DyMAS-Net: Dynamic Multi-Scale Adaptive Sampling Network for Efficient Medical Image Segmentation

Siqi Wang, Qingxue Zhao, Di Wu, Jiakang Gao, and Jun Tian *

College of Software, Nankai University, Tianjin, China
jtian@nankai.edu.cn

Abstract. Achieving high-precision medical image segmentation while maintaining computational efficiency remains a critical challenge for clinical applications. Existing methods often struggle to balance multi-scale feature fusion, lightweight design and contextual modeling, particularly for complex medical scenes with ambiguous boundaries. To address these limitations, We propose DyMAS-Net, a lightweight framework integrating multi-scale convolution, adaptive dynamic sampling, and dual attention mechanisms. Key innovations include: (i) Hierarchical Multi-Scale Convolution Block (HMCB) combining grouped depthwise convolutions with hybrid attention to capture cross-scale dependencies; (ii) Adaptive Dynamic Sampling Module (ADSM) that dynamically adjusts receptive fields through learnable position offsets and scope prediction, enabling context-aware upsampling with minimal computational overhead; (iii) Dual Attention Fusion Unit (DAFU) integrating channel-spatial attention for global context modeling and depthwise separable gating for local feature refinement. Extensive evaluations across 7 medical image segmentation tasks (breast cancer, thyroid nodules, skin lesions) show DyMAS-Net achieves state-of-the-art performance with an average Dice score of 87.19%, outperforming TransUnet and SwinUnet by 3.02% and 2.77%, respectively. Remarkably, it attains this with only 6.24M parameters and 8.87G FLOPs, 93. 3% fewer parameters than TransUnet. The framework's efficiency-accuracy balance enables practical deployment in resource-constrained environments, thus promoting health equity.

Keywords: Medical Image Segmentation · Lightweight Architecture · Multi-Scale Feature Fusion · Adaptive Dynamic Sampling.

1 Introduction

Accurate medical image segmentation is foundational for computer-aided diagnosis and image-guided interventions, playing a pivotal role in clinical diagnosis and treatment planning. However, complex medical scenarios involving multi-scale structures and ambiguous boundaries, such as tumor infiltration and low-contrast lesions in ultrasound, pose significant challenges. Current methods

* Jun Tian is the corresponding author.

struggle to balance computational efficiency with the capacity to capture intricate structural variations, particularly in resource-constrained clinical environments [20]. This limitation exacerbates health inequities, as advanced diagnostic tools remain inaccessible to underserved regions with limited computational resources [6].

Convolutional Neural Networks (CNNs), as exemplified by architectures such as U-Net [21], demonstrate strong capabilities in local feature extraction. Subsequent variants like U-Net++ [26] and Attention U-Net [16] further refine hierarchical feature fusion. However, their constrained receptive fields fundamentally limit their capacity to capture long-range dependencies, which are crucial for comprehensive contextual understanding. While recent transformer-based approaches like Swin-Unet [2] address this limitation through self-attention mechanisms, they encounter prohibitively high computational complexity when processing high-resolution medical imaging data. Recent efforts to prioritize computational efficiency include lightweight networks such as MobileNetV3 [14] and UNeXt [23], which achieve computational efficiency through depthwise separable convolutions. However, these models frequently sacrifice multi-scale feature fusion and long-range dependency modeling. To address this, EMCAD [19] introduces an efficient multi-scale convolutional attention decoder. Additionally, methods such as ConvUNeXt [13] and CMUNeXt [22] utilize large convolutional kernels but rely on fixed kernel-based interpolation methods, leading to feature misalignment during upsampling. Although dynamic upsampling methods like CARAFE [24] enhance spatial adaptation, their computational overhead remains prohibitive for real-time applications. A critical gap persists in designing models that harmonize adaptive multi-scale processing, efficient attention mechanisms, and dynamic spatial adaptation without compromising computational efficiency.

To address these challenges, we propose DyMAS-Net, a lightweight framework that integrates three synergistic components: (i) Hierarchical Multi-Scale Convolution Block (HMCB): A multi-branch architecture combining grouped depthwise convolutions with hybrid attention. By fusing features from parallel multi-scale kernels (3×3, 5×5, 7×7) and prioritizing salient features through channel shuffling, HMCB enhances cross-scale dependency modeling while maintaining computational efficiency. (ii) Adaptive Dynamic Sampling Module (ADSM): A novel upsampling mechanism that dynamically adjusts spatial sampling positions and receptive field ranges through learnable offsets. Unlike fixed-kernel interpolation methods, ADSM adaptively captures context-aware features during decoding, effectively mitigating feature misalignment in ambiguous boundary regions. (iii) Dual Attention Fusion Unit (DAFU): A dual-branch attention module integrating depthwise separable gating for local feature refinement and cascaded channel-spatial attention for global context aggregation. DAFU bridges semantic gaps between encoder and decoder features by suppressing irrelevant noise and amplifying boundary-sensitive information.

Our primary contributions are: (1) We propose HMCB, a hierarchical convolution block that optimizes multi-scale feature fusion via hybrid attention and kernel grouping, significantly enhancing cross-scale dependency modeling. (2)

We design ADSM, a dynamic sampling module that adaptively adjusts spatial sampling positions and ranges, overcoming limitations of fixed-kernel interpolation during upsampling in medical imaging. (3) We introduce DAFU, a dual attention mechanism that synergizes local feature refinement and global context modeling to resolve feature misalignment in ambiguous boundary regions. (4) We develop DyMAS-Net, a unified framework that integrates the above innovations, achieving state-of-the-art performance across diverse medical segmentation tasks while maintaining a lightweight design.

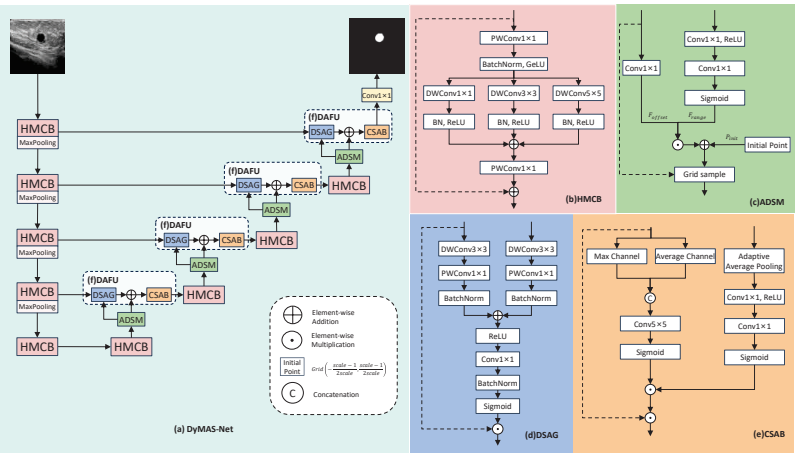

Fig. 1. Overview of DyMAS-Net Architecture. (a) DyMAS-Net, (b) Hierarchical Multi-Scale Convolution Block (HMCB), (c) Adaptive Dynamic Sampling Module (ADSM), (d) Depthwise Separable Attention Gate (DSAG), (e) Channel-Spatial Attention Block (CSAB), and (f) Dual Attention Fusion Unit (DAFU).

2 Method

In this section, we describe three core components: the Hierarchical Multi-Scale Convolution Block (HMCB), the Adaptive Dynamic Sampling Module (ADSM), and the Dual Attention Fusion Unit (DAFU). These components are systematically integrated into a lightweight framework designed for medical image segmentation named DyMAS-Net.

2.1 Hierarchical Multi-Scale Convolution Block (HMCB)

To address the critical need for efficient multi-scale feature fusion in medical image segmentation, we propose the HMCB, as illustrated in Fig. 1(b). The HMCB

employs a multi-branch architecture that processes features at multiple scales while maintaining computational efficiency. For input x, HMCB first expands the channel dimensions through a pointwise convolution ($PWConv$), followed by parallel depthwise convolutions ($DWConv$) with heterogeneous kernel sizes k_i:

$$F_i = DWConv_{k_i}(GeLU(BN(PWConv(x)))), \quad k_i \in \{3, 5, 7\} \tag{1}$$

where $PWConv$ denotes the channel expansion operation, and $DWConv_{k_i}$ represents depthwise convolution with kernel size k_i. The outputs from the parallel branches are fused through either element-wise summation or channel concatenation. And the final output combines the fused features with a residual connection to ensure stable gradient flow:

$$y = PWConv(\bigoplus_{i=1}^{3} F_i) + x \tag{2}$$

where $\bigoplus$ denotes the fusion operation, which can be either addition or concatenation depending on the configuration, and $PWConv$ compresses the channels back to the original dimension. The HMCB achieves efficient multi-scale feature extraction by leveraging grouped depthwise convolutions and residual connections, substantially reducing computational complexity compared to standard multi-scale approaches while maintaining the ability to capture features at different scales.

2.2 Adaptive Dynamic Sampling Module (ADSM)

ADSM (Fig. 1(c)) is a dynamic point-wise sampling mechanism, designed to dynamically adjust spatial sampling positions and receptive field ranges during the upsampling process. Unlike traditional fixed-kernel interpolation methods, ADSM employs learnable offsets to adaptively capture context-aware features, effectively mitigating feature misalignment in ambiguous boundary regions. Given an input feature map $x \in \mathbb{R}^{C \times H \times W}$, where C denotes the number of channels, with H and W representing the height and width of the feature map, respectively. ADSM first generates a set of learnable offsets $F_{offset} \in \mathbb{R}^{2G \times H \times W}$ through a convolutional layer, where G is the number of groups requiring manual configuration. The generated offsets F_{offset} are then constrained by a range predictor, which is implemented as a lightweight network, to generate spatial masks that adaptively regulate the offset magnitudes. This process can be formulated as:

$$F_{offset} = Conv(x), \quad F_{range} = \sigma(Conv(ReLU(Conv(x))))) \tag{3}$$

where σ is the sigmoid activation function. The final sampling positions are computed by combining the initial positions p_{init} with the modulated offsets. The feature map is then sampled based on the adjusted positions:

$$y = GridSample(x, p_{init} + F_{offset} \odot F_{range}) \tag{4}$$

where *GridSample* performs bilinear interpolation based on the calculated offsets, and $\odot$ denotes element-wise multiplication. This adaptive sampling mechanism allows ADSM to dynamically adjust to the local structure of the feature map, enhancing the model's ability to capture fine-grained details and reducing feature misalignment during upsampling.

2.3 Dual Attention Fusion Unit (DAFU)

DAFU (Fig. 1(f)) integrates two attention mechanisms to refine feature representations: Depthwise Separable Attention Gate (DSAG) and Channel-Spatial Attention Block (CSAB). The DSAG(Fig. 1(d)) operates on decoder features g and encoder features x, and selectively refines the features based on their relevance. The final output of DSAG is obtained by applying the attention weights to the encoder feature:

$$x' = x \cdot \sigma(W_g g + W_x x) \tag{5}$$

where W_g and W_x are learned filters, and σ denotes the Sigmoid function. The CSAB (Fig. 1(e)) jointly models channel and spatial contexts: channel attention weights are generated via adaptive average pooling and convolutional layers, while spatial attention combines max and average pooling features, with their product enhancing discriminative boundaries and suppressing noise. This dual mechanism resolves feature misalignment in ambiguous regions.

2.4 Overall Architecture

DyMAS-Net (Fig. 1(a)) utilizes an encoder-decoder architecture, where the encoder progressively extracts features at multiple scales using the HMCB, and the decoder upsamples the feature maps using ADSM. Skip connections are established between corresponding encoder and decoder layers to preserve spatial information. The encoder consists of multiple HMCB layers, each followed by max pooling, while the decoder consists of upsampling layers and fusion blocks (DSAG, HMCB, and CSAB) to refine the segmentation outputs.

3 Experiments and Results

3.1 Datasets

To validate the effectiveness of our method, we conducted experiments on 7 medical image segmentation datasets covering three clinical scenarios:

Breast Ultrasound Image Segmentation. Three breast cancer ultrasound datasets were used: BUSI [1] (437 benign and 210 malignant images), Breast-Lesions-USG [17] (256 lesion images), and STU [27] (42 lesion images).

Thyroid Ultrasound Image Segmentation. We employed two thyroid nodule ultrasound datasets: DDTI [18] containing 637 lesion images and TN3K [9] with 3,493 lesion images.

Table 1. Performance comparison of different methods on medical image segmentation tasks using Dice coefficient.

Methods	Params(M)	FLOPs(G)	Breast Cancer			Thyroid Lesion		Skin Lesion		Avg.
			BUSI	USG	STU	DDTI	TN3K	ISIC17	ISIC18	
CA-Net [10]	2.79	5.99	73.02	71.03	77.91	87.20	81.99	79.21	82.46	78.97
UNeXt [23]	1.47	0.57	72.14	67.39	83.53	86.39	77.53	86.89	88.89	80.39
CE-Net [11]	29.00	8.92	74.71	67.42	82.85	87.99	82.84	87.78	88.55	81.73
CPFNet [8]	30.65	8.07	77.04	72.40	79.75	87.35	83.52	86.83	87.95	82.12
CMUNeXt [22]	3.15	7.40	76.52	76.41	88.22	88.33	82.54	83.15	87.28	83.21
AttUNet [16]	34.87	66.64	75.62	77.92	86.98	89.48	83.70	84.86	86.79	83.62
UNet [21]	31.03	54.74	74.91	77.98	87.56	89.23	83.21	85.62	87.51	83.72
DSEUNet [15]	69.07	53.82	78.02	79.99	84.40	90.15	85.21	83.22	87.32	84.04
TransUnet [5]	93.23	32.24	**80.81**	85.31	90.33	90.28	86.34	76.76	79.35	84.17
SwinUnet [2]	27.15	5.92	78.73	80.61	81.73	90.13	84.96	86.83	87.95	84.42
TransFuse [25]	26.18	11.53	78.39	74.11	86.97	89.89	85.43	86.89	89.69	84.48
NUNet [4]	77.05	40.18	78.74	76.23	82.24	90.18	84.96	88.96	90.22	84.50
LGANet [12]	29.11	5.65	80.55	81.37	80.47	**91.15**	**86.44**	86.69	89.36	85.15
CARAFE [24]	8.17	14.78	76.41	79.97	88.50	90.87	85.59	87.04	87.76	85.16
AAUNet [3]	221.60	330.79	80.52	80.34	84.42	90.83	85.06	**89.21**	**90.44**	85.83
DyMAS-Net(Ours)	6.24	8.87	78.45	**85.34**	**90.76**	91.00	86.18	89.17	89.42	**87.19**

Skin Disease Dermoscopy Image Segmentation. Two ISIC challenge for skin disease analysis datasets were selected: ISIC2017 [7] (2,750 lesion images), and ISIC2018 [6] (2,594 images).

These datasets originate from diverse clinical centers and exhibit significant variations in image characteristics and lesion morphology, enabling comprehensive evaluation of segmentation robustness.

3.2 Implementation Details

We implemented our model using PyTorch 1.11.0 on an NVIDIA Tesla V100-SXM2-32GB GPU. All input images were resized to 256×256 pixels to ensure consistency. Data augmentation techniques including random rotation ($\pm 15°$) and horizontal/vertical flipping were applied during training. A hybrid loss combining Binary Cross-Entropy (BCE) and Dice coefficient was employed as the final loss function. The model was trained for 250 epochs using the Adam optimizer with an initial learning rate of 1×10^{-4} and a batch size of 32. Training and validation sets were split in a 4:1 ratio. Segmentation performance was evaluated using Dice coefficient, with additional comparisons of model complexity through parameter counts (M) and FLOPs (G).

3.3 Performance Comparison

We classified the compared models into CNN-based and Transformer-based architectures, with further distinction between lightweight and non-lightweight models. As shown in Table 1, our DyMAS-Net achieves state-of-the-art performance with only 6.24M parameters and 8.87G FLOPs, demonstrating superior efficiency-accuracy balance compared to both lightweight and non-lightweight counterparts.

The CNN-based models include the classic UNet[21] and its variants. Among lightweight designs, UNeXt [23] achieves 80.39% average Dice, while CMUNeXt [22] reaches 83.21%. Notably, our model surpasses CARAFE [24], a dynamic upsampling method with 8.17M parameters and 14.78G FLOPs, by 2.03% in average Dice while using 23.62% fewer parameters. Compared to the non-lightweight DSEUNet [15], DyMAS-Net improves segmentation accuracy by 3.15% with only 9.03% of its parameters.

For Transformer-based approaches, hybrid architectures like TransFuse [25] achieve 84.48% Dice, while pure Transformer models such as SwinUnet [2] attain 84.42%. Remarkably, DyMAS-Net outperforms the heavyweight TransUnet [5] by 3.02% Dice using merely 6.7% of its parameters. Our model outperforms LGANet [12] and AAUNet [3], both of which achieve suboptimal segmentation performance despite utilizing substantially higher computational cost. This advancement demonstrates a remarkable efficiency-accuracy trade-off, making it more practical for resource-constrained scenarios. Fig. 2 visually confirms our model's superior boundary details preservation capability compared to existing methods.

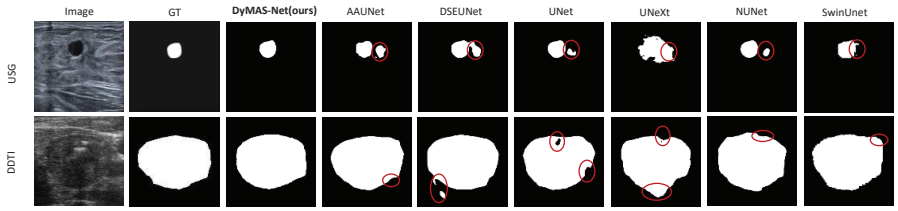

Fig. 2. Visual comparison of segmentation results by representative methods across different datasets.

3.4 Ablation Study

To evaluate the contribution of each proposed component, we conducted comprehensive ablation studies on five medical image segmentation datasets (BUSI, USG, STU, DDTI, TN3K) using different combinations of our key modules. As detailed in Table 2, the baseline configuration with only HMCB achieves competitive performance with fewer parameters, establishing an efficient foundation for multi-scale feature extraction. When using ADSM alone, the model requires significantly higher computational resources for marginal accuracy gains, indicating that dynamic sampling without multi-scale context modeling leads to suboptimal efficiency. Similarly, the standalone DAFU configuration exhibits high parameter cost with limited performance improvements, emphasizing the necessity of synergistic module integration.

The pairwise combinations reveal critical synergies: integrating HMCB with DAFU improves accuracy by 1.44% while maintaining parameter efficiency, validating the complementary nature of multi-scale convolution and attention-based feature refinement. Conversely, the ADSM+DAFU combination achieves higher accuracy but incurs more FLOPs, underscoring HMCB's role in balancing computational complexity. The complete DyMAS-Net configuration achieves optimal performance with only 6.24M parameters, demonstrating a 2.80% improvement in accuracy over the baseline HMCB while using approximately 70% fewer parameters than ADSM / DAFU configurations alone.

This evidences three critical design insights: 1) HMCB's hierarchical multi-scale processing forms an efficient backbone. 2) ADSM contributes the most significant accuracy gains through adaptive dynamic spatial sampling, particularly enhancing boundary delineation; 3) DAFU bridges semantic gaps between encoder-decoder features, improving small lesion segmentation through dual attention mechanisms. The progressive performance gains confirm that our modular design achieves effective efficiency-accuracy trade-offs through complementary feature enhancement pathways.

Table 2. Ablation studies of key components (HMCB, ADSM, DAFU).

HMCB	ADSM	DAFU	Params (M)	FLOPs (G)	Avg. Dice
✓	×	×	5.61	7.48	83.55
×	✓	×	21.01	13.05	83.74
×	×	✓	21.17	14.17	83.32
✓	×	✓	5.89	7.98	84.99
×	✓	✓	21.39	14.31	85.11
✓	✓	✓	6.24	8.87	**86.35**

4 Conclusions

This paper introduces DyMAS-Net, a lightweight framework that effectively balances segmentation accuracy and computational efficiency for medical imaging analysis. By integrating hierarchical multi-scale convolution blocks, adaptive dynamic sampling mechanisms, and dual attention fusion unit, our approach effectively handles ambiguous boundaries and multi-scale structural variations in complex clinical scenarios. Extensive evaluations across diverse medical imaging datasets demonstrate that DyMAS-Net achieves competitive performance compared to existing approaches while maintaining significantly reduced computational demands, making it particularly suitable for resource-constrained clinical deployments. Despite these strengths, our current evaluations primarily focus on ultrasound and dermoscopy images, suggesting a need for further validation on a broader range of medical imaging modalities, and for systematic investigation of its robustness under real-world noise and varied protocols. Future work will

explore hybrid architectures to further enhance segmentation capabilities while maintaining operational efficiency. We believe this work advances accessible diagnostic solutions, contributing to equitable healthcare delivery in underserved regions through efficient medical imaging technologies.

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

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)
2. 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)
3. Chen, G., Li, L., Dai, Y., Zhang, J., Yap, M.H.: Aau-net: an adaptive attention u-net for breast lesions segmentation in ultrasound images. IEEE Transactions on Medical Imaging **42**(5), 1289–1300 (2022)
4. Chen, G., Li, L., Zhang, J., Dai, Y.: Rethinking the unpretentious u-net for medical ultrasound image segmentation. Pattern Recognition **142**, 109728 (2023)
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. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., 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)
7. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: 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: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). pp. 168–172. IEEE (2018)
8. Feng, S., Zhao, H., Shi, F., Cheng, X., Wang, M., Ma, Y., Xiang, D., Zhu, W., Chen, X.: Cpfnet: Context pyramid fusion network for medical image segmentation. IEEE transactions on medical imaging **39**(10), 3008–3018 (2020)
9. Gong, H., Chen, J., Chen, G., Li, H., Li, G., Chen, F.: Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. Computers in biology and medicine **155**, 106389 (2023)
10. Gu, R., Wang, G., Song, T., Huang, R., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T., Zhang, S.: Ca-net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. IEEE transactions on medical imaging **40**(2), 699–711 (2020)
11. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: Ce-net: Context encoder network for 2d medical image segmentation. IEEE transactions on medical imaging **38**(10), 2281–2292 (2019)
12. Guo, Q., Fang, X., Wang, L., Zhang, E., Liu, Z.: Lganet: Local-global augmentation network for skin lesion segmentation. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2023)

13. Han, Z., Jian, M., Wang, G.G.: Convunext: An efficient convolution neural network for medical image segmentation. Knowledge-based systems **253**, 109512 (2022)
14. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1314–1324 (2019)
15. Li, J., Wang, J., Lin, F., Wu, W., Chen, Z.M., Heidari, A.A., Chen, H.: Dseunet: A lightweight unet for dynamic space grouping enhancement for skin lesion segmentation. Expert Systems with Applications **255**, 124544 (2024)
16. 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)
17. Pawłowska, A., Ćwierz-Pieńkowska, A., Domalik, A., Jaguś, D., Kasprzak, P., Matkowski, R., Fura, Ł., Nowicki, A., Żołek, N.: Curated benchmark dataset for ultrasound based breast lesion analysis. Scientific Data **11**(1), 148 (2024)
18. Pedraza, L., Vargas, C., Narváez, F., Durán, O., Muñoz, E., Romero, E.: An open access thyroid ultrasound image database. In: 10th International symposium on medical information processing and analysis. vol. 9287, pp. 188–193. SPIE (2015)
19. Rahman, M.M., Munir, M., Marculescu, R.: Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11769–11779 (2024)
20. Richardson, S., Lawrence, K., Schoenthaler, A.M., Mann, D.: A framework for digital health equity. NPJ digital medicine **5**(1), 119 (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. Tang, F., Ding, J., Quan, Q., Wang, L., Ning, C., Zhou, S.K.: Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024)
23. 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)
24. Wang, J., Chen, K., Xu, R., Liu, Z., Loy, C.C., Lin, D.: Carafe: Content-aware reassembly of features. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3007–3016 (2019)
25. Zhang, Y., Liu, H., Hu, Q.: Transfuse: Fusing transformers and cnns 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. 14–24. Springer (2021)
26. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4. pp. 3–11. Springer (2018)
27. Zhuang, Z., Li, N., Joseph Raj, A.N., Mahesh, V.G., Qiu, S.: An rdau-net model for lesion segmentation in breast ultrasound images. PloS one **14**(8), e0221535 (2019)