EMF-former: An Efficient and Memory-Friendly Transformer for Medical Image Segmentation

Zhaoquan Hao, Hongyan Quan(✉), and Yinbin Lu

School of Computer Science and Technology, East China Normal University,
Shanghai, China
hyquan@cs.ecnu.edu.cn

Abstract. Medical image segmentation is of significant importance for computer-aided diagnosis. In this task, methods based on Convolutional Neural Networks (CNNs) have shown good performance in extracting local features. However, they cannot capture global dependencies, which is crucial for medical image. On the other hand, Transformer-based methods can establish global dependencies through self-attention, providing a supplement to local convolution. However, the expensive matrix multiplication in the self-attention of a vanilla transformer and the memory usage is still a bottleneck. In this work, we propose a segmentation model named EMF-former. By combining DWConv, channel shuffle and PW-Conv, we design a Depthwise Separable Shuffled Convolution Module (DSPConv) to reduce the parameter count of convolutions. Additionally, we employ an efficient Vector Aggregation Attention (VAA) that substitutes key-value interactions with element-wise multiplication after broadcasting two vectors to reduce computational complexity. Moreover, we substitute the parallel multi-head attention module with the Serial Multi-Head Attention Module (S-MHA) to reduce feature redundancy and memory usage in multi-head attention. Combining the above modules, EMF-former could perform the medical image segmentation efficiently with fewer parameter counts, lower computational complexity and lower memory usage while preserving segmentation accuracy. We conduct experimental evaluations on ACDC and Hippocampus dataset, achieving mIOU values of 80.5% and 78.8%, respectively.

Keywords: Transformer · Light-weight · Medical Image Segmentation

1 Introduction

As one of the fundamental yet crucial research directions in medical image analysis, medical image segmentation aims to classify the pixels of a given medical image into regions, organs, or lesions using algorithms. The segmentation results not only enable the detection of abnormalities in human body regions but also serve as guidance for clinical practitioners.

Automated medical image segmentation can serve as an excellent assistant diagnostic tool for medical experts. Existing medical image segmentation algorithms can be categorized into two types: methods based on Convolutional Neural Networks (CNNs) and based on Transformer networks.

As one of the mainstream approaches in computer vision, Convolutional Neural Networks (CNNs) have been widely adopted in the field of medical image segmentation. Most segmentation methods utilizing CNNs in medical image segmentation are based on improvements of UNet [21] or its variants [29]. Following the success of U-Net, numerous variants based on the U-Net architecture have been developed, including Unet++[29], MHUNet [1].

With the rapid progress of vision tasks [27], researchers have started to focus on lightweight segmentation models. Additionally, models like DC-Unet [18] and EGE-UNet [22] have also proposed the lightweight methods in medical image, specifically in polyp segmentation and skin lesion segmentation. However, due to the focus of convolutional kernels on local regions, CNNs are unable to model global dependencies, resulting in suboptimal performance sometimes.

In recent years, with the emergence of Transformer, there have been successful attempts to apply them to computer vision tasks. Vision Transformer (ViT) [7] pioneered the use of Transformer encoders for image classification. Furthermore, ViT havs also been applied to the field of medical image segmentation, such as TransUNet [5], UNETR [8] and Swin-Unet [4] which have achieved high-quality organ segmentation results. Similarly, Segformer [25], nnformer [28] and CiT-Net [11], efficiently achieve better performance in the field of medical image segmentation. However, due to the reliance of self-attention in ViT, they lead to significant computational costs and memory usage that can not be ignored.

Therefore, how to achieve lightweight Transformer to improve the segmentation efficiency is also an important research direction. The methods [20], [26], [27] propose token sparsification, which reduces the computational cost and number of parameters of matrix multiplication. CCNet [9] reduces the computational complexity by designing an attention module. When these lightweight methods are applied to medical image segmentation [15], [13], they also show good results and accomplish segmentation tasks such as organ and blood vessel segmentation.

However, currently, there is limited research on designing structures specifically for multi-head attention. They still employ parallel multi-head attention, which fails to address the feature redundancy among different heads [12], [10] and the memory usage caused by parallel attention calculation.

In this work, we have further improved a convolutional operation and introduced a more efficient attention that reduced the expensive computational cost and memory usage associated traditional Multi-Head attention.

In contrast to the existing approaches, our method provides the following main contributions: (1) We propose a DSPConv module, which effectively reduces the number of parameters. The DSPConv module consists of DWConv, channel shuffle and PWConv. And the convolution operation is performed on few channels of the feature map. (2) We introduce an efficient attention named Vector Aggregation Attention (VAA) that reduces the complexity of attention computation. (3) The Serial Multi-Head Attention Module is proposed, to reduce memory usage by calculating attention serially. Meanwhile, in the process of calculation, some heads will be ignored, aiming to reduce computational redundancy among different heads. (4) Finally, by combining these modules, we

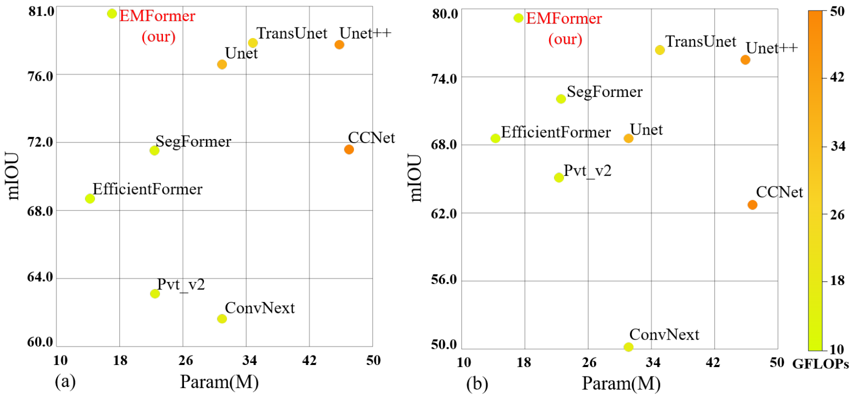

Fig. 1. (a) and (b) respectively show the visualization of experimental results on ACDC and Hippocampus datasets. The X-axis represents the number of parameters, while the Y-axis represents mIoU. The color depth represents computational complexity.

created EMF-former. The experiments results demonstrated improvements in overall metrics, validating the effectiveness of EMF-former.

2 Method

The visual comparison of some metrics is shown in Fig 1. The overall architecture of EMF-former is shown in Fig 2(a).

2.1 DSPConv module

In this work, we propose a convolutional module named DSPConv by combining DWConv, channel shuffle, and PWConv. This convolutional module aims to reduce the number of parameters in convolutional operations and feature redundancy while ensuring accurate feature extraction. Inspired by the findings of Chen et al. [6], we apply the method from the field of nature images to the field of medical images and try to improve the approach, which selectively applies DWConv only to a subset of channels in the feature maps. Compared to regular Conv, DSPConv has fewer parameters, and can ensure information exchange among different channels through channel shuffle and PWConv.

Specifically, the proposed DSPConv module is shown in Fig 2(b3). In the process of getting the output $O \in \mathbb{R}^{H\times W\times C_1}$, firstly, our DSPConv module uses DWConv with a kernel size of K on first $1/4$ of channel C_1. Additionally, to ensure information interaction among different channels, we employ a channel shuffle, shuffling the channels after the DWConv convolution operation and performing PWConv on the remaining $3/4$ of channel C_1 to facilitate information exchange. Therefore, the number of parameters required for our DSPConv is

$$K \times K \times \tfrac{1}{4}C_1 + \tfrac{3}{4}C_1 \approx K \times K \times \tfrac{1}{4}C_1, \tag{1}$$

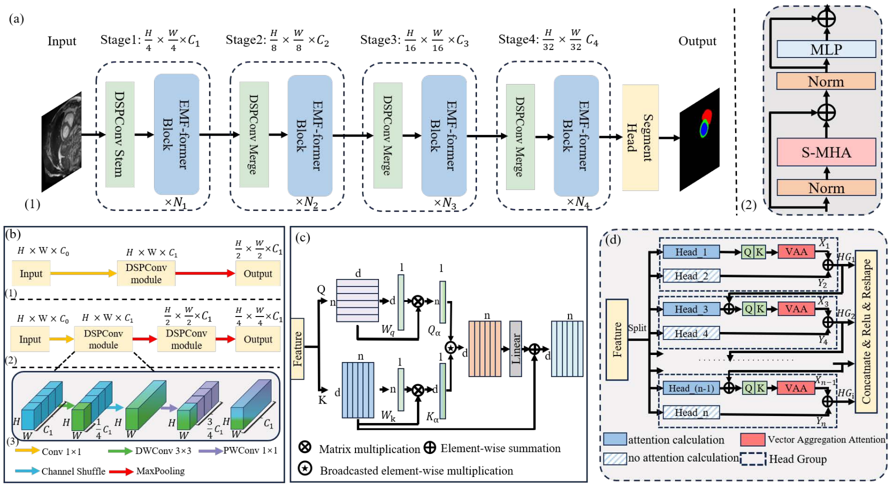

Fig. 2. (a1) The overall architecture of our EMF-former. (a2) Details of a EMF-former Block. (b) Overview of DSPConv Modules. (c) Overview of Vector Aggregation Attention(VAA). (d) Overview of Serial Multi-Head Attention Module(S-MHA).

which is lower than the regular Conv with $K \times K \times (C_1)^2$. And the FLOPs are

$$H \times W \times K^2 \times \tfrac{1}{4}C_1 + H \times W \times (\tfrac{3}{4}C_1)^2 \approx H \times W \times (\tfrac{3}{4}C_1)^2, \qquad (2)$$

which is fewer than regular Conv with $H \times W \times K^2 \times (C_1)^2$. Moreover, as the input channel size c increases, the difference in the total count becomes even larger. By implementing this approach, we can reduce feature redundancy while achieving lightweight results.

The DSPConv module used in the DSPConv Stem is shown in Fig 2(b2). With this module, the feature map can be downsampled 4 times. The other is the module used in the DSPConv Merge as shown in Fig 2(b1), where the feature map can be downsampled 2 times.

2.2 Vector Aggregation Attention

For attention computation, Q, K, V$\in \mathbb{R}^{N\times d}$ denote the query, key, and value matrices, respectively(N = H × W, where H and W are the height and width of the feature map). An attention function transforms each query as a weighted sum of values, which is then multiplied by the V matrix to obtain the attention score. This process requires matrix multiplication among Q, K, V, which all have dimension $\mathbb{R}^{N\times d}$, and results in the complexity of $O\left(N^2 d\right)$, as:

$$Attn(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{C}}\right)V, \qquad (3)$$

To reduce the computational complexity, in this work, we introduce an efficient Vector Aggregation Attention(VAA) inspired by the approach proposed by Shaker et al. [23] and Lin et al. [16], as shown in Fig 2(c). We replace the matrix multiplication of Q and K with element-wise multiplication of two broadcasted vectors. Additionally, we replace the key-value interaction with a fully connected layer. This method not only enables the computation of global attention but also reduces computational complexity compared to self-attention.

Specifically, after generating Q and K using a linear layer within a single head, where Q and K are both $\in \mathbb{R}^{N \times d}$, with N representing the number of tokens and d can be interpreted as the length of each token. We utilize two learnable vectors, $W_q \in \mathbb{R}^{d \times 1}$ and $W_k \in \mathbb{R}^{1 \times N}$, to multiply with Q and K. This generates two global attention vectors, $Q_\alpha \in \mathbb{R}^{N \times 1}$ and $K_\alpha \in \mathbb{R}^{1 \times d}$, as:

$$\begin{aligned} Q_\alpha &= QW_q, \\ K_\alpha &= W_k K, \end{aligned} \tag{4}$$

where, Q_α can be understood as aggregating the features of all dimensions of each token, while K_α can be understood as aggregating all tokens into a single token. Subsequently, we perform a broadcasting operation to obtain two matrices with the same dimensions $\mathbb{R}^{N \times d}$. These matrices are then multiplied element-wise and through the Linear layer to compute the global attention, as:

$$VAA_{Attn}(Q, K) = L\left(\frac{Q_\alpha K_\alpha}{\sqrt{C}}\right) + K. \tag{5}$$

Therefore, our proposed VAA avoids directly performing matrix multiplication on the Q matrix and K matrix and reduces the computational complexity to $O\left(N\right)$. And L is the Linear layer which replaces the V matrix.

2.3 Serial Multi-Head Attention Module

Meanwhile, the multi-head attention has feature redundancy among different heads [12], [10]. It leads to the fact that the multi-head attention not only occupies any memory and computational resources, but many of its components are used to extract redundant global features, which makes the overall efficiency limited.

To address the problem, we designed the Serial Multi-Head attention module (S-MHA), as shown in Fig 2(d), and each of the two different heads are combined to form a Head Group. Then the computation results of different Head Groups are connected to the next Head Group for summation, and the attention computation is performed again. In addition, we try to introduce the work by Chen et al. [6] into Transformer. Specifically, instead of performing the attention computation for the second head in each Head Group, we directly sum the results of the computation with the first head.

The reason is that we hypothesize that since there is feature redundancy in multi-head attention, we can refrain from performing attention calculations on some heads. Subsequent experiments prove that our conception is correct. We set that the multi-head attention has a maximum of 8 heads (n=8), formally, this attention can be formulated as:

$$\begin{cases} X_i = VAA_{Attn}(Q,K)_{Head_i}, & i \in \{1,3,5,7\}\,; \\ Y_i = Head_i, & i \in \{2,4,6,8\}\,; \\ HG_i = X_{2i-1} + Y_{2i}, & i \in \{1,2,3,4\}\,, \end{cases} \tag{6}$$

where X_i denotes the attention output of the i-th head (i is odd). HG_i denotes the output of the i-th Head Group. It is worth noting that since no attention calculation is performed on the $Head_i$ (i is even), the output can be considered as the $Head_i$ itself, so Y_i denotes the output of the i-th head (i is even).

To reduce redundancy among heads and encourage the Q, K layers to learn projections on features with richer information. We add the output of the Head Group to the subsequent head:

$$Head_{2i+1} = Head_{2i+1} + HG_i, i \in \{1,2,3\}\,, \tag{7}$$

$Head_{2i+1}$ will be used as a new input feature for the ($2i+1$)-th head. Eventually, we concatenate the output of the Head Groups to get the output:

$$SMHA = Concat\,[HG_1, HG_2, HG_3, HG_4]\,. \tag{8}$$

Overall, with the above operations, the memory and computational resources can be saved because the Serial Multi-Head Attention Module does not need to be computed across multi-head at the same time. It is also possible to learn richer features among different heads to improve model performance.

3 Experiment

3.1 Dataset and Implementation Details

Automated Cardiac Diagnostic Challenge (ACDC) [3] is a dataset for automated cardiac diagnostics that contains a total of 100 patients. It related to three organs, left ventricle (LV), right ventricle (RV) and myocardium (MYO). We split the dataset to 70 training samples, 10 validation samples and 20 test samples.

The MSD Hippocampus dataset (Hippocampus) [2] is one of the segmentation tasks in the MSD (Medical Segmentation Decathlon), where the goal is to segment the hippocampus of the brain from a 2D MRI image. We extracted 50 training samples, 20 test samples and 10 validation samples from the dataset.

We resized the images and labels to a size of 256 × 256, used a batch size of 4, initialized the learning rate to 0.0015, employed the SGD optimizer, and trained the network for 200 epochs with cross-entropy loss. Our EMF-former parameters are set as follows: dims is {64, 128, 256, 512}, num_head is {1, 2, 4, 8}, depths is {3, 4, 6, 3}. These values correspond to four stages of EMF-former. Our framework was implemented in Pytorch and all experiments were performed on NVIDIA GeForce RTX 2080Ti GPUs. We use the following common performance evaluation metrics, including pixel accuracy (Acc), mean IoU (mIOU, the average of the intersection and merger ratios).

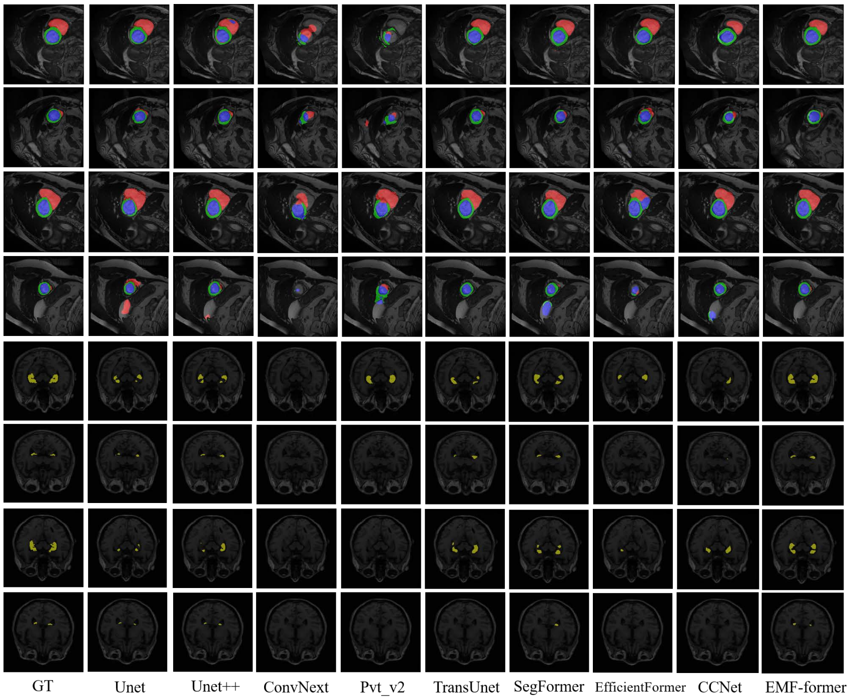

Fig. 3. The visual comparison on ACDC and Hippocampus dataset.

3.2 Comparisons With Other Methods

In order to demonstrate the effectiveness of our proposed EMF-former, we conducted comparative tests based on the ACDC dataset and the Hippocampus dataset on ConNext [17], Unet [21], Unet++ [29], TransUnet [5], Segformer [25], Pvt_v2 [24], ccnet [9], and EfficientFormer [14], respectively.

The quantitative results of the segmentation task on the ACDC dataset are shown in Table 1, and the visual comparison are illustrated in Fig 3. Our method achieves a balance between model complexity and accuracy. Specifically, our EMF-former model achieves an mIOU of 80.5% and an ACC of 87.77%.

The results on the Hippocampus dataset are shown in Table 1, and the visual comparison results are illustrated in Fig 3. Similarly, our method achieves a balance between model accuracy and complexity, with an mIOU of 78.8% and an ACC of 82.75%. Moreover, we observe that our method can guarantee accurate segmentation of small targets, comparing to other methods.

3.3 Ablation Studies

To thoroughly demonstrate the effectiveness of different modules in our model, we conducted a series of ablation experiments on the ACDC dataset.

Table 1. Quantitative comparison with previous methods on ACDC dataset and Hippocampus datasets. Blue indicates the best result, and red indicates the second-best.

Method	**ACDC**			**Hippocampus**		**Mem(M)**
	Param(M)	**ACC(%)**	**mIOU(%)**	**ACC(%)**	**mIOU(%)**	
ConvNext	31.26	70.97	61.6	50.00	49.9	344.7
Unet	31.03	76.37	74.6	71.45	68.4	547.7
Unet++	45.48	85.92	75.4	78.05	73.6	2446.7
TransUnet	35.31	86.15	77.9	79.21	72.0	455.9
Pvt_v2	22.97	73.95	63.1	68.99	65.2	361.8
SegFormer	23.34	80.42	71.6	75.65	74.4	356.0
EfficientFormer	14.37	76.85	68.4	64.65	62.9	226.2
CCNet	47.42	80.65	71.3	73.77	68.5	787.2
EMF-former	17.34	87.77	80.5	82.75	78.8	324.8

Table 2. Ablation experiments of DSPConv, VAA and S-MHA in EMF-former in ACDC dataset. The SHU refers to the convolutional modules proposed in ShuffleNet, and the SWI refers to the additive attention proposed in SwiftFormer.

Backbone	**DSPConv**	**VAA**	**S-MHA**	**Param(M)**	**mIOU(%)**
Segformer				23.34	71.6
Segformer	✓			21.17	73.2
Segformer		✓		19.94	74.2
Segformer		✓	✓	18.19	76.6
Segformer	✓	✓		18.75	75.7
Segformer	SHU	✓	✓	17.04	77.2
Segformer	✓	SWI	✓	17.01	78.3
EMF-former	✓	✓	✓	17.34	80.5

As shown in Table 2, we can observe that the proposed DSPConv module, Vector Aggregation Attention and Serial Multi-Head Attention Module demonstrate excellent performance. Combining these three modules allows EMF-former to achieve the best medical image segmentation results. Furthermore, when we replace the DSPConv convolution operation with the convolution structure proposed in ShuffleNet [19], which is also a lightweight model, the mIOU value decreases. And we replace VAA with additive attention proposed in Swiftformer [23], which has a similar attention calculation to our VAA, the mIOU value decreases. This further confirms the effectiveness of the modules in EMF-former.

4 Conclusion

In this work, we propose a segmentation model based on Transformer, named EMF-former. We utilize the DSPConv module to significantly reduce the parameter count. Then we introduce an effective Vector Aggregation Attention that replaces expensive matrix multiplication operations with element-wise interactions between two broadcasted vectors. We also replace key-value interactions with Linear layers. Additionally, we proposed the Serial Multi-Head Attention Module, which enables the efficient utilization of computational and spatial resources. Experimental results demonstrate that EMF-former ensures segmentation accuracy on several 2D medical image while achieving a lightweight effect.

In future work, we are interested in designing novel segmentation heads to further reduce model size, enhance the model capacity and efficiency.

Disclosure of Interests. We have no competing interests relevant to the content of this article.

References

1. Ahmad, P., Jin, H., Alroobaea, R., Qamar, S., Zheng, R., Alnajjar, F., Aboudi, F.: Mh unet: A multi-scale hierarchical based architecture for medical image segmentation. IEEE Access **9**, 148384–148408 (2021)
2. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. Nature communications **13**(1), 4128 (2022)
3. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE transactions on medical imaging **37**(11), 2514–2525 (2018)
4. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
5. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
6. Chen, J., Kao, S.h., He, H., Zhuo, W., Wen, S., Lee, C.H., Chan, S.H.G.: Run, don't walk: Chasing higher flops for faster neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12021–12031 (2023)
7. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
8. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
9. Huang, Z., Wang, X., Wei, Y., Huang, L., Shi, H., Liu, W., Huang, T.S.: Ccnet: Criss-cross attention for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(6), 6896–6908 (2023)
10. Jiao, J., Tang, Y.M., Lin, K.Y., Gao, Y., Ma, J., Wang, Y., Zheng, W.S.: Dilateformer: Multi-scale dilated transformer for visual recognition. IEEE Transactions on Multimedia (2023)
11. Lei, T., Sun, R., Wang, X., Wang, Y., He, X., Nandi, A.: Cit-net: Convolutional neural networks hand in hand with vision transformers for medical image segmentation. arXiv preprint arXiv:2306.03373 (2023)
12. Li, J., Tu, Z., Yang, B., Lyu, M.R., Zhang, T.: Multi-head attention with disagreement regularization. arXiv preprint arXiv:1810.10183 (2018)
13. Li, X., Jiang, Y., Li, M., Yin, S.: Lightweight attention convolutional neural network for retinal vessel image segmentation. IEEE Transactions on Industrial Informatics **17**(3), 1958–1967 (2020)
14. Li, Y., Yuan, G., Wen, Y., Hu, J., Evangelidis, G., Tulyakov, S., Wang, Y., Ren, J.: Efficientformer: Vision transformers at mobilenet speed. Advances in Neural Information Processing Systems **35**, 12934–12949 (2022)

15. Lin, X., Yu, L., Cheng, K.T., Yan, Z.: Batformer: Towards boundary-aware lightweight transformer for efficient medical image segmentation. IEEE Journal of Biomedical and Health Informatics (2023)
16. Lin, Y., Fang, X., Zhang, D., Cheng, K., Chen, H.: Boosting convolution with efficient mlp-permutation for volumetric medical image segmentation (2023)
17. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11976–11986 (2022)
18. Lou, A., Guan, S., Loew, M.: Dc-unet: rethinking the u-net architecture with dual channel efficient cnn for medical image segmentation. In: Medical Imaging 2021: Image Processing. vol. 11596, pp. 758–768. SPIE (2021)
19. Ma, N., Zhang, X., Zheng, H.T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Proceedings of the European conference on computer vision (ECCV). pp. 116–131 (2018)
20. Pan, J., Bulat, A., Tan, F., Zhu, X., Dudziak, L., Li, H., Tzimiropoulos, G., Martinez, B.: Edgevits: Competing light-weight cnns on mobile devices with vision transformers. In: European Conference on Computer Vision. pp. 294–311. Springer (2022)
21. 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. pp. 234–241. Springer (2015)
22. Ruan, J., Xie, M., Gao, J., Liu, T., Fu, Y.: Ege-unet: an efficient group enhanced unet for skin lesion segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 481–490. Springer (2023)
23. Shaker, A., Maaz, M., Rasheed, H., Khan, S., Yang, M.H., Khan, F.S.: Swiftformer: Efficient additive attention for transformer-based real-time mobile vision applications. arXiv preprint arXiv:2303.15446 (2023)
24. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pvt v2: Improved baselines with pyramid vision transformer. Computational Visual Media **8**(3), 415–424 (2022)
25. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems **34**, 12077–12090 (2021)
26. Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z.H., Tay, F.E., Feng, J., Yan, S.: Tokens-to-token vit: Training vision transformers from scratch on imagenet. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 558–567 (2021)
27. Zhang, Q., Yang, Y.B.: Rest: An efficient transformer for visual recognition. Advances in neural information processing systems **34**, 15475–15485 (2021)
28. Zhou, H.Y., Guo, J., Zhang, Y., Yu, L., Wang, L., Yu, Y.: nnformer: Interleaved transformer for volumetric segmentation. arXiv preprint arXiv:2109.03201 (2021)
29. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)