; Bradshaw, T.; Daw, N.C.; Navid, F.; Shulkin, B.L.; Cho, S.Y. Multi-level otsu method to define metabolic tumor volume in positron emission tomography. *Am. J. Nucl. Med. Mol. Imaging* **2018**, *8*, 373–386. [\[PubMed\]](#)
41. 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, 8–20 August 2021; pp. 93–99. [\[CrossRef\]](#)
42. Huang, W.-B.; Wen, D.; Yan, Y.; Yuan, M.; Wang, K. Multi-target osteosarcoma MRI recognition with texture context features based on CRF. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 3978–3983. [\[CrossRef\]](#)
43. Gou, F.; Wu, J. Novel data transmission technology based on complex IoT system in opportunistic social networks. *Peer-to-Peer Netw. Appl.* **2022**, *185*, 1–18. [\[CrossRef\]](#)
44. Xiong, W.; Chen, H.; Jiao, Y.; Yang, M.; Zhou, X. A user cache management and cooperative transmission mechanism based on edge community computing in opportunistic social networks. *IET Commun.* **2022**, *16*, 2045–2058. [\[CrossRef\]](#)
45. 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. [\[CrossRef\]](#)
46. 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. [\[CrossRef\]](#)
47. Gou, F.; Wu, J. Triad link prediction method based on the evolutionary analysis with IoT in opportunistic social networks. *Comput. Commun.* **2021**, *181*, 143–155. [\[CrossRef\]](#)
48. Han, K.; Hhan, M.; Cheon, J.H. Improved Homomorphic Discrete Fourier Transforms and FHE Bootstrapping. *IEEE Access* **2019**, *7*, 57361–57370. [\[CrossRef\]](#)
49. Deng, Y.; Gou, F.; Wu, J. Hybrid data transmission scheme based on source node centrality and community reconstruction in opportunistic social networks. *Peer-to-Peer Netw. Appl.* **2021**, *14*, 3460–3472. [\[CrossRef\]](#)
50. Wu, J.; Yu, L.; Gou, F. Data transmission scheme based on node model training and time division multiple access with IoT in opportunistic social networks. *Peer-to-Peer Netw. Appl.* **2022**, *15*, 2719–2743. [\[CrossRef\]](#)
51. 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. [\[CrossRef\]](#)
52. Yu, G.; Wu, J. Efficacy prediction based on attribute and multi-source data collaborative for auxiliary medical system in developing countries. *Neural Comput. Appl.* **2022**, *34*, 5497–5512. [\[CrossRef\]](#)
53. Suriyan, K.; Ramaingam, N.; Rajagopal, S.; Sakkarai, J.; Asokan, B.; Alagarsamy, M. Performance analysis of peak signal-to-noise ratio and multipath source routing using different denoising method. *Bull. Electr. Eng. Infor-Matics* **2022**, *11*, 286–292. [\[CrossRef\]](#)
54. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
55. 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. [\[CrossRef\]](#)
56. Liu, F.; Gou, F.; Wu, J. An Attention-Preserving Network-Based Method for Assisted Segmentation of Osteosarcoma MRI Images. *Mathematics* **2022**, *10*, 1665. [\[CrossRef\]](#)
57. 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.
58. 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.
59. 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. [\[CrossRef\]](#)
60. 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. [\[CrossRef\]](#)

61. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. [[CrossRef](#)]
62. 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.

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.