Interaction-aware Contrastive Multi-task Learning for Joint Segmentation and Malignancy Classification in Thyroid Ultrasound

Chaochao Lin[1], Said Boumaraf[2], Xiabi Liu[1,*], Qianglin Liu[3], Lijuan Niu[3,*], and Naoufel Werghi[4]

[1] School of Computer Science & Technology, Beijing Institute of Technology, Beijing, China
liuxiabi@bit.edu.cn
[2] Space Telecommunications Exploitation Center, Algerian Space Agency, Algiers, Algeria
[3] Department of Ultrasound, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
niulijuan8197@126.com
[4] Department of Computer Science, Khalifa University, Abu Dhabi, UAE

Abstract. Ultrasound-based thyroid malignancy assessment requires both accurate nodule delineation and discriminative diagnosis. Joint segmentation and classification is therefore a promising solution to reduce error propagation and improve diagnostic consistency. However, existing joint pipelines often couple tasks mainly through shared parameters or prediction-level cues, providing limited feature-level interaction between branches. This weak coupling can lead to attention–boundary mismatch and poorly aligned cross-task representations. We propose IC-MTL, an interaction-aware contrastive multi-task framework that explicitly models bidirectional cross-task interaction and introduces supervised contrastive regularization on the post-interaction representations to jointly encourage lesion-aligned attention and class-separable diagnostic features. IC-MTL is evaluated on the public DDTI benchmark and a real-world clinical dataset. Across both datasets, IC-MTL improves performance over staged baselines and recent state-of-the-art joint models. Qualitative analyses further indicate improved lesion-aligned attention and more separable representations, supporting more reliable clinical risk stratification.

Keywords: Ultrasound · Thyroid nodules · Joint segmentation and classification · Cross-task interaction · Supervised contrastive learning · Multi-task learning

1 Introduction

Thyroid cancer is a prevalent endocrine malignancy with a growing global burden [1, 6]. Ultrasound (US) is the first-line imaging modality for thyroid nodule evaluation, but malignancy assessment remains operator-dependent and subject to inter-observer variability. In clinical workflows, radiologists typically delineate the nodule and assess malignancy risk based on morphological cues such as margin and shape. Therefore, computer-aided systems that provide both nodule segmentation and malignancy prediction may facilitate more consistent reporting and help reduce unnecessary fine-needle aspiration (FNA). Recent advances in deep learning (DL) have enabled automated analysis of thyroid nodule US images for both segmentation and diagnosis [21, 5, 12, 14]. A common practice is to adopt a staged segmentation-then-classification workflow [20]. However, this design is vulnerable to error propagation [18]. Inaccurate boundaries or incomplete lesion coverage in segmentation can directly bias the downstream malignancy prediction. The Thyroid Imaging Reporting and Data System (TI-RADS) emphasizes morphology-based descriptors (e.g., margin and shape) for malignancy assessment [19]. Since these cues depend on accurate delineation, joint segmentation and classification is a natural choice.

Multi-task learning (MTL) [22, 4] provides a general framework for joint optimization, and several ultrasound approaches adopt a shared encoder with task-specific branches for nodule segmentation and malignancy classification [23, 11, 3, 8, 7, 2]. Existing designs can be broadly categorized into implicit coupling via shared parameters and explicit task coupling that uses segmentation outputs/features to form ROI-aware diagnostic cues, e.g., ROI pooling, masking, or mask-guided attention. Representative examples include Zhou *et al.* [23], which uses iterative training to optimize multiple tasks on ultrasound volumes, and Luo *et al.* [11], which integrates segmentation cues into classification via attention. Other designs have leveraged multi-scale encoder–decoder features with dedicated classification branches [3], multi-stage training with intra- and inter-task consistency constraints [8], transformer-based global feature fusion [2], and attention-based MTL architectures [7]. Attention mechanisms and uncertainty-aware weighting have also been explored to improve robustness and interpretability in joint ultrasound analysis [10]. While these approaches reduce reliance on staged processing, the coupling is often weak or limited to prediction-level cues. Consequently, boundary cues from segmentation may not benefit classification, and malignancy semantics may not refine segmentation. This limited information exchange can lead to attention–boundary mismatch and less trustworthy predictions [18].

Motivated by these limitations, we propose IC-MTL, an interaction-aware contrastive multi-task framework for joint thyroid nodule segmentation and malignancy classification in ultrasound. The key novelty is to explicitly couple the two tasks in a bidirectional manner and to regularize the resulting post-interaction features with supervised contrastive objectives, rather than relying on implicit sharing alone. Specifically, IC-MTL integrates a bidirectional Cross-task Interaction (CI) module with Supervised Contrastive Multi-task Regularization (SCMR). CI specifies how boundary-sensitive cues and malignancy-related semantics are exchanged across branches. SCMR provides the optimization signal that shapes the post-interaction features, promoting separable benign/malignant embeddings and discriminative foreground/background representations.

2 Methodology

Fig. 1(a) illustrates the overall architecture of IC-MTL, which consists of a shared encoder and two task-specific heads for segmentation and classification. Given an input US image, the model outputs a nodule mask and a malignancy probability. Two tightly coupled designs distinguish IC-MTL from conventional joint models. (i) A bidirectional Cross-task Interaction (CI) module defines the computation that exchanges features between the two branches, enabling boundary cues to inform diagnosis and malignancy semantics to refine delineation. (ii) A supervised contrastive multi-task regularizer (SCMR) provides the corresponding optimization signal on post-interaction features, encouraging the interaction to produce the intended effect, benign/malignant separation for classification and foreground/background separation for segmentation.

Cross-task Interaction As is well known, segmentation emphasizes boundary- and shape-related localization cues, whereas classification requires malignancy-discriminative semantics. Therefore, segmentation and benign/malignant classification should be complementary. However, in typical joint architectures with a shared encoder, this complementarity is only implicitly exploited, which may lead to an attention–boundary mismatch. The classifier can overfit to background context or spurious textures, while the segmentation branch is optimized for boundary delineation. To explicitly align the two tasks, we introduce a bidirectional Cross-task Interaction (CI) module by controlling what is transferred in each direction. As shown in Fig. 1(b), segmentation-to-classification injects boundary-aware cues into diagnosis, and classification-to-segmentation transfers malignancy-relevant semantics to refine delineation, jointly suppressing irrelevant activations and improving consistency between attention and boundaries.

Both directions can be described with a generic operator form. Let x_{src} and x_{tgt} denote the source and target features, respectively. Our CI module performs a learned cross-task transformation:

$$x'_{tgt} = \phi(x_{tgt},\, \psi(x_{src}))\,, \tag{1}$$

where $\psi : \mathcal{X} \to \mathcal{M}$ is a cross-task message generator and $\phi : \mathcal{X} \times \mathcal{M} \to \mathcal{X}$ is a message-passing update. In general, we adopt a gated residual update:

$$\psi(x_{src}) \triangleq m, \qquad \phi(x_{tgt}, m) \triangleq x_{tgt} + \mathcal{U}(x_{tgt} \odot m)\,, \tag{2}$$

where m is a learnable modulation field, $\mathcal{U}(\cdot)$ is a lightweight mixing operator (e.g., a 1×1 convolution followed by a nonlinearity), and $\odot$ denotes element-wise multiplication. In this work, we denote $m(\cdot)$ as

$$m = f(x_{src}) \odot g(x_{src}), \qquad f(x) = \sigma_1(W_f x + b_f), \qquad g(x) = \sigma_2(W_g x + b_g), \tag{3}$$

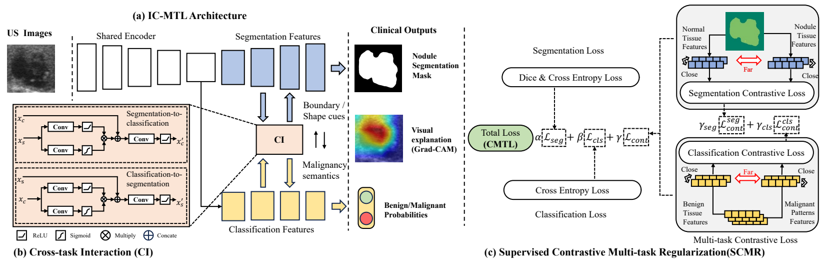


Fig. 1: Overview of our (a) IC-MTL architecture with (b) cross-task interaction (CI) and (c) supervised contrastive multi-task regularization (SCMR).

$$\mathcal{U}(z) = \sigma_3(W_k\,[x_{tgt},\, z] + b_k)\,, \tag{4}$$

where $[\cdot, \cdot]$ denotes channel-wise concatenation. This operator is applied in both directions by setting $(x_{src}, x_{tgt}) \in \{(x_s, x_c), (x_c, x_s)\}$, yielding the updated features x'_c and x'_s, respectively.

Supervised Contrastive Multi-task Regularization To train IC-MTL, we adopt a multi-objective optimization scheme (Fig. 1(c)) that regularizes the post-interaction representations for both tasks. The overall objective is

$$\min_{\theta} \quad \mathcal{L}(\theta) = \alpha\,\mathcal{L}_{seg}(\theta) + \beta\,\mathcal{L}_{cls}(\theta) + \gamma\,\mathcal{L}_{cont}(\theta), \tag{5}$$

where θ denotes the model parameters and α, β, and γ balance the segmentation, classification, and contrastive terms. $\mathcal{L}_{seg}$ is the sum of Dice and cross-entropy for mask prediction, and $\mathcal{L}_{cls}$ is binary cross-entropy for malignancy classification.

SCMR is additionally imposed to encourage intra-class compactness and inter-class separability at both image- and pixel-level embeddings:

$$\mathcal{L}_{cont}(\theta) = \gamma_{seg}\,\mathcal{L}_{cont}^{seg}(\theta) + \gamma_{cls}\,\mathcal{L}_{cont}^{cls}(\theta), \tag{6}$$

where γ_{seg} and γ_{cls} control the relative emphasis on pixel-level and image-level supervised contrastive regularization, respectively.

To align the representation geometry of segmentation and classification after cross-task interaction, SCMR performs supervised contrastive learning at both image and pixel levels. At the sample level, it pulls embeddings from the same diagnosis class closer while pushing benign and malignant embeddings apart. At the pixel level, it similarly increases the margin between nodule foreground and background representations. By adding the same supervised contrastive principle to both branches, SCMR aligns the geometry of the post-interaction feature spaces and reduces cross-task inconsistency, while preserving discriminability for each task. For an embedding set $\{u_i\}$ with labels $\{\ell_i\}$, define $P(i) = \{p \neq i \mid \ell_p = \ell_i\}$ and $A(i) = \{a \neq i\}$. The supervised contrastive loss is

$$\mathcal{L}_{supcon}(\{u_i, \ell_i\}) = \sum_i \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\langle u_i, u_p \rangle / \tau)}{\sum_{a \in A(i)} \exp(\langle u_i, u_a \rangle / \tau)}. \tag{7}$$

Accordingly, SCMR regularizes both branches by setting $\mathcal{L}_{cont}^{cls} = \mathcal{L}_{supcon}(\{z_i, y_i\})$ and $\mathcal{L}_{cont}^{seg} = \mathcal{L}_{supcon}(\{s_i, \tilde{y}_i\})$, where z_i and y_i denote the sample-level embedding and benign/malignant label, and s_i and $\tilde{y}_i$ denote the pixel embedding and its foreground/background label, respectively. Here τ is a temperature and $\langle \cdot, \cdot \rangle$ denotes cosine similarity. By coupling these two constraints after interaction, SCMR jointly shapes the multi-task feature space to be discriminative yet consistent across tasks.

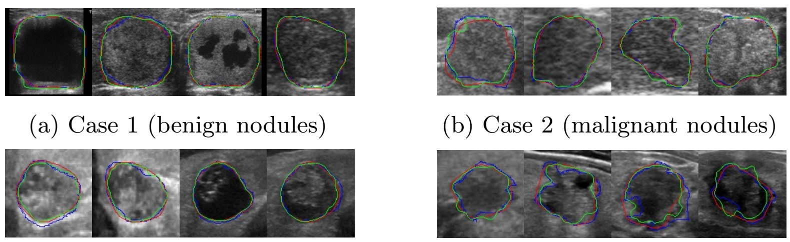
3 Experiments and discussions

Datasets and Experimental Setup IC-MTL is evaluated on two thyroid ultrasound datasets, a public benchmark (DDTI) and a real-world clinical dataset collected from the Department of Ultrasound, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. The Thyroid Digital Image Database (DDTI) [15] contains 347 images from 299 patients with XML annotations. The clinical dataset comprises de-identified ultrasound videos from 116 patients (457 videos; 30 fps). Frames are sampled every 10 frames, resulting in 4,598 images with pixel-level masks and benign/malignant labels annotated by radiologists and pathology (FNA cytology or surgical histopathology) serves as the reference standard. The use of the clinical dataset was approved by the Ethical Committee of National Cancer Center/Chinese Academy of Medical Sciences Cancer Hospital (approval number: NCC2024C-155). The requirement for informed consent was waived by the committee. Images are cropped by the annotated ROI, resized to 128×128, and normalized to $[0, 1]$. Five-fold patient-level cross-validation is used to prevent information leakage across splits. Classification is evaluated with ACC, F1, and AUC, and segmentation with DSC.

The shared encoder is implemented with U-Net convolutional blocks [16]. Training is performed in PyTorch on an NVIDIA RTX 2080Ti GPU (12GB) with RMSprop (weight decay 10^{-8}, momentum 0.999) for 150 epochs (batch size 4, and initial learning rate 10^{-6}). The learning rate is decayed by $\times 0.1$ if a weighted validation score (0.8×ACC + 0.2×DSC) does not improve for 5 epochs. Dynamic loss re-weighting is used. Specifically, we adopt a two-stage schedule that prioritizes segmentation first to stabilize the spatial representation, and then increases the emphasis on classification after the segmentation becomes reliable. When DSC< 0.8, $(\alpha, \beta, \gamma, \gamma_{seg}, \gamma_{cls}) = (1, 0.2, 1, 0.5, 0)$; once DSC$\geq 0.8$, $(\alpha, \beta, \gamma, \gamma_{seg}, \gamma_{cls}) = (0.35, 1, 1, 0, 0.5)$. The DSC threshold and the two sets of weights are chosen empirically on the validation folds to balance convergence stability and final performance. The contrastive temperature is set to $\tau = 0.5$ according to [9].

Table 1: Comparative results on DDTI and local datasets. Best results are bolded, while second-best results are underlined.

Dataset	Model	DSC	ACC	F1	AUC
DDTI	Chowdary *et al.* [3]	89.99±1.56	87.33±1.90	92.16±1.44	81.57±5.78
	Luo *et al.* [11]	92.14±1.95	<u>88.33±3.54</u>	<u>92.73±2.40</u>	<u>83.00±5.03</u>
	He *et al.* [7]	92.07±0.63	86.33±3.98	91.62±2.31	77.49±12.8
	IC-MTL (ours)	**92.38±0.76**	**90.00±2.36**	**93.86±1.22**	**85.92±6.77**
Local	Chowdary *et al.* [3]	84.41±2.35	83.84±5.46	90.40±4.16	73.82±11.8
	Luo *et al.* [11]	84.85±3.46	85.83±5.45	91.63±4.21	62.92±20.5
	He *et al.* [7]	**88.28±0.93**	87.73±2.02	92.51±1.58	82.17±7.05
	IC-MTL (ours)	<u>87.77±0.81</u>	**88.65±2.05**	**93.32±1.47**	**84.66±6.95**



(a) Case 1 (benign nodules)

(b) Case 2 (malignant nodules)

(c) Case 3 (benign nodules)

(d) Case 4 (malignant nodules)

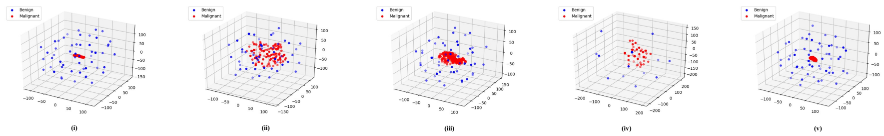
Fig. 2: The segmentation results (Seg-then-Cls in red, IC-MTL in green, and the ground truth in blue).

Comparison with State-of-the-Art and Generalization IC-MTL is compared with representative joint segmentation–classification methods [3, 11, 7] on the public DDTI benchmark and an independent local clinical dataset. Table 1 reports segmentation and diagnosis performance measured by DSC and ACC/F1/AUC, respectively. On DDTI, IC-MTL achieves the best overall performance, improving diagnostic accuracy over prior joint models to ACC 90.00% and AUC 85.92%, while preserving strong segmentation quality with DSC 92.38%. These results suggest that explicit cross-task coupling via bidirectional interaction and contrastive regularization enhances malignancy discrimination and improves segmentation–classification consistency, thereby mitigating the error propagation observed in staged pipelines. On the local dataset, IC-MTL remains competitive in classification, achieving ACC 88.65% and AUC 84.66%, indicating improved robustness under distribution shift due to differences in scanners and acquisition settings. Although He *et al.* reports a slightly higher DSC, 88.28% compared with 87.77%, the difference is marginal, and IC-MTL provides stronger diagnostic performance with higher ACC/F1/AUC, which is desirable for clinical risk stratification. Overall, consistent improvements across two datasets indicate better joint performance and good generalization.

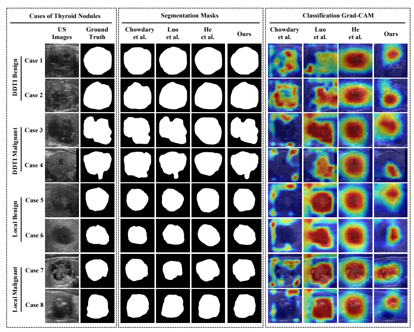
Table 2: Ablation study on the DDTI dataset.

Model	DSC	ACC	F1	AUC
Cls-only	–	86.00±2.79	91.19±1.94	79.70±6.94
Seg-then-Cls	91.91±0.24	82.73±7.44	88.28±7.42	64.29±16.1
w/o CI & SCMR	91.30±0.98	87.33±2.53	92.24±1.55	82.43±4.43
w/o CI	91.60±0.86	87.33±0.91	92.15±1.07	81.52±7.87
w/o Seg→Cls	91.86±1.13	87.33±5.48	92.04±3.80	80.39±7.58
w/o Cls→Seg	91.72±0.58	89.00±2.53	93.26±1.36	82.70±6.94
w/o SCMR	92.13±0.92	88.00±4.31	91.38±2.02	80.75±7.02
w/o SCMR (Cls)	92.32±0.49	84.67±3.42	90.42±2.68	78.57±4.43
w/o SCMR (Seg)	90.72±2.80	85.33±5.19	90.74±3.4	77.10±13.3
IC-MTL	**92.38±0.76**	**90.00±2.36**	**93.86±1.22**	**85.92±6.77**

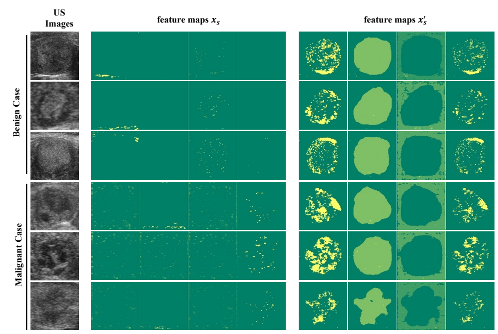
Ablation Studies An ablation study is conducted on DDTI to quantify the contribution of each component to segmentation and malignancy classification. The complete IC-MTL model achieves the strongest overall results across the reported metrics, suggesting complementary benefits from CI and SCMR. Joint learning is compared with two baselines: Cls-only, which predicts malignancy from encoder features, and the two-stage Seg-then-Cls pipeline, which classifies masked ROIs. As shown in Table 2, Seg-then-Cls underperforms Cls-only, with ACC decreasing from 86.00 to 82.73, indicating that hard masking may remove discriminative context and amplify the impact of segmentation errors. The joint variant without the proposed modules, denoted w/o CI & SCMR, improves diagnostic performance to ACC 87.33 and AUC 82.43, supporting the advantage of learning both tasks within a unified model. Qualitative comparisons of predicted masks are provided in Fig. 2. The Seg-then-Cls baseline, using a standard U-Net [16] for segmentation, often produces oversimplified and nearly circular masks, whereas IC-MTL yields more accurate contours, particularly for malignant nodules with irregular shapes. This observation is consistent with improved localization of diagnostic cues such as irregular margins. Removing CI degrades both segmentation and classification performance, with DSC decreasing to 91.60 and ACC to 87.33, indicating that explicit cross-task message passing is beneficial. Disabling either interaction direction also reduces performance. Removing Seg→Cls lowers AUC to 80.39, while removing Cls→Seg lowers DSC to 91.72. These results suggest that boundary cues assist diagnosis and malignancy-related semantics refine delineation. SCMR further regularizes post-interaction representations. Without SCMR, AUC decreases from 85.92 to 80.75. Removing the classification contrastive term, w/o SCMR (Cls), results in a marked drop in ACC to 84.67, whereas removing the segmentation contrastive term, w/o SCMR (Seg), primarily affects DSC, which decreases to 90.72. This pattern suggests that the two contrastive terms contribute to complementary aspects of multi-task feature learning.



(a) t-SNE visualization of features before the fully connected layers. (i) Chowdary *et al.*; (ii) Luo *et al.*; (iii) He *et al.*; (iv) IC-MTL before CI; (v) IC-MTL after CI.



(b) Predicted masks and Grad-CAM heatmaps.



(c) Segmentation-branch feature maps before and after CI.

Fig. 3: Qualitative analyses of IC-MTL.

Qualitative Analysis Qualitative analyses are conducted to examine the effect of the proposed interaction and contrastive objectives in terms of feature separability, model attention, and segmentation-branch activations. As summarized in Fig. 3, IC-MTL produces more class-discriminative representations and more boundary-consistent attention than prior joint models. The classification features are visualized using t-SNE [13] to assess the separability between benign and malignant embeddings. As shown in Fig. 3a, representative prior methods exhibit substantial class overlap, whereas IC-MTL yields tighter intra-class clustering and clearer separation. In addition, features after CI show reduced overlap compared with the pre-interaction features, suggesting that incorporating segmentation-derived cues improves diagnostic discriminability. Grad-CAM [17] is used to interpret the diagnostic branch. In Fig. 3b, competing joint models sometimes attend to background structures or diffuse regions, while IC-MTL produces more compact and lesion-aligned heatmaps. The highlighted regions also better coincide with the predicted masks and clinically relevant morphology such as irregular margins, indicating improved attention–boundary consistency enabled by CI. Segmentation branch features are further inspected before and after interaction, as shown in Fig. 3c. Without CI, the segmentation feature often responds to noisy textures, whereas the post-interaction features exhibit sharper and more localized responses near lesion boundaries. This observation is consistent with transferring malignancy-related semantics to refine delineation, providing an intuitive explanation for the improved joint performance.

4 Conclusions

This work presents IC-MTL, an interaction-aware contrastive multi-task framework for joint thyroid nodule segmentation and malignancy classification in ultrasound. By coupling the two tasks through bidirectional cross-task interaction and supervised contrastive regularization, IC-MTL encourages attention–boundary consistency and more discriminative diagnostic representations. Experiments on the public DDTI benchmark and an independent real-world clinical dataset show improved diagnostic performance over strong staged and joint baselines while maintaining competitive segmentation quality. These findings suggest that IC-MTL may support joint delineation and malignancy risk estimation for thyroid ultrasound assessment and may help reduce unnecessary fine-needle aspiration. Future work may include broader multi-center validation, improved robustness to device and domain shift, and prospective evaluation in real-world clinical workflows.

Acknowledgments. This study was funded by National Natural Science Foundation of China (grant number 82171965 and 82471999).

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

References

1. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **74**(3), 229–263 (2024)
2. Chen, F., Han, H., Wan, P., Liao, H., Liu, C., Zhang, D.: Joint segmentation and differential diagnosis of thyroid nodule in contrast-enhanced ultrasound images. IEEE Transactions on Biomedical Engineering (2023)
3. Chowdary, J., Yogarajah, P., Chaurasia, P., Guruviah, V.: A multi-task learning framework for automated segmentation and classification of breast tumors from ultrasound images. Ultrasonic imaging **44**(1), 3–12 (2022)
4. Crawshaw, M.: Multi-task learning with deep neural networks: A survey. arXiv preprint arXiv:2009.09796 (2020)
5. Ding, J., Huang, Z., Shi, M., Ning, C.: Automatic thyroid ultrasound image segmentation based on u-shaped network. In: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). pp. 1–5. IEEE (2019)
6. Gong, Y., Jiang, Q., Zhai, M., Tang, T., Liu, S.: Thyroid cancer trends in china and its comparative analysis with g20 countries: Projections for 2020–2040. Journal of Global Health **14** (2024)
7. He, Q., Yang, Q., Su, H., Wang, Y.: Multi-task learning for segmentation and classification of breast tumors from ultrasound images. Computers in Biology and Medicine **173**, 108319 (2024)

8. Kang, Q., Lao, Q., Li, Y., Jiang, Z., Qiu, Y., Zhang, S., Li, K.: Thyroid nodule segmentation and classification in ultrasound images through intra-and inter-task consistent learning. Medical image analysis **79**, 102443 (2022)
9. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in neural information processing systems **33**, 18661–18673 (2020)
10. Liu, W., Zhang, S., Shi, J., Dong, Y., Xu, L., Jia, G., Duan, W., Jin, K., Cao, X., Zhang, J., et al.: A multi-task deep learning framework for intraoperative diagnosis of thyroid cancer metastasis using whole slide images. International Journal of Medical Informatics p. 106176 (2025)
11. Luo, Y., Huang, Q., Li, X.: Segmentation information with attention integration for classification of breast tumor in ultrasound image. Pattern Recognition **124**, 108427 (2022)
12. Ma, J., Wu, F., Jiang, T., Zhao, Q., Kong, D.: Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. International journal of computer assisted radiology and surgery **12**, 1895–1910 (2017)
13. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
14. Nguyen, D.T., Kang, J.K., Pham, T.D., Batchuluun, G., Park, K.R.: Ultrasound image-based diagnosis of malignant thyroid nodule using artificial intelligence. Sensors **20**(7), 1822 (2020)
15. 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)
16. 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)
17. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)
18. Tang, L., Tian, C., Yang, H., Cui, Z., Hui, Y., Xu, K., Shen, D.: Ts-dsann: Texture and shape focused dual-stream attention neural network for benign-malignant diagnosis of thyroid nodules in ultrasound images. Medical Image Analysis **89**, 102905 (2023)
19. Tessler, F.N., Middleton, W.D., Grant, E.G., Hoang, J.K., Berland, L.L., Teefey, S.A., Cronan, J.J., Beland, M.D., Desser, T.S., Frates, M.C.: Acr thyroid imaging, reporting and data system (ti-rads): White paper of the acr ti-rads committee. Journal of the American College of Radiology **14**(5), 587–595 (2017)
20. Yadav, N., Dass, R., Virmani, J.: Deep learning-based cad system design for thyroid tumor characterization using ultrasound images. Multimedia Tools and Applications **83**(14), 43071–43113 (2024)
21. Ying, X., Yu, Z., Yu, R., Li, X., Yu, M., Zhao, M., Liu, K.: Thyroid nodule segmentation in ultrasound images based on cascaded convolutional neural network. In: Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part VI 25. pp. 373–384. Springer (2018)
22. Zhao, Y., Wang, X., Che, T., Bao, G., Li, S.: Multi-task deep learning for medical image computing and analysis: A review. Computers in Biology and Medicine **153**, 106496 (2023)

23. Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., Yap, P.T., Shen, D.: Multi-task learning for segmentation and classification of tumors in 3d automated breast ultrasound images. Medical Image Analysis **70**, 101918 (2021)