TinyU-Net: Lighter yet Better U-Net with Cascaded Multi-Receptive Fields

Junren Chen[1], Rui Chen[2], Wei Wang[3], Junlong Cheng[1], Lei Zhang[1](✉), and Liangyin Chen[1](✉)

[1] College of Computer Science, Sichuan University, Chengdu, Sichuan, China
{zhanglei, chenliangyin}@scu.edu.cn
[2] Department of Electronic Engineering, Tsinghua University, Beijing, China
[3] School of Automation, Chengdu University of Information Technology, Chengdu, Sichuan, China

Abstract. The lightweight models for automatic medical image segmentation have the potential to advance health equity, particularly in limited-resource settings. Nevertheless, their reduced parameters and computational complexity compared to state-of-the-art methods often result in inadequate feature representation, leading to suboptimal segmentation performance. To this end, we propose a Cascade Multi-Receptive Fields (CMRF) module and develop a lighter yet better U-Net based on CMRF, named TinyU-Net, comprising only 0.48M parameters. Specifically, the CMRF module leverages redundant information across multiple channels in the feature map to explore diverse receptive fields by a cost-friendly cascading strategy, improving feature representation while maintaining the lightweightness of the model, thus enhancing performance. Testing CMRF-based TinyU-Net on cost-effective medical image segmentation datasets demonstrates superior performance with significantly fewer parameters and computational complexity compared to state-of-the-art methods. For instance, in the lesion segmentation of the ISIC2018 dataset, TinyU-Net is 52×, 3×, and 194× fewer parameters, respectively, while being +3.90%, +3.65%, and +1.05% higher IoU score than baseline U-Net, lightweight UNeXt, and high-performance TransUNet, respectively. Notably, the CMRF module exhibits adaptability, easily integrating into other networks. Experimental results suggest that TinyU-Net, with its outstanding performance, holds the potential to be implemented in limited-resource settings, thereby contributing to health equity. The code is available at https://github.com/ChenJunren-Lab/TinyU-Net.

Keywords: Medical image segmentation · Lightweight neural networks · Multi-receptive fields · Health equity

1 Introduction

Achieving health equity necessitates advancements in medical technology alongside concerted efforts to ensure their accessibility across diverse healthcare settings [1]. In the field of medical image segmentation, where sophisticated models like vision transformer (ViT) [2] and U-Net family [3,4,5] have shown remarkable efficacy, their impressive performance is accompanied by a heavy burden of parameters and computations, constraining its broad adoption and implementation in limited-resource settings [6]. Hence, studies that focus solely on enhancing segmentation performance, without considering resource constraints, fail to ensure equitable access to the benefits of automatic medical image segmentation across all healthcare facilities. This issue is particularly critical in underserved regions with restricted access to computational resources, ultimately leading to health inequities. Indeed, lightweight models offer a solution that aligns with health equity, ensuring affordable universal accessibility.

Recently, lightweight medical image segmentation has garnered significant attention due to its applicability in limited-resource settings. UNeXt [7] catalyzed this interest by multilayer perceptron (MLP) and depthwise separable convolution (DSC) [8,9], resulting in parameters and floating-point operations (FLOPs) suitable for limited-resource settings. ConvUNeXt [10] further advanced U-Net by integrating lightweight attention mechanisms and utilizing large convolutional kernels, reducing parameters while maintaining segmentation superiority. U-Lite [11], employing axial depthwise convolutions, expands the model's receptive field while easing computational burdens to suit limited-resource settings. CMUNeXt [12] harnesses large kernels and inverted bottleneck designs to seamlessly integrate distant spatial and location information, effectively extracting global contextual cues for swift and precise diagnostic assistance in real-world scenarios. FBSNet [13], featuring a dual-branch structure, captures extensive receptive field information and establishes local pixel dependencies to preserve details, facilitating real-time semantic segmentation deployable on edge devices. Despite offering promising avenues for addressing global healthcare disparities, existing lightweight networks often struggle to surpass current state-of-the-art models, owing to the reduced parameter counts and computational complexity resulting in inadequate feature representation. Therefore, *there is a greedy demand for a model that is lighter yet better in medical image segmentation.*

How to chase high-performance while being lightweight? Modern feature extraction modules based on multi-receptive fields enrich the complexity of the model, exemplified by techniques like feature pyramid [14] and parallel multi-group convolution [15], which have demonstrated notable efficacy in enhancing performance, thereby serving as a source of motivation. Nevertheless, the implementation of these multi-receptive field techniques inevitably escalates costs, rendering the models less lightweight. This is detrimental to clinical practice in limited-resource settings, ultimately challenging global health equity [16]. To comprehensively address this challenge, we propose a novel Cascade Multi-Receptive Fields (CMRF) module. Specifically, the CMRF module utilizes the redundant information across multiple channels in the feature map, exploring diverse receptive fields in the feature map through a cost-friendly cascading strategy. It enhances feature representation by fusing information from various receptive fields in a layer while maintaining a lightweight design, thereby improving performance. Further, building upon the CMRF module, we have constructed the lighter yet better TinyU-Net, a novel lightweight U-shaped network for medical image segmentation.

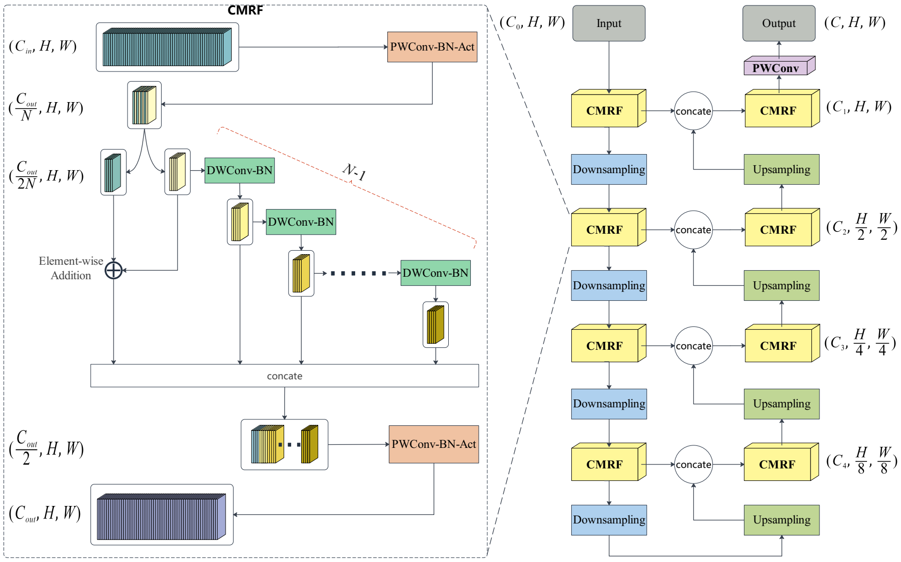

Fig. 1. Details of the CMRF (left part of the figure) and the architecture of the TinyU-Net (right part of the figure). The CMRF module fuses information from various receptive fields using a cost-friendly cascading strategy, aiming to uphold lightweight design while improving feature representation. The building modules, rooted in CMRF, constitute the architecture of our TinyU-Net, which adopts the U-shaped framework.

In summary, our contributions are: (1) Proposing the CMRF module, improving feature representation by fusing information from multi-receptive fields in a layer through a cost-friendly cascading strategy. (2) When applying the proposed CMRF to other models, segmentation performance consistently improves while reducing both parameters and computational complexity. (3) Proposing TinyU-Net, boasting a mere 0.48M parameters and small computational complexity, yet yielding exceptional performance for medical image segmentation.

2 Method

Fig. 1 illustrates the specifics of our proposed CMRF and outlines the architecture of TinyU-Net with the CMRF serving as its foundational building module. We begin by introducing CMRF (see Section 2.1), a novel lightweight module facilitating feature representation from multi-receptive fields through a cost-friendly cascading strategy. Next, we introduce TinyU-Net (see Section 2.2), a tiny variant of U-Net based on CMRF designed to optimize lightweight medical image segmentation models without intricate embellishments.

2.1 CMRF

As shown in the left part of Fig. 1, our proposed CMRF module adeptly incorporates depthwise convolution (DWConv) and pointwise convolution (PWConv) from DSC [8,9]. To make the CMRF module lightweight, we reduced the channels in the intermediate layers compared to the number of input and output channels [17,18]. Given the input feature map denoted as $X_{in} \in \mathbb{R}^{C_{in} \times H \times W}$, we utilize the PWConv-BN-Act block to mine the feature information and yield feature map $X' \in \mathbb{R}^{\frac{C_{out}}{N} \times H \times W}$, while regulating the output channel count. Where BN and Act represent batch normalization and non-linear activation, respectively. In this study, GELU [19] is adopted as the activation function. Furthermore, inspired by lightweight modules such as Ghost [20] and PConv [21] to further optimize the CMRF module for lighter weight, we utilize the information redundancy across multiple channels from the feature map in a cost-efficient way. Specifically, starting from the leftmost channel of the feature map $X' \in \mathbb{R}^{\frac{C_{out}}{N} \times H \times W}$ and numbering from 1 to $\frac{C_{out}}{N}$, the feature map $X' \in \mathbb{R}^{\frac{C_{out}}{N} \times H \times W}$ is divided into $X'_{odd} \in \mathbb{R}^{\frac{C_{out}}{2N} \times H \times W}$ and $X'_{even} \in \mathbb{R}^{\frac{C_{out}}{2N} \times H \times W}$ based on the parity of channel numbers. These are subjected to linear operations and cascade operations, respectively. On the one hand, to enhance the richness of feature information, we fusion $X'_{odd} \in \mathbb{R}^{\frac{C_{out}}{2N} \times H \times W}$ with $X'_{even} \in \mathbb{R}^{\frac{C_{out}}{2N} \times H \times W}$ through element-wise addition to yield $X''_{linear} \in \mathbb{R}^{\frac{C_{out}}{2N} \times H \times W}$, drawing inspiration from the mixup [22,23] data augmentation. On the other hand, we employ a cascade strategy on $X'_{even} \in \mathbb{R}^{\frac{C_{out}}{2N} \times H \times W}$, leveraging DWConv-BN block as a component, of which there are $N-1$, to mining information with various receptive field, while retaining the output of each DWConv-BN block. Where DWConv represents depthwise convolution with a convolutional kernel size of 3, where both the input and output channel quantities are $\frac{C_{out}}{2N}$. To utilize the results of the aforementioned process and enrich feature representation, we concatenate the feature maps along the channel direction, yielding the feature map $X''' \in \mathbb{R}^{\frac{C_{out}}{2} \times H \times W}$. Furthermore, considering the need for a lightweight design in the CMRF module, to fully and efficiently leverage the information from channels with various receptive field information, we further append a PWConv-BN-Act block to yield a feature map $X_{out} \in \mathbb{R}^{C_{out} \times H \times W}$ to facilitate the fusion of information from multi-receptive fields while regulating the output channel count.

2.2 TinyU-Net

As shown in the right part of Fig. 1, given our novel CMRF as the crucial building module, we further propose TinyU-Net, a tiny variant of the U-Net family to make medical image segmentation lighter and better. We aim to keep the CNN-based architecture as simple as possible, without bells and whistles, to make it cost-friendly in general. Similar to the U-Net framework, the TinyU-Net architecture comprises four CMRF-Downsampling blocks functioning as the encoder, four Upsampling-concat-CMRF blocks serving as the decoder, four skip connections, and a PWConv acting as the category adjuster, collectively constituting a U-shaped network. Note that the Downsampling operator employs a max-pooling in which size and stride are both set to two, for downsampling by 2×. The Upsampling operator uses the bicubic interpolation algorithm for upsampling by 2×. Considering lightweight design principles, we utilize a PWConv to adjust the output channels to accommodate the segmentation categories.

In this study, we referred to the output outcomes of individual layers in the standard U-Net. Subsequently, we determined the channel numbers for the output feature maps of the CMRF as $C_1 = 64$, $C_2 = 128$, $C_3 = 256$, and $C_4 = 512$ from the topmost to the bottommost layer. To summarize, when provided with the input $Y_{in} \in \mathbb{R}^{C_0 \times H \times W}$, the corresponding output is denoted as $Y_{out} \in \mathbb{R}^{C \times H \times W}$ for TinyU-Net. Where $\left\{C_0, C, H, W, \frac{H}{2}, \frac{W}{2}, \frac{H}{4}, \frac{W}{4}, \frac{H}{8}, \frac{W}{8}\right\} \in \mathbb{N}^{+}$ and C represents the number of categories of segmentation including background.

3 Experiments

Datasets. Given our goal of promoting and applying our work in medical institutions with limited resources, such as community hospitals, we chose ISIC2018 [24] and NCP [25] datasets for verifying our methods. The ISIC2018 dataset consists of images with skin disease lesions. Our experiments were conducted using the officially partitioned data, which consists of 2594 training images, 100 validation images, and 1000 test images. The NCP lesion segmentation dataset comprises CT slice images from the China Consortium of Chest CT Image Investigation (CC-CCII). A total of 750 CT slices from 150 COVID-19 patients were manually segmented into background (BG), lung field (LF), ground-glass opacity (GGO), and consolidation (CL). In the experiments, we divided the data into training, validation, and test sets with a ratio of 6:2:2.

Implementation details. We implemented the TinyU-Net model on an NVIDIA 4090 GPU with 24 GB of memory using the PyTorch framework. We utilized the Adam optimizer with a learning rate of 1×10^{-4} and a first-order momentum decay rate of 0.9. Additionally, we employed a cosine annealing learning rate scheduler with a minimum learning rate set to 1×10^{-6}. The models were trained using the sum of cross entropy and dice loss as the final loss function. The input image resolution was set to 256×256 (i.e., $H = W = 256$), and training was stopped after 300 iterations. We conducted experiments on the same dataset using the provided source code by the authors, applying data augmentation strategies and ceasing data augmentation after 180 epochs. Furthermore, in our experiments, we designated the initial input channel as 3 (i.e., $C_0 = 3$). Notably, in our experiments, configuring the number of DWConv-BN blocks to 7 (i.e., $N = 8$) yields optimal performance (see Section 4 for ablation study).

Performance comparison. We defined whether a model was lightweight based on the parameter count, with 5M parameters as the threshold. We compared our

TinyU-Net with both currently popular lightweight models [7,10,11,12,13] and the non-lightweight state-of-the-art models [3,4,5,26,27]. Note that we evaluated segmentation performance using metrics such as Intersection over Union (IoU) and Dice. Furthermore, we provided comparisons of models in the number of parameters (in M) and FLOPs (in G) which can represent computational complexity.

Ablation study. To further validate the effectiveness of the CMRF module, we directly replaced the feature extraction blocks of the encoder and decoder in models [12,26]. Additionally, we performed ablation experiments to evaluate the impact of varying the number of DWConv-BN blocks.

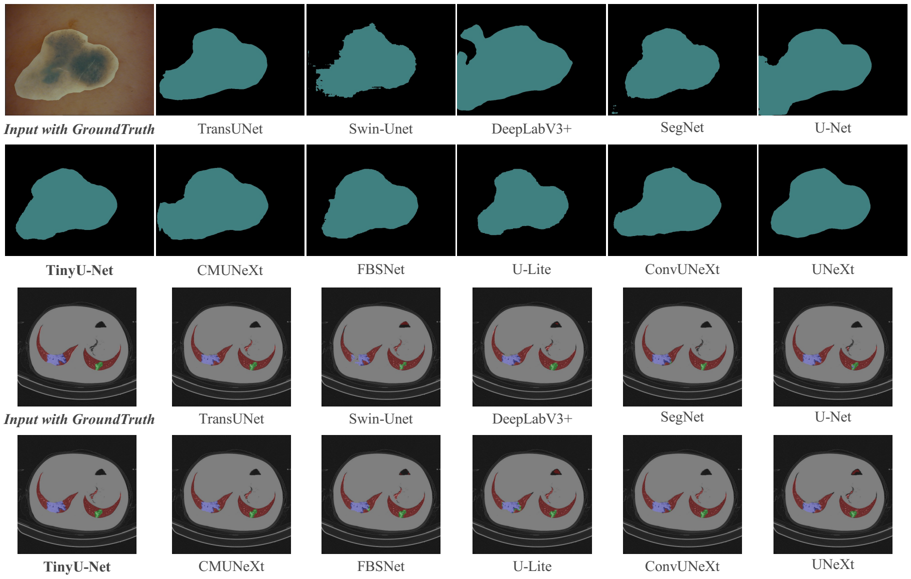

Fig. 2. Comparative qualitative results on ISIC2018 (top two rows) and NCP (bottom two rows) datasets.

4 Results and Discussion

Comparative quantitative results. Tables 1 and 2 display the experimental findings concerning the ISIC2018 and NCP datasets, respectively. Our proposed TinyU-Net achieved the highest mean IoU (mIoU) and mean Dice (mDice) scores with the fewest parameters. Specifically, TinyU-Net demonstrates a noteworthy improvement compare to baseline U-Net, achieving a +2.80% increase in mIoU scores on the ISIC2018 dataset and a +1.49% enhancement in mIoU scores on the NCP dataset. This achievement is realized by utilizing parameters that

Table 1. Comparative quantitative results on ISIC2018 dataset.

Model	Params (M)	FLOPs (G)	IoU (%) Mean	IoU (%) Lesion	IoU (%) BG	Dice (%) Mean	Dice (%) Lesion	Dice (%) BG
U-Net [3]	24.89	112.91	82.69	76.14	89.23	90.38	86.45	94.31
SegNet [26]	29.44	80.34	83.64	77.53	89.74	91.06	87.35	94.59
DeepLabV3+ [27]	5.81	13.22	83.43	77.35	89.50	91.02	87.23	94.46
TransUNet [4]	93.23	64.48	84.66	78.99	90.32	91.75	88.26	94.91
Swin-Unet [5]	27.15	15.46	83.39	77.28	89.49	90.98	87.19	94.45
UNeXt [7]	1.47	1.15	82.85	76.39	89.31	90.50	86.61	94.35
ConvUNeXt [10]	3.51	14.51	83.96	77.95	89.96	91.24	87.61	94.71
U-Lite [11]	0.88	1.52	84.13	78.12	90.13	91.29	87.72	94.81
CMUNeXt [12]	3.15	14.84	84.62	78.82	90.42	91.62	88.16	94.97
FBSNet [13]	0.60	5.76	84.03	78.12	89.93	91.35	87.72	94.70
TinyU-Net	0.48	3.33	85.49	80.04	90.94	92.18	88.91	95.25

Table 2. Comparative quantitative results on NCP dataset. GL, GGO, LF, and BG denote consolidation, ground-glass opacity, lung field, and background, respectively.

Model	Params (M)	FLOPs (G)	IoU (%) Mean	IoU (%) CL	IoU (%) GGO	IoU (%) LF	IoU (%) BG	Dice (%) Mean	Dice (%) CL	Dice (%) GGO	Dice (%) LF	Dice (%) BG
U-Net [3]	24.89	112.93	78.78	65.21	56.68	93.65	99.58	86.95	78.94	72.35	96.72	99.79
SegNet [26]	29.45	80.49	76.04	63.40	48.44	92.75	99.57	84.73	77.60	65.27	96.24	99.79
DeepLabV3+ [27]	5.81	13.22	77.28	64.73	52.41	92.52	99.46	85.80	78.59	68.78	96.11	99.73
TransUNet [4]	93.23	64.51	79.50	65.60	58.98	93.82	99.58	87.50	79.22	74.19	96.81	99.79
Swin-Unet [5]	27.15	15.46	72.98	54.74	46.33	91.48	99.35	82.32	70.75	63.32	95.55	99.67
UNeXt [7]	1.47	1.15	76.33	62.19	51.03	92.62	99.47	85.05	76.69	67.58	96.17	99.74
ConvUNeXt [10]	3.51	14.52	79.36	66.97	57.18	93.70	99.59	87.38	80.22	72.76	96.75	99.80
U-Lite [11]	0.88	1.52	75.66	60.00	50.52	92.61	99.51	84.51	75.00	67.13	96.16	99.75
CMUNeXt [12]	3.15	14.84	79.55	67.44	57.44	93.73	99.59	87.52	80.55	72.97	96.77	99.80
FBSNet [13]	0.61	5.80	77.92	64.76	53.83	93.48	99.59	86.26	78.61	69.99	96.63	99.79
TinyU-Net	0.48	3.34	80.27	68.75	58.80	93.93	99.59	88.05	81.48	74.05	96.87	99.80

are 52× smaller and showcasing 34× fewer FLOPs compared to the baseline U-Net. The small number of parameters and FLOPs in TinyU-Net stem from the lightweight design of our proposed CMRF module and the naive U-Net-like architecture of TinyU-Net. Notably, despite UNeXt standing out with 1.15 GFLOPs among the compared lightweight models, its performance is suboptimal. In light of this phenomenon, we believe that the model with low computational complexity (i.e., low FLOPs) might encounter challenges in attaining optimal performance. As quantitatively shown in Table 1, the non-lightweight TransUNet ranks second to TinyU-Net, yet its parameters and FLOPs are 194× and 19× greater than those of TinyU-Net, respectively. Furthermore, the results in Table 1 highlight that TinyU-Net achieves an impressive 80.04% in IoU score for skin lesion segmentation. Specifically, TinyU-Net demonstrates a +3.9% improvement in IoU score compared to the baseline U-Net and achieves

Table 3. Ablation results (IoU (%)) for CMRF on ISIC2018 and NCP datasets. GL, GGO, and LF denote consolidation, ground-glass opacity, and lung field, respectively.

Model	Feature extraction block	Params (M)	FLOPs (G)	ISIC2018 Lesion	NCP CL	NCP GGO	NCP LF
SegNet [26]	Original	29.44	80.34	77.53	63.40	48.44	92.75
	CMRF	**0.64**	**7.02**	**78.42**	**63.66**	**50.60**	**92.80**
CMUNeXt [12]	Original	3.15	14.84	78.82	67.44	57.44	93.73
	CMRF	**1.97**	**12.23**	**79.16**	**67.48**	**57.58**	**93.74**

Table 4. Ablation results (mIoU (%)) for the number of DWConv-BN blocks on NCP datasets.

N	1	2	4	8	16
mIoU (%)	79.24	79.58	79.77	**80.27**	79.56

a +1.05% higher IoU score than TransUNet. TinyU-Net's remarkable performance improvement can be attributed to proposed CMRF module. This module enhances feature representation by mining information from multi-receptive fields in a network layer through the cascading strategy. In Table 2, it is noteworthy that the lightweight CMUNeXt outperformed the non-lightweight TransUNet and closely trailed TinyU-Net in terms of performance. Its parameters and FLOPs are 7× and 4× higher than those of TinyU-Net, respectively. We posit that a potential explanation is that models characterized by high parameters and computational complexity (i.e., high FLOPs) might not exhibit a performance advantage when confronted with a limited amount of data. Our experimental results imply that lightweight models hold promise for efficient medical image segmentation in limited-resource settings.

Comparative qualitative results. Fig. 2 offers qualitative examples, illustrating the capability of our proposed TinyU-Net to accurately delineate the segmentation of diverse lesions in both skin diseases and NCP with less dataset, and mitigating issues of under-segmentation and over-segmentation. Significantly, TinyU-Net, despite being extremely lightweight, demonstrates competitive segmentation predictions when compared to other methods.

Ablation results. As shown in Table 3, directly replacing the feature extraction blocks of encoder and decoder in SegNet [26] and CMUNeXt [12] with CMRF modules significantly reduces model parameters and FLOPs while enhancing segmentation performance. We attribute this phenomenon to the CMRF's capability to fuse information from multi-receptive fields by a cost-friendly cascading strategy, thereby improving feature representation. Furthermore, the results demonstrate that the CMRF module can easily integrate into other networks due to its customizable adjustment of input and output channel quantities. Table 4 demonstrates Tiny-UNet's optimal segmentation at $N = 8$ on NCP datasets. We think that excessively large receptive fields may pick up irrelevant background information, while overly small ones might not fully explore the whole lesion.

5 Conclusions

To pursue optimized performance while preserving lightweight characteristics in neural networks, we introduce a novel CMRF module that fuses information from multi-receptive fields in a layer based on a cost-friendly cascading strategy to improve feature representation. The CMRF proposed is general and flexible, easily extendable to other networks, in our belief. Building upon our CMRF, we further propose a lightweight TinyU-Net, a simple yet effective U-shaped network, for medical image segmentation. Our TinyU-Net achieves state-of-the-art performance with a small number of parameters and FLOPs on cost-effective medical image datasets. We believe that the proposed method can adapt well to limited-resource settings, thereby fostering global health equity.

Acknowledgments. This study was partially supported by the Science and Technology plan transfer payment project of Sichuan Province (No. 2021ZYSF007), the Sichuan University and Yibin Municipal People's Government University and City strategic cooperation special fund project (No. 2020CDYB-29), and the Key Research and Development Program of Science and Technology Department of Sichuan Province (No. 2021KJT0012-2021YFS0067).

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

References

1. Richardson, S., Lawrence, K., Schoenthaler, A.M. et al. A framework for digital health equity. npj Digit. Med. 5, 119. (2022)
2. Dosovitskiy, Alexey, et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. (2020).
3. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
4. Chen J, Lu Y, Yu Q, et al.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306. (2021)
5. Cao H, Wang Y, Chen J, et al.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: ECCV Workshops. pp. 205-218. (2022)
6. Laibacher, Tim, Tillman Weyde, and Sepehr Jalali.: M2u-net: Effective and efficient retinal vessel segmentation for real-world applications. In: CVPR workshops. pp. 115-124. (2019)
7. Valanarasu J M J, Patel V M.: Unext: Mlp-based rapid medical image segmentation network. In: MICCAI. pp. 23-33. Springer (2022)
8. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: CVPR. pp. 1251-1258. (2017)
9. Howard, Andrew, et al.: Searching for mobilenetv3. In: ICCV. pp. 1314-1324. (2019)
10. Han Z, Jian M, Wang G G.: ConvUNeXt: An efficient convolution neural network for medical image segmentation. Knowledge-Based Systems, 253: 109512. (2022)
11. Dinh, Binh-Duong, et al.: 1M parameters are enough? A lightweight CNN-based model for medical image segmentation. 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE. pp. 1279-1284. (2023)

12. Tang, Fenghe, et al.: Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. arXiv preprint arXiv:2308.01239. (2023)
13. Gao, Guangwei, et al.: FBSNet: A fast bilateral symmetrical network for real-time semantic segmentation. IEEE Transactions on Multimedia. (2022)
14. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2881-2890. (2017)
15. Szegedy, Christian, et al.: Inception-v4, inception-resnet and the impact of residual connections on learning. Proceedings of the AAAI conference on artificial intelligence. Vol. 31. No. 1. (2017)
16. Walter, J.R., Xu, S., Rogers, J.A. From lab to life: how wearable devices can improve health equity. Nat Commun 15, 123. (2024)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770-778. (2016)
18. Kadri, R., Bouaziz, B., Tmar, M., Gargouri, F.: Depthwise Separable Convolution ResNet with attention mechanism for Alzheimer's detection. In 2022 International Conference on Technology Innovations for Healthcare (ICTIH). pp. 47-52. IEEE. (2022)
19. Hendrycks, Dan, and Kevin Gimpel.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415. (2016)
20. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C.: Ghostnet: More features from cheap operations. In: CVPR. pp. 1580-1589. (2020)
21. 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: CVPR. pp. 12021-12031. (2023)
22. Zhang, Hongyi, et al.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412. (2017)
23. Ge, Zheng, et al.: Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430. (2021)
24. Codella, Noel, et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368. (2019)
25. Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., et al: Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. Cell, 181(6), 1423-1433. (2020)
26. Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481-2495. (2017)
27. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: ECCV. pp. 801-818. (2018)