DPMNet : *Dual-Path MLP-based Network for Aneurysm Image Segmentation*

Shudong Wang[1], Xue Zhao[1], Yulin Zhang[2], Yawu Zhao[1], Zhiyuan Zhao[1], Hengtao Ding[1], Tianxing Chen[3], and Sibo Qiao[4](✉)

[1] College of Computer Science and Technology, China University of Petroleum, Qingdao, Shandong, China
[2] College of Mathematics and System Science, Shandong College of Science and Technology, Qingdao, Shandong, China
[3] University of Washington, Seattle, America
[4] College of Software, Tiangong University, Tianjin, China
siboqiao@126.com

Abstract. MLP−based networks, while being lighter than traditional convolution− and transformer−based networks commonly used in medical image segmentation, often struggle with capturing local structures due to the limitations of fully−connected (FC) layers, making them less ideal for such tasks. To address this issue, we design a Dual−Path MLP−based network (DPMNet) that includes a global and a local branch to understand the input images at different scales. In the two branches, we design an Axial Residual Connection MLP module (ARC−MLP) to combine it with CNNs to capture the input image's global long−range dependencies and local visual structures simultaneously. Additionally, we propose a Shifted Channel−Mixer MLP block (SCM−MLP) across width and height as a key component of ARC−MLP to mix information from different spatial locations and channels. Extensive experiments demonstrate that the DPMNet significantly outperforms seven state−of−the−art convolution− , transformer−, and MLP−based methods in both Dice and IoU scores, where the Dice and IoU scores for the IAS−L dataset are 88.98% and 80.31% respectively. Code is available at `https://github.com/zx123868/DPMNet`.

Keywords: Medical image segmentation · Intracranial aneurysms · MLP

1 Introduction

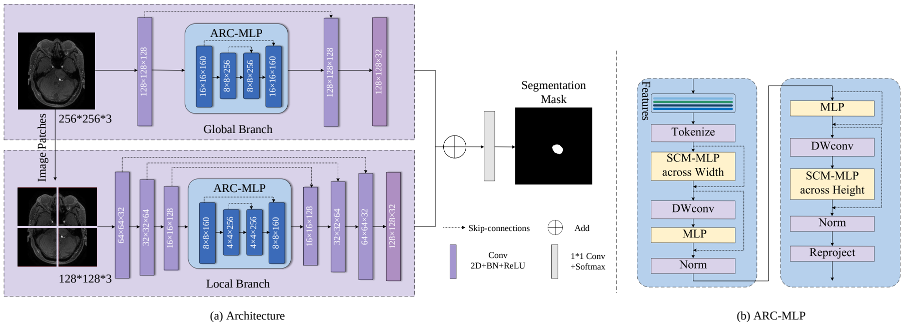
Intracranial aneurysms (IAS) is a prevalent disease characterized by a significant mortality risk [1]. However, the intricate and highly variable nature of these vascular structures demands a high level of expertise in neurovascular anatomy to achieve accurate segmentation. Consequently, there is a strong demand for an automated computer−aided system to perform the segmentation of aneurysms, aiming to enhance the early diagnosis rate of aneurysms. The success of Convolutional Neural Networks (CNNs) in image segmentation can be attributed to the inherent local relationships within images, where a pixel is more strongly connected to its nearby neighbors than those far away. One notable architecture in this domain is the UNet [11], which is based on an encoder−decoder structure. Following UNet [11], several significant enhancements like ResUNet [19], AttUNet [8], UNet++ [21], UNet3+ [6], WRANet [20] and DualANet [18] have been proposed. On top of that, we also aim to have the capability to capture long−range dependencies. ViT [5] emerged as a groundbreaking achievement in incorporating transformers into the field of computer vision (CV). Other transformer−based networks like TransUNet [3], MedT [14], Swin Transformer [7], and DPC−MSGATNet [9] are also widely used for medical image segmentation.

Recently, MLP−based networks, being lighter than traditional convolution− and transformer−based networks, have also been found to be competent in image segmentation. MLP−Mixer [13], as an all MLP−based network, utilizes two variants of MLP layers: channel−mixing MLPs and token−mixing MLPs, interleaved to facilitate interaction across input dimensions. UNeXt [15] combines CNNs and MLPs to craft a streamlined model that balances maintaining good performance with reducing both parameters and computational requirements. However, MLP−based networks often struggle with capturing local structures due to the limitations of fully−connected (FC) layers [4]. To address this issue, we propose DPMNet, which combines convolutional and MLP elements. While retaining the fundamental encoder−decoder setup of UNet, complete with skip connections, we use a dual−path structure and adapt the module design for improved performance. Specifically, DPMNet comprises two branches: global and local. These branches work concurrently, processing both the entire IAS view and image patches to grasp multi−scale representations. Each branch of DPMNet consists of two key stages: the convolutional stage and the Axial Residual Connection MLP (ARC−MLP) stage. In addition, we use the axial shift strategy in SCM−MLP to introduce a sense of locality to the block [16]. We evaluate DPMNet on three types of IAS datasets and two public medical datasets, demonstrating its superior performance compared to the latest generic segmentation architectures.

Our contributions can be summarized as follows: (1) We propose DPMNet, a dual−path MLP−based network for aneurysm image segmentation. We design an Axial Residual Connection MLP (ARC−MLP) module to combine it with CNNs. This integration allows for the simultaneous capture of both the input image's global long−range dependencies and local visual structures. (2) We design a Shifted Channel−Mixer MLP (SCM−MLP) block to enhance interactions between different channels and locations. (3) Our DPMNet significantly outperforms seven state−of−the−art methods in both Dice and IoU scores on five datasets.

2 DPMNet

The Overall Architecture: As shown in Fig. 1, DPMNet comprises two branches: global and local. The two branches use an encoder−decoder architecture with two stages: the Convolutional stage and the Axial Residual Connection MLP stage (ARC−MLP). Each convolutional block has a 3×3 convolutional layer, a batch normalization layer, and ReLU activation. The ARC−MLP is composed of four residual connections and two groups of sequentially connected Shifted Channel−Mixer MLP blocks (SCM−MLP) , operating in the width and height dimensions, along with DWconv, a MLP, and a normalization layer.



(a) Architecture

(b) ARC-MLP

Fig. 1. The overview of DPMNet and ARC−MLP.

Global/Local branch: To comprehend input images across various scales, we simultaneously use a global branch and a local branch to capture long−distance spatial dependencies among image patches and the image's high−level semantic details. In the global branch, we removed two layers of convolutional blocks as we observed that the shifting operation in the SCM−MLP block is sufficient to capture the image's local visual structures. In the local branch, the image is divided into four patches of size $I/2 \times I/2$, where I represents the image's original dimensions. Considering to decrease the computational complexity while giving it good performance, in the local branch we divide the original image into four patches. Each patch then undergoes individual processing through the local branch of DPMNet, and the resulting output feature maps are subsequently resampled based on their original locations to derive the final output feature maps. We adjust the output feature dimensions of both branches to 128×128×32 and add them together. This merged output is then passed through a 1×1 convolutional layer to produce the final segmentation mask.

Shifted Channel−Mixer MLP Block: To address the challenge that a fully−connected (FC) layer lacks local context due to the loss of spatial information, we use an axial shift strategy in the Shifted Channel−Mixer MLP block (SCM−MLP). As depicted in Fig. 2, we initiate an axial shift [17] after padding and chunking the feature map. B, C, H, and W represent the batch size, number of channels, height, and width of the feature map, respectively. N denotes the size of the shift. In this context, we're assuming $B = 1$, $C = 4$, and $N = 4$. The shift range spans from -2 to 2. This shift strategy enables the MLP to concentrate on specific locations within the convolutional features, introducing a sense of locality to the block. This idea is inspired by the Swin Transformer [7], which incorporates window−based attention to add greater locality into a primarily global model. The height and width shift operations facilitate communication between distinct spatial locations. After the shifting operation, we concatenate them in a particular dimension. Additionally, drawing inspiration from MLP−Mixer [13], the tokens resulting from these operations are fed into a channel−mixed MLP after layer normalization. The channel−mixed MLP comprises two fully connected layers and a GELU nonlinearity, promoting communication between different channels. In summary, our approach not only blends information from various spatial locations but also mixes information across different channels. This dual mixing enhances the feature information, contributing to a more comprehensive understanding of the data and improved segmentation accuracy.

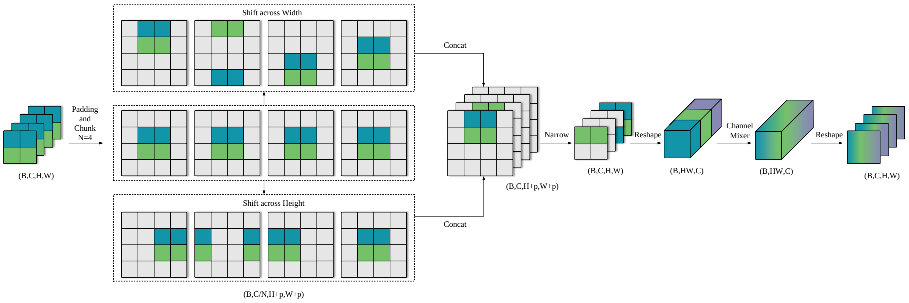


Fig. 2. The architecture of Shifted Channel−Mixer MLP (SCM−MLP) Block.

Axial Residual Connection MLP stage: To capture both global long−range dependencies within the input image as well as the finer details of its local visual structures, we design the ARC−MLP to combine it with CNNs [10]. In the ARC−MLP stage, we first project the features into 1D tokens to facilitate subsequent full connection layer operations. In the tokenization process, we employ a kernel size of 3 to extract patches, which are then flattened and resized to match the embedding dimension. These tokens undergo processing through an SCM−MLP (across width) to reintroduce localization, compensating for the absence resulting from the substitution of partial convolution with fully connected layers. Then we use a depth wise convolutional layer (DWConv) between the SCM−MLP and the MLP. The role of DWConv is to encode positional information within the MLP features in a resource−efficient manner. It allows the model to understand and represent the spatial relationships of features in a more streamlined and computationally efficient way, optimizing the network's performance. The MLP feeds tokens that preserve the original features into the fully connected layer. Followed by a layer normalization (LN), we then pass tokens through another group of MLP, DWconv, SCM−MLP (across height), and layer normalization. We add four residual connections behind each of the four MLP blocks. Finally we pass the output features to the next block.

The computation in the ARC−MLP stage can be summarized as:

$$X_T = Tokenize\left(X\right); X_{shift} = SCM - MLP_W\left(X_T\right), \tag{1}$$

$$Y = f\left(LN\left(X_T + MLP\left(DWConv\left(X_T + X_{shift}\right)\right)\right)\right), \tag{2}$$

$$Y_T = Tokenize\left(Y\right); Y_{shift} = SCM - MLP_H\left(Y_T\right), \tag{3}$$

$$Y = f\left(LN\left(Y_T + MLP\left(DWConv\left(Y_T + Y_{shift}\right)\right)\right)\right), \tag{4}$$

where X denotes the original input features, H denotes height, W denotes width, $Tokenize\left(.\right)$ denotes converting input features into tokens that can be fed into the fully connected layer, *SCM-MLP* $\left(.\right)$ denotes the shifted channel−mixer MLP, $MLP\left(.\right)$ denotes feeding into an MLP layer, $DWConv$ denotes depth−wise convolution and LN denotes layer normalization.

3 Experiments

Datasets: We adhere to the ethical guidelines outlined in the Declaration of Helsinki during our research, with approval from the Ethics Committee of the Affiliated Hospital of Qingdao University. Our dataset comprises 3D−TOF−MRA images from 679 patients, consisting of 579 with unruptured cystic aneurysm (IAS positive) and 100 without (IAS negative). Skilled medical professionals manually annotated the aneurysms as the ground truth. To assess the model's segmentation performance across various aneurysm sizes, we categorized them into three datasets based on their diameters: IAS−L, IAS−M, and IAS−S. These categories represent aneurysms with diameters of over 7mm, between 3mm and 7mm, and less than 3mm, respectively. The datasets consist of 1186 images for IAS−L, 1838 images for IAS−M, and 672 images for IAS−S. Furthermore, we employed two publicly medical datasets, MSD Lung Tumours [2,12] and MSD Colon Cancer [2,12], containing 1613 and 1278 images, respectively, to assess the performance of our DPMNet. The image slices in all datasets were adjusted to a 256 × 256 size.

Implementation Details: All networks were implemented based on the Pytorch framework and our experiments were conducted on NVIDIA 4090 24G GPUs. We've incorporated a combination of binary cross−entropy (BCE) and dice loss to train DPMNet. The loss L between the predicted $\hat{y}$ and the target y is formulated as:

$$L = 0.4BCE\left(\hat{y}, y\right) + 0.6Dice\left(\hat{y}, y\right) \tag{5}$$

We use AdamW as our optimizer, starting with a learning rate of 0.0001. Additionally, we employ the CosineAnnealingLR scheduler, setting a minimum learning rate of 1e-5. The batch size is fixed at 4. We train DPMNet for a total of 400 epochs. To evaluate our approach, we employed dice similarity coefficient (Dice) and Intersection over Union (IoU) as metrics. We conducted three random 80−20 splits on the dataset and calculated the mean and standard deviation of the results to provide a comprehensive evaluation.

Table 1. Comparative experimental results on the three IAS datasets.

Networks	IAS−L		IAS−M		IAS−S	
	Dice	IoU	Dice	IoU	Dice	IoU
UNet [11]	83.25±0.39	73.43±0.43	65.62±0.39	54.11±0.42	50.78±1.00	39.56±1.03
AttUNet [8]	83.57±0.68	73.80±0.62	59.23±1.98	48.09±2.09	51.14±0.65	39.69±0.51
UNet++ [21]	82.01±0.49	72.37±0.43	63.30±0.52	52.11±0.62	49.07±1.02	38.33±0.92
WRANet [20]	83.63±0.08	73.84±0.24	62.95±0.98	51.65±0.76	45.38±0.74	35.14±0.57
DualANet [18]	81.99±1.04	71.90±1.12	66.24±0.20	54.35±0.25	50.22±1.57	38.68±1.44
TransUNet [3]	80.52±0.34	70.88±0.43	60.82±0.88	49.93±0.72	46.63±0.70	35.89±0.74
UNeXt [15]	86.93±0.20	77.24±0.29	72.44±0.70	57.72±0.77	52.30±2.23	36.62±1.87
DPMNet	**88.98±0.10**	**80.31±0.17**	**79.14±0.18**	**66.00±0.22**	**64.55±0.44**	**48.67±0.39**

Comparative Results: To ensure a comprehensive performance assessment, we selected several SOTA segmentation methods based on different architectural paradigms: CNN, Transformer, and MLP. UNet [11], AttUNet [8], UNet++ [21], WRANet [20], and DualANet [18] all belong to the CNN−based category. On the other hand, TransUNet [3] belongs to the transformer−based models, while UNeXt [15] represents the MLP−based models. Table 1 records the results of the comparative experiments conducted on the three IAS datasets, while Table 2 presents the results obtained from the two public datasets. It's obvious that our proposed DPMNet significantly outperforms all other SOTA methods in both Dice and IoU scores.

In comparison to the baseline model UNet [11], our DPMNet exhibited improvements in Dice scores by 5.73%, 13.52%, and 13.77%, and in IoU scores by 6.88%, 11.89%, and 9.11%, across the three IAS datasets. Likewise, when compared to UNet [11] on the MSD Lung Tumours dataset [2,12] and the MSD Colon Cancer dataset [2,12], DPMNet yields increased Dice scores by 8.44% and 9.93%, and higher IoU scores of 11.8% and 11.97%. For the IAS−S dataset

Table 2. Comparative experimental results on the MSD Lung Tumours dataset and the MSD Colon Cancer dataset.

Networks	Params (in M)	GFLOPs	MSD Lung Tumours [2,12]		MSD Colon Cancer [2,12]	
			Dice	IoU	Dice	IoU
UNet [11]	34.53	262.09	82.64±0.39	73.93±0.31	72.30±0.63	61.00±0.70
AttUNet [8]	34.88	266.53	82.95±0.21	74.31±0.25	71.96±0.91	60.97±1.01
UNet++ [21]	9.16	139.61	84.14±1.02	75.42±0.12	75.24±0.35	63.98±0.27
WRANet [20]	34.88	267.16	83.15±0.51	74.79±0.46	71.95±0.17	60.59±0.19
DualANet [18]	2.58	22.04	80.48±1.78	71.20±2.01	67.90±0.40	55.64±0.42
TransUNet [3]	93.23	228.91	79.76±0.84	71.06±0.76	66.17±1.43	54.76±1.10
UNeXt [15]	**1.47**	**2.29**	90.25±0.20	82.53±0.29	80.98±0.13	68.52±0.19
DPMNet	31.72	33.6	**91.08±0.05**	**83.86±0.09**	**84.10±0.18**	**72.97±0.25**

which aneurysm diameters less than 3mm, the UNet++ [21], TransUNet [3], and WRANet [20] methods displayed subpar segmentation performance as they struggled to capture detailed representations in smaller tumor regions. On the other four datasets, the TransUNet [3] model achieves the worst performance, possibly because of its heavy reliance on vast data for feature learning, leading to challenges in accurately delineating tumor boundaries. In contrast, our DPMNet demonstrated adaptability to different diameters and types of tumors, attributed to long−range and local feature information extraction provided by the dual−path structure and the ARC−MLP module. Fig. 3 presents comparative experimental results on the IAS−L dataset. To enhance the visualization of the differences between the comparative models' predictions, we cropped the images and adjusted the masks to occupy a larger proportion of the images.

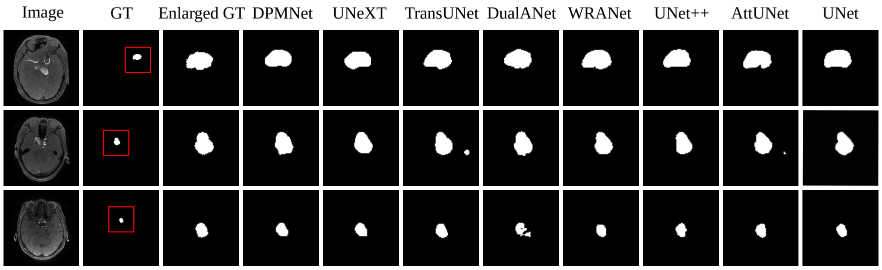


Fig. 3. Comparative experimental results on the IAS−L dataset.

4 Discussion

Ablation Study: To analyze the contributions of each component in our DPMNet, we conducted detailed ablation studies focusing on the SCM−MLP block and branches. The results presented in Table 3 reveal that incorporating the SCM−MLP block in DPMNet enhances segmentation performance on the IAS−L dataset, exhibiting improvements of 0.35% in Dice and 0.42% in IoU score compared to DPMNet without SCM−MLP. This underscores the significance of the SCM−MLP block in effectively integrating feature information from diverse spatial locations and channels. Furthermore, DPMNet surpasses the performance of both the local and global branches individually. Specifically, the Dice scores improve by 0.69% and 0.28%, while IoU scores see enhancements of 1.02% and 0.24%, respectively. This signifies that DPMNet captures more comprehensive representations by combining global context and local visual cues, outperforming a single−branch approach. The ablation study reinforces the notion that each component of DPMNet contributes meaningfully to enhancing overall performance.

Table 3. Ablation studies on the IAS−L dataset.

Model	Params(in M)	GFLOPs	Dice	IoU
Local Branch	20.49	31.43	88.29±0.15	79.29±0.23
Global Branch	9.4	30.9	88.70±0.18	80.07±0.16
DPMNet w/o SCM−MLP	3.01	28.18	88.63±0.10	79.89±0.13
DPMNet	31.72	33.6	88.98±0.10	80.31±0.17

5 Conclusions and Future Works

In this paper, we proposed a new deep dual−path network architecture DPMNet for improving aneurysm segmentation performance, which can assist clinicians in analazing aneurysm morphology and promote IAS diagnosis. We incorporated an Axial Residual Connection MLP (ARC−MLP) module and a Shifted Channel−Mixer MLP (SCM−MLP) block to enhance extraction of semantic information. Experimental results demonstrate the effectiveness of our approach in achieving state−of−the−art performance.

Regarding limitations, our DPMNet relies on labeled data for supervised training. Typically, annotating the IAS views' data is intricate and demands considerable time from experienced cardiologists. In the future, we aim to adopt a semi−supervised approach for training the model, significantly lessening our dependence on labeled data.

Acknowledgments. This work is supported by the National Key Research and Development Program of China (2021YFA1000102, 2021YFA1000103) and Chinese College Students' Innovation and Entrepreneurship Training Program (202310424302).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Agid, R., Andersson, T., Almqvist, H., Willinsky, R., Lee, S.K., Farb, R., Söderman, M., et al.: Negative ct angiography findings in patients with spontaneous subarachnoid hemorrhage: when is digital subtraction angiography still needed? American journal of neuroradiology **31**(4), 696–705 (2010)
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**, 4128 (2022)
3. 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)
4. Ding, X., Xia, C., Zhang, X., Chu, X., Han, J., Ding, G.: Repmlp: Re-parameterizing convolutions into fully-connected layers for image recognition. arXiv preprint arXiv:2105.01883 (2021)
5. 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)
6. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.W., Wu, J.: Unet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 1055–1059. IEEE (2020)
7. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
8. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
9. Qiao, S., Pang, S., Luo, G., Sun, Y., Yin, W., Pan, S., Lv, Z.: Dpc-msgatnet: dual-path chain multi-scale gated axial-transformer network for four-chamber view segmentation in fetal echocardiography. Complex & Intelligent Systems **9**, 4503–4519 (2023)
10. Qiao, S., Pang, S., Xie, P., Yin, W., Yu, S., Gui, H., Wang, M., Lyu, Z.: Hcmmnet: Hierarchical conv-mlp-mixed network for medical image segmentation in metaverse for consumer health. IEEE Transactions on Consumer Electronics **70**(1), 2078–2089 (2024)
11. 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)
12. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M., Golia-Pernicka, J., Heckers, S.H., Jarnagin, W.R., McHugo, M.K., Napel, S., Vorontsov, E., Maier-Hein, L., Cardoso, M.J.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms (2019)
13. Tolstikhin, I.O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., et al.: Mlp-mixer: An all-

mlp architecture for vision. Advances in neural information processing systems **34**, 24261–24272 (2021)

14. Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I., Patel, V.M.: Medical transformer: Gated axial-attention for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 36–46. Springer (2021)
15. Valanarasu, J.M.J., Patel, V.M.: Unext: Mlp-based rapid medical image segmentation network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 23–33. Springer (2022)
16. Wang, H., Zhu, Y., Green, B., Adam, H., Yuille, A., Chen, L.C.: Axial-deeplab: Stand-alone axial-attention for panoptic segmentation. In: European conference on computer vision. pp. 108–126. Springer (2020)
17. Yu, T., Li, X., Cai, Y., Sun, M., Li, P.: S2-mlp: Spatial-shift mlp architecture for vision. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 297–306 (2022)
18. Zhang, Y., Han, Z., Liu, L., Wang, S.: Duala-net: A generalizable and adaptive network with dual-branch encoder for medical image segmentation. Computer Methods and Programs in Biomedicine **243**, 107877 (2024)
19. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters **15**(5), 749–753 (2018)
20. Zhao, Y., Wang, S., Zhang, Y., Qiao, S., Zhang, M.: Wranet: wavelet integrated residual attention u-net network for medical image segmentation. Complex & Intelligent Systems pp. 1–13 (2023)
21. 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)