Original Article

Multiple myeloma segmentation net (MMNet): an encoder-decoder-based deep multiscale feature fusion model for multiple myeloma segmentation in magnetic resonance imaging

Xin Zhao[1], Lili Chen[1], Nannan Zhang[2], Yuchan Lv[2], Xue Hu[2]

[1]School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing, China; [2]The Department of Blood Transfusion, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China

Contributions: (I) Conception and design: X Zhao, L Chen; (II) Administrative support: L Chen, X Hu; (III) Provision of study materials or patients: N Zhang, Y Lv, X Hu; (IV) Collection and assembly of data: N Zhang, Y Lv, X Hu; (V) Data analysis and interpretation: X Zhao, L Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Lili Chen, PhD. School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, 66 Xuefu Avenue, South Bank District, Chongqing 400074, China. Email: clili522@cqjtu.edu.cn; Xue Hu, PhD. The Department of Blood Transfusion, The First Affiliated Hospital of Chongqing Medical University, 1 Youyi Road, Yuzhong District, Chongqing 400074, China. Email: huxue@hospital.cqmu.edu.cn.

Background: Patients with multiple myeloma (MM), a malignant disease involving bone marrow plasma cells, shows significant susceptibility to bone degradation, impairing normal hematopoietic function. The accurate and effective segmentation of MM lesion areas is crucial for the early detection and diagnosis of myeloma. However, the presence of complex shape variations, boundary ambiguities, and multiscale lesion areas, ranging from punctate lesions to extensive bone damage, presents a formidable challenge in achieving precise segmentation. This study thus aimed to develop a more accurate and robust segmentation method for MM lesions by extracting rich multiscale features.

Methods: In this paper, we propose a novel, multiscale feature fusion encoding-decoding model architecture specifically designed for MM segmentation. In the encoding stage, our proposed multiscale feature extraction module, dilated dense connected net (DCNet), is employed to systematically extract multiscale features, thereby augmenting the model's sensing field. In the decoding stage, we propose the CBAM-atrous spatial pyramid pooling (CASPP) module to enhance the extraction of multiscale features, enabling the model to dynamically prioritize both channel and spatial information. Subsequently, these features are concatenated with the final output feature map to optimize segmentation outcomes. At the feature fusion bottleneck layer, we incorporate the dynamic feature fusion (DyCat) module into the skip connection to dynamically adjust feature extraction parameters and fusion processes.

Results: We assessed the efficacy of our approach using a proprietary dataset of MM, yielding notable advancements. Our dataset comprised 753 magnetic resonance imaging (MRI) two-dimensional (2D) slice images of the spinal regions from 45 patients with MM, along with their corresponding ground truth labels. These images were primarily obtained from three sequences: T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), and short tau inversion recovery (STIR). Using image augmentation techniques, we expanded the dataset to 3,000 images, which were employed for both model training and prediction. Among these, 2,400 images were allocated for training purposes, while 600 images were reserved for validation and testing. Our method showed increase in the intersection over union (IoU) and Dice coefficients by 7.9 and 6.7 percentage points, respectively, as compared to the baseline model. Furthermore, we performed comparisons with alternative image segmentation methodologies, which confirmed the sophistication and efficacy of our proposed model.

Conclusions: Our proposed multiple myeloma segmentation net (MMNet), can effectively extract multiscale features from images and enhance the correlation between channel and spatial information.

Furthermore, a systematic evaluation of the proposed network architecture was conducted on a self-constructed, limited dataset. This endeavor holds promise for offering valuable insights into the development of algorithms for future clinical applications.

Keywords: Multiple myeloma segmentation (MM segmentation); multiscale features; atrous convolution; attention mechanism; magnetic resonance imaging (MRI)

Submitted Apr 02, 2024. Accepted for publication Jul 26, 2024. Published online Sep 24, 2024.
doi: 10.21037/qims-24-683
View this article at: https://dx.doi.org/10.21037/qims-24-683

Introduction

Multiple myeloma (MM) is a malignant disorder marked by the abnormal proliferation of cloned plasma cells. As the global population ages, the incidence of MM continues to rise. This condition can manifest as persistent unexplained bone pain, and it has the potential to lead to renal failure and recurrent bacterial infections, particularly pneumococcal pneumonia. These manifestations significantly impact the health and overall quality of life of afflicted individuals (1). Currently, the diagnosis of MM primarily relies on blood tests, bone marrow puncture and biopsy (2), and imaging examinations. However, both blood examination and bone marrow puncture are invasive methods and unsuitable for large-scale screening and systemic detection. Furthermore, bone marrow puncture only provides information about a single site lesion. In contrast, imaging detection methods offer a noninvasive approach for quantifying overall tumor information and observing multiple lesions and bone tissue simultaneously. This has aided in overcoming the limitations of puncture procedures, which can only capture partial tumor information. Therefore, accurately segmenting the areas of MM lesions from images is crucial for precise diagnosis and treatment planning. This process not only benefits clinicians in better understanding the extent and severity of the lesions but also provides crucial evidence for subsequent treatment decisions. Moreover, previous studies have shown that precise measurements of the volume and various morphological features of MM lesions can help predict disease progression (3), forecast the development of local bone destruction (4), and contribute to the construction of advanced tumor growth models (5). It has been established that imaging indicators can be an effective supplement to the prognosis of MM (6-8). In 2019, Rasche *et al.* (9) reported that high-risk genetic progression markers (such as RAS mutation) were mainly found in large lesions with a diameter >2.5 cm. Accurate lesion segmentation can facilitate the advancement of similar research endeavors.

The imaging methods recommended by the International Myeloma Working Group include bone marrow magnetic resonance imaging (MRI) (10), computed tomography (CT) (11), and positron emission tomography computed tomography (PET-CT) (12). Among these, MRI is the preferred diagnostic tool for initial assessments, as MRI's outstanding soft-tissue contrast enables the direct imaging of the bone marrow with high sensitivity. This not only facilitates the detection of bone destruction but also allows for the assessment of tumor burden. *Figure 1* illustrates the MRI imaging of a patient's spinal region and the MM lesions. This comprehensive visualization aids in understanding the extent of the disease and contributes to better-informed diagnostic and treatment decisions. The manifestations of MM lesions can generally be classified into three categories: localized, diffuse, and salt-and-pepper types. Different types of lesions often imply varying sizes and complex shapes of the affected areas, posing challenges for the accurate segmentation of MM. In recent years, several researchers have endeavored to segment MM plasma cells from microscopic images. Qiu *et al.* (13) proposed a deep learning framework called semantic cascade mask region-based convolutional neural network (R-CNN) for detecting and segmenting myeloma cells. This framework integrates with the proposed feature selection pyramid network, uses a mask aggregation module to refine high-certainty instance masks, merges them into a single segmentation map, and employs the results from additional semantic segmentation branches to enhance segmentation performance. Paing *et al.* (14) introduced a method for computer-assisted detection and segmentation of myeloma cells from microscopic images of bone marrow aspirates, employing different mask R-CNN models for instance segmentation on different images and applying

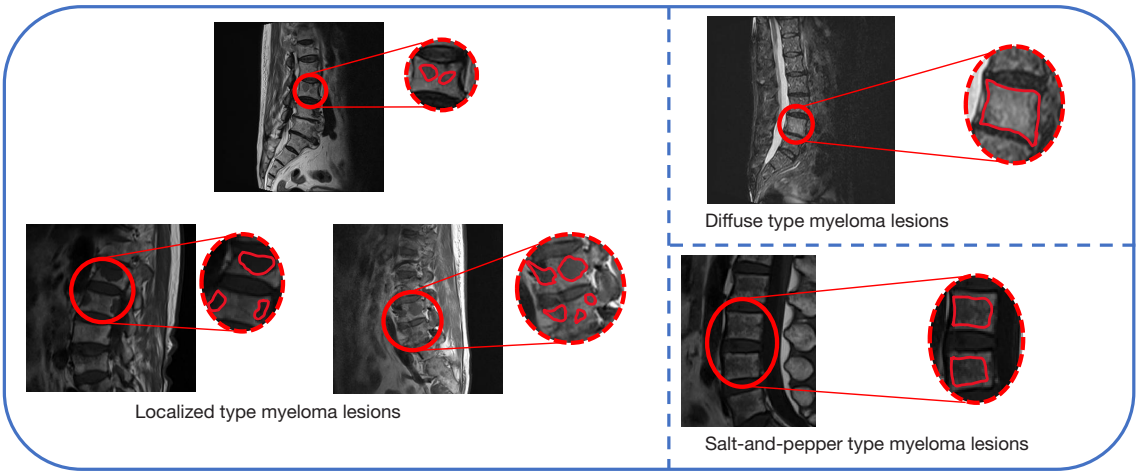

Figure 1 Example of multiple myeloma lesion area in patients.

deep augmentation to enhance model performance. Additionally, some researchers have used medical imaging techniques for corresponding analyses of MM. Algorithms have been developed for the automatic segmentation of disseminated bone marrow and subsequent radiomic analysis of 30 different bone marrow spatial (BMS) image sets from MRI, achieving automatic and comprehensive characterization of bone marrow. Typical bone marrow patterns in MM have been shown to correlate with radiomic features of corresponding BMS (15). Researchers have also automatically segmented pelvic bone marrow from T1-weighted whole-body MRI. In this approach, imageomic features are extracted from these segmentations, and a random forest model is trained to predict the presence of plasma cell infiltration (PCI) and cytogenetic abnormalities (16). Wennmann *et al.* (17) trained a no-new-Net (nnU-Net) to automatically segment pelvic bone marrow from whole-body apparent diffusion coefficient (ADC) maps in multicenter datasets, achieving a quality comparable to that of manual segmentation. Automatically extracted ADC values were significantly correlated with bone marrow PCI and thus demonstrated potential value for automatic staging, risk stratification, or treatment response assessment. A whole-body imaging approach for segmenting lesion areas in patients MM using MRI can assist physicians in better determining lesion locations and conditions. Our proposed model has been correspondingly improved in terms of feature extraction, upsampling, and feature fusion, enabling better capture of global image information and applicability for segmenting complex and variable MM lesions.

Traditional medical image segmentation methods have conventionally relied on manually designed features and rules, often employing techniques such as thresholding segmentation (18), region growing (19), and edge detection (20). However, these approaches encounter challenges when faced with complex structures and significant grayscale variations in medical images. Deep learning, particularly CNNs, has revolutionized medical image segmentation by enabling models to learn features and patterns directly from data, yielding remarkable outcomes. CNNs are particularly adept at capturing hierarchical representations of images, starting from low-level features such as edges and textures to high-level semantic features that are crucial for accurate segmentation. This hierarchical feature extraction capability is crucial in medical imaging in which subtle differences in texture or shape hold critical diagnostic information. CNN architectures, such as U-Net (21), DeepLab (22) series, and V-Net (23), have emerged as prominent research focal points in the realm of medical image segmentation. The U-Net structure, leveraging both encoder and decoder components, facilitates the simultaneous capture of global and local features, rendering it adaptable for a diversity of medical imaging tasks.

To enhance U-Net's performance, numerous improvements and variants have been proposed. Residual connections (24), as seen in the residual U-Net (ResU-Net) (25), mitigate the gradient disappearance problem, accelerate model training, and improve overall performance. U-Net++ (26), by incorporating more intricate connection modes, further enhances information transmission and

feature extraction, leading to improved performance. Attention U-Net (27) introduces an attention mechanism, directing the network to focus more on crucial areas, thereby enhancing image segmentation accuracy. Chen *et al.* (28). introduced the Transformer (29) into medical image segmentation, presenting Transformer U-Net (Trans-UNet), a model capable of precisely locating pixel information and overcoming the inherent locality of traditional convolution operations. Despite the progress achieved by these methods in enhancing segmentation network performance, the need for a series of upsampling and downsampling operations to enlarge the receptive field for pixel-level prediction remains. However, these operations inevitably result in information loss and underutilization. Furthermore, obtaining a sufficient amount of global information remains challenging, limiting the segmentation network's capacity to achieve higher accuracy.

Multiscale features play a pivotal role in the domains of computer vision and image processing. As image analysis tasks grow in complexity, it has been increasingly recognized that single-scale feature extraction may be inadequate for responding effectively to variations in diverse scenes and objects. Consequently, multiscale feature extraction has attracted considerable attention, as it may represent a potent tool for addressing challenges in image processing tasks. In real-world applications, objects may manifest with varying sizes, shapes, and orientations, posing a challenge for traditional single-scale feature extraction methods in capturing this diversity. Modern methods for extracting multiscale features include a range of structures and techniques, including pyramid structures (30), attention mechanisms, and convolution kernels with distinct receptive fields. Scholars have made notable contributions to the field of multiscale feature extraction. For instance, Xia *et al.* (31) introduced the multi-scale context-attention network (MC-Net), a multiscale context attention network that integrates multiscale and context attention modules. This network demonstrates proficiency in capturing both local and global semantic information surrounding the target. Additionally, Yang *et al.* (32) proposed a multiscale attention network designed specifically for the automatic segmentation of glomerular electron dense deposits in electron microscope images. This method employs fully convolutional networks, incorporating multiscale skip connections and attention mechanisms. The multiscale skip connection merges feature maps of varying scales, while the attention mechanism concentrates on prominent structures, resulting in discriminative feature representations. Yan *et al.* (33) introduced a feature variation (FV) module adept at adaptively adjusting the global attributes of features to enhance feature representation ability, demonstrating effectiveness in segmenting coronavirus disease 2019 (COVID-19) infection. The experimental results included Dice similarity coefficients (DSCs) of 0.987 for lung segmentation and 0.726 for COVID-19 segmentation. Li *et al.* (34) developed the multi-scale fusion U-Net (MF U-Net), a multiscale fusion network designed for breast cancer lesion segmentation. The model incorporates a wavelet fusion module (WFM) to segment irregular and blurred breast lesions, a multiscale dilated convolution module (MDCM) to manage segmentation difficulties caused by large-scale changes in breast lesions, and focal DSC loss.

The contributions of this article can be summarized as follows:

(I) A new encoder-decoder architecture for MM segmentation is proposed, which enhances the network's ability to extract multiscale features, enriches the diversity of features, and effectively improves the accuracy of the model in lesion segmentation.

(II) We propose dilated dense connected net (DCNet), an innovative feature extraction module designed to replace the conventional convolutional block, thereby extending both the depth and width of the network. This modification aims to bolster the sensory field of the model, facilitating multiscale feature extraction.

We introduce a dynamic feature fusion (DyCat) module that amalgamates the strengths of dynamic convolution (35) and the attentional feature fusion (AFF) module (36). This module possesses the capability to flexibly reprocess feature maps in both the encoder and decoder stages, enabling adaptive adjustments to the feature fusion process and consequently enhancing the overall fusion effect.

We propose the CBAM-atrous spatial pyramid pooling (CASPP) module, which integrates dilated convolution (37) with both channel and spatial attention mechanisms (38), specifically designed for the decoding stage. This module effectively prioritizes channel and spatial information while extracting multiscale features. The resulting output from the CASPP module is then concatenated with the final feature map, leading to a significant enhancement in both the performance and generalization ability of the model.

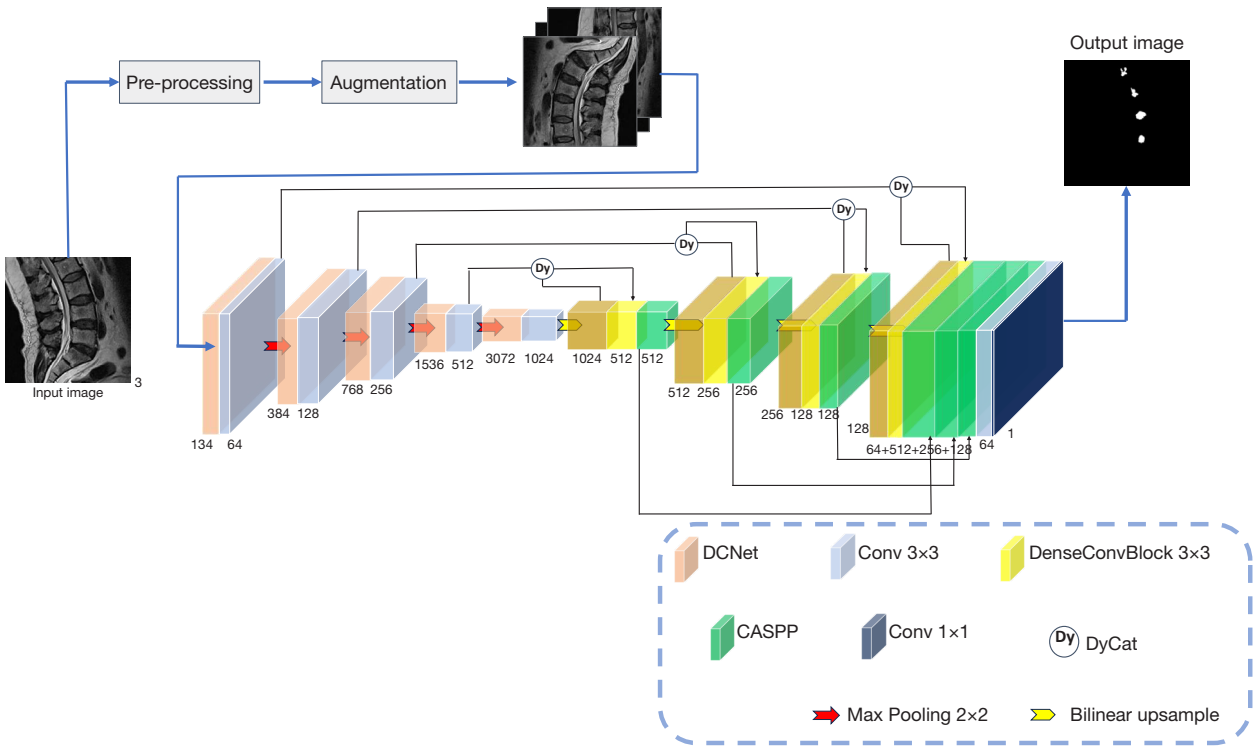

Figure 2 An overview of our proposed MMNet model. The cube represents the feature map. The proposed DCNet, DyCat, and CASPP modules are integrated into the framework in order. DCNet, dilated dense connected net; Conv, convolution; CASPP, CBAM-atrous spatial pyramid pooling; DyCat, dynamic feature fusion; MMNet, multiple myeloma segmentation net.

Methods

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of The First Affiliated Hospital of Chongqing Medical University (No. K2023-314). The requirement for individual consent was waived due to the retrospective nature of the analysis.

Using the U-Net model as a foundation, this paper introduces a novel MM segmentation model incorporating the principles of the dense convolution module (39), dynamic convolution, atrous convolution, atrous spatial convolution pooling pyramid (ASPP), and attention mechanism. The comprehensive segmentation process and network architecture are depicted in (*Figure 2*). The multiple myeloma segmentation net (MMNet) is an end-to-end implementation designed to leverage both global and local information, enhancing the semantic information's comprehensiveness and diversity. As with the original U-Net, MMNet consists of three fundamental components: encoder, decoder, and bottleneck layer. During the encoding stage, our proposed multiscale feature extraction module is sequentially employed. In the decoding stage, the proposed CASPP module is used to preserve multiscale features, subsequently fusing them with the final generated feature map. Within the bottleneck layer of feature fusion, our proposed DyCat module is employed to achieve more effective feature extraction and more comprehensive fusion.

DCNet

In the MR images of most patients myeloma, MM lesions of various scales are typically present, often characterized by their diminutive size. This poses a significant challenge for basic image segmentation models in accurately delineating all the lesions. Consequently, the extraction of features from lesion regions at different scales is a pivotal task for MM segmentation algorithms. Given that receptive fields of varying sizes correspond to distinct capabilities in capturing long-range dependencies, our designed DCNet feature extraction module primarily comprises two DC

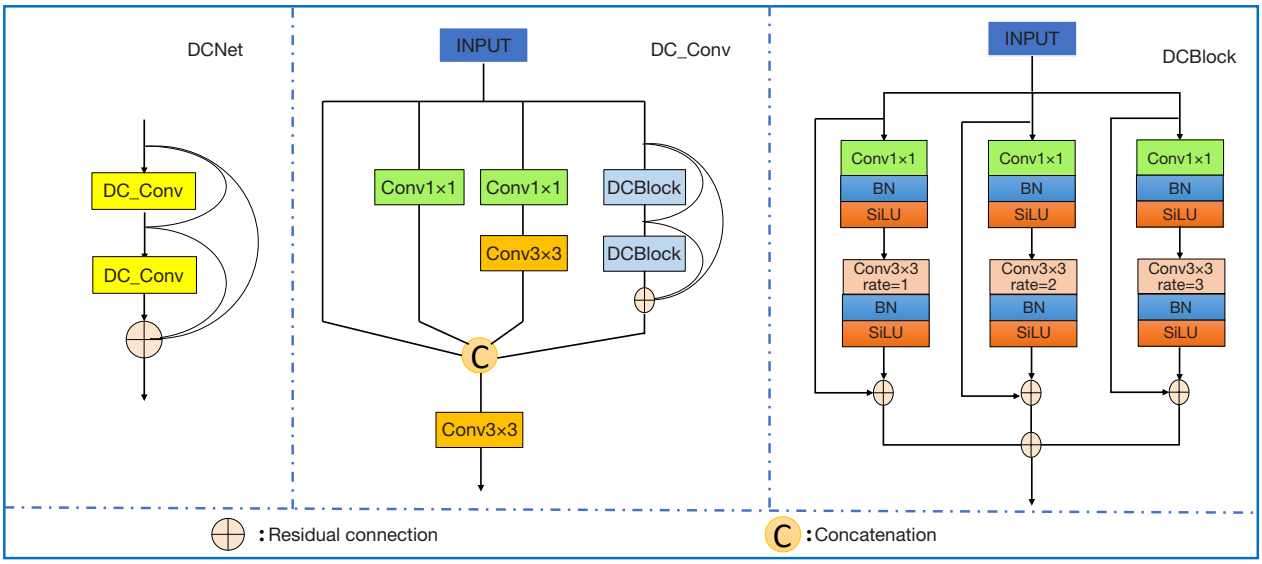

Figure 3 DCNet is constructed from the fundamental component DC_Conv, which in turn encompasses the critical subcomponent, DCBlock. DCNet, dilated dense connected net; DC_Conv, DC convolution; Conv, convolution; BN, batch normalization; SiLU, sigmoid linear unit.

convolution (DC_Conv) modules in a densely connected configuration. The structure of the DC_Conv module is depicted in *Figure 3*. This module harnesses the advantages of multiscale receptive fields and can be partitioned into four branches: the first branch, featuring two DC blocks (DCBlocks), is densely connected; a 1×1 convolution forms another branch, introducing additional nonlinearity to the feature map and enhancing generalization ability; a third branch incorporates a 1×1 convolution followed by a 3×3 convolution in series, facilitating the extraction of deeper features from the feature map and effectively amalgamating features at different levels; the fourth branch is a residual connection employed to mitigate issues of gradient explosion and disappearance during training. Finally, the output feature maps of each branch undergo concatenation processing, and the features are further extracted through a 3×3 convolution, ultimately outputting the specified dimension.

Based on the receptive field enhancement (RFE) module in YOLO-FaceV2 (40), the envisioned DCBlock structure comprises three branches. Following a 1×1 convolution operation, each branch employs dilated convolutions with expansion rates of 1, 2, and 3, respectively, incorporating residual connections to fortify the fusion of multiscale features and augment the receptive field. Ultimately, a weighting mechanism is applied to each branch's feature, ensuring a balanced representation across the diverse branches.

For clarity, we establish a dense connection between two DC_Conv blocks to form the feature extraction module DCNet within MMNet. Dense connections can effectively enhance the capabilities of DCNet. This means that there are direct connections between each DC_Conv block, allowing features to be transmitted more efficiently from one block to the next without the loss of information or features. Consequently, this enhances the dimensionality of the feature maps and the depth of the network while avoiding overfitting. During the feature extraction stage, we sequentially employ the proposed feature extraction module, augmenting the feature map dimension and receptive field. This approach enables the fusion of multiscale features, thereby improving the accuracy of myeloma lesion area recognition.

The architectural depiction of the model during the encoding stage, along with the dimension information regarding the input and output feature maps of its internal modules, is presented in (*Table 1*).

DyCat module

In the conventional U-Net architecture, the skip connection straightforwardly concatenates the encoder and decoder feature maps, potentially resulting in constrained information transmission. This paper introduces a novel

Table 1 The architecture and dimensions of the encoding phase

Layers	Parameter	Input dimension	Output dimension
DCNet 1	Tate (1, 2, 3)	320×320×3	320×320×134
Conv 1	3×3	320×320×134	320×320×64
Max pooling 1	2×2, stride2	320×320×64	160×160×64
DCNet 2	Rate (1, 2, 3)	160×160×64	160×160×384
Conv 2	3×3	160×160×384	160×160×128
Max pooling 2	2×2, stride2	160×160×128	80×80×128
DCNet 3	Rate (1, 2, 3)	80×80×128	80×80×768
Conv 3	3×3	80×80×768	80×80×256
Max pooling 3	2×2, stride2	80×80×256	40×40×256
DCNet 4	Rate (1, 2, 3)	40×40×256	40×40×1,536
Conv 4	3×3	40×40×1,536	40×40×512
Max pooling 4	2×2, stride2	40×40×512	20×20×512
DCNet 5	Rate (1, 2, 3)	20×20×512	20×20×3,072
Conv 5	3×3	20×20×3,072	20×20×1,024

Layers, the modules used in the encoding stage; Parameters, the internal parameters of each module; Input dimension, the size and dimension of the input feature map of each module; Output dimension, the size and dimension of the output feature map of each module; DCNet, dilated dense connected net; Conv, convolution.

dynamic feature connection method, DyCat, which leverages dynamic convolution and the AFF feature fusion module to dynamically adapt convolution parameters and the feature fusion process. This dynamic approach ensures a more comprehensive and effective fusion of features between the encoder and decoder. Not only does it enhance the model's performance, but it also exhibits remarkable efficacy in complex lesion scenarios.

The DyCat mechanism is illustrated in *Figure 4*, showing the output feature map of the encoder stage passing through two branches. One branch employs a 1×1 convolution to extract more abstract feature information and reduce computational complexity, while the other branch employs three 3×3 dynamic convolutions connected by a residual connection. Dynamic convolution enables the dynamic adjustment of convolution kernel weights based on input data characteristics. This adaptability to feature output of the encoder stage enhances flexibility in feature processing, allowing the model to better capture local features within the image. Similarly, for the feature map of the decoder stage, three 3×3 dynamic convolutions connected by residuals are employed to adaptively extract features. Subsequently, the AFF attention feature fusion mechanism is employed to fuse the feature maps of the two components. This fusion process, distinct from simple summation or concatenation, dynamically combines semantic and scale-inconsistent features more effectively.

The AFF module was introduced by Dai *et al.* in 2021 (36), and it can effectively integrate the features of different layers or branches and improve the performance of the model. Its principal element is the multiscale channel attention module (MS-CAM). Channel attention is derived through two branches with distinct scales, addressing the challenge in the effective fusion of features at different scales. MS-CAM employs point-by-point convolution for channel scale processing, foregoing the use of convolution kernels with varying sizes. The calculation formula for channel attention in local features is expressed as follows:

$$L(X) = B\left(PWConv_2\left(\delta\left(B\left(PWConv_1(X)\right)\right)\right)\right) \quad [1]$$

First, the number of channels of input features is reduced to half of the original by $PWConv_1 1*1$ point-by-point convolution. Subsequently, the batch normalization (BatchNorm) layer is applied for normalization (B), and the rectified linear unit (ReLU) activation function is used

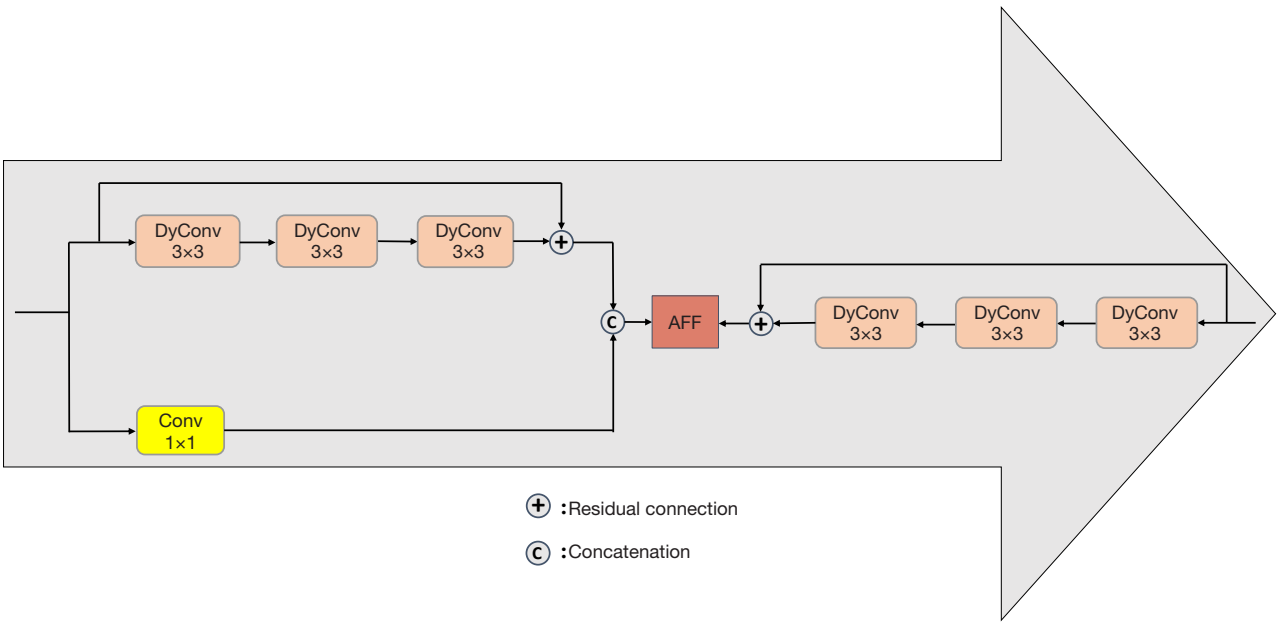

Figure 4 Illustration of the DyCat module. DyCat is mainly composed of dynamic convolution, residual connection, and AFF feature fusion modules. DyConv, dynamic convolution; AFF, attentional feature fusion module.

for nonlinear processing (δ). The number of channels is then restored to the original number of input channels via the convolution of $PWConv_1 1*1$. The computed weight value is used to execute the attention operation on the input feature, yielding the output. The specific formula is provided below:

$$X' = X \oplus M(X) = X \oplus \sigma\left(L(X) \oplus g(X)\right) \quad [2]$$

The symbol $\oplus$ represents the multiplication of the corresponding elements of the two feature maps, and the calculation formula of the global feature channel attention adopts $g(X)$. The difference between it and $L(X)$ is that the input X is first subjected to a global average pooling operation. Given two features X and Y for feature fusion, the calculation method of AFF is as follows:

$$Z = 2 \cdot X \cdot M\left(X + Y\right) + 2 \cdot Y \cdot \left(1 - M\left(X + Y\right)\right) \quad [3]$$

where Z is the output feature after feature fusion, M is the MS-CAM, and $+$ is the initial feature integration.

CASPP module

To enhance the model's feature capture during upsampling, promote channel correlation, achieve precise spatial positioning, and ensure that the final output feature map encompasses comprehensive and effective multiscale features, we introduce the CASPP module. The CASPP module's design stems from a nuanced understanding of the challenges neural networks face in handling multiscale and channel correlations. Although the ASPP module effectively integrates multiscale information through atrous convolution, it fails to emphasize channel and spatial positioning correlations. To address this limitation, we introduce channel and spatial attention modules following the ASPP, enabling adaptive focus on channel and spatial information. This refinement enhances the model's ability to capture crucial features and accurately identify lesion areas.

Specifically, as depicted in *Figure 5*, the feature map undergoes transformation into a specified dimension through a 3×3 convolution after the DyCat module. Subsequently, it enters an ASPP module that includes expansion rates of 1, 6, 12, and 18 for further extracting the multiscale features. These multiscale features are subsequently routed through the channel and spatial attention modules sequentially. As distinct channels may offer information on various facets of the lesion and given the dynamic nature of the lesion's location, shape, and size within the image for MM segmentation,

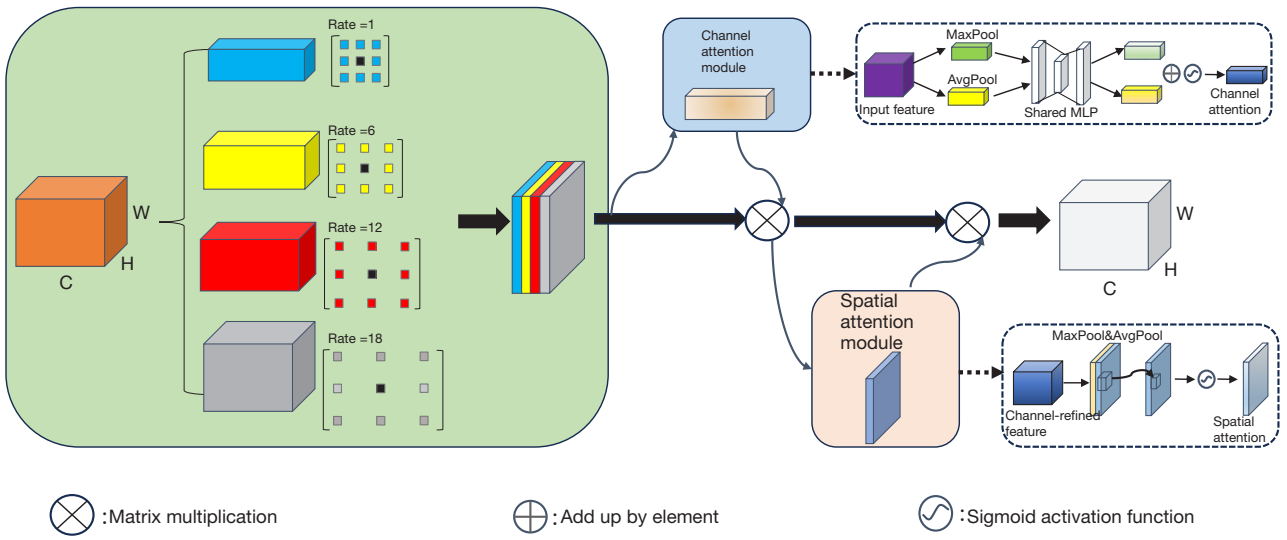

Figure 5 CASPP structure diagram. The feature map is applied to the channel and spatial attention mechanism after four types of dilated convolutions with different expansion rates are implemented. W, width; H, height; C, channel; MLP, multilayer perceptron; CASPP, CBAM-atrous spatial pyramid pooling.

the integration of spatial attention renders the network more adaptable to diverse spatial layouts, consequently enhancing segmentation accuracy. This design excels in its comprehensive attention to multiscale, channel, and spatial information, thereby augmenting the network's proficiency in comprehending and interpreting intricate pathological images.

A summary of the model's architecture in the whole decoding stage and the dimension information of the input and output of the feature map of its internal module are provided in *Table 2*.

Binary cross-entropy and Dice loss

The loss function employed in this experiment is the binary cross-entropy and Dice loss (BCEDiceLoss), a widely used metric in image segmentation tasks. BCE loss is a prevalent metric in binary classification problems, encompassing binary image segmentation. It measures the discrepancy between the predicted probability and the actual binary label. The formula is provided below:

$$BCELoss = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log\left(\hat{y}_i\right) + (1-y_i)\cdot \log\left(1-\hat{y}_i\right)\right] \quad [4]$$

where N is the number of samples, y_i is the true label of the i th sample, and $\hat{y}_i$ is the prediction probability of the i th sample.

Dice loss is commonly employed in image segmentation tasks to assess the similarity between the predicted region and the ground truth region. The formula is expressed as follows:

$$DiceLoss = 1 - \frac{2\times \text{Intersection}}{Union + \text{Intersection}} \quad [5]$$

where *Intersection* is the intersection of the predicted region and the real region, and *Union* is the union of the predicted region and the real region. Consequently, the formula of BCEDiceLoss can be expressed as follows:

$$BCEDiceLoss = \lambda \cdot BCELoss + \mu \cdot DiceLoss \quad [6]$$

where λ and μ are parameters that weigh two loss functions, usually between 0 and 1, and BCEDiceLoss combines the two aspects of BCELoss and DiceLoss and comprehensively considers the accuracy of the prediction probability and the similarity between the predicted region and the real region. This enables the model to learn the characteristics of the MM lesion area more comprehensively in the optimization process, which helps to improve the segmentation performance of the model in complex scenes. In addition, in this study, the weights of BCELoss and DiceLoss were set to 0.5 and 1, respectively. Since the boundary of the lesion in the MM image is blurred, giving more weight to the Dice loss helps to improve the sensitivity of the model to boundary pixels.

Table 2 The architecture and dimensions of the decoding phase

Layers	Parameter	Input dimension	Output dimension
DyCat 1	3×3 & 1×1	40×40×512 & 40×40×512	40×40×1,024
DenseConvBlock 1	3×3	40×40×1024	40×40×512
CASPP 1	Rate (1, 6, 12, 18)	40×40×512	40×40×512
Upsample 1	Scale factor 2	40×40×512	80×80×512
DyCat 2	3×3 & 1×1	80×80×256 & 80×80×512	80×80×768
DenseConvBlock 2	3×3	80×80×768	80×80×256
CASPP 2	Rate (1, 6, 12, 18)	80×80×256	80×80×256
Upsample 2	Scale factor 2	80×80×256	160×160×256
DyCat 3	3×3 and 1×1	160×160×128 & 160×160×256	160×160×384
DenseConvBlock 3	3×3	160×160×384	160×160×128
CASPP 3	Rate (1, 6, 12, 18)	160×160×128	160×160×128
Upsample 3	Scale factor 2	160×160×128	320×320×128
DyCat 4	3×3 and 1×1	320×320×64 & 320×320×128	320×320×192
DenseConvBlock 4	3×3	320×320×192	320×320×64
Conv 6	3×3	320×320×960	320×320×64
Conv 7	1×1	320×320×64	320×320×1

Layers, the modules used in the decoding stage; Parameters, the internal parameters of each module; Input dimension, the size and dimension of the input feature map of each module; Output dimension, the size and dimension of the output feature map of each module; DyCat, dynamic feature fusion; CASPP, CBAM-atrous spatial pyramid pooling; Conv, convolution.

Datasets

The experimental data were collected by The First Affiliated Hospital of Chongqing Medical University. The region of interest was manually delineated by a physician assistant with 5 years of experience and a radiologist and verified by a musculoskeletal radiologist with 14 years of experience. MRI was performed using 1.5- or 3.0-T MRI device (MAGNETOM Essenza or Skyra, Siemens Healthineers, Erlangen, Germany). The imaging protocol was as follows: a sagittal turbo spin echo (TSE) T1-weighted sequence (repetition time/echo time =490/10 ms, field of view =32 cm, matrix size =320×320, slice thickness =3 mm), a sagittal TSE T2-weighted sequence (repetition time/echo time =2,900/102 ms, field of view =32 cm, matrix size =400×400, slice thickness =3 mm), a sagittal T2-weighted Dixon sequence (repetition time/echo time =3,000/82 ms, field of view =32 cm, matrix size =300×300, slice thickness =3 mm), and an axial TSE T2-weighted sequence (repetition time/echo time =3,740/108 ms, field of view =18 cm, matrix size =300×206, slice thickness =4 mm). The image data consisted of three sequences: T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), and short tau inversion recovery (STIR). Typically, areas of bone destruction or marrow infiltration appear as low signal on T1WI and high signal on T2WI. On STIR sequences, due to suppression of bone marrow fat signal, lesions exhibit higher signal intensity compared to that on T2WI; thus, by combining images from different sequences, we could better leverage the data, improving the model's training effectiveness, generalization ability, and robustness, thereby enhancing its capability to handle the diverse data encountered in real-world applications. These data encompassed MRI images of 45 patients with MM, with most patients undergoing three different sequences, while a few underwent one or two sequences. This resulted in 753 original MR images and their corresponding ground-truth labels. Given the typically large size of the original images, for optimal utilization of computing resources and accelerated convergence, we resized all original MR images to 320×320 after converting them to standard image formats. To mitigate the impact of limited data on model performance and enhance robustness,

we applied data augmentation techniques to the original dataset. These techniques included random horizontal and vertical flips, image rotation, and scaling, which expanded the dataset to 3,000 images. In this augmented dataset, 80% of the images were used for training, and 20% were used for validation and testing. The data processing flowchart is shown in *Figure 6*.

We thoroughly analyzed the heterogeneity of MM lesion images in both the training and testing datasets. The lesions exhibited significant variations in both size and number. The statistical analysis results are presented in *Table 3*. We calculated the statistical features for lesion size and quantity in each of the mask images, including mean, standard deviation, and the coefficient of variation. The lesion size was measured in pixels. Specifically, the lesion size was determined by calculating the number of pixels within the lesion area on the mask image.

In the training set, the coefficients of variation for lesion size and quantity were 1.46 and 1.26, respectively. In the test set, these coefficients were 1.35 for lesion size and 0.77 for lesion quantity. Generally, a coefficient of variation exceeding 1 indicates substantial variations in lesion size and number among different patients in both training and test sets. This heterogeneity poses significant challenges to lesion segmentation algorithms. The distribution of lesions is illustrated in (*Figure 7*), which includes box plots and histograms of lesion size and quantity in both the training and test sets. Box plots depict the distribution of lesion size and quantity data across patients, offering insights into their range, central tendency, and outliers. Histograms display the frequency distribution of lesion size and quantity data, partitioning them into intervals; the count of lesions and patients within each interval are thus presented in bar chart format.

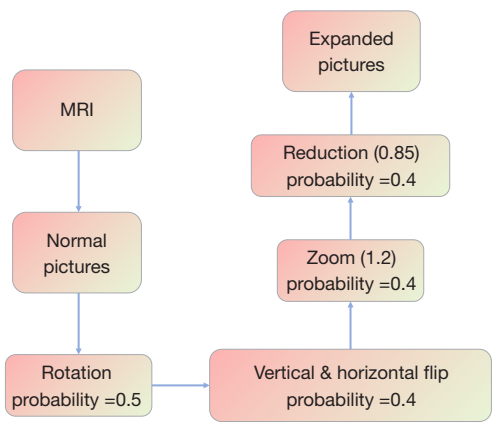

Figure 6 Data expansion process. MRI, magnetic resonance imaging.

Implementation procedure

The hardware and software configurations, along with the hyperparameter settings employed in the experiments, are summarized in *Table 4*. Each experiment adopted uniform hyperparameters for training, validation, and testing. The model achieving the highest intersection over union (IoU) value on the validation set was kept as the final model. Notably, to maintain the integrity of the experimental outcomes, no pretrained weight parameters were used before or after model optimization.

Statistical analysis

In model performance evaluation, rigorous statistical analysis is crucial for ensuring the reliability and validity of results. In our study, we employed independent samples *t*-tests to compare P values across various metrics from different experiments, including ablation studies, model comparison experiments, and module comparison experiments. Subsequently, we applied the Benjamini-Hochberg procedure to correct each P value for multiple comparisons to control the false-discovery rate. The P value is a statistical measure used to determine the significance of results, representing the probability of observing the data

Table 3 Analysis of heterogeneity in size and shape of multiple myeloma lesions in our self-constructed dataset

Index	Training set		Test set	
	Lesion size (pixels)	Lesion count	Lesion size (pixels)	Lesion count
Mean value	193.58	2.60	222.60	1.88
Standard deviation	282.49	3.29	300.19	1.45
Coefficient of variation	1.46	1.26	1.35	0.77

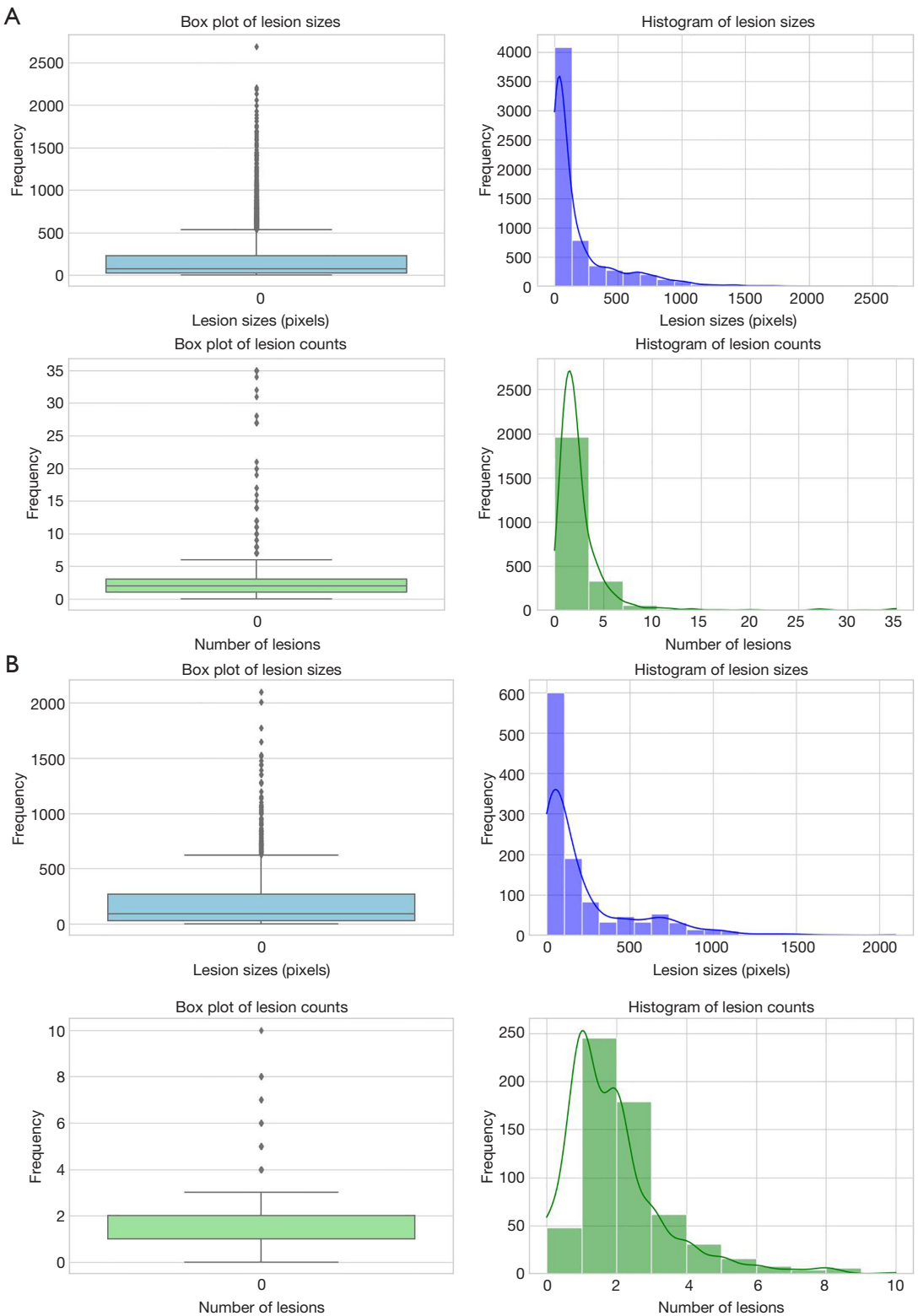

Figure 7 Box plots and histograms of lesion size and quantity in the training and test sets. (A) Box plots and histograms depicting lesion sizes and counts in the training set. (B) Box plots and histograms for lesion sizes and counts in the test set. The box plots illustrate the distribution of lesion sizes and counts, while the histograms display the frequency distribution of lesion sizes and counts.

Table 4 The hardware, software configuration, and super parameter setting used in the experiment

Name	Configuration information
Development language	Python 3.8.10
Framework	PyTorch 1.11.0, cuda 11.3, cudnn 8200
CPU	Intel (R) Xeon (R) Platinum 8255C CPU @ 2.50 GHz
GPU	V100-SXM2-32 GB GPU
Loss function	BCEDiceLoss
Optimizer	SGD
Learning rate	0.001
Number of epochs	250

CPU, central processing unit; GPU, graphics processing unit; SGD, stochastic gradient descent.

or more extreme data under the null hypothesis. In our study, lower P values indicated that the observed differences between models were unlikely to be due to chance, suggesting statistical significance. The independent samples *t*-test was used to compare the means of two independent groups to assess whether there was statistical evidence of a significant difference between the means of the populations in question. The formula for the *t*-test statistic is as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad [7]$$

Where $\bar{X}_1$ and $\bar{X}_2$ are the sample means, S_1^2 and S_2^2 are the sample variances, and n_1 and n_2 are the sample sizes. The Benjamini-Hochberg procedure was employed to control the false-discovery rate in multiple testing, allowing us to make more confident assertions about the significance of our research findings and thereby enhancing the credibility of our conclusions.

Results

In this section, the evaluation indicators used in the experiment are described. Following this is a detailed analysis of the experimental results of MMNet on this dataset and the complex comparison with other advanced algorithms. In addition, the series of ablation experiments carried out to verify the effectiveness of the proposed module is outlined.

Evaluation metrics

The choice of evaluation metrics directly mirrors the algorithm's quality. This study primarily employed five indicators to assess algorithmic performance, including IoU, Hausdorff distance, precision, recall, and Dice coefficient. Theoretically, a smaller Hausdorff distance index indicates superior segmentation performance, whereas values closer to 1 for the other indices signify enhanced segmentation performance. The formulae for each evaluation index are outlined as follows:

$$IoU = \frac{Area\_of\_Overlap}{Area\_of\_Union} \quad [8]$$

$$d_H(A,B) = \max\left(h(A,B), h(B,A)\right) \quad [9]$$

Here, $d_H(A,B)$ denotes the Hausdorff distance of two point sets A and B, where $h(A,B)$ denotes the nearest distance from point set A to point set B: $h(A,B) = \max_{a\in A} \min_{b\in B} \|a-b\|$. Similarly, $h(B,A)$ denotes the nearest distance from point set B to point set A.

$$Precision = \frac{TP}{TP + FP} \quad [10]$$

$$Recall = \frac{TP}{TP + FN} \quad [11]$$

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \quad [12]$$

Ablation experiments for evaluation module performance

For a quantitative analysis of the MMNet algorithm's detection performance on the myeloma lesion segmentation dataset, this study used UNet as the benchmark model but did not employ pretrained weight parameters for the models pre- or postenhancement. While maintaining consistent experimental configurations, the input image resolution was fixed at 320×320. Incrementally, the DCNet, DyCat, and CASPP modules were added to original UNet model. The optimal model file generated by each module served as the benchmark model, and ablation experiments were conducted on the same test set to assess each module's impact on lesion segmentation performance. The comprehensive experimental results are detailed in *Table 5*. The data marked with an asterisk in the table denote the optimal value for each evaluation index.

Table 5 Ablation studies on the different architectures

Models	Method	IoU	HD	Precision	Recall	Dice
Model 1	Baseline	0.638±0.032	51.35±5.15	0.852±0.019	0.707±0.033	0.752±0.027
Model 2	Baseline + DCNet	0.682±0.029	44.67±4.88	0.845±0.020	0.766±0.029	0.787±0.024
Model 3	Baseline + DyCat	0.664±0.031	50.14±5.61	0.854±0.019*	0.738±0.032	0.771±0.026
Model 4	Baseline + CASPP	0.685±0.031	44.38±5.00*	0.823±0.025	0.786±0.027	0.789±0.025
Model 5	Baseline + DCNet + DyCat	0.692±0.027	46.54±5.33	0.850±0.019	0.775±0.026	0.796±0.022
Model 6	Baseline + DCNet + CASPP	0.687±0.029	45.64±4.89	0.827±0.025	0.785±0.025	0.789±0.024
Model 7	Baseline + DyCat + CASPP	0.680±0.029	44.45±4.65	0.819±0.023	0.778±0.025	0.783±0.023
Model 8	Baseline + DCNet + DyCat + CASPP	0.716±0.025*	46.68±4.94	0.847±0.020	0.814±0.022*	0.819±0.019*

Each evaluation index is expressed as mean ± 95% confidence interval. *, best result in the table. IoU, intersection over union; HD, Hausdorff distance; Dice, Dice similarity coefficient; DCNet, dilated dense connected net; DyCat, dynamic feature fusion; CASPP, CBAM-atrous spatial pyramid pooling.

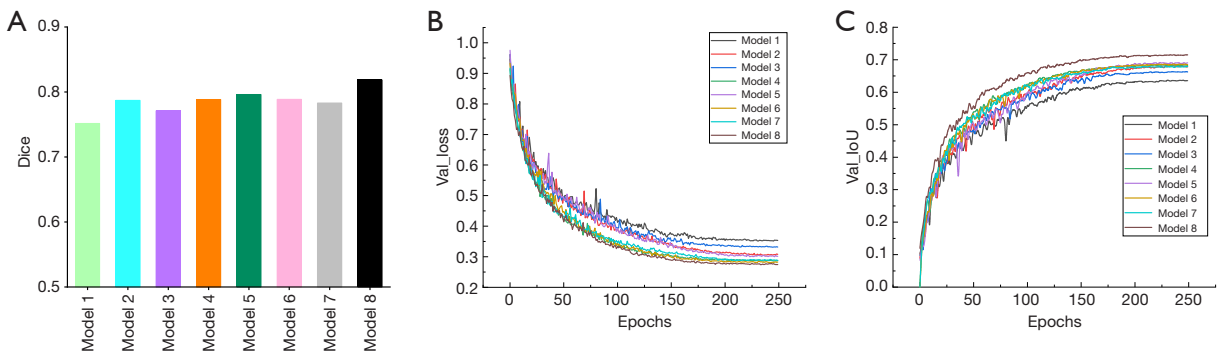

Figure 8 The fluctuations in IoU and loss for each model in the ablation experiment, along with visualizations depicting the magnitude of the Dice coefficient. (A) The Dice coefficient histogram. (B) Loss downward trend graph. (C) IoU rising trend graph. Dice, Dice similarity coefficient; IoU, intersection over union.

The experimental results indicated that compared to the baseline model, the models with each of the DCNet, DyCat, and CASPP modules individually added led to increases in IoU of 4.46, 2.63, and 4.71 percentage points and improvements in Dice coefficient of 3.56, 1.96, and 3.7 percentage points, respectively. Some model structures also exhibited enhancements in indicators such as Hausdorff distance, precision, and recall. Furthermore, the judicious combination of different modules significantly elevated image segmentation accuracy. The IoU and Dice coefficient indicators for combinations such as those of DCNet with DyCat and of DCNet with CASPP surpassed those of the individual modules. Notably, the performance of the MMNet model, formed by integrating all three modules, outperformed the combination of any two modules in terms of IoU, Dice coefficient, and recall rate. These results prove that the three proposed modules individually enhance the accuracy and efficiency of MM lesion region segmentation, substantiating the effectiveness of MMNet.

The histogram of the Dice coefficient, the loss change curve, and the IoU change curve of each model in the ablation experiment are visualized in *Figure 8A-8C*. Notably, the MMNet model demonstrated a tendency toward stability after 200 iterations, exhibiting a final loss value lower than that of any model in the ablation experiment. Additionally, its final IoU value and Dice coefficient surpassed those of all models in the ablation experiment.

As a visual depiction of the impact of various modules on

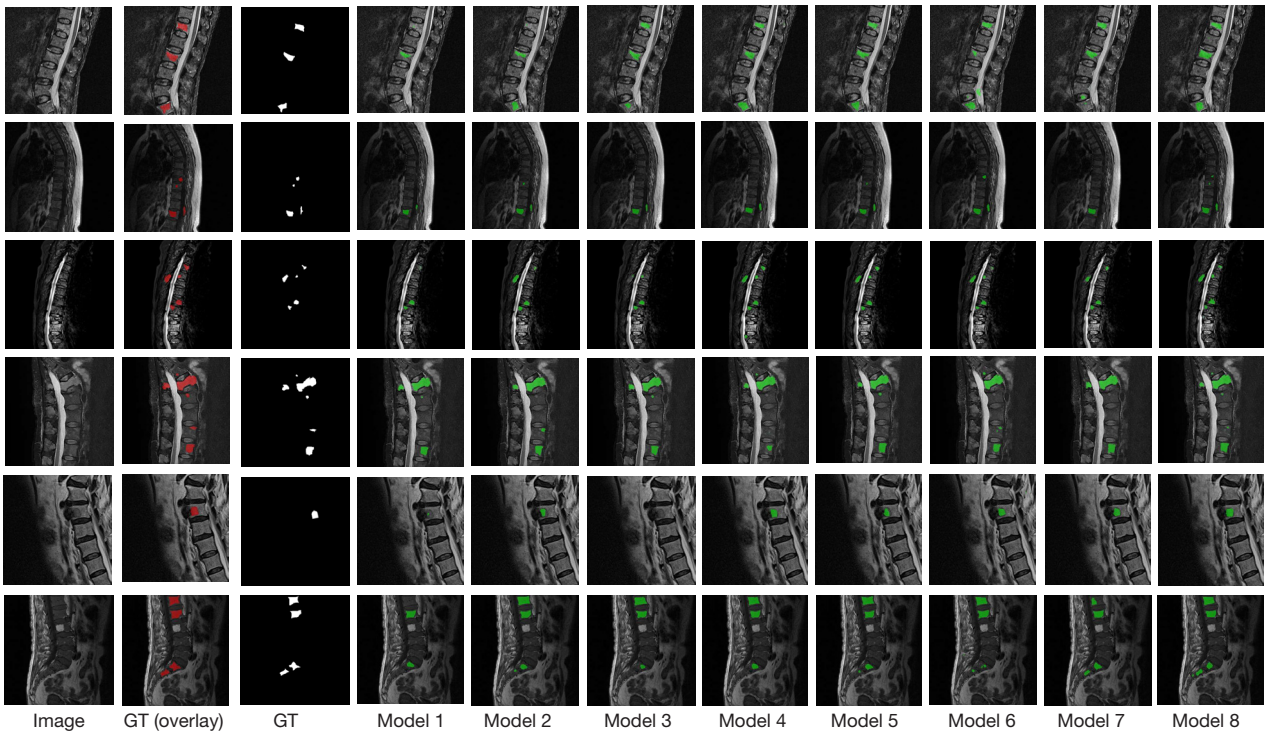

Figure 9 Visualization of segmentation outcomes for each model within the ablation study. Red represents ground truth labels, while green represents the model's predicted regions. GT, ground truth.

the segmentation performance of MM, the segmentation results of each model are displayed in *Figure 9*. The baseline model, lacking additional modules, exhibited the poorest segmentation performance. Model 2, incorporating a method for extracting multiscale features and expanding the receptive field, notably enhanced lesion area perception. The introduction of the DyCat and CASPP modules effectively mitigated the issue of insufficient segmentation, as is particularly evident in columns 5, 6, and 7 of *Figure 9*, which show that the segmentation of small lesion areas and nuanced details significantly improved with the introduction of the combination of these two modules. Ultimately, by incorporating all modules into MMNet (Model 8), the network amalgamated the advantages of the three modules, preserving richer details and semantic information, resulting in the successful identification of almost all lesion areas. Based on both evaluation indices and visual segmentation analysis, it is evident that the proposed modules indeed enhance the performance in myeloma segmentation.

Furthermore, to validate the significant differences between various model variants and MMNet in the ablation study, we employed an independent samples *t*-test to compare the P values across different metrics. Each P value was then corrected for multiple testing using the Benjamini-Hochberg procedure. As shown in *Table 6*, there were notable differences between MMNet and its variants, particularly for the original model (Model 1). After correction, the P values for the Dice coefficients of these two models remained below 0.001, indicating a significant difference compared to MMNet.

Ablation experiments evaluating the internal parameters of the module

The performance of deep learning models is intricately linked to their structural design, with the selection of hyperparameters emerging as a critical determinant of model efficacy. In our proposed DCNet and CASPP modules, we employ a fusion of dilated convolutions featuring varying expansion rates. Dilated convolutions introduce intervals within the convolution kernel, thereby expanding the receptive field without inflating parameter counts. Through adjustment of the expansion rate, we can modulate the sampling density of convolution operations

Table 6 The P value of the significance test (*t*-test) of each evaluation index between MMNet and other model variants in the ablation experiment (corrected for multiple testing using Benjamini-Hochberg adjustment)

Model	Method	P value				
		IoU	HD	Precision	Recall	Dice
Model 1	Baseline	1.04×10^{-3}	2.66×10^{-1}	7.03×10^{-1}	1.68×10^{-5}*	9.78×10^{-4}*
Model 2	Baseline + DCNet	2.12×10^{-1}	8.75×10^{-1}	9.07×10^{-1}	9.51×10^{-2}	1.95×10^{-1}
Model 3	Baseline + DyCat	3.74×10^{-2}	5.37×10^{-1}	5.95×10^{-1}	3.41×10^{-3}	3.13×10^{-2}
Model 4	Baseline + CASPP	4.63×10^{-1}	7.06×10^{-1}	4.63×10^{-1}	3.48×10^{-1}	3.48×10^{-1}
Model 5	Baseline + DCNet + DyCat	4.57×10^{-1}	9.72×10^{-1}	9.72×10^{-1}	2.15×10^{-1}	4.57×10^{-1}
Model 6	Baseline + DCNet + CASPP	2.35×10^{-1}	8.72×10^{-1}	2.35×10^{-1}	2.40×10^{-1}	2.35×10^{-1}
Model 7	Baseline + DyCat + CASPP	3.22×10^{-1}	4.44×10^{-1}	3.22×10^{-1}	3.22×10^{-1}	3.22×10^{-1}

*, P values less than 0.001. MMNet, multiple myeloma segmentation net; IoU, intersection over union; HD, Hausdorff distance; Dice, Dice similarity coefficient; DyCat, dynamic feature fusion; DCNet, dilated dense connected net; CASPP, CBAM-atrous spatial pyramid pooling.

Table 7 The influence of different dilation rates on the DCNet module

Dilation rate	Baseline + DCNet				
	IoU	HD	Precision	Recall	Dice
1, 2, 3	0.682±0.029*	44.67±4.88*	0.845±0.020	0.766±0.029*	0.787±0.024*
2, 4, 8	0.652±0.031	48.13±5.22	0.846±0.021*	0.727±0.032	0.762±0.027
6, 12, 18	0.590±0.034	52.97±4.98	0.836±0.019	0.661±0.037	0.710±0.031

Each evaluation index is expressed as mean ± 95% confidence interval. *, best result in the table. DCNet, dilated dense connected net; Dilation rate, the different combinations of different dilation rates in the module; IoU, intersection over union; HD, Hausdorff distance; Dice, Dice similarity coefficient.

Table 8 The influence of different dilation rates on the CASPP module

Dilation rate	Baseline + CASPP				
	IoU	HD	Precision	Recall	Dice
1, 2, 3, 4	0.677±0.030	44.17±4.31*	0.817±0.024	0.777±0.027	0.782±0.025
1, 6, 12, 18	0.685±0.031*	44.38±5.00	0.823±0.025*	0.786±0.027*	0.789±0.025*
6, 12, 18, 24	0.674±0.028	44.75±5.15	0.819±0.024	0.775±0.025	0.782±0.024

Each evaluation index is expressed as mean ± 95% confidence interval. *, best result in the table. CASPP, CBAM-atrous spatial pyramid pooling; Dilation rate, different combinations of different dilation rates in the module; IoU, intersection over union; HD, Hausdorff distance; Dice, Dice similarity coefficient.

on input, enabling flexible control over receptive field size. To scrutinize the impact of distinct dilation rates on CASPP and DCNet module performance, we examined the combinations of varied dilation rates within these modules. This analysis aimed to delineate performance disparities attributable to internal parameters and refine the module design. Detailed outcomes of these investigations are presented in *Tables* 7,*8*.

Ablation experiment of CASPP module

The CASPP module is composed of ASPP, channel, and spatial attention mechanisms. To ascertain whether the CASPP module could outperform its constituent

Table 9 Ablation study for the CASPP module

Modules	IoU	HD	Precision	Recall	Dice
ASPP	0.676±0.031	43.99±5.02*	0.854±0.019*	0.751±0.031	0.780±0.026
Attention mechanisms	0.651±0.032	47.66±4.91	0.794±0.027	0.756±0.029	0.760±0.027
CASPP	0.685±0.031*	44.38±5.00	0.823±0.025	0.786±0.027*	0.789±0.025*

Each evaluation index is expressed as mean ± 95% confidence interval. *, best result in the table. CASPP, CBAM-atrous spatial pyramid pooling; IoU, intersection over union; HD, Hausdorff distance; Dice, Dice similarity coefficient; ASPP, atrous spatial convolution pooling pyramid.

Table 10 Comparison of segmentation results of different network models in the multiple myeloma dataset

Method	IoU	HD	Precision	Recall	Dice
U-Net	0.638±0.032	51.35±5.15	0.852±0.019*	0.707±0.033	0.752±0.027
U-Net++	0.623±0.033	54.39±5.61	0.840±0.022	0.698±0.034	0.739±0.028
Atten-UNet	0.613±0.033	51.73±5.31	0.819±0.026	0.693±0.032	0.729±0.029
UNeXt	0.452±0.035	63.92±4.64	0.685±0.037	0.537±0.037	0.584±0.036
MA-Net	0.612±0.033	54.50±5.13	0.818±0.023	0.694±0.035	0.729±0.029
MultiRes-UNet	0.557±0.035	66.80±5.16	0.825±0.029	0.633±0.039	0.688±0.035
SCOAT-Net	0.601±0.030	51.53±4.88	0.839±0.021	0.670±0.031	0.726±0.027
MMNet (ours)	0.716±0.025*	46.68±4.94*	0.847±0.020	0.814±0.022*	0.819±0.019*

Each evaluation index is expressed as mean ± 95% confidence interval. *, best result in the table. IoU, intersection over union; HD, Hausdorff distance; Dice, Dice similarity coefficient; Atten-UNet, attention U-Net; MA-Net, multi-scale attention net; SCOAT-Net, spatial- and channel-wise coarse-to-fine attention network; MMNet, multiple myeloma segmentation net.

components individually, we conducted a series of ablation studies. The findings from these experiments are presented in *Table 9*. For a fair comparison, an equivalent number of standalone ASPP and attention modules were implemented into the model architecture at the same location as that of the CASPP module. The outcome revealed that the singular application of ASPP or attention modules yielded inferior results to that of the integrated CASPP module. This suggests that the CASPP module synergistically leverages the strengths of both ASPP and attention mechanisms, fostering a mutually reinforcing effect. This strategic integration enables the model to exhibit heightened sensitivity to both spatial and channel information while simultaneously maintaining focus on multiscale features. Such a configuration significantly enhances the overall performance of the model.

Comparative experiments

To ascertain whether the MMNet algorithm outperforms other algorithms in MM segmentation, we conducted comparative experiments, with MMNET being contrasted against other advanced image segmentation algorithms, including U-Net, U-Net++, UNeXt (41), multi-scale attention net (MA-Net) (42), Atten-UNet, MultiResUNet (43), and spatial- and channel-wise coarse-to-fine attention network (SCOAT-Net) (44). The results are presented in *Table 10*. To ensure the fairness of the comparative experiment, none of the models employed pretrained weight parameters, and the experimental results in the table are derived from the respective retraining of the code. Besides the discrepancies in the network model, all other implementation details align with those in the MMNet model.

Table 10 reveals that the proposed algorithm attained the top scores in four evaluation indicators, IoU, Hausdorff distance (HD), recall, and Dice, and ranked second in precision. These results could be attributed to the stem from U-Net's emphasis on general segmentation tasks, with a focus primarily on common rules, which overlooks the

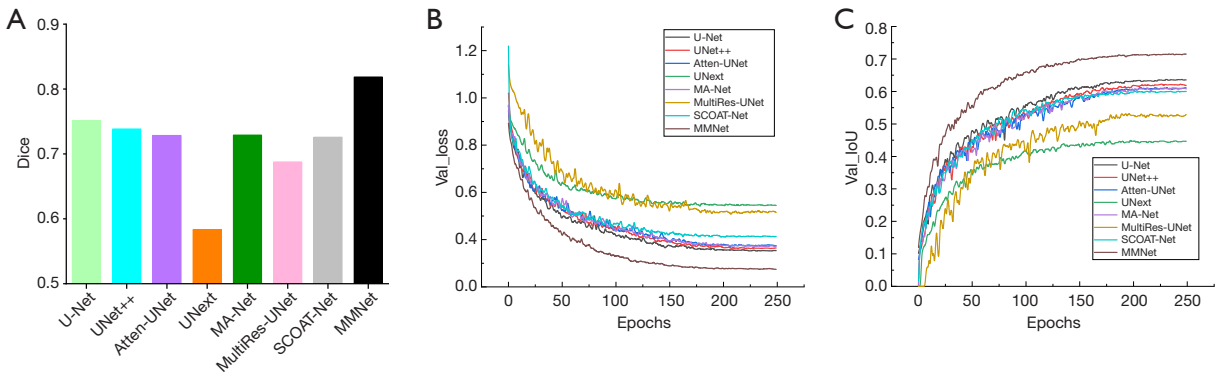

Figure 10 The fluctuations in IoU and loss for each model in the comparative experiment, along with visualizations depicting the magnitude of the Dice coefficient. (A) Dice coefficient histogram. (B) Loss downward trend graph. (C) IoU rising trend graph. Dice, Dice similarity coefficient; IoU, intersection over union; Atten-UNet, attention U-Net; MA-Net, multi-scale attention net; SCOAT-Net, spatial- and channel-wise coarse-to-fine attention network; MMNet, multiple myeloma segmentation net.

unique features present in myeloma images. Furthermore, the proposed algorithm, leveraging multiscale features, dynamic convolution, and an attention mechanism, exhibited clear performance advantages, particularly in measurements. These advantages aid in overcoming challenges posed by variations in size or shape, enabling proficiency in accurately segmenting the more nuanced features of myeloma. There are numerous evaluation indices available for image segmentation tasks, with IoU and Dice often being considered the most important. IoU represents the intersection and union ratio between the segmentation result and the ground truth label, offering a direct measure of the segmentation algorithm's effectiveness. The Dice coefficient, a statistical index used to measure the similarity of two sets, is frequently employed to quantify the similarity between two samples. In our experiment, the IoU and Dice indicators demonstrated an impressive improvement of 26.45% and 23.5%, respectively, compared to the lowest-performing model (UNeXt), underscoring its relatively superior accuracy in distinguishing between lesions and background areas. These outcomes underscore the efficacy of our proposed method in myeloma lesion segmentation.

To enable a comprehensive comparison of model disparities, we generated a loss change curve, an IoU change curve, and a histogram illustrating the Dice coefficient variations across different models during the comparative experimental validation phase. These visualizations are presented in *Figure 10*.

To visually and qualitatively illustrate the disparities in the segmentation performance of MM among different comparison models, *Figure 11* showcases the segmentation outcomes of each comparison model. In *Figure 11*, the first and second columns display the original MR image and its corresponding ground truth label, respectively; columns 3–9 demonstrate the segmentation results of U-Net, U-Net++, Atten-UNet, MultiResUNet, MA-Net, UNeXt, and SCOAT-Net. The last column illustrates the segmentation results of the MMNet model proposed in this paper. As anticipated, certain competitive networks (such as U-Net++) exhibited subpar segmentation performance in images featuring small lesions, struggling to accurately identify the lesion area. For other networks, such as MA-Net and UNeXt, the absence of multiscale features hampered segmentation when images containing lesions of varying scales were present, making them less effective than our proposed MMNet. Additionally, MMNet demonstrated a robust competitive edge in detecting images with indistinct boundaries.

The MR images in the fourth row in *Figure 11* illustrate extramedullary infiltration MM. In this context, our proposed model performed excellently, achieving accurate segmentation without confusion with the surrounding soft tissue. This further demonstrates the high generalization capability and reliability of our model in segmenting MM lesions.

To validate the significant differences between MMNet and the other image segmentation models, we conducted independent samples *t*-tests to compare the P values of these metrics, which was followed by Benjamini-Hochberg multiple testing correction. The results are presented in *Table 11*. Specifically, for the IoU, Dice, and recall metrics, the P values for each model were predominantly

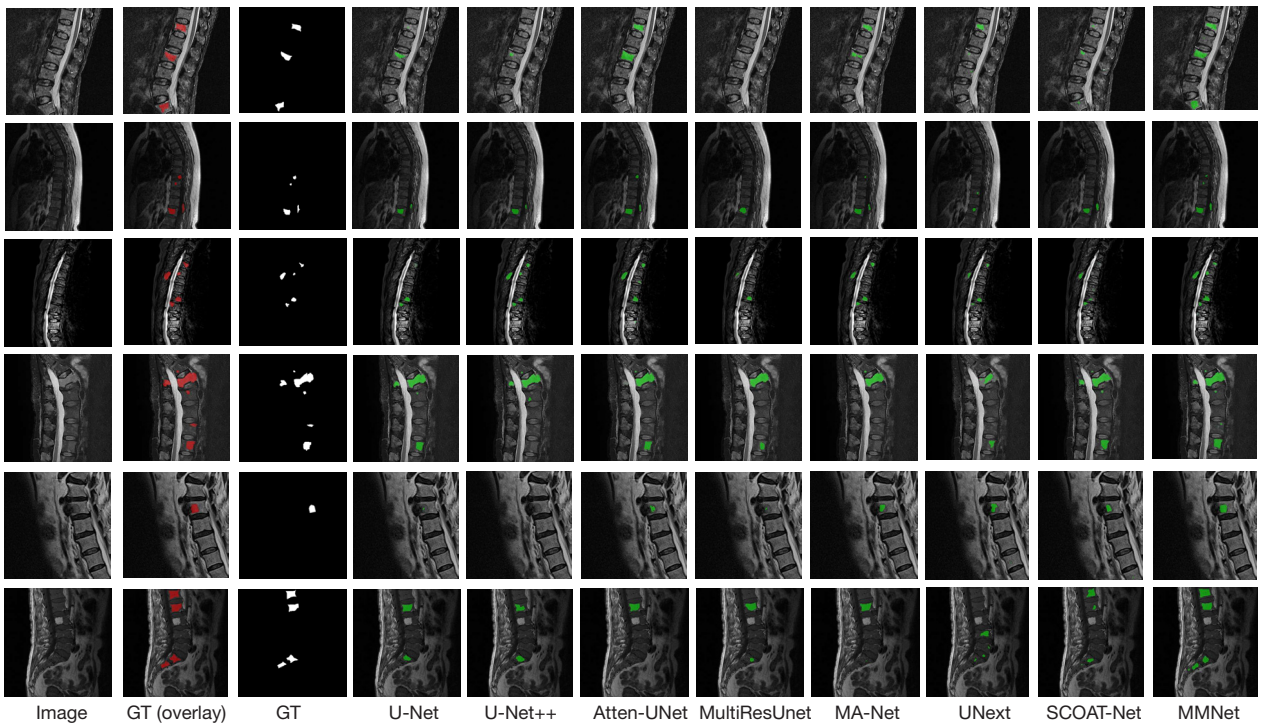

Figure 11 The efficacy of various segmentation models was visually assessed using a custom multiple myeloma segmentation dataset. From left to right respectively are the original image and the ground truth label, followed by the segmentation results of U-Net, U-Net++, Attention-UNet, MultiResUnet, MA-Net, UNeXt, SCOAT-Net, and the proposed MMNet model. Red represents ground truth labels, while green represents the model's predicted regions. GT, ground truth; Atten-UNet, attention U-Net; MA-Net, multi-scale attention net; SCOAT-Net, spatial- and channel-wise coarse-to-fine attention network; MMNet, multiple myeloma segmentation net.

Table 11 The P values from the significance test (*t*-test) between MMNet and other compared methods in different metrics (corrected for multiple testing via the Benjamini-Hochberg adjustment)

Method	P value				
	IoU	HD	Precision	Recall	Dice
U-Net	1.04×10^{-3}	2.66×10^{-1}	7.03×10^{-1}	1.68×10^{-5}*	9.78×10^{-4}*
U-Net++	1.29×10^{-4}*	7.04×10^{-2}	6.89×10^{-1}	3.91×10^{-6}*	1.29×10^{-4}*
Atten-UNet	1.95×10^{-4}*	1.37×10^{-1}	1.87×10^{-1}	8.56×10^{-6}*	1.95×10^{-4}*
UNeXt	1.58×10^{-22}*	2.11×10^{-6}*	3.96×10^{-10}*	1.08×10^{-22}*	4.78×10^{-19}*
MA-Net	2.45×10^{-5}*	4.90×10^{-2}	7.64×10^{-2}	3.82×10^{-6}*	2.45×10^{-5}*
MultiRes-UNet	4.86×10^{-12}*	9.66×10^{-6}*	3.33×10^{-2}	2.95×10^{-13}*	1.03×10^{-10}*
SCOAT-Net	6.71×10^{-7}*	2.26×10^{-1}	6.25×10^{-1}	8.02×10^{-10}*	2.09×10^{-6}*

*, P values less than 0.001. MMNet, multiple myeloma segmentation net; IoU, intersection over union; HD, Hausdorff distance; Dice, Dice similarity coefficient; Atten-UNet, attention U-Net; MA-Net, multi-scale attention net; SCOAT-Net, spatial- and channel-wise coarse-to-fine attention network.

Table 12 Comparative analysis of performance: DCNet versus other backbone networks

Backbone	IoU	HD	Precision	Recall	Dice
MobileNet	0.467±0.033	60.17±4.45	0.687±0.034	0.574±0.036	0.603±0.033
EfficientNet	0.533±0.034	59.94±5.35	0.749±0.031	0.627±0.036	0.663±0.033
DenseNet	0.644±0.033	50.25±5.19	0.850±0.023	0.713±0.033	0.755±0.028
ResNet	0.650±0.032	48.25±4.95	0.853±0.021*	0.721±0.033	0.760±0.028
DCNet ours)	0.682±0.029*	44.67±4.88*	0.845±0.020	0.766±0.029*	0.787±0.024*

Each evaluation index is expressed as mean ± 95% confidence interval. *, best result in the table. DCNet, dilated dense connected net; Backbones, different feature extraction networks; IoU, intersection over union; HD, Hausdorff distance; Dice, Dice similarity coefficient.

Table 13 P values from significance tests (*t*-tests) between the DCNet and other feature extraction models across different metrics (corrected for multiple testing via the Benjamini-Hochberg adjustment)

Method	P value				
	IoU	HD	Precision	Recall	Dice
MobileNet	3.65×10^{-18}*	5.93×10^{-6}*	1.13×10^{-13}*	1.08×10^{-14}*	2.11×10^{-16}*
EfficientNet	1.10×10^{-9}*	4.43×10^{-5}*	9.08×10^{-7}*	2.23×10^{-8}*	1.18×10^{-8}*
DenseNet	1.80×10^{-1}	1.87×10^{-1}	7.32×10^{-1}	1.03×10^{-1}	1.80×10^{-1}
ResNet	3.04×10^{-1}	3.73×10^{-1}	5.80×10^{-1}	2.65×10^{-1}	3.04×10^{-1}

*, P values less than 0.001. DCNet, dilated dense connected net; IoU, intersection over union; HD, Hausdorff distance; Dice, Dice similarity coefficient.

below 0.001. This indicates the accuracy and robustness of MMNet, further substantiating the veracity of these observed differences.

Comparative experiments with other backbone networks

To assess the comparative performance of DCNet against other feature extraction networks in extracting features from MM images, we conducted a comprehensive evaluation. We selected several widely employed backbone networks for comparison, including MobileNet (45), EfficientNet (46), DenseNet, and ResNet (47). The results of our horizontal comparison experiment are presented in *Table 12*. Our proposed DCNet outperformed all other networks, exhibiting superior scores across various metrics such as IoU, Hausdorff distance, recall, and Dice. These results suggest that DCNet excels in capturing information across different scales, thus substantiating its superiority and effectiveness in feature extraction for MM images.

Moreover, to validate the significant differences of DCNet compared to other feature extraction networks, we employed an independent samples *t*-test to compare the P values of these metrics and applied Benjamini-Hochberg correction for multiple testing. The detailed results are presented in *Table 13*. For MobileNet and EfficientNet, the corrected P values of all metrics remained relatively low. In contrast, for DenseNet and ResNet, the corrected P values were comparatively higher. These findings suggest that DCNet may exhibit performance differences across various metrics when compared to DenseNet and ResNet. These differences were more pronounced when the comparison involved MobileNet and EfficientNet, further highlighting the potential advantages of DCNet as a feature extraction network for MM lesion segmentation.

Discussion

Conducting multiscale feature extraction on images of MM patients is pivotal for the efficient segmentation of MM lesions. We employed an encoding-decoding network structure and integrated newly designed modules, DCNet and CASPP, to efficiently extract multiscale features from

images and enhance the correlation between channel and spatial information. Furthermore, MMNet incorporates our newly proposed DyCat module in the feature fusion process, enabling dynamic feature selection and adaptive adjustment of the fusion process. Ablation experiments demonstrated the effectiveness of the individual modules proposed in our method. Our approach achieved an IoU score of 0.716±0.025 and a Dice coefficient score of 0.819±0.019 on the self-constructed dataset.

To the best of our knowledge, most current segmentation tasks in the field of MM focus on the segmentation of myeloma plasma cells within microscopic images. Paing *et al.* (14) developed a variety of mask R-CNN models using different image types, including raw microscopic images, contrast-enhanced images, and stained cell images, for instance segmentation of MM cells. They applied deep augmentation techniques to enhance the performance of the Mask R-CNN model. Bozorgpour *et al.* (48) designed a two-stage deep learning approach for detecting and segmenting MM plasma cells, which was evaluated in the SegPC2021 Grand Challenge and achieved second place in the final testing phase among all participating teams. Moving beyond microscopic images, Wennmann *et al.* (17) trained an nnU-Net on a multicenter dataset to automatically segment bone marrow from whole-body ADC images, achieving segmentation quality comparable to that of manual methods. The automatically extracted ADC values were significantly correlated with PCI, thus demonstrating potential value for automatic staging, risk stratification, and treatment response assessment. In our study, we used the spinal MR images of patients to segment MM lesions. This imaging approach provided improved localization and assessment of lesion conditions. Additionally, we proposed a model with enhanced features in feature extraction, upsampling, and feature fusion, which more effectively captures global image information and is suitable for segmenting the complex and variable lesions associated with MM. In future research, we will further explore the application of MMNet in other medical image segmentation tasks. We plan to extend its application to whole-body imaging datasets and 3D datasets to validate its performance in different dimensions and more complex scenarios. Moreover, we will focus on investigating MMNet's potential in multimodal medical image processing, integrating various types of imaging data (such as MRI, CT, and PET) for joint analysis, with the aim of improving segmentation accuracy and diagnostic precision.

Conclusions

This study focused on the automatic and precise segmentation of MM in MRI. Through an analysis of the distinctive characteristics of MM lesions, we propose an innovative automatic segmentation method, MMNet. Through the integration of three innovative modules, DCNet, CASPP, and DyCat, the network performance is improved in the coding, decoding, and feature fusion components, respectively. This enhances the ability to extract multiscale features, expands the depth and width of the network, enriches the diversity of features and the feature fusion process, and effectively improves the accuracy of the model in lesion segmentation. Extensive experimental results confirmed the ability of our proposed module to achieve precise segmentation of MM lesions. Furthermore, our findings confirmed the superior performance of our proposed model in MM segmentation when compared to other advanced image segmentation models.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (No. 62376044), the Natural Science Foundation of Chongqing (No. CSTB2022NSCQ-MSX0801), and the Chongqing Medical Scientific Research Project (Joint Project of Chongqing Health Commission and Science and Technology Bureau).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-24-683/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of The First Affiliated Hospital of Chongqing Medical University (No. K2023-314). The requirement for individual consent was waived due to the retrospective nature of the analysis.

References

1. Dimopoulos MA, Terpos E. Multiple myeloma. Ann Oncol 2010;21 Suppl 7:vii143-50.
2. Zhang C, Zhang Y. Bone marrow particle enrichment analysis for the laboratory diagnosis of multiple myeloma: A case study. J Clin Lab Anal 2020;34:e23372.
3. Wennmann M, Kintzelé L, Piraud M, Menze BH, Hielscher T, Hofmanninger J, Wagner B, Kauczor HU, Merz M, Hillengass J, Langs G, Weber MA. Volumetry based biomarker speed of growth: Quantifying the change of total tumor volume in whole-body magnetic resonance imaging over time improves risk stratification of smoldering multiple myeloma patients. Oncotarget 2018;9:25254-64.
4. Wennmann M, Hielscher T, Kintzelé L, Menze BH, Langs G, Merz M, Sauer S, Kauczor HU, Schlemmer HP, Delorme S, Goldschmidt H, Weinhold N, Hillengass J, Weber MA. Analyzing Longitudinal wb-MRI Data and Clinical Course in a Cohort of Former Smoldering Multiple Myeloma Patients: Connections between MRI Findings and Clinical Progression Patterns. Cancers (Basel) 2021;13:961.
5. Piraud M, Wennmann M, Kintzelé L, Hillengass J, Keller U, Langs G, Weber MA, Menze BH. Towards quantitative imaging biomarkers of tumor dissemination: A multi-scale parametric modeling of multiple myeloma. Med Image Anal 2019;57:214-25.
6. Hillengass J, Usmani S, Rajkumar SV, Durie BGM, Mateos MV, Lonial S, et al. International myeloma working group consensus recommendations on imaging in monoclonal plasma cell disorders. Lancet Oncol 2019;20:e302-12.
7. Gariani J, Westerland O, Natas S, Verma H, Cook G, Goh V. Comparison of whole body magnetic resonance imaging (WBMRI) to whole body computed tomography (WBCT) or (18)F-fluorodeoxyglucose positron emission tomography/CT ((18)F-FDG PET/CT) in patients with myeloma: Systematic review of diagnostic performance. Crit Rev Oncol Hematol 2018;124:66-72.
8. Cho HJ, Jung SH, Jo JC, Lee YJ, Yoon SE, Park SS, et al. Development of a new risk stratification system for patients with newly diagnosed multiple myeloma using R-ISS and (18)F-FDG PET/CT. Blood Cancer J 2021;11:190.
9. Rasche L, Kortüm KM, Raab MS, Weinhold N. The Impact of Tumor Heterogeneity on Diagnostics and Novel Therapeutic Strategies in Multiple Myeloma. Int J Mol Sci 2019;20:1248.
10. Chiarilli MG, Delli Pizzi A, Mastrodicasa D, Febo MP, Cardinali B, Consorte B, Cifaratti A, Panara V, Caulo M, Cannataro G. Bone marrow magnetic resonance imaging: physiologic and pathologic findings that radiologist should know. Radiol Med 2021;126:264-76.
11. Al-Sabbagh A, Ibrahim F, Szabados L, Soliman DS, Taha RY, Fernyhough LJ. The Role of Integrated Positron Emission Tomography/Computed Tomography (PET/CT) and Bone Marrow Examination in Staging Large B-Cell Lymphoma. Clin Med Insights Oncol 2020;14:1179554920953091.
12. Nanni C. PET/CT in multiple myeloma. Nucl Med (Stuttg) 2017;41:216-20.
13. Qiu X, Lei H, Xie H, Lei B. Segmentation of Multiple Myeloma Cells Using Feature Selection Pyramid Network and Semantic Cascade Mask RCNN. 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), Kolkata, India, 2022:1-4.
14. Paing MP, Sento A, Bui TH, Pintavirooj C. Instance Segmentation of Multiple Myeloma Cells Using Deep-Wise Data Augmentation and Mask R-CNN. Entropy (Basel) 2022;24:134.
15. Wennmann M, Klein A, Bauer F, Chmelik J, Grözinger M, Uhlenbrock C, et al. Combining Deep Learning and Radiomics for Automated, Objective, Comprehensive Bone Marrow Characterization From Whole-Body MRI: A Multicentric Feasibility Study. Invest Radiol 2022;57:752-63.
16. Wennmann M, Ming W, Bauer F, Chmelik J, Klein A, Uhlenbrock C, et al. Prediction of Bone Marrow Biopsy Results From MRI in Multiple Myeloma Patients Using Deep Learning and Radiomics. Invest Radiol 2023;58:754-65.
17. Wennmann M, Neher P, Stanczyk N, Kahl KC, Kächele J, Weru V, et al. Deep Learning for Automatic Bone Marrow Apparent Diffusion Coefficient Measurements From Whole-Body Magnetic Resonance Imaging in Patients With Multiple Myeloma: A Retrospective Multicenter Study. Invest Radiol 2023;58:273-82.

18. Sezgin M, Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging 2004;13:146-68.
19. Adams R, Bishof L. Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence 1994;16:641-7.
20. Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986;8:679-98.
21. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A. editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, Springer, 2015;9351:234-41.
22. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018;40:834-48.
23. Milletari F, Navab N, Ahmadi SA. 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 2016:565-71.
24. Shafiq M, Gu ZQ. Deep Residual Learning for Image Recognition: A Survey. Appl Sci 2022;12:8972.
25. Xiao X, Lian S, Luo Z, Li S, editors. Weighted Res-UNet for High-Quality Retina Vessel Segmentation. 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 2018:327-31.
26. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018) 2018;11045:3-11.
27. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, Mcdonagh S, Hammerla NY, Kainz B. Attention U-Net: Learning Where to Look for the Pancreas. arXiv: 1804.03999. 2018.
28. Chen J, Lu Y, Yu Q, Luo X, Zhou Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv 2102.04306. 2021.
29. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention Is All You Need. Part of Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017.
30. Lin TY, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature Pyramid Networks for Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:2117-25.
31. Xia H, Ma M, Li H, Song S. MC-Net: multi-scale context-attention network for medical CT image segmentation. Appl Intell 2021;52:1508-19.
32. Yang J, Hu X, Pan H, Chen P, Xia S. Multi-scale attention network for segmentation of electron dense deposits in glomerular microscopic images. Microsc Res Tech 2022;85:3256-64.
33. Yan Q, Wang B, Gong D, Luo C, Zhao W, Shen J, Ai J, Shi Q, Zhang Y, Jin S, Zhang L, You Z. COVID-19 Chest CT Image Segmentation Network by Multi-Scale Fusion and Enhancement Operations. IEEE Trans Big Data 2021;7:13-24.
34. Li JY, Cheng LL, Xia TJ, Ni HM, Li J. Multi-Scale Fusion U-Net for the Segmentation of Breast Lesions. IEEE Access 2021;9:137125-39.
35. Chen Y, Dai X, Liu M, Chen D, Yuan L, Liu Z. Dynamic Convolution: Attention over Convolution Kernels. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020:11030-9.
36. Dai Y, Gieseke F, Oehmcke S, Wu Y, Barnard K. Attentional Feature Fusion. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2021:3560-9.
37. Yu F, Koltun V, Funkhouser T. Dilated Residual Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:472-80.
38. Woo SH, Park J, Lee JY, Kweon IS. CBAM: Convolutional Block Attention Module. Proceedings of the European Conference on Computer Vision (ECCV), 2018:3-19.
39. Huang G, Liu Z, Laurens VDM, Weinberger KQ. Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:4700-8.
40. Yu Z, Huang H, Chen W, Su Y, Liu Y, Wang XY. YOLO-FaceV2: A Scale and Occlusion Aware Face Detector. arXiv: 2208.02019. 2022.
41. Valanarasu JMJ, Patel VM. UNeXt: MLP-Based Rapid Medical Image Segmentation Network. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S. editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. Lecture Notes in Computer Science, Springer, 2022;13435:23-33.
42. Fan TL, Wang GL, Li Y, Wang HR. MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. IEEE Access 2020;8:179656-65.
43. Ibtehaz N, Rahman MS. MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image

segmentation. Neural Netw 2020;121:74-87.
44. Zhao S, Li Z, Chen Y, Zhao W, Xie X, Liu J, Zhao D, Li Y. SCOAT-Net: A novel network for segmenting COVID-19 lung opacification from CT images. Pattern Recognit 2021;119:108109.
45. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv: 1704.04861. 2017.
46. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning, PMLR 97:6105-6114, 2019.
47. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:770-8.
48. Bozorgpour A, Azad R, Showkatian E, Sulaiman A. Multi-scale Regional Attention Deeplab3+: Multiple Myeloma Plasma Cells Segmentation in Microscopic Images. arXiv: 2105.06238. 2021.

Cite this article as: Zhao X, Chen L, Zhang N, Lv Y, Hu X. Multiple myeloma segmentation net (MMNet): an encoder-decoder-based deep multiscale feature fusion model for multiple myeloma segmentation in magnetic resonance imaging. Quant Imaging Med Surg 2024;14(10):7176-7199. doi: 10.21037/qims-24-683