Uncertainty-aware Diffusion and Reinforcement Learning for Joint Plane Localization and Anomaly Diagnosis in 3D Ultrasound

Yuhao Huang[1], Yueyue Xu[2], Haoran Dou[3], Jiaxiao Deng[1], Xin Yang[1], Hongyu Zheng[2], and Dong Ni[1,4,5,6](✉)

[1] Medical Ultrasound Image Computing (MUSIC) Lab, School of Biomedical Engineering, Medical School, Shenzhen University, Shenzhen, China
nidong@szu.edu.cn
[2] The People's Hospital of Guangxi Zhuang Autonomous Region, Nanning, China
[3] Centre for CIMIM, Manchester University, Manchester, UK
[4] School of Artificial Intelligence, Shenzhen University, Shenzhen, China
[5] National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen, China
[6] School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China

Abstract. Congenital uterine anomalies (CUAs) can lead to infertility, miscarriage, preterm birth, and an increased risk of pregnancy complications. Compared to traditional 2D ultrasound (US), 3D US can reconstruct the coronal plane, providing a clear visualization of the uterine morphology for assessing CUAs accurately. In this paper, we propose an intelligent system for simultaneous automated plane localization and CUA diagnosis. Our highlights are: 1) we develop a denoising diffusion model with local (plane) and global (volume/text) guidance, using an adaptive weighting strategy to optimize attention allocation to different conditions; 2) we introduce a reinforcement learning-based framework with unsupervised rewards to extract the key slice summary from redundant sequences, fully integrating information across multiple planes to reduce learning difficulty; 3) we provide text-driven uncertainty modeling for coarse prediction, and leverage it to adjust the classification probability for overall performance improvement. Extensive experiments on a large 3D uterine US dataset show the efficacy of our method, in terms of plane localization and CUA diagnosis. Code is available at GitHub.

Keywords: CUA · 3D Ultrasound · Diffusion · Reinforcement Learning

1 Introduction

Congenital uterine anomalies (CUAs) are one of the leading causes of infertility, abnormal fetal positioning, and preterm birth [3]. 2D ultrasound (US) is commonly used for initial anomaly screening, but its limited spatial information may lead to misdiagnosis. In comparison, 3D US provides a more detailed visualization of uterine morphology and the structural spatial relationship [15]. Moreover,

it can reconstruct the coronal plane, which is infeasible for 2D US in practice, to show clear uterine fundus and cavity features for improving diagnostic accuracy. However, manually localizing the standard plane and performing efficient diagnosis is challenging and biased due to the vast search space, orientation variability, anatomical complexity and internal invisibility of 3D US. Thus, developing an automatic approach to address both tasks simultaneously is essential to alleviate the burden on sonographers and reduce operator dependency.

Plane localization in 3D medical images. Compared with the one-step method that directly transforms the volume data into plane parameters, recent iterative methods have shown better localization performance [17]. Several reinforcement learning (RL) based methods [1,25,26,29,13], equipped with different strategies, including registration, early-stop termination, neural architecture search, and tangent-based formulation, have achieved satisfactory localization accuracy in 3D US. Most recently, Dou et al. [7] proposed a denoising diffusion model with spherical tangent to drive a continuous and robust transformation. They also explored the localization inconsistency score to perform threshold-based binary classification (normal/abnormal) and obtain 90.67% accuracy.

Deep learning-based 3D medical classification. Previous studies have adopted different techniques, including multi-task learning [27], multi-plane attention [14], and long-range modeling [9], for 3D classification. Yang et al. [24] introduced the MedMNIST v2 dataset and benchmarked different deep models. Currently, although there are limited studies directly targeting 3D US, various US video classification approaches [21,22] achieved good performance on different tasks and have the potential to handle 3D tasks. However, most existing methods relied only on global volumetric knowledge or local information from slices/planes. Moreover, they only output the final classification categories without any intermediate results, limiting their interpretability and clinical availability. Thus, they may not suit our CUA diagnosis task.

In this work, we propose a joint framework based on uncertainty-aware diffusion and RL to automatically localize the coronal plane and diagnose CUAs in 3D US. We believe this is the first work to integrate these two essential tasks in one unified framework, matching the clinical workflow in uterine examinations. Our contribution is three-fold. First, we adopt adaptive multi-scale conditions to guide the denoising process in the diffusion model for accurate plane detection. Second, we leverage the RL agent with unsupervised rewards to provide slice summary, balancing learning difficulty and information preservation. Third, we utilize the uncertainty score powered by text conditions to refine the original classification probability and further improve performance. Validated on the large uterine US dataset demonstrates that our method outperforms strong competitors in both localization and diagnosis tasks, showing satisfactory performance.

2 Methodology

Fig. 1 shows the schematic view of our proposed method. In stage 1, we first add *Gaussian* noise to the target plane parameter p_0, and adaptively perform noisy

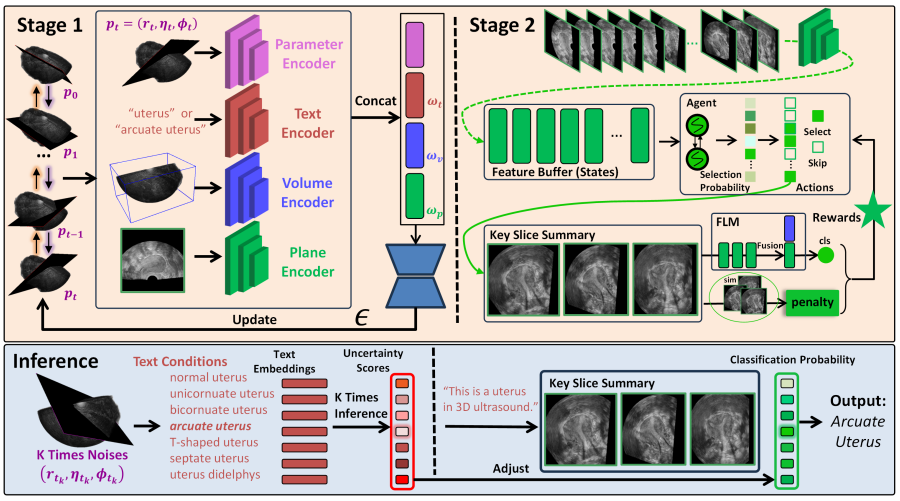

Fig. 1. Overview of our proposed framework.

plane refinement conditioned on multi-scale guidance (i.e., plane, volume and text). In stage 2, we leverage the RL strategy to select the key slices from the above iterative process to build the slice summary and boost model learning. During inference, we obtain the text-driven uncertainty scores to adjust the original classification probabilities, finally improving the overall performance.

2.1 Adaptive Conditional Diffusion Model for Plane Localization

To achieve accurate and efficient plane localization in 3D US, inspired by [7], we formulate this task as a conditional generative problem using the diffusion model [20]. The plane function is represented using tangent-point in spherical coordinates [7], i.e., $p = (r, \eta, \theta)$, to alleviate the angular sensitivity. Give a SP parameter p_0, during the forward diffusion process with t steps, p_t can be defined as $\sqrt{\alpha_t}p_0 + \sqrt{1-\alpha_t}\epsilon$, where $\alpha_t = \prod_{s=1}^{t}(1-\beta_s)$ with β_s is the noise schedule hyperparameter. Then, the reverse denoising process can be written as:

$$p_{t-1} = \sqrt{\alpha_{t-1}}\left(\frac{p_t - \sqrt{1-\alpha_t}\epsilon_\theta(p_t, t, c)}{\sqrt{\alpha_t}}\right) + \sqrt{1-\alpha_{t-1}}\epsilon_\theta(p_t, t, c), \quad (1)$$

where c is the condition set including volume and plane (c_v, c_p) features by convolution encoders, and text embeddings c_t from BiomedCLIP [28]. Texts aim to equip the localization model with semantic information for following classification [5]. Specifically, during training, we randomly provide text prompts like 'This is a uterus in 3D ultrasound' or 'This is a $\mathcal{U}_{gt}$ uterus in 3D ultrasound'. $\mathcal{U}_{gt}$ represents the ground truth (GT) of CUA category. We also introduce the adaptive parameters to control the weights of different conditions. The features from different encoders and the timestep embeddings were first concatenated and

mapped to the weights $(\omega_v, \omega_p, \omega_t)$ using linear projections with *Sigmoid*. Then, the training loss can be defined as: $\mathcal{L} = \mathbb{E}_{x_0,c,t,\epsilon\sim\mathcal{N}(0,1)}\left[\|\epsilon - \epsilon_\theta(x_t, t, c(\omega)\|^2\right]$.

During inference, we provide the category-free text 'This is a uterus ultrasound image'. Then, K=*8* randomly noise plane parameters p_k are generated as initial inputs. After the denoising process, K predicted parameters are obtained, and the final predicted plane can be determined by the averaged parameter.

2.2 RL with Unsupervised Reward for Key Slice Summary

The denoising process will output a series of 2D planes $s_1, s_2, ...s_T$ with features $f_1, f_2, ...f_T$. These bring vital local information to perform classification. However, not all slices can contribute positively, since some may contain diagnosis-irrelevant features. Moreover, using only the final slice may seriously lose vital local information embedded in the denoising process. Hence, an algorithm that can select key slices should be designed to discard unimportant features, preserve slices of fruitful anatomical knowledge, and improve classification accuracy.

Inspired by [12], we introduce an RL-based solution to extract representative slices and remove redundant information. Specifically, we input T features from the buffer into the designed RL agent and obtain the plane-wise selection probability $prob_t$. The agent includes a Bi-LSTM and a fully connected layer with *sigmoid*. Then, the action $a_t \in \{0, 1\}$ can be sampled by $Bernoulli(prob_t)$, where 1 indicates that the t^{th} plane should be selected. A key slice set S can be constructed by selecting the final prediction and planes with $a_t = 1$. For evaluating the quality of $\mathcal{S}$, a reward function is required to balance the redundancy and retention of anatomical information while controlling the set size ($|\mathcal{S}|$):

$$R = \underbrace{\frac{1}{|\mathcal{S}|(|\mathcal{S}|-1)}\sum_{i\neq j} sim(f_i, f_j)}_{R_{sim}} + \underbrace{\sum_{c=1}^{C} y_c log(\hat{y}_c)}_{R_{cls}} - \underbrace{\alpha \cdot max(0, e^{\gamma(|\mathcal{S}|-\mathcal{S}_{max})})}_{R_{penalty}}, \quad (2)$$

where R_{sim} calculates the cosine similarities (sim) between features to constrain redundancy, R_{cls} uses the classification loss to feedback anatomical retention (C and y_c represents the category and the corresponding probabilities). Specifically, by taking the maximum value at the time dimension, the features of key slices of size $(\mathcal{S}, \mathcal{D})$ will be fused to $\hat{f}_{\mathcal{S}} = (1, \mathcal{D})$, which then be combined with the volume features and inputted into a linear layer with *Softmax* for classification output (named fusion linear module, FLM). $R_{penalty}$ penalizes the situations of storing over $\mathcal{S}_{max} = 5$ slices. Since no key slice labels are available, the reward functions are conducted in an unsupervised manner. Our goal is to learn a policy (π_ζ) that maximizes the expected reward $J(\zeta)$. We follow REINFORCE algorithm [23] and approximate the gradient by running N=100 episodes and averaging the results:

$$\nabla_\zeta J(\zeta) \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(R_n - b)\nabla_\zeta \log \pi_\zeta(a_t|f_t(s_t)), \quad (3)$$

where R_n is the reward at n episode and b is the moving average of previous rewards to accelerate convergence. Then, with the L2 regularization, we update ζ by $\zeta_{i+1} = \zeta_i - \rho\nabla_\zeta(\eta\sum_{i,j}\zeta_{i,j}^2 - J(\zeta))$. Moreover, we optimize the FLM by cross-entropy loss. To ease learning, the training of the agent and FLM alternates.

2.3 Uncertainty-aware Strategy for Classification Adjustment

The text-to-image diffusion model has proven its zero-shot classification ability in medical tasks [8]. Specifically, classification is performed by comparing reconstruction errors across images generated for each possible condition (e.g., sick/healthy). Here, we assume that giving the correct text condition will bring less uncertainty. During testing, we generate texts (c_{t_i}) 'This is a $\mathcal{U}$ in 3D ultrasound', for all uterine types ($\mathcal{U}$). The uncertainty-aware score for c_{t_i} is:

$$S_{unc}(c_{t_i}) = \sum_{i=1}^{K} \left\| p_k(c_{t_i}) - \frac{1}{K}\sum_{i=1}^{K} p_k(c_{t_i}) \right\|^2, \tag{4}$$

where $p_k(c_{t_i})$ indicates the predicted plane parameter under c_{t_i}. Then, the coarse classification result represents the category corresponding to the smallest uncertainty value, i.e., $\arg\min_c S_{unc}(c)$. We also design a simple yet effective way to connect the uncertainty score and original classification probability (p_o). Specifically, we first perform normalization on the uncertainty score, and obtain the uncertainty probability (p_u), then the adjusted probability (p_a) is written as:

$$p_{\mathrm{a}}(y_i) = \frac{p_o(y_i)\cdot\frac{1}{p_u(y_i)}}{\sum_{j=1}^{N}(p_o(y_j)\cdot\frac{1}{p_u}(y_j)}. \tag{5}$$

3 Experimental Results

Materials and Implementation Details. Our private large 3D uterus dataset in pelvic US contains 677 volumes, annotated with coronal standard planes, and uterine categories: normal (N, 345), unicornuate (U, 52), bicornuate (B, 5), arcuate (A, 145), T-shaped (T, 15), septate uterus (S, 96), and uterus didelphys (D, 19), by experienced sonographers under strict quality control using the Pair annotation software package [18]. The average volume size is 367×189×334, with spacing equal to 0.4^3 mm^3. We randomly split the volumes into 406, 68, and 203 for training, validation, and testing, respectively.

In this study, we implemented our method in *PyTorch*, using one NVIDIA 4090 GPU with 24GB memory. In stage 1, the denoising diffusion model with cosine noise scheduler was developed based on *diffusers* package by *HuggingFace*. The time steps are set to 1000 for training, and 100 for validation and testing. The training includes $200K$ interactions with batch size=8 and learning rate (lr)=5e-4. The architectures of parameter/plane/volume encoders follow the design in [7], here, we used different multi-layer perceptions (MLPs) to map them

Table 1. Comparison on plane localization task. The best results are shown in bold.

	ITN [17]	RL_{WSDT} [25]	RL_{NAS} [25]	RL_{FT} [29]	$DIFF_{MSG}$ [7]	Ours
Ang	30.36±18.45	17.71±15.69	13.22±12.76	12.65±11.32	8.84±6.77	**7.12±5.16**
Dis	1.24±1.99	1.55±1.44	1.20±1.09	1.26±1.10	1.01±0.84	**0.86±0.78**
SSIM	0.57±0.16	0.63±0.14	0.71±0.11	0.72±0.09	0.74±0.13	**0.76±0.12**
NCC	0.61±0.14	0.69±0.13	0.76±0.13	0.75±0.13	0.80±0.19	**0.82±0.16**

Table 2. Comparison on CUA classification task. The best results are shown in bold.

	Method					Acc	Pre	Rec	F1	AUC
2D/2.5D	ResNet18 [10]					51.23	32.68	48.52	34.40	0.8753
	ResNet18-2.5D [10]					70.44	45.75	40.10	35.73	0.9350
	ResNet18-LSTM [10]					75.86	63.84	39.98	39.50	0.9628
Video	I3D [2]					70.44	74.52	59.77	61.54	0.9600
	VST [19]					88.18	63.19	65.60	63.20	0.9912
	UniFormerV2 [16]					89.16	77.02	63.65	64.11	0.9929
3D	ResNet18 [10]					74.38	73.62	58.51	59.43	0.9715
	ResNet50 [10]					76.35	71.11	60.79	55.92	0.9766
	MedicalNet18 [4]					84.24	73.88	73.54	70.20	0.9800
	MedicalNet50 [4]					85.22	96.54	65.20	67.81	0.9806
	DenseNet121 [11]					74.88	89.93	81.26	78.83	0.9768
	ViT [6]					76.35	74.84	74.16	71.34	0.9732
	nnMamba [9]					86.70	87.95	89.51	83.33	0.9903
	AP	PP	SS	GF	UA	Acc	Pre	Rec	F1	AUC
Ours	✓	✗	✗	✗	✗	91.13	97.86	88.42	91.96	0.9944
	✓	✗	✗	✓	✗	92.12	72.78	74.86	71.46	0.9971
	✗	✓	✗	✓	✗	87.19	71.17	74.41	68.42	0.9867
	✗	✗	✓	✓	✗	92.61	98.17	91.11	93.91	0.9968
	✗	✗	✗	✗	✓	92.12	97.60	87.26	91.28	0.9961
	✗	✗	✓	✓	✓	**94.58**	**98.34**	**98.76**	**95.21**	**0.9989**

to $\mathbb{R}^{32}$. The text encoder followed BiomedCLIP [28], obtaining $\mathbb{R}^{32}$ features via MLP. Through concatenation, they were input into the UNet-based denoising network (with the positional encoding of timestep added to all blocks). In stage 2, we trained the RL agent and FLM with lr=5e-4 in $100K$ interactions. Alternate training was conducted every $1K$ iterations. Both stages used AdamW optimizers and conducted validation every $2K$ iterations, and models with the best validation performance were saved for final evaluation. All slices were resized to 224^2, and volume size standardization was not required since we processed one volume per iteration. They were then scaled to 0–1 by dividing 255.

Quantitative and Qualitative Analysis. For plane localization, we analyzed the performance in terms of spatial differences (*Ang*, *Dis*) and content similarities (*SSIM*, *NCC*). *Ang* (°) calculates the angles between normal vectors

Table 3. Different localization methods with (w) and without (w/o) slice summary (SS) for classification. The improved metrics are highlighted in bold.

	RL_{WSDT}		RL_{NAS}		RL_{TF}		DIFF_{MSG}		Ours	
	w/o SS	w SS	w/o SS	w SS	w/o SS	w SS	w/o SS	w SS	w/o SS	w SS
Acc	79.81	**81.28**	83.74	**85.22**	86.70	**87.68**	89.66	**90.64**	91.13	**93.10**
F1	64.74	**69.05**	70.85	**73.69**	74.76	**78.71**	78.46	**83.35**	91.96	**93.38**

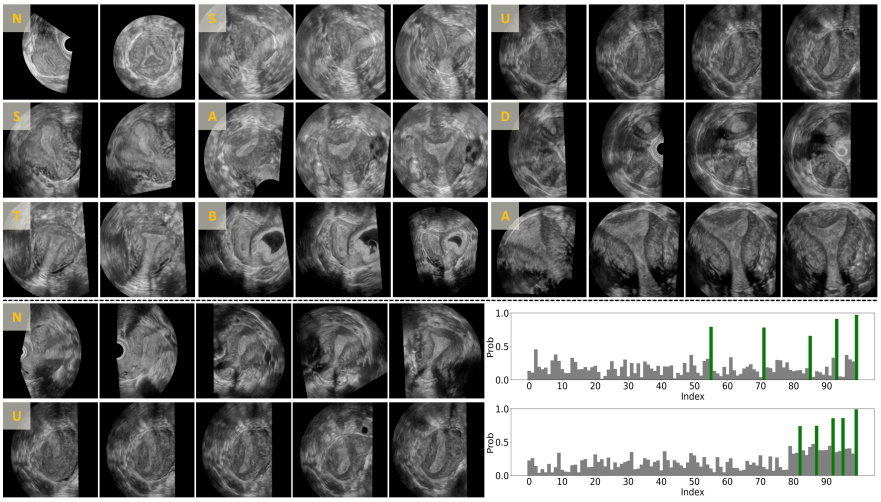

Fig. 2. Typical summary examples with uterine types in the top-left corners. Below are two summaries with index selection probabilities: green (select) and gray (skip).

of two planes, and *Dis* (mm) denotes the difference between their Euclidean distances towards the volume origin. Details for *SSIM* and *NCC* refer to [7]. For classification, Accuracy (Acc, %), Precision (Pre, %), Recall (Rec, %), F1-score (F1, %), and AUC were included for comprehensive analysis. Besides, we used the paired t-test to assess statistical significance for plane localization metrics, and the Chi-square test for overall classification performance.

Table 1 compares our method with five iterative approaches, i.e., one traditional CNN [17], three RL methods [25,26,29], and one diffusion-based solution [7]. Results show that our proposed method significantly outperforms all the competitors ($p < 0.05$) in terms of spatial and content similarities. In Table 2, we evaluated various typical and advanced methods and their variants, containing three 2D/2.5D traditional networks [10], three video-based approaches [2,19,16], and seven 3D models [10,4,11,6,9]. Results show that they all obtain *Acc* and *F1* less than 90%. Our proposed method shows significant improvement over them in all evaluation metrics (last row, $p < 0.05$), validating its effectiveness. Table 2 also provides the ablation studies to test the contribution of each proposed strategy. *AP*, *PP*, and *SS* represent using all planes, the final predicted plane, and

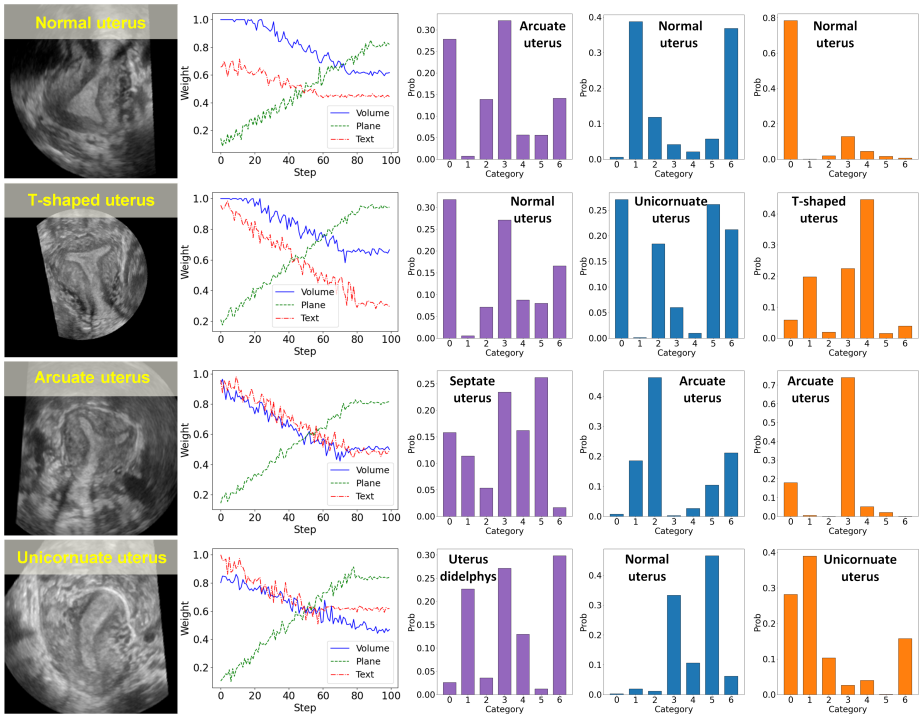

Fig. 3. Visualization results. c1: predicted planes with GTs (yellow). c2: denoising curves of condition weights at different time steps. c3-5: original, uncertainty and adjusted probabilities with predicted classification results (c: column).

slice summary as classification inputs, respectively. GF means the global volumetric features, and UA denotes the uncertainty-aware strategy. We observe that even with GF, using only one plane degrades the performance compared to using all planes. Besides, SS can remove abundant information, extract key features, and improve performance (F1: 22.45%↑). Notably, with UA only, our model achieves satisfactory performance with Acc, F1 and Pre>90%. Adding SS and GF can further enhance the performance on all evaluation metrics. We further prove that SS is general to different plane localization methods, with significantly improved performance (Table 3, all p-values<0.05).

We also visualize the key slice summaries in Fig. 2. It shows that the selected slices are representative and diverse, capturing the vital anatomical information of the corresponding CUA types. Besides, the denoising curves in Fig. 3 prove that global conditions (volume/text) play a vital role in the early denoising stages, driving coarse localization, while local plane conditions have minimal impact initially. As denoising progresses, the model gradually shifts focus to local information to achieve fine plane localization. Bar plots in the last three columns (c) show examples where raw or uncertainty-based predictions were incorrect (c3-4), and how uncertainty adjustment corrects the results (c5).

4 Conclusion

In this paper, we propose a novel joint framework for automated plane localization and CUA diagnosis in 3D US. We first introduce the adaptive denoising strategy to weight conditions and improve localization. Then, we propose RL with unsupervised rewards for key slice summary construction, thereby enhancing model learning. Last, we leverage the text condition-driven uncertainty-aware score to adjust the classification probability and boost the overall performance. In the future, we will extend the method to more modalities and tasks.

Acknowledgments. This work was supported by the grant from National Natural Science Foundation of China (12326619, 62171290, 82201851); Science and Technology Planning Project of Guangdong Province (2023A0505020002); Frontier Technology Development Program of Jiangsu Province (BF2024078); Guangxi Province Science Program (2024AB17023), Key Research and Development Program (AB23026042) and Natural Science Foundation (2025GXNSFAA069471).

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

References

1. Alansary, A., et al.: Automatic view planning with multi-scale deep reinforcement learning agents. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 277–285. Springer (2018)
2. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017)
3. Chan, Y.Y., Jayaprakasan, K., Zamora, J., Thornton, J.G., et al.: The prevalence of congenital uterine anomalies in unselected and high-risk populations: a systematic review. Human reproduction update **17**(6), 761–771 (2011)
4. Chen, S., Ma, K., Zheng, Y.: Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:1904.00625 (2019)
5. Clark, K., Jaini, P.: Text-to-image diffusion models are zero shot classifiers. Advances in Neural Information Processing Systems **36** (2024)
6. Dosovitskiy, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
7. Dou, H., Huang, Y., Huang, Y., Yang, X., Zhen, C., Zhang, Y., Xiong, Y., Huang, W., Ni, D.: Standard plane localization using denoising diffusion model with multi-scale guidance. Computer Methods and Programs in Biomedicine p. 108619 (2025)
8. Favero, G.M., Saremi, P., Kaczmarek, E., Nichyporuk, B., Arbel, T.: Conditional diffusion models are medical image classifiers that provide explainability and uncertainty for free. arXiv preprint arXiv:2502.03687 (2025)
9. Gong, H., et al.: nnmamba: 3d biomedical image segmentation, classification and landmark detection with state space model. In: ISBI. pp. 1–5. IEEE (2025)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)

11. Huang, G., Liu, Z., et al.: Densely connected convolutional networks. In: CVPR. pp. 4700–4708 (2017)
12. Huang, R., Ying, Q., et al.: Extracting keyframes of breast ultrasound video using deep reinforcement learning. Medical Image Analysis **80**, 102490 (2022)
13. Huang, Y., Zou, Y., Dou, H., Huang, X., Yang, X., Ni, D.: Localizing standard plane in 3d fetal ultrasound. In: 3D Ultrasound, pp. 239–269. CRC Press (2023)
14. Jang, J., et al.: M3t: three-dimensional medical image classifier using multi-plane and multi-slice transformer. In: CVPR. pp. 20718–20729 (2022)
15. Kougioumtsidou, A., Mikos, T., Grimbizis, G.F., et al.: Three-dimensional ultrasound in the diagnosis and the classification of congenital uterine anomalies using the eshre/esge classification: a diagnostic accuracy study. Archives of Gynecology and Obstetrics **299**, 779–789 (2019)
16. Li, K., Wang, Y., He, Y., Li, Y., et al.: Uniformerv2: Spatiotemporal learning by arming image vits with video uniformer. arXiv preprint arXiv:2211.09552 (2022)
17. Li, Y., Khanal, B., Hou, B., Alansary, A., Cerrolaza, J.J., et al.: Standard plane detection in 3d fetal ultrasound using an iterative transformation network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 392–400. Springer (2018)
18. Liang, J., Yang, X., Huang, Y., Li, H., He, S., Hu, X., Chen, Z., Xue, W., Cheng, J., Ni, D.: Sketch guided and progressive growing gan for realistic and editable ultrasound image synthesis. Medical image analysis **79**, 102461 (2022)
19. Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., Hu, H.: Video swin transformer. In: CVPR. pp. 3202–3211 (2022)
20. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
21. Sun, A., Zhang, Z., et al.: Boosting breast ultrasound video classification by the guidance of keyframe feature centers. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 441–451. Springer (2023)
22. Wang, Y., Li, Z., et al.: Key-frame guided network for thyroid nodule recognition using ultrasound videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 238–247. Springer (2022)
23. Williams, R.J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning **8**, 229–256 (1992)
24. Yang, J., Shi, R., et al.: Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. Scientific Data **10**(1), 41 (2023)
25. Yang, X., Dou, H., Huang, R., Xue, W., Huang, Y., et al.: Agent with warm start and adaptive dynamic termination for plane localization in 3d ultrasound. IEEE Transactions on Medical Imaging **40**(7), 1950–1961 (2021)
26. Yang, X., Huang, Y., Huang, R., Dou, H., Li, R., Qian, J., Huang, X., Shi, W., Chen, C., Zhang, Y., et al.: Searching collaborative agents for multi-plane localization in 3d ultrasound. Medical Image Analysis **72**, 102119 (2021)
27. Yang, Z., Fan, T., Smedby, Ö., Moreno, R.: 3d breast ultrasound image classification using 2.5 d deep learning. In: 17th International Workshop on Breast Imaging (IWBI 2024). vol. 13174, pp. 443–449. SPIE (2024)
28. Zhang, S., Xu, Y., Usuyama, N., Xu, H., Bagga, J., Tinn, R., Preston, S., et al.: Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. arXiv preprint arXiv:2303.00915 (2023)
29. Zou, Y., Dou, H., Huang, Y., Yang, X., Qian, J., Zhen, C., et al.: Agent with tangent-based formulation and anatomical perception for standard plane localization in 3d ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 300–309. Springer (2022)