Organ-invariant Tumor Representation Learning via Disentangled Adversarial MAE for Multi-organ Ultrasound Imaging

Xiangyu Xiong[1], Dengqiang Jia[1], Yue Sun[1], Hao Yang[1], Jiaju Huang[1], Zhuoneng Zhang[1], and Tao Tan[1⋆]

[1]Intelligent Medical Computing Laboratory, Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR, China
taotanjs@gmail.com

Abstract. In multi-organ ultrasound (US) imaging, data variability is substantial due to heterogeneous organ-specific imaging characteristics and the presence of diverse tumor types. Existing masked autoencoder (MAE) approaches struggle to disentangle tumor-related features from organ-dependent variations and do not exploit textual priors about organs, thereby limiting their capacity to learn organ-invariant tumor representations. To address these challenges, we propose a novel disentangled adversarial masked autoencoder (DA-MAE), which partitions the encoder into distinct tumor and organ branches to capture complementary information. An organ-invariant adversarial training strategy is employed to suppress organ-specific information within the tumor branch, promoting organ-invariant tumor modeling. In parallel, text prompts provide organ-level priors that implicitly guide the decoder to reconstruct the input image, facilitating more accurate learning of organ-specific representations in the organ branch. Evaluated on five-organ US benchmarks, our method achieves superior performance in downstream classification with an AUC ranging from 78.12% to 99.97% and segmentation with a Dice ranging from 69.38% to 86.12%. Compared to the current best-performing MAE-based method, our method consistently improves classification AUC by 0.17–1.93% and segmentation Dice by 0.58–1.51%. Our method validates the effectiveness of disentangled representation learning with adversarial and text-guided strategies for accurate tumor analysis in multi-organ US imaging scenarios. Our codes are available at https://github.com/yXiangXiong/DA-MAE.

Keywords: Multi-organ ultrasound · Organ-invariant adversarial · Text prompt · Masked autoencoder.

1 Introduction

Ultrasound (US) imaging has emerged as a cornerstone of modern clinical diagnostics due to its non-invasive nature, real-time capabilities, cost-effectiveness,

⋆ Corresponding author

and absence of ionizing radiation [5], making it particularly suitable for multi-organ assessments in resource-limited settings. In clinical practice, automatic US image analysis methods have recently been extensively utilized for the classification and segmentation of various organs [24], enabling early detection of abnormalities such as tumors or cysts. However, blurred tissue boundaries and complex speckle noise lead to low generalization ability in tumor identification and boundary delineation, ultimately impacting treatment planning and patient outcomes. To achieve precise multi-organ classification and segmentation, advanced representation learning models are required to generalize across heterogeneous US datasets from multiple centers.

Self-supervised learning (SSL) techniques, particularly masked autoencoders (MAEs) [15] and their variants, have learned robust feature representations from unlabeled US data. BusM2AE [38] introduces a multi-scale MAE, addressing the limitations of the vanilla MAE in handling breast tumors of varying sizes through token-level and feature-level multi-scale masking strategies. Similarly, URFM [20] establishes a general-purpose foundation model pre-trained on multi-organ US images using representation-based masked image modeling, where high-level semantic representations from BiomedCLIP [39] serve as reconstruction targets instead of raw pixels. In parallel, UA-MAE [35] incorporates uncertainty awareness into the vanilla MAE for breast tumor segmentation, dynamically prioritizing diagnostically critical regions during masking. These state-of-the-art (SOTA) methods address the adaptability of the MAE to US-specific challenges, including speckle noise, low contrast, and anatomical heterogeneity.

Current MAE-based pre-training on multi-organ US data encounters challenges, particularly in disentangling tumor commonalities from organ-specific features and incorporating organ-specific guidance during multi-organ image reconstruction. The high data variability caused by the distinct imaging characteristics of different organs results in entangled representations where tumor-related anomalies are not sufficiently decoupled from normal organ structures. Tumor features tend to exhibit poor transferability across organs. This limits generalization in downstream tumor segmentation and detection. Additionally, standard pixel-level reconstruction lacks explicit organ-specific guidance, hindering the model's ability to incorporate organ priors in heterogeneous US environments, which leads to suboptimal feature learning.

To overcome these limitations, we propose a novel disentangled masked autoencoder using organ-invariant adversarial training (DA-MAE). If separate SSL parts were used per organ, our method could not capture shared semantic invariance and would lose cross-organ transferability. Therefore, we use a single SSL training scheme for all downstream tasks to disentangle common tumor patterns from organ-specific styles in one shared feature space. Specifically, DA-MAE integrates orthogonal loss to promote feature decoupling between tumor and organ branches, adversarial classification with gradient reversal to enable organ-invariant tumor modeling, and BiomedCLIP-derived text prompts for organ-specific guidance during reconstruction. All components in our method are not simply combined modules, but complementary designs centered on dual-

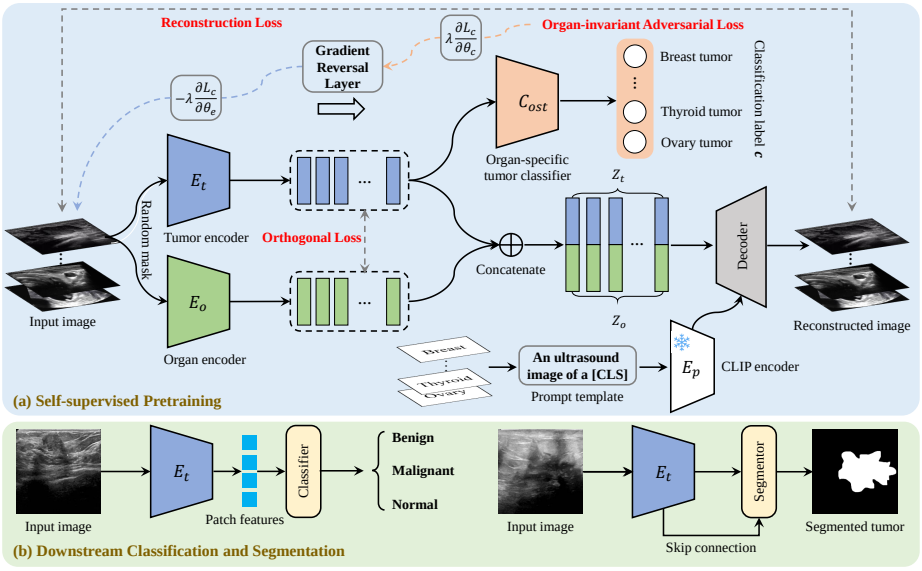

Fig. 1. Overview of proposed DA-MAE. (a) In self-supervised pretraining phase, we introduce two encoders to achieve disentanglement and an organ-specific tumor classifier to realize organ-invariant adversarial training. We also employ a CLIP encoder to help decoder reconstruct the input image. (b) The weights of pre-trained tumor encoder are transferred to initialize the downstream classification and segmentation encoders.

encoder feature disentanglement. Each module plays an indispensable role in disentangling task-shared pathological invariants from domain-specific anatomical confounders, and ablation experiments verify their progressive complementary effects. This design achieves superior downstream performance in multi-organ classification and segmentation, yielding 0.17-12.05% improvements in AUC and 0.58-2.91% improvements in Dice on five-organ US benchmarks compared to SOTA MAE-based methods.

Contributions: (1) We effectively disentangle tumor commonalities from organ-specific representations through feature decorrelation and organ-invariant adversarial training, which suppresses organ-specific information in the tumor branch. (2) We integrate BiomedCLIP-derived organ-specific text prompts into the reconstruction process, providing explicit semantic guidance to enhance cross-organ generalization and improve reconstruction fidelity. (3) We conduct extensive evaluations on five-organ US datasets, demonstrating that our DA-MAE learns organ-invariant tumor representations across multi-organ US scenarios and outperforms the SOTA supervised methods and MAE variants in both downstream classification and segmentation tasks.

2 Method

In this section, we show the proposed DA-MAE in Fig. 1. At first, we delineate the core modules in the self-supervised pre-training stage. Next, we introduce the downstream classification and segmentation networks. Lastly, we provide the training objectives for DA-MAE.

2.1 Disentangled Adversarial Masked Learning

MAE [15] is a self-supervised pre-training method in which parts of the input image are masked, and the model learns to reconstruct them. This forms the backbone for feature extraction of encoders.

Tumor-Common Encoder. This tumor-common encoder, denoted as E_t, extracts tumor features shared across organs. The patch sequence of the input image is first linearly mapped into patch embeddings, and then the positional encoding information is added to them. The mapped patches are masked at a specific ratio r. The remaining visible patches are fed into the tumor-common encoder to extract tumor latent features.

Organ-Specific Encoder. This organ-specific encoder, denoted as E_o, extracts the unique features of each organ. Following the pipeline of the tumor-common encoder, the patch sequence of the input image is first linearly mapped into patch embeddings, and then the positional encoding information is added to them. The mapped patches are masked at a specific ratio r, and the masked positions are the same as those in the tumor-common encoder. The remaining visible patches are fed into the organ-specific encoder to extract organ latent features.

Organ-specific Tumor Classifier. This organ-specific tumor classifier, denoted as C_{ost}, predicts the organ-specific tumor label of an input image, while the tumor encoder produces features that hinder the classifier's accurate prediction. This adversarial training encourages the tumor encoder to produce organ-invariant tumor representations. Specifically, the tumor-common encoder and organ-specific tumor classifier are connected via a gradient reversal layer (GRL) [6] to enable adversarial optimization, aligning tumor features across organ dimensions so that they are not organ-discriminative.

Prompt-Guided Decoder. To reconstruct the original input image, the output latent features of the two encoders are concatenated and projected to the decoder embedding dimension. Moreover, we condition the decoder with textual prompts from the powerful biomedical text encoder of BiomedCLIP [39]. Specifically, we create high-quality text prompts for the five organs and input them into the text encoder to compute text embeddings. Lastly, we project the text embeddings to the decoder embedding dimension and fuse them with the concatenated latent features as the decoder input.

2.2 DA-MAE Adaptation in Downstream Tasks

After self-supervised pre-training, we attached task-specific heads for the downstream classification and segmentation tasks, respectively. As shown in Fig. 1(b),

the classification network consists of the pre-trained encoder followed by a randomly initialized fully connected layer, while the segmentation network comprises the pre-trained tumor encoder, which is skip connected to a randomly initialized UNETR [14] decoder.

2.3 Pre-training and Downstream Objectives

Reconstruction Loss. We adopt the mean squared error (MSE) as the loss function and compute the loss only on masked patches between the pixel decoder prediction and the original input.

$$\mathcal{L}_{recon} = \frac{1}{N_m} \sum_{i=1}^{N_m} \|x_i - \hat{x}_i\|_2^2, \tag{1}$$

where N_m is the number of masked patches in an image, x_i is the input patch as the reconstruction target, and $\hat{x}_i$ is the decoder output for the masked patch.

Orthogonal Loss. We apply the orthogonal loss to ensure that the latent features, i.e., tumor feature embedding z_t and organ feature embedding z_o, from both encoders are as uncorrelated or orthogonal as possible.

$$\mathcal{L}_{ortho} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{z_t^{(i)} \cdot z_o^{(i)}}{\|z_t^{(i)}\|_2 \|z_o^{(i)}\|_2} \right|, \tag{2}$$

where N is the number of images in a batch.

Adversarial Loss. During the forward pass, the GRL acts as an identity function, passing the tumor feature embedding z_t to the organ-specific tumor classifier C_{ost} unchanged: $GRL(z_t) = z_t$. During backpropagation, however, the GRL multiplies the incoming gradients from the organ-specific tumor classifier by -1. This reverses the gradient direction, encouraging the tumor encoder to learn features that cause the organ-specific tumor classifier to fail. The organ-invariant adversarial loss is computed based on this reversed gradient flow:

$$\mathcal{L}_{adv} = CE\big(C_{ost}\big(GRL(z_t)\big), c\big), \tag{3}$$

where $CE(\cdot, \cdot)$ denotes the cross-entropy loss between the predicted and the ground-truth of organ-specific tumor labels.

The overall learning target of the self-supervised pre-training stage can be computed as follows:

$$\mathcal{L} = \mathcal{L}_{recon} + \lambda_{ortho}\mathcal{L}_{ortho} + \lambda_{adv}\mathcal{L}_{adv}, \tag{4}$$

where the λ_{ortho} and λ_{adv} are the weights for orthogonal loss and organ-invariant adversarial loss, respectively. After pre-training, we finetune the tumor-common encoder for downstream classification using cross-entropy loss and for segmentation using a combination of soft-dice loss and cross-entropy loss.

Table 1. Quantitative comparison of our DA-MAE and two SOTA supervised and three SOTA MAE-based methods for downstream classification and segmentation.

Classification						Segmentation					
Dataset	Method	AUC↑	F1↑	Recall↑	Prec↑	Dataset	Method	Dice↑	IoU↑	HD95↓	Sens↑
BUSG	HoVerTran	62.65	58.20	58.85	62.09	BUSG	SynNet	61.72	71.90	5.66	87.83
	BusM2AE	66.07	40.66	45.39	41.39		BusM2AE	66.47	74.85	5.63	89.30
	USFM	71.73	**66.74**	66.89	66.64		USFM	67.18	74.93	5.45	**89.33**
	URFM	76.19	61.03	61.24	64.52		UA-AE	67.87	75.46	5.67	89.16
	Ours	**78.12**	65.36	**69.15**	**69.62**		Ours	**69.38**	**76.49**	**5.40**	88.08
BUSI	HoVerTran	86.50	70.50	71.46	69.68	BUSI	SynNet	71.36	77.53	5.49	87.47
	BusM2AE	87.76	68.45	66.58	71.17		BusM2AE	74.32	79.86	5.11	87.04
	USFM	88.77	70.60	68.04	75.07		USFM	74.86	80.03	5.08	87.75
	URFM	89.06	69.90	68.09	73.01		UA-AE	76.55	81.33	4.99	88.44
	Ours	**90.07**	**77.70**	**75.52**	**81.23**		Ours	**77.13**	**81.51**	**4.93**	**88.92**
TN3K	HoVerTran	74.45	65.92	65.74	66.17	TN3K	SynNet	74.32	78.69	5.42	87.15
	BusM2AE	74.59	60.46	60.75	69.22		BusM2AE	75.43	79.53	5.32	88.41
	USFM	76.42	62.80	62.40	67.43		USFM	76.75	80.15	5.28	88.63
	URFM	76.92	65.39	64.67	69.52		UA-AE	77.54	80.78	5.22	89.13
	Ours	**78.69**	**70.15**	**69.94**	**70.40**		Ours	**78.27**	**81.29**	**5.10**	**89.54**
OCYST	HoVerTran	99.39	93.99	93.81	94.31	MMOTU	SynNet	82.60	84.42	4.99	91.74
	BusM2AE	99.54	97.45	97.66	97.37		BusM2AE	84.13	85.85	4.78	92.36
	USFM	99.68	98.29	98.29	98.29		USFM	84.95	86.09	4.81	92.67
	URFM	99.80	**98.30**	**98.44**	98.21		UA-AE	85.23	86.64	4.67	92.89
	Ours	**99.97**	98.29	98.15	**98.48**		Ours	**86.12**	**87.16**	**4.62**	**93.23**
GSGB	HoVerTran	86.61	73.83	73.30	74.50	OKUS	SynNet	74.77	81.01	5.17	89.76
	BusM2AE	88.97	78.08	77.92	78.25		BusM2AE	75.62	81.75	5.15	90.52
	USFM	89.57	80.46	**82.38**	79.37		USFM	76.16	82.40	4.84	91.24
	URFM	89.82	78.40	77.62	79.39		UA-AE	76.58	82.76	**4.81**	91.43
	Ours	**91.08**	**81.60**	82.00	**81.24**		Ours	**77.70**	**82.96**	5.02	**91.94**

3 Experiments

Dataset. We introduce a five-organ pre-training dataset comprising 59,997 US images: (1) 9,277 breast images from ExpUNet [11], BUSIS [40], BUSI-WHU [17], US3M [36], UDIAT [37], BCMID [30], BUS-UCLM [34], BUS-BRA [7], TUS [29], BUID [3], and BUET-BUSD [32]; (2) 14,160 thyroid images from ThyNodule [16], AUITD [4], DDTI [28], TG3K [10], and TUD [23]; (3) 16,452 ovary images from PCOSGen [13, 12] and PCOS-Dataset [18]; (4) 10,692 gallbladder images from UIdataGB [33]; and (5) 9,416 kidney images collected from KS-US [21]. We create downstream task-specific datasets across the same organs: (1) breast datasets comprising BUSI [2] with 780 images for benign, malignant, and normal tumor classification and segmentation, BUSG [27] with 256 images for benign and malignant tumor classification and segmentation; (2) thyroid datasets comprising TN3K [9, 8] with 3,493 images for benign and normal nodule classification and thyroid nodule segmentation; (3) ovary datasets comprising MMOTU [25] with 1,202 images for segmentation and OCYST [1] with 577 images for complex and simple cyst classification; (4) gallbladder datasets comprising GSGB [22] with 5,138 images for biliary and non-biliary atresia classification; and (5) kidney datasets comprising OKUS [31] with 534 images for segmentation.

Implementation Details. During the training process, the images are uni-

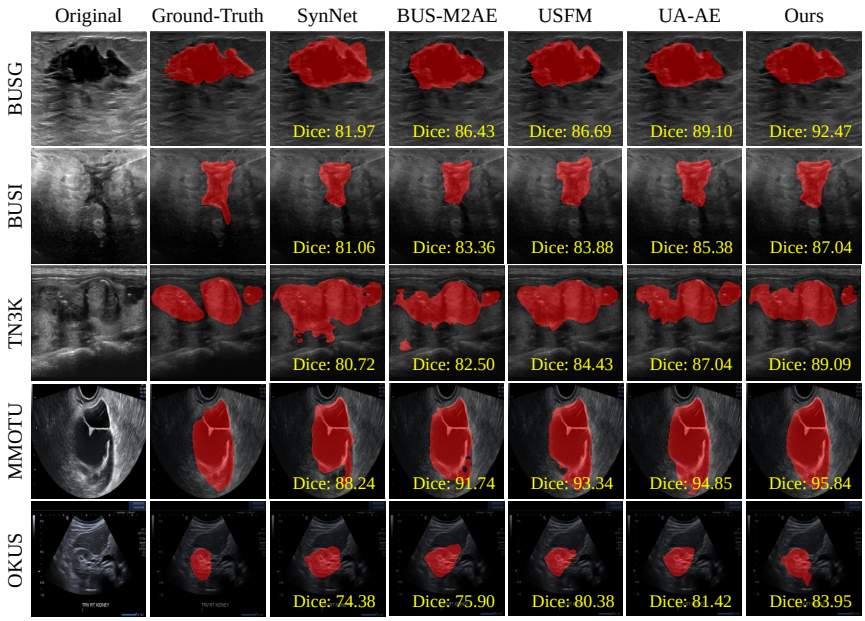

Fig. 2. Qualitative comparison of our DA-MAE and a supervised method, SynNet, and three MAE-based methods, including BUS-M2AE, USFM, and UA-AE.

formly resized to 224×224, and random horizontal flipping is applied. All experiments are carried out on NVIDIA 4090 GPUs. For the self-supervised pre-training, we set the masking ratio r to 65%, the batch size to 256, and the training epochs to 800. We employ the AdamW optimizer with an initial learning rate of 1.5e-4, a weight decay of 0.05, $\beta_1 = 0.9$, and $\beta_2 = 0.95$. We set the values of λ_{ortho} and λ_{adv} in Eq. 4 to 0.5 and 0.02, respectively. For the downstream tasks, we set the batch size to 64 and the training epochs to 200. We employ the AdamW optimizer with an initial learning rate of 1e-3, a weight decay of 0.05, $\beta_1 = 0.9$, and $\beta_2 = 0.95$. We perform five-fold cross-validation on 80% of the data, then select the best hyperparameter configuration to retrain a final model on all non-test data, and finally evaluate it on the 20% test set. The tumor and organ encoders and decoder follow the ViT-B/16 architecture of standard MAE. The organ-specific tumor classifier adopts a two-layer multilayer perceptron (MLP).

Comparison with State-of-the-Art Methods. Table 1 shows that our DA-MAE consistently outperforms two SOTA supervised methods, including HoVer-Tran [26] and SynNet [41], as well as four SOTA MAE-based methods, including BUS-M2AE [38], USFM [19], URFM [20], and UA-AE [35]. In downstream classification tasks, DA-MAE achieves near-maximum values in AUC, F1, Recall, and Precision metrics across five US datasets. On average, DA-MAE outperforms URFM by 1.29% AUC, 4.02% F1, 4.94% Recall, and 3.26% Precision. Specifically, on the typical BUSG and BUSI datasets, it outperforms URFM by 7.91% higher Recall, 7.80% higher F1, and 8.22% higher Precision. In downstream seg-

Table 2. The ablation results of our DA-MAE for downstream classification and segmentation performance on two breast US datasets and a thyroid US dataset.

Baseline MAE	Distentangling Dual-Encoder	Organ-invariant Adversarial	Prompt -Guided	BUSG		BUSI		TN3K	
				AUC↑	Dice↑	AUC↑	Dice↑	AUC↑	Dice↑
✓				66.82	63.50	86.57	71.84	74.57	74.57
✓	✓			68.75	64.19	88.02	74.24	76.33	74.86
✓	✓		✓	74.55	64.97	88.48	74.66	77.36	75.97
✓	✓	✓		76.79	68.29	89.17	76.01	78.20	77.95
✓	✓	✓	✓	**78.12**	**69.38**	**90.07**	**77.13**	**78.69**	**78.27**

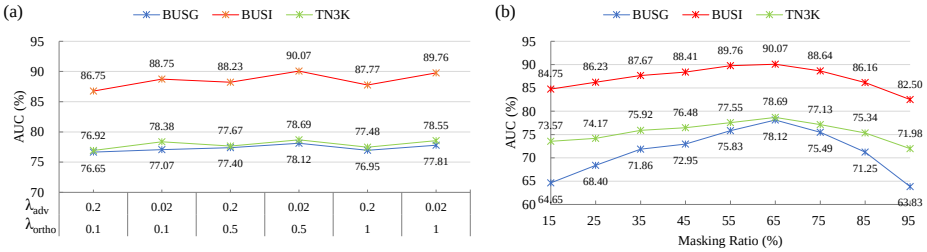

Fig. 3. Parameter experiments on downstream classification task. (a) Effect of orthogonal loss weight λ_{ortho} and adversarial loss weight λ_{adv}. (b) Effect of masking ratio.

mentation tasks, DA-MAE consistently leads in core Dice and IoU metrics. It outperforms UA-AE by 0.58–1.51% in Dice, specifically by 1.12% higher Dice on the OKUS dataset. Fig. 2 further shows that the segmented results of DA-MAE are closer to the ground-truth masks compared to the SOTA methods.

Ablation Study. Our DA-MAE improved AUC by 1.71%–6.31% and increased Dice by 1.13%–4.96% across different module combinations. As shown in Table 2, the disentangled dual-encoder slightly outperforms the standard MAE. Adding the prompt-guided decoder yields further noticeable gains, indicating that organ semantic guidance helps the model learn more discriminative features. The largest consistent improvement comes from combining the disentangled dual-encoder with organ-invariant adversarial training, suggesting that suppressing organ-specific information effectively promotes better tumor-focused representations. Finally, integrating all components delivers the strongest downstream performance, confirming that the synergistic interaction among these modules can learn robust and generalizable representations from unlabeled US images.

Parameter Study. We study loss weights λ_{ortho} and λ_{adv} on downstream classification across three US datasets, with results represented in Fig. 3(a). The highest performance is achieved when $\lambda_{ortho} = 0.5$ and $\lambda_{adv} = 0.02$ are used. This demonstrates that a strong orthogonality constraint encourages meaningful disentanglement of tumor and organ features, whereas a small adversarial weight avoids distortion of organ features during the early pre-training stages. We further study the impact of the masking ratio r in pre-training on downstream

classification performance across three US datasets. As shown in Fig. 3(b), performance improves with the masking ratio up to a clear peak at 65%, followed by consistent degradation at higher ratios. This demonstrates that overly aggressive masking can destroy semantically meaningful context, whereas low masking ratios fail to force the model to learn sufficiently robust representations.

4 Conclusion

In this work, we propose an SSL framework for multi-organ US imaging that captures tumor-common features and organ-specific characteristics. Integrating orthogonal loss, organ-invariant adversarial training, and BiomedCLIP text prompts, it addresses the challenge of disentangling tumor signals from organ variations. Our framework surpasses SOTA on five-organ US benchmarks with 1.93% AUC and 1.51% Dice gains. Our disentanglement paradigm can be readily extended to CT, MR, and other clinical imaging scenarios that inherently contain shared pathological characteristics and domain-specific anatomical variations.

Acknowledgments. This study is supported in part by Macao Polytechnic University Grant (RP/FCA-16/2025), Shenzhen Medical Research Fund (D2501013), and Science and Technology Development Fund of Macao (0014/2025/RIC) and (0041/2023/RIB2).

Disclosure of Interests. The authors declare no competing interests.

References

1. Acharya, U.R., Akter, A., et al.: Use of nonlinear features for automated characterization of suspicious ovarian tumors using ultrasound images in fuzzy forest framework. International Journal of Fuzzy Systems **20**(4), 1385–1402 (2018)
2. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)
3. Ardakani, A.A., Mohammadi, A., Mirza-Aghazadeh-Attari, M., Acharya, U.R.: An open-access breast lesion ultrasound image database: Applicable in artificial intelligence studies. Computers in Biology and Medicine **152**, 106438 (2023)
4. Azouz, M.: Algerian ultrasound images thyroid dataset: Auitd (2022), `https://www.kaggle.com/datasets/azouzmaroua/algeria-ultrasound-images-thyroid-dataset-auitd`
5. Cheng, H.D., Shan, J., et al.: Automated breast cancer detection and classification using ultrasound images: A survey. Pattern recognition **43**(1), 299–317 (2010)
6. Ganin, Y., Ustinova, E., Ajakan, H., et al.: Domain-adversarial training of neural networks. Journal of Machine Learning Research **17**(59), 1–35 (2016)
7. Gómez-Flores, W., Gregorio-Calas, M.J., Coelho de Albuquerque Pereira, W.: Busbra: A breast ultrasound dataset for assessing computer-aided diagnosis systems. Medical Physics **51**(4), 3110–3123 (2024)
8. Gong, H., Chen, G., Wang, R., Xie, X., Mao, M., Yu, Y., et al.: 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)

9. Gong, H., Cheng, H., Xie, Y., Tan, S., et al.: Less is more: Adaptive curriculum learning for thyroid nodule diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 248–257. Springer (2022)
10. Gong, H., et al.: Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. Computers in biology and medicine **155**, 106389 (2023)
11. Guo, Y., Duan, X., Wang, C., Guo, H.: Segmentation and recognition of breast ultrasound images based on an expanded u-net. Plos one **16**(6), e0253202 (2021)
12. Handa, P., Saini, A., Dutta, S., Pathak, H., Choudhary, N., Goel, N., Dhanao, J.K.: Pcosgen-test dataset (2025). https://doi.org/10.5281/zenodo.14591782, https://doi.org/10.5281/zenodo.14591782
13. Handa, P., Saini, A., Dutta, S., Pathak, H., Choudhary, N., Goel, N., Dhanao, J.K.: Pcosgen-train dataset (2025). https://doi.org/10.5281/zenodo.14592001, https://doi.org/10.5281/zenodo.14592001
14. Hatamizadeh, A., Tang, Y., Nath, V., et al.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
15. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16000–16009 (2022)
16. Hou, X., et al.: An ultrasonography of thyroid nodules dataset with pathological diagnosis annotation for deep learning. Scientific Data **11**(1), 1272 (2024)
17. Huang, J., et al.: Busi_whu: Breast cancer ultrasound image dataset (2023)
18. Indirani, A.: Pcos dataset (nov 2024), https://doi.org/10.6084/m9.figshare.27682557
19. Jiao, J., Zhou, J., Li, X., Xia, M., Huang, Y., Huang, L., et al.: Usfm: A universal ultrasound foundation model generalized to tasks and organs towards label efficient image analysis. Medical Image Analysis **96**, 103202 (2024)
20. Kang, Q., Lao, Q., Gao, J., Bao, W., et al.: Urfm: a general ultrasound representation foundation model for advancing ultrasound image diagnosis. iScience (2024)
21. Kaur, G., Singh, S.: Kidney Ultrasound Images "Stone" and "No Stone" (2024), https://doi.org/10.17632/h6jc4xm4py.1
22. Li, R.: Neonatal Biliary Atresia Ultrasound Diagnosis for ghost gallbladders (2025), https://doi.org/10.17632/pw2w73mg8v.2
23. Li, X., Fu, C., Xu, S., Sham, C.W.: Thyroid ultrasound image database and marker mask inpainting method for research and development. Ultrasound in Medicine & Biology **50**(4), 509–519 (2024)
24. Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., et al.: Deep learning in medical ultrasound analysis: a review. Engineering **5**(2), 261–275 (2019)
25. Lyu, S., et al.: Unsupervised cross-domain semantic segmentation on multimodality ovarian tumor ultrasound data. Pattern Recognition p. 112311 (2025)
26. Mo, Y., Han, C., Liu, Y., Liu, M., Shi, Z., Lin, J., et al.: Hover-trans: Anatomy-aware hover-transformer for roi-free breast cancer diagnosis in ultrasound images. IEEE Transactions on Medical Imaging **42**(6), 1696–1706 (2023)
27. Pawłowska, A., Ćwierz-Pieńkowska, A., et al.: Curated benchmark dataset for ultrasound based breast lesion analysis. Scientific Data **11**(1), 148 (2024)
28. 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)
29. Rodtook, A., Kirimasthong, K., Lohitvisate, W., Makhanov, S.S.: Automatic initialization of active contours and level set method in ultrasound images of breast abnormalities. Pattern Recognition **79**, 172–182 (2018)

30. Seddik Tawfik, N., Ghatwary, N., Elgendy, A., Nasr, O., ye, x.: Bcmid: Breast cancer multimodal imaging dataset (Mar 2025)
31. Singla, R., et al.: The open kidney ultrasound data set. In: International Workshop on Advances in Simplifying Medical Ultrasound. pp. 155–164. Springer (2023)
32. Tasnim, J., Hasan, M.K.: Cam-qus guided self-tuning modular cnns with multi-loss functions for fully automated breast lesion classification in ultrasound images. Physics in Medicine & Biology **69**(1), 015018 (2023)
33. Turki, A., Obaid, A.M., Bellaaj, H., et al.: Uidatagb: Multi-class ultrasound images dataset for gallbladder disease detection. Data in Brief **54**, 110426 (2024)
34. Vallez, N., Bueno, G., Deniz, O., Rienda, M.A., Pastor, C.: Bus-uclm: Breast ultrasound lesion segmentation dataset. Scientific Data **12**(1), 242 (2025)
35. Xiong, X., et al.: Ua-mae: An uncertainty-aware masked autoencoder for breast lesion segmentation in ultrasound images. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 4295–4298. IEEE (2025)
36. Yan, P., Gong, W., Li, M., Zhang, J., Li, X., Jiang, Y., Luo, H., Zhou, H.: Tdf-net: Trusted dynamic feature fusion network for breast cancer diagnosis using incomplete multimodal ultrasound. Information Fusion **112**, 102592 (2024)
37. Yap, M.H., Pons, G., Marti, J., Ganau, S., Sentis, M., Zwiggelaar, R., et al.: Automated breast ultrasound lesions detection using convolutional neural networks. IEEE journal of biomedical and health informatics **22**(4), 1218–1226 (2017)
38. Yu, L., Gou, B., et al.: Bus-m2ae: Multi-scale masked autoencoder for breast ultrasound image analysis. Computers in Biology and Medicine **191**, 110159 (2025)
39. Zhang, S., et al.: Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. arXiv:2303.00915 (2023)
40. Zhang, Y., Xian, M., Cheng, H.D., Shareef, B., Ding, J., Xu, F., et al.: Busis: A benchmark for breast ultrasound image segmentation. Healthcare **10**(4) (2022)
41. Zhao, G., Zhu, X., Wang, X., Yan, F., Guo, M.: Syn-net: A synchronous frequency-perception fusion network for breast tumor segmentation in ultrasound images. IEEE Journal of Biomedical and Health Informatics **29**(3), 2113–2124 (2025)