

Two-Stage Cross-Domain Cervical Abnormality Screening with Cytopathological Image Synthesis and Knowledge Distillation

Jincheng Li^{1*}, Yuzhi He^{2*}, Yihui Zhan¹, Xinmei Zhang¹, Yifei Sun³, Zelin Liu⁴, Lichi Zhang⁴, Minye Shao⁵, and Lili Zhao^{1†}

¹ School of Artificial Intelligence and Computer Science, Nantong University, Nantong 226019, China

² School of Telecommunications Engineering, Xidian University, Xi'an 710126, China

³ College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China

⁴ School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China

⁵ Department of Computer Science, Durham University, Durham, UK

*These authors contributed equally to this work.

†Corresponding author: ylzh@ntu.edu.cn

Abstract. Cross-domain diagnosis remains a major challenge in cervical cell pathology due to pronounced domain shifts across institutions and the subtle visual differences among disease stages, which jointly impair model generalization. To address these issues, this paper proposes a two-stage framework for cross-domain cervical cell detection. In the first stage, we propose the Spatially-Continuous Unpaired Neural Schrödinger Bridge (SC-UNSB), which constructs a synthetic intermediate domain to mitigate cross-domain distribution shifts by modeling image translation as an entropy-regularized optimal transport process. In the second stage, we propose a dual-level feature alignment strategy within a knowledge distillation, which progressively aligns shallow structural features and deep semantic representations to facilitate the transfer of domain-invariant knowledge from the source to the target model. Experimental results demonstrate that the proposed method effectively mitigates domain shift and category ambiguity, improving the cross-domain detection performance. Code is available at <https://github.com/ZhanYiHui06/MICCAI2026>

Keywords: Cross-domain cervical cell · Schrödinger bridge · Knowledge distillation.

1 Introduction

As the fourth most common cancer globally, cervical cancer remains a leading cause of mortality among women, making early detection critical for prognosis and survival [1,2]. Recently, object detection models based on deep learning

have demonstrated robust performance on datasets of cervical cytology [3,4,5,6]. However, their generalization across institutions remains limited. Variations in sample preparation workflows and staining protocols across medical institutions often lead to substantial performance degradation in cross-institutional evaluations. Moreover, cervical carcinogenesis is a continuous biological process, where visual differences between lesion stages are inherently subtle. Under cross-domain, such fine-grained distinctions become even more difficult to discern, increasing the risk of misclassification and potential clinical harm [7,8].

Domain adaptation techniques have been introduced to mitigate cross-domain discrepancies. Among them, knowledge distillation provides a mechanism for transferring representations from a model trained on a source domain to a target-domain model [9,10]. Existing distillation-based adaptation methods typically perform feature alignment at a single representation level or treat shallow and deep features independently [11,12,13]. Such strategies may not fully account for the hierarchical nature of cytological representations, where structural information and semantic abstraction contribute differently to cross-domain generalization. Furthermore, when appearance discrepancies between domains are substantial, feature-level alignment alone may be insufficient, motivating the integration of image-level translation mechanisms.

To address these challenges, this paper proposes a two-stage framework for cross-domain cervical cell detection. Our contributions can be summarized as follows:

- To mitigate severe appearance discrepancies across institutions, we propose the Spatially-Continuous Unpaired Neural Schrödinger Bridge (SC-UNSB), a spatially continuous image translation framework that constructs an intermediate domain by enforcing dense statistical normalization within the Schrödinger Bridge formulation. This design explicitly addresses boundary-induced artifacts in ultra-high-resolution cytopathological images.
- To reduce semantic misalignment under cross-domain settings, we introduce a dual-level feature alignment strategy within a knowledge distillation framework. By jointly aligning shallow structural representations and high-level semantic embeddings, the proposed method facilitates progressive transfer of domain-invariant knowledge.
- Extensive experiments on two large-scale cervical cytology datasets demonstrate that the proposed framework consistently improves cross-domain detection performance, achieving up to 26.9% mAP and 45.8% mAP50.

2 Methodology

2.1 Spatially-Continuous Unpaired Neural Schrödinger Bridge

To mitigate cross-domain appearance discrepancies, we build upon the Unpaired Neural Schrödinger Bridge (UNSB) [14], which formulates unpaired image translation as an entropy-regularized optimal transport problem. Given source and

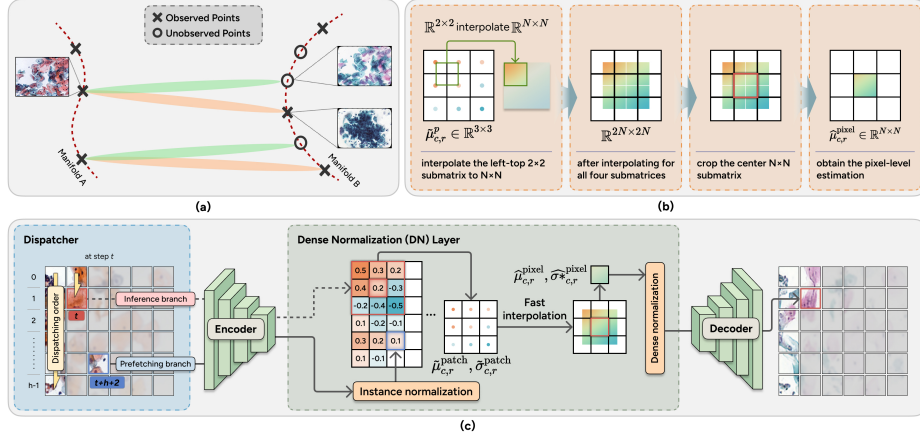


Fig. 1. Overview of SC-UNSB. (a) Learning process of the Schrödinger Bridge for entropy-regularized transport between source and target distributions. (b) Dense pixel-level moment estimation, where an $N \times N$ statistical field is interpolated from a 3×3 neighborhood of patch-level statistics. (c) Architecture of the dispatcher and the Dense Normalization (DN) layer. An ultra-high-resolution image is divided into patches. The dispatcher coordinates prefetching and inference branches, and pixel-wise mean $\hat{\mu}_{r,c}$ and variance $\hat{\sigma}_{r,c}$ are obtained via fast interpolation from neighboring patches to ensure spatially continuous normalization.

target distributions π_0 and π_1 , the Schrödinger Bridge seeks a diffusion process $\{X_t\}_{t \in [0,1]}$ by solving

$$\min_{\mathbb{Q}} \mathcal{KL}(\mathbb{Q} \parallel \mathbb{W}) \quad \text{s.t.} \quad \mathbb{Q}_0 = \pi_0, \mathbb{Q}_1 = \pi_1, \quad (1)$$

where $\mathbb{W}$ denotes the reference Wiener measure. In practice, UNSB approximates this objective via a discrete-time Markov process and adversarial learning.

While UNSB effectively aligns global distributions, directly applying it to ultra-high-resolution cytopathological images requires patch-wise translation, which introduce statistical discontinuity and boundary artifacts across neighboring regions. In standard generators, Instance Normalization (IN) computes feature statistics $\mu^{(k)}$ and $\sigma^{(k)}$ independently for each patch $x^{(k)}$:

$$\text{IN}(x^{(k)}) = \gamma \frac{x^{(k)} - \mu^{(k)}}{\sigma^{(k)}} + \beta \quad (2)$$

This independence, however, triggers a Boundary Drift Error: the statistical discontinuity at the patch interface $\partial\Omega_k$ causes the optimal transport manifold to become spatially fractured. Such fractures manifest as tiling artifacts, which can distort cell structures and jeopardize clinical diagnosis.

To rectify the transport path, we propose the Spatially-Continuous Unpaired Neural Schrödinger Bridge (SC-UNSB). Instead of treating statistical moments as scalars, we re-parameterize them as continuous functions of pixel coordinates

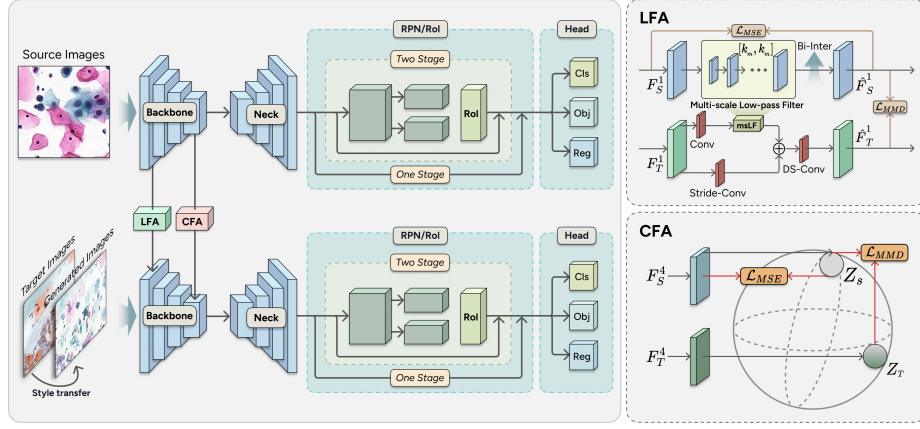


Fig. 2. Model of the dual-level feature alignment. The source model guides the target model via LFA in the frequency domain for structural patterns, and CFA for high-level semantics.

$p = (u, v)$ within the generator. We use a Dense Normalization (DN) module into the UNSB generator, as illustrated in Fig. 1. For a specific pixel p in patch k , its statistics are estimated by interpolating the coarse statistics from the current patch and its neighbors $\mathcal{N}_k$. Let $\mathcal{S}_k = \{(\mu^{(n)}, \sigma^{(n)}) \mid n \in \mathcal{N}_k\}$ be the set of regional statistics. We define a bilinear interpolation operator Φ_{interp} to construct the smooth statistical fields:

$$\hat{\mu}(p) = \Phi_{\text{interp}}(\{\mu^{(n)}\}_{n \in \mathcal{N}_k}, p), \quad \hat{\sigma}(p) = \Phi_{\text{interp}}(\{\sigma^{(n)}\}_{n \in \mathcal{N}_k}, p) \quad (3)$$

Consequently, the generator G_θ is modified to be position-aware, utilizing these dense fields for normalization:

$$G_{\text{SC}}(x^{(k)}, t) = \text{Conv} \left(\gamma \odot \frac{x^{(k)} - \hat{\mu}(p)}{\hat{\sigma}(p)} + \beta \right) \quad (4)$$

By integrating $\hat{\mu}(p)$ and $\hat{\sigma}(p)$, we implicitly constrain the search space of the Schrödinger Bridge to the subspace of spatially continuous functions. This ensures that the generated cell structures maintain topological consistency across patch boundaries, effectively eliminating tiling artifacts while preserving high-frequency biological details.

2.2 Distillation-based Feature Alignment

We propose a dual-level feature alignment strategy within a knowledge distillation framework, as illustrated in Fig. 2. The proposed framework consists of two complementary components: Loose Feature Alignment (LFA), which focuses on preserving shallow structural representations at early stages, and Compact

Feature Alignment (CFA), which emphasizes the transfer of high-level semantic knowledge.

To reduce the marginal distribution discrepancy between source and target domains, we adopt $\mathcal{L}_{MMD}$ as the cross-domain alignment objective. Meanwhile, a reconstruction constraint $\mathcal{L}_{MSE}$ is imposed to preserve the original source-domain representations during alignment. The overall feature alignment losses are defined as

$$\mathcal{L}_{MMD} = \frac{1}{B} \left\| \sum_{i=1}^B \phi(\hat{F}_S^i) - \sum_{j=1}^B \phi(\hat{F}_T^j) \right\|_2^2, \mathcal{L}_{MSE} = \|F_S - \hat{F}_S\|_2^2 \quad (5)$$

where $\phi(\cdot)$ denotes the feature mapping function, $\hat{F}_S$ and $\hat{F}_T$ are the aligned source and target features, and B is the batch size.

Loose Feature Alignment (LFA) operates on shallow features to preserve structural information that is less sensitive to semantic variation but vulnerable to domain shift. We transform features into the frequency domain using a multi-scale low-pass filter (MSLF). For the source model, average pooling with different kernel sizes extracts low-frequency components, followed by bilinear interpolation:

$$\hat{F}_S = MSLF(F_S) = \Phi(AvgPool_{k_a \times k_a}(F_S)) \quad (6)$$

where $AvgPool_{K_a \times K_a}$ denotes the average pooling function with a kernel size of $K_a \times K_a$, and $\Phi(\cdot)$ represents the bilinear interpolation operation.

Regarding the target domain model feature $\hat{F}_T$, considering the discrepancies in shallow features between the source and target domain models, we design a learnable low-pass filter. This filter is composed of a multi-scale low-pass filter, a convolutional downsampling module, and depthwise separable convolution (DSConv). The learnable target domain feature $\hat{F}_T$ in the frequency domain can be expressed as:

$$\hat{F}_T = Conv_{3 \times 3}(Concat[DownSample_{s \times s}(F_T), MSLF(F_T)]) \quad (7)$$

where $DownSample_{s \times s}$, $Concat$, and $Conv_{3 \times 3}$ denote the convolutional downsampling module, the feature concatenation operation, and the 3×3 convolution operation, respectively.

After obtaining the frequency-domain features of the source and target domain models through the MSLF transformation, the loss for LFA can be calculated using the loss formula, which can be expressed as $\mathcal{L}_{LFA} = \mathcal{L}_{MMD} + \mathcal{L}_{MSE}$

Compact Feature Alignment (CFA) aligns high-level semantic representations from the penultimate layer. Since encoder outputs may differ in dimension, a 1×1 convolution projects features into a unified embedding space Z . Alignment in this compact space promotes transfer of class-discriminative knowledge. Similarly, the loss for CFA can be expressed as $\mathcal{L}_{CFA} = \mathcal{L}_{MMD} + \mathcal{L}_{MSE}$

By maintaining the original training paradigm of the target domain model while employing a coarse-to-fine strategy—transitioning from loose to compact

feature alignment—the target domain model is enabled to acquire knowledge from the source domain model, with a specific emphasis on domain-invariant features. The overall training loss can be expressed as:

$$\mathcal{L} = \underbrace{\mathcal{L}_{cls} + \mathcal{L}_{loc}}_{task} + \alpha \cdot (\underbrace{\mathcal{L}_{LFA} + \mathcal{L}_{CFA}}_{feature\ alignment}) \quad (8)$$

where $\mathcal{L}_{cls}$ and $\mathcal{L}_{loc}$ represent the classification and localization losses for the cell detection task, respectively, and α serves to balance the various loss components.

3 Experiment

Datasets and Settings: We conducted extensive experiments using the CRIC dataset[16], which contains 7,839 cervical microscopy bounding boxes, as the source domain ($\mathcal{D}_s$), and the ComparisonDetector dataset[17], which contains 7,410 clustered cell images from Pap smear slides, as the target domain ($\mathcal{D}_t$). The ($\mathcal{D}_t$) splits are reconstructed for strict cross-domain evaluation and are released with the code. Models are trained using stochastic gradient descent with a learning rate of 0.001 and momentum 0.9. Pseudo-labels with confidence above 0.9 supervise the student model. Evaluation covers image generation quality and detection performance on the target-domain test set, which is strictly excluded from training.

For generation, we compare SC-UNSB with five representative unpaired translation methods that map $\mathcal{D}_s$ to the intermediate domain ($\mathcal{D}_i$). We leverage FID [22] and KID [23] scores to assess the diversity and quality of the generated images. Additionally, NIQE [24] and HIST are utilized to evaluate the naturalness

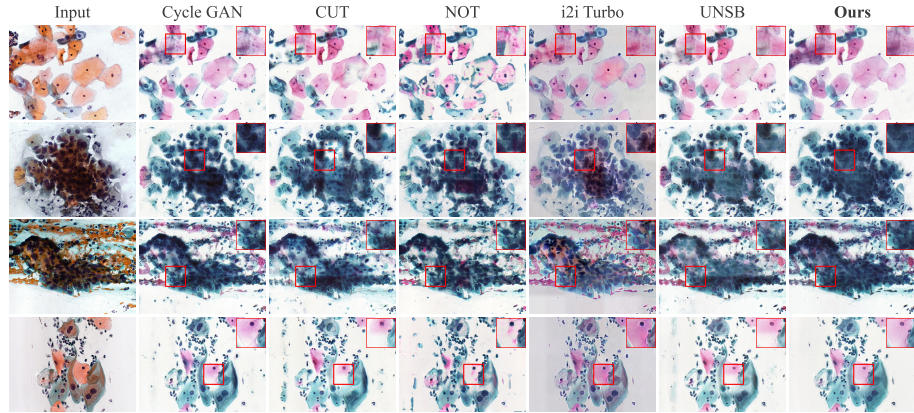


Fig. 3. Qualitative comparison between different generators in transferring $\mathcal{D}_s$ images into $\mathcal{D}_i$ images. Each method produces four examples in individual rows. Zoom in to check image details.

Table 1. Comparison with other methods. The best, second best and third best results are highlighted in **dark**, **medium** and **light** green, respectively.

Generator	$\mathcal{D}_s$	CycleGAN[18]	CUT[19]	NOT[20]	i2i-Turbo[21]	UNSB[14]	SC-UNSB
FID ↓	241.05	147.28	132.61	177.43	154.65	143.43	135.31
KIDx100 ↓	10.896	1.831	1.365	5.104	2.514	2.411	1.807
NIQE ↓	14.72	13.51	14.6	12.86	16.39	13.94	11.38
HIST ↑	0.384	0.695	0.722	0.571	0.473	0.701	0.754
Detector	RetinaNet						
mAP ↑	4.6%	10.5%	14.0%	9.3%	10.3%	17.8%	20.2%
mAP50 ↑	9.3%	22.4%	29.3%	18.0%	26.8%	35.4%	41.5%
Detector	RetinaNet + CFA						
mAP ↑	7.8%	12.1%	16.3%	9.6%	13.9%	21.2%	22.7%
mAP50 ↑	19.1%	23.7%	33.8%	20.4%	30.1%	38.7%	43.1%
Detector	RetinaNet + LFA						
mAP ↑	5.1%	12.3%	19.4%	12.1%	12.4%	23.8%	24.1%
mAP50 ↑	11.5%	24.0%	35.4%	18.4%	28.4%	41.1%	43.7%
Detector	RetinaNet + LFA + CFA						
mAP ↑	12.6%	18.8%	22.7%	11.3%	15.1%	24.1%	26.9%
mAP50 ↑	26.6%	31.8%	40.3%	21.9%	30.9%	42.6%	45.8%

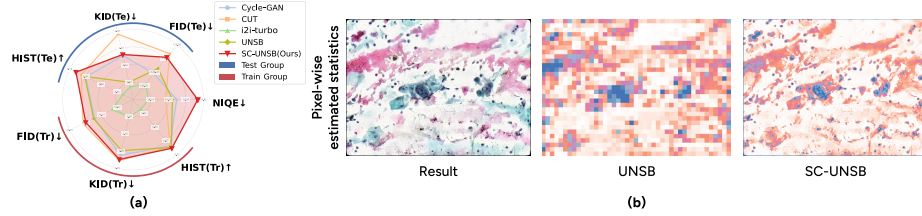


Fig. 4. (a) Radar chart comparison of multiple quantitative metrics; (b) Visual comparison of pixel-wise estimated statistics.

and realism of the synthesized results. For detection, RetinaNet is trained on $\mathcal{D}_i$ and $\mathcal{D}_t$, followed by knowledge distillation. Performance is evaluated using mAP[25] and mAP50.

Generation: As shown in Table 1, the high FID of source images indicates a substantial domain gap. While CycleGAN and UNSB partially reduce this gap, their relatively high FID and NIQE scores indicate limited alignment in distribution statistics and perceptual quality. SC-UNSB achieves the lowest NIQE and competitive FID/KID scores, along with the highest HIST value, these results confirm that enforcing spatially continuous statistical fields effectively stabilizes the transport path and reduces structural artifacts introduced by patch-wise translation. Qualitative results in Fig. 3 further demonstrate that SC-UNSB effectively mitigates tiling artifacts and boundary inconsistency, producing spatially coherent cell structures that better support downstream detection. Fig. 4 provides a quantitative and structural comparison of the estimated statistical fields. The radar chart shows that SC-UNSB achieves consistent improvements

across multiple quality metrics rather than optimizing a single criterion. The visualization of pixel-wise statistics reveals that the baseline exhibits abrupt transitions at patch boundaries, leading to fragmented normalization patterns. In contrast, SC-UNSB produces smooth and spatially coherent moment fields. Such continuity preserves cellular morphology and staining consistency, which is critical for downstream detection. Additional hyperparameter analyses are reported in Table 2.

Table 2. Ablation study of different interpolating granularities.

Interpolating Granularity	NIQE↓	SC-UNSB (testing set)			SC-UNSB (training set)		
		FID↓	KID↓ × 100	HIST↑	FID↓	KID↓ × 100	HIST↑
1 pixel	11.38	135.31	1.807	0.754	66.291	1.534	0.711
16 pixels	11.59	147.93	2.22	0.717	70.77	1.8	0.700
32 pixels	11.62	162.59	3.630	0.713	77.1	2.37	0.698
64 pixels	11.34	148.29	2.91	0.708	76.87	2.6	0.688
128 pixels	11.40	143.37	2.47	0.704	70.6	2.07	0.682
256 pixels	13.94	143.43	2.411	0.701	68.232	1.623	0.707

Detection: As shown in Table 1, training RetinaNet on the $(\mathcal{D}_i)$ improves cross-domain detection compared with direct source training. Incorporating LFA or CFA further enhances performance, and their combination achieves the best results, reaching 26.9% mAP and 45.8% mAP50 with SC-UNSB. The improvement is most evident when using SC-UNSB-generated data, suggesting that spatially coherent synthesis facilitates more effective feature alignment. Table 3 shows that the proposed alignment strategy outperforms conventional distillation methods, which focus on prediction-level transfer without explicitly addressing cross-domain feature mismatch. Additional hyperparameter analyses are reported in Table 3 as well.

Table 3. Comparison with different Knowledge Distillation methods and ablation study of Hyperparameter of ours.

Metric	Knowledge Distillation				Hyperparameter Analysis			
	KD[26]	DKD[27]	SPD[28]	Ours	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$
mAP ↑	21.7%	18.6%	22.3%	26.9%	25.3%	26.9%	26.1%	24.5%
mAP50 ↑	40.6%	36.7%	43.4%	45.8%	44.2%	45.8%	45.4%	45.1%

4 Conclusion

This work presents a two-stage framework for cross-domain cervical cell detection that explicitly addresses both appearance-level domain shift and representation-level feature misalignment. By constructing a spatially coherent intermediate domain through SC-UNSB and introducing dual-level feature alignment within a distillation framework, the proposed approach enhances the transfer of domain-invariant knowledge across institutions. These results highlight the potential of combining generative domain bridging with progressive feature alignment to enable cross-domain diagnosis in cervical cytopathology.

Acknowledgments This work was supported by the Natural Science Foundation of Jiangsu Province (BK20251838) and the Nantong Science and Technology Program Project (JC2024055).

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

References

1. Cohen, P.A., Jhingran, A., Oaknin, A., Denny, L.: Cervical cancer. *The Lancet* **393**(10167), 169–182 (2019)
2. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., et al.: Erratum: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca Cancer J Clin* **70**(4), 313 (2020)
3. Fei, M., Shen, Z., Song, Z., Wang, X., Cao, M., Yao, L., Zhao, X., Wang, Q., Zhang, L.: Distillation of multi-class cervical lesion cell detection via synthesis-aided pre-training and patch-level feature alignment. *Neural Networks* **178**, 106405 (2024)
4. Li, J., Dong, D., Zheng, M., Zhang, J., Hang, Y., Zhang, L., Zhao, L.: High-Precision Mixed Feature Fusion Network Using Hypergraph Computation for Cervical Abnormal Cell Detection. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 250–259. Springer (2025)
5. Fei, M., Song, Z., Shen, Z., Liu, M., Wang, Q., Zhang, L.: Weakly Semi-supervised Cervical Lesion Cell Detection via Twin-Memory Augmented Multiple Instance Learning. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 637–647. Springer (2025)
6. Hu, Y., Chen, Q., Liao, L., Lin, W., Wu, H., Wang, L.: Controllable Image Synthesis Workflow for Enhancing Cervical Cell Detection. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 85–95. Springer (2025)
7. Zhu, C., Lin, J., Tan, G., Zhu, N., Li, K., Wang, C., Li, S.: Advancing Ultrasound Medical Continuous Learning with Task-Specific Generalization and Adaptability. In: *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 3019–3025. IEEE (2024)
8. Zhu, C., Lin, Y., Chen, S., Wang, Y., Lin, J.: MedEyes: Learning Dynamic Visual Focus for Medical Progressive Diagnosis. In: *Proceedings of the AAAI Conference on Artificial Intelligence* **40**(16), 13916–13924 (2026)

9. Lan, Q., Hsu, Y.-C., Khan, N.S., Jiang, X.: ReCo-KD: Region-and Context-Aware Knowledge Distillation for Efficient 3D Medical Image Segmentation. *arXiv preprint arXiv:2601.08301* (2026)
10. Lan, Q., Tian, Q.: ACAM-KD: Adaptive and Cooperative Attention Masking for Knowledge Distillation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3957–3966 (2025)
11. Pan, Q., Xue, Y., Yang, B.: A Deformable-Based Source-Free Unsupervised Domain Adaptation Method for Cervical Cell Detection. In: *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE (2025)
12. Shen, Z., Fei, M., Wang, X., Cai, J., Wang, S., Zhang, L., Wang, Q.: Two-stage cytopathological image synthesis for augmenting cervical abnormality screening. *arXiv preprint arXiv:2402.14707* (2024)
13. Tong, S., Gao, S., Liu, K., Huang, Z., Xu, H., Ying, H., Wu, J.: Uncertainty-Aware Multi-expert Knowledge Distillation for Imbalanced Disease Grading. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 624–634. Springer (2025)
14. Kim, B., Kwon, G., Kim, K., Ye, J.C.: Unpaired image-to-image translation via neural schrödinger bridge. *arXiv preprint arXiv:2305.15086* (2023)
15. De Bortoli, V., Thornton, J., Heng, J., Doucet, A.: Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in neural information processing systems* **34**, 17695–17709 (2021)
16. Rezende, M.T., Silva, R., Bernardo, F.d.O., Tobias, A.H., Oliveira, P.H., Machado, T.M., Costa, C.S., Medeiros, F.N., Ushizima, D.M., Carneiro, C.M., et al.: Cric searchable image database as a public platform for conventional pap smear cytology data. *Scientific data* **8**(1), 151 (2021)
17. Liang, Y., Tang, Z., Yan, M., Chen, J., Liu, Q., Xiang, Y.: Comparison detector for cervical cell/clumps detection in the limited data scenario. *Neurocomputing* **437**, 195–205 (2021)
18. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232 (2017)
19. Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y.: Contrastive learning for unpaired image-to-image translation. In: *European conference on computer vision*, pp. 319–345. Springer (2020)
20. Korotin, A., Selikhanovych, D., Burnaev, E.: Kernel neural optimal transport. *arXiv preprint arXiv:2205.15269* (2022)
21. Parmar, G., Park, T., Narasimhan, S., Zhu, J.-Y.: One-step image translation with text-to-image models. *arXiv preprint arXiv:2403.12036* (2024)
22. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* **30** (2017)
23. Chen, R., Huang, W., Huang, B., Sun, F., Fang, B.: Reusing discriminators for encoding: Towards unsupervised image-to-image translation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8168–8177 (2020)
24. Ho, M.-Y., Wu, M.-S., Wu, C.-M.: Ultra-high-resolution unpaired stain transformation via kernelized instance normalization. In: *European Conference on Computer Vision*, pp. 490–505. Springer (2022)

25. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision, pp. 740–755. Springer (2014)
26. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
27. Zhao, B., Cui, Q., Song, R., Qiu, Y., Liang, J.: Decoupled knowledge distillation. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp. 11953–11962 (2022)
28. Wei, S., Luo, C., Luo, Y.: Scaled decoupled distillation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15975–15983 (2024)