scientific reports

Check for updates

OPEN

Dual-channel compression mapping network with fused attention mechanism for medical image segmentation

Xiaokang Ding[1,3], Ke'er Qian[1✉], Qile Zhang[2,3✉], Xiaoliang Jiang[1] & Ling Dong[1]

Accurate image segmentation is the key to quantitative analysis and recognition of pathological tissues in medical imaging technology, which can provide important technical support for medical diagnosis and treatment. However, the task of lesion segmentation is particularly challenging due to the difficulty in identifying edges, the complexity of different tissues, and the variability in their shapes. To address these challenges, we propose a dual-channel compression mapping network (DCM-Net) with fused attention mechanism for medical image segmentation. Firstly, a dual-channel compression mapping module is added to U-Net's standard convolution blocks to capture inter-channel information. Secondly, we replace the traditional skip path with a fusion attention mechanism that can better present context information in high-level features. Finally, the combination of squeeze-and-excitation module and residual connection in the decoder part can improve the adaptive ability of the network. Through extensive experiments on various medical image datasets, DCM-Net has demonstrated superior performance compared to other models. For instance, on the ISIC database, our network achieved an Accuracy of 91.42%, True Positive Rate (TPR) of 88.93%, Dice of 86.09%, and Jaccard of 76.02%. Additionally, on the pituitary adenoma dataset from Quzhou People's Hospital, DCM-Net reached an Accuracy of 97.07%, TPR of 93.09%, Dice of 92.29%, and Jaccard of 87.73%. These results demonstrate the effectiveness of DCM-Net in providing accurate and reliable segmentation, and it shows valuable potential in the field of medical imaging technology.

Medical image segmentation is an important subject in modern image analysis, and it is the key to accurately identify and separate different tissues, structures or pathological regions in medical images. This capability provides clinicians with a solid foundation for precise quantitative analysis and visualization, which are essential for accurate diagnosis, effective treatment planning, and continuous patient monitoring[1]. With the widespread application of imaging techniques such as ultrasound, MRI, and CT, the acquisition of medical images has become increasingly rapid and high-resolution. While this technological advancement has significantly improved diagnostic capabilities, it has also introduced new challenges, including the management of large volumes of data and the increased complexity of the information contained within these images[2]. Traditional manual image segmentation methods are the gold standard, but they are time-consuming and limited by operator experience and potential biases that make it difficult to meet the precision and efficiency requirements of modern medicine. Therefore, current research is intensely focused on developing efficient, automatic, and accurate medical image segmentation algorithms to ensure that they can be reliably used in different imaging modes and clinical scenarios.

Classical segmentation algorithms, such as fuzzy clustering and level set, have been instrumental in the field of medical imaging. However, these technologies have some problems such as long segmentation time, low precision and manual setting of some important parameters. In addition, external factors such as sharpness and brightness in image acquisition will also affect the segmentation quality. In contrast, deep learning algorithms have revolutionized the field of medical imaging by automating feature extraction and achieving numerous breakthroughs. Furthermore, the researchers are working on designing more accurate, robust, and adaptive segmentation models to better handle external factors such as brightness changes and noise interference. For example, fully convolutional network (FCN)[3] and U-Net[4] have become standard frameworks with their powerful

[1]College of Mechanical Engineering, Quzhou University, Quzhou 324000, China. [2]Department of Rehabilitation, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou People's Hospital, Quzhou 324000, China. [3]Xiaokang Ding and Qile Zhang contributed equally to this work. ✉email: moneyke2@qzc.edu.cn; zjzhangqile@163.com

feature extraction capabilities and end-to-end learning methods. These models achieve fine target differentiation by learning the mapping from the original image to the segmentation mask, significantly improving the accuracy and stability of segmentation. Despite significant progress in the field, medical image segmentation still faces many challenges that require ongoing exploration of more efficient and robust algorithms to meet diverse clinical needs. Specifically, the complexity of anatomical variations among different patient populations, noise and artifacts in imaging modalities, and the need for real-time processing in certain medical scenarios all underscore the need for further innovation in algorithm design. Additionally, ensuring the generalizability of segmentation models across various imaging modalities and disease types remains a critical goal in translating these advances into widespread clinical applications. Therefore, the continuous exploration and development of more efficient and robust algorithms are essential to meet the diverse and evolving needs of clinical practice, which is essential to improve the accuracy, reliability and applicability of medical image segmentation.

Since the introduction of the U-Net, there has been a surge in research proposing its utilization across various advanced techniques, including attention mechanisms[5,6], dual U-Net structures[7,8], and multi-scale feature fusion[9,10]. Among them, Nawaz et al.[11] utilized the CornerNet to compute a reliable set of features aimed at precisely identifying the locations of melanoma lesions. Once the feature set was obtained, the fuzzy k-means algorithm was employed to carry out the segmentation process with a higher degree of precision. Sun et al.[12] integrated Transformer and DA-block into U-Net architecture and proposed a segmentation algorithm named DA-TransUNet. This innovative approach combines Transformers' powerful feature extraction capabilities with the efficiency and precision of DA-Block to deliver more accurate and reliable results for clinical applications. Zhang et al.[13] introduce CT-Net, a cutting-edge model designed to efficiently extract local and global representations for comprehensive analysis of medical images through its innovative asymmetric asynchronous branch-parallel architecture. Zhao et al.[14] developed an innovative approach by combining the principles of text attention with the diffusion theory, which enhance the model's ability to focus on and interpret key regions within the target region. Inspired by U-shaped network, Zhang et al.[15] introduced parallel expansion pooling module, large kernel convolution, and pyramid architecture into the U-Net, which not only accelerates the learning speed but also significantly improves its ability to distinguish and process complex patterns within medical images. Ansari et al.[16] proposed a novel pyramid scene parsing module that operates on the skip connections within fixed-width neural networks. This module is specifically designed to extract features at multiple scales while capturing rich context associations from different levels of the network.

Inspired by the above methods, this article proposes a two-channel compression mapping network with fused attention mechanism for medical image segmentation. A comprehensive series of experiments were conducted on two medical image segmentation databases, ISIC-2018 and a pituitary adenoma dataset, to evaluate the performance of the proposed DCM-Net. The results of these experiments demonstrated that DCM-Net consistently outperformed the most advanced existing methods in terms of segmentation accuracy and reliability. These findings demonstrate the superiority of DCM-Net in medical image segmentation tasks and its potential to improve diagnostic accuracy and clinical decision making processes. Our main contributions are given as:

(1) To enhance the feature extraction capabilities of the network, a dual-channel compressed mapping module is incorporated into the convolutional block. This module facilitates a richer and more detailed representation of the input data, which can improve the overall performance of the segmentation process.
(2) The implementation of fusion attention mechanisms in the connections between the encoder and decoder significantly enhances the model's ability to present contextual information within high-level features. These attention mechanisms selectively focus on the most relevant parts of a feature map, which can enable a neural network to integrate important context clues throughout the encoding and decoding phases.
(3) In the decoder part of the network, the combination of the SE module and residual connections is employed to improve the adaptive capabilities. These components enhance the robustness and flexibility of the algorithm, while suppressing less useful features and solving the gradient disappearance problem.

Methods
Overview of DCM-Net
Our DCM-Net is mainly composed of three primary stages: encoding, decoding, and encoding-decoding connection. Each of these stages plays a pivotal role in ensuring the network's effective operation and high performance in image segmentation tasks. To enhance the functionality of these stages, we have designed several specialized modules, we have designed some modules, including dual-channel compression mapping module, fusion attention mechanism, and squeeze-and-excitation module, as shown in Fig. 1. For the encoder part, after adding the dual-channel compressed mapping, the information is compressed by parallel channels to obtain more efficient features, and then redundant and miscellaneous items are removed by mapping. For the decoder part, we combine the SE module with the residual connection. The SE module enhances the input feature mapping by recalibrating the channel feature response, while the residual connection provides a bypass around the nonlinear transformation to mitigate problems such as gradient disappearance or gradient explosion. Finally, in the connecting part of the encoder and decoder, we introduce a fusion attention mechanism to replace the traditional skip connections. This mechanism comprises of attention gates, spatial attention and channel attention, which can effectively integrate features from different levels of the network to solve the semantic gap between the encoder and the decoder. By meticulously refining each component and ensuring seamless integration between stages, DCM-Net is capable of achieving superior segmentation performance.
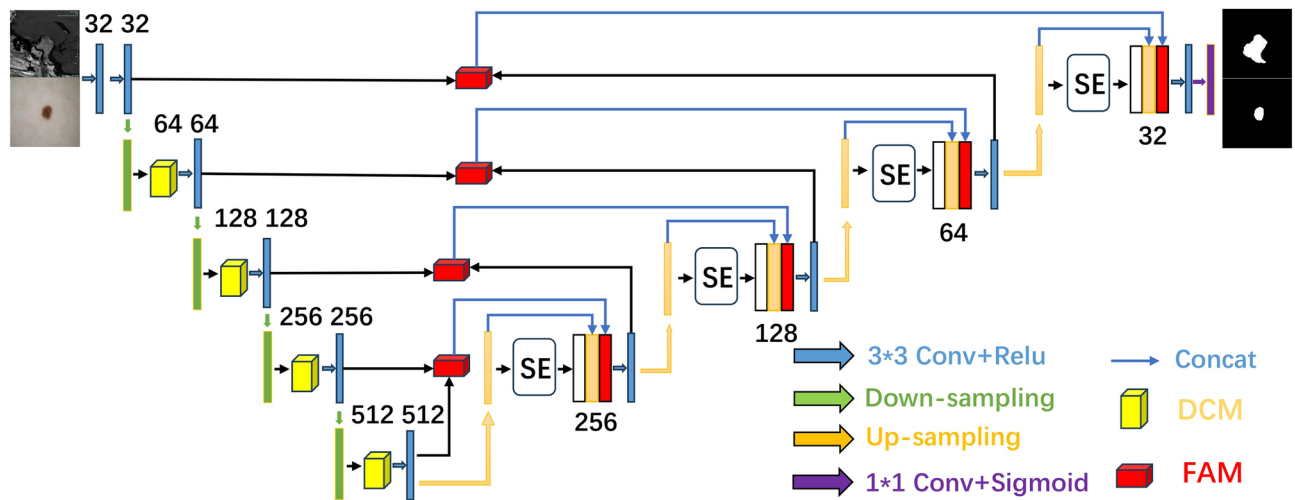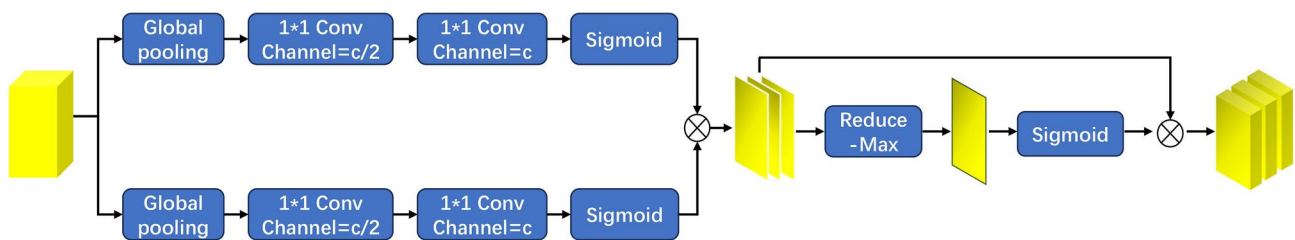
Fig. 1. Architecture of DCM-Net.

Fig. 2. Structure of dual-channel compression mapping module.

Dual-channel compression mapping module

In the U-Net architectures, traditional convolution blocks often fall short in effectively extracting features and capturing long-range context information. To address these limitations, we designed a dual-channel compression mapping module, as illustrated in Fig. 2. Specifically, the feature map is split into two identical parallel feature extraction paths, each path incorporates a global pooling operation followed by two fully connected layers. The global pooling operation is critical because it reduces the spatial dimension of the feature graph and helps suppress less relevant information. In the subsequent stage within each path, the first fully connected layer employs a 1×1 convolution kernel. This layer performs a crucial role by halving the number of channels, thereby effectively capturing and highlighting the dependencies between channels. The second fully connected layer has the task of restoring the number of channels to the original number, which ensures that the information processed through the initial compression is fully reintegrated. The output of these parallel paths is then merged to take full advantage of their respective strengths and capabilities. In addition, the resulting combined output undergoes a minimization process, specifically designed to highlight the most crucial features across the channels. To further refine the model's ability to learn and adapt, a sigmoid activation function is applied, each followed by batch normalization (BN) and ReLU activation[17], to make the model better capture complex patterns and improve the prediction accuracy. Finally, element-by-element multiplication is performed between the feature maps of spatially compression and channel compression. This step ensures that greater activation values are assigned to the most important features, thereby increasing the model's focus on key spatial information and improving the overall effectiveness of the feature extraction process. Through this comprehensive and multi-layered approach, the dual-channel compression mapping module substantially elevates the model's capability to perform precise and accurate feature extraction, which is crucial for high-quality medical image segmentation.

Fused attention mechanism

Inspired by the ability of the human visual system to selectively focus on significant information, attention mechanism[18,19] has been introduced into the field of computer vision and has achieved remarkable success. In the context of image segmentation, attention mechanism can effectively identify and focus on key areas in complex scenes, which improves the accuracy and speed of segmentation process. Despite their success, traditional attention mechanisms face significant challenges when applied to the complex and often noisy field of medical images. This misallocation of attention can lead to suboptimal segmentation outcomes in which important pathological features can be overlooked or misunderstood. To address these challenges, we introduce attention gates[20,21] mechanism, which dynamically control the importance of different spatial location features in
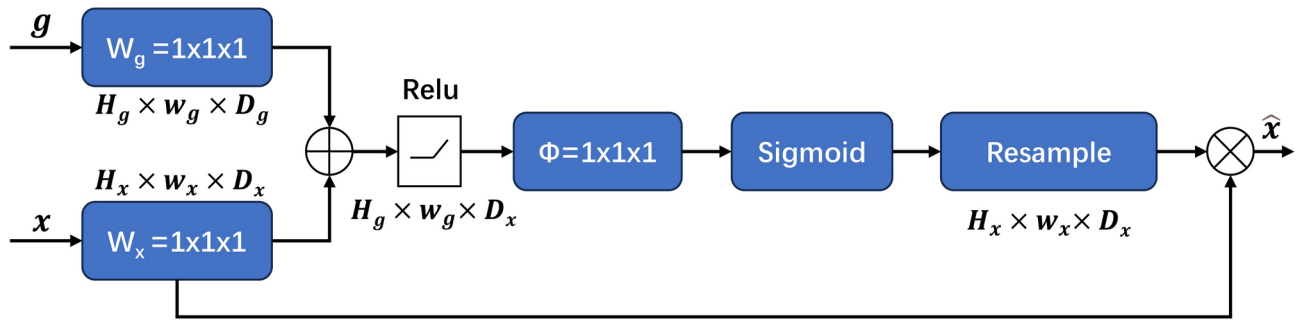
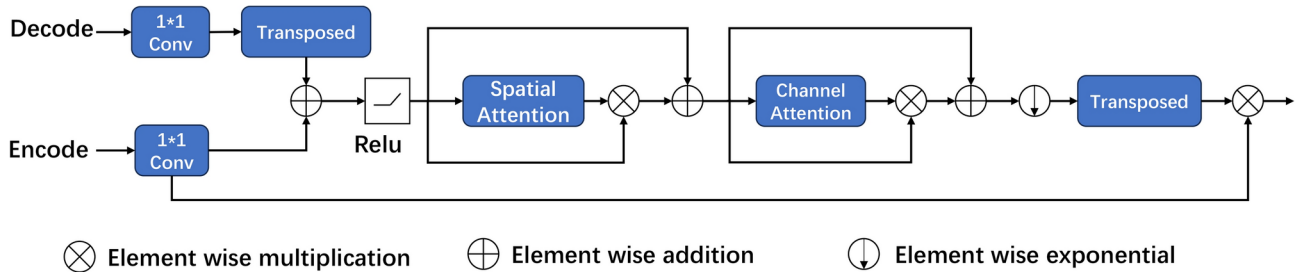**Fig. 3**. Structure of attention gates module.

Fig. 4. Structure of fused attention mechanism module.

an image by generating a gated signal, as shown in Fig. 3. This mechanism effectively enhances the model's ability to recognize and focus on task-relevant areas, and it ensures that the most relevant areas in the image receive proper attention. In the specific task of image segmentation, the addition of attention gate helps to integrate the information of different layers of the network more effectively. Especially in the case of skipping connections, it enables the network to more efficiently utilize feature information at different scales.

Building upon the foundation of the attention gate structure, we introduce an innovative hybrid attention mechanism seamlessly integrates both spatial attention and channel attention, as illustrated in Fig. 4. By combining the synergies created by these different forms of attention, a more adaptable and robust feature extraction process can be achieved, as well as better handling the different complexities and nuances of medical images.

Different from the original attention gate structure, the process of up-sampling is ranked after the 1×1 convolutional layer in the fused attention mechanism. We use deconvolution up-sampling method, the principle of which is to restore the original size of the feature map by filling zero values around the feature map. To prevent the convolution layer from extracting useless features, the convolutional layer is scheduled before up-sampling. Both inputs are convolved with 1×1 simultaneously to ensure a consistent size by up-sampling. After that, the two elements are added as dots and activated using the ReLU function, with the result represented by $F^{H \times W \times C}$. In the next step, $F^{H \times W \times C}$ is processed with spatial attention[22,23], which mainly weights outputs from different regions to enhance specific target regions of interest. Following this, the feature map is subjected to channel attention[24,25], which directs its focus towards extracting valuable information from a channel perspective. In the final stage, each element undergoes exponential calculation to enhance its significance. Subsequently, the size of these elements is adjusted through up-sampling to match the input size of the attention gate. This multiplication step completes the hybrid process of the entire attention mechanism, which ensures that the network can effectively utilize spatial and channel attention features.

As illustrated in Fig. 5, spatial attention performs global max pooling and global average pooling of the channel dimension on $F^{H \times W \times C}$, resulting in two H×W×1 feature maps. Then, the results of global max pooling and global average pooling are concatenated according to the channel to obtain the feature map size of H×W×2. Subsequently, the 7×7 convolution operation is performed on the concatenated results to obtain the feature map size of H×W×1. Finally, the spatial attention weight matrix $M_S(F)$ is obtained by sigmoid activation function, and its calculation formula is as follows:

$$M_S(F) = \sigma \left(F^{7 \times 7}([AvgPool(F); MaxPool(F)]) \right) \tag{1}$$

As illustrated in Fig. 6, the initial step of channel attention involves global maximum pooling and global average pooling of the input in the spatial dimension. Subsequently, the outcomes of these operations are directed to a shared multi-layer perceptron (MLP)[26] for learning purposes, yielding two 1×1×C feature maps. Notably, the number of neurons in the first layer of MLP is C/r with ReLU activation function while there are C neurons in its second layer. Finally, following an addition operation on the output of MLP, it undergoes mapping through a
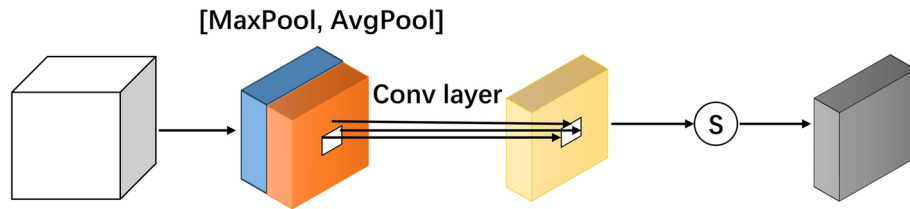
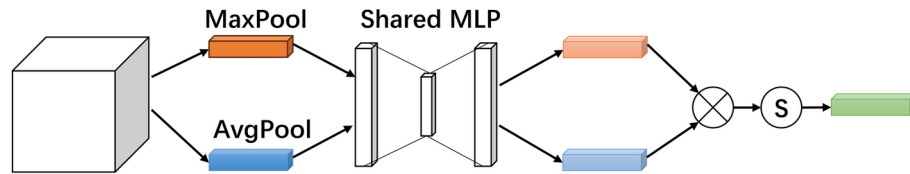**Fig. 5**. Structure of spatial attention module.

Fig. 6. Structure of channel attention module.
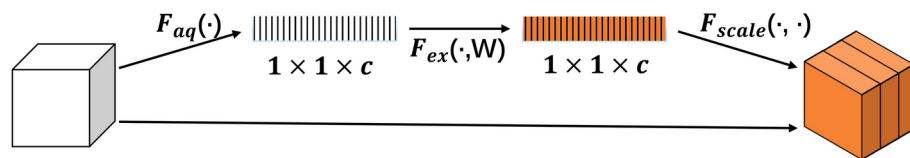

Fig. 7. Structure of squeeze-and-excitation module.

sigmoid activation function to obtain the channel attention weight matrix $M_C(F)$, and its calculation formula is as follows:

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{2}$$

SE module and residual connection

As shown in Fig. 7, the squeeze-and-excitation module[27,28] comprises two key components: the squeeze and the excitation stages. The squeeze component is designed to reduce the spatial dimensions of the input feature map, effectively capturing essential global spatial information. In contrast, the excitation component is responsible for learning and deriving adaptive weights for each channel in the input feature map. Subsequently, these derived weights are utilized to perform a multiplication operation with the input feature map, resulting in the generation of the final output feature map. Given its capacity to dynamically adjust channel weights and thereby enhance model performance, the SE module is integrated into the decoder section of the network. In the decoder, the up-sampling results are processed by SE module and residual connection respectively, and then the two output results are concatenated. Additionally, a regular connection is employed to combine these outputs with those from a mixed attention mechanism. It is worth noting that this residual connection facilitates direct cross-layer connections and plays a pivotal role in improving feature learning as well as enhancing overall model performance.

Loss function

In various medical images, the characteristics, locations, and sizes of lesions can vary significantly. This diversity extends to the datasets used, which can differ widely in terms of their composition and complexity. Consequently, the segmentation ability of a model can vary across different datasets. In the process of model training, dice loss function[29,30] is used in this paper, and the formula is as follows:

$$L_{dice}(y, p) = 1 - \frac{2\sum_{i=1}^{N} p_i y_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i} \tag{3}$$

where N is the number of all pixels, p_i and y_i are the predicted and label masks of pixel I.
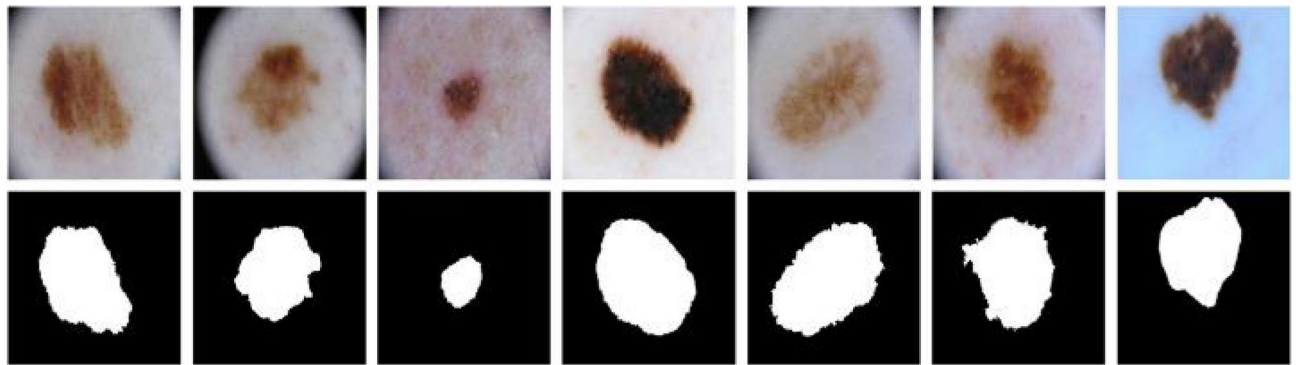
Fig. 8. Some images and their corresponding labels on ISIC-2018.
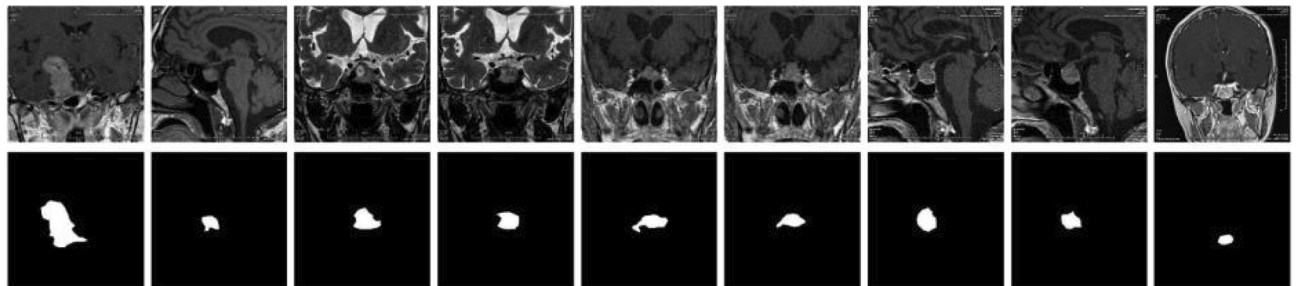

Fig. 9. Some images and their corresponding labels on pituitary adenoma.

Experimental results

Dataset

To verify the segmentation ability of the proposed DCM-Net, we conducted a series of experiments on two distinct and challenging datasets: International Skin Imaging Collaboration (ISIC-2018) dataset and pituitary adenoma dataset. The details of the dataset are as follows:

ISIC-2018 dataset

The ISIC dataset, which was released by the International Skin Imaging Collaboration, serves as a comprehensive repository of dermoscopic images intended for research and analysis within the field of dermatology. This extensive dataset covers multiple years, with each annual release offering distinct characteristics and applications. For our specific study, we concentrated on utilizing the 2018 edition of the ISIC dataset. Within the ISIC-2018 dataset, there are 3694 original dermoscopic images paired with corresponding binary label images that represent a diverse range of skin diseases. These images feature intricate backgrounds and noise elements such as hair and blood vessels, presenting significant challenges for accurate lesion segmentation and recognition. In our experimental setup, we assigned 2594 images for training purposes, 1000 images for testing, and 100 images for validation. To ensure resilience against potential influences from minor environmental variables that could impact segmentation accuracy assessments, we standardized all image pixel dimensions to 256×256 with three channels. Figure 8 offers a visual representation displaying selected original dermoscopic images alongside their corresponding binary label counterparts from within the ISIC-2018 dataset. The dataset can be obtained from: https://challenge.isic-archive.com/data/#2018.

Pituitary adenoma dataset

The pituitary adenoma images used in our study were obtained from Quzhou People's Hospital. These images are typically acquired using head X-ray plain films, while CT scans are employed for larger tumors. The scanning process covers the entire head, resulting in a more complex background compared to ISIC-2018 images. Considering the data set imbalance[31], we used 1400 images for training, 305 for validation, and 400 for testing. In order to ensure the consistency of training, the pixel value of each image was adjusted to 256×256 and 3 channels. Figure 9 showcases partial original images alongside their corresponding binary label images from the pituitary adenoma dataset.

The pituitary adenoma dataset from Quzhou People's Hospital is private. All methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were approved by Quzhou People's Hospital. Informed consent was obtained from all subjects and their legal guardian.

Evaluation metrics

To further confirm the performance of the DCM-Net, we adopt the evaluation metrics commonly used in medical image segmentation, including Accuracy[32,33], true positive rate(TPR)[34,35], Dice[36,37], and Jaccard[38,39], and they can be given as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$TPR = \frac{TP}{TP + FN} \tag{5}$$

$$Dice = \frac{2TP}{2TP + FN + FP} \tag{6}$$

$$Jaccard = \frac{TP}{TP + FN + FP} \tag{7}$$

Parameter selection

The experimental environment is built under Python3.7, and Keras is used to build the network framework. All models were trained on a Windows 64-bit system using an NVIDIA Quadro RTX 6000 with 8GB of RAM and an additional 24 GB of extended RAM. To ensure the efficiency and effectiveness of the training process, several key parameters were meticulously configured. In the training process, Adam optimizer is used, the learning rate is 0.5e−3, the number of network iteration training is set to 200, and the batch size is 16 during each training. The selection of the optimizer and learning rate is shown in Table 1. During convolution, padding is set to the 'SAME' mode, and the initial number of convolution kernels is 16. Additionally, a dropout rate of 0.2 was applied to introduce regularization, helping to prevent overfitting by randomly dropping 20% of the neurons during training. Figure 10 shows the loss and accuracy curve of DCM-Net during training and verification.

Comparison with other methods

In order to confirm the outstanding capability of the DCM-Net in medical image segmentation, we initially trained the DCM-Net with U-Net[4], OD-Segmentation[40], CL-Net[41], EE-Net[42], DR-Vnet[43], SK-U-Net[44], Double-Net[45], Nested-Net[46], Connected-Net[47], X-Net[48] and SA-Net[49] on the ISIC 2018 dataset. These comparative models' codes were obtained from open-source papers and had demonstrated good performance on various medical image datasets such as lungs, retinas, mandibles, etc. During training, all experiments followed a single variable principle, with consistent conditions maintained except for model code to ensure experiment accuracy. Table 2 presents the performance of all models on the ISIC 2018 dataset. Based on Table 2, it is evident that, apart from DCM-Net, the DR-Vnet model exhibits the highest evaluation metrics. By combining residual connections with dense blocks, it achieves precise image segmentation. The other models demonstrate relatively poor segmentation capabilities, indicating significant susceptibility to image background and noise interference, resulting in degraded model performance. In comparison to all open-source academic models, the proposed DCM-Net outperforms them across all four indicators. Particularly noteworthy are the Dice and Jaccard values which significantly surpass those of other models at 86.09% and 76.02%, respectively. While the Accuracy and TPR values are similar to those of other models, they also rank highest at 91.42% and 88.93%. These metrics show improvements over the original U-Net by 0.74%, 2.44%, 3.31%, and 4.77%, respectively. These findings suggest that the new modules utilized in DCM-Net, including DCM, FAM, residual connections, and SE, collectively enhance the model's segmentation capabilities.

In addition to the aforementioned comparative data that demonstrates the outstanding segmentation capability of DCM-Net, we also visualized the results for a more intuitive comparison of all models' performance,

	Accuracy	TPR	Dice	Jaccard
Optimizer (learning rate = 0.001)				
Adam	0.9142	0.8893	0.8609	0.7602
Adamax	0.9128	0.8879	0.8586	0.7582
RMSprop	0.9090	0.8843	0.8523	0.7526
Rprop	0.9137	0.8888	0.8601	0.7595
SGD	0.8965	0.8721	0.8312	0.7340
Learning rate (optimizer = Adam)				
0.0005	0.9091	0.8848	0.8602	0.7596
0.001	0.9142	0.8893	0.8609	0.7602
0.005	0.8829	0.8588	0.8092	0.7146
0.01	0.8752	0.8513	0.7973	0.7040
0.05	0.8352	0.8124	0.7356	0.6496
0.1	0.8334	0.8107	0.7325	0.6469

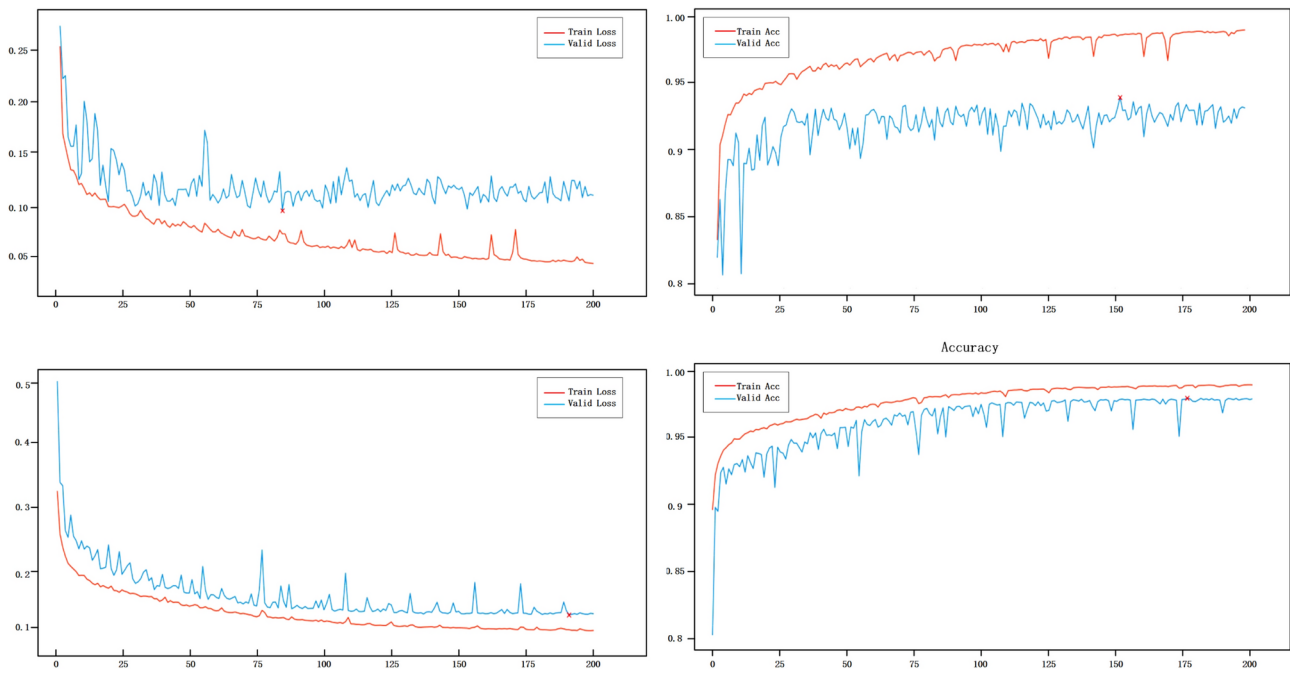**Table 1.** Results of experiments with different optimizer and learning rate on ISIC-2018.

Fig. 10. Loss and accuracy curves of DCM-Net.

Method	Accuracy	TPR	Dice	Jaccard
U-Net	0.9068	0.8549	0.8278	0.7125
OD-segmentation	0.9132	0.8574	0.8195	0.7002
CL-Net	0.8979	0.8450	0.8138	0.6925
EE-Net	0.9079	0.8789	0.8455	0.7369
DR-Vnet	0.9107	0.8659	0.8473	0.7421
SK-U-Net	0.9045	0.8564	0.8343	0.7216
Double-Net	0.9077	0.8834	0.8383	0.7268
Nested-Net	0.9096	0.8463	0.8348	0.7236
Connected-Net	0.9049	0.8538	0.8465	0.7409
X-Net	0.8917	0.8584	0.8178	0.6966
SA-Net	0.8992	0.8461	0.8194	0.7010
DCM-Net	0.9142	0.8893	0.8609	0.7602

Table 2. Results of comparative experiments on ISIC-2018.

as shown in Fig. 11. As can be seen from the study results in Fig. 11, it is evident that U-Net's performance in segmenting images with subtle lesions or irregular edges is not optimal. Similarly, Double-Net, Nested-Net, Connected-Net, and X-Net (rows 9, 10, 11, and 12 in Fig. 11) have all advanced their connection methods by adopting a dual U-Net structure for more intricate analysis of extracted features.However, these models are limited in their ability to extract multi-scale features and consequently struggle to accurately process more complex image segmentation tasks. In response to these limitations, SK-U-Net and SA-Net (rows 8 and 13 in Fig. 11) have introduced new modules selective kernel and spatial attention, respectively. These innovations aim to enhance the scope of receptive field and capture additional features. Although they address the issue of feature extraction to some extent, they do not lead to significant improvements in segmentation effectiveness. Among models other than DCM-Net, DR-Vnet stands out for its incorporation of dense blocks and residual connections along with a channel compression module. Our DCM-Net integrates the strengths of the aforementioned models while leveraging the DCM module for channel compression to further enhance image feature extraction. Furthermore, our approach incorporates the FAM, which not only utilizes spatial attention but also integrates attention gate and channel attention mechanisms for improved context integration. Additionally, we combine the SE module with residual connections resulting in favorable outcomes. In the final row of Fig. 11, it can be observed that following the application of DCM-Net, the edge intricacies in the blurred images are effectively showcased. For images displaying minimal skin cancer alterations, DCM-Net also aims to enhance them to the greatest extent possible.
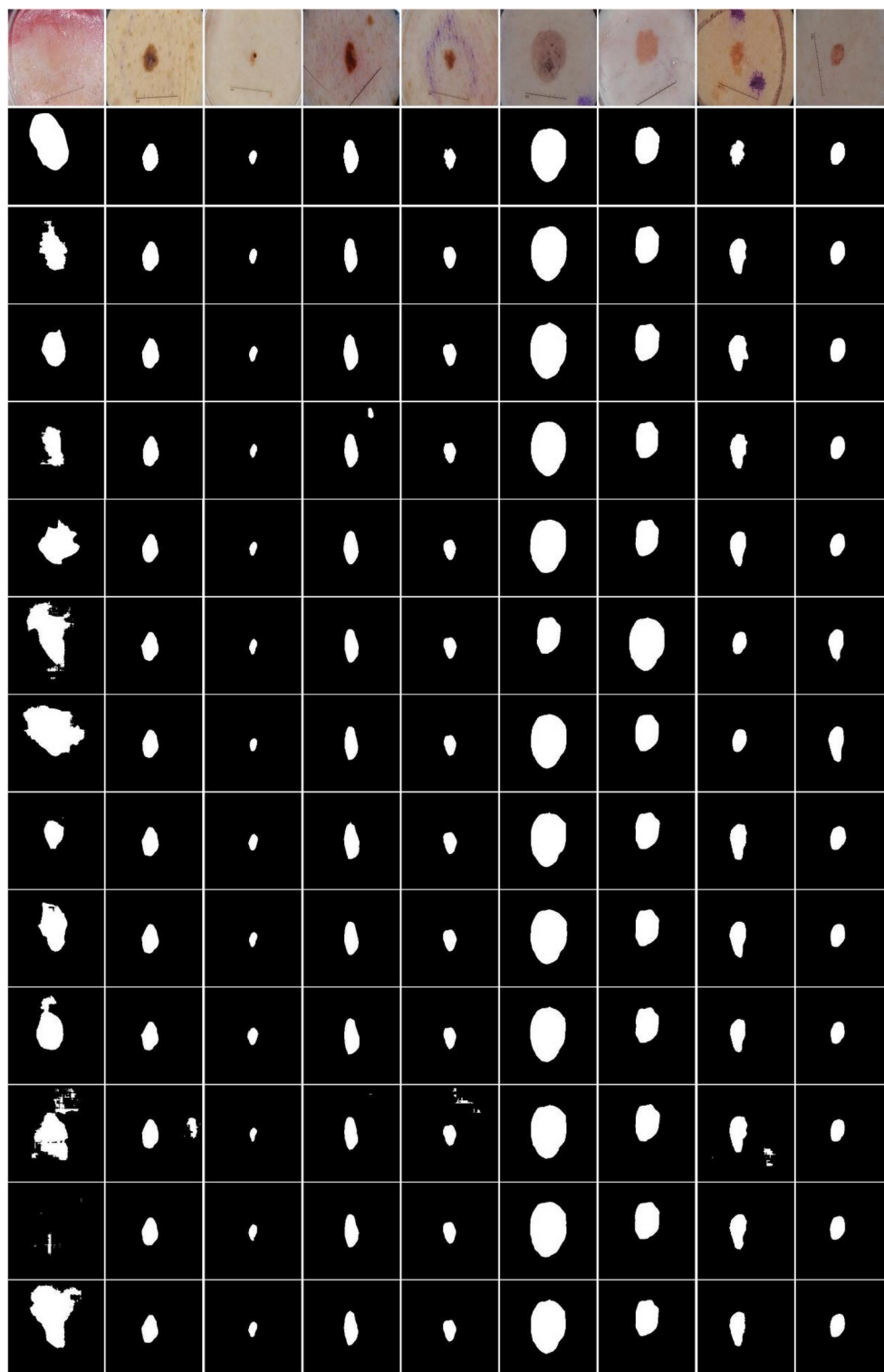
Fig. 11. Visual segmentation results of different methods on ISIC-2018 dataset. First and second rows: original images and their corresponding ground truth. The third to last rows are results of U-Net, OD-Segmentation, CL-Net, EE-Net, Dr-Vnet, SK-U-Net, Double-Net, Nested-Net, Connected-Net, X-Net, SA-Net and DCM-Net.

We additionally performed a comparative analysis using the pituitary adenoma dataset, employing identical evaluation metrics. The corresponding data findings are detailed in Table 3. It can be seen from the data in the table that the output of the model has been adjusted accordingly due to changes in the data set. Although segmenting pituitary tumor images is more challenging than skin cancer images, there are twice as many pictures in the training set for pituitary tumors compared to skin cancer. As a result, all of the models have shown

Method	Accuracy	TPR	Dice	Jaccard
U-Net	0.9688	0.9277	0.8947	0.8109
OD-segmentation	0.9706	0.9301	0.9320	0.8768
CL-Net	0.9697	0.9205	0.9275	0.8691
EE-Net	0.9695	0.9274	0.9287	0.8701
DR-Vnet	0.9616	0.8940	0.8827	0.7929
SK-U-Net	0.9694	0.9189	0.9237	0.8625
Double-Net	0.9698	0.9229	0.8992	0.8237
Nested-Net	0.9699	0.9250	0.9269	0.8675
Connected-Net	0.9696	0.9192	0.9294	0.8715
X-Net	0.9680	0.9109	0.9167	0.8513
SA-Net	0.9685	0.9096	0.9128	0.8454
DCM-Net	0.9707	0.9309	0.9329	0.8773

Table 3. Results of comparative experiments on pituitary adenoma.

improvements in terms of Acc, TPR, Dice, and Jaccard values. Table 3 indicates that other models like OD-Segmentation also demonstrate impressive performance on this dataset. However, when considering all four metrics, DCM-Net continues to outperform others with Accracy, TPR, Dice and Jaccard values reaching 97.07%, 93.09%, 93.29%, and 87.73%. The data indicate that DCM-Net can also process images with more complex backgrounds and get better results. Meanwhile, it also shows from the side that it has stability in medical image segmentation, whether it is skin cancer images or pituitary tumor images, it can achieve the best segmentation effect. We also visualized the results on pituitary adenoma dataset, as shown in Fig. 12, from which it can be seen the excellent performance of DCM-Net.

Computational efficiency

In Table 4, we present a detailed analysis of the relevant parameters and computational efficiency of various segmentation methods applied on the ISIC-18 dataset. Models such as U-Net, DR-Vnet, X-Net, SA-Net, and Connected-Net demonstrate notable advantages due to their reduced parameter requirements. However, this simplicity often comes at the cost of lower accuracy in lesion detection. Conversely, models like OD-Segmentation and CL-Net incorporate residual modules that enhance feature representation, albeit with increased training time and parameter demands. Similarly, EE-Net employs repeated multi-scale fusion to achieve superior performance, which requires significant computational resources. Despite the long training time required, our model stands out by achieving higher detection accuracy. These findings underscore the importance of balancing computational efficiency with the ultimate objective of accurate and reliable segmentation when selecting a model. The slow training speed of our model is mainly due to the dual-channel structure and a variety of complex attention experiments. In future experiments, we will consider using system-on-chip[50] and other indicators to make it superior to other models in terms of efficiency.

Ablation experiment

In this section, we perform ablation studies on the DCM-Net in ISIC-2018 dataset. We integrated DCM, FAM, SE, and residual connections within the U-Net architecture, initially conducting individual experiments to assess their separate contributions to the model. Subsequently, we combine the three in pairs to examine the resultant effects. Once all experiments were conducted, we evaluated using Accracy, TPR, Dice, and Jaccard evaluation metrics. Evaluation results are presented in Table 5, with visualizations depicted in Fig. 13.

Effect of DCM
The first is to add the DCM module to the encoder part in U-Net. From Table 5, it is not difficult to see that after using the DCM module, Accuracy increases by 0.17%, TPR increases by 0.25%, Dice increases by 0.90%, and Jaccard increases by 1.37%. From the visualization results in Fig. 13, it can be obtained that after the addition of DCM module, the model segmentation is more detailed and the accuracy of segmentation is improved. The main reason for this is that by adding DCM to the encoder, the ability of the model to process low-level features is improved, the useless information among them is discarded and the effective information is retained. Furthermore, we changed the element multiplication part of the DCM to the element addition, and the result is shown in Table 6. It can be seen from the results that the performance of element multiplication is better than that of element addition.

Effect of FAM
In the subsequent experiments, FAM is used at the encoder and decoder junction, as shown in Fig. 4, compared with the original U-Net, the segmentation performance of the model has been greatly improved. This is mainly due to the comprehensive use of multiple attention mechanisms in the hybrid attention mechanism, so that the model's attention is focused on effective information. In addition, in row 5 of Fig. 13, we can see that this module is more proficient in the segmentation of small and medium scale skin cancerization compared with the large-scale skin cancerization. In summary, it shows that the adopted hybrid attention mechanism combined with the original U-Net has a certain effect.
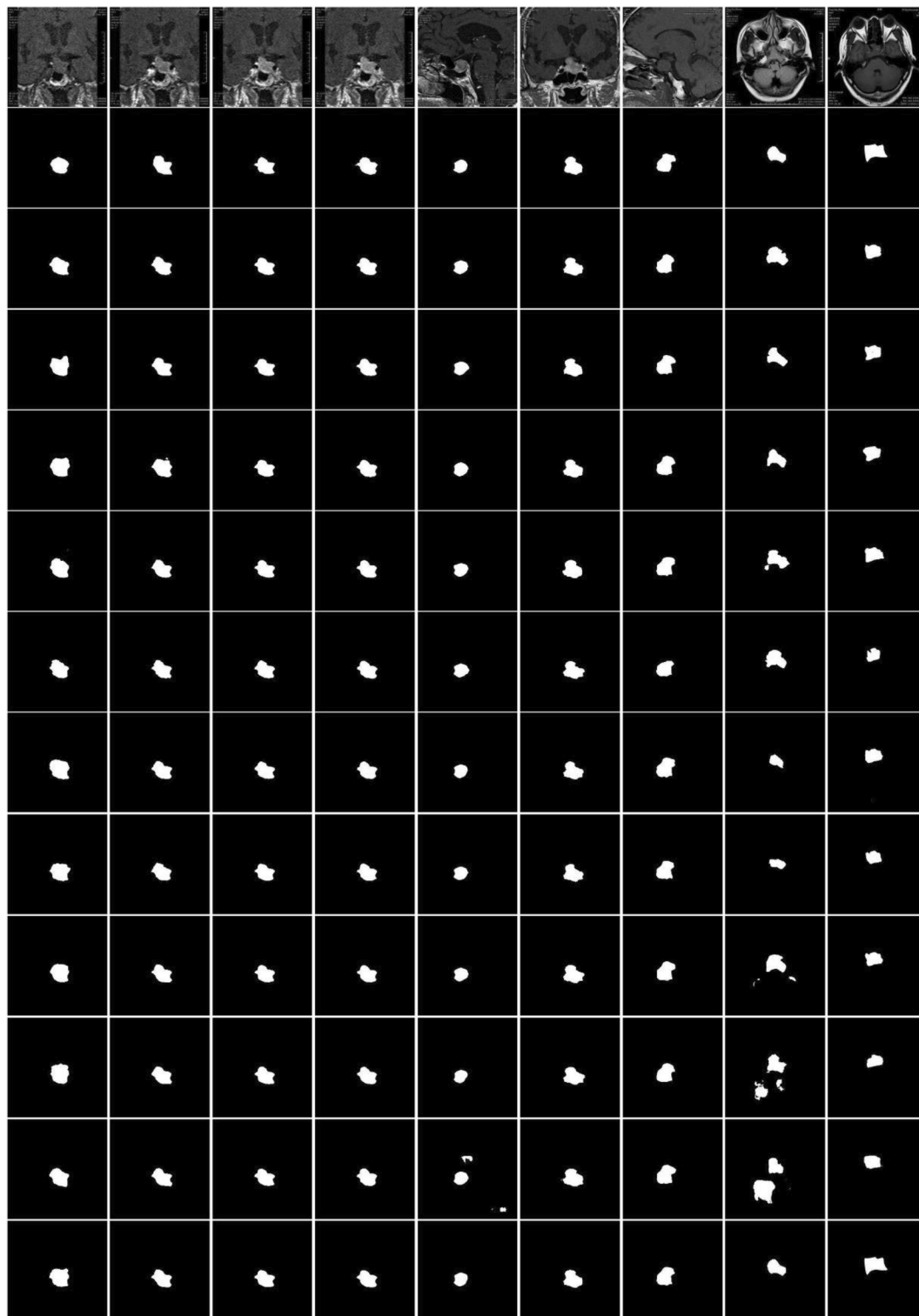
Fig. 12. Visual segmentation results of different methods on pituitary adenoma dataset. First and second rows: original images and their corresponding ground truth. The third to last rows are results of U-Net, OD-Segmentation, CL-Net, EE-Net, Dr-Vnet, SK-U-Net, Double-Net, Nested-Net, Connected-Net, X-Net, SA-Net and DCM-Net.

Effect of SE and residual connection
The last module to perform the experiment is the SE and residual connection. We introduce it into the U-Net framework in the decoder part. The SE in this module can make up for the deficiency of the original framework for information extraction. Furthermore, through the residual connection, more original information and

Method	Parameter (M)	Time (ms)
U-Net	2.06	269
OD-segmentation	19.82	365
CL-Net	7.72	292
EE-Net	29.98	339
DR-Vnet	20.25	592
SK-U-Net	1.85	195
Double-Net	22.28	696
Nested-Net	22.19	359
Connected-Net	3.17	382
X-Net	0.70	382
SA-Net	0.51	198
DCM-Net	2.72	334

Table 4. Parameters and computational efficiency of different methods on the ISIC-18 dataset.

Method	Accuracy	TPR	Dice	Jaccard
U-Net	0.9068	0.8549	0.8278	0.7125
U-Net+DCM	0.9085	0.8574	0.8368	0.7262
U-Net+FAM	0.9105	0.8580	0.8534	0.7503
U-Net+(SE+Res)	0.9081	0.8503	0.8430	0.7376
U-Net+DCM+FAM	0.9111	0.8758	0.8542	0.7510
U-Net+DCM+(SE+Res)	0.9100	0.8841	0.8518	0.7461
U-Net+FAM+(SE+Res)	0.9127	0.8786	0.8587	0.7584
DCM-Net	0.9142	0.8893	0.8619	0.7602

Table 5. Ablation experiments on ISIC-2018.

gradients are retained, and the problem of gradient explosion is also solved. From the sixth row of Fig. 13 and Table 5, the above conclusions are further confirmed; all the indicators have increased after the addition of this module, indicating that it has a certain role in improving the segmentation ability of the model.

Conclusion

In the field of medical image segmentation, identifying image edges has always been an important factor limiting the segmentation effect. To solve this problem, we proposed an innovative dual-channel compressive mapping network aimed at effectively overcoming the challenges of medical image segmentation. The network consists of three key parts: DCM, FAM, and SE. Firstly, DCM can efficiently compress information in both channels and enhance information processing capabilities. Then, FAM enables the model to focus on effective information, helping to improve segmentation accuracy. Finally, with the help of SE and residual connection, the decoder obtains successfully compressed and activated high-level semantic features. After a series of experiments and evaluations on the medical image datasets, our DCM-Net achieved significant improvements in segmentation accuracy and computational efficiency. The experimental results show that the method has good application prospects in the field of medical image segmentation. Meanwhile, we are keenly aware of the limitations current models face in computational efficiency, which seriously impact their usefulness in real-world clinical applications. In future work, we aim to optimize computational efficiency through techniques such as model compression and lightweight architecture design.
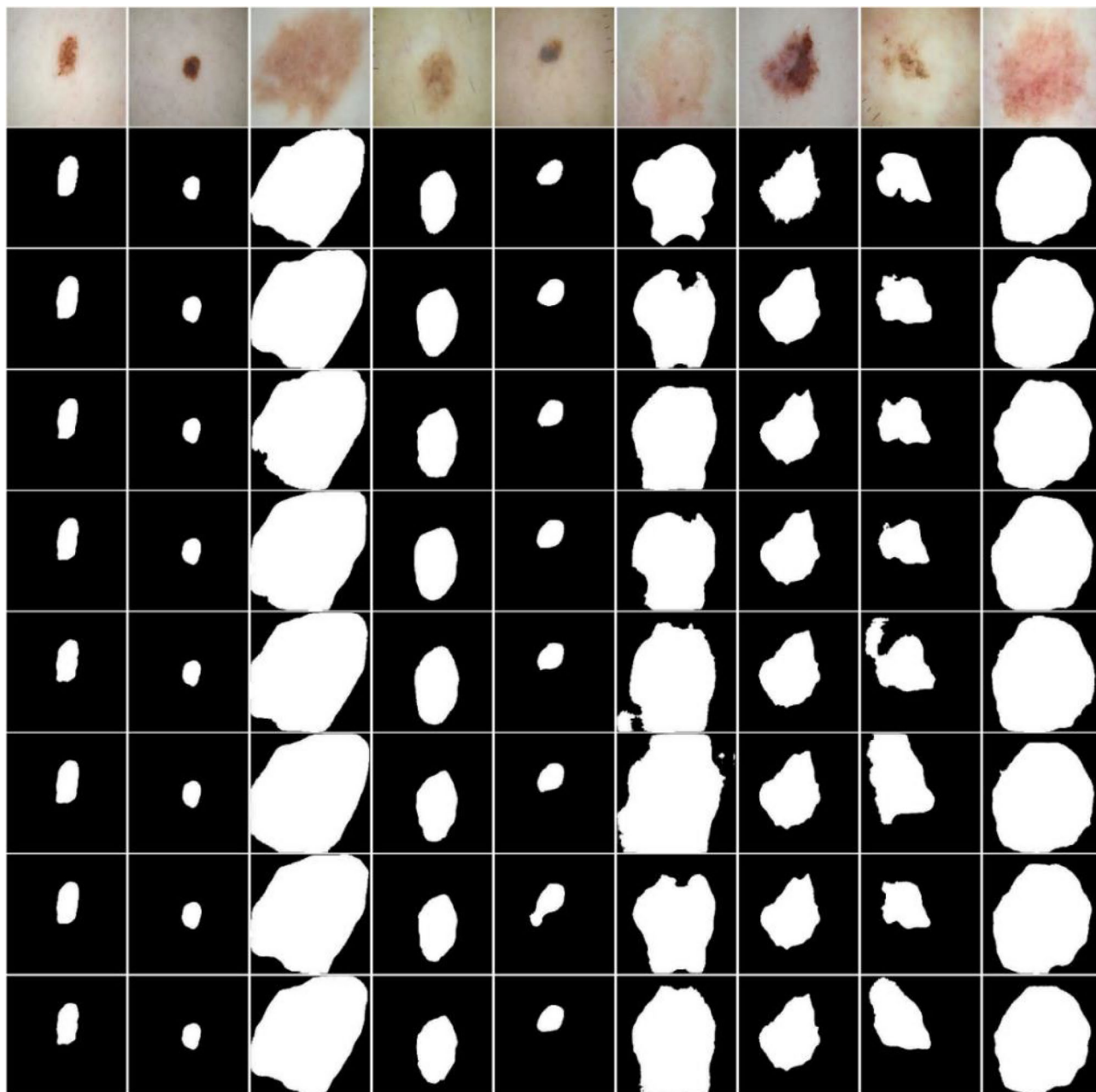
Fig. 13. Visual segmentation results of different methods on ISIC-2018 dataset. First and second rows: original images and their corresponding ground truth. The third to last rows are results of U-Net, U-Net+DCM, U-Net+FAM, U-Net+(SE+Res), U-Net+DCM+FAM, U-Net+DCM+(SE+Res), U-Net+FAM+(SE+Res) and DCM-Net.

Method	Accuracy	TPR	Dice	Jaccard
Element multiplication	0.9085	0.8574	0.8368	0.7262
Element addition	0.8993	0.8838	0.8305	0.7165

Table 6. Multiplication and addition experiments on ISIC-2018.

Data availability

The ISIC-2018 datasets generated and/or analysed during the current study are available in the ISIC repository, https://challenge.isic-archive.com/data/#2018. The pituitary adenoma datasets generated and/or analyzed

during the current study are not publicly available due to the laboratory policy but are available from the corresponding author on reasonable request.

References

1. Saeed, T. et al. Neuro-xai: Explainable deep learning framework based on deeplabv3+ and Bayesian optimization for segmentation and classification of brain tumor in MRI scans. *J. Neurosci. Methods* **410**, 110247 (2024).
2. Nazir, K. et al. 3d kronecker convolutional feature pyramid for brain tumor semantic segmentation in MR imaging. *Comput. Mater. Contin.* **76**, 2861–2877 (2023).
3. Shelhamer, E., Long, J. & Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 640–651 (2017).
4. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18* 234–241 (2015).
5. Rajamani, K. T., Rani, P., Siebert, H., ElagiriRamalingam, R. & Heinrich, M. P. Attention-augmented u-net (aa-u-net) for semantic segmentation. *SIViP* **17**, 981–989 (2023).
6. Khan, W. R. et al. A hybrid attention-based residual Unet for semantic segmentation of brain tumor. *Comput. Mater. Contin.* **76**, 647–664 (2023).
7. Yin, H. et al. Dfbu-net: Double-branch flat bottom u-net for efficient medical image segmentation. *Biomed. Signal Process. Control* **90**, 105818 (2024).
8. Ahmed, M. R. et al. Doubleu-netplus: A novel attention and context-guided dual u-net with multi-scale residual feature fusion network for semantic segmentation of medical images. *Neural Comput. Appl.* **35**, 14379–14401 (2023).
9. Ansari, M. Y. et al. A lightweight neural network with multiscale feature enhancement for liver CT segmentation. *Sci. Rep.* **12**, 14153 (2022).
10. Cao, X. et al. Demf-net: A dual encoder multi-scale feature fusion network for polyp segmentation. *Biomed. Signal Process. Control* **96**, 106487 (2024).
11. Nawaz, M. et al. Mseg-net: A melanoma mole segmentation network using cornernet and fuzzy k-means clustering. *Comput. Math. Methods Med.* **2022**, 7502504 (2022).
12. Sun, G. et al. Da-transunet: Integrating spatial and channel dual attention with transformer u-net for medical image segmentation. *Front. Bioeng. Biotechnol.* **12**, 1398237 (2024).
13. Zhang, N. et al. Ct-net: Asymmetric compound branch transformer for medical image segmentation. *Neural Netw.* **170**, 298–311 (2024).
14. Zhao, Y., Li, J., Ren, L. & Chen, Z. Dtan: Diffusion-based text attention network for medical image segmentation. *Comput. Biol. Med.* **168**, 107728 (2024).
15. Zhang, J., Luan, Z., Ni, L., Qi, L. & Gong, X. Msdanet: A multi-scale dilation attention network for medical image segmentation. *Biomed. Signal Process. Control* **90**, 105889 (2024).
16. Ansari, M. Y., Yang, Y., Meher, P. K. & Dakua, S. P. Dense-psp-unet: A neural network for fast inference liver ultrasound segmentation. *Comput. Biol. Med.* **153**, 106478 (2023).
17. Mohanty, S. & Dakua, S. P. Toward computing cross-modality symmetric non-rigid medical image registration. *IEEE Access* **10**, 24528–24539 (2022).
18. Ma, Z. & Li, X. An improved supervised and attention mechanism-based u-net algorithm for retinal vessel segmentation. *Comput. Biol. Med.* **168**, 107770 (2024).
19. Feng, Y., Zhu, X., Zhang, X., Li, Y. & Lu, H. Pamsnet: A medical image segmentation network based on spatial pyramid and attention mechanism. *Biomed. Signal Process. Control* **94**, 106285 (2024).
20. Chen, J. et al. Attention gate and dilation u-shaped network (gdunet): An efficient breast ultrasound image segmentation network with multiscale information extraction. *Quant. Imaging Med. Surg.* **14**, 2034–2048 (2024).
21. Hussain, T. & Shouno, H. Magres-unet: Improved medical image segmentation through a deep learning paradigm of multi-attention gated residual u-net. *IEEE Access* **12**, 40290–40310 (2024).
22. Shen, N. et al. Multi-organ segmentation network for abdominal CT images based on spatial attention and deformable convolution. *Expert Syst. Appl.* **211**, 118625 (2023).
23. Fu, Z., Li, J. & Hua, Z. Msa-net: Multiscale spatial attention network for medical image segmentation. *Alex. Eng. J.* **70**, 453–473 (2023).
24. Wang, C. et al. Cfatransunet: Channel-wise cross fusion attention and transformer for 2d medical image segmentation. *Comput. Biol. Med.* **168**, 107803 (2024).
25. Wu, Y., Wang, G., Wang, Z., Wang, H. & Li, Y. Triplet attention fusion module: A concise and efficient channel attention module for medical image segmentation. *Biomed. Signal Process. Control* **82**, 104515 (2023).
26. Liu, X., Hu, Y. & Chen, J. Hybrid CNN-transformer model for medical image segmentation with pyramid convolution and multi-layer perceptron. *Biomed. Signal Process. Control* **86**, 105331 (2023).
27. Zhang, L., Xu, C., Li, Y., Liu, T. & Sun, J. Mcse-u-net: Multi-convolution blocks and squeeze and excitation blocks for vessel segmentation. *Quant. Imaging Med. Surg.* **14**, 2426–2440 (2024).
28. Jiang, S., Chen, X. & Yi, C. Ssa-unet: Whole brain segmentation by u-net with squeeze-and-excitation block and self-attention block from the 2.5 d slice image. *IET Image Process.* **18**, 1598–1612 (2024).
29. Mehrtash, A., Wells, W. M., Tempany, C. M., Abolmaesumi, P. & Kapur, T. Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE Trans. Med. Imaging* **39**, 3868–3878 (2020).
30. Li, Y.-Z. et al. Rsu-net: U-net based on residual and self-attention mechanism in the segmentation of cardiac magnetic resonance images. *Comput. Methods Programs Biomed.* **231**, 107437 (2023).
31. Ansari, M. Y., Chandrasekar, V., Singh, A. V. & Dakua, S. P. Re-routing drugs to blood brain barrier: A comprehensive analysis of machine learning approaches with fingerprint amalgamation and data balancing. *IEEE Access* **11**, 9890–9906 (2022).
32. Diniz, J. O. B., Ferreira, J. L., Diniz, P. H. B., Silva, A. C. & Paiva, A. C. A deep learning method with residual blocks for automatic spinal cord segmentation in planning CT. *Biomed. Signal Process. Control* **71**, 103074 (2022).
33. Fan, X., Zhou, J., Jiang, X., Xin, M. & Hou, L. Csap-unet: Convolution and self-attention paralleling network for medical image segmentation with edge enhancement. *Comput. Biol. Med.* **172**, 108265 (2024).
34. Li, J., Gao, G., Yang, L. & Liu, Y. A retinal vessel segmentation network with multiple-dimension attention and adaptive feature fusion. *Comput. Biol. Med.* **172**, 108315 (2024).
35. Li, J. et al. Class-aware attention network for infectious keratitis diagnosis using corneal photographs. *Comput. Biol. Med.* **151**, 106301 (2022).
36. Yang, Y., Feng, C. & Wang, R. Automatic segmentation model combining u-net and level set method for medical images. *Expert Syst. Appl.* **153**, 113419 (2020).

37. Selvaraj, A. & Nithiyaraj, E. Cedrnn: A convolutional encoder-decoder residual neural network for liver tumour segmentation. *Neural Process. Lett.* **55**, 1605–1624 (2023).
38. Abdel-Nabi, H., Ali, M. Z. & Awajan, A. A multi-scale 3-stacked-layer coned u-net framework for tumor segmentatio in whole slide images. *Biomed. Signal Process. Control* **86**, 105273 (2023).
39. Hu, K. et al. Dsc-net: A novel interactive two-stream network by combining transformer and CNN for ultrasound image segmentation. *IEEE Trans. Instrum. Meas.* **72**, 3322993 (2023).
40. Wang, L. et al. Automated segmentation of the optic disc from fundus images using an asymmetric deep learning network. *Pattern Recogn.* **112**, 107810 (2021).
41. Khan, N., Haq, I. U., Ullah, F. U. M., Khan, S. U. & Lee, M. Y. Cl-net: Convlstm-based hybrid architecture for batteries' state of health and power consumption forecasting. *Mathematics* **9**, 3326 (2021).
42. Wang, L. et al. Ee-net: An edge-enhanced deep learning network for jointly identifying corneal micro-layers from optical coherence tomography. *Biomed. Signal Process. Control* **71**, 103213 (2022).
43. Karaali, A., Dahyot, R. & Sexton, D. J. Dr-vnet: Retinal vessel segmentation via dense residual unet, in *International Conference on Pattern Recognition and Artificial Intelligence* 198–210 (2022).
44. Byra, M. et al. Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network. *Biomed. Signal Process. Control* **61**, 102027 (2020).
45. Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P. & Johansen, H. D. Doubleu-net: A deep convolutional neural network for medical image segmentation, in *2020 IEEE 33rd International Symposium on Computer-based Medical Systems (CBMS)* 558–564 (2020).
46. Gouizi, F. & Megherbi, A. C. Nested-net: A deep nested network for background subtraction. *Int. J. Multimed. Inf. Retr.* **12**, 5 (2023).
47. Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C. & Elmaghraby, A. S. Connected-unets: A deep learning architecture for breast mass segmentation. *NPJ Breast Cancer* **7**, 151 (2021).
48. Bullock, J., Cuesta-Lázaro, C. & Quera-Bofarull, A. Xnet: A convolutional neural network (CNN) implementation for medical X-ray image segmentation suitable for small datasets, in *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, Vol. 10953, 453–463 (2019).
49. Zhang, Q.-L. & Yang, Y.-B. Sa-net: Shuffle attention for deep convolutional neural networks, in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2235–2239 (2021).
50. Zhai, X. et al. Real-time automated image segmentation technique for cerebral aneurysm on reconfigurable system-on-chip. *J. Comput. Sci.* **27**, 35–45 (2018).

Author contributions

Conceptualization, methodology, and writing by X.D. and Q.Z.; validation and experiments by K.Q. and L.D.; writing review and editing by X.J. All authors reviewed the manuscript.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.Q. or Q.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.