HASD: Hierarchical Adaption for pathology Slide-level Domain-shift

Jingsong Liu[‡1,3], Han Li[‡1,2], Chen Yang[2,4], Michael Deutges[2,4], Ario Sadafi[2,4], Xin You[2], Katharina Breininger[5], Nassir Navab[2,3], and Peter J. Schüffler[1,3] [*]

[1] Institute of Pathology, TU Munich, Munich, Germany
[2] Computer Aided Medical Procedures (CAMP), TU Munich, Munich, Germany
[3] Munich Center for Machine Learning (MCML), Munich, Germany
[4] Institute of AI for Health, Helmholtz Munich, Munich, Germany
[5] Center for AI and Data Science (CAIDAS), Julius-Maximilians-Universität Würzburg, Würzburg, Germany

Abstract. Domain shift is a critical problem for pathology AI as pathology data is heavily influenced by center-specific conditions. Current pathology domain adaptation methods focus on image patches rather than WSI, thus failing to capture global WSI features required in typical clinical scenarios. In this work, we address the challenges of slide-level domain shift by proposing a Hierarchical Adaptation framework for Slide-level Domain-shift (HASD). HASD achieves multi-scale feature consistency and computationally efficient slide-level domain adaptation through two key components: (1) a hierarchical adaptation framework that integrates a Domain-level Alignment Solver for feature alignment, a Slide-level Geometric Invariance Regularization to preserve the morphological structure, and a Patch-level Attention Consistency Regularization to maintain local critical diagnostic cues; and (2) a prototype selection mechanism that reduces computational overhead. We validate our method on two slide-level tasks across five datasets, achieving a 4.1% AUROC improvement in a Breast Cancer HER2 Grading cohort and a 3.9% C-index gain in a UCEC survival prediction cohort. Our method provides a practical and reliable slide-level domain adaption solution for pathology institutions, minimizing both computational and annotation costs.

Keywords: Domain shift · Pathology slide-level tasks

1 Introduction

Digital pathology has experienced significant attention in recent years, driven by the widespread adoption of whole-slide imaging (WSI) scanners and rapid progress in artificial intelligence (AI) research [15]. This trend has facilitated notable advancements in many AI-based pathology tasks [25, 6]. However, a major challenge in pathology is **domain shift**, arising from differences in pathology

---

[*] Corresponding author: `peter.schueffler@tum.de`
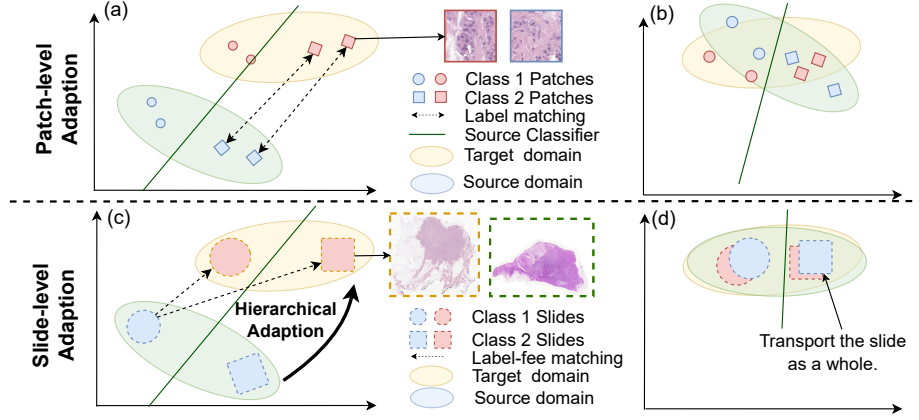[1] [‡]These authors contributed equally to this work.

Fig. 1: (a)(b): Existing methods rely on patch-level labels from the source domain and pseudo labels from the target domain for patch-level domain adaptation. (c)(d): In slide-level tasks, patches lack individual labels. We propose a **hierarchical** adaption framework that leverages **domain-level** alignment solver with **slide-level** structure and **patch-level** diagnostic features preserved.

centers, including **imaging variances** (varying staining protocols, digitization procedures), **label discrepancies** (patient populations), and other factors. Domain shift can downgrade model performances when transferring a model to a different clinical site, potentially compromising patient safety [11, 13].

Domain adaptation (DA) is a method to mitigate domain shift commonly used in histology image analysis, leveraging adversarial training or generative models [8, 9, 13, 17, 18, 22, 5]. By extracting domain-invariant features or synthesizing target-like samples, DA helps to align distributions between different datasets, improving model robustness and generalization. However, most methods focus on **patch-level adaption**, making them misaligned with real-world applications where clinical decisions require a global understanding of the slide (e.g., cancer grading or survival prediction) that patch-level adaptation methods fail to capture. An intuitive approach to extend patch-level DA methods to the slides-level is by decomposing slide-level tasks into many patch-level predictions [21, 7], applying DA at the patch level, and then aggregating the slides-level results. However, this often requires extensive patch-level annotations from pathologists, which is very time-consuming. Another naive approach aggregates patch embeddings via simple integration techniques (e.g., mean pooling) [1]. However, this treats all patches equally, failing to emphasize diagnostically relevant regions.

Instead, we propose to realize **slide-level** pathology DA by adopting the more adaptive attention-based aggregation (ABMIL). This presents two major challenges: **(1) The distortion of multi-scale feature distributions**: The

success of ABMIL relies on the domain-level feature distributions, intra-slide patch relations, and patch-level attention patterns. Without a structured adaptation framework, DA can distort slide morphology and misalign critical diagnostic features. **(2) Computational and structural challenges:** The large number of patches per slide, varying across different slides, introduces significant computational overhead, necessitating an efficient adaptation strategy.

To solve these problems, we propose in this work an efficient Hierarchical Adaption framework for Slide-level Domain-shift (HASD). Specifically, HASD achieves multi-scale feature consistency and computationally efficient DA through two key components (Fig. 1): **(1) Hierarchical Adaptation Framework**: Domain-level Alignment Solver, where an entropic Sinkhorn-Knopp solver [3] ensures effective alignment of feature distributions between domains; Slide-level Geometric Invariance Regularization, ensuring that slides are adapted as a whole without structural distortions; Patch-level Attention Consistency Regularization, which stabilizes domain adaptation by ensuring critical diagnostic features remain consistently focused across domains. **(2) Efficient Prototype Selection**: To mitigate the computational burden of the HASD, we select the K most informative prototypes per slide, reducing redundancy while preserving essential slide-level information. Patch-level features are extracted using a pre-trained foundation model (UNI) [2], ensuring robust representation learning.

We demonstrate our method's DA capabilities in addressing imaging variances and label discrepancies with two slide-level tasks across five datasets. Compared to SOTA methods, our hierarchy adaption framework achieves an average **4.1% AUROC** gain for Breast Cancer Grading and a **3.9% C-index** gain in UCEC Survival Prediction task, **without** requiring additional pathologist annotations. As such, our method provides a practical solution for pathology institutions seeking to transfer models from a source center to a target center while addressing domain shift, ensuring reliable adaptation with minimal computational overhead and annotation costs for slide-level tasks.

2 Methods

2.1 Problem Definition

Multiple Instance Learning. In a slide-level task with a cohort of N slides, each slide is treated as a bag B containing a different number of P patches extracted by a pretrained encoder into vectors $\{r_1, r_2, ..., r_p\} \in \mathbb{R}^M$. Under the MIL assumption, each slide B carries a single label y (e.g., a cancer diagnosis), while individual patches lack direct labels. ABMIL [10] learns an attention mechanism $Atten(\cdot)$ to aggregate the set of patch features into a slide-level prediction $\hat{y}$.

Slide-Level Domain Shift. We consider two domains: a source domain $\mathbf{R}^{src}$ where the model is trained, and a target domain $\mathbf{R}^{tgt}$ where the model is tested. Slide-level domain shift cause the distribution of features in $\mathbf{R}^{src}$ to differ from
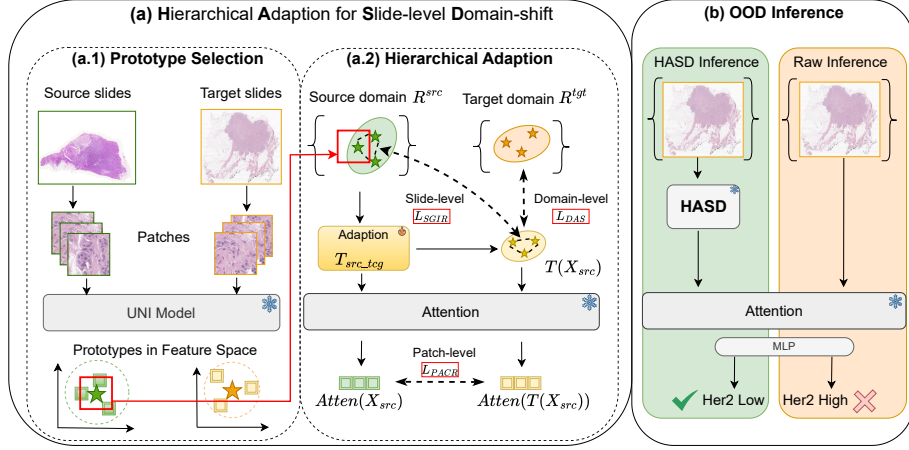
Fig. 2: HASD is designed for scenarios where UNI model extracts the patch features from both centers, and the attention aggregation with the MLP classifier are trained on the source domain. (a): The training process of HASD includes the selection of label-free prototypes in each slide and domain adaption in the hierarchical framework, including domain-, slide- and patch-level regularization. (b): Inference with or without HASD.

that in $\mathbf{R}^{tgt}$. Formally, we represent the source and target domains as:

$$\mathbf{R}^{src} = \{\mathbf{r}^{src}_{1,1}, \mathbf{r}^{src}_{1,2}, \ldots, \mathbf{r}^{src}_{n,p}\}, \quad \mathbf{R}^{tgt} = \{\mathbf{r}^{tgt}_{1,1}, \mathbf{r}^{tgt}_{1,2}, \ldots, \mathbf{r}^{tgt}_{n,p}\},$$

where $\mathbf{r}^{src}_{n,p}$ denotes the pth patch vector of the nth slide.

Optimal Transport (OT) [20] provides a solution for learning a mapping between $\mathbf{R}^{src}$ and $\mathbf{R}^{tgt}$ while minimizing the overall cost: $\gamma\, c\,(\mathbf{R}^{src}, \mathbf{R}^{tgt})$, with $c(\cdot, \cdot)$ measures the feature distance (e.g., cosine similarity) and γ matches samples from source to target.

Traditional OT is not well-suited for pathology slide-level domain adaptation, as it incurs significant computational overhead [24] and treats each patch as an isolated point, thereby ignoring the structured slide morphology and failing to preserve multi-scale feature consistency [14]. In the following section, we introduce our method to address these challenges.

2.2   Hierarchical Alignment Framework

Domain-level Alignment Solver (DAS). Instead of directly using OT for domain adaptation, we propose optimizing a transformation function T that maps $\mathbf{R}^{src}$ into a transition domain while minimizing the transport costs between $T(\mathbf{R}^{src})$ and $\mathbf{R}^{tgt}$. To efficiently compute the transport costs, we adopt a Sinkhorn–Knopp (SK) solver [3] by introducing an entropic regularization term

$\mathcal{H}(\gamma)$. Besides image variance, label discrepancy also poses a further challenge in slide-level domain adaptation. To this end, we further introduce a partial mass relaxation $\mathcal{R}_{\mathrm{partial}}(\gamma)$, allowing for samples to be excluded from matching by relaxing the marginal constraints via Kullback–Leibler (KL) divergence. Accordingly, we learn the transformation T under the guidance of the total transport cost:

$$\mathcal{L}_{\mathrm{DAS}}(T, \gamma) \; = \; \sum_{i,j} \gamma_{ij}\, c\left(T\!\left(\mathbf{R}^{src}\right), \mathbf{R}^{tgt}\right) \; + \; \varepsilon\mathcal{H}(\gamma)) + \tau\mathcal{R}_{\mathrm{partial}}(\gamma),$$

where γ and τ control the strength of entropic regularization and partial relaxation, respectively. In our method, γ is fixed as 0.001, and we only activate the partial mass relaxation term (i.e. set $\tau \neq 0$) when label prevalence discrepancies occur across centers. In such cases (Fig. 4 for an example), we refer to the resulting approach as **Partial DAS**.

Slide-level Geometric Invariance Regularization (SGIR). While DAS allows us to handle domain-level disparities, it operates on each patch independently, and we risk losing the local relationship between patches that belong to the same slide. Therefore, rather than treating patches as isolated points during alignment, we treat the slide as a whole and incorporate a slide-level constraint to enforce structural consistency in feature space. For this, we define our SGIR loss as:

$$\mathcal{L}_{\mathrm{SGIR}}(T) \; = \; \sum_{n}^{N} \left\| G(B_n^{src}) - G(T(B_n^{src})) \right\|_F, \quad \text{where} \quad B_n^{src} = \{ r_{n,1}^{src}, \ldots, r_{n,p}^{src} \},$$

where $\| \cdot \|_F$ denotes the Frobenius norm and the Gram matrix $G(X) = XX^T$. By preserving patch pairwise distance within the same slide, we ensure the slide-level structural consistency.

Patch-level Attention Consistency Regularization (PACR). While domain-level and slide-level constraints preserve global structure, the adaptation can be over-warped, causing clinically essential patch-level local cues to be overlooked [14]. To address this, we introduce PACR to preserve the aggregator's patch-level attention distributions learned on the source domain. Intuitively, a patch with high attention in the source domain should maintain high attention after transformation by T. Formally, we define

$$\mathcal{L}_{\mathrm{PACR}}(T) \; = \; \sum_{n}^{N} \sum_{p}^{P} \left\| Atten(T(r_{n,p}^{src})) \; - \; Atten(r_{n,p}^{src}) \right\|^2.$$

We thus form the total optimization objective as:

$$\mathcal{L}_{total}(T, \gamma) = \mathcal{L}_{DAS}(T, \gamma) + \lambda_1 \mathcal{L}_{\mathrm{SGIR}}(T) + \lambda_2 \mathcal{L}_{\mathrm{PACR}}(T),$$

with λ_1 and λ_2 balancing the domain alignment and regularization.

Overall, with our hierarchical adaption framework, we learn a transformation function $T : \mathbf{R}^{src} \rightarrow \mathbf{R}^{tgt}$, by minimizing the total objective function L_{total}, so that the model trained on $\mathbf{R}^{src}$ generalizes better to $\mathbf{R}^{tgt}$.

2.3   Label-free Prototype Selection

To be computationally efficient, we cluster each slide's patches into k groups and use the clusters' centroids as label-free prototypes (see ablation for k-selection) before applying HASD. This strategy significantly reduces complexity while preserving crucial slide-level information for alignment. An overview of our method is illustrated in Fig 2.

3   Experiments and Results

We validate our method with two slide-level tasks across five datasets:

Breast Cancer HER2 Status. The prediction of the HER2 status in breast cancer from hematoxylin & eosin (**H&E**) stained tumor tissue is a challenging slide-level task [12]. We evaluate our method on multi-institutional breast cancer slides from three centers: Yale Hospital (**Yale**, n=192) [7], The Cancer Genome Atlas (**TCGA**, n=182) [23], and a local hospital (**INT**, n=77).

UCEC Survival Prediction. Uterine corpus endometrial carcinoma (UCEC) is a significant form of gynecological cancer [19]. Reliable survival prediction for personalized treatment planning is limited by domain shift across different centers. To evaluate our method, we employ UCEC cases from two independent centers: **TCGA-UCEC** (n=504) and **CPTAC-UCEC** (n=205) [4]. We use disease-specific survival (**DSS**) and the Concordance Index (**C-Index**) to evaluate performances. In this cohort, domain shift arises not only from image variances (scanners, standings, etc.) but also from **label prevalence variances** (Fig. 4a), posing an additional challenge for adaptation.

We conduct the domain adaption experiments in a **cross-validation** manner: for each center, 80% data is used to train a source-domain model, which is then evaluated on (1) the rest 20% in-domain test splits (**ID**) or (2) the entirety (100%) of each out-of-domain target datasets in the same cohort, completely held-out from the source training split (**OOD**, denoted as $source \rightarrow target$).

3.1   Quantitative Evaluation

We exhaustively compare HASD in eight OOD and five ID setups with the following four existing methods: (1) Attention Based Multi Instance Learning (**ABMIL**). We include a native ABMIL approach trained on the source data and directly applied to the target data. (2) **Patch-level Stain Normalization.**

Widely used to mitigate domain shift on patch-level. We apply the Reinhard stain normalization [16], instead of Macenko due to its much higher runtime, to transform all target patches into the source domain's color space. (3) Multi Instance SimpleShot **(MI-SimpleShot).** Handles domain shifts by constructing class prototypes from slide features obtained by mean pooling of patch-level features [2]. (4) **Supervised Contrastive Domain Adaptation (SCDA).** Builds upon MI-SimpleShot with supervised contrastive learning on the source domain to refine class prototypes [1].

Table 1: ID vs. OOD performance gap.

Her2 Grading Task

Methods	ID Average	OOD Average
MI-SimpleShot	81.6	70.8(10.8↓)
SCDA	82.5	72.0(10.5↓)
ABMIL	85.8	75.8(10.0↓)
ReinhardNorm	85.3	75.3(10.0↓)
Ours	**86.1**	**79.9(6.2↓)**

Survival Prediction Task

Methods	ID Average	OOD Average
MI-SimpleShot	64.9	52.9(12.0↓)
SCDA	65.3	54.7(10.6↓)
ABMIL	68.1	57.5(10.6↓)
ReinhardNorm	**68.5**	56.3(12.2↓)
Ours	68.1	**61.4(6.7↓)**

Fig. 3: OOD Performances on HER2 Grading (red) and Survival Prediction (cyan).
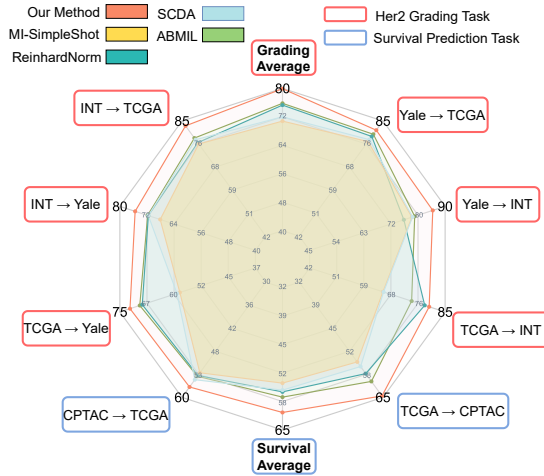
Measuring Slide-level Domain Adaption. We report the performance of all evaluated methods under various OOD setups (see Fig. 3). In the HER2 Grading task, our method consistently outperforms others except for setup $TGCA \rightarrow TUM$, where stain normalization achieves the best result. Our model surpasses the next-best method by **6.0% AUC in the most favorable case**, with an average improvement of **4.1% across all setups**. For the Survival Prediction task, our method, enhanced by partial DAS to address label prevalence differences, consistently outperforms all baselines, achieving an average **CI improvement 3.9%** over the next-best method.

ID vs. OOD Gap Robustness Analysis. Evaluating the ID vs. OOD gap is essential for assessing **domain shift robustness**. A smaller gap indicates better generalization. Table 1 summarizes the differences. For HER2 Grading, MI-SimpleShot suffers the largest drop (10.8% AUC), while our method maintains the smallest gap (6.2% AUC) with the highest OOD score, demonstrating

Table 2: Ablation Study for Parameter Selection.

DAS	# Prototype k	PACR	SGIR	Grading Average	Survial Prediction Average
✓	1			75.4	58.2
✓	5			76.2(0.8↑)	58.6(0.4↑)
✓	10			76.1(0.1↓)	59.1(0.5↑)
✓	10	✓		79.6(3.5↑)	58.6(0.5↓)
✓	10	✓	✓	79.9(0.3↑)	60.1(1.5↑)
Partial DAS	10	✓	✓	-	61.4(1.3↑)

strong generalization. In Survival Prediction, domain shift is further influenced by **label prevalence differences**. All other methods drops $> 10\%$ CI, while our method, aided by **partial DAS**, achieves the best OOD performance with a smaller gap (6.7% CI), proving its effectiveness.

3.2   Ablation study

We summarize the impact of hierarchical components, including DAS, PACR, and SGIR, and the number of prototypes on both slide-level tasks in Table 2. Due to GPU memory constraints, the prototype number is limited to 10.

3.3   Quantitative Evaluation of Image & Label Variance Alignment

We illustrate how our method addresses **image variances** and alleviates **label prevalence variance** in t-SNE plots (Fig. 4) visualizing the aggregated slide-level features in the Survival Prediction task before and after adaptation.
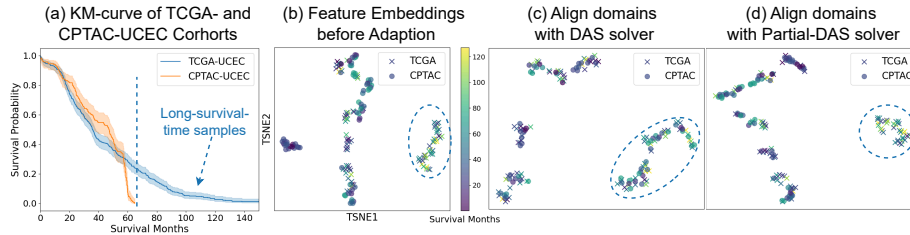

Fig. 4: T-SNE for the Survival task. (a) The TCGA data contain more long-surviving patients than CPTAC. (b) A clear domain shift exists between centers. (c) Balanced SB solver misaligns long-surviving TCGA samples with CPTAC, leading to poor adaptation. (d) Partial SB solver allows unmatched long-survival samples to remain unaligned, preserving clinical label relevance.

4   Conclusion

Domain shift is a critical problem for pathology AI as pathology data is heavily influenced by center-specific conditions. Current pathology domain adaptation methods focus on image patches rather than WSI, thus failing to capture global WSI features required in typical clinical scenarios. We, for the first time, introduced a slide-level domain adaption method HASD, which incorporates a hierarchical adaptation design to maintain multi-scale feature consistency and a prototype selection mechanism to achieve computationally efficiency. Our method outperformed conventional DA methods in two tasks (Her2 grading and Survival) by 4.1% and 3.9%, when models were trained and tested in different centers, while preserving in-domain performance. We believe that our research on slide-level domain adaptation will improve the generalizability of future clinical AI models. Code is available at anonymous.

References

1. Carretero, I., Meseguer, P., del Amor, R., Naranjo, V.: Enhancing whole slide image classification through supervised contrastive domain adaptation. arXiv preprint arXiv:2412.04260 (2024)
2. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**(3), 850–862 (2024)
3. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems **26** (2013)
4. Edwards, N.J., Oberti, M., Thangudu, R.R., Cai, S., McGarvey, P.B., Jacob, S., Madhavan, S., Ketchum, K.A.: The cptac data portal: a resource for cancer proteomics research. Journal of proteome research **14**(6), 2707–2713 (2015)
5. Falahkheirkhah, K., Lu, A., Alvarez-Melis, D., Huynh, G.: Domain adaptation using optimal transport for invariant learning using histopathology datasets. arXiv preprint arXiv:2303.02241 (2023)
6. Fan, J., Lv, T., Di, Y., Li, L., Pan, X.: Pathmamba: Weakly supervised state space model for multi-class segmentation of pathology images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 500–509. Springer (2024)
7. Farahmand, S., Fernandez, A.I., Ahmed, F.S., Rimm, D.L., Chuang, J.H., Reisenbichler, E., Zarringhalam, K.: Deep learning trained on hematoxylin and eosin tumor region of interest predicts her2 status and trastuzumab treatment response in her2+ breast cancer. Modern Pathology **35**(1), 44–51 (2022)
8. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. Journal of machine learning research **17**(59), 1–35 (2016)
9. Ganz, J., Puget, C., Ammeling, J., Parlak, E., Kiupel, M., Bertram, C.A., Breininger, K., Klopfleisch, R., Aubreville, M.: Assessment of scanner domain shifts in deep multiple instance learning. In: BVM Workshop. pp. 137–142. Springer (2024)
10. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)

11. de Jong, E.D., Marcus, E., Teuwen, J.: Current pathology foundation models are unrobust to medical center differences. arXiv preprint arXiv:2501.18055 (2025)
12. Liu, S., Zhu, C., Xu, F., Jia, X., Shi, Z., Jin, M.: Bci: Breast cancer immuno-histochemical image generation through pyramid pix2pix. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 1815–1824 (June 2022)
13. Naranjo, V.: Domain adaptation for unsupervised cancer detection: An application for skin whole slides images from an interhospital dataset. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer (2024)
14. Peyré, G., Cuturi, M., et al.: Computational optimal transport: With applications to data science. Foundations and Trends® in Machine Learning 11(5-6), 355–607 (2019)
15. Pocevičiūtė, M., Eilertsen, G., Garvin, S., Lundström, C.: Detecting domain shift in multiple instance learning for digital pathology using fréchet domain distance. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 157–167. Springer (2023)
16. Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Computer graphics and applications 21(5), 34–41 (2001)
17. Reisenbüchler, D., Luttner, L., Schaadt, N.S., Feuerhake, F., Merhof, D.: Unsupervised latent stain adaptation for computational pathology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 755–765. Springer (2024)
18. Sharma, Y., Syed, S., Brown, D.E.: Mani: Maximizing mutual information for nuclei cross-domain unsupervised segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 345–355. Springer (2022)
19. Somasegar, S., Bashi, A., Lang, S.M., Liao, C.I., Johnson, C., Darcy, K.M., Tian, C., Kapp, D.S., Chan, J.K.: Trends in uterine cancer mortality in the united states: a 50-year population-based analysis. Obstetrics & Gynecology 142(4), 978–986 (2023)
20. Thorpe, M.: Introduction to optimal transport. Notes of Course at University of Cambridge (2018)
21. Wang, P., Li, P., Li, Y., Xu, J., Jiang, M.: Classification of histopathological whole slide images based on multiple weighted semi-supervised domain adaptation. Biomedical Signal Processing and Control 73, 103400 (2022)
22. Wang, Z., Zhu, X., Su, L., Meng, G., Zhang, J., Li, A., Wang, M.: Instance-aware feature alignment for cross-domain cell nuclei detection in histopathology images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. pp. 499–508. Springer (2021)
23. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M.: The cancer genome atlas pan-cancer analysis project. Nature genetics 45(10), 1113–1120 (2013)
24. Yeaton, A., Krishnan, R.G., Mieloszyk, R., Alvarez-Melis, D., Huynh, G.: Hierarchical optimal transport for comparing histopathology datasets. arXiv preprint arXiv:2204.08324 (2022)
25. Yin, C., Liu, S., Wong, V.W.S., Yuen, P.C.: Histosyn: Histomorphology-focused pathology image synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 200–210. Springer (2024)