

Research Article

BA-GCA Net: Boundary-Aware Grid Contextual Attention Net in Osteosarcoma MRI Image Segmentation

Jia Wu ^{1,2,3,4}, Zikang Liu,¹ Fangfang Gou,¹ Jun Zhu ^{3,4}, Haoyu Tang,^{3,4} Xian Zhou ⁵,
and Wangping Xiong ⁵

¹School of Computer Science and Engineering, Central South University, Changsha 410083, China

²Research Center for Artificial Intelligence, Monash University, Melbourne, Clayton VIC 3800, Australia

³The First People's Hospital of Huaihua, Huaihua, Hunan, China

⁴Collaborative Innovation Center for Medical Artificial Intelligence and Big Data Decision Making Assis-10 Tance, Hunan University of Medicine, Changsha, China

⁵Jiangxi University of Chinese Medicine, Nanchang 330004, JiangXi, China

Correspondence should be addressed to Jia Wu; jiawu@ieee.org and Jun Zhu; zhujun190301@163.com

Received 30 April 2022; Revised 22 June 2022; Accepted 5 July 2022; Published 30 July 2022

Academic Editor: Jianli Liu

Copyright © 2022 Jia Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Osteosarcoma is one of the most common bone tumors that occurs in adolescents. Doctors often use magnetic resonance imaging (MRI) through biosensors to diagnose and predict osteosarcoma. However, a number of osteosarcoma MRI images have the problem of the tumor shape boundary being vague, complex, or irregular, which causes doctors to encounter difficulties in diagnosis and also makes some deep learning methods lose segmentation details as well as fail to locate the region of the osteosarcoma. In this article, we propose a novel boundary-aware grid contextual attention net (BA-GCA Net) to solve the problem of insufficient accuracy in osteosarcoma MRI image segmentation. First, a novel grid contextual attention (GCA) is designed to better capture the texture details of the tumor area. Then the statistical texture learning block (STLB) and the spatial transformer block (STB) are integrated into the network to improve its ability to extract statistical texture features and locate tumor areas. Over 80,000 MRI images of osteosarcoma from the Second Xiangya Hospital are adopted as a dataset for training, testing, and ablation studies. Results show that our proposed method achieves higher segmentation accuracy than existing methods with only a slight increase in the number of parameters and computational complexity.

1. Introduction

Osteosarcoma is the most common malignant bone tumor which occurs most frequently in children and adolescents between the ages of 10 and 30 years, with the highest incidence during the adolescent growth spurt [1]. In the past few years, neoadjuvant chemotherapy and biosensors have greatly developed, making the treatment of osteosarcoma easier. Nevertheless, without an early diagnosis, patients with advanced osteosarcoma will develop metastasis and recurrence disease, whose 5-year survival rate still keeps less than 20% [2, 3]. Therefore, how to clearly and accurately diagnose osteosarcoma has become the key to prevention and treatment.

Magnetic resonance imaging (MRI) can display the structure of soft tissue clearly and has higher contrast as well as resolution than other imaging methods [4], which makes the tumor area easier to distinguish. It is considered to be the best imaging method to evaluate the relationship between the primary osteosarcoma lesion and its surrounding areas [5]. Traditionally, the diagnosis of osteosarcoma is based on manual histopathological analysis with biosensors on MRI images by doctors. However, it has great disadvantages. In developing countries where the medical level is relatively backward, the doctor-patient ratio remains low, with each doctor handling the diagnosis and treatment of about 60 patients per day on average [6–8]. In addition, one patient will produce more than 600 MRI images during one

diagnosis with biosensors, making analysis laborious and time-consuming [9–14]. To make matters worse, doctors' high-intensity work makes their manual judgments susceptible to inter- and intra-observer variations and results in inaccurate segmentation of osteosarcoma areas [15–17]. Furthermore, due to the heterogeneity of osteosarcoma [1], sophisticated diagnoses using MRI biosensors often require experienced radiologists, which is extraordinarily challenging for some developing countries with backward allocation of medical resources [18–20].

In order to solve the problems of manually segmenting lesions by doctors, researchers have designed a variety of automatic segmentation models and applied them to medical image segmentation to help doctors diagnose diseases and predict lesion areas, thus reducing the pressure on medical resources in developing countries and enabling diagnosis of diseases to achieve high accuracy at low computational costs. A fully convolutional network (FCN) [21] uses skip layers to achieve end-to-end and joint learning of semantic as well as location. It is the most classic model in medical image segmentation. U-Net [22] crops the output feature maps from shallow layers and concatenates them to the ones from deep layers to fuse and exploit the low-level and high-level features, improving the network's performance on neural structure and cell segmentation. In the specific field of osteosarcoma segmentation, [23] uses a recurrent convolutional neural network (RCNN) combining CNN and GRU and achieves better performance with a small number of histopathological osteosarcoma images. MSFCN [24] and MSRN [25] add multiple supervised structures to the network to promote learning and improve the overall osteosarcoma segmentation accuracy.

Considerable progress has been made in the research on osteosarcoma segmentation models. However, a few segmentation problems in MRI images using biosensors have unfortunately been overlooked: (i) The boundaries between osteosarcoma and normal tissues in some images are not clear enough and the lesion area is indistinguishable from other soft tissues. Therefore, the low-contrast boundaries may be blurred during convolution operations, resulting in segmentation failure. (ii) In transverse section images, the osteosarcoma area is often small, and the model is prone to spatial shift in the process of down-sampling and up-sampling, which leads to difficulty in localization and decrease in accuracy. (iii) Some osteosarcoma images have complicated and irregular shape boundaries, and the model cannot identify small gaps between osteosarcoma and normal tissues, causing the identification of the entire region as a lesion area and the loss of segmentation details. These problems contribute to poor performance on a number of difficult tasks with vague foreground-background boundaries or small and complex osteosarcoma regions, which have become a significant factor that affects the accuracy of segmentation.

In order to solve the problems mentioned above, we propose a novel boundary-aware grid contextual attention network (BA-GCA Net), which effectively improves the performance of the network on MRI images with blurred osteosarcoma boundaries and complex foreground

structure. First, we propose a plug-and-play grid contextual attention structure. The structure splits the input feature map into patches, exploits the local contextual information to learn the positional features inside the image patches, and enhances the network's capability to capture the details of osteosarcoma boundaries and texture. For the problem that certain osteosarcoma MRI images using biosensors have intricate shape boundaries or fuzzy tumor texture, a statistical texture learning block (STLB) is integrated into the network. STLB learns the low-level features and applies them to the task. The texture enhancement module (TEM) in the STLB first enhances the texture in the low-level feature map and produces a clearer texture map, which is conducive to more accurate segmentation of the osteosarcoma region. Then the pyramid texture feature extraction module (PTFEM) is used in the STLB to further extract and utilize the enhanced texture. Due to the small tumor areas in some osteosarcoma MRI images, the subtle spatial shift of the prediction map will lead to poor segmentation performance. To this end, we use a spatial transformer block (STB) in the network to make it invariant to spatial shifts. STB localizes and regresses the input feature map, learns an affine transformation matrix, and applies the transformation to the feature map. It spatially adjusts the prediction map so that the positioning of osteosarcoma areas is more accurate.

In general, our contributions can be summarized as follows:

- (1) We propose a novel BA-GCA Net, which can better learn detailed features in the input image and fully exploit texture features in a spatially invariant way to improve the segmentation accuracy of osteosarcoma MRI images.
- (2) In order to pay more attention to local details in the input image, we propose a plug-and-play grid contextual attention (GCA) structure, which reshapes the image into patches and applies local and global contextual attention to them to enhance the perception of local details in osteosarcoma areas.
- (3) Inspired by U-Net, low-level features in the input image have rich texture details and therefore a statistical texture learning block (STLB) is used to learn texture features in the low-level feature map and utilize them in deeper layers of the network, improving the segmentation accuracy in tasks where the tumor area has blurred boundaries or small gaps.
- (4) To improve the model's ability in locating the osteosarcoma lesion area, we use an STB in the network to learn an affine transformation matrix and adjust the prediction map. Moreover, a boundary loss is designed in the loss function to facilitate STB to learn positional boundary information, which promotes the segmentation performance on images where the region of osteosarcoma is small.
- (5) Over 80,000 MRI images of osteosarcoma output by biosensors from the Second Xiangya Hospital are adopted as a dataset for experiments. Results show that our proposed method outperforms other models

on accuracy with only a slight increase in the number of parameters and computational complexity compared with the backbone network and achieves a balance in terms of accuracy and computational efficiency, which is helpful to doctors in judging the osteosarcoma lesion area and reducing workload.

2. Related Works

2.1. Osteosarcoma Image Segmentation. Accurate diagnosis and prediction of osteosarcoma are the keys to increasing the survival rate of the patients and making precise follow-up treatment plans. Numerous researchers have studied in osteosarcoma image segmentation before. Reference [26] uses similarity mapping and slope value to analyze the time-intensity curves of regions of interest (ROI) and fuses the anatomic information of traditional MRI sequences with the numerical information of dynamic MRI sequences to obtain a better description of osteosarcoma regions. Reference [27] proposes a dynamic clustering algorithm DCHS based on Harmony Search (HS) and Fuzzy C-means (FCM) to automatically segment osteosarcoma MRI images, using a subset of Haralick texture features and pixel intensity values as a feature space to DCHS to delineate tumor volume, achieving a Dice measurement of an average of 0.72.

With the rapid development of deep learning and computer vision [28], a great number of deep learning-based models have been designed by researchers in the segmentation of osteosarcoma images as auxiliary diagnosis methods. Multiple supervised fully convolutional network (MSFCN) [24] adds supervision layers to the output layers of different sizes in the VGG model and uses the output information of the multiple supervision layers to produce the prediction map. Multiple supervised residual network (MSRN) [25] integrates residual structure into the network on the basis of multiple supervised structure, which improves the performance of a deep neural network on osteosarcoma image segmentation tasks. Wnet++ [29] uses two cascaded U-Nets and dense skip connections to realize automatic segmentation of tumor areas. In addition, Wnet++ adopts multi-scale input to alleviate information loss caused by down-sampling and introduces an attention mechanism to better represent tumor features, which increases the accuracy of segmentation. In osteosarcoma MRI image segmentation, there are often blurred foreground-background boundaries or small and complex segmentation regions. Therefore, the local semantic details and boundary information are of particular importance. Different from the above methods, we design BA-GCA Net, which embeds modules into the semantic segmentation framework to enhance the model's ability to extract rich local semantic, texture statistics, and boundary information and improves the model in performance on intricate segmentation tasks.

2.2. Boundary Prediction Enhancement Methods. In osteosarcoma MRI image segmentation, the prediction of boundaries is crucial to the model's performance. Previously, a great number of researchers in the field of medical

image segmentation have devoted to solving the problem of segmentation boundaries [23–28]. Structure boundary preserving segmentation [30] obtains the structured boundary information of an image through a key point selection algorithm, a boundary preserving block, and a shape boundary-aware evaluator. BFP [31] utilizes a boundary-aware feature propagation module to transfer low-level boundary information. InverseForm [32] enables boundary loss function to learn spatial transformation distance through a pretrained inverse transformation network. Some other works [33–35] have improved the boundary loss function and achieved good results.

Unfortunately, the above methods neither take full advantage of the rich low-level statistical boundary texture features of the input image nor solve the problem of spatial shift of the prediction boundary that may exist in small osteosarcoma lesion areas. Unlike the above methods, we use a statistical texture learning block (STLB) [36] to quantify and count the low-level texture information output by the shallow layers in the network. Due to the segmentation of some small tumor areas in the osteosarcoma MRI images, spatial shifts may occur during down-sampling and up-sampling. Therefore, we integrate the STB [37] into the deep layer of the network to enhance the spatial transformation invariance. Combined with the boundary loss function, STB will automatically learn the spatial shift of the prediction map and adjust it adaptively.

2.3. Attention Mechanisms. Attention mechanism has been proved to raise the model's capability of giving more weight to useful features to improve semantic analysis and has achieved good results in a variety of computer vision tasks [38–42].

SENet [38] compresses feature maps in spatial dimension and generates a channel-wise attention. Based on SENet, CBAM [39] additionally introduces a spatial attention through channel pooling and large-scale convolution and has certain improvements in classification and detection. SANet [40] divides the segmentation task into two subtasks that are pixel-level prediction and pixel grouping and combines multi-scale prediction and pixel-grouping spatial attention to improve performance. OCR [41] learns the relationship between pixel and object region features based on coarse segmentation maps and enhances the description of pixel features. Coordinate attention (CA) [42] rethinks the attention mechanism and produces the attention map with positional information by compressing the spatial features of the image into attention weights in horizontal and vertical directions.

The above methods can extract image context in an efficient way, but cannot pay extra attention to the local details of the image that are indispensable for pixel-to-pixel osteosarcoma MRI image segmentation. Different from the above methods, our proposed grid contextual attention (GCA) combines local and global contextual attention, which can exploit the global contextual features of the feature map and learn local contextual features in the meantime.

3. Methods

As is mentioned above, the shortage of medical resources and the backward medical level in some developing countries make diagnosis of osteosarcoma more formidable. Moreover, the blurred and low-contrast tumor areas as well as small and intricate structural boundaries may lead to fuzzy or even wrong prediction of the lesion region from automatic segmentation models and influence clinical diagnoses. To this end, we add GCA, STLB, and STB to the network to reduce segmentation errors in osteosarcoma MRI images using biosensors while keeping the number of parameters and computational complexity at a low level to ensure low diagnostic cost. The overall structure of the network is shown in Figure 1.

For an osteosarcoma image generated by MRI biosensors, it is first fed into the backbone to extract high-level and low-level features. We integrate GCA at the top of the ResNet building blocks after the first two layers to improve the feature extraction ability of the model. The low-level features produced by the first two layers are fed into STLB to enhance and analyze texture statistics. The output of STLB is concatenated in channel dimension with the high-level features from the backbone added with GCA. Then the fused output is fed into STB to perform an affine transformation to the prediction map and produce the final output. Canny [43] operator is applied to the prediction result and ground truth to extract the boundaries. Segmentation loss and boundary loss are calculated using segmentation masks and boundaries, respectively, to form the compound loss function.

3.1. Grid Contextual Attention. In the pixel-wise osteosarcoma MRI image segmentation, understanding the local details in the image often helps in more accurate segmentation as well as less uncertain prediction. Therefore, we design a grid contextual attention (GCA) structure based on both local and global contextual features. The structure is shown in Figure 2.

For an input feature map $X \in R^{C \times H \times W}$, its global contextual information is obtained by:

$$\begin{aligned} \hat{A}^h &= \text{Excit}(\text{Avg}_w(X)), \\ \hat{A}^w &= \text{Excit}(\text{Avg}_h(X)), \end{aligned} \quad (1)$$

where $\hat{A}^h \in R^{C \times H \times 1}$, $\hat{A}^w \in R^{C \times 1 \times W}$, Avg_h , and Avg_w denote average pooling of the feature map in height and width directions, respectively, and Excit denotes the activation transformation of the input as:

$$\text{Excit}(x) = \text{Conv}_{\text{Sigmoid}}(\text{Conv}_{\text{RELU}}(x)). \quad (2)$$

In the part of local attention, GCA splits the feature map into patches, each of which is denoted as $P_{i,j} \in R^{C \times P_h \times P_w}$, where P_h and P_w represent the patch size in height and width directions and $i \in \{1, 2, \dots, (H/P_h)\}$, $j \in \{1, 2, \dots, (W/P_w)\}$.

Each patch $P_{i,j}$ is passed through average pooling in the width and the height directions respectively to get local attention $A_{i,j}^h \in R^{C \times P_h \times 1}$ and $A_{i,j}^w \in R^{C \times 1 \times P_w}$. Then the patches are concatenated by:

$$\begin{aligned} A_j^h &= \text{Concat}(A_{1,j}^h, A_{2,j}^h, \dots, A_{H/P_h,j}^h), \\ A_i^w &= \text{Concat}(A_{i,1}^w, A_{i,2}^w, \dots, A_{i,W/P_w}^w), \end{aligned} \quad (3)$$

where $i \in \{1, 2, \dots, (H/P_h)\}$, $j \in \{1, 2, \dots, (W/P_w)\}$, $A_j^h \in R^{C \times H \times 1}$, and $A_i^w \in R^{C \times 1 \times W}$.

In order to enable localized patches to get global contextual information, each A_j^h and A_i^w are performed element-wise product with \hat{A}^h and \hat{A}^w , respectively, to get \tilde{A}_j^h and \tilde{A}_i^w .

After getting the global contextual information, compression and expansion are applied to \tilde{A}_j^h and \tilde{A}_i^w , where compression uses a shared $\text{Conv}_{\text{Norm-RELU}}$ to fuse attention maps in height and width directions and expansion applies $\text{Conv}_{\text{Sigmoid}}$ to \tilde{A}_j^h and \tilde{A}_i^w separately to extract contextual features. Then \tilde{A}_j^h and \tilde{A}_i^w are divided into patches again to obtain $\tilde{A}_{i,j}^h$ and $\tilde{A}_{i,j}^w$, which are subsequently multiplied to get the attention maps $A_{i,j}^m \in R^{C \times P_h \times P_w}$. Specifically, $A_{i,j}^m$ is written as:

$$A_{i,j}^m = \tilde{A}_{i,j}^h \times \tilde{A}_{i,j}^w, \quad (4)$$

where \times denotes matrix multiplication in spatial dimension.

Finally, $A_{i,j}^m$ and $P_{i,j}$ are applied element-wise product to get the reweighted patches and the patches are concatenated, obtaining the output feature map $X_{\text{out}} \in R^{C \times H \times W}$.

Compared with coordinate attention (CA) [42], GCA can better learn the local detailed features in the osteosarcoma MRI images while maintaining the global semantic features, which improves the segmentation accuracy in some blurred tumor images. In addition, to flexibly adjust the patch size and compensate for the loss of information between patches by using different patch sizes, padding-crop operation is designed in GCA. Through adaptive padding, the input feature map can be divided into patches of any size and be cropped back to the original input size after the attention operation.

3.2. Statistical Texture Learning Block. In osteosarcoma segmentation tasks, the rich contextual information contained in low-level features plays a crucial role in segmentation performance. To solve the problem of blurred boundaries as well as complex and irregular tumor shapes in osteosarcoma MRI images using biosensors, we use statistical texture learning block (STLB) [36] to fully exploit and utilize the texture features and combine the rich low-level features with the high-level features in the deeper layer of the network.

SFNet can combine low-level and high-level features with semantic flow. At the same time, STLB can explore the statistical features of osteosarcoma image texture [44, 45]. It not only learns the structural texture information, but also learns the global statistical information of the image, which is helpful for osteosarcoma segmentation. In this section, the 1d and 2d quantization and counting operator (QCO) are first introduced for statistical description of the texture information. Then two modules in STLB are introduced: the texture enhancement module (TEM) based on 1d-QCO to enhance the osteosarcoma

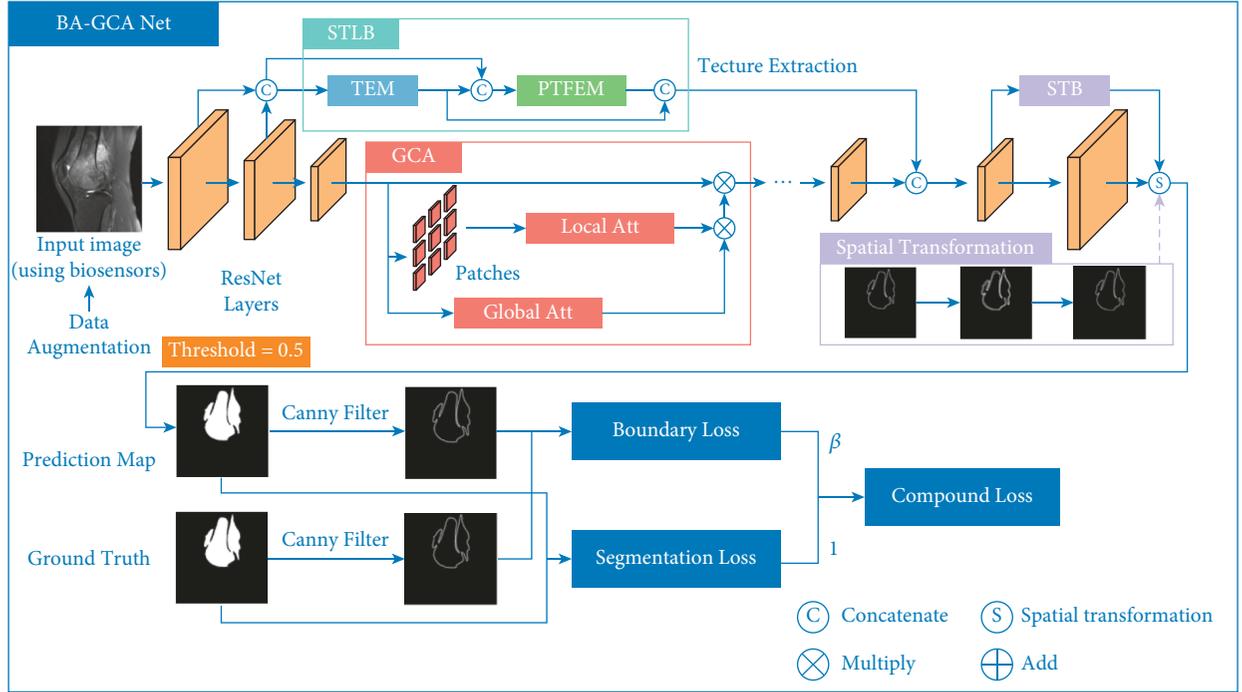


FIGURE 1: The overall structure of BA-GCA Net.

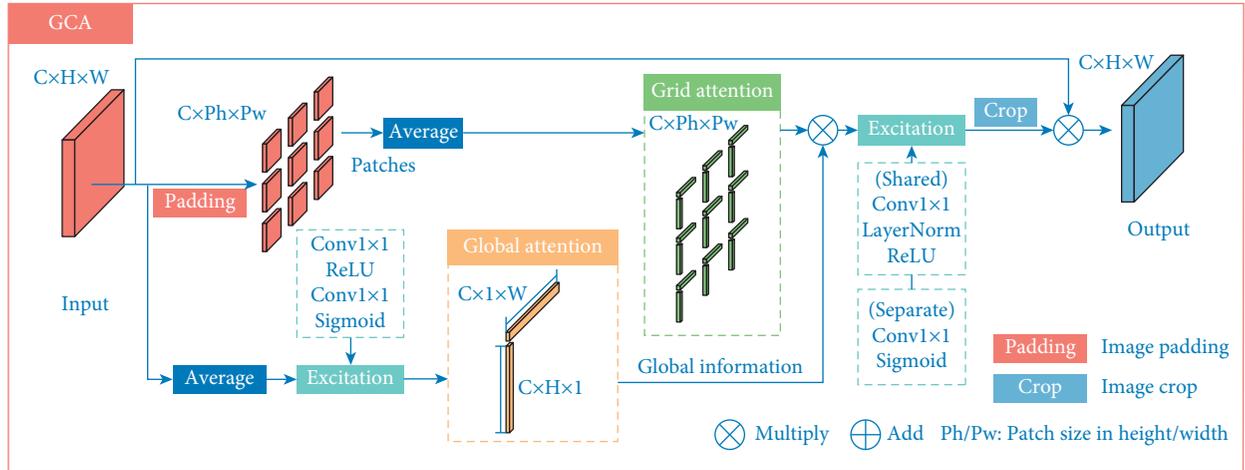


FIGURE 2: The structure of grid contextual attention (GCA).

texture features and the pyramid texture feature extraction module (PTFEM) based on 2d-QCO to further exploit the texture features.

3.2.1. 1d-QCO. The structure of 1d-QCO is shown in Figure 3.

For an input feature map $X \in R^{C \times H \times W}$, 1d-QCO applies global average pooling to X to get the average feature $a \in R^{C \times 1 \times 1}$. Then the cosine similarity between each pixel $X_{i,j}$ in X and a is calculated to get $S \in R^{1 \times H \times W}$, where $i \in \{1, 2, \dots, H\}$ and $j \in \{1, 2, \dots, W\}$. Each position $S_{i,j}$ is denoted as:

$$S_{i,j} = \frac{a \cdot X_{i,j}}{\|a\|_2 \cdot \|X_{i,j}\|_2}. \quad (5)$$

Then S is reshaped to $S \in R^{HW}$ and quantized to obtain the N levels $L = [L_1, L_2, \dots, L_N]$. The n th level L_n is written as:

$$L_n = \frac{n}{N} (\max(S) - \min(S)) + \min(S), \quad (6)$$

where N is a hyperparameter and $n \in \{1, 2, \dots, N\}$.

For the similarity $S_i \in R$ of each pixel in the feature map, we encode it to $E^i \in R^N$, where $i \in \{1, 2, \dots, HW\}$ and each dimension $n \in \{1, 2, \dots, N\}$ of E_i is calculated by:

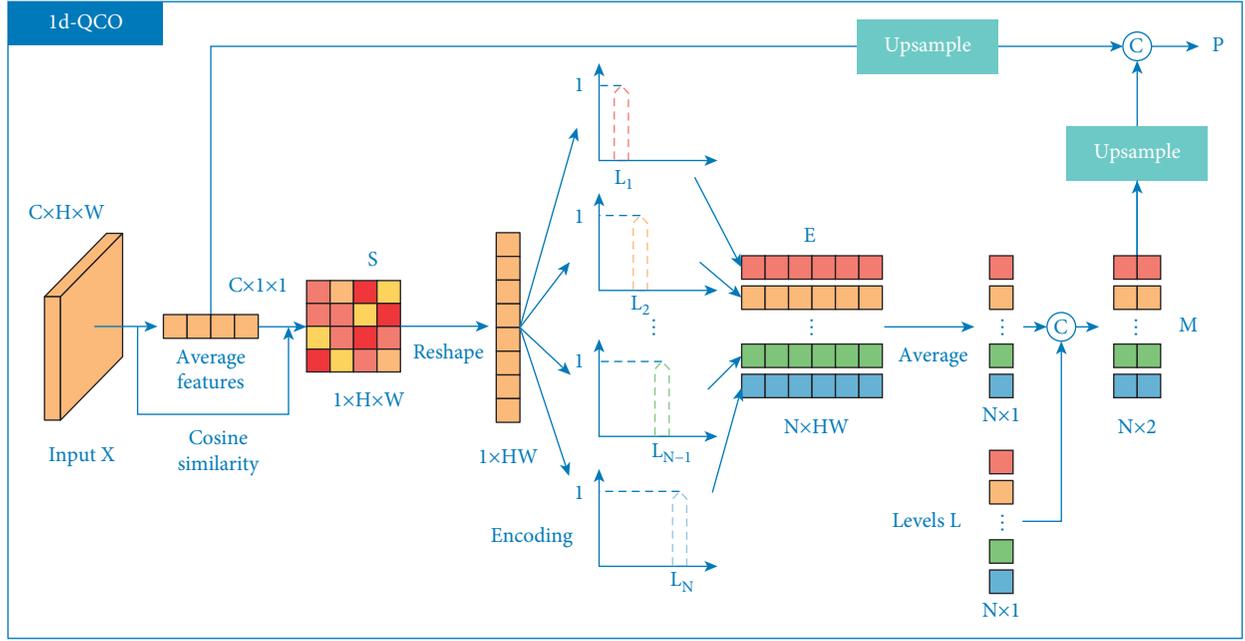


FIGURE 3: The structure of 1d-QCO.

$$E_{i,n} = \begin{cases} 1 - |L_n - S_i|, & -\frac{0.5}{N} \leq L_n - S_i < \frac{0.5}{N}, \\ 0, & \text{other.} \end{cases} \quad (7)$$

The encoding map $E \in R^{N \times HW}$ consists of each pixel's encoded value. Compared with one-hot encoding or argmax operation, the quantization encoding is smoother and robust to gradient vanishing.

1d-QCO then applies counting operation to the encoding map E to get the counting map $M \in R^{N \times 2}$. Concretely, M is calculated by:

$$M = \text{Concat} \left(L, \frac{\sum_{i=1}^{HW} E_{i,n}}{\sum_{n=1}^N \sum_{i=1}^{HW} E_{i,n}} \right), \quad (8)$$

where Concat denotes concatenate operation in channel dimension.

Thereafter, the average feature a is up-sampled to $a \in R^{N \times C}$ and concatenated to the up-sampled M to produce $P \in R^{N \times C_1}$, which is calculated by:

$$P = \text{Concat}(\text{Up}(M), a). \quad (9)$$

The output of 1d-QCO includes the encoding map E as well as the statistical texture information P of osteosarcoma.

3.2.2. 2d-QCO. 1d-QCO contains the statistical texture features of the osteosarcoma images. However, it cannot learn positional relationships between pixels. Therefore, a 2d-QCO is proposed.

Similar to 1d-QCO, 2d-QCO calculates cosine similarity and level encoding of the input feature map $X \in R^{C \times H \times W}$ to get the encoding map $E \in R^{N \times HW}$ and quantization levels L . Then E is reshaped to $E \in R^{N \times 1 \times H \times W}$. For the encoding of

each adjacent pixel pair $E_{i,j} \in R^{N \times 1}$ and $E_{i,j+1} \in R^{N \times 1}$, the encoded value $\hat{E}_{i,j} \in R^{N \times N}$ that carries adjacent information is calculated by:

$$\hat{E}_{i,j} = E_{i,j} \times E_{i,j+1}^T, \quad (10)$$

where T and \times denote matrix transpose and multiplication, respectively. Then we get the encoding map $\hat{E} \in R^{N \times N \times H \times W}$ that contains adjacent features of the input.

In the counting process, the counting map $M \in R^{N \times N \times 3}$ is produced by:

$$M = \text{Concat} \left(\hat{L}, \frac{\sum_{i=1}^H \sum_{j=1}^W \hat{E}_{m,n,i,j}}{\sum_{m=1}^N \sum_{n=1}^N \sum_{i=1}^H \sum_{j=1}^W \hat{E}_{m,n,i,j}} \right), \quad (11)$$

where $\hat{L} \in R^{N \times N \times 2}$ represents the pairwise combination of all the quantization levels and $\hat{L}_{m,n} = [L_m, L_n]$.

In 2d-QCO, the average feature is written as $a \in R^{N \times N \times C}$ and the final output $P \in R^{N \times N \times C_1}$ is obtained by:

$$P = \text{Concat}(\text{Up}(M), a). \quad (12)$$

3.2.3. Texture Enhancement Module. The low-level texture such as structural boundaries in osteosarcoma images are often blurred and of low contrast. To this end, a texture enhancement module (TEM) is employed to sharpening the structural texture and make the low-level features easier to learn. The structure of TEM is shown in Figure 4.

Inspired by the histogram quantization method in traditional image processing algorithms, the statistical information in each quantization level is treated as a node in the graph adjacency matrix. Unlike the traditional method of defining a diagonal matrix artificially, TEM uses graph reasoning to construct the adjacency matrix and reconstruct

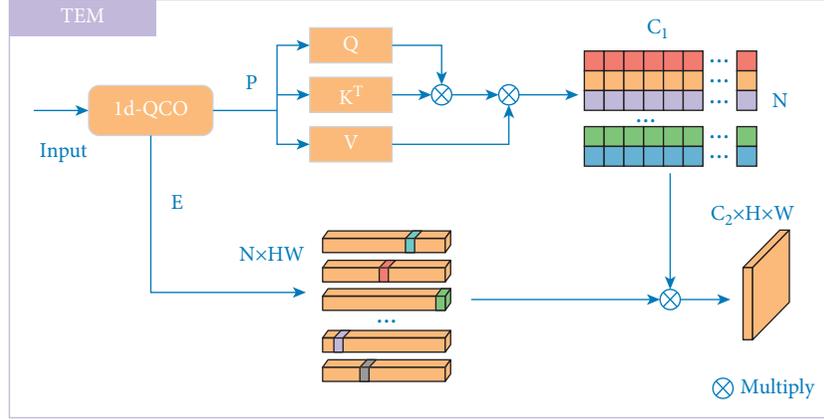


FIGURE 4: The structure of the texture enhancement module.

the quantization level L to get L' . Concretely, the process can be written as:

$$\begin{aligned} G &= \text{Softmax}(\text{Conv}(P^T) \times \text{Conv}(P)), \\ L' &= \text{Conv}(P) \times G. \end{aligned} \quad (13)$$

Finally, the output $O \in R^{C_2 \times H \times W}$ is obtained using the reconstructed quantization levels L' and the encoding map E . $O \in R^{C_2 \times H \times W}$ is denoted as:

$$O = \text{Reshape}(L' \times E), \quad (14)$$

where Reshape represents reshaping the output to $O \in R^{C_2 \times H \times W}$.

3.2.4. Pyramid Texture Feature Extraction Module. The features of statistical texture in osteosarcoma MRI images are effectively enhanced through TEM. Then a pyramid texture feature extraction module (PTFEM) is proposed to extract and exploit rich texture features of the boundaries. The structure of PTFEM is shown in Figure 5.

Inspired by the conventional gray-level co-occurrence matrix algorithm, the input feature map of osteosarcoma is first passed through 2d-QCO to get the statistical co-occurrent features $P \in R^{C' \times N \times N}$ and then the texture features $T \in R^{C'}$ is calculated using a MLP and a level-wise average operation. Specifically, the process is denoted as:

$$\begin{aligned} P' &= \text{MLP}(P), P' \in R^{C' \times N \times N}, \\ T &= \frac{\sum_{m=1}^N \sum_{n=1}^N P'_{k,m,n}}{N \times N}, k \in \{1, 2, \dots, C'\}. \end{aligned} \quad (15)$$

Some previous works such as FPN [46] and DeepLabV3+ [47] found that the employment of multi-scale structure can improve the model's performance. Inspired by these works, PTFEM integrates 2d-QCO with different scales into the structure to better extract and utilize the osteosarcoma texture features.

3.3. Spatial Transformer Block. Due to the small and complicated osteosarcoma lesion areas in some MRI images produced by biosensors, even a slight spatial shift of the

prediction map can produce poor results, which in turn lead to wrong diagnosis. To this end, we use an STB [37] to make the backbone invariant to spatial transformation and more robust to the osteosarcoma images with small and intricate tumor regions.

For the high-resolution segmentation map of osteosarcoma, we assume that the error of the map to the ground truth can be reduced by homography transformation. Therefore, we use STB to learn this spatial transformation. The structure of STB is shown in Figure 6.

For an input feature map $X \in R^{C \times H \times W}$, STB uses a set of down-sampling convolutions $F_{\text{down-sample}}$ and fully connected layers $F_{\text{regression}}$ to produce an affine transformation matrix $M_{\text{affine}} \in R^{2 \times 3}$. Concretely, M_{affine} is denoted as:

$$M_{\text{affine}} = F_{\text{regression}}(F_{\text{down-sample}}(X)). \quad (16)$$

Simultaneously, the input feature map is applied a 1×1 convolution as well as a softmax activation in another branch to get the initial prediction map $\text{pred} \in R^{2 \times H \times W}$, which can be described as:

$$\text{pred} = \text{Softmax}(\text{Conv}(X)). \quad (17)$$

We denote the affine matrix as $M_{\text{affine}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$. For the coordinates (x_n^s, y_n^s) of each pixel in the initial prediction map pred and the coordinates (x_n^t, y_n^t) of each pixel in the final prediction map $\text{pred}' \in R^{2 \times H \times W}$, where $n \in \{1, 2, \dots, HW\}$, the affine transformation is defined as:

$$\begin{pmatrix} x_n^s \\ y_n^s \\ 1 \end{pmatrix} = M_{\text{affine}} \begin{pmatrix} x_n^t \\ y_n^t \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{pmatrix} x_n^t \\ y_n^t \\ 1 \end{pmatrix}. \quad (18)$$

In order to apply spatial transformation to the initial prediction map, STB samples each (x_n^s, y_n^s) to obtain the final output pred' . Specifically, the process can be represented as:

$$\text{pred}'_{c,n} = \sum_{i=1}^H \sum_{j=1}^W \text{pred}_{c,i,j} \times I(x_n^s - j; \Phi_x) \times I(y_n^s - i; \Phi_y), \quad (19)$$

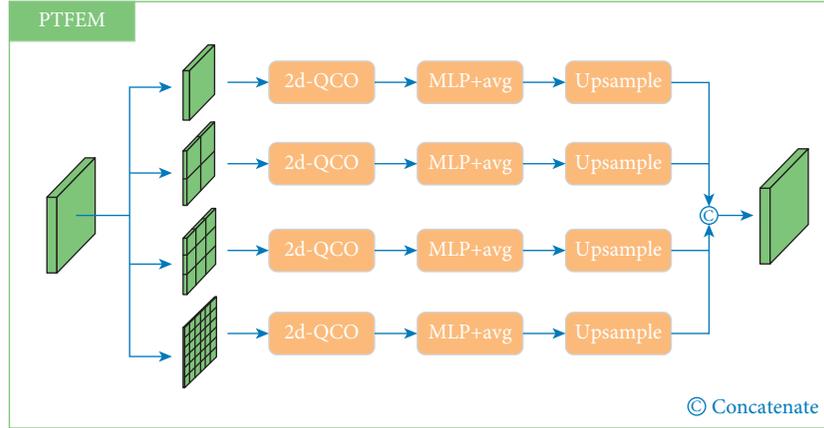


FIGURE 5: The structure of pyramid texture feature extraction module.

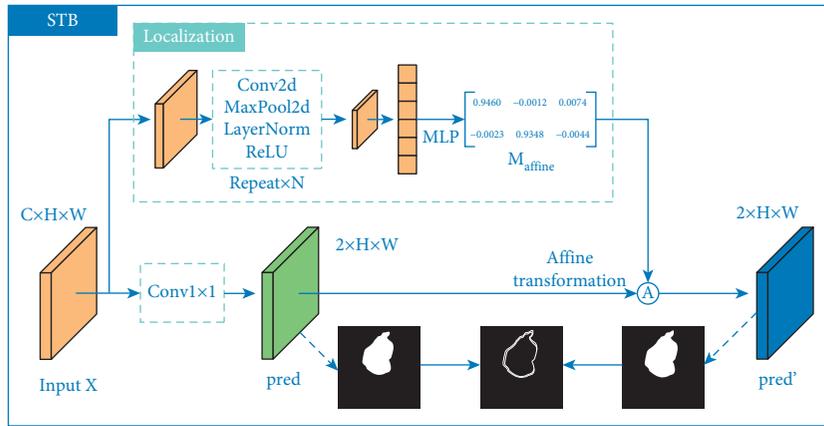


FIGURE 6: The structure of spatial transformer block (STB).

where I denotes to bilinear interpolation, Φ_x and Φ_y denote sampling parameters, and $c \in \{1, 2, \dots, C\}$ represents the channel index of the feature map.

3.4. Compound Loss Function. To improve both the prediction accuracy and the boundary perception ability of the model for osteosarcoma, we design a compound loss function based on focal loss [48]. The loss function consists of a segmentation loss as well as a boundary loss.

3.4.1. Weighted Segmentation Focal Loss. For the prediction map y_{pred} and the ground truth y_{gt} , the weighted segmentation focal loss is defined as:

$$FL_{\text{seg}} = -\alpha y_{\text{gt}} (1 - y_{\text{pred}})^\gamma \log(y_{\text{pred}}) \cdot (1 - \alpha) (1 - y_{\text{gt}}) y_{\text{pred}}^\gamma \log(1 - y_{\text{pred}}), \quad (20)$$

where α and γ represent the balance weight and the exponential hyperparameter, respectively.

3.4.2. Weighted Boundary Focal Loss. In osteosarcoma MRI image segmentation, the boundary plays an essential role in improving the segmentation performance. Therefore, we

introduce a weighted boundary focal loss to facilitate the model to learn boundary information.

First, the segmentation head and the ground truth are applied Canny [43] operator to produce the prediction and ground truth boundary b_{pred} and b_{gt} . Then we perform edge sharpening on the normalized b_{gt} with a threshold of 0.5 and obtain a clear ground truth boundary b'_{gt} . The weighted boundary focal loss is calculated using b_{pred} and b'_{gt} as:

$$FL_{\text{boundary}} = -\alpha b'_{\text{gt}} (1 - b_{\text{pred}})^\gamma \log(b_{\text{pred}}) - (1 - \alpha) (1 - b'_{\text{gt}}) b_{\text{pred}}^\gamma \log(1 - b_{\text{pred}}). \quad (21)$$

The compound loss function is defined as:

$$\text{Loss} = FL_{\text{seg}} + \beta FL_{\text{boundary}}, \quad (22)$$

where β is the hyperparameter of weight. After experiments, one of the suitable values of β is 0.2, which is used in this article.

The compound loss function enables the model aware of the boundaries of osteosarcoma and also promotes STB to learn the spatial transformation between the prediction output and the ground truth, which in turn makes the model more robust to tumor segmentation. BA-GCA Net is trained in combination of compound loss function, enhancing the model's ability to boundary localization. As a method to

TABLE 1: Patient statistics.

Characteristics	Total	Training set	Test set
Age			
<15	48 (23.5%)	38 (23.2%)	10 (25.0%)
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%)
SES			
Low SES	78 (38.2%)	66 (40.2%)	12 (30.0%)
High SES	126 (61.8%)	98 (59.8%)	28 (70.0%)
Surgery			
Yes	181 (88.8%)	146 (89.0%)	35 (87.5%)
No	23 (11.2%)	18 (11.0%)	5 (12.5%)
Grade			
Low grade	41 (20.1%)	15 (9.1%)	26 (65.0%)
High grade	163 (79.9%)	149 (90.9%)	14 (35.0%)
Location			
Axial	29 (14.2%)	21 (12.8%)	8 (20.0%)
Extremity	138 (67.7%)	109 (66.5%)	29 (72.5%)
Other	37 (18.1%)	34 (20.7%)	3 (7.5%)

assist doctors in diagnosis, it reduces the doctors' workload and improves the accuracy of diagnosis.

4. Experiments

In this section, we employ over 80,000 MRI images of osteosarcoma output by biosensors from 204 cases as the dataset for experiments to evaluate the model and perform ablation studies, which are provided by the Ministry of Education Mobile Health Information-China Mobile Joint Laboratory and the Second Xiangya Hospital of Central South University [49].

4.1. Dataset. We have collected statistics about patients and the results are shown in Table 1. We randomly select 80% of the images for training and the remaining 20% for evaluation. To be specific, there are a total of 204 case samples, of which 164 are in the training set and 40 are in the test set.

Due to the confidentiality of data between hospitals and the privacy of patients, the dataset is relatively hard to obtain, which leads to the overfitting problem of the model. To promote the robustness of the model to new data, we perform data augmentation on the training set. We rotate the images at 3 angles (0, 90, and 180), flip the image on different axes (no flip, up-down, and left-right), perform Gaussian blurring, add Gaussian noise (using different variances), and apply salt and pepper noise (using different proportions) to augment the training set.

4.2. Evaluation Metrics. In order to evaluate the performance of the model on osteosarcoma MRI image segmentation, in this section, we introduce accuracy, precision,

TABLE 2: Hyperparameters of the model.

stage	Hyperparameter	Value
GCA block	Strategy	Residual
ST block	Weight Bias	[1, 0, 0, 0, 1, 0]
	Interpolation method	Bilinear
STL block	num_levels	128
Loss function	α	Based on ratios in batch
	β	0.2
	γ	1.25
Training	Initializer	kaiming_uniform
	Epochs	200
	Base learning rate	0.0001
	Optimizer	Adam
	Learning rate decay	$(1 - (\text{epoch}/\text{total_epochs}))^{0.9}$
	Up-sampling	Bilinear

recall, F1-score, Dice similarity coefficient (DSC) [50], and Intersection of Union (IOU) as the evaluation metrics and the confusion matrix with true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) to explain the performance of the model [51]. The evaluation metrics are defined as follows:

Accuracy (Acc) is used to evaluate the proportion of the model's right prediction and it is denoted as [52]:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (23)$$

precision (Pre) is to calculate the percentage of true osteosarcoma areas among the prediction areas [53]. Precision can be written as:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (24)$$

recall (Rec) evaluates the percentage of the prediction osteosarcoma areas among the true areas [54]. Concretely, recall is denoted as:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (25)$$

F1 indicates the robustness of segmentation and is defined as [55]:

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \quad (26)$$

For a simpler description of DSC as well as IOU, we denote the prediction of tumor area as y_{pred} and the ground truth as y_{gt} .

DSC represents the similarity between y_{pred} and y_{gt} . DSC can be written as

$$\text{DSC} = \frac{2|y_{\text{pred}} \cap y_{\text{gt}}|}{|y_{\text{pred}}| + |y_{\text{gt}}|}. \quad (27)$$

IOU measures the degree of overlap between the model's prediction map and the ground truth and it is denoted as [56]

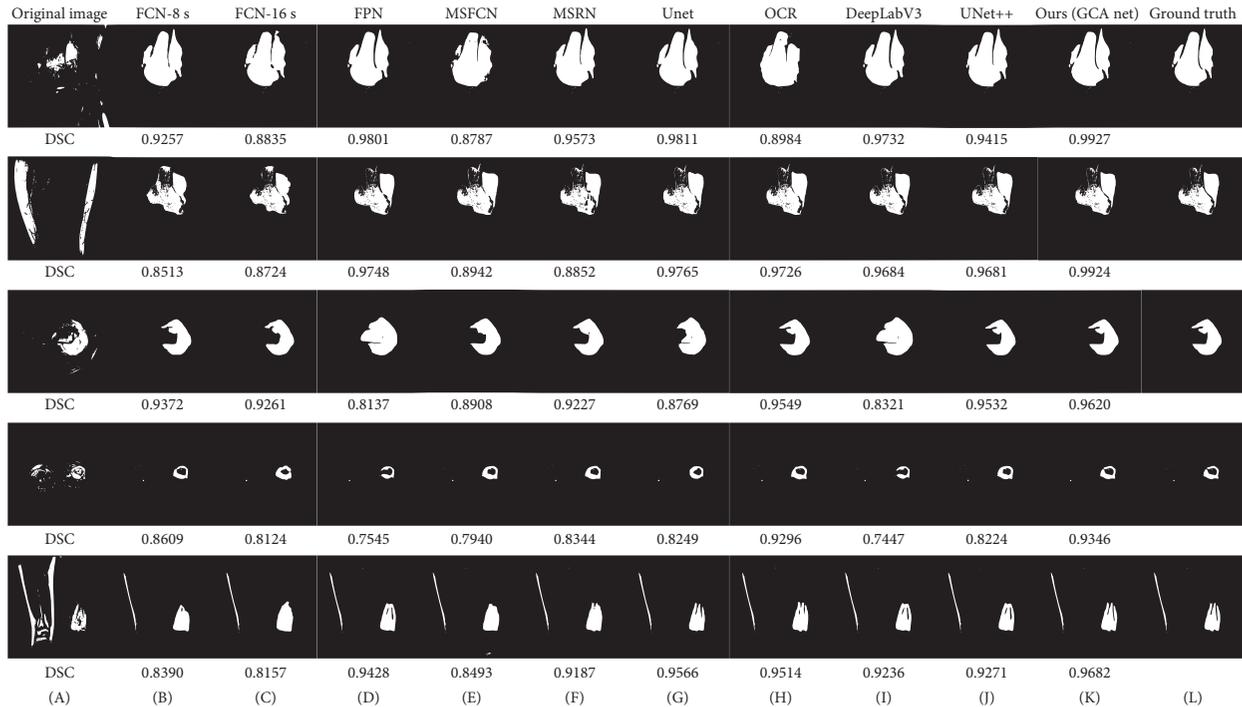


FIGURE 7: The segmentation effects of models on some osteosarcoma MRI images.

TABLE 3: Performance of models.

Model	Pre	Rec	F1	DSC	IOU	#params	FLOPs
FCN-16s [21]	0.922	0.882	0.900	0.859	0.824	134.3 M	190.35 G
FCN-8s [21]	0.892	0.914	0.902	0.875	0.831	134.3 M	190.08 G
MSFCN [24]	0.881	0.936	0.906	0.874	0.841	23.38 M	1524.34 G
MSRN [25]	0.893	0.945	0.918	0.887	0.853	14.27 M	1461.23 G
FPN [46]	0.914	0.924	0.919	0.888	0.852	48.20 M	141.45 G
U-Net [22]	0.922	0.924	0.923	0.892	0.867	17.26 M	160.16 G
OCR [41]	0.897	0.908	0.901	0.891	0.827	27.35 M	125.67 G
DeepLabV3 [4]	0.926	0.925	0.925	0.909	0.870	39.63 M	170.45 G
UNet++ [59]	0.924	0.924	0.924	0.908	0.868	18.16 M	165.23 G
SVM [60]	0.756	0.764	0.760	0.734	0.702	—	—
DRN [50]	0.916	0.922	0.917	0.909	0.843	17.66 M	76.93 G
DRN + CA [42]	0.918	0.923	0.919	0.910	0.851	18.06 M	77.21 G
Ours (DRN + GCA)	0.927	0.924	0.925	0.913	0.866	18.11 M	77.34 G
Ours (DRN + GCA + STLb)	0.925	0.934	0.929	0.916	0.873	18.47 M	82.67 G
Ours (DRN + GCA + STLb + STB)	0.938	0.937	0.937	0.927	0.880	19.88 M	149.70 G

$$\text{IOU} = \frac{y_{\text{pred}} \cap y_{\text{gt}}}{y_{\text{pred}} \cup y_{\text{gt}}} \quad (28)$$

Furthermore, we adopt #params as the number of parameters of the model and use floating point operations (FLOPs) to evaluate the computational complexity [57, 58].

4.2.1. Training Details. The Dilated ResNet-D-22(DRN-D-22) [52] is chosen as the backbone and BA-GCA Net is designed based on it. Note that osteosarcoma MRI images often have large individual differences, and the relationship between pixels in one image plays a more crucial role than that among images. To this end, we replace batch norm in

the backbone to layer norm. Other hyperparameters of the model are shown in Table 2.

4.3. Comparison with Other Methods. Figure 7 shows the performance of each model on osteosarcoma MRI image segmentation. Column (A) represents the original image, columns (B)-(K) are the prediction output of each model, and column (L) is the ground truth. Note that BA-GCA Net outperforms other models in some difficult tasks, such as the second image that has low contrast and blurred boundaries, the fourth image of transverse section with small and complicated tumor area, and the last image that contains small gaps in the region of osteosarcoma. Results show that our proposed BA-GCA Net has better performance than some

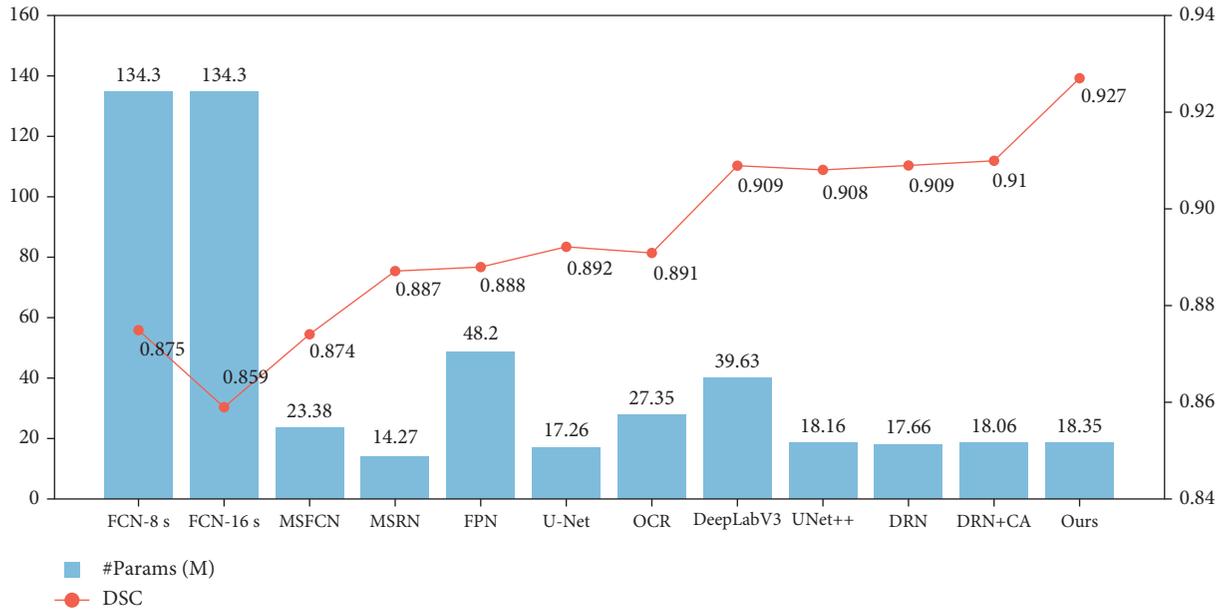


FIGURE 8: Comparison of #params and DSC between models.

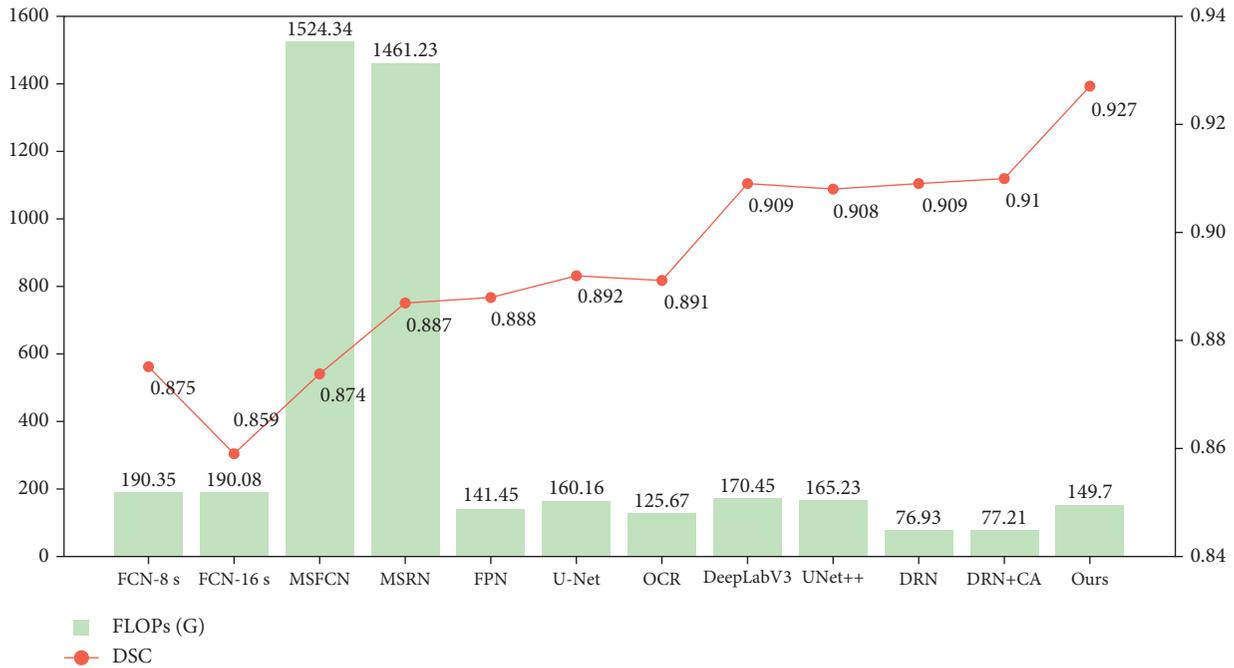


FIGURE 9: Comparison of FLOPs and DSC between models.

latest methods such as DeepLabV3 and UNet++ in capturing boundary details and recognizing blurred osteosarcoma lesion areas. Furthermore, BA-GCA Net is more robust to segmentation and DSC of its prediction remains above 0.93 in difficult segmentation tasks. Compared with other models, BA-GCA Net shows an advantage in processing the low-contrast and complex osteosarcoma images and localizing the small tumor regions, which is helpful for clinical diagnosis.

The quantitative evaluation results of BA-GCA Net and other comparative models on the test set are shown in Table 3.

From the results we know that BA-GCA Net achieves higher precision, F1-score, DSC, and IOU than latest methods, which means our proposed method performs better overall on the test set. By integrating GCA, STLb, and STB into the backbone, the DSC of the model has increased by 0.004, 0.003, and 0.011 and the IOU of the model has increased by 0.023, 0.007, and 0.007, respectively. It is proved that the three blocks effectively strengthen the backbone with an increase of DSC by 0.018 and increase of IOU by 0.037. More detailed analysis will be introduced in the Ablation Study section.

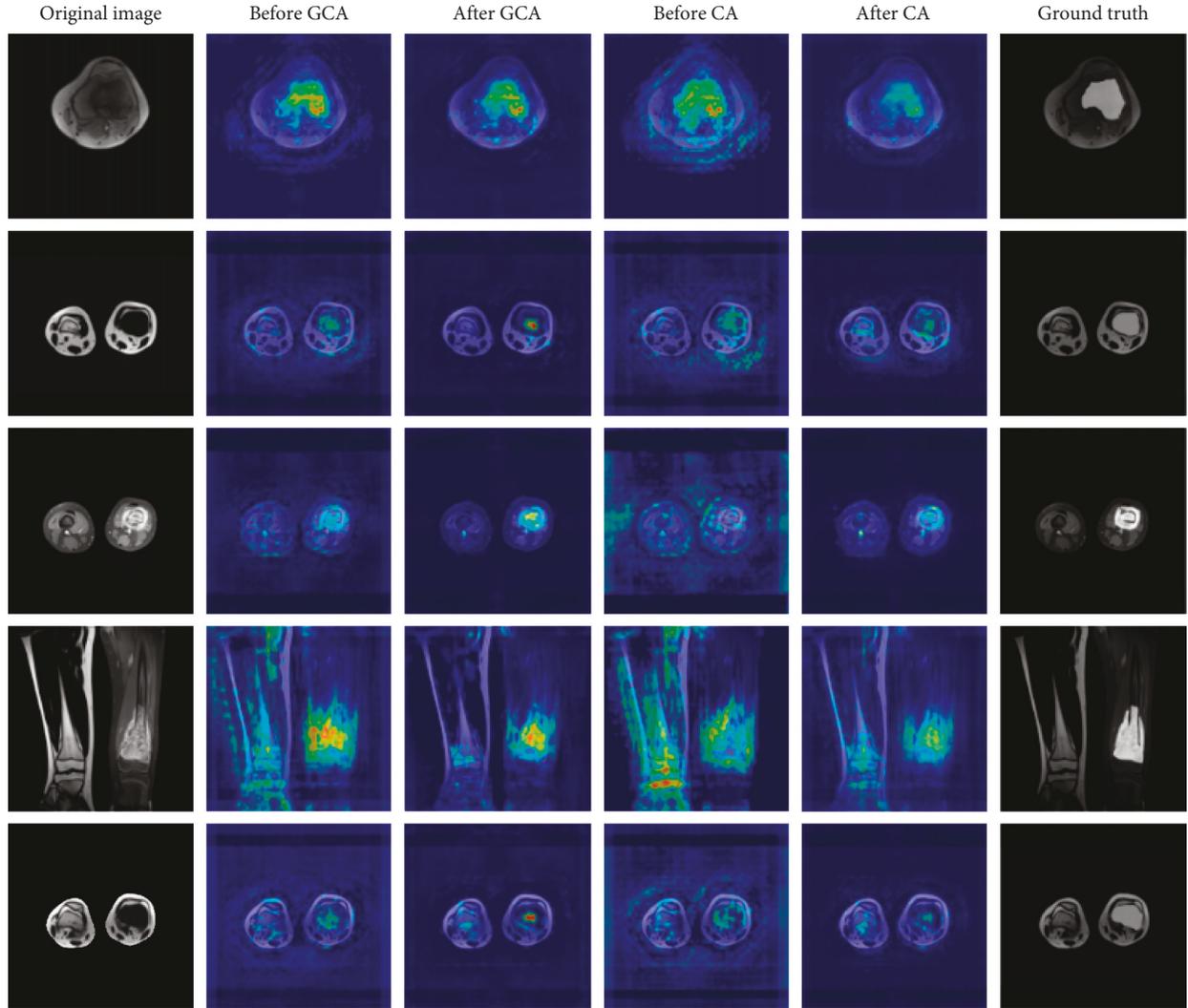


FIGURE 10: Visualization results of GCA and CA.

The relationship between the number of parameters and DSC of different models is shown in Figure 8. It shows that BA-GCA Net achieves the highest DSC among all the models. Compared with UNet++, BA-GCA Net has a 0.019 higher DSC with only a slight increase in the number of parameters by 0.19 M. Moreover, BA-GCA Net outperforms the DRN added with coordinate attention (CA) [42] on DSC increasing by 0.017 and has only an increase in the number of parameters by 1.82 M.

The comparison of FLOPs and DSC between models is shown in Figure 9. We can know from the results that the computational cost of our proposed method is higher than DRN added with CA by 72.49GFLOPs and lower than U-Net by 10.46GFLOPs, which achieves a good balance between the accuracy and the computational complexity.

5. Ablation Study

In this section, ablation studies on the three blocks are introduced respectively in order to better analyze the role as well as the necessity of each block in BA-GCA Net. By

comparing the performance of the model with and without the block, and visualizing the outputs in the middle layers, we can check out whether a block plays the role as expected.

5.1. Ablation of GCA. Attention mechanism is able to calibrate the input feature map so that the model can focus on the regions of interest. To illustrate that GCA can learn this calibration more precisely, in this section, we employ Seg-Grad-CAM [61] to visualize GCA as well as CA [42]. By summing the partial derivatives of the output target region with respect to the feature maps before and after the last GCA block (or CA block), respectively, and taking the mean of the derivatives for each channel as the weight of the feature map, we can visualize the influence of the attention block. The visualization results of GCA and CA are shown in Figure 10. Yellow and red indicate that the model gives higher weight to the region, while green and blue are the opposite. The visualization only computes the partial derivatives of the osteosarcoma region in the ground truth. The images show that GCA is more sensitive to the osteosarcoma areas and locates the tumor regions more precisely.

TABLE 4: Performance of backbone with GCA and CA.

Model	param add (M)	#params (M)	DSC	IOU
DRN-D-22	—	17.66	0.909	0.843
+CA	+0.4	18.06	0.910 (+0.001)	0.851 (+0.008)
+GCA	+0.45	18.11	0.913 (+0.004)	0.866 (+0.023)

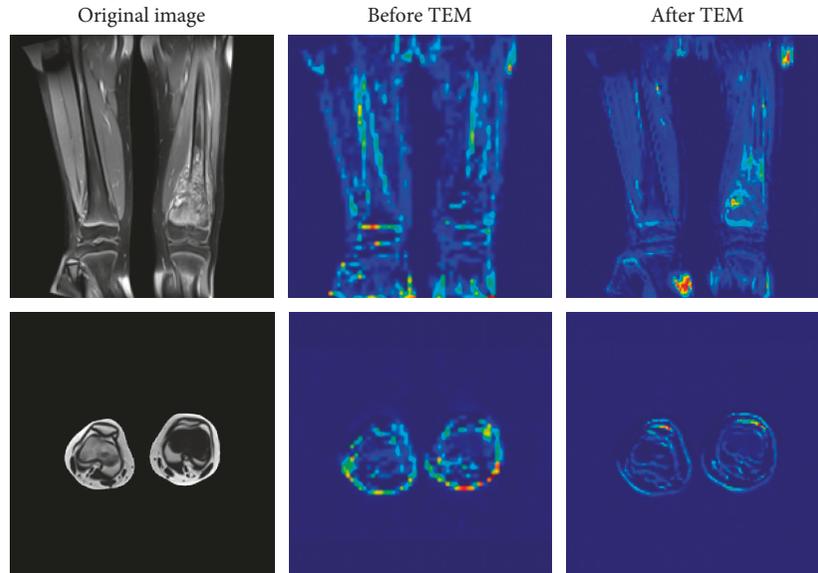


FIGURE 11: Visualization of feature maps before and after TEM.

TABLE 5: Comparison of performance before and after using STLB.

Model	Param Add (M)	#params (M)	DSC	IOU
DRN-D-22	—	17.66	0.909	0.843
+STLB	+0.35	18.01	0.914 (+0.005)	0.862 (+0.019)

Compared with CA, our proposed structure can better observe the tumor details and calibrate the feature map.

The DSC and IOU indicators and number of parameters of DRN, DRN with CA, and DRN with GCA are shown in Table 4, respectively. Compared with CA, the number of parameters of GCA is only 0.05 M more, but the DSC and IOU have increased by 0.003 and 0.015, respectively.

5.2. Ablation of STLB. STLB provides an effective method for better extracting and utilizing low-level texture features of the input images. A key role in STLB is the texture enhancement module (TEM). To analyze what TEM learns during training, we visualize the input and output feature maps of the TEM, mapping the grayscale values to colors from blue to red, which is shown in Figure 11. Note that through quantization, counting, and texture enhancement, the boundaries and some texture of the tissues such as bones, muscles, and osteosarcomas are exploited and sharpened to produce clearer feature maps, which is helpful for PTFEM to extract the spatial correlation features between pixels.

Table 5 shows the changes of DSC, IOU, and number of parameters of the model before and after adding STLB. Compared with DRN-D-22, the DRN added with STLB achieves an increase in DSC and IOU by 0.005 and 0.019, respectively, and the number of parameters increases by only 0.35 M. STLB enhances and utilizes the texture details and statistical features of osteosarcoma images, thereby improving the performance on low-contrast and blurred tumor segmentation tasks.

5.3. Ablation of STB. As is mentioned above, STB learns to perform affine transformation on the prediction map to better locate the tumor area. In this section, we examine the segmentation effect before and after affine transformation, as shown in Figure 12. Column (D) shows the comparison of boundaries before and after STB. Canny filter is used to extract the boundary. The red one is the output boundary before STB, and the blue one is the opposite. From the changes of DSC, we can conclude that STB performs beneficial spatial transformation on the input feature map,

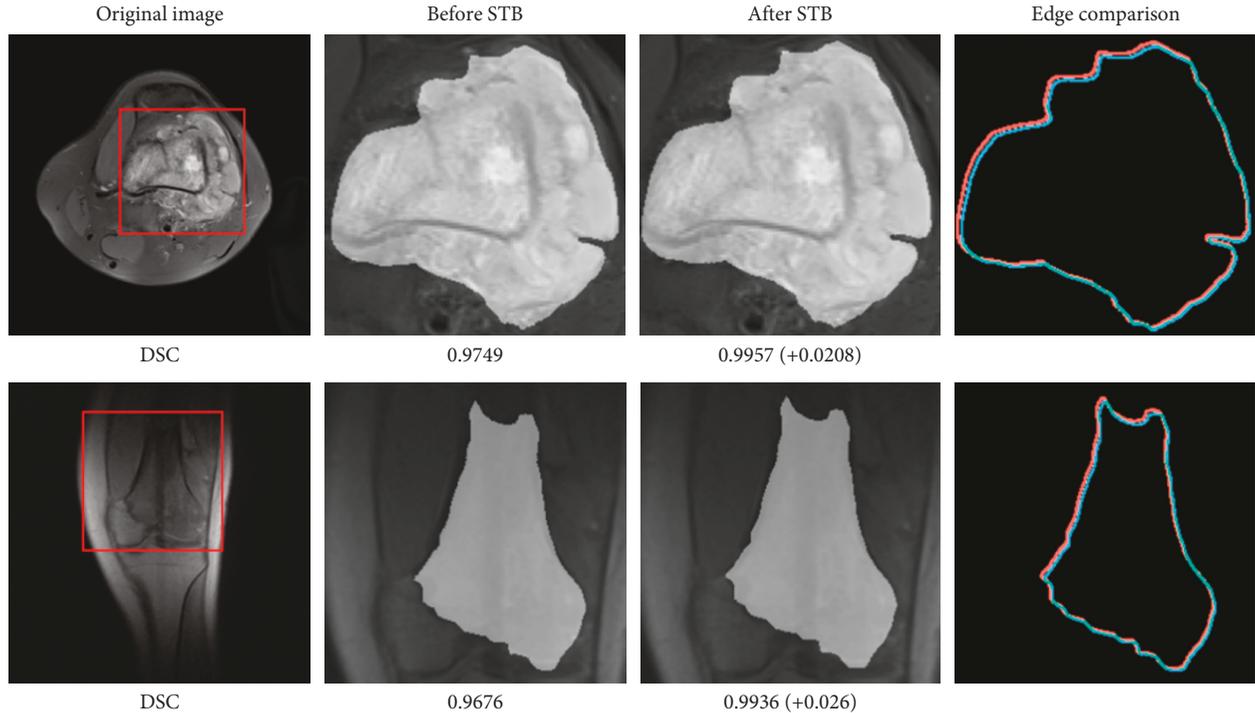


FIGURE 12: Visualization of prediction boundaries before and after STB.

making the prediction area more accurate. By integrating STB into the model, the spatial shift and deformation caused by down-sampling and up-sampling are corrected, and the DSC and IOU has increased by 0.011 and 0.007, respectively.

6. Conclusion

In this article, we use a novel GCA, STLb, and STB to improve the model's segmentation performance on some difficult tasks such as complex osteosarcoma boundaries, small tumor areas, and low-contrast images produced by MRI biosensors. We propose a BA-GCA Net with the three blocks and employ over 80,000 MRI images of osteosarcoma from the Second Xiangya Hospital in China to train and test the model. In order to check out the function of each block, we conduct ablation studies. The visual analysis of the results helps us understand how each block works and its effectiveness. The test results show that our proposed BA-GCA Net achieves 0.927 DSC and 0.880 IOU, which is better than other existing models. The number of parameters and computational cost are only 19.88 M and 149.70GFLOPs, respectively, which means the model reaches a balance between accuracy and computational consumption. The model can assist doctors in judging the area of osteosarcoma at a relatively low cost, reduce the workload of doctors, and improve the efficiency of diagnosis.

In the future, in view of the difficulty in obtaining the clinical data of osteosarcoma, we will introduce few-shot learning into our method, so that the model can use fewer samples to obtain similar results. It will solve the problem of insufficient generalization of hospital self-trained models

due to the data incompatibility between hospitals, improve the robustness of the model, and reduce training costs.

Data Availability

The 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 conflicts of interest.

Acknowledgments

This work was supported by the Hunan Provincial Natural Science Foundation of China (2018JJ3299, 2018JJ3682, and 2019JJ40440).

References

- [1] M. Barani, M. Mukhtar, A. Rahdar, S. Sargazi, S. Pandey, and M. Kang, "Recent advances in nanotechnology-based diagnosis and treatments of human osteosarcoma," *Biosensors*, vol. 11, no. 2, p. 55, 2021.
- [2] D. Zheng, W. Liu, W. Xie et al., "AHA1 upregulates IDH1 and metabolic activity to promote growth and metastasis and predicts prognosis in osteosarcoma," *Signal Transduction and Targeted Therapy*, vol. 6, no. 1, p. 25, 2021.
- [3] F. Sadoughi, P. Maleki Dana, Z. Asemi, and B. Yousefi, "DNA damage response and repair in osteosarcoma: defects,

- regulation and therapeutic implications,” *DNA Repair*, vol. 102, Article ID 103105, 2021.
- [4] J. Wu, P. Xiao, Z. Zhou et al., “An artificial intelligence multiprocessing scheme for the diagnosis of osteosarcoma MRI images,” *IEEE Journal of Biomedical and Health Informatics*, pp. 1–12, 2021.
 - [5] T. Ouyang, S. Yang, F. Gou, Z. Dai, and J. Wu, “Rethinking U-net from an attention perspective with transformers for osteosarcoma MRI image segmentation,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–24, Article ID 7973404, 2022.
 - [6] Y. Tan, W. Jia, and G. Fang, “A staging auxiliary diagnosis model for non-small cell lung cancer based on intelligent medical system,” *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 6654946, 13 pages, 2021.
 - [7] Q. Zhuang, Z. Dai, and J. Wu, “Deep active learning framework for lymph node metastasis prediction in medical support system,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–17, Article ID 4601696, 2022.
 - [8] B. Lv, F. Liu, F. Gou, and J. Wu, “Multi-scale tumor localization based on priori guidance-based segmentation method for osteosarcoma MRI images,” *Mathematics*, vol. 10, no. 12, p. 2099, 2022.
 - [9] J. Wu, Y. Tan, Z. Chen, and M. Zhao, “Data decision and drug therapy based on non-small cell lung cancer in a big data medical system in developing countries,” *Symmetry*, vol. 10, no. 5, p. 152, 2018.
 - [10] R. Cui, Z. Chen, J. Wu, Y. Tan, and G. Yu, “A multiprocessing scheme for PET image pre-screening, noise reduction, segmentation and lesion partitioning,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1699–1711, 2021.
 - [11] Y. Shen, F. Gou, and J. Wu, “Node screening method based on federated learning with IoT in opportunistic social networks,” *Mathematics*, vol. 10, p. 1669, 2022.
 - [12] J. Wu, F. Gou, and X. Tian, “Disease control and prevention in rare plants based on the dominant population selection method in opportunistic social networks,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–16, Article ID 1489988, 2022.
 - [13] J. Wu, J. Xia, and F. Gou, “Information transmission mode and IoT community reconstruction based on user influence in opportunistic social networks,” *Peer-to-Peer Networking and Applications*, vol. 15, no. 3, pp. 1398–1416, 2022.
 - [14] Y. Jiao, H. Qi, and J. Wu, “Capsule network assisted electrocardiogram classification model for smart healthcare,” *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 543–555, 2022.
 - [15] F. Liu, F. Gou, and J. Wu, “An attention-preserving network-based method for assisted segmentation of osteosarcoma MRI images,” *Mathematics*, vol. 10, p. 1665, 2022.
 - [16] J. Wu, X. Tian, and Y. Tan, “Hospital evaluation mechanism based on mobile health for IoT system in social networks,” *Computers in Biology and Medicine*, vol. 109, pp. 138–147, 2019.
 - [17] L. Chang, J. Wu, N. Moustafa, A. K. Bashir, and K. Yu, “AI-driven synthetic biology for non-small cell lung cancer drug effectiveness-cost analysis in intelligent assisted medical systems,” *IEEE Journal of Biomedical and Health Informatics*, vol. 1, p. 1, 2021.
 - [18] J. Wu, L. Chang, and G. Yu, “Effective data decision-making and transmission system based on mobile health for chronic disease management in the elderly,” *IEEE Systems Journal*, vol. 15, no. 4, pp. 5537–5548, 2021.
 - [19] J. Wu, Q. Zhuang, and Y. Tan, “Auxiliary medical decision system for prostate cancer based on ensemble method,” *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 6509596, 11 pages, 2020.
 - [20] F. Gou and J. Wu, “Triad link prediction method based on the evolutionary analysis with IoT in opportunistic social networks,” *Computer Communications*, vol. 181, pp. 143–155, 2022.
 - [21] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceeding of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, June 2015.
 - [22] O. Ronneberger, P. Fischer, and T. Brox, “U-net: convolutional networks for biomedical image segmentation,” *Lecture Notes in Computer Science*, Springer, Berlin, Germany, pp. 234–241, 2015.
 - [23] R. A. Nabid, M. L. Rahman, and M. F. Hossain, “Classification of Osteosarcoma Tumor from Histological Image Using Sequential RCNN,” in *Proceeding of the 2020 11th International Conference on Electrical and Computer Engineering (ICECE)*, Dhaka, Bangladesh, December 2020.
 - [24] L. Huang, W. Xia, B. Zhang, B. Qiu, and X. Gao, “MSFCN-multiple supervised fully convolutional networks for the osteosarcoma segmentation of CT images,” *Computer Methods and Programs in Biomedicine*, vol. 143, pp. 67–74, 2017.
 - [25] R. Zhang, L. Huang, W. Xia, B. Zhang, B. Qiu, and X. Gao, “Multiple supervised residual network for osteosarcoma segmentation in CT images,” *Computerized Medical Imaging and Graphics*, vol. 63, pp. 1–8, 2018.
 - [26] Y. Zhao, F. Hong, and Li. Minglu, “Multimodality MRI information fusion for osteosarcoma segmentation,” in *Proceeding Of The IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, 2003*, Kyoto, Japan, October 2003.
 - [27] R. Mandava, O. M. Alia, B. C. Wei, D. Ramachandram, M. E. Aziz, and I. L. Shuaib, “Osteosarcoma segmentation in MRI using dynamic Harmony Search based clustering,” in *2010 International Conference of Soft Computing and Pattern Recognition*, Cergy-Pontoise, France, December 2010.
 - [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
 - [29] L. Shuai, X. Gao, and J. Wang, “Wnet ++: A Nested W-Shaped Network with Multiscale Input and Adaptive Deep Supervision for Osteosarcoma Segmentation,” in *Proceeding of the 2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT)*, Xi’an, China, August 2021.
 - [30] H. J. Lee, J. U. Kim, S. Lee, H. G. Kim, and Y. M. Ro, “Structure boundary preserving segmentation for medical image with ambiguous boundary,” in *Proceeding of the 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA, June 2020.
 - [31] H. Ding, X. Jiang, A. Q. Liu, N. M. Thalmann, and G. Wang, “Boundary-aware feature propagation for scene segmentation,” in *Proceeding of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), October 2019.
 - [32] S. Borse, Y. Wang, Y. Zhang, and F. Porikli, “InverseForm: a loss function for structured boundary-aware segmentation,” in *Proceeding of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, June 2021.

- [33] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. Ben Ayed, "Boundary loss for highly unbalanced segmentation," *Medical Image Analysis*, vol. 67, Article ID 101851, 2021.
- [34] F. Gou and J. Wu, "Data transmission strategy based on node motion prediction IoT system in opportunistic social networks," *Wireless Personal Communications*, pp. 1–12, 2022.
- [35] D. Karimi and S. E. Salcudean, "Reducing the Hausdorff Distance in medical image segmentation with convolutional neural networks," *IEEE Transactions on Medical Imaging*, vol. 39, no. 2, pp. 499–513, 2020.
- [36] L. Zhu, D. Ji, S. Zhu, W. Gan, W. Wu, and J. Yan, "Learning statistical texture for semantic segmentation," in *Proceeding of the 2021IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, June 2021.
- [37] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with convolutional neural networks," *International Journal of Computer Vision*, vol. 116, no. 1, pp. 1–20, 2015.
- [38] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proceeding of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018.
- [39] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block Attention module," *Computer Vision ECCV 2018*, Springer, vol. 3, pp. 3–19, Berlin/Heidelberg, Germany, 2018.
- [40] Z. Zhong, Z. Q. Lin, R. Bidart et al., "Squeeze-and-Attention networks for semantic segmentation," in *Proceeding of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 2020.
- [41] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," *Computer Vision – ECCV 2020*, Springer, Berlin/Heidelberg, Germany, pp. 173–190, 2020.
- [42] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceeding of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, June 2021.
- [43] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [44] Z. Zhu, M. Xu, S. Bai, T. Huang, and X. Bai, "Asymmetric non-local neural networks for semantic segmentation," in *Proceeding of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), October 2019.
- [45] X. Li, A. You, Z. Zhu et al., "Semantic flow for fast and accurate scene parsing," *Computer Vision – ECCV 2020*, Springer, Berlin/Heidelberg, Germany, pp. 775–793, 2020.
- [46] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceeding of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.
- [47] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Computer Vision – ECCV 2018*, Springer, Berlin/Heidelberg, Germany, pp. 833–851, 2018.
- [48] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceeding of the 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, October 2017.
- [49] J. Wu, S. Yang, F. Gou et al., "Intelligent segmentation medical assistance system for MRI images of osteosarcoma in developing countries," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 6654946, 17 pages, 2022.
- [50] F. Gou and J. Wu, "Message transmission strategy based on recurrent neural network and attention mechanism in IoT system," *Journal of Circuits, Systems, and Computers*, vol. 31, no. 07, 2022.
- [51] X. Zhan, H. Long, F. Gou, X. Duan, G. Kong, and J. Wu, "A convolutional neural network-based intelligent medical system with sensors for assistive diagnosis and decision-making in non-small cell lung cancer," *Sensors*, vol. 21, p. 7996, 2021.
- [52] L. Li, F. Gou, H. Long, K. He, and J. Wu, "Effective data optimization and evaluation based on social communication with AI-assisted in opportunistic social networks," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 4879557, 19 pages, 2022.
- [53] L. Li, F. Gou, and J. Wu, "Modified data delivery strategy based on stochastic block model and community detection in opportunistic social networks," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–16, Article ID 5067849, 2022.
- [54] J. Wu, F. Gou, W. Xiong, and X. Zhou, "A reputation value-based task-sharing strategy in opportunistic complex social networks," *Complexity*, vol. 2021, pp. 1–15, Article ID 8554351, 2021.
- [55] Y. Deng, F. Gou, and J. Wu, "Hybrid data transmission scheme based on source node centrality and community reconstruction in opportunistic social networks," *Peer-to-Peer Networking and Applications*, vol. 14, no. 6, pp. 3460–3472, 2021.
- [56] G. Yu, Z. Chen, J. Wu, and Y. Tan, "Medical decision support system for cancer treatment in precision medicine in developing countries," *Expert Systems with Applications*, vol. 186, Article ID 115725, 2021.
- [57] Y. Shen, F. Gou, and Z. Dai, "Osteosarcoma MRI image-assisted segmentation system base on guided aggregated bilateral network," *Mathematics*, vol. 10, no. 7, p. 1090, 2022.
- [58] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Transactions on Medical Imaging*, vol. 25, no. 11, pp. 1451–1461, 2006.
- [59] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.
- [60] Y. Qin, X. Li, J. Wu, and K. Yu, "A management method of chronic diseases in the elderly based on IoT security environment," *Computers & Electrical Engineering*, vol. 102, 2022.
- [61] K. Vinogradova, A. Dibrov, and G. Myers, "Towards interpretable semantic segmentation via gradient-weighted class activation mapping (student abstract)," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 10, pp. 13943–13944, 2020.