ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10446716

DEEP FUSION OF SHIFTED MLP AND CNN FOR MEDICAL IMAGE SEGMENTATION

Chengyu Yuan[1,2,3], Hao Xiong[4,✉], Guoqing Shangguan[1], Hualei Shen[1,2,3,✉], Dong Liu[1,2,3], Haojie Zhang[5,6], Zhonghua Liu[7], Kun Qian[5,6], Bin Hu[5,6], Björn W. Schuller[8,9], Yoshiharu Yamamoto[10], Shlomo Berkovsky[4]

[1] College of Computer and Information Engineering, Henan Normal University, China
[2] Key Laboratory of Artificial Intelligence and Personalized Learning in Education, Henan, China
[3] Big Data Engineering Lab of Teaching Resources & Assessment of Education Quality, Henan, China
[4] Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Australia
[5] Key Laboratory of Brain Health Intelligent Evaluation and Intervention (BIT), Ministry of Education, China
[6] School of Medical Technology, Beijing Institute of Technology, China
[7] School of Information Engineering, Zhejiang Ocean University, China
[8] GLAM – Group on Language, Audio, & Music, Imperial College London, U. K.
[9] Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany
[10] Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Japan

ABSTRACT

Medical image segmentation is an important task in modern analysis of medical images. Current methods tend to extract either local features with convolutions or global features with Transformers. However, few of them are able to effectively fuse global and local features to facilitate segmentation. In this work, we propose a novel hybrid network that involves three main branches: the Multi-Layer Perception (MLP) branch, the Convolutional Neural Network (CNN) branch, and a Fusion branch. The MLP and CNN branches aim to learn global and local features, respectively. To fuse these, the fusion branch introduces a novel hierarchical fusion that performs multi-layered fusions that generate high-level representations to enhance segmentation. Our evaluation with two datasets shows strong performance of the proposed method compared to state-of-the-art baselines.

***Index Terms*—** Medical image segmentation, MLP, CNN, hierarchical fusion

1. INTRODUCTION

Medical image segmentation semantically segments images and is crucial in clinical analysis and decision making. The existing works can be broadly classified as: CNN-based, Transformer-based, and MLP-based methods.

CNN-based models. UNet is the most widely used CNN network for medical image segmentation [1]. Inspired by UNet, many variants have been proposed to further enhance segmentation performance [2, 3, 4, 5]. Rather than capturing long-range dependencies, these CNN-based methods only extract local features due to the limited receptive field of convolutions. However, long-range dependencies also introduce vital information into segmentation.

Transformer-based models. Vision Transformer has been an emerging topic in computer vision [6, 7]. For medical image segmentation, typical Transformers such as Medical Transformer [8], attention gated networks [9], and attention UNet [10] have been proposed. The superiority of these methods can be attributed to their strong capability in global information extraction. However, the attention mechanisms in Transformers involve complex computations, and consequently require demanding computational resources.

MLP-based models. Recently, MLP-based approaches [11, 12, 13] have been recognised as efficient alternatives to Transformers. They exploit MLPs to learn channel-wise features capturing global information. Besides, the linear MLP operation is simple and can substantially reduce the computational complexity. Accordingly, MLP-based methods, including UNeXt [14], PHNet [15] and USMLP [16], have been ex-

This work is partially supported by National Natural Science Foundation of China (No. 62072160, 62227807, and 62272044), Science and Technology Research Project of Henan (No. 232102211024), China, the Teli Young Fellow Program from the Beijing Institute of Technology, China and the Grants-in-Aid for Scientific Research (No. 20H00569) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. ✉*Corresponding authors*: Hao Xiong (hao.xiong@mq.edu.au), Hualei Shen (shenhualei@htu.edu.cn).

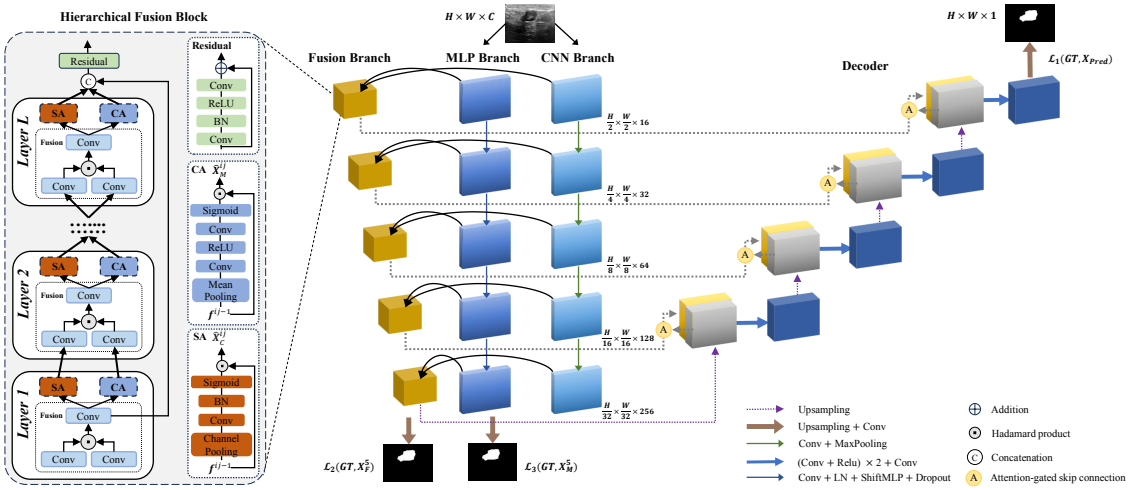


Fig. 1: The proposed network consists of a decoder and a hybrid encoder, including a Fusion branch, an MLP branch, and a CNN branch. The MLP branch and CNN branch capture the global and local features, respectively. The fusion branch, containing hierarchical fusion blocks, fuses the global and local features extracted by the MLP and CNN branches.

ploited in medical image segmentation. These methods incorporate MLPs into the CNN architecture. The resultant mixed structure of MLP and CNN facilitates the learning of both local and global features.

To mitigate the aforementioned issues, we propose a novel hybrid network containing three separate branches – the MLP branch, CNN branch, and Fusion branch – for medical image segmentation. The MLP and CNN branches work in parallel, aiming to capture global and local features, respectively. In the fusion branch, we introduce a novel hierarchical fusion block to fuse the extracted global and local features for segmentation. In the hierarchical fusion block, its first layer initially fuses the features from MLP and CNN branches, for which it then exploits channel attention and spatial attention to identify higher-level global information and local details, and feeds these into the subsequent layers. In each layer, we repeatedly fuse the identified higher-level features from the previous layer, and then apply channel attention and spatial attention. As the number of fusion layers increases, the feature representation becomes more and more high-level. At last, high-level features generated by the last layer are concatenated with low-level features from the first layer to yield robust segmentation.

2. METHOD

Fig. 1 illustrates the proposed network with a hybrid encoder and decoder. The hybrid encoder has three branches including a Fusion branch, an MLP branch, and a CNN branch. We will introduce them in following sub-sections.

2.1. Hybrid Encoder

Let $I \in \mathbb{R}^{H\times W\times C}$ denote an input medical image, with height, width, and channel of H, W, and C, respectively. The image I is first fed into both the MLP branch and the CNN branch.

MLP Branch: The MLP branch captures global information and includes five MLP blocks. Each MLP block is equipped with a convolutional (Conv) layer, a layer normalisation (LN), an axial shifted MLP layer [12], and a dropout. For Conv, we set its kernel size to 3×3, stride to 2, and padding to 1. For the axial shifted MLP, we exploit the shift step of 5 to shift the feature maps along the width and height axis. The dropout rate is set to 0.1. Each MLP block reduces the feature resolution by 2.

CNN Branch: The CNN branch aims to extract local features and it also contains five CNN blocks, for which each block is equipped with a convolutional (Conv) layer, a batch normalisation (BN) layer, a max pooling operator and a ReLU activation function. For Conv, we set its kernel to 3×3, stride to 1, and padding to 1. After each CNN block, the feature resolution is reduced by 2.

Fusion Branch: The fusion branch also has five hierarchical fusion blocks. The i^{th} block aims to fuse the features produced by its corresponding i^{th} MLP block and i^{th} CNN block. Fig. 1 illustrates the proposed hierarchical fusion block with L layers. For block i, we obtain the fused feature $\mathbf{f}^{ij}$ of the j^{th} layer via:

$$\mathbf{f}^{ij} = \begin{cases} Conv(W_1^{ij} X_M^i \odot W_2^{ij} X_C^i), & j = 1 \\ Conv(W_1^{ij} \hat{X}_M^{ij} \odot W_2^{ij} \hat{X}_C^{ij}), & j > 1, \end{cases} \quad (1)$$

where W_1^{ij}, W_2^{ij} are parameters of the convolutional layer, and $Conv(\cdot)$ and $\odot$ represent convolution and the Hadamard

product, respectively. The first layer ($j = 1$) fuses the global feature X_M^i and local feature X_C^i extracted by the i^{th} MLP and CNN blocks. We then perform hierarchical fusion in subsequent layers ($j > 1$). In each layer, we concurrently apply channel attention [17] and spatial attention [18] to identify the global feature $\hat{X}_M^{ij}$ and local feature $\hat{X}_C^{ij}$ from the fused feature produced by the previous layer, and then perform fusion (as per Eq. 1) to $\hat{X}_M^{ij}$ and $\hat{X}_C^{ij}$.

The spatial attention is defined as:

$$\hat{X}_C^{ij} = \sigma(BN(Conv(CP(\mathbf{f}^{ij-1})))) \odot \mathbf{f}^{ij-1}, \quad (2)$$

where σ is the sigmoid function, $CP(\cdot)$ is the cross channel pooling operator, and BN represents batch normalisation. Besides, the channel attention is:

$$\hat{X}_M^{ij} = \sigma(Conv(ReLU(Conv(MP(\mathbf{f}^{ij-1}))))) \odot \mathbf{f}^{ij-1}, \quad (3)$$

where $MP(\cdot)$ refers to the mean pooling operator, and $ReLU(\cdot)$ is the ReLU operator.

After L layered fusion and attention operations, we obtain the high-level representations of global information $\hat{X}_M^{iL}$ and local information $\hat{X}_C^{iL}$. Combined with the fused feature $\mathbf{f}^{i1}$ of the first layer, we obtain the final fused feature X_F^i

$$X_F^i = Res([\mathbf{f}^{i1}, \hat{X}_M^{iL}, \hat{X}_C^{iL}]), \quad (4)$$

where $Res(\cdot)$ and $[\cdot]$ are the residual block and concatenation operators, respectively.

2.2. Decoder

As shown in Fig. 1, the decoder of our network primarily concatenates features from the encoder and performs upsampling to generate the segmentation map. We utilise attention-gated skip-connection [9] to connect the encoder and decoder.

2.3. Loss Function

Finally, the loss function $\mathcal{L}$ of our network is defined as:

$$\mathcal{L} = \alpha\mathcal{L}_1(GT, X_P) + \beta\mathcal{L}_2(GT, X_F^5) + \gamma\mathcal{L}_3(GT, X_M^5), \quad (5)$$

where GT refers to the ground truth, X_P is the predicted segmentation map, and X_F^5, X_M^5 are features produced by the last Fusion and MLP blocks (5^{th} block in branch). The terms $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$ are:

$$\mathcal{L}_1 = BCE_w(GT, X_P) + IoU_w(GT, X_P), \quad (6)$$

$$\mathcal{L}_2 = BCE_w(GT, X_F^5) + IoU_w(GT, X_F^5), \quad (7)$$

$$\mathcal{L}_3 = BCE_w(GT, X_M^5) + IoU_w(GT, X_M^5). \quad (8)$$

Here, $BCE_w(\cdot)$ and $IoU_w(\cdot)$ are weighted binary cross entropy and weighted intersection over union functions [19], respectively.

3. EXPERIMENTS AND RESULTS

3.1. Data

We used the Breast UltraSound Images (BUSI) [21] and International Skin Imaging Collaboration (ISIC 2018) [22] datasets in the evaluation. BUSI contains 780 images and their ground-truth segmentation maps. All these are categorised as normal, benign, or malignant cases of breast cancer. We utilise only benign and malignant images, including 647 cases in total, as the benchmark set. ISIC2018 provides $2,594$ skin images and the corresponding lesion segmentation maps.

3.2. Implementation Details and Evaluation Metrics

We trained our model on one NVIDIA A100 GPU using the Adam optimiser with a momentum of 0.9, learning rate of 0.001, and batch size of 16. The maximum training epoch is set to 1000. We performed 5-fold cross validation on each dataset using an Intel® Xeon® CPU E5-2620.

We compare the computational complexity of the evaluated methods using the number of parameters, floating point operators (FLOPs), and average CPU inference time. To measure segmentation performance, we utilise the Dice similarity coefficient (DSC) and mean intersection over union (mIoU):

$$DSC = 2\frac{|X \cap Y|}{|X| + |Y|}, IoU = \frac{|X \cap Y|}{|X \cup Y|}, \quad (9)$$

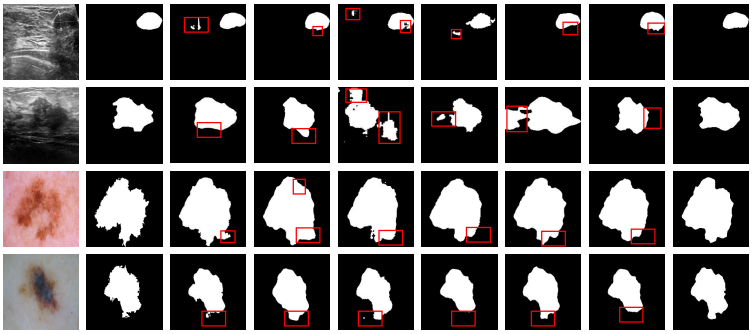
where X denotes the predicted segmentation, Y is the ground truth, and $mIoU$ is the mean IoU across all the classes.

3.3. Results

We compare our method to several baselines, including CNN based methods (U-Net [1], UNet++ [3], and ResUNet [2]) and transformer/MLP based methods (MedT [8], TransFuse [20], and UNeXt [14]). As shown in Table 1, our method outperforms the others on both datasets in terms of DSC and mIoU. It is noteworthy that while our method is inferior to UNeXt in terms of CPU inference time, this metric is not the main focus of this work. With about $1s$ inference time, our model has the potential to be deployed in a computational resource-constrained environment.

Also, we observe that our method is slightly better than the second best one for each dataset. For example, for BUSI the DSC score of the second best TransFuse is 0.006 lower than ours. However, its DSC score on the ISIC2018 dataset drops substantially compared to our method. This indicates that our performance is more stable as we consistently achieve best results on both datasets.

In Fig. 2, we visually compare the segmentation maps generated by our method and those generated by other methods. In general, our approach generates more accurate segmentation maps compared to the ground truth.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 2: Segmentation examples of our method and other baselines on BUSI (top two rows) and ISIC2018 (bottom two rows): (a) original image, (b) ground truth, (c) U-Net, (d) U-Net++, (e) ResUNet, (f) MedT, (g) TransFuse, (h) UNeXt, and (i) the proposed method. The segmentation artefacts of the comparison methods are marked by red bounding boxes.

Table 1: Quantitative evaluation of the proposed method and state-of-the-art baselines on BUSI and ISIC2018.

Methods	Params (M)↓	FLOPs (G) ↓	Inference Time CPU (s) ↓	BUSI		ISIC2018	
				DSC ↑	mIoU ↑	DSC ↑	mIoU ↑
U-Net [1]	31.04	55.84	4.519	$0.7367_{\pm 0.0237}$	$0.6462_{\pm 0.0232}$	$0.7945_{\pm 0.0188}$	$0.7103_{\pm 0.0218}$
UNet++ [3]	9.16	34.91	5.957	$0.7417_{\pm 0.0223}$	$0.6008_{\pm 0.0287}$	$0.8883_{\pm 0.0152}$	$0.8036_{\pm 0.0229}$
ResUNet [2]	13.04	80.98	6.509	$0.6442_{\pm 0.0323}$	$0.5380_{\pm 0.0340}$	$0.8663_{\pm 0.0061}$	$0.7926_{\pm 0.0088}$
MedT [8]	1.60	21.24	12.264	$0.6724_{\pm 0.0386}$	$0.5685_{\pm 0.0438}$	$0.8487_{\pm 0.0220}$	$0.7674_{\pm 0.0271}$
TransFuse [20]	26.17	8.65	1.609	$0.7921_{\pm 0.0166}$	$0.7075_{\pm 0.0213}$	$0.8898_{\pm 0.0044}$	$0.8202_{\pm 0.0056}$
UNeXt [14]	**1.47**	**0.58**	**0.298**	$0.7605_{\pm 0.0037}$	$0.6250_{\pm 0.0039}$	$0.8932_{\pm 0.0070}$	$0.8111_{\pm 0.0111}$
Ours	5.66	1.57	1.229	$\mathbf{0.7980}_{\pm 0.0097}$	$\mathbf{0.7132}_{\pm 0.0119}$	$\mathbf{0.8983}_{\pm 0.0047}$	$\mathbf{0.8316}_{\pm 0.0059}$

Table 2: Effectiveness of each branch for segmentation on BUSI and ISIC2018.

Architecture	BUSI		ISIC2018	
	DSC ↑	mIoU ↑	DSC ↑	mIoU ↑
w/o MLP Branch	$0.7562_{\pm 0.0199}$	$0.6650_{\pm 0.0248}$	$0.8915_{\pm 0.0146}$	$0.8225_{\pm 0.0041}$
w/o CNN Branch	$0.7895_{\pm 0.0182}$	$0.7040_{\pm 0.0224}$	$0.8968_{\pm 0.0024}$	$0.8304_{\pm 0.0031}$
w/o Fusion Branch	$0.7851_{\pm 0.0125}$	$0.6997_{\pm 0.0146}$	$0.8943_{\pm 0.0032}$	$0.8266_{\pm 0.0039}$
Ours	$\mathbf{0.7980}_{\pm 0.0097}$	$\mathbf{0.7132}_{\pm 0.0119}$	$\mathbf{0.8983}_{\pm 0.0047}$	$\mathbf{0.8316}_{\pm 0.0059}$

Table 3: Sensitivity of the number of layers in the hierarchical fusion block on BUSI and ISIC2018. The fusion block with 3 layers achieves the best performance.

Layer Numbers	BUSI		ISIC2018	
	DSC ↑	mIoU ↑	DSC ↑	mIoU ↑
1 Layer	$0.7895_{\pm 0.0143}$	$0.7049_{\pm 0.0133}$	$0.8962_{\pm 0.0038}$	$0.8292_{\pm 0.0042}$
2 Layers	$0.7920_{\pm 0.0170}$	$0.7094_{\pm 0.0185}$	$0.8973_{\pm 0.0023}$	$0.8315_{\pm 0.0030}$
3 Layers	$\mathbf{0.7980}_{\pm 0.0097}$	$\mathbf{0.7132}_{\pm 0.0119}$	$\mathbf{0.8983}_{\pm 0.0047}$	$\mathbf{0.8316}_{\pm 0.0059}$
4 Layers	$0.7918_{\pm 0.0146}$	$0.7064_{\pm 0.0171}$	$0.8943_{\pm 0.0034}$	$0.8295_{\pm 0.0044}$
5 Layers	$0.7865_{\pm 0.0291}$	$0.7004_{\pm 0.0321}$	$0.8906_{\pm 0.0033}$	$0.8219_{\pm 0.0039}$

3.4. Ablation Studies

We first evaluate the effectiveness of three individual branches. Table 2 shows that removing either the MLP (**w/o MLP Branch**) or CNN branch (**w/o CNN Branch**) decreases DSC and mIoU on both datasets. That indicates that both global and local information is important to ensure robust segmentation. Meanwhile, **w/o Fusion Branch** replaces the proposed fusion branch by simply concatenating the global and local features. We observe that such a concatenation decreases segmentation accuracy, which further shows that our proposed fusion branch attains a more efficient fusion.

We also evaluate how the number of layers in our hierarchical fusion block affects the performance. As shown in Table 3, our fusion module with 3 layers achieves the best results. It is also worth noting that adding more layers to the fusion may introduce more parameters, degrading the inference speed. Hence, 3 layers can be regarded as the optimal trade-off between accuracy and the number of parameters.

4. CONCLUSION

This paper proposed a hybrid network for medical image segmentation. It contains an MLP branch and a CNN branch that learn both global and local features from the input image. The fusion branch with hierarchical fusion blocks is designed to effectively produce high-level representations from features extracted by the MLP and CNN branches, and then to fuse them with low-level features to enhance segmentation. Experimental results on two public datasets show the superiority of our method over several state-of-the-art baselines.

5. REFERENCES

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2015, pp. 234–241, Springer.

[2] Xiao Xiao, Shen Lian, Zhiming Luo, and Shaozi Li, "Weighted res-UNet for high-quality retina vessel segmentation," in *Proceedings of the 9th International Conference on Information Technology in Medicine and Education*. 2018, pp. 327–331, IEEE.

[3] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "UNet++: redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.

[4] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D Johansen, "Resunet++: An advanced architecture for medical image segmentation," in *2019 IEEE International Symposium on Multimedia*. 2019, pp. 225–2255, IEEE.

[5] Hasib Zunair and A Ben Hamza, "Sharp U-Net: Depthwise convolutional network for biomedical image segmentation," *Computers in Biology and Medicine*, vol. 136, pp. 104699, 2021.

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, and Sylvain Gelly, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proceedings of the International Conference on Learning Representations*, 2020.

[7] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, and Philip HS Torr, "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *2021 IEEE Conference on Computer Vision and Pattern Recognition*. 2021, pp. 6881–6890, IEEE.

[8] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel, "Medical transformer: Gated axial-attention for medical image segmentation," in *Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2021, pp. 36–46, Springer.

[9] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, 2019.

[10] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert, "Attention U-Net: Learning where to look for the pancreas," *arXiv, 1804.03999*, 2018.

[11] Ilya O Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, and Jakob Uszkoreit, "MLP-mixer: An all-MLP architecture for vision," *Advances in Neural Information Processing Systems*, vol. 34, pp. 24261–24272, 2021.

[12] Dongze Lian, Zehao Yu, Xing Sun, and Shenghua Gao, "As-MLP: An axial shifted MLP architecture for vision," in *International Conference on Learning Representations*, 2022.

[13] Tan Yu, Xu Li, Yunfeng Cai, Mingming Sun, and Ping Li, "S^2-MLP: Spatial-shift MLP architecture for vision," in *Proceedings of the 2022 IEEE Winter Conference on Applications of Computer Vision*. 2022, pp. 297–306, IEEE.

[14] Jeya Maria Jose Valanarasu and Vishal M Patel, "Unext: MLP-based rapid medical image segmentation network," in *Proceedings of the 25th International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2022, pp. 23–33, Springer.

[15] Yi Lin, Xiao Fang, Dong Zhang, Kwang-Ting Cheng, and Hao Chen, "A permutable hybrid network for volumetric medical image segmentation," *arXiv, 2303.13111*, 2023.

[16] Jiaming Luo, Yongzhe Tang, Jie Wang, and Hongtao Lu, "USMLP: U-shaped sparse-MLP network for mass segmentation in mammograms," *Image and Vision Computing*, vol. 137, pp. 104761, 2023.

[17] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *2018 IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 7132–7141, IEEE.

[18] Yading Yuan and Yeh-Chi Lo, "Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 519–526, 2019.

[19] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand, "Basnet: Boundary-aware salient object detection," in *2019 IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 7471–7481, IEEE.

[20] Yundong Zhang, Huiye Liu, and Qiang Hu, "Transfuse: Fusing transformers and cnns for medical image segmentation," in *Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2021, pp. 14–24, Springer.

[21] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, pp. 104863, 2020.

[22] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, and Harald Kittler, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proceedings of the 15th International Symposium on Biomedical Imaging (ISBI 2018)*. 2018, pp. 168–172, IEEE.