DPGS-Net: Dual Prior-Guided Cross-Domain Adaptive Framework for Ultrasound Image Segmentation

Weijie Zhang[1*], Lingfeng Xie[1*], Kun Zeng[2], Xiaonan Luo[3], and Yongyi Gong[1✉]

[1] School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou, China
20231010049@mail.gdufs.edu.cn, gongyongyi@gdufs.edu.cn
[2] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
[3] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guangxi, China

Abstract. Ultrasound image segmentation plays a critical role in medical-assisted diagnosis but suffers from inherent limitations, such as high noise, artifacts, and morphological diversity. Existing methods struggle to generalize with small-sample data due to feature contradictions from varying acquisition angles, limiting multi-center clinical use. To address these issues, we propose a dual prior-guided two-stage segmentation framework. In the first stage, the prior classification of small-sample data guides domain adaptation pretraining on large-scale datasets, employing dynamic class balancing to mitigate data distribution bias. The second stage features a multi-level feature fusion architecture with three core modules: First, we design a Multi-branch Convolutional Parallel Attention (MCPA) module that extracts contextual features via parallel dual attention to adaptively select multi-scale features. Next, we propose a Multi-scale Fusion Dilated Convolution (MFDC) module that enhances the encoder's capability to capture lesion boundaries across different receptive fields through hierarchical dilated convolutions. Finally, we introduce an Enhanced Feature Decoding module (EFD) in the decoder, embedding a cross-layer compensation mechanism using shallow high-resolution features to recover spatial details lost. Furthermore, we propose an interactive dual-stream architecture that bridges prior-guided classification and segmentation tasks, where complementary features are fused through cross-task attention to optimize holistic semantic consistency and robustness. Experiments on the public dataset demonstrate our method's superiority over mainstream approaches. Ablation studies validate the effectiveness of our method, providing a solution for high-precision, high-availability small-sample ultrasound image segmentation. Code is on Github: https://github.com/notchXie/DPGS-Net.

*W. Zhang and L. Xie contribute equally to this work.
✉The corresponding author is Yongyi Gong.

Keywords: Ultrasound Image · Segmentation · Deep Learning · Multi-task Learning · Pretraining and Fine-tuning

1 Introduction

Ultrasound imaging, known for its non-invasive, real-time, and cost-effective nature, is crucial for early screening, diagnosis, and treatment planning in healthcare [5,16]. Unlike CT or MRI, it does not use ionizing radiation and is portable, making it widely used in clinical settings. Early ultrasound segmentation relied on traditional methods like edge detection and manual annotations, which were slow, labor-intensive, and dependent on physician expertise. This often led to inconsistent results and limited scalability across different centers. As a result, there was a push for improved solutions. While deep learning has revolutionized medical image segmentation, ultrasound segmentation still faces several challenges: noise and artifacts caused by the equipment, tissue variability due to patient differences and imaging conditions, and the limited availability of annotated data because of privacy concerns. These issues, especially the scarcity of annotated samples, hinder model training and reduce segmentation accuracy and clinical reliability.

The encoder-decoder architecture, exemplified by U-Net [19], has become a cornerstone of medical segmentation by fusing multi-scale features through skip connections [1]. U-Net++ [29] enhanced skip connection flexibility, while attention mechanisms [17] improved target region focus. The DeepLab series [5] advanced multi-scale context modeling via dilated convolutions and spatial pyramid pooling. Transformer-based models, such as TransUNet [4], SETR [28], and Swin-Unet [3], leveraged self-attention for long-range dependencies, and UCTransNet [21] refined skip connection channel selection. However, feature contradictions from acquisition variations, complex backgrounds, and lesion diversity challenge small-sample generalization and cross-domain adaptability, as U-Net variants struggle with noise interference, as shown in Fig. 1. To address these issues, researchers have explored multi-task learning [14, 22], few-shot meta-learning [12], domain adaptation [8, 20], unsupervised adaptation [27], prior shape mapping [26], and pretraining-fine-tuning paradigms [6]. Despite these

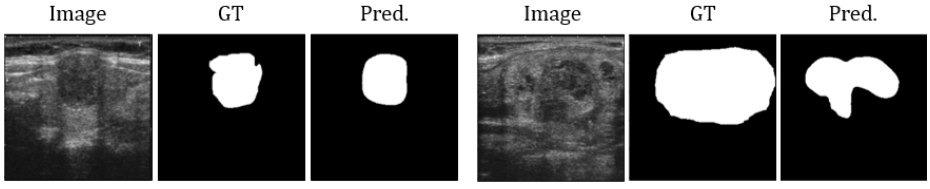

Fig. 1. The left image shows a U-Net [19] segmentation using extracted contrast features, while the right image reveals notable errors from U-Net due to feature contradictions in the ultrasound, despite using the same features.

efforts, no single method fully addresses ultrasound segmentation's complexity, underscoring the need for an integrated framework.

Thus, we proposes a dual prior-guided cross-domain adaptive framework for small-sample ultrasound image segmentation. A prior-classification-guided domain adaptation pretraining strategy addresses feature contradictions and data scarcity by leveraging large-sample datasets of the same category. A multi-level feature fusion architecture integrates MCPA, MFDC, and EFD modules to enhance feature extraction for complex backgrounds and lesion boundaries. A dual-stream feature interaction mechanism jointly optimizes prior classification and segmentation tasks, improving semantic consistency and cross-task adaptability. Experiments on the DDTI dataset demonstrate superior performance, with ablation studies validating each module's effectiveness. The contributions of this paper are summarized as follows:

- A general domain adaptation pretraining strategy is proposed, guided by prior classification knowledge and leveraging large-sample datasets of the same category to address feature contradictions and domain adaptation challenges in small-sample scenarios.
- A multi-level feature fusion segmentation network is designed, integrating MCPA, MFDC, and EFD modules to significantly enhance feature extraction and segmentation accuracy.
- A dual-stream interaction mechanism is constructed for prior classification and segmentation tasks, adaptively guiding the segmentation network through cross-task feature fusion to optimize robustness and generalization.

2 Method

We propose a Dual Prior-Guided Two-Stage Segmentation Framework (DPGS-Net). As shown in Fig. 2, Stage I employs domain adaptation pretraining, leveraging prior classification knowledge from a large-sample dataset of the same category. Stage II integrates three core modules into U-shaped segmentation network for multi-level feature fusion, enhanced by a classification network trained with prior knowledge, forming a multi-task dual-stream feature interaction for secondary guidance. Details follow below.

2.1 Prior Classification-Guided Domain Adaptation Pretraining

As shown in Fig. 2, let the source domain (large-sample dataset) be $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and the target domain (small-sample dataset) be $D_t = \{(x_j^t, y_j^t)\}_{j=1}^{N_t}$, where $N_t \ll N_s$. To address data scarcity and class imbalance in the target domain, we construct a prior classification task using surface positional prior information. Spatial distribution features $F_{pos} = E_{pos}(x_j^t)$ are extracted to generate class labels $C_t = \{c_k\}_{k=1}^{K}$, ensuring each class covers significant spatial patterns. A dynamic class-balancing strategy is employed to build the test set D_{test}^t: for each class c_k, $n_k = \alpha_k \cdot N_t$ samples are randomly sampled, with $\sum_{k=1}^{K} n_k = N_t'$

and α_k dynamically adjusted based on class distribution entropy, ensuring minority class coverage exceeds a threshold τ. Feature consistency validation removes samples with significant divergence from the source domain, yielding a balanced and representative D^t_{test}.

After obtaining D^t_{test}, we perform domain adaptation pretraining. First, a base segmentation model M_s is pretrained on the source domain dataset D_s. The pretrained model is fine-tuned on D^t_{test} by minimizing L_{seg}. Early stopping prevents overfitting, and performance is recorded after each fine-tuning session. Multiple fine-tuning experiments are conducted, and the best-performing model on D^t_{test} is selected as the domain-adapted pretrained model M_{adapt}, which adapts to the target domain distribution and provides a feature extraction foundation for Stage II.

2.2 Multi-Level Feature Fusion Segmentation Network

Multi-branch Convolutional Parallel Attention Module As shown in Fig. 4, the MCPA module adaptively extracts shallow features and fuses cross-task information via multi-scale convolutions and a dual-attention mechanism [13,23]. The input feature $F_{\text{in}} \in \mathbf{R}^{H\times W\times C}$ is processed through 1×1, 5×5, and 9×9 convolutions to extract multi-scale features F_1, F_5, and F_9. These features are used to compute channel-attention and spatial-attention features:

$$F_{\text{CA}} = \sum_{k\in\{1,5,9\}} \left(F_k \otimes \sigma(\text{MLP}(\text{GAP}(F_k)))\right), \tag{1}$$

$$F_{\text{SA}} = \sum_{k\in\{1,5,9\}} \left(F_k \otimes \sigma(\text{Conv}_{3\times3}(\text{Concat}(\text{AvgPool}(F_k), \text{MaxPool}(F_k))))\right), \tag{2}$$

where GAP is global average pooling, MLP is a multi-layer perceptron, and σ is the Sigmoid function. The concatenated feature $F_A = \text{Concat}(F_{\text{CA}}, F_{\text{SA}})$ is further processed through 3×3, 5×5, and 7×7 depthwise separable convolutions [7] to extract multi-scale features F_{DW3}, F_{DW5}, and F_{DW7}. The final output is obtained by fusing these features: $\hat{F} = F_{\text{DW3}} + F_{\text{DW5}} + F_{\text{DW7}}$.

Multi-Scale Fusion Dilated Convolution Module The MFDC module enhances blurred boundary sensitivity and multi-scale context capture via hierarchical convolutions. As shown in Fig. 3, the input feature $F_{\text{in}} \in \mathbf{R}^{H\times W\times C}$ is processed as:$F_1 = \sum_{k\in\{3,5,7\}} \text{Conv}_{k\times k}(F_{\text{in}})$, $F_2 = \sum_{r\in\{1,2,3\}} \text{DilatedConv}_r(F_1)$, $F_3 = \text{Conv}_{1\times1}(\text{Concat}(F_1, F_2))$, where DilatedConv_r denotes dilated convolution [25] with rate r. The fused features F_1 and F_2 are concatenated, compressed, and summed via residual connections to generate the final output:$\hat{F} = F_1 + F_2 + F_3$.

Enhanced Feature Decoding Module To mitigate spatial detail loss in decoding, the EFD module employs lightweight convolutions and residual connections for improved feature reconstruction. As shown in Fig. 3, the input feature

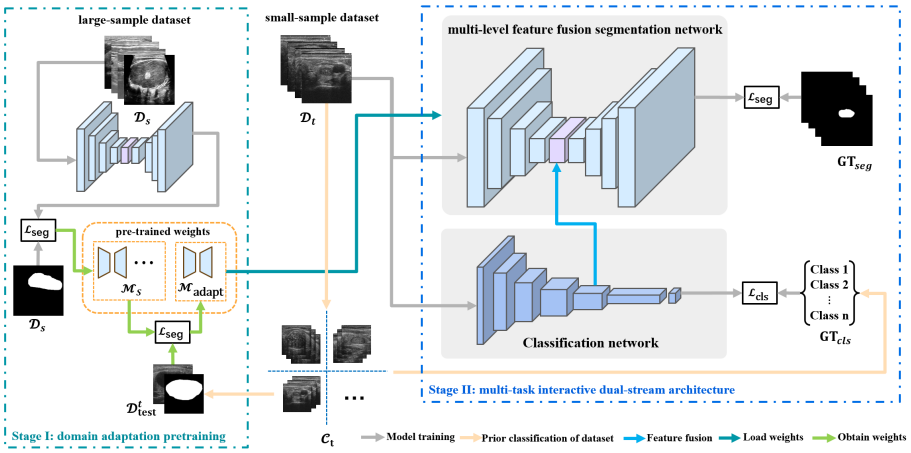

Fig. 2. DPGS-Net: stage I is the domain adaptation pretraining strategy guided by prior classification knowledge. stage II is the multi-task dual-stream feature interaction mechanism. The multi-level feature fusion segmentation network and C_t are applied in both stages, while M_{adapt} would be loaded to the segmentation network of stage II.

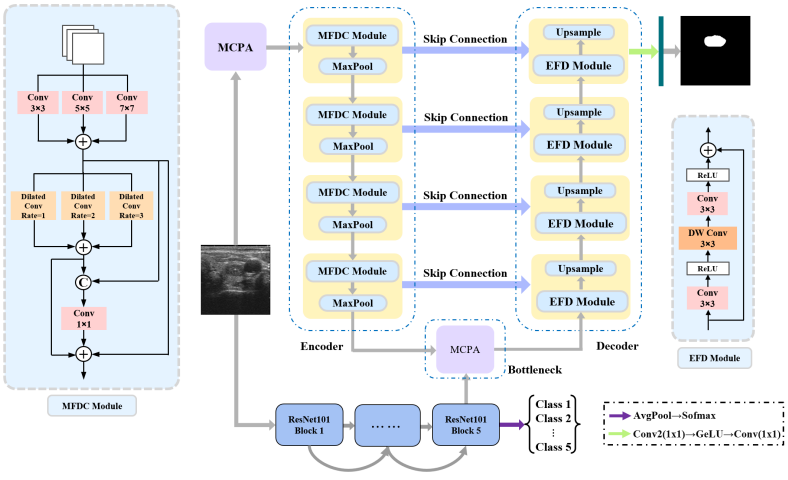

Fig. 3. The backbone network used for target domain segmentation in Stage II consists of the multi-level feature fusion segmentation network and the multi-task dual-stream feature interaction mechanism.

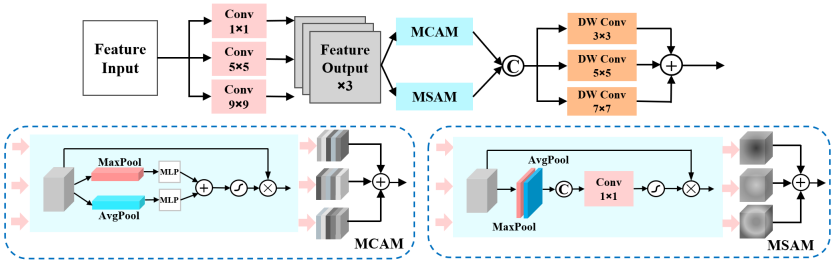

Fig. 4. Multi-branch Convolutional Parallel Attention Module. MCAM: Multi-branch Channel Attention Mechanism; MSAM: Multi-branch Spatial Attention Mechanism.

$F_{\text{in}} \in \mathbf{R}^{H\times W\times C}$ undergoes a series of 3×3 convolutions and depthwise separable convolutions, enhanced by ReLU activations, and is fused with initial features via residual connections to generate the decoded output:

$$\hat{F} = \text{ReLU}\left(F_{\text{in}} + \text{Conv}_{3\times 3}\left(\text{DWConv}\left(\text{Conv}_{3\times 3}\left(F_{\text{in}}\right)\right)\right)\right), \tag{3}$$

2.3 Multi-Task Dual-Stream Feature Interaction Mechanism

Dual-Task Parallel Training To address feature contradictions via prior classification guidance, we propose a multi-task dual-stream feature interaction mechanism, optimizing both classification and segmentation tasks with prior feature guidance. As shown in Fig. 3, a pretrained ResNet101 network [11] is used for prior classification, where labels C_t are generated using surface positional information. High-dimensional features $F_{\text{cls}} \in \mathbf{R}^{H\times W\times C}$ from the penultimate layer of ResNet101, which match the spatial dimensions of $F_{\text{bottleneck}} \in \mathbf{R}^{H\times W\times C}$ from the segmentation network, are fed into the MCPA module for adaptive cross-task feature fusion $F_{\text{fused}} = \text{MCPA}(\text{Concat}(F_{\text{bottleneck}}, F_{\text{cls}}))$, where F_{fused} guides the segmentation reconstruction in the decoder.

Loss Function The total loss L_{total} combines segmentation loss L_{seg} and classification loss L_{cls} with weights λ_1 and λ_2:

$$L_{total} = \lambda_1 L_{\text{seg}} + \lambda_2 L_{\text{CE}} \tag{4}$$

where the segmentation loss is: $L_{\text{seg}} = \mu_1 \times L_{\text{BCE}} + \mu_2 \times L_{\text{dice}}$ with hyperparameters:$\mu_1 = 0.4, \mu_2 = 0.6, \lambda_1 = 0.5, \lambda_2 = 0.5$. L_{seg} is the weighted sum of binary cross-entropy and Dice losses, and L_{CE} is the cross-entropy loss for classification.

3 Experiments

3.1 Datasets and Implementation Details

We utilized the DDTI dataset [18] as the small-sample target domain and the TN3K dataset [9] as the large-sample source domain. DDTI, split into 511 train-

ing, 63 validation, and 63 test images, mimics small-sample feature contradictions due to its size and bias. TN3K, with 2,879 images for training/validation and 614 for testing, offers broader device and lesion variety.

Experiments ran on a NVIDIA RTX 4090 GPU, with input images resized to 256*256, using a batch size of 12 over 50 epochs. The Adam optimizer (1=0.9, 2=0.999) was applied, starting with a learning rate of 1e-3 for pretraining on TN3K, then lowered to 5e-4 for domain adaptation fine-tuning with dynamic class balancing on DDTI, and finally adjusted to 2e-4 for DDTI training.

3.2 Experimental Results

Comparative Experiments We conducted comparison experiments with mainstream models and state-of-the-art methods, including U-Net [19], U-Net++ [29], TransUNet [4], Swin-Unet [3], FCN [15], DeepLabv3+ [5], TRFE+ [10], BPAT-UNet [2], DAC-Net [24], as well as our baseline model (Ours) and the model with domain adaptation pretraining (Ours+DAP). Experiments were conducted on the DDTI test set, with training details restored as described in the original papers, and some parameters set to match. Model performance was comprehensively evaluated using five metrics: Dice coefficient (Dice), Intersection over Union (IoU), Precision, Recall, and Mean Pixel Accuracy (MPA). The experimental results are shown in Table 1.

The experimental results show that our method (Ours+DAP) outperforms all comparison methods on all evaluation metrics. The model with domain adaptation pretraining achieved significant advantages across all metrics. For example, in terms of Dice coefficient, our method achieved 86.52%, which is 4.57 percentage points higher than DeepLabv3+ and 6.49 percentage points higher than the state-of-the-art method TRFE+. Our method also demonstrated significantly better performance in Recall and other metrics, validating its high coverage and low false-positive rate even in the presence of complex background interference. Even without domain adaptation pretraining, our model (Ours) still outperforms

Table 1. Performance comparison of our method with mainstream models and state-of-the-art methods on the DDTI dataset. DAP: Domain Adaptation Pretraining.

Methods	**Dice (%)**	**IoU (%)**	**Precision (%)**	**Recall (%)**	**MPA (%)**
U-Net [19]	71.66	59.09	67.25	86.12	91.97
U-Net++ [29]	78.61	66.79	83.16	78.53	94.41
TransUNet [4]	68.07	54.13	67.82	79.29	90.89
Swin-Unet [3]	66.41	51.81	65.85	78.18	90.23
FCN [15]	68.94	55.89	74.44	71.63	92.45
DeepLabv3+ [5]	81.95	71.39	87.61	81.59	95.05
TRFE+ [10]	80.03	69.85	82.44	83.71	94.41
BPAT-UNet [2]	78.93	67.02	84.11	78.82	94.50
DAC-Net [24]	68.53	54.73	82.78	84.56	89.45
Ours	82.31	72.07	83.71	85.45	95.11
Ours+DAP	**86.52**	**77.63**	**89.76**	**86.30**	**96.09**

other models in Dice, IoU, and MPA, thanks to the multi-level feature fusion architecture and the multi-task dual-stream feature interaction mechanism.

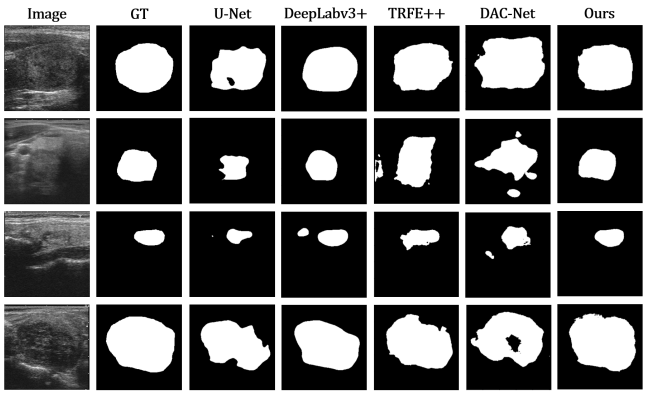

Fig. 5. A qualitative comparison of segmentation results produced by our method and other methods on the DDTI dataset.

3.3 Ablation Study

Table 2. Ablation study results. SN: Multi-Level Feature Fusion Segmentation Network; MT: Multi-Task Dual-Stream Feature Interaction Mechanism; PT: General Pretraining; DAP: Domain Adaptation Pretraining.

Configuration	Dice (%)	IoU (%)	Precision (%)	Recall (%)	MPA (%)
SN	79.48	68.87	86.90	76.76	94.84
SN + MT	82.31	72.07	83.71	85.45	95.11
SN + PT	83.93	73.91	88.68	82.46	95.48
SN + MT + PT	84.46	74.84	**90.81**	82.35	95.84
SN + MT + DAP	**86.52**	**77.63**	89.76	**86.30**	**96.09**

To further validate the effectiveness of each module in our framework, we designed ablation experiments by removing key methods and modules and evaluating their performance. The ablation results are shown in Table 2. Each proposed method contributes to the optimization of the baseline model, with domain adaptation pretraining providing the most significant improvement over general pretraining. These results comprehensively validate the effectiveness of our proposed methods and modules, demonstrating that the designs for multi-task feature interaction, multi-scale feature fusion, and domain adaptation pretraining effectively enhance model performance.

4 Conclusion

This paper proposes a dual prior-guided two-stage segmentation framework that combines domain adaptation pretraining strategies with a multi-level feature fusion architecture and a multi-task dual-stream feature interaction mechanism. This framework addresses feature contradictions, insufficient cross-domain generalization, and data distribution bias in small-sample datasets ultrasound image segmentation tasks. Comparative experiments show that our method significantly outperforms current mainstream methods, and ablation studies further validate the effectiveness and generalizability of our approach. The proposed methods and strategies are also applicable to other models, providing a universal and effective solution for multi-center clinical deployment with significant clinical translation potential.

Acknowledgments. The work was supported by Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019A1515011078). The corresponding author is Yongyi Gong.

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

References

1. Azad, R., Aghdam, E.K., Rauland, A., Jia, Y., Avval, A.H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J.P., Adeli, E., Merhof, D.: Medical image segmentation review: The success of u-net. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)
2. Bi, H., Cai, C., Sun, J., Jiang, Y., Lu, G., Shu, H., Ni, X.: Bpat-unet: Boundary preserving assembled transformer unet for ultrasound thyroid nodule segmentation. Computer methods and programs in biomedicine **238**, 107614 (2023)
3. 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)
4. 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)
5. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
6. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., Sun, H., He, J., Zhang, S., Zhu, M., Qiao, Y.: Sam-med2d (2023), https://arxiv.org/abs/2308.16184
7. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1251–1258 (2017)
8. Feng, W., Ju, L., Wang, L., Song, K., Zhao, X., Ge, Z.: Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 623–631 (2023)

9. Gong, H., Chen, G., Wang, R., Xie, X., Mao, M., Yu, Y., Chen, F., Li, G.: Multi-task learning for thyroid nodule segmentation with thyroid region prior. In: 2021 IEEE 18th international symposium on biomedical imaging (ISBI). pp. 257–261. IEEE (2021)
10. 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)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
12. Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A.: Meta-learning in neural networks: A survey. IEEE transactions on pattern analysis and machine intelligence **44**(9), 5149–5169 (2021)
13. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
14. Li, G., Zhang, Y., Luo, Y.: Multi-task cascaded attention network for brain tumor segmentation and classification. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2340–2344. IEEE (2024)
15. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
16. Ning, Z., Zhong, S., Feng, Q., Chen, W., Zhang, Y.: Smu-net: Saliency-guided morphology-aware u-net for breast lesion segmentation in ultrasound image. IEEE transactions on medical imaging **41**(2), 476–490 (2021)
17. 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. arXiv preprint arXiv:1804.03999 **10** (2018)
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. 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)
20. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7472–7481 (2018)
21. Wang, H., Cao, P., Wang, J., Zaiane, O.R.: Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 36, pp. 2441–2449 (2022)
22. Wang, J., Wang, Y., Zhang, X., Cheng, F.: Multi-task based category assisted gastrointestinal tumor segmentation. In: 2024 9th International Conference on Intelligent Computing and Signal Processing (ICSP). pp. 1092–1096. IEEE (2024)
23. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
24. Yang, Y., Huang, H., Shao, Y., Chen, B.: Dac-net: A light-weight u-shaped network based efficient convolution and attention for thyroid nodule segmentation. Computers in Biology and Medicine **180**, 108972 (2024)

25. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
26. Zhang, R., Lu, W., Guan, C., Gao, J., Wei, X., Li, X.: Shan: Shape guided network for thyroid nodule ultrasound cross-domain segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 732–741. Springer (2024)
27. Zheng, B., Zhang, R., Diao, S., Zhu, J., Yuan, Y., Cai, J., Shao, L., Li, S., Qin, W.: Dual domain distribution disruption with semantics preservation: Unsupervised domain adaptation for medical image segmentation. Medical Image Analysis **97**, 103275 (2024)
28. Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P.H., et al.: Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6881–6890 (2021)
29. 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)