Modality-Agnostic Domain Generalizable Medical Image Segmentation via Self-Coordinated Focusing

Wentao Tang[0000−0001−9345−9483], Hongmin Deng[0000−0002−4851−2051], Linbo Qing*[0000−0003−3555−0005], and Rui Tan[0009−0003−6887−322X]

School of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China
qing_lb@scu.edu.cn

Abstract. The domain gap caused by variations in image appearance is a primary obstacle to transferring medical image segmentation models from laboratory research to clinical practice. To address this issue, domain generalization techniques have been introduced to enhance the transferability of the model. However, due to the static nature of existing perception and feature extraction mechanisms, current methods still lack sufficient adaptability in practical generalization scenarios. Inspired by the highly generalizable dynamic information focusing and fusion capabilities of the human visual system, this paper proposes the Self-Coordinated Focusing Network (SCFNet), a novel model designed for modality-agnostic medical image segmentation with domain generalization ability. The model comprises two core components: Dynamic Focusing Convolution (DFC) and the Interactive Attention Fusion Decoding (IAFD) module. DFC dynamically adjusts convolutional aggregation weights and receptive-field scales by pre-perceiving target types and their spatial scale distributions, enabling input-adaptive, target-aware feature perception. Furthermore, IAFD enhances feature fusion efficiency and information interaction during decoding by decomposing semantic information and introducing interactive attention across semantic hierarchies. We evaluate SCFNet on ten datasets spanning four imaging modalities. Experimental results demonstrate that SCFNet consistently outperforms multiple state-of-the-art methods in multi-modal settings, exhibiting superior segmentation accuracy and stronger generalization capability.

Keywords: Domain Generalization · Medical Image Segmentation · Dynamic Convolution.

1 Introduction

Medical image segmentation is a critical link between imaging acquisition and clinical decision-making, assisting in disease diagnosis and treatment. Accurate delineation of early-stage tumors is particularly important for timely detection

* Corresponding author.

and improved patient outcomes [4]. However, given the complexity and inconsistent quality of medical imagery, manual segmentation is labor-intensive and prone to inter-observer variability. These challenges have created a strong clinical need for reliable automated segmentation methods [11].

The rapid development of deep learning has provided effective solutions for automatic medical image segmentation. As a data-driven approach, deep learning can learn discriminative features through training. However, due to factors such as the limited scale of medical imaging datasets and significant inter-domain shifts, the fundamental assumption that training and testing data follow the same distribution is often violated [8]. As a result, models tend to overfit the training data, leading to insufficient generalization ability and posing challenges to their translation into real-world clinical applications.

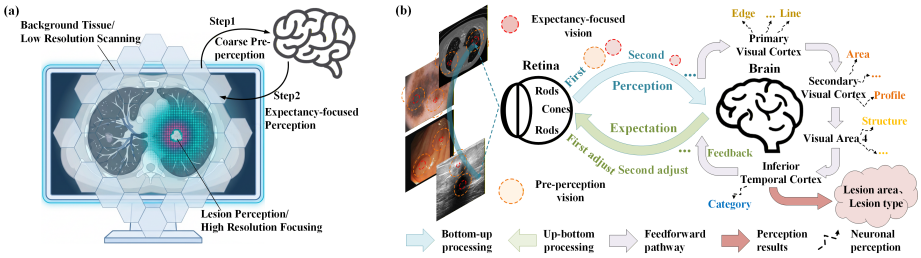

Fig. 1. The perception mechanism of human vision. (a) The focusing mechanism in human visual cognition. (b) The coordinated focusing and information processing in the human visual system.

To address this, we analyze the regulatory mechanisms underlying generalized perception of differential information and find a close correspondence to dynamic perceptual focusing in human vision. As illustrated in Fig. 1, human visual cognition operates as an iterative loop of perception, feedback, and dynamic focusing. This behavior arises from differences in receptive fields and perceptual patterns across sensory cells, together with the cognitive system's efficiency in coordinating and reallocating attention. Through this "global-to-local, interactive, and adaptive focusing" mechanism, human vision can flexibly choose appropriate perceptual strategies to recognize diverse targets across scenes.

Inspired by this, we propose the Self-Coordinated Focusing Network (SCFNet) for medical image segmentation. By leveraging the Dynamic Focusing Convolution (DFC) and Interactive Attention Fusion Decoding (IAFD) modules to integrate the principles of the dynamic focusing and information interaction mechanisms of human vision, the proposed method achieves input-adaptive dynamic perception. Extensive experimental results across diverse modalities and clinical scenarios fully demonstrate the superiority of our approach in both segmentation performance and generalization capability. Our contributions can be summarized as follows:

1. A segmentation model called SCFNet is proposed for medical image segmentation tasks. The method achieves high-performance segmentation across multiple modalities and unseen data domains by utilizing efficient dynamic focusing for feature perception and effective interactive fusion for information transfer.
2. A convolutional algorithm DFC is developed. The algorithm is based on the pre-perception of the type and spatial scale distribution characteristics of input information. By dynamically adjusting the convolution aggregation weights and receptive fields depending on the input, appropriate regulation of feature extraction during perceptual operations can be achieved.
3. An IAFD module is designed to decompose information from low-level semantic features using wavelet transformation, and facilitates efficient information integration in feature decoding through cross-level feature interaction attention and fusion of dominant information.

2 Method

Fig. 2 presents the detailed implementation of SCFNet. As illustrated in Fig. 2(a), the model adopts an overall U-shaped encoder-decoder architecture. For feature perception, SCFNet integrates DFC modules at each stage to achieve input-adaptive aggregation weights and scales. In terms of feature fusion, the IAFD module is introduced to significantly enhance information interaction efficiency during the decoding process by leveraging its efficient cross-semantic interaction capabilities.

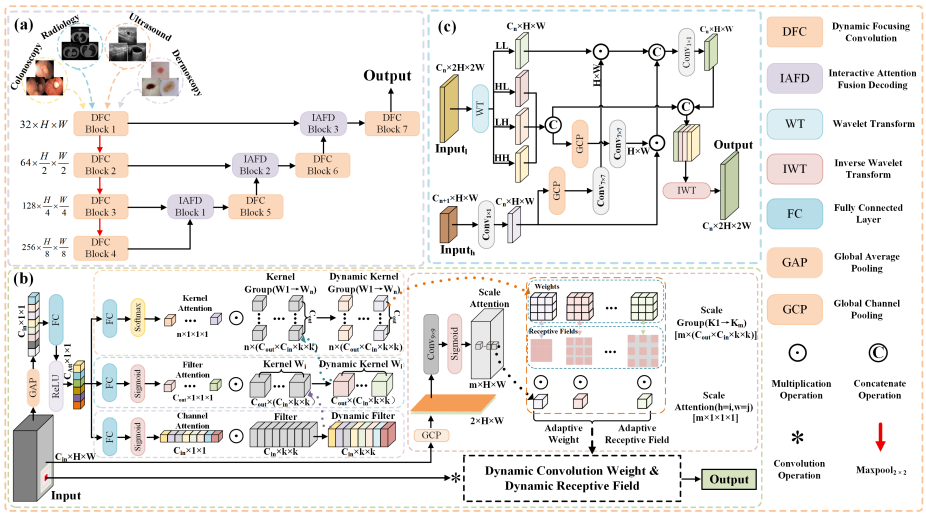

Fig. 2. Details of the self-coordinated focusing segmentation method. (a) The framework of SCFNet. (b) The working principle of DFC. (c) The details of IAFD.

2.1 Dynamic focusing convolution

Fig. 2(b) illustrates the architecture of the DFC. The entire operation consists of two main functional parts: feature focusing and scale focusing. These correspond to the perception pattern and scale adjustment in human cognitive mechanisms, enabling the dynamic adjustment of aggregation weights and receptive fields based on the inputs to enhance cross-domain generalization.

Feature focusing extends the dynamic weight generation method of classic dynamic convolution in [20], but introduces additional dynamic deployment across input channels and filter dimensions to enhance the method's performance. As shown in Equations (1)-(4), the input feature tensor X is first aggregated through a spatial global average pooling (GAP) operation. The aggregated features are then mapped through fully connected layers and activation functions, before being fed into three attention branches. Finally, the input encoded tensor F_{ce} is mapped through fully connected layers in each branch and transformed via a Softmax or Sigmoid function to generate dynamic weights for the channel, filter, and kernel groups. The entire process can be formalized as:

$$F_{ce} = \delta\left(\mathrm{FC}\left(\mathrm{GAP}\left(X\right)\right)\right), \tag{1}$$

$$CA = \sigma\left(\mathrm{FC}_{ca}\left(F_{ce}\right)\right), \tag{2}$$

$$FA = \sigma\left(\mathrm{FC}_{fa}\left(F_{ce}\right)\right), \tag{3}$$

$$KA = \mathrm{Softmax}\left(\mathrm{FC}_{ka}\left(F_{ce}\right)\right), \tag{4}$$

where FC represents the fully connected layer, δ and σ denote ReLU and Sigmoid activation functions respectively, and CA, FA, and KA represent the dynamic attention scalars across the input channels, filters, and kernel groups.

Scale focusing is inspired by the receptive-scale adjustment in human cognition. It extracts scale-related information through spatial pre-perception and then performs an appropriate scale-focusing combination at each spatial location based on the target's scale distribution. Given an input tensor X, the features are first compressed via global channel pooling (GCP) to reduce dimensionality and encode spatial scale information. The encoded feature is then passed through a 9×9 convolution to capture the spatial distribution of scales, and an activation mapping is applied to obtain the dynamic scale weights SA. The scale attention SA scalars can be expressed in Equation (5):

$$SA = \sigma\left(\mathrm{Conv}_{9\times 9}\left(\mathrm{GCP}\left(X\right)\right)\right), \tag{5}$$

where $\mathrm{Conv}_{9\times 9}$ denotes the 9×9 convolution operation, and GCP represents the combined operation of global channel max-pooling and global channel average-pooling. Based on these multi-dimensional descriptors, DFC dynamically formulates scale-specific kernels W^k, a spatially-adaptive weight composite $\mathcal{W}_{ij}$, and a dynamic receptive footprint $\mathcal{R}_{ij}(x)$ to simulate human adaptive vision. The final output y_{ij} at coordinate (i, j) is expressed as:

$$W^k = \sum_{g=1}^{n_k} KA_g^k \odot FA_g^k \odot CA_g^k \odot W_g^k, \tag{6}$$

$$\mathcal{W}_{ij} = \bigcup_{k=1}^{m} \left(SA_{ij}^{k} \cdot W^{k} \right), \tag{7}$$

$$\mathcal{R}_{ij}(x) = \bigcup_{k=1}^{m} \mathcal{N}(x, \mathbf{p}_{ij}, d_k), \tag{8}$$

$$y_{ij} = \mathcal{A}\left(\mathcal{W}_{ij}, \mathcal{R}_{ij}(x) \right), \tag{9}$$

where n_k is the number of kernel groups, W_g^k is the baseline kernel, $\odot$ denotes element-wise multiplication, m denotes scale groups, $\mathcal{N}$ represents a local sampling grid centered at $\mathbf{p}_{ij}$ with dilation rate d_k, and $\mathcal{A}$ is the aggregation operator.

2.2 Interactive Attention Fusion Decoding Module

As illustrated in Fig. 2(c), the method decomposes low-level features and establishes interactions between each decomposed component and high-level features, enabling efficient fusion of dominant information across hierarchical levels. Specifically, in the low-level branch, the input tensor X_l is first processed by a wavelet transform to decompose the information into one low-frequency component and three high-frequency components. The transformation process is expressed as follows:

$$(LL, LH, HL, HH) = \mathrm{WT}\left(X_l \right), \tag{10}$$

where WT represents the Haar wavelet transform, LL denotes the low-frequency part of the tensor, and LH, HL, and HH represent the high-frequency parts of the tensor. Subsequently, the three high-frequency components are concatenated along the channel dimension and then passed through an attention mapping consisting of global channel pooling and convolution operations to generate the low-level semantic attention tensor Att_l. The Att_l can be expressed as:

$$Att_l = \sigma\left(\mathrm{Conv}_{7\times7}\left(\mathrm{GCP}\left(\mathrm{Concat}\left(HL, LH, HH \right) \right) \right) \right), \tag{11}$$

where $\mathrm{Conv}_{7\times7}$ refers to the convolution operation with a kernel size of 7, GCP denotes the global channel pooling operation, and Concat indicates the concatenation operation along the channel dimension. In the high-level input branch, the input tensor X_h undergoes a channel compression via a 1×1 convolution. On one hand, it is passed through a mapping of channel pooling and convolution to generate the high-level semantic attention tensor. On the other hand, it is processed with the low-level semantic attention tensor to generate the high-level output tensor Y_h. The generated high-level semantic attention tensor Att_h then enters the low-level branch, where it interacts with LL to produce the low-level output tensor Y_l. The information interaction process can be formalized as:

$$Att_h = \sigma\left(\mathrm{Conv}_{7\times7}\left(\mathrm{GCP}\left(\mathrm{Conv}_{1\times1}\left(X_h \right) \right) \right) \right), \tag{12}$$

$$Y_h = \mathrm{Conv}_{1\times1}\left(X_h \right) \odot Att_l, \tag{13}$$

$$Y_l = Att_h \odot LL. \tag{14}$$

After the interaction, the features Y_l and Y_h are concatenated along the channel dimension, and then information fusion is completed solely through a 1×1 convolution to generate LL'. Finally, the inverse wavelet transform IWT, based on LL', LH, HL, and HH, maps and generates the output feature Y. The computation process can be formalized as:

$$LL' = \mathrm{Conv}_{1\times1}\left(\mathrm{Concat}\left(Y_h, Y_l\right)\right), \tag{15}$$

$$Y = \mathrm{IWT}\left(LL', LH, HL, HH\right). \tag{16}$$

3 Experimental results

3.1 Datasets

This study uses 10 datasets across four imaging modalities to evaluate both intra-domain segmentation and cross-center generalization. To explicitly simulate real-world domain shifts, we rigorously partition the datasets into seen (source) and unseen (target) clinical settings.

For the **seen clinical settings**, each source dataset is randomly split for training (70%) and testing (30%). Conversely, to strictly validate out-of-distribution robustness, models are evaluated on the **unseen clinical settings** in a zero-shot manner, utilizing weights trained exclusively on their corresponding source domains. Specifically, the cross-center modality shifts are constructed as follows: (1) **Dermoscopy:** Train on the multi-center ISIC2018 [5] (2,594 cases) and test generalization on the unseen PH2 [12] (200 cases, Pedro Hispano Hospital). (2) **Radiology:** Train on COVID19-1 [10] (1,277 cases) and test on the cross-center COVID19-2 [16] (2,535 cases, Qaem Hospital). (3) **Ultrasound:** Train on BUSI [14] (645 cases, Baheya Hospital) and test on STU [23] (42 cases, Shantou University). (4) **Colonoscopy:** Train on CVC-ClinicDB [2] (612 cases) and Kvasir [9] (1,000 cases); assess generalization against resolution-induced shifts on CVC-ColonDB [18] (380 cases) and ETIS [17] (196 cases).

3.2 Experimental settings and evaluation metrics

SCFNet is implemented using the PyTorch framework and trained on an NVIDIA RTX 3090 Ti GPU. The Adam optimizer is used with an initial learning rate of 10^{-4} and a weight decay of 10^{-8}. The training process spans a total of 300 epochs, with the learning rate scaled by a factor of 0.8 every 80 epochs. Crucially, aligning with the mathematical formulation of DFC, the number of scale groups and kernel groups are explicitly configured to $m = 3$ and $n_k = 2$, respectively.

For the dermoscopy, radiology, and colonoscopy tasks, input images are resized to 256×256 with a batch size of 16. For the ultrasound task, the image size is set to 512×512 with a batch size of 6. The standard Dice Loss is adopted as the objective function. Segmentation performance is evaluated using two widely adopted spatial overlap metrics: the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU).

Table 1. Quantitative comparison of segmentation results between SCFNet and other methods with seen clinical settings.

Method	Dermoscopy		Radiology		Ultrasound		Colonoscopy			
	ISIC2018		COVID19-1		BUSI		CVC-ClinicDB		Kvasir	
	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU
UNet[15]	90.92	83.93	79.87	71.46	78.59	68.73	89.56	82.27	87.68	79.50
R2UNet[1]	89.09	81.14	64.34	57.69	74.60	64.88	75.29	65.37	76.97	66.07
CENet[7]	92.26	86.06	77.43	69.35	82.10	72.97	89.02	81.77	85.53	76.45
CANet[6]	91.68	85.12	80.04	71.40	83.35	74.51	87.63	79.81	82.14	71.66
TransUNet[3]	92.17	86.05	82.57	74.83	81.77	73.00	91.90	85.93	89.05	81.23
DCSAUNet[19]	91.41	84.70	83.42	76.29	82.04	73.69	89.67	82.77	85.04	75.56
M2SNet[21]	92.38	86.28	82.73	75.73	81.76	72.94	93.95	89.15	88.44	80.25
MADGNet[13]	91.60	85.12	83.72	76.70	82.55	73.61	94.75	90.45	88.17	79.88
DYGLNet[22]	92.54	86.57	78.10	71.60	82.67	75.10	90.06	83.29	87.88	79.76
SCFNet(Ours)	**93.10**	**87.52**	**84.14**	**76.85**	**84.38**	**75.75**	**94.79**	**90.50**	**89.52**	**81.88**

Table 2. Quantitative comparison of segmentation results between SCFNet and other methods with unseen clinical settings.

Method	Dermoscopy		Radiology		Ultrasound		Colonoscopy			
	PH2		COVID19-2		STU		CVC-ColonDB		ETIS	
	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU
UNet[15]	91.27	84.26	77.65	69.87	69.01	58.18	76.11	66.80	70.02	61.09
R2UNet[1]	91.61	84.77	65.96	59.31	75.84	64.44	60.46	53.73	65.67	58.23
CENet[7]	91.92	85.30	75.71	67.90	88.49	80.75	79.52	70.78	74.31	66.57
CANet[6]	92.44	86.16	80.04	72.42	86.32	77.48	65.72	55.71	63.72	54.78
TransUNet[3]	92.67	86.55	81.10	73.27	88.70	80.91	76.95	68.08	77.01	68.15
DCSAUNet[19]	91.11	84.00	78.86	71.39	87.45	79.12	71.15	61.80	69.73	62.18
M2SNet[21]	92.47	86.25	78.86	71.39	83.33	73.57	76.19	66.88	71.72	63.72
MADGNet[13]	92.05	85.48	74.26	66.11	86.06	77.02	82.15	73.50	78.21	69.81
DYGLNet[22]	93.23	87.54	74.85	68.38	88.16	80.21	78.27	68.92	70.09	61.52
SCFNet(ours)	**93.31**	**87.64**	**85.12**	**77.63**	**91.24**	**84.67**	**83.09**	**74.58**	**79.98**	**71.86**

3.3 Comparison with State-of-the-art models

Tables 1 and 2 present quantitative comparisons between SCFNet and nine state-of-the-art segmentation methods across multi-modal datasets, where the bold and underlined values denote the best and second-best results, respectively. Overall, SCFNet achieves the best segmentation performance across all modalities in both seen and unseen clinical settings. Notably, while some methods perform comparably to SCFNet in seen settings, the gap becomes substantial under domain shift. For instance, MADGNet is close to SCFNet on the seen radiology data, but on the unseen dataset the DSC and IoU gaps widen to 10.86% and 11.52%, respectively. This distinct advantage in unseen clinical settings highlights the superior generalization and transferability of our approach.

Qualitative comparisons between SCFNet and nine state-of-the-art methods across multiple imaging modalities are shown in Fig. 3. Overall, SCFNet con-

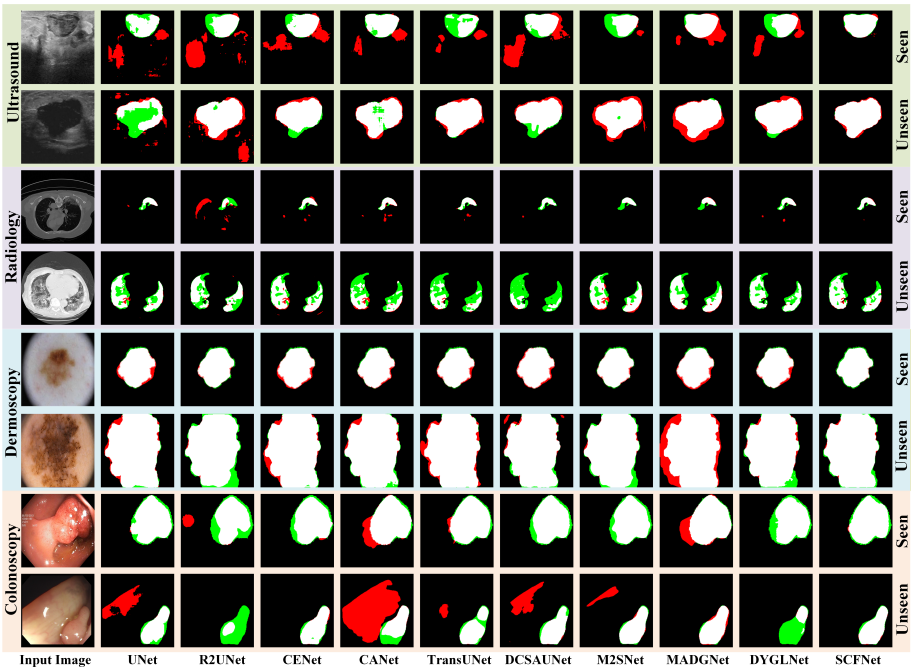

Fig. 3. Qualitative comparison of segmentation results between SCFNet and other methods. The red represents false positive and green represents false negative.

sistently produces segmentation results that are closest to the ground truth in both seen and unseen clinical settings. From a cross-modality and cross-domain generalization perspective, most existing methods suffer from the static nature of their perception mechanisms, leading to pronounced performance degradation when imaging characteristics change. In contrast, SCFNet maintains stable performance in cross-modality and cross-domain segmentation by dynamically regulating perception in an input-adaptive manner. These results demonstrate that SCFNet is more robust to variations in imaging characteristics and achieves accurate segmentation with strong generalization across multi-modal data.

4 Conclusion

In this paper, we propose SCFNet, a modality-agnostic generalized medical image segmentation model. The primary novelty of this model lies in two core modules: DFC and IAFD. DFC performs input-adaptive dynamic adjustment of the convolutional aggregation weights and the aggregation receptive field, guided by prior awareness of the input morphology and its spatial scale distribution. IAFD combines frequency decomposition with an interactive attention mechanism across semantic hierarchies, facilitating the exchange and integration of complementary information during decoder fusion. Extensive experiments across

modalities and clinical scenarios show that SCFNet achieves state-of-the-art segmentation performance. Future work will target more complex cross-domain clinical scenarios with more pronounced distribution shifts, with the goal of continuously improving the model's generalization capability and robustness in unseen domains.

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

References

1. Alom, M.Z., Yakopcic, C., Taha, T.M., Asari, V.K.: Nuclei segmentation with recurrent residual convolutional neural networks based u-net (R2U-Net). In: IEEE National Aerospace and Electronics Conference. pp. 228–233. IEEE (2018)
2. Bernal, J., Javier Sanchez, F., Fernandez-Esparrach, G., Gil, D., Rodriguez, C., Vilarino, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics **43**, 99–111 (2015)
3. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (2024)
4. Coates, A.S., Winer, E.P., Goldhirsch, A., Gelber, R.D., Gnant, M., Piccart-Gebhart, M., Thürlimann, B., Senn, H.J.: Tailoring therapies—improving the management of early breast cancer: St gallen international expert consensus on the primary therapy of early breast cancer 2015. Annals of Oncology Official Journal of the European Society for Medical Oncology **26**(8), 1533 (2015)
5. 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)
6. 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 (2021)
7. 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)
8. Hu, S., Liao, Z., Zhang, J., Xia, Y.: Domain and content adaptive convolution based multi-source domain generalization for medical image segmentation. IEEE Transactions on Medical Imaging **42**(1), 233–244 (2023)
9. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D.: Kvasir-SEG: A segmented polyp dataset. In: International Conference on Multimedia Modeling. pp. 451–462. Springer (2020)
10. Ma, J., Ge, C., Wang, Y., An, X., Gao, J., Yu, Z., Zhang, M., Liu, X., Deng, X., Cao, S., et al.: COVID-19 CT lung and infection segmentation dataset. Zenodo (2020)
11. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**(9), 60–88 (2017)

12. Mendonça, T., Ferreira, P.M., Marques, J.S., Marcal, A.R.S., Rozeira, J.: PH2 - a dermoscopic image database for research and benchmarking. In: International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 5437–5440. IEEE (2013)
13. Nam, J.H., Lee, S.C., Syazwanyl, N.S., Kim, S.J.: Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11480–11491. IEEE (2024)
14. Pawlowska, A., Karwat, P., Zolek, N.: Dataset of breast ultrasound images. Data in Brief **48** (2023)
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
16. Samant, P.: Lung CT nodule/lesion segmentation (pidata-new-names). Kaggle (2022)
17. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. International Journal of Computer-Assisted Radiology and Surgery **9**(2), 283–293 (2014)
18. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. IEEE Transactions on Medical Imaging **35**(2), 630–644 (2016)
19. Xu, Q., Ma, Z., HE, N., Duan, W.: DCSAU-Net: A deeper and more compact split-attention u-net for medical image segmentation. Computers in Biology and Medicine **154**, 106626 (2023)
20. Yang, B., Bender, G., Le, Q.V., Ngiam, J.: Condconv: Conditionally parameterized convolutions for efficient inference. In: Advances in Neural Information Processing Systems. (2019)
21. Zhao, X., Jia, H., Pang, Y., Lv, L., Tian, F., Zhang, L., Sun, W., Lu, H.: M^2SNet: Multi-scale in multi-scale subtraction network for medical image segmentation. arXiv preprint arXiv:2303.10894 (2023)
22. Zhao, Y., Wang, C., Hao, Y., Li, L., Liao, T.: DyGLNet: Hybrid global-local feature fusion with dynamic upsampling for medical image segmentation. Pattern Recognition **173**, 112792 (2026)
23. Zhuang, Z., Li, N., Raj, A.N.J., Mahesh, V.G.V., Qiu, S.: An rdau-net model for lesion segmentation in breast ultrasound images. PLOS One **14**(8), e0221535 (2019)