

U-Net Based Dual-Pooling Segmentation of Bone Metastases in Thoracic SPECT Bone Scintigrams

Yang He¹, Qiang Lin^{2*}, Yongchun Cao², Zhengxing Man²

¹Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, China

²School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou, China
Email: 748428952@qq.com, *qiang.lin2020@hotmail.com, 408690991@qq.com, wwuh@qq.com

How to cite this paper: He, Y., Lin, Q., Cao, Y.C. and Man, Z.X. (2024) U-Net Based Dual-Pooling Segmentation of Bone Metastases in Thoracic SPECT Bone Scintigrams. *Journal of Computer and Communications*, 12, 60-71.
<https://doi.org/10.4236/jcc.2024.124006>

Received: March 5, 2024

Accepted: April 12, 2024

Published: April 15, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In order to enhance the performance of the CNN-based segmentation models for bone metastases, this study proposes a segmentation method that integrates dual-pooling, DAC, and RMP modules. The network consists of distinct feature encoding and decoding stages, with dual-pooling modules employed in encoding stages to maintain the background information needed for bone scintigrams diagnosis. Both the DAC and RMP modules are utilized in the bottleneck layer to address the multi-scale problem of metastatic lesions. Experimental evaluations on 306 clinical SPECT data have demonstrated that the proposed method showcases a substantial improvement in both DSC and Recall scores by 3.28% and 6.55% compared the baseline. Exhaustive case studies illustrate the superiority of the methodology.

Keywords

Tumor Bone Metastasis, Bone Scintigram, Lesion Segmentation, CNN, Dual Pooling

1. Introduction

Bone metastasis is a prevalent complication of malignant tumors, commonly observed in parenchymal tumors such as breast, prostate, and lung cancers, with an incidence ranging from 30% to 75% [1]. During metastasis, tumor cells interact with osteoblasts, osteoclasts and bone stromal cells, potentially leading to pain, fractures, dysfunction, and psychological distress as they disrupt bone tissue, significantly impacting the life of tumor patients [2]. However, early symp-

toms of bone metastasis are often subtle, and the onset of pain often signifies that the optimal window for treatment has passed. So far, patients can only mitigate the risk of mortality through screening, early diagnosis and prompt intervention [3].

Single Photon Emission Computed Tomography (SPECT) is a cost-effective functional imaging modality with high detection sensitivity and has been commonly used clinically as a screening tool for bone metastases. However, due to the disparities between imaging modalities and the human body, SPECT imaging results are afflicted with drawbacks such as poor spatial resolution, low specificity, and susceptibility to noise interference [4]. Although the fidelity and quantitative accuracy of SPECT images have markedly enhanced with advancements in imaging technologies and innovations in contrast agents, discrepancies in medical personnel skills and the influence of various subjective and objective factors may lead to incorrect diagnostic outcomes. To enhance the efficiency and accuracy of doctors' diagnoses, scientists are endeavoring to develop assisted diagnosis systems to automate medical image analysis, wherein machine learning algorithms, particularly deep learning methods, play a pivotal role. Convolutional Neural Networks (CNNs) can autonomously learn features without human intervention and have demonstrated the efficacy of this approach [5] [6] [7] [8] [9].

Medical image segmentation is an extension of image segmentation in the medical field, serving as the foundation for extracting pathological regions, conducting clinical trials, measuring specific tissues, and facilitating three-dimensional reconstruction through the delineation of regions of interest. Researchers have explored the utilization of image segmentation techniques to segment lesions, yielding numerous achievements across various modalities such as CT [10] [11], MRI [12] [13], and ultrasound [14]. However, studies on SPECT images, characterized by poor imaging quality and limited manipulable objects, have been relatively scarce, mostly focusing on binary classification or binary segmentation tasks [15] [16] [17] [18] [19]. Lin *et al.* [18] devised a Res-U-Net model by integrating residual modules [20] into the U-Net [21] base model and conducted experiments on an extended dataset of 2280 SPECT lung cancer thorax images, achieving CPA, Recall, and IoU values of 0.7721, 0.6788, and 0.6103, respectively. In another study, Cao *et al.* [19] addressed the poor imaging quality of SPECT images by incorporating data fusion techniques to enhance data quality. They also introduced the concepts of dense connectivity and deep supervision into the model, resulting in improved segmentation performance.

It is noteworthy that the majority of current research efforts in the domain of SPECT images are built upon existing networks developed for traditional natural images or other modalities. Consequently, there is a paucity of designs tailored specifically for SPECT images themselves. Building upon these observations, our work focuses on adapting existing networks to SPECT and amalgamating them with existing methodologies. Specifically, we analyzed the clinical diagnostic pattern of nuclear medicine physicians and used the dual-pooling module to

achieve a balance between highlighting the foreground and maintaining the background information, as physicians need to compare the hotspots with their periphery to screen the suspicious hotspots, which makes the network more in line with the clinical diagnostic patterns; secondly, we paid attention to the multiscale problem of the bone metastatic lesions, and we borrowed the solution of CE-Net [22] and tried to reconstruct the region of the receptive field; Finally, we evaluate the constructed network's segmentation performance on clinical data. The primary contributions of this work are outlined as follows:

Firstly, we tackle the research problem of lesion segmentation in SPECT images, recognizing significant research potential in this direction. Secondly, we endeavor to enhance the segmentation performance of the network by improving its adaptability to SPECT images. Finally, we employ a set of previously unobserved SPECT images for clinical diagnosis of lung cancer patients as a test set to evaluate the model's performance.

2. Materials and Methods

2.1. Materials

All the data used in the experiment were sourced from the clinical data generated during the bone metastasis examination of patients by the Department of Nuclear Medicine at Gansu Cancer Hospital from January 2014 to December 2019. A total of 410 images of 205 patients were collected. ^{99m}Tc -MDP bone imaging agent was injected into patients without allergic reaction, and after several hours of metabolism, the patient's anteroposterior data were captured with Siemens multispectral gamma camera. The final results for each body position are presented as a 16-bit unsigned matrix of size 1024×256 , where each element value represents the intensity of the radiation value at the current position and the distance between pixels is 2.16 mm. We selected 306 images from 410 images, and extracted the patients' thoracic regions.

2.2. Method

CNNs have dominated the field of computer vision and spawned a series of semantic segmentation models. Among them, the U-Net network stands out for its effectiveness in addressing the challenge of medical segmentation with small sample sizes, and it has gradually become a standard architecture in the field. **Figure 1** illustrates the proposed model architecture based on U-Net in this work.

On the whole, U-Net is an encoder-decoder architecture, consisting of a contracting path on the left and an expansive path on the right. The contracting path extracts image features by pixel by stacking multiple combinations of convolution and pooling. Similarly, the expansive path stacks the combination of multiple convolutions and up-sampling operations to restore the original size of the image, and fuses the features corresponding to the two paths through skip connections to retain more location information.

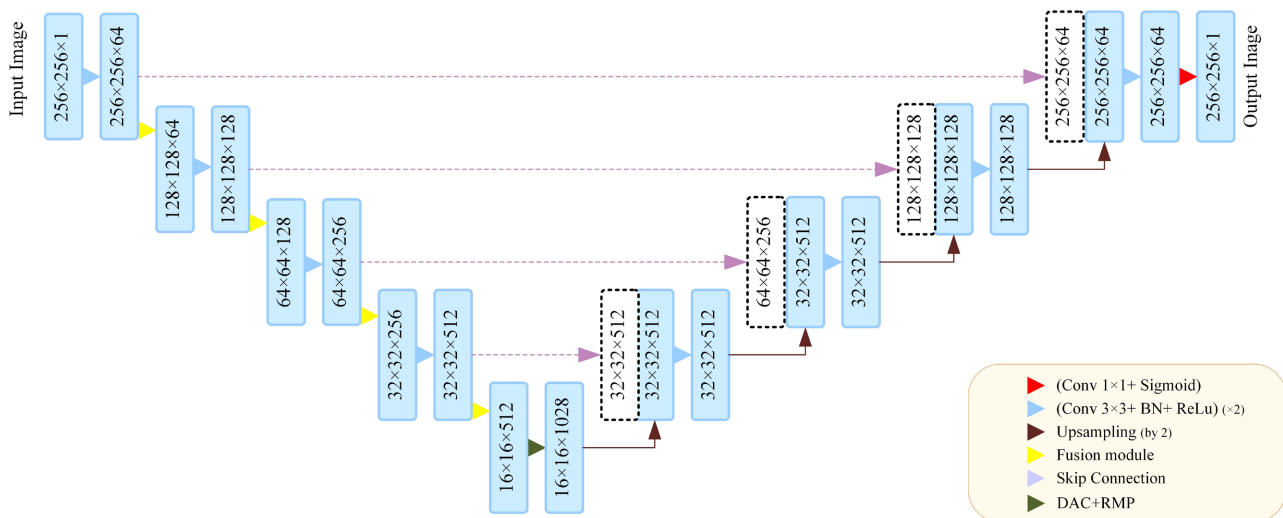


Figure 1. The overall architecture diagram of the proposed model.

2.2.1. Dual-Pooling and Feature Fusion

CNN is a special type of DNN. Generally, a pooling layer is inserted after the convolution block. Pooling layer provides some form of invariance for the network, and reduces the computational complexity of the upper layer by eliminating some connections between convolution layers, so as to reduce the computational cost and control over fitting.

In deep neural networks, various methods are employed to implement pooling operations, with max pooling and average pooling standing out as the most prevalent techniques in CNN. Average pooling performs down-sampling by dividing the input into rectangular pooled regions and calculating the average value of each region, while max pooling is to maximize the target region. The difference between the two methods is that the feature maps obtained by max pooling are more sensitive to texture feature information. Average pooling is to average the images in the pooled area, and the feature information obtained in this way is more sensitive to background information. At present, semantic segmentation networks often use max pooling for down-sampling, such as U-Net, U-Net++ [23] and other models. These networks are not designed for SPECT images originally.

However, in the routine diagnosis of SPECT images, the determination of the hotspots needs to take into account the background. For example, the ratio of target background ratio to measure the tumor tissue radioactivity and the designated normal tissue radioactivity is a semi quantitative indicator for the determination of the focus. Therefore, the use of average pooling in the down-sampling process has better adaptability to SPECT images, while taking into account that in the case of small samples, texture features can reduce the data demand for the model. Therefore, the max and average pooling are used for the network feature maps to conduct down-sampling at the same time, and the sampling results are fused to provide richer features for the model.

The classical feature fusion method generally adopts the method of addition

or channel splicing. The former directly adds two features element-by-element, similar to the superposition of information; The latter splices the two feature vectors to increase the feature dimensions of the image, while the information contained in the features in each channel remains unchanged. However, considering that the two pooling results come from the same feature maps in the application process and have a certain degree of spatial consistency, the model is based on the idea of coordinate attention [24] and changed to a dual input module. Specifically, coordinate attention mechanism coordinate attention decomposes channel attention into two one-dimensional feature encoding processes, which aggregate features along two spatial directions. In this way, remote dependencies can be captured in one spatial direction, while accurate location information can be saved in another spatial direction. In order to effectively combine the two inputs, we encode the max pooling results in two directions. The weight of the results is multiplied by the average pooling results. The modified feature fusion module is shown in Figure 2.

2.2.2. Multi-Scale Feature Extraction

Bone metastatic lesions exhibit considerable variability in size. We introduced modules designed in CE-Net and redesigned the size of the receptive field.

DAC module: This module improves the convolution operation by using four cascading branches with receptive fields of 3, 7, 9, 19 to encode the high-level semantic feature mapping, and finally adding the original feature through skip connection. Generally, the convolution of large receptive fields can extract and generate more abstract features for large objects, while the convolution of small receptive fields is better for small objects. By combining cavity convolutions with different dilation rates, DAC blocks can extract features of objects with different sizes.

RMP module: One of the challenges in medical image segmentation is the huge variation in target size. For example, metaphase or advanced tumors may be much larger than early tumors. Residual multi-core pooling encodes global context information by building multiple pooling branches of different sizes. In this experiment, we used 2×2 , 3×3 , 4×4 and 5×5 pooling to construct the

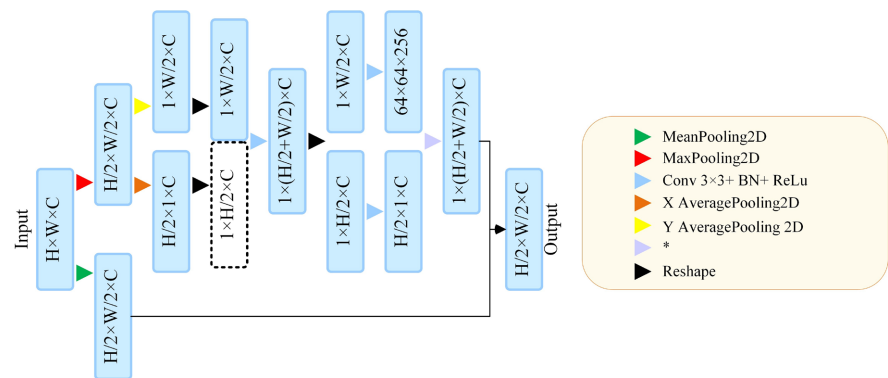


Figure 2. Schematic diagram of the fusion module structure.

module. To reduce the dimensionality and computational cost of the weights, each pooling operation is followed by a 1×1 convolution to modify the number of channels to 1. The low-dimensional feature maps are then up-sampled to obtain the features of the original feature map size. These results will be combined with the original features on the channel dimension to generate feature maps with more information.

In the experiment, we add the above two modules to the bottleneck part of the U-Net network without making other major adjustments to the encoding and decoding parts. It is worth noting that the DAC module uses a large number of dilated convolutions to expand the range of receptive fields, which may lead to the existence of grid problems and loss of information continuity, which is fatal for pixel level classification tasks. For this reason, the DAC has been redesigned, as shown in **Figure 3(a)** and **Figure 3(b)**. The difference between the two is that in the former we directly modified the 1×1 convolution in the last layer to 3×3 and the latter combination of convolution kernels was optimized by constructing a perceptual field combination of 3, 5, 7 and 9 and comparing it with the previous two.

3. Experimental Results and Discussion

In this section, we present the overall results of the experiment. In order to explore the reasons for the improvement of performance, we used some visual methods to display the final results in the experiment as much as possible

3.1. Experimental Setup

The experimental evaluation metrics used in the experiment include CPA, DSC (Dice Similarity Coefficient) and Recall, which are defined in Equations (1)-(3).

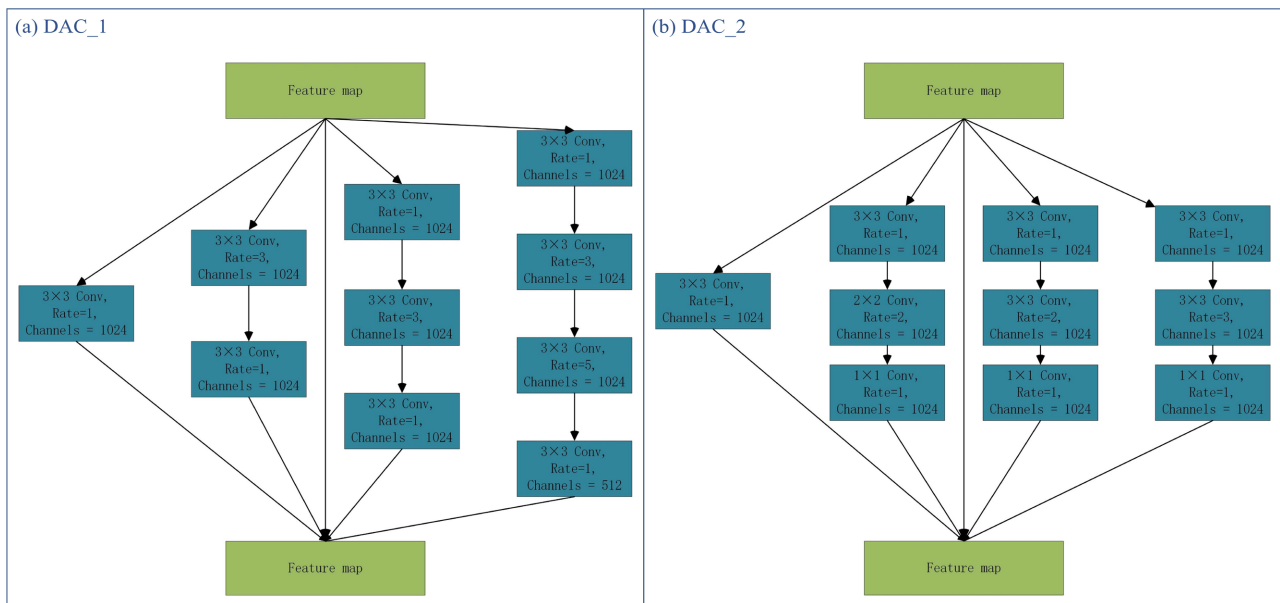


Figure 3. The updated DAC modules used in the experiment.

$$DSC = \frac{2TP}{2TP + FP + FN}, \quad (1)$$

$$Recall = \frac{TP}{TP + FN}, \quad (2)$$

$$CPA = \frac{TP}{TP + FP}, \quad (3)$$

where TP = True Positive, FP = False Positive, and FN = False Negative.

The initial data set was randomly divided into training set and test set according to the ratio of 8:2.224 images were used to train the model, and the remaining 62 images were used to test the performance of the model.

The parameter settings of the segmentation models used are provided in **Table 1**. The experiments are run in TensorFlow 2.0 platform on an Intel Core i7-9700 PC with 32GB RAM running Windows 10.

A trained segmentation model was run 5 times on the test set in order to reduce the effects of randomness as much as possible. For each of the defined evaluation metrics above, the resultant experimental results reported are the average of the 5 running results.

3.2. Results

The experimental results are shown in **Table 2**. Compared to the baseline model, the proposed method shows a large improvement in DSC and Recall, by 3.28% and 6.25%, respectively. The great improvement in Recall indicates that the method is superior in the detection of metastatic lesions, which can further enhance the reliability of the method. On the other hand, networks that incorporate background information are consistently superior to the network solely utilizing the max pooling layer. This enhancement is particularly noticeable in improving the recall metric, which reaches an average improvement of 3.81%. The improved performance can be attributed to the inclusion of background information,

Table 1. Parameter settings used in the experiments.

Para.	Learning rate	Optimizer	Batch size	Epoch
Value	1e-4	Adam	32	200

Table 2. Experimental scores of defined evaluation metrics obtained by different methods.

Method	Pooling	DAC	RMP	DSC	Recall	CPA
U-Net	Max	×	×	0.6086	0.5157	0.7467
	Average	×	×	0.6255	0.5646	0.7023
Ours	Fusion	×	×	0.6291	0.5431	0.7476
		Raw	✓	0.6363	0.5634	0.7318
		DAC_1	✓	0.6363	0.5419	0.7553
		DAC_2	✓	0.6414	0.5812	0.7169

which furnishes valuable reference features crucial for hotspot determination, aligning more closely with the diagnostic practices of clinical physicians.

Additionally, models that incorporated the DAC module-based model provided a larger receptive field area, enhancing the model's ability to detect multifocal lesion features. DAC_2, denoted as the design approach with no checkerboard effect and reduced redundancy, exhibits superior performance. This outcome offers a valuable reference point for optimizing future bone imaging models.

3.3. Discussion

In this section, we illustrate the efficacy of the methods through several case studies, demonstrating their capability to enhance segmentation performance. Furthermore, we conduct an analysis to elucidate the factors contributing to instances of relatively low segmentation performance by the models.

3.3.1. The Effect of Average Pooling

Figure 4 depicts the impact of average pooling on the segmentation performance in SPECT images. As shown in Case #1 in **Figure 4**, the human spine region has formed a large-scale concentration phenomenon, but there are many missed reports in the prediction results of the original U-Net model, such as the lesions on the upper left, which is not ideal for clinical auxiliary diagnosis. After replacing with average pooling, the model will incorporate the background information into the reference when extracting features, improving the prediction ability of the model for the lesions that are difficult to observe. But at the same time, the model's high sensitivity to SPECT data is also easy to be amplified, resulting in more false positives, as shown in Case #2 in **Figure 4**. In these two figures, the U-Net model based on max pooling can better predict the location and contour of the lesions, but after changing the pooling strategy, some local hot spots will also be predicted as lesions, which improves the false alarm rate of the model results.

3.3.2. The Effect of New DAC Modules

Figure 5 compares the prediction results of different DAC module models after

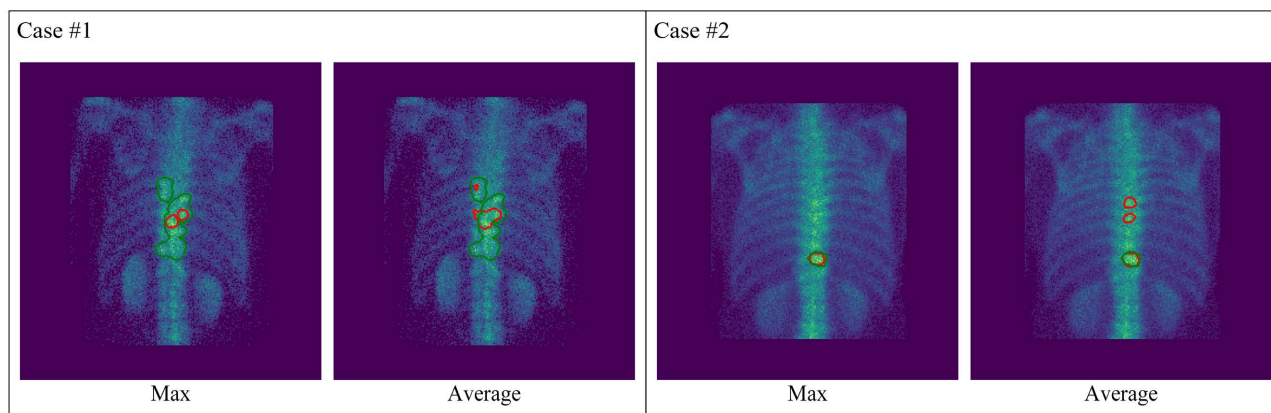


Figure 4. Illustrations of how average pooling affects the segmentation performance, with the green and red circles indicating the ground truth (lesions) and model's predictions.

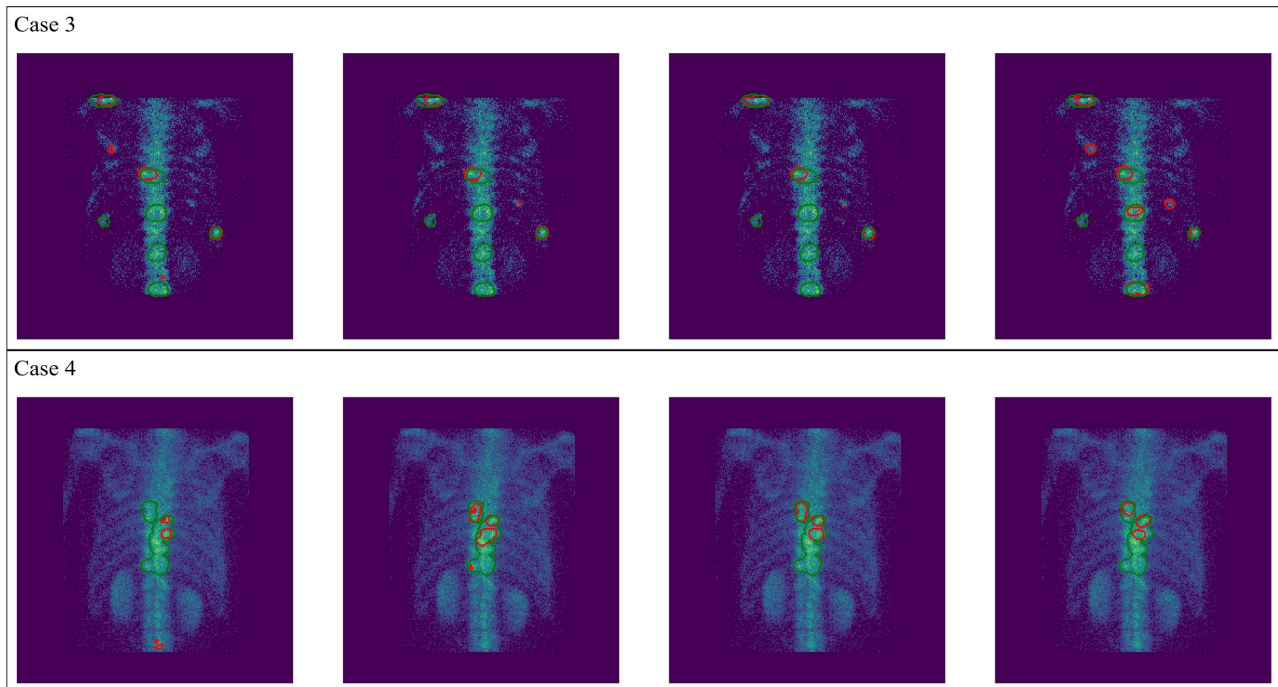


Figure 5. Illustrations of how different DAC modules affect the segmentation performance, from left to right are none, raw DAC, DAC_1, and DAC_2.

feature fusion. The redesigned cavity convolution combination achieves the best results. From the prediction results, the vertebral body region is a hot area, which adds difficulties to the prediction of the model. The redesigned cavity convolution combination can improve the prediction ability of the model in this area, which is important for the cone, which is a part of the bone metastasis prone area. Secondly, the midpoint floc prediction in the model prediction results is relatively reduced, which can ensure the continuity of the prediction area. However, the redesigned cavity convolution still falls short in terms of receptive field range compared to the previous two, as a large receptive field range can better help the model predict large scale regions.

4. Conclusion

In view of the improvement of automatic segmentation performance of SPECT lung cancer bone metastases, this paper reconsiders the down-sampling strategy in the traditional U-Net network, draws on the clinical diagnostic patterns of nuclear medicine physicians, and establishes an analogous mapping between the relevant patterns and the sampling method. Using clinical SPECT data for model evaluation, the average DSC reached 0.64143, which shows the research significance of average pooling and enhanced information retention in SPECT image segmentation. In the near future, we plan to expand this work in the following directions. First, a more complete SPECT image data set is established, and the data set available for training can be further expanded through data enhancement to test the feasible development direction of the depth semantic segmenta-

tion model; Secondly, we will work on a large number of false positives in the model to improve the overall performance of the model. Finally, dilated convolution has significant advantages in segmentation, and it remains to be investigated how to efficiently design the receptive field configuration of the network to fit the research work on bone imaging.

Ethics Approval

The study was approved by the Ethics Committee of Gansu Provincial Tumor Hospital (Approval No.: A202106100014).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Shi, Y., Jiang, Z. and Zhang, L. (2010) Chinese Expert Consensus Statement on Clinical Diagnosis and Treatment of Malignant Tumor Bone Metastasis and Bone Related Diseases. *Chinese-German Journal of Clinical Oncology*, **9**, 1-12. <https://doi.org/10.1007/s10330-009-0188-2>
- [2] Dong, Z., Zhao, J., Liu, C., *et al.* (2019) Expert Consensus on the Diagnosis and Treatment of Bone Metastasis in Lung Cancer. *Chinese Journal of Lung Cancer*, **22**, 187-207.
- [3] He, J., Li, N., Chen, W., *et al.* (2021) China Guideline for the Screening and Early Detection of Lung Cancer (2021, Beijing). *Clinical Medicine of China*, **30**, 193-207.
- [4] Collins, F.S. and Varmus, H.A. (2015) New Initiative on Precision Medicine. *The New England Journal of Medicine*, **372**, 793-795. <https://doi.org/10.1056/NEJMp1500523>
- [5] Lu, S., Lu, Z. and Zhang, Y.D. (2019) Pathological Brain Detection Based on Alex-Net and Transfer Learning. *Journal of Computational Science*, **30**, 41-47. <https://doi.org/10.1016/j.jocs.2018.11.008>
- [6] Li, W. (2015) Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Network. *Journal of Computing and Communication*, **3**, 146-151. <https://doi.org/10.4236/jcc.2015.311023>
- [7] Menze, B.H., Jakab, A., Bauer, S., *et al.* (2014) The Multimodal Brain Tumor Image Segmentation Benchmark (Brats). *IEEE Transactions on Medical Imaging*, **34**, 1993-2024. <https://doi.org/10.1109/TMI.2014.2377694>
- [8] Cheng, J., Liu, Y., Xu, F., *et al.* (2013) Superpixel Classification Based Optic Disc and Optic Cup Segmentation for Glaucoma Screening. *IEEE Transactions on Medical Imaging*, **32**, 1019-1032. <https://doi.org/10.1109/TMI.2013.2247770>
- [9] Song, T.H., Sanchez, V., ElDaly, H., *et al.* (2017) Dual-Channel Active Contour Model for Megakaryocytic Cell Segmentation in Bone Marrow Trepine Histology Images. *IEEE Transactions on Biomedical Engineering*, **64**, 2913-2923. <https://doi.org/10.1109/TBME.2017.2690863>
- [10] Ghosh, S., Das, I., Das, N., *et al.* (2019) Understanding Deep Learning Techniques for Image Segmentation. *ACM Computing Surveys*, **52**, Article No. 73. <https://doi.org/10.1145/3329784>

- [11] Christ, P.F., Elshaer, M.E., Ettliger, F., et al. (2016) Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. *International Conference on Medical Image Computing and Computer Assisted Intervention*, Athens, 17-21 October 2016, 415-423. https://doi.org/10.1007/978-3-319-46723-8_48
- [12] Vorontsov, E., Tang, A., Pal, C., et al. (2018) Liver Lesion Segmentation Informed by Joint Liver Segmentation. *International Symposium on Biomedical Imaging*, Washington DC, 4-7 April 2018, 1332-1335. <https://doi.org/10.1109/ISBI.2018.8363817>
- [13] Poudel, R.P., Lamata, P. and Montana, G. (2017) Recurrent Fully Convolutional Neural Networks for Multi-Slice MRI Cardiac Segmentation. *1st International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016*, Athens, 17 October 2016, 83-94. https://doi.org/10.1007/978-3-319-52280-7_8
- [14] Cui, Z., Yang, J. and Qiao, Y. (2016) Brain MRI Segmentation with Patch-Based CNN Approach. *Chinese Control Conference*, Chengdu, 27-29 July 2016, 7026-7031. <https://doi.org/10.1109/ChiCC.2016.7554465>
- [15] Luo, M., Lin, Q. and Li, T. (2021) SPECT Bone Imaging Thyroid Lesion Segmentation. *International Conference on Big Data and Intelligent Algorithms*, Orlando, 15-18 December 2021, 9-11.
- [16] Gao, R., Lin, Q., Man, Z., et al. (2021) Deep Learning-Based Segmentation of Arthritis Lesions in SPECT Images. *Journal of Northwest Minzu University (Natural Science)*, **42**, 22-30+37.
- [17] Che, G., Cao, Y., Zhu, A., et al. (2021) Segmentation of Bone Metastases Based on Attention Mechanism. *IEEE International Conference on Power Electronics, Computer Applications*, Shenyang, 22-24 January 2021, 259-263. <https://doi.org/10.1109/ICPECA51329.2021.9362531>
- [18] Lin, Q., Luo, M., Gao, R., et al. (2020) Deep Learning Based Automatic Segmentation of Metastasis Hotspots in Thorax Bone SPECT Images. *PLOS ONE*, **15**, e0243253. <https://doi.org/10.1371/journal.pone.0243253>
- [19] Cao, Y., Liu, L., Chen, X., et al. (2023) Segmentation of Lung Cancer-Caused Metastatic Lesions in Bone Scan Images Using Self-Defined Model with Deep Supervision. *Biomedical Signal Processing and Control*, **79**, Article ID: 104068. <https://doi.org/10.1016/j.bspc.2022.104068>
- [20] He, K., Zhang, X., Ren, S., et al. (2016) Deep Residual Learning for Image Recognition. *Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [21] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [22] Gu, Z., Cheng, J., Fu H., et al. (2019) CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, **38**, 2281-2292. <https://doi.org/10.1109/TMI.2019.2903562>
- [23] Zhou, Z., Siddiquee, M.M., Tajbakhsh, N., et al. (2018) UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018*, Granada, 20 September 2018, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1
- [24] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate Attention for Efficient Mobile

Network Design. *Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 13708-13717.

<https://doi.org/10.1109/CVPR46437.2021.01350>