Construction of multi-scale feature fusion segmentation model of MRI knee images based on dual attention mechanism weighted aggregation

Xinghui Gai[a], Huifang Cai[a,1], Junying Wang[a,1], Xinyue Li[a], Yan Sui[b], Kang Liu[b] and Dewu Yang[a,*]
[a]*Department of Medical Technique, Beijing Health Vocational College, Beijing, China*
[b]*Department of Radiology, Fuxing Hosptital Affiliated to Capital Medical University, Beijing, China*

Abstract.
BACKGROUND: Early diagnosis of knee osteoarthritis is an important area of research in the field of clinical medicine. Due to the complexity in the MRI imaging sequences and the diverse structure of cartilage, there are many challenges in the segmentation of knee bone and cartilage. Relevant studies have conducted semantic fusion processing through splicing or summing forms, which results in reduced resolution and the accumulation of redundant information.
OBJECTIVE: This study was envisaged to construct an MRI image segmentation model to improve the diagnostic efficiency and accuracy of different grade knee osteoarthritis by adopting the Dual Attention and Multi-scale Feature Fusion Segmentation network (DA-MFFSnet).
METHODS: The feature information of different scales was fused through the Multi-scale Attention Downsample module to extract more accurate feature information, and the Global Attention Upsample module weighted lower-level feature information to reduce the loss of key information.
RESULTS: The collected MRI knee images were screened and labeled, and the study results showed that the segmentation effect of DA-MFFSNet model was closer to that of the manually labeled images. The mean intersection over union, the dice similarity coefficient and the volumetric overlap error was 92.74%, 91.08% and 7.44%, respectively, and the accuracy of the differential diagnosis of knee osteoarthritis was 84.42%.
CONCLUSIONS: The model exhibited better stability and classification effect. Our results indicated that the Dual Attention and Multi-scale Feature Fusion Segmentation model can improve the segmentation effect of MRI knee images in mild and medium knee osteoarthritis, thereby offering an important clinical value and improving the accuracy of the clinical diagnosis.

Keywords: Coronary atherosclerosis, morphological processing, regional growth, auxiliary diagnosis, vascular stenosis

1. Background

Knee osteoarthritis (KOA) is a chronic joint disease characterized by articular cartilage degeneration and secondary bone hyperplasia. The thinning or local coloboma of the articular cartilage causes pain and

---

[1]These authors contributed equally to this work and should be considered co-first authors.
[*]Corresponding author: Dewu Yang, Department of Medical Technique, Beijing Health Vocational College, Beijing 102402, China. E-mail: yangdewu@wjw.beijing.gov.cn.

abnormal joint function, and in severe cases, it may lead to impairment of the mobility [1]. In terms of early diagnosis, evaluating the degree of knee cartilage coloboma, including the area, volume and shape can provide an important basis for intervention treatment and health management. The classification of cartilage injury can be divided into 4 grades: I, II, III and IV. Patients with cartilage injury can take corresponding treatment measures under the guidance of doctors depending on the underlying medical conditions. At the same time, locating an accurate and effective site and shape presentation of the coloboma is an important premise for ensuring the scientific surgical treatment [2].

At present, MRI is one of the main examination methods employed in the clinical diagnosis of knee osteoarthritis. Compared with the conventional X-ray and CT, MRI has the advantages of absence of radiation damage, high specificity and multi-parameter imaging, and has superior performance in the early detection of lesions in cartilage, bone marrow, meniscus, synovium and ligaments. MRI is affected by sequence parameters, wherein tissue contrast presents certain differences, tissue boundaries are blurred, cartilage may be small or long, and background tissue interferences are relatively large. This makes it difficult to identify the anatomy and segment the structure of MRI knee images [3]. Clinical doctors conduct data analysis on the MRI images one by one, and mostly use manual tracing methods to identify the shape, calculate the area and analyze the relevant parameter values of the cartilage defect area. However, this not only consumes basic hospital resources, but also limits the diagnostic accuracy due to the influence of personnel factors. Therefore, it is of great clinical value to segment the MRI knee images by using an intelligent model.

The traditional segmentation algorithm is mainly based on the gray level distribution for selecting, clustering and calculating, which may be affected by the type of tissue and structure or shape of the knee joint. As a result, the segmentation quality is more sensitive to noise, contrast and sharpness. On this basis, most machine learning algorithms use the model training methods to extract information about key anatomical structures or pathological regions by presetting interesting features of images, but this method has high requirements like uniformity and contrast of feature information. The deep learning algorithm of convolutional neural network has high capabilities of feature extraction and information expression, and has become a hot research content in the field of medical image segmentation in recent years [4, 5]. Tanzila Saba et al. described the related image enhancement and segmentation techniques for the detection of knee diseases and analyzed several approaches for the features extraction and segmentation in knee bone cancer [6]. Han Lihong et al. compared the application value of different algorithms of convolutional neural network in MRI image analysis of patients with severe stroke, and summarized the advantages of U-Net deep learning in MRI image segmentation [7]. Jianshe Shi et al. proposed that the automatic segmentation of cardiac MRI based on multi-input fusion network could improve the training speed, thereby improving the efficiency of diagnosis [8]. Huang Tongyuan et al. conducted a study on brain tumor segmentation by magnetic resonance imaging based on DO-UNet model. They achieved automatic segmentation by using attention mechanism and multi-scale fusion algorithm, which further improved the segmentation [9].

Deep learning algorithm has shown increasingly powerful advantages in feature recognition and target segmentation of MRI images. However, in the process of feature fusion, interactions between the high-level features and low-level features are often ignored by means of splicing or summing alone, resulting in reduced information resolution and accumulation of useless information, thus decreasing the segmentation effect of small and narrow knee joint cartilage. Aiming at the existing problems in knee joint segmentation models [10], this study adopted the Dual attention and multi-scale feature fusion segmentation network (DA-MFFSnet) by referring to the relevant research of MRI image deep learning model. The U-shaped architecture of Multi-scale Attention Downsample and Global Attention Upsample was conducted to
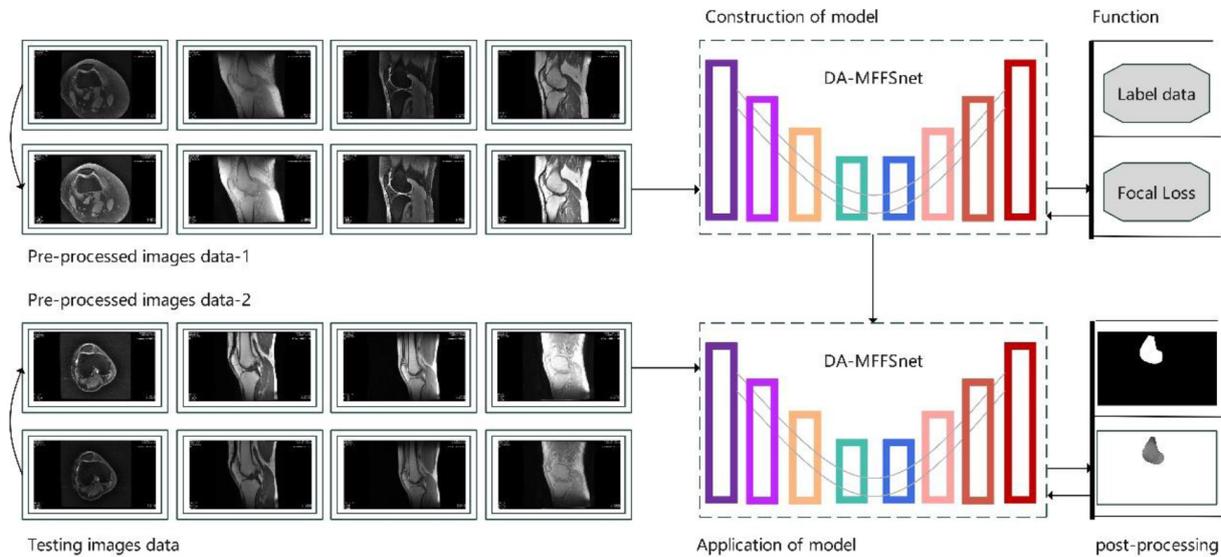
Fig. 1. Schematic representation of the MRI knee segmentation model.

extract highly accurate feature information in the coding process and to reduce the redundant information in the decoding process. The high-quality MRI knee images segmentation can provide more accurate artificial intelligence services for clinical diagnosis and treatment.

2. Methods

Feature information enhancement and clustering of the knee MRI images is the key to segmentation and recognition of bone and cartilage. In this study, an intelligent model of auxiliary diagnosis was constructed using following three stages: the enhancement pre-processing (advance enhancement processing), the network segmentation and the enhancement post-processing. The specific processing flow is shown in Fig. 1. Initially, the original knee MRI images were screened and pre-enhanced. Subsequently, the image sets were manually annotated, and the model was trained based on dual attention mechanism and multi-scale feature fusion algorithm. Finally, the test samples were intelligently segmented and post-enhanced to evaluate the reliability of the model from the auxiliary diagnosis level.

2.1. Advance enhancement processing of images

Image gray value distribution is an important factor affecting the contrast parameters. Histogram equalization is the most common preprocessing method to enhance image contrast. The main principle is to change the image histogram distribution into an approximate uniform distribution, and adaptive histogram equalization can be used when local features of the image are considered. In order to avoid discontinuity and excessive enhancement caused by the Adaptive Histogram Equalization, this study adopted the Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithm to achieve contrast enhancement and noise suppression [11]. The basic principle was to set a threshold value, and when a grayscale histogram of an MRI image exceeded the threshold, it was clipped, and the part exceeding the threshold was evenly distributed to each grayscale level, as shown in Fig. 2.
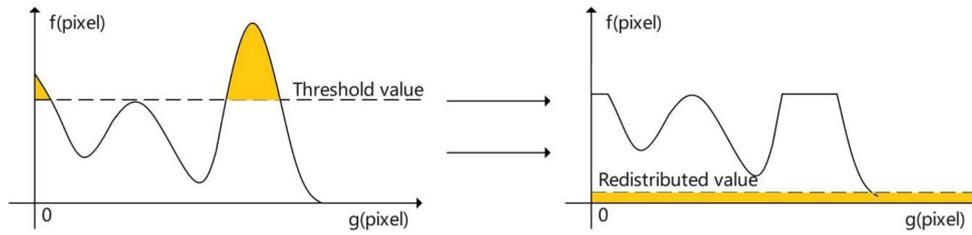
Fig. 2. Processing of restricted histogram distribution in the CLAHE algorithm.
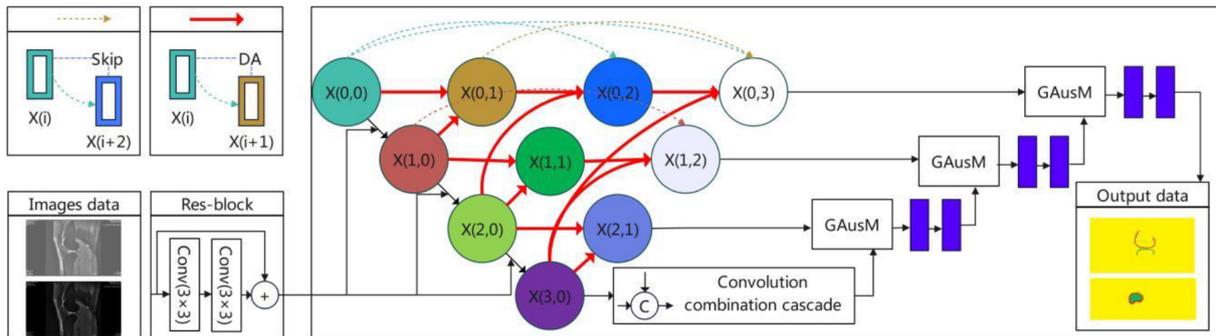

Fig. 3. Network architecture of the DA-MFFSnet model.

2.2. Network of image segmentation

The network architecture of the DA-MFFSnet model (shown in Fig. 3) used skip connections to carry out information transfer between the coding layer and the decoding layer. The encoder in the U-shaped architecture was used for multi-level feature extraction, the decoder was used for feature upsample and image recovery, and the intermediate convolutional cascade layer was used to improve the sensitivity field of the image.

2.2.1. Multi-scale attention subsampling

In convolutional neural networks, high-level features are concerned with the location information of the organs of interest, while low-level features are more concerned with the edge information of the space. The subsampling module uses channel attention and spatial attention to extract multi-scale fusion features, which makes the network segmentation more accurate. In this study, $X(i)$ was set as the feature of the i stage of the encoder, and DA represented the dual attention mechanism of the feature of the i stage and the feature of the $i + 1$ stage. The convolution operation of phase feature images was realized through spatial attention, the convolution kernel was 1×1, and the activation function was ReLU. Channel attention used the global average pooling to perform the Squeeze operation (dimension compression) and assigning weights to the output features, while Sigmoid activation function was used for normalization [12].

2.2.2. Convolutional concatenated intermediate layer

Because pooling operation reduces the resolution of feature maps, this study connected the multi-scale feature information by convolutional concatenated method, thus enlarging the sensitivity field and making it encode higher-level feature information. The specific process consisted of 2 steps: First, the feature pyramid was constructed through four convolution layers with expansion rates of 3, 6, 12 and 18, with a

convolution kernel of 3×3 and channels of 64. Subsequently, a 3×3 depth separable convolution was carried out to achieve information integration.

2.2.3. Global attention upsampling

Global attention upsampling was performed on the dual attention downsampled feature information and convolutional concatenated intermediate layer output information. First, global average pooling was performed on high-level features, and the convolution kernel was 1×1. The normalization and ReLU nonlinear activation function were performed to reduce the channel number and processing pressure. Then, output features were multiplied by pixel units with low-level features. The feature presentation of the region of interest was focused, the accumulation of redundant information was weakened, and the high-level features and the low-level features before output were concatenated [13].

2.3. The loss function

Focal loss function was added in different semantic layers for a better focus of the model training on features of interest in sample images. The calculation method was shown in Eq. (1):

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \tag{1}$$

where, p_t is the probability that the model is predicted as background with value range of [0, 1], α_t is the modulation parameter for the imbalance of positive and negative sample numbers, γ is the modulation parameter for the imbalance of distinguishable and indistinguishable sample numbers with a value range of [0, 5]. $(1 - p_t)^\gamma$ is the modulation factor, when p_t tends to 1, it indicates that the sample is distinguishable. When $(1 - p_t)^\gamma$ tends to 0, a small contribution is made to the loss, thereby reducing the loss proportion of distinguishable samples. When p_t tends to 0, it indicates that the sample is indistinguishable. When $(1 - p_t)^\gamma$ tends to 1, a larger contribution is made to the loss, thereby increasing the loss proportion of indistinguishable samples.

2.4. Image post-enhancement processing

The process of MRI image segmentation of the knee is generally affected by the irregularity of the anatomical structure and the randomness of image speckle noise. In this study, algorithms of the erosion of the image and dilation of morphological opening operation were used to smooth the contour of the segmentation unit, break the narrow neck and eliminate small protrusions within the effective range. The image opening operation called morphologyEx (cv2.MORPH_OPEN) function from the OpenCV-Python and the convolution kernel = np.ones ((3, 3), np.uint8) were employed. The image gradient operation was used to subtract the image after dilation from the image after erosion to enhance the image contour. The morphologyEx (cv2.MORPH_GRADIENT) function from OpenCV-Python and the convolution kernel = np.ones ((5, 5), np.uint8) were employed in this study.

2.5. Auxiliary diagnostic evaluation

The physical characteristics, geometric features and morphological signs of the anatomical structure in MRI images of the knee joint are the key factors in the evaluation of articular cartilage degeneration [14] and secondary bone hyperplasia [15]. The gray mean and standard deviation of the segmentation area are selected as physical indexes. The arithmetic mean deviation, maximum height and average width of the contour are geometric features, and the existence of sag and bulge are morphological features.

Among them, the physical characteristics reflect the distribution characteristics of gray values in the segmentation area, the geometric characteristics reflect the smoothness of the segmentation boundary, and the morphological signs analyze the pathological morphology of the anatomical structure from the perspective of imaging diagnosis of cartilage defect or bone hyperplasia. In this study, the weight of indicators was assigned by the set valued statistics, and the condition of knee osteoarthritis was divided into four levels of normal, mild, moderate and severe according to the diagnostic criteria recommended by the American Association of Rheumatology. An auxiliary diagnostic evaluation model for knee osteoarthritis was constructed [16].

3. Data collection

3.1. General information

From January 2020 and December 2021, 100 patients were randomly selected with suspected knee osteoarthritis admitted to the radiology department of the hospital. After arthroscopy, surgical treatment, and other clinical diagnosis, 59 patients were identified with knee osteoarthritis and 41 cases were normal. There were 53 males and 47 females, aged 42 to 65 years, with a mean age of (51.28 ± 4.69) years, and a mean course of disease of (3.01 ± 1.29) years from 1 to 6 years. Inclusion criteria: Patients underwent magnetic resonance knee imaging, exhibited clinical symptoms of knee pain, swelling and limited activity, and they and their families were aware of the study content and agreed to participate. Exclusion criteria: Patients with a history of severe knee trauma or surgery, MRI images with significant technical defects, and patients with knee disease more than 6 years or severe deformity.

3.2. Examination methods

The knee joint MRI was performed with GE Discovery 750W 3.0T magnetic resonance equipment with AW4.7 image processing workstation. The examined knee joint was placed in the coil with a center alignment to the lower margin of the patellar. The scanning sequence was FSE-T1WI and FSE-PDWI in sagittal position, FSE-PDWI in coronal position and FSE-PDWI in transverse position, and the ZTE sequence was added. Among these, ZTE sequence can achieve the purpose of displaying short T2 components, articular cartilage, and the hierarchical structure of cartilage that cannot be displayed by conventional sequences. This can assist to observe the early damage and provide more enhanced image comparison. Compared with conventional sequences, ZTE imaging has its unique features. Since TE is zero, the articular cartilage can be clearly displayed and interference from adjacent articular fluid is removed.

3.3. Evaluation index

In this study, 2,100 knee MRI images of 100 patients and corresponding label data were used for grouping and training according to different sequences. The ratio of training, verification and test samples was 80:12:8, and the entire label data were labeled by experts with more than 10 years of clinical work experience. Based on manual labeling, the Mean Intersection over Union (MIoU), Volume Overlap Error (VOE), Dice similarity coefficient (DSC), and Mean Pixel Accuracy (MPA) were used to verify the performance of the index algorithm. The equations used for the calculation are shown in Eqs (2)–(5):

$$MIoU = \frac{1}{2} \left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP} \right) \tag{2}$$

(1) Original image   (2)Label   (3) U-Net   (4)U-Net++   (5) attention-U-Net (6)DA-MFFSNet

Fig. 4. Segmentation effect of different algorithms for knee joint.

$$\text{VOE} = 1 - \frac{\text{TN}}{\text{TN} + \text{FN} + \text{FP}} \tag{3}$$

$$\text{DSC} = \frac{2\text{TP}}{\text{FP} + 2\text{TP} + \text{FN}} \tag{4}$$

$$\text{MPA} = \frac{1}{2} \left(\frac{\text{TP}}{\text{TP} + \text{FP}} + \frac{\text{TN}}{\text{TN} + \text{FN}} \right) \tag{5}$$
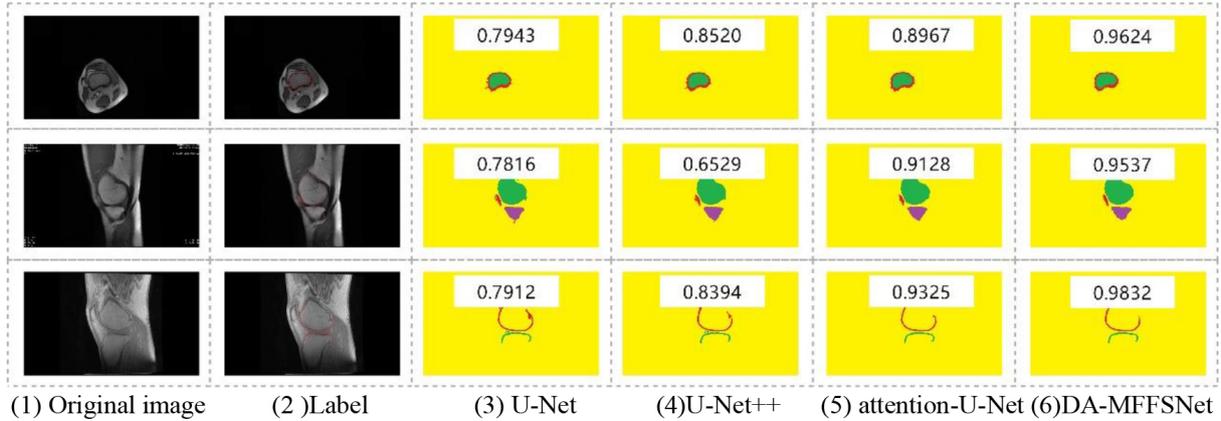
where, TP (true positive) represents those actual labels that were also classified as labels at the time of prediction, FP (false positive) represents the parts that were actually the backgrounds, but were predicted as labels, FN (false negative) represents the actual labels that were classified as backgrounds during the prediction, and TN (true negative) represents the actual backgrounds of those predicted to be the backgrounds.

4. Results and discussion

4.1. Results of knee joint segmentation

The experimental data of U-Net, U-Net++, attention U-Net and DA-MFFSNet algorithms were compared based on the manually labeled images, as shown in Fig. 4, where (1) is the original image, (2) is the label image, (3) is the U-NET segmentation image, (4) is the U-net++ segmentation image, (5) is the attention U-Net segmentation image, (6) is the DA-MFFSNet segmentation image, and the corresponding value is the average pixel accuracy (MPA). The segmentation results show that the DA-MFFSNet algorithm was more beneficial in the segmentation of the edge structure and provided better details of the knee bone and cartilage. Compared with other supervised learning algorithms, DA-MFFSNET algorithm improved the segmentation effect of MRI knee images, and the effect was closer to manual labeled images.

4.2. Performance of attention module

In this study, the dual attention mechanism was used to segment the knee joint features. In order to evaluate the performance of the attention module, Concat was used to replace the attention module while

Table 1
Influence of attention module on the image segmentation performance

Algorithm	DA	GAusM	MioU (%)	DSC (%)	VOE (%)
Concat (−)	−	−	81.53	79.84	14.26
DA-MFFSNet (−+)	−	√	83.29	82.55	13.78
DA-MFFSNet (+−)	√	−	86.01	85.43	11.95
DA-MFFSNet (++)	√	√	92.74	91.48	7.44

maintaining other network parameters. The results are shown in Table 1, "–" indicates that the module was not used, and "√" indicates that the module was used. The average crossover ratio and the dice similarity coefficient of the DA-MFFSNet model was 92.74% and 91.08%, respectively, which was 11.21% and 11.64% higher than that of the Concat model. The volume overlap error of the DA-MFFSNet model was 7.44%, which was 6.82% lower than the Concat model, indicating that the dual attention module improved the accuracy of feature segmentation of knee joint.

4.3. Auxiliary diagnostic level

In this study, the physical characteristics, geometric features and morphological signs of the MRI segmentation images of knee joint was selected as the auxiliary diagnostic data. The C4.5 decision tree algorithm was used to train the diagnostic model, and the optimal evaluation model was obtained through parameter adjustment. Taking 168 images of the test sample set as an example, the ROC curve analysis was performed according to clinical diagnostic criteria. The accuracy of differentiation of normal, mild, moderate and severe knee osteoarthritis was 92.15%, 79.55%, 81.23% and 84.76%, respectively, and the AUC of normal and severe knee osteoarthritis was 0.8904 and 0.8517, respectively. The results indicate a better diagnostic stability and classification effect.

5. Conclusions

With the continuous innovation and wide application in clinical practice, MRI can be used in the diagnosis of knee joint trauma and surrounding tissue diseases, such as cartilage diseases, meniscus abnormalities, ligament injuries, joint effusion and bone damage, etc. Manual tracing or semi-automatic segmentation methods are often used for the objective and quantitative evaluation of the cartilage defects. However, considering the anatomical structure, the slender and narrow cartilage and defect areas are very difficult to distinguish, thereby increasing the difficulty in diagnosing the disease. Traditional image segmentation algorithms provide a difficulty in identifying cartilage defect lesions from backgrounds with low contrast. Furthermore, the method of image-by-image tracing is not only a waste of labor cost [17], but also susceptible to the influence of the years and professional experience of clinicians. As a result, the construction of feature segmentation models for different clinical diagnostic purposes has become a research hotspot in MRI knee image analysis.

In order to solve the above problems, this study constructed a multi-scale feature fusion segmentation model using dual attention mechanism to improve the segmentation accuracy of the bone and cartilage tissue of the knee joint. Compared with U-Net, U-NET++ and other segmentation algorithms, the0 employed DA-MFFSNet module included Multi-scale Attention Downsample module and the Global Attention Upsample module, and was integrated in U-type coding-decoding network to realize effective extraction of spatial information and detailed features. The results showed that the segmentation performance of the module was better than other algorithms. It was effective and robust for the segmentation of

cartilage with low tissue contrast and small target. At the same time, considering the complexity of MRI sequences, image features with the same organizational structure are quite different, which limits the accuracy of segmentation results. In future studies, the training sample size will be increased according to different anatomical structure types and imaging sequences, convolution kernel size will be adjusted, algorithm parameters will be optimized, and auxiliary diagnostic evaluation indicators will be added to improve the quality of MRI knee image segmentation. For different sequences of MRI, the algorithm will be optimized from the sequence dimension to avoid the influence of sequence factors on the accuracy of the algorithms.

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] Mirzaii-Dizgah MR, Mirzaii-Dizgah MH, MirzaiiDizgah I, Karami M, Forogh B. Osteoprotegerin changes in saliva and serum of patients with knee osteoarthritis. Revista Espanola de Cirugia Ortopedica y Traumatologia. 2021; 66(1): 47-51.

[2] Thomas Abbey C, Simon Janet E, Evans Rachel, Turner Michael J, Vela Luzita I, Gribble Phillip A. Knee surgery is associated with greater odds of knee osteoarthritis diagnosis. Journal of Sport Rehabilitation. 2019; 28(7): 716-723.

[3] Majidi H, Niksolat F, Anbari K. Comparing the accuracy of radiography and sonography in detection of knee osteoarthritis: A diagnostic study. Open Access Macedonian Journal of Medical Sciences. 2019; 7(23): 4015-4018.

[4] Karim Md. R, Jiao J, Doehmen T, Cochez M, Beyan O, Rebholz Schuhmann D, Decker S. DeepKneeExplainer: Explainable Knee Osteoarthritis Diagnosis from Radiographs and Magnetic Resonance Imaging. IEEE ACCESS. 2021; 9: 39757-39780.

[5] Ahmed SM, Mstafa RJ. A comprehensive survey on bone segmentation techniques in knee osteoarthritis research: from conventional methods to deep learning. Diagnostics. 2022; 12(3): 611-611.

[6] Saba T, Rehman A, Mehmood Z, Kolivand H, Sharif M. Image enhancement and segmentation techniques for detection of knee joint diseases: A survey. Current Medical Imaging Reviews. 2018; 14(5): 704-715.

[7] Shi JS, Ye YG, Zhu DX, Su LT, Huang YF, Huang JH. Automatic segmentation of cardiac magnetic resonance images based on multi-input fusion network. Computer Methods and Programs in Biomedicine. 2021; 209(prepublish): 106323.

[8] Huang TY, Liu Y. Research on the magnetic resonance imaging brain tumor segmentation algorithm based on DO-UNet. International Journal of Imaging Systems and Technology. 2022; 33(1): 143-157.

[9] Liu F, Wang HB, Liang SNi, Jin Z, Wei SC, Li XJ. MPS-FFA: A multiplane and multiscale feature fusion attention network for Alzheimer's disease prediction with structural MRI. Computers in Biology and Medicine. 2023; 157: 106790-106790.

[10] Lu JF, Ren HP, Shi MT, Cui C, Zhang SQ, Emam M, Li L. A novel hybridoma cell segmentation method based on multi-scale feature fusion and dual attention network. Electronics. 2023; 12(4): 979-979.

[11] Mojdeh M, Tavakoli Tafti K, Soltani P. Evaluation of histogram equalization and contrast limited adaptive histogram equalization effect on image quality and fractal dimensions of digital periapical radiographs. Oral Radiology. 2022; 39(2): 418-424.

[12] Hou GM, Qin JH, Xiang XY, Tan Y, Neal N. X. AF-Net: A medical image segmentation network based on attention mechanism and feature fusion. Computers, Materials & Continua. 2021; 69(2): 1877-1891.

[13] Liang BT, Tang C, Xu M, Wu TB, Lei ZK. Fusion network based on the dual attention mechanism and atrous spatial pyramid pooling for automatic segmentation in retinal vessel images. Journal of the Optical Society of America. A, Optics, Image Science, and Vision. 2022; 39(8): 1393-1402.

[14] Francisco X, André V, Cristina V, et al. Magnetic resonance imaging is able to detect patellofemoral focal cartilage injuries: A systematic review with meta-analysis. Knee Surgery, Sports Traumatology, Arthroscopy: Official Journal of the ESSKA. 2022; 31(6): 2469-2481.

[15] Yuen J, Miller KJ, Klassen BT, Lehman VT, Lee KH, Kaufmann TJ. Hyperostosis in combination with low skull density ratio: A potential contraindication for magnetic resonance imaging-guided focused ultrasound thalamotomy. Mayo Clinic Proceedings: Innovations, Quality Outcomes. 2022; 6(1): 10-15.

[16] Arunrukthavon P, Heebthamai D, Benchasiriluck P, Chaluay S, Chotanaphuti T, Khuangsirikul S. Can urinary CTX-II be a biomarker for knee osteoarthritis? Arthroplasty. 2020; 2(2): 185-199.

[17]   Andersen S, Hittle B, Keith JP, Powell K, Wiet G. Pipeline for automated processing of clinical cone-beam computed tomography for patient-specific temporal bone simulation: Validation and clinical feasibility. Otology Neurotology. 2023; 44(2): e88-e94.