OPTIMAL TRANSPORT WITH FOUNDATION MODELS FOR MULTI-STAIN CBIR IN DIGITAL PATHOLOGY

Maxime Amodei[⋆] *Raphaël Marée*[⋆] *Pierre Geurts*[⋆]

[⋆] Institut Montefiore, University of Liège, Belgium

ABSTRACT

Combining multiple staining in histopathology is vital to relate complementary information within a tissue. We present a novel framework that enables local correspondence in the form of patch retrieval (CBIR) in multi-stain whole-slide images without the need for explicit registration. By leveraging foundation models as meaningful patch encoders, we are able to retrieve an H&E patch in a dataset of patches extracted from IHC slides. Using Fused Gromov-Wasserstein optimal transport for weak supervision, we learn a lightweight correction module that aligns embeddings and improves retrieval accuracy by up to 30% (depending on patch size and model), at negligible cost, despite staining and structural variations. Our method paves the way for efficient multi-stain integration in digital pathology workflows.

***Index Terms*—** CBIR, histopathology, multi-stain, IHC, annotations, deep learning.

1. INTRODUCTION

Multiple stainings in histopathology reveal complementary pieces of information that benefit downstream tasks at the expense of a higher cost in staining and annotation. Recent research aims at facilitating annotation and data analysis tasks in such multi-stain contexts [1]. With this practical objective in mind, in this paper, we investigate a novel approach that retrieves patches from one staining to another by locating in *whole-slide images* (WSIs) a small region around a query point (multi-stain content-based image retrieval).

We propose to evaluate, in the multi-stain context, multiple recent foundation models which have been trained independently to provide concise representations of patches and WSIs. We use these models as patch encoders, providing an embedding vector to facilitate the search. To improve the accuracy of this patch retrieval framework in a multi-stain context, we design *correction modules*, which adjust the embeddings from one staining and align them closer to the embeddings of the other domain, thus avoiding pixel-perfect image-to-image translation. We propose to build a training set for the correctors using Fused Gromov Wasserstein (FGW) optimal transport [2]: this algorithm naturally composes structural interactions between objects of the same domain and similarities between objects of different domains to create a globally coherent mapping.

Our contributions are threefold: (1) a cross-stain patch retrieval framework based on foundation-model embeddings, (2) an OT-based approach to automatically and efficiently generate noisy supervision for embedding correction, and (3) a systematic analysis of parameter effects on retrieval accuracy across stainings on a publicly-available dataset. Our code is available at `https://github.com/maxime915/cbir-mm-fgw`.

The paper is organized as follows. Related works are reviewed in Section 2. We provide details of our retrieval approach and correction modules in Section 3. The experimental setup is described in Section 4 and the results are presented in Section 5. Section 6 synthesizes our main findings and discusses future work directions.

2. RELATED WORKS

Our method borrows ideas from content-based image retrieval for its structure, image-to-image translation for the correction implementation, multi-stain image registration, and foundation models.

CBIR is a popular area of research in histopathology [3, 4, 5, 6, 7]. Finding similar patches or WSIs to a given query can help with diagnosis or annotations [8]. Most of these works, however, either focus on single-stain datasets or do not assess the cross-stain robustness of these systems. Our contribution with respect to this literature is to highlight that feature extractors based on foundation models are not robust for cross-stain retrieval and to propose a lightweight correction method to improve their performance. Adapting embedding vectors from one staining is akin to **image-to-image translation**, domain adaptation, stain transfer, and multi-modal learning, leading to many interesting approaches. Multiple approaches [9, 10, 11] learn a form of H&E to IHC stain translation but require a dataset of matching patches. We design our approach at multiple resolutions and on datasets of paired WSIs (not patches), avoiding the need for matching patches. Madelaine [12] learns a multi-stain representation in a much more complex setup than our correction modules.

Another related problem is **multi-stain image registration**, whose goal is to align WSIs of the same or close sections obtained with different staining [13, 14, 15]. These methods offer a potential solution to our problem when the query patch must be retrieved from a re-stained image of the same section from which it was extracted. The problem we target, however, is more general. Furthermore, research is still ongoing on this topic, and the performance and execution speed of these methods vary significantly [16, 17].

We focus our attention on **foundation models** suitable for histopathology images. Multiple models have been designed in recent years, based on different deep learning models, often providing support for other modalities (text, videos) depending on their intended use [18]. They are trained with diverse materials [19]. Although no model dominates across all application scenarios, [20] has shown that domain-aligned pretraining yields better downstream performance.

3. METHODS

Our general multi-stain patch retrieval framework is described in Section 3.1. To overcome the limitation from the difference in staining, we propose a first correction method, doing a naive correspondence between H&E and PHH3 in Section 3.2. We then propose a way to improve the multi-stain correspondence in Section 3.3 and its practical implementation in Section 3.4.

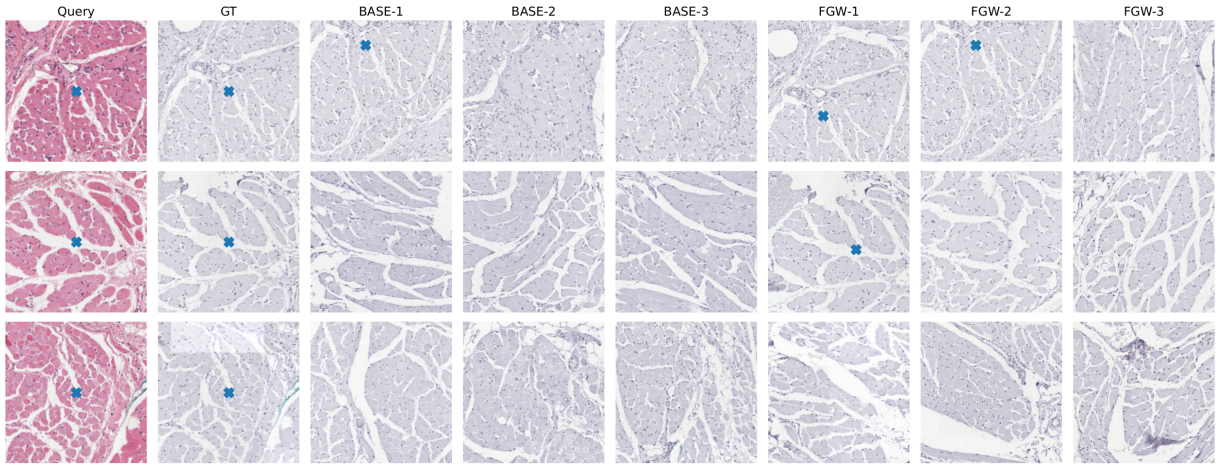

Fig. 1. Examples of queries, ground truth, and UNI-2-h results at 500µm. From left to right: query, matching ground truth, first three patches retrieved without correction, and first three retrieved with FGW correction. Landmarks, when present, are marked with a cross.

3.1. Multi-Stain Content-Based Patch Retrieval

We seek to find an object represented by a query patch Q inside a database of target patches $\{T_i\}_{i=1}^N$, all sharing the same stain, different from that of the query patch. We consider the case where all patches are square with an imposed size. Rather than working with pixels, we compute an embedding vector for each patch using a foundation model FM, and use the cosine similarity S_{cos} (i.e., the inner product of ℓ_2-normalized embeddings) to obtain the set $\mathcal{R}_k(Q)$ of the k most similar target patches to Q:

$$\mathcal{R}_k(Q) = \mathrm{Top_k}\{S_{cos}(\mathcal{C}(\mathrm{FM}(Q)), \mathrm{FM}(T_i))\}_{i=1}^N.$$

Differences in staining can substantially degrade retrieval accuracy, if the embedding FM is not robust to modality shifts. To achieve this robustness, a correction module $\mathcal{C}$ is introduced to align the embedding of the query with the target embeddings. Two instances of this correction module are proposed below.

3.2. Modality Correction by Embedding Renormalization

To address differences between staining modalities, we first adopt a simple per-dimension embedding normalization strategy. Query embeddings are re-centered and rescaled so that their component-wise means and standard deviations align with those of the target modality, estimated from randomly sampled patches. Although effective as a first-order correction, this method ignores correlations between embedding dimensions and assumes stable semantic meaning of these dimensions across modalities, which is oversimplistic.

3.3. Supervision from Optimal Transport

Building on the idea of matching the distribution of embeddings in the two modalities from the previous section, we relax the assumption of independence and consider optimal transport to provide a faithful coupling of the two distributions. To reduce the complexity of the optimal transport computation, we assume a dataset of paired (but not necessarily aligned) WSIs of the same tissue. While a high-resolution registration of the WSIs would yield a similar result, our approach has the potential to work even in cases where registration is challenging, and only requires tuning a single hyperparameter: the number of sampled patches in each image (see below).

We propose the Fused Gromov-Wasserstein Optimal Transport [2] as it is designed to leverage both intra-domain structural similarity and inter-domain direct similarity when optimizing the mapping of objects from one domain to another. From n randomly sampled patches, we use Euclidean distances between patch centers in each image to provide a global structure, and the embeddings of the foundation model as a base for the similarity. Both metrics are rescaled to have a maximal value of 1.0, thus making the tradeoff between structural and direct similarity easier to tweak. Formally, from patches $\{x_i\}$ in one domain and $\{y_i\}$ unrelated patches in the other domain, we denote by D_{ik} the rescaled euclidean distance between the two patches x_i and x_k in the first image, D'_{kj} the distance between y_j and y_l, and C_{ij} the cosine distance between the embeddings of x_i and y_j, the optimal transport plan is given by Equation (1).

$$T^\star = \arg\min_T \alpha \sum_{ijkl} T_{ij}T_{kl}|D_{ik} - D'_{jl}|^2 + (1-\alpha)\sum_{ij} T_{ij}C_{ij}, \tag{1}$$

with $T_{ij} \geq 0$ and $\sum_i T_{ij} = \sum_j T_{ij} = 1/n$. From the optimal transport plan $T^\star_{ij}$, we associate x_i to $y_{j^\star}$ where $j^\star = \arg\max_j T_{ij}$, thus yielding n pairs of patches per pair of images for our supervised dataset. Because this method does not provide a mapping for a new query, only for the sampled patches, we then learn a dedicated correction function as sketched in Section 3.4. The optimization problem in (1) is solved using the Python Optimal Transport (POT) library [21, 22]. This method does not require any learning: the mapping is done once for each resolution and all machine learning happens in Section 3.4.

Depending on the patch size and the number of patches n, it is possible that some patches do not have a close match in the other image: this is seemingly a limitation of our method. While the matching may not be perfect, we hope that an approximate matching will be enough to build an effective mapping as described in Section 3.4. We will check the quality of the matching and its impact on retrieval performance in Section 5.3.

3.4. Learning a correction function

Relying on the supervised dataset obtained in Section 3.3, we formulate the correction as a regression task between the embeddings of the two modalities. For embeddings with d components, we propose

a residual MLP with two layers of size d and leaky ReLU activations, initialized with identity weights ($\mathbf{I}_d$) and null biases ($\mathbf{0}_d$). In addition, we learn the correction in both directions at the same time and impose a cycle consistency constraint in both directions (both using an ℓ_2 loss, with a regularization of 0.1), drawing inspiration from [23]. The objective is minimized using Adam with a weight decay of 0.01 and a learning rate varying from 10^{-3} to 10^{-6} over 200 epochs of cosine annealing. We further split the set of embedding pairs into a training set used to learn the correction and a validation set on which we perform early stopping.

4. EXPERIMENTAL PROTOCOL

The quantitative evaluation of CBIR system in general is challenging, as there is usually no explicit information about which patch should be retrieved for a given query. Given our multi-stain setting, we propose here to exploit an existing annotated dataset of paired multi-stained images. CBIR performance will then be assessed by using annotations in one stain as queries and measuring the capacity of the system to retrieve the same annotation in the other stain. The dataset is described in Section 4.1 and the protocol in Section 4.2.

4.1. Dataset

The HyReCo dataset [1] contains 54 pairs of WSI stained in H&E and then cleaned and re-stained in PHH3, each with 43 annotations in the form of matching landmarks in the two images. HyReCo also contains other WSIs in other IHC stains, but they were discarded as they contain too few landmarks for evaluation. All images have a resolution of 0.41 pixels per micron, with dimensions in the order of 10^4 microns and 10^5 pixels. The dataset is split such that the first 27 pairs of WSI compose the training set, and the next 11 the test set, on which we compare our methods.

4.2. Protocol

From a pair of *query* (here, H&E) and *target* (here, PHH3) images, we extract square patches centered around the landmark annotations in the query image and uniformly tessellate the target image into tiles of the same size as the query patches. They overlap such that every vertex or midpoint of a tile in a tiling without overlap becomes the center of a new tile, thus reducing the worst-case distance from any point to the center of a tile. In case a tile goes beyond the image, a black background is used. We ignore the landmarks in the tessellation to avoid introducing an overoptimistic bias. As a consequence, except by chance, there is no tile extracted from the target image that matches exactly the query patch.

With this setup, the retrieval method selects the k tiles in the target images that are the most similar to the patch in the query image. We measure performance using the top-k accuracy, i.e., the proportion of queries where the landmark is included in at least one of the k selected tiles. Given the way tiles were extracted, there are always two tiles that contain each landmark. An example of predictions is shown in Figure 1.

The foundation models in our study include UNI [24] and UNI-2, two DINOv2-based models pretrained self-supervised on histopathology images. UNI-2 uses a multi-stain dataset, while UNI-1 uses only H&E slides, producing 1536- and 1024-dimensional embeddings, respectively. We also evaluate Gigapath [25] (H&E and IHC; 1536-D, Apache 2.0), H-optimus-0 [26] (H&E; 1536-D, Apache 2.0), and Virchow [27] (H&E; 1280-D, Apache 2.0).

5. RESULTS

Section 5.1 discusses the impact of the stain when the feature extractors are uncorrected, while Section 5.2 compares the two proposed correction methods to this baseline. Section 5.3 presents additional experiments to improve our understanding of the behavior of the optimal transport correction method. Eventually, Section 5.4 investigates the impact of rotation and displacement on retrieval performance. The main results are collected in Table 1.

5.1. Impact of staining on uncorrected models

To study the impact of the WSI staining on retrieval performance, we compare the top-1 and top-5 accuracies in the normal case and in the case where the same H&E WSI is used for both the query and target patches (resp. the *BASE*, for baseline, and *UM*, for unimodal, columns in Table 1). Table 1 shows a very important gap between the two for all models and tile sizes, highlighting that cross-stain retrieval is very challenging and justifying the need for a correction method. In the multimodal case, UNI-1 scores higher than other models for all tile sizes, even UNI-2-h. This might be explained by their difference in training data. Since UNI-2-h has been trained on multiple stains data, unlike UNI-1, it might produce more stain-specific features, which is detrimental for cross-stain CBIR. Unsurprisingly, the accuracies increase with the tile size, independently of the other variables. We also tested VGG as a simpler baseline, which performed slightly worse than Virchow but omit it from the table for brevity.

5.2. Correction methods

We compare in Table 1 the correction methods using renormalization (*NORM*, Section 3.2) and using FGW pairing (*FGW*, Section 3.4) against the *baseline* without any correction and the uni-modal upper bound (*UM*, using H&E as the query and target images). For both corrections, we extract $n = 1000$ random patches from each training image, and fix $\alpha = 0.5$ (the default) for FGW. While both methods provide an improvement over the baseline, the correction methods learned on FGW paired patches fare significantly better than renormalization for almost all configurations. The improvement is very significant for the smallest patches, with a top-1 accuracy jump of about 40% with respect to BASE and of more than 30% with respect to renormalization. While UNI-1 was the best performer uncorrected, the combo UNI-2-h + FGW stands out as the best option overall.

5.3. Evaluating the Optimal Transport Pairing

We would like here to assess the quality of the pairing returned by the optimal transport method, to better understand the good performance observed in the previous subsection. To do so, we extract n randomly sampled patches in both the H&E WSI and the PHH3 image (as described in Section 3.3) and introduce one additional patch centered around each landmark in the H&E set, and an equal number of randomly sampled patches in the PHH3 set. Using the extracted transport plan, we observe the proportion of H&E landmark patches that are matched to an PHH3 patch that includes the landmark and report it as the accuracy of our method. For low enough values of n, there might be no patch in the PHH3 set that includes a given landmark. As an upper bound, we thus also report the *ideal accuracy*, corresponding to the proportion of landmarks in H&E for which the matching landmark in PHH3 is included in at least one patch. The results are shown in Table 2 for UNI-2-h. We fix $n = 1000$ for all

		500µm					1000µm					5000µm				
		GP	H-0	U-1	U-2	Vir	GP	H-0	U-1	U-2	Vir	GP	H-0	U-1	U-2	Vir
Top-1	BASE	21.30	4.94	**34.88**	25.31	3.70	2.47	17.90	**52.47**	34.57	6.17	41.98	51.23	**79.01**	66.67	40.74
	NORM	37.04	10.19	**42.59**	35.49	7.10	11.11	29.01	**58.95**	44.44	16.36	67.59	64.51	**83.95**	75.62	55.86
	FGW	25.62	33.33	59.26	**74.38**	35.49	32.72	45.99	71.60	**72.84**	44.75	68.83	67.90	**89.81**	83.33	60.49
	UM	**99.07**	95.99	98.46	96.30	78.70	**99.38**	97.53	99.07	98.46	86.73	**99.07**	**99.69**	**99.69**	**99.69**	96.30
Top-5	BASE	37.65	12.04	**51.54**	44.44	8.33	9.26	36.73	**75.62**	58.95	18.52	77.78	83.33	**95.37**	90.43	75.31
	NORM	53.09	22.22	**61.73**	54.01	15.74	26.54	52.47	**79.94**	70.37	33.33	88.58	91.67	**96.91**	93.83	87.35
	FGW	40.74	49.38	78.09	**90.12**	57.10	54.32	69.14	91.05	**91.98**	71.30	93.52	88.89	**98.46**	96.91	85.19
	UM	**100.00**	**100.00**	99.69	**100.00**	91.98	**100.00**	99.69	**100.00**	**100.00**	97.22	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Table 1. Top-1 and top-5 accuracies across configurations. The baseline (BASE, no correction) yields the lowest performance, while the uni-modal setting (UM) achieves the highest. Among correction methods (renormalization: NORM; FGW pairing: FGW), the best result per model and tile size is underlined, and the best-performing model among **gigapath**, **H-optimus-0**, **UNI-1**, **UNI-2-h**, and **Virchow** is shown in bold. FGW achieves the highest scores at all resolutions. Overall, corrected UNI-1 and UNI-2-h consistently outperform the others.

	500µm	1000µm	5000µm
acc(%)	10.86	28.55	82.91
acc*(%)	18.38	53.71	100.00
t(s)	0.81	0.77	0.71

Table 2. FOT pairing accuracy using UNI-2-h.

	500µm	1000µm	5000µm
T_1	41.67	43.21	80.56
T_5	64.20	71.91	95.06

Table 3. Top-1 and top-5 accuracy with a correction for UNI-2-h trained using a random pairing.

Tile size	BASE	FGW 500µm	FGW 1000µm	FGW 5000µm
500µm	25.31	**74.38**	69.14	19.14
1000µm	34.57	57.72	**72.84**	27.78
5000µm	66.67	71.91	73.15	**83.33**

Table 4. Top-1 accuracy of UNI-2-h with correctors obtained at different resolutions.

configurations. Both the expected accuracy (acc(%)) and the accuracy of a simulated optimal pairing (acc*(%)) increase steadily with tile size. However, the gap between them remains significant, showing that our method has room for improvement. Note that pairing patches fully at random would lead to an average accuracy lower than 2.5% for all patch sizes, meaning that FGW pairing is nevertheless much better than random.

To further measure the impact of the FGW pairing, as an ablation experiment, we trained a new correction model using a fully random pairing of the patches in place of the FGW pairing. Top-1 and top-5 accuracies with this correction model are shown in Table 3. Comparing with the results in Table 1, the FGW pairing consistently scores higher than the random one. Random pairing actually achieves very close performances to the renormalization correction. Despite not being perfect, the FGW pairing is nevertheless crucial to reach optimal performance.

So far, a separate correction module was trained for each patch size. Hoping to avoid this expense, we investigate the performance of each size-specific correction module in all three resolutions for UNI-2-h and compare them against the baseline (no correction, BASE) using the Top-1 accuracy in Table 4. Unfortunately, changing the resolution often leads to worse results than the baseline, and is always significantly worse than the appropriate corrector.

5.4. Studying the impact of rotation and displacement

We introduce small disturbances in the experiment: before extracting the pixels corresponding to a patch in the H&E image, we apply a rotation by an angle θ and a small displacement d from the center of the landmark in a random direction. The landmark coordinates are kept the same, and we record the top-1 accuracy for angles in increments of 45° and $d \in \{0, 0.35w, 0.71w\}$ where w is the tile width. We fix the tile size to 1000µm and use UNI-2-h.

Table 5 shows a notable decrease in accuracy for all non-zero angles without a sharp drop. The drop going from $d = 0$ to $d = 0.35w$ is significantly lower than the drop going to $d = 0.71w$, for all rotations. Interestingly, uni-modal retrieval seems to be less affected by rotations than multi-stain retrieval, while the FGW correction and the baseline are equally impacted on the other hand.

d	0			$0.35w$			$0.71w$		
θ(°)	BASE	FGW	UM	BASE	FGW	UM	BASE	FGW	UM
0	**34.57**	**72.84**	**98.46**	**30.25**	**61.73**	**82.41**	**14.20**	**18.52**	19.75
45	29.94	65.58	97.53	26.85	54.94	81.79	11.73	16.67	18.83
90	28.70	**55.25**	95.68	23.77	**46.60**	82.10	10.80	16.36	**21.60**
135	**26.54**	59.57	**95.37**	**20.99**	50.31	**80.56**	**10.19**	**15.12**	**18.52**
180	29.94	61.42	**95.37**	25.93	49.38	**80.56**	11.11	17.28	18.83

Table 5. Top-1 accuracy for UNI-2-h at 1000µm without correction (BASE), with FGW correction, and in the uni-modal setting (UM).

6. CONCLUSION

In this paper, we investigated the use of foundation models as encoders for cross-stain patch retrieval and found that, while they perform well in single-stain contexts, their representations degrade notably in multi-stain settings and under orientation changes. These limitations highlight that current foundation models lack sufficient stain invariance for reliable multi-stain retrieval. To address this, we proposed a lightweight correction module integrated into the retrieval pipeline, effectively reducing the cross-stain performance gap for multiple foundation models. By leveraging fused Gromov-Wasserstein optimal transport, we demonstrated that pixel-perfect aligned images are not required to learn an effective cross-domain mapping. While the accuracy could be further improved, our findings suggest that patch-level retrieval can serve as an effective method to help with multi-stain analysis. As future work, we plan to extend our approach to multiple IHC stains, to improve the OT pairing accuracy, and to enhance the search accuracy in smaller tiles.

7. ACKNOWLEDGEMENTS

This work was supported by Service Public de Wallonie Recherche under grant n° 2010235 - ARIAC by DIGITALWALLONIA4.AI and by IMI 2 Joint Undertaking under grant agreement No 945358 (BIGPICTURE).

The present research benefited from computational resources made available on Lucia, the Tier-1 supercomputer of the Walloon Region, infrastructure funded by the Walloon Region under the grant agreement n°1910247.

8. REFERENCES

[1] Jeroen van der Laak, Johannes Lotz, Nick Weiss et al., "Hyreco – hybrid re-stained and consecutive histological serial sections (cc-by-sa 4.0)," IEEE Dataport, July 2021.

[2] Vayer Titouan, Nicolas Courty, Romain Tavenard et al., "Optimal transport for structured data with application on graphs," in *Proceedings of the 36th International Conference on Machine Learning*, Kamalika Chaudhuri and Ruslan Salakhutdinov, Eds. 09–15 Jun 2019, vol. 97 of *Proceedings of Machine Learning Research*, pp. 6275–6284, PMLR.

[3] Oscar Jimenez-del Toro, Sebastian Otálora, Manfredo Atzori et al., *Deep Multimodal Case–Based Retrieval for Large Histopathology Datasets*, p. 149–157, Springer International Publishing, 2017.

[4] Isaiah Lahr, Saghir Alfasly, Peyman Nejat et al., "Analysis and validation of image search engines in histopathology," *IEEE Reviews in Biomedical Engineering*, vol. 18, pp. 350–367, 2025.

[5] Saghir Alfasly, Ghazal Alabtah, Sobhan Hemati et al., "Validation of histopathology foundation models through whole slide image retrieval," *Scientific Reports*, vol. 15, no. 1, Feb. 2025.

[6] Yushan Zheng, Zhiguo Jiang, Haopeng Zhang et al., "Histopathological whole slide image analysis using context-based cbir," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1641–1652, July 2018.

[7] H. C. Akakin and M. N. Gurcan, "Content-based microscopic image retrieval system for multi-image queries," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 758–769, July 2012.

[8] H.R. Tizhoosh and Liron Pantanowitz, "On image search in histopathology," *Journal of Pathology Informatics*, vol. 15, pp. 100375, Dec. 2024.

[9] Fangda Li, Zhiqiang Hu, Wen Chen et al., *Adaptive Supervised PatchNCE Loss for Learning H&E-to-IHC Stain Translation with Inconsistent Groundtruth Image Pairs*, p. 632–641, Springer Nature Switzerland, 2023.

[10] Xianchao Guan, Yifeng Wang, Ye Zhang et al., "Ot-stainnet: Optimal transport driven semantic matching for weakly paired h&e-to-ihc stain transfer," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 3, pp. 3194–3202, Apr. 2025.

[11] Shengyi Hua, Fang Yan, Tianle Shen et al., "Pathoduet: Foundation models for pathological slide analysis of h&e and ihc stains," *Medical Image Analysis*, vol. 97, pp. 103289, Oct. 2024.

[12] Guillaume Jaume, Anurag Vaidya, Andrew Zhang et al., *Multistain Pretraining for Slide Representation Learning in Pathology*, p. 19–37, Springer Nature Switzerland, Oct. 2024.

[13] Chandler D. Gatenbee, Ann-Marie Baker, Sandhya Prabhakaran et al., "Virtual alignment of pathology image series for multi-gigapixel whole slide images," *Nature Communications*, vol. 14, no. 1, July 2023.

[14] Marek Wodzinski, Niccolò Marini, Manfredo Atzori et al., "Regwsi: Whole slide image registration using combined deep feature- and intensity-based methods: Winner of the acrobat 2023 challenge," *Computer Methods and Programs in Biomedicine*, vol. 250, pp. 108187, June 2024.

[15] Marek Wodzinski and Henning Müller, "Deephistreg: Unsupervised deep learning registration framework for differently stained histology samples," *Computer Methods and Programs in Biomedicine*, vol. 198, pp. 105799, Jan. 2021.

[16] Behnaz Elhaminia, Abdullah Alsalemi, Esha Nasir et al., "From traditional to deep learning approaches in whole slide image registration: A methodological review," *Journal of Pathology Informatics*, vol. 19, pp. 100512, Nov. 2025.

[17] Philippe Weitz, Masi Valkonen, Leslie Solorzano et al., "The acrobat 2022 challenge: Automatic registration of breast cancer tissue," *Medical Image Analysis*, vol. 97, pp. 103257, Oct. 2024.

[18] Dong Li, Guihong Wan, Xintao Wu et al., "A survey on computational pathology foundation models: Datasets, adaptation strategies, and evaluation tasks," *arXiv preprint arXiv:2501.15724*, 2025.

[19] Mohsin Bilal, Aadam, Manahil Raza et al., "Foundation Models in Computational Pathology: A Review of Challenges, Opportunities, and Impact," Feb. 2025, arXiv preprint.

[20] Saghir Alfasly, Peyman Nejat, Sobhan Hemati et al., "Foundation models for histopathology—fanfare or flair," *Mayo Clinic Proceedings: Digital Health*, vol. 2, no. 1, pp. 165–174, Mar. 2024.

[21] Rémi Flamary, Nicolas Courty, Alexandre Gramfort et al., "Pot: Python optimal transport," *Journal of Machine Learning Research*, vol. 22, no. 78, pp. 1–8, 2021.

[22] Rémi Flamary, Cédric Vincent-Cuaz, Nicolas Courty et al., "Pot python optimal transport (version 0.9.5)," 2024.

[23] Jun-Yan Zhu, Taesung Park, Phillip Isola et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

[24] Richard J. Chen, Tong Ding, Ming Y. Lu et al., "Towards a general-purpose foundation model for computational pathology," *Nature Medicine*, vol. 30, no. 3, pp. 850–862, Mar. 2024.

[25] Hanwen Xu, Naoto Usuyama, Jaspreet Bagga et al., "A whole-slide foundation model for digital pathology from real-world data," *Nature*, vol. 630, no. 8015, pp. 181–188, May 2024.

[26] Charlie Saillard, Rodolphe Jenatton, Felipe Llinares-López et al., "H-optimus-0," GitHub repository, 2024.

[27] Eugene Vorontsov, Alican Bozkurt, Adam Casson et al., "A foundation model for clinical-grade computational pathology and rare cancers detection," *Nature Medicine*, 2024.