

Graph-Semantic Guided Learning for Virtual Immunohistochemistry Staining on Consecutive Histology Sections

Fanhao Qiu¹, Yangyang Zhang², Zhengxia Wang^{1*}

¹School of Computer Science and Technology, Hainan University, Haikou, China

²School of Information and Communication Engineering, Hainan University, Haikou, China
fanhaosc@hainanu.edu.cn, zyy929375916@163.com, zxiawang@hainanu.edu.cn

Abstract

Virtual Immunohistochemistry (IHC) staining technology employs generative models to directly synthesize IHC images from Hematoxylin and Eosin (H&E) images, reducing reliance on chemical staining while improving diagnostic efficiency and reducing costs. However, existing virtual staining methods relying on adjacent sections face two critical challenges: insufficient mining of pathological semantics and the spatial misalignment of pathological semantics due to physical discrepancies between sections. To address these, we propose GSGStain, a **Graph-Semantic Guided Learning** for virtual **Staining**. Our method innovatively transforms the problem from pixel space to graph space, enabling semantic noise correction for spatial misalignment features. Specifically, to capture the rich pathological semantics, we construct a cell graph from the H&E image to encode tissue architecture, annotating nodes with noisy biomarker semantic features derived from misaligned adjacent IHC sections. Furthermore, to correct for the semantic misalignment, a Graph Semantic Rectification Module (GSRM) then refines these features using graph contextual reasoning, while a graph semantic consistency loss ensures alignment between generated IHC images and rectified semantics. Additionally, we propose a dual-branch discriminator to compel the generator to match the empirical distribution of real images, significantly improving generation quality. Extensive experiments on two public benchmarks demonstrate that GSGStain significantly outperforms state-of-the-art methods in both image quality and pathological consistency. This work establishes a new paradigm for semantically robust virtual staining.

Introduction

Immunohistochemistry (IHC) staining serves as a cornerstone of cancer diagnostics, providing molecular-level visualization of key protein biomarkers critical for tumor subtyping and prognosis prediction (Zaha 2014). Despite its clinical importance, traditional IHC staining remains constrained by its labor-intensive workflow, high costs, and irreversible tissue depletion. These issues are particularly prominent in rapid diagnosis and resource-limited scenarios (Li et al. 2023). Recent advances in computational pathology have introduced virtual staining technology, which uses

deep learning models to generate IHC images directly from routine Hematoxylin and Eosin (H&E) stained slides (Zhang et al. 2022; Pati et al. 2024; Guan et al. 2025a), with the potential to revolutionize diagnostic workflows by reducing the reliance on chemical staining. The critical challenges in virtual IHC staining lie in achieving accurate stain style transfer while ensuring pathological fidelity during the image-to-image translation. This requires the model to learn semantically precise biomarker mappings, as erroneous mappings could lead to severe clinical misjudgment (Wang et al. 2025a).

Given the physical difficulty of re-staining the same tissue section, existing virtual staining methods predominantly adopt a weakly supervised paradigm. These methods typically utilize clinically common H&E and IHC image pairs from adjacent tissue sections for training (Liu et al. 2022; Li et al. 2023; Ma et al. 2024; Wang et al. 2024, 2025a). Since spatially adjacent sections exhibit strong correlation in pathological information, many state-of-the-art approaches focus on deriving effective supervisory signals from these weakly paired images. For example, some methods enforce statistical consistency in protein expression levels between the generated and reference images using focal optical density maps (Chen et al. 2024a; Peng et al. 2025). Meanwhile, Peng et al. (2024a) extract prior knowledge from adjacent IHC slides to serve as explicit supervisory targets.

Despite their efforts, these methods still face two fundamental limitations. The first is the insufficient mining of pathological semantics. Current approaches typically rely on indirect statistical constraints (Peng et al. 2025) or coarse-grained supervision (e.g., patch-level positive/negative labels (Zeng et al. 2022)), which do not directly link supervisory signals to cell-level visual features. For instance, HER2 positivity assessment depends on fine-grained characteristics, particularly membrane staining intensity and the proportion of positive expression, which may vary even within the same pathological grade (Huang et al. 2022). However, existing methods, which depend on macroscopic statistics or regional labels, struggle to accurately reproduce these diagnostically critical visual cues. Consequently, the generated staining results may exhibit ambiguity, limiting their clinical interpretability like (Chen et al. 2024a).

A more critical challenge stems from the spatial misalignment of pathological semantics. The inherent physical dis-

*Corresponding Author

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

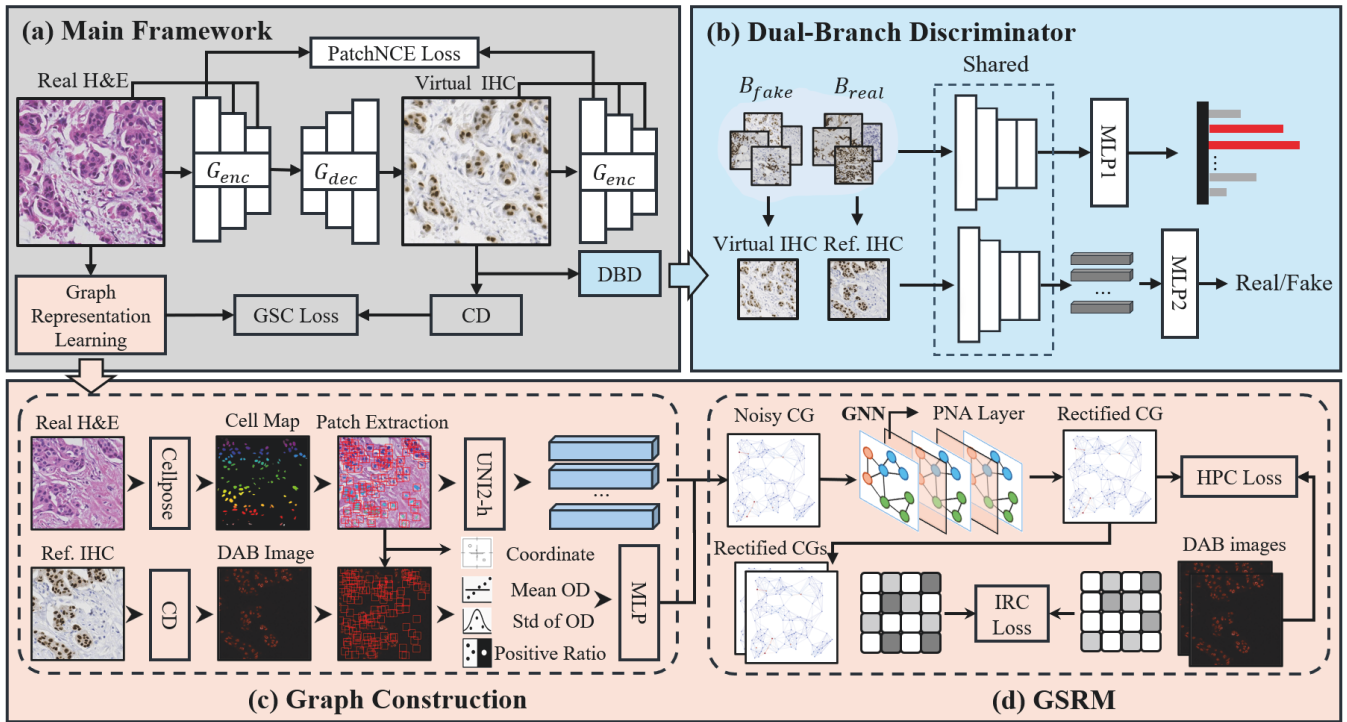


Figure 1: An overview of the GSGStain framework. **(a)** The main generator, consisting of an encoder (G_{enc}) and a decoder (G_{dec}), translates a real H&E image into a virtual IHC, guided by a PatchNCE loss for content preservation and a graph semantic consistency (GSC) loss for semantic accuracy. CD denotes color deconvolution. **(b)** A dual-branch discriminator (DBD) provides adversarial supervision. **(c)** The graph construction generates noisy cell graph (CG) by defining cell nodes from the H&E image and assigning them multi-modal features derived from both H&E and reference IHC images. **(d)** The GSRM refines the noisy CG into a rectified CG, trained with HPC and IRC losses. This final rectified CG provides the high-fidelity semantic target for the GSC loss, completing the guidance loop for the generator.

crepancies between adjacent tissue sections inevitably cause spatial inconsistencies between input H&E images and their corresponding IHC reference (Zhang et al. 2022). This mismatch poses a fundamental problem: enforcing semantic similarity between ostensibly corresponding regions may adversely affect feature representation learning. For example, while the contrastive learning-based method ASP (Li et al. 2023) employs adaptive weighting to reduce the influence of misaligned regions, its core mechanism still relies on pulling features from spatially corresponding patches closer together. This approach still easily causes feature confusion when severe misalignment occurs. Alternative solutions like PyramidP2P (Liu et al. 2022) or DSFF-GAN (Ma et al. 2024) attempt multi-scale alignment strategies, but these often compromise high-frequency details. Consequently, such methods frequently produce blurry outputs or introduce artifacts, limiting their diagnostic utility.

To address these issues, we propose GSGStain, an innovative **Graph-Semantic Guided** framework for virtual **Staining** as shown in Figure 1. Our method fundamentally transforms the virtual staining task by shifting the paradigm from pixel space to graph space, which enables explicit identification and correction of semantically inconsistent features caused by spatial misalignment. Specifically, to fully extract patho-

logical information, we first construct a cell graph from the real H&E image to precisely encode the tissue microenvironment. The nodes of this graph are then initialized with semantic features extracted from the spatially misaligned adjacent IHC image. While these features provide a potent supervisory signal, they are inherently noisy. Furthermore, to address this semantic misalignment, we propose a Graph Semantic Rectification Module (GSRM), which leverages contextual information of the graph to infer and clean up the noisy features on each node. Moreover, a novel graph semantic consistency (GSC) loss constrains the generated IHC image to align with the rectified, ideal graph semantics. Finally, we design a dual-branch discriminator to compel the generator to learn a more realistic image distribution by distinguishing pools of real images from those containing a fake one. The main contributions of this paper are as follows:

- We propose GSGStain, a novel virtual staining paradigm that reframes the problem from pixel space to graph space and introduces the GSRM to address the semantic noise caused by the spatial misalignment of adjacent slices. Experimental results on multiple public datasets demonstrate that our method is superior to existing techniques in terms of both image quality and staining accu-

racy.

- We design a novel GSC loss that utilizes the high-level semantics refined by the GSRM as a reliable supervisory signal, thereby guiding the generator to achieve higher pathological fidelity.
- We introduce a dual-branch discriminator that, by enforcing alignment between the global statistical features of generated and real images, effectively enhances the overall realism and visual quality of the staining results.

Related Works

Virtual Immunohistochemistry staining

Existing virtual IHC methods are mainly categorized into supervised, unsupervised and weakly-unsupervised methods, each with distinct requirements for the alignment of H&E and IHC image pairs. Supervised methods aim to achieve high-fidelity, pixel-to-pixel mapping but rely on perfectly aligned pairs, which are difficult to obtain due to tissue processing constraints, inherent deformations, and biological variability (Huang et al. 2022; Peng et al. 2024a). Some techniques, such as PyramidP2P (Liu et al. 2022), attempt to address this by employing complex preprocessing steps like elastic registration, but these methods can introduce errors and computationally demanding. Others depend on expensive multimodal imaging setups, such as fluorescence imaging (de Haan et al. 2021), further limiting their practicality. In contrast, unsupervised methods, such as CycleGAN and CUT (Zhu et al. 2017; Park et al. 2020; Li et al. 2024), leverage style transfer to generate IHC images without requiring strictly aligned pairs. While these methods offer greater flexibility, they often struggle to ensure biological plausibility, as they lack explicit supervision on biomarker expression patterns. Recent efforts have sought to mitigate this issue by incorporating auxiliary supervision, such as latent feature extraction (Boyd et al. 2022). However, their performance remains constrained by the accuracy of these auxiliary networks and often requires labor-intensive expert input.

To address these limitations, weakly supervised methods have emerged as a promising alternative. By leveraging guidance signals from coarsely aligned adjacent tissue sections, they strike a balance between the rigid requirement of supervised learning and the lack of pathological constraints in unsupervised methods. These methods preserve biological consistency by incorporating regional-level constraints from coarsely aligned H&E-IHC pairs, maintaining key biomarker patterns while relaxing the need for pixel-perfect alignment (Li et al. 2023; Chen et al. 2024a; Wang et al. 2024; Peng et al. 2024b; ?). Compared to traditional strategies, this paradigm avoids computationally intensive registration and reduces reliance on labor-intensive expert annotations. Building upon this framework, our proposed method further advances the approach by constructing a cell graph from adjacent section pairs, enabling precise generation guidance without requiring pixel-perfect alignment.

Graph Learning in Digital Pathology

Graph-based analysis effectively models tissue microenvironments by capturing morphology, spatial topology, and in-

tercellular interactions. The most common approach is the cell graph (CG), where nodes represent cells and edges encode their relationships (Bilgin et al. 2007). While early CGs used hand-crafted node features (Bilgin et al. 2007; Zhou et al. 2019), modern approaches leverage deep learning for richer representations (Pati et al. 2022), with connectivity typically established via methods like k-Nearest Neighbors. These graphs are then analyzed by GNNs to enable applications such as cellular community detection (Javed et al. 2020) and whole-slide image classification (Adnan, Kalra, and Tizhoosh 2020). Recent advances have further improved biological interpretability by incorporating multi-scale and hierarchical structures (Pati et al. 2020), which facilitates the natural integration of domain-specific prior knowledge.

Method

This section presents the proposed GSGStain framework for generating virtual IHC images from H&E images, as shown in Figure 1. This pipeline begins by transforming the histology image into a context-rich cell graph to enable fine-grained pathological modeling. To address the semantic noise introduced by spatial misalignment, we then introduce a Graph Semantic Rectification Module (GSRM), tasked with actively to rectify the noisy initial features in the graph space. Then, we design a graph semantic consistency (GSC) loss to provide the generator with the direct and cell-by-cell supervisory signal. Finally, a dual-branch discriminator is introduced to provide adversarial constraints from the dual perspectives of local texture and batch distribution. The following subsections will elaborate on the implementation details of these modules.

Graph Construction

To capture rich histological context, we represent the H&E image I_H as a cell-centric graph $\mathcal{G} = (V, E, \mathbf{H})$, where V is the set of nodes representing individual cells, E is the set of edges encoding their interactions, and $\mathbf{H} \in \mathbb{R}^{|V| \times d}$ is the node feature matrix of dimension d as shown in Figure 1 (c). The detailed construction is as follows.

Node Definition and Feature Extraction: Nodes ($v \in V$) are defined by the centroids of cell instances segmented from I_H using the Cellpose (Stringer, Michaelos, and Pachitariu 2020). Each node v is assigned a multi-modal feature vector $\mathbf{h}_v \in \mathbf{H}$ composed of three parts: (i) **Contextual Morphological Features** (\mathbf{h}_v^c): To resolve the scale mismatch and context loss from using small cell masks, we extract a fixed-size square patch centered on each node’s centroid. This patch, capturing the cell and its microenvironment, is processed by UNI2-h (Chen et al. 2024b) encoder to yield \mathbf{h}_v^c . (ii) **Spatial Features** (\mathbf{h}_v^s): We incorporate a scale-invariant spatial vector, \mathbf{h}_v^s , by normalizing the node’s centroid coordinates. (iii) **Initial IHC Features** (\mathbf{h}_v^i): A set of quantifiable features (mean OD, std of OD, Positive Ratio) are extracted from the corresponding patch region in the adjacent IHC slide I_I to form the initial IHC characteristic \mathbf{h}_v^i . The handcrafted features are processed by a MLP to unify their dimensionality. The final initial feature for a node v is the concatenation $\mathbf{h}_v = (\mathbf{h}_v^c, \mathbf{h}_v^s, \mathbf{h}_v^i)$.

Graph Topology: The graph topology models intercellular interactions based on spatial proximity and morphological similarity. We first apply a k-Nearest Neighbors (kNN) algorithm to the node centroids to establish the edge set E . Each edge $(u, v) \in E, \{u, v \in V\}$ is then assigned a weight w_{uv} determined by the cosine similarity between the morphological features \mathbf{h}_u^c and \mathbf{h}_v^c . Finally, the topology is pruned by removing edges that exceed a predefined spatial distance threshold.

Graph Semantic Rectification Module

Following the construction of the graph \mathcal{G} with initial noisy IHC features \mathbf{h}_v^i , we introduce the Graph Semantic Rectification Module (GSRM), which leverages high-confidence morphological context \mathbf{h}_v^c to rectify the low-confidence semantics introduced by spatial misalignment as shown in Figure 1 (d). To this end, we employ a Principal Neighbourhood Aggregation (PNA) network (Corso et al. 2020) as the backbone architecture for the GSRM. PNA can capture the full statistical distribution of neighboring features, enabling the GSRM to distinguish between different microenvironmental patterns. Through L layers of message passing, the GSRM takes the initial graph as input and outputs a semantically rectified graph, \mathcal{G}_{rec} , where each node v carries a rectified IHC semantic feature vector $\tilde{\mathbf{h}}_v^i$. The feature update for a node v at layer l to $l + 1$ follows the PNA message passing scheme.

To effectively train the GSRM, we firstly propose the hierarchical pathological consistency (HPC) loss. Specially, we define $S_{pred}^{(0)}$ as the $K \times K$ grid of statistical features aggregated from the rectified node semantics $\tilde{\mathbf{h}}_v^i$, and $S_{gt}^{(0)}$ as the corresponding grid from the real IHC image. We then construct two pyramids by iteratively applying average pooling with a stride of 2. The \mathcal{L}_{HPC} is the weighted sum of L2 losses between the pyramid levels:

$$\mathcal{L}_{HPC} = \sum_{s=0}^{S-1} w_s \cdot \|S_{pred}^{(s)} - S_{gt}^{(s)}\|^2 \quad (1)$$

where w_s is the weight item and $\|\cdot\|$ denotes the L2 norm. We further introduce an inter-sample relational constraint L_{IRC} to ensure that the feature space learned by the GSRM maintains the intrinsic structure of the real data distribution. Specifically, by calculating the pairwise cosine similarities among all real fingerprints within the batch, we construct a ground-truth relationship matrix M_{gt} . Similarly, we construct a predicted relationship matrix M_{pred} . The elements of these matrices are defined as:

$$M_{gt}[i, j] = sim(S_{gt}^{(0)}(i), S_{gt}^{(0)}(j)) \quad (2)$$

$$M_{pred}[i, j] = sim(S_{pred}^{(0)}(i), S_{pred}^{(0)}(j)) \quad (3)$$

where $sim(\cdot)$ denote the cosine similarity. The L_{IRC} loss aims to minimize the difference between these two relationship matrices:

$$\mathcal{L}_{IRC} = \frac{1}{B} \sum_{i=1}^B \sum_{j=1}^B |M_{gt}^{ij} - M_{pred}^{ij}| \quad (4)$$

where B denotes the batchsize.

The training objective of the GSRM is to minimize the weighted sum of these two complementary losses:

$$L_{GSRM} = \lambda_{HPC} L_{HPC} + \lambda_{IRC} L_{IRC} \quad (5)$$

where λ_{HPC} and λ_{IRC} are hyperparameters of the loss.

Dual-Branch Discriminator

To ensure high-fidelity image generation, we designed a novel dual-branch discriminator D that supervises the generator G on both local texture and batch distribution. This architecture consists of a shared feature extraction backbone and two independent heads. The local branch functions as a standard PatchGAN discriminator (Isola et al. 2017), providing standard adversarial losses \mathcal{L}_D^{adv} and \mathcal{L}_G^{adv} . The contextual branch is designed for batch-level evaluation; its prediction head outputs a continuous confidence score for each image, which we denote as the function $s(\cdot)$.

This contextual branch is trained with a ranking loss. For a training batch of N real reference images $\mathcal{B}_{real} = \{I_1^r, \dots, I_N^r\}$ and N generated images $\mathcal{B}_{fake} = \{\hat{I}_1^f, \dots, \hat{I}_N^f\}$, the discriminator's loss \mathcal{L}_D^R uses a hinge formulation to maximize the margin between the average real and fake scores:

$$\mathcal{L}_D^R = \mathbb{E}_{I_H, I_I} [\max(0, 1 - (\bar{s}_{real} - \bar{s}_{fake}))] \quad (6)$$

where $\bar{s}_{real} = \frac{1}{B} \sum_{i=1}^B s(I_i^r)$ and $\bar{s}_{fake} = \frac{1}{B} \sum_{i=1}^B s(\hat{I}_i^f)$. D is jointly optimized by \mathcal{L}_D^{adv} and \mathcal{L}_D^R . Correspondingly, the generator's loss \mathcal{L}_G^R aims to increase the confidence scores of its generated images:

$$\mathcal{L}_G^R = -\mathbb{E}_{I_H} [\bar{s}_{fake}] \quad (7)$$

This objective compels the generator to learn the batch distributional properties of the real images in order to deceive the contextual branch.

Framework

In our GSGStain framework, the generator translates an H&E image I_H to a virtual IHC image I_G . To ensure the generated image is not only visually realistic but also pathologically correct, its rendering must adhere to the high-confidence semantic guidance from the GSRM. We therefore designed the graph semantic consistency loss L_{GSC} , which applies direct, cell-level constraints on the generator's output. The computation involves comparing the ideal and actual semantics for each cell node v . We take the ideal IHC semantic vector $\tilde{\mathbf{h}}_v^i$ provided by the GSRM. We then extract the semantic vector \mathbf{h}_g by applying an identical feature extraction pipeline to the corresponding location in the generated IHC image \hat{I}_I . The loss is defined as the average L2 distance between these two vectors over all nodes:

$$L_{GSC} = \mathbb{E}_{v \in V} [\|\tilde{\mathbf{h}}_v^i - \mathbf{h}_g\|^2] \quad (8)$$

The entire GSGStain framework is optimized end-to-end via a unified objective function that integrates all the aforementioned constraints:

$$\mathcal{L}_G^{total} = \mathcal{L}_G^{adv} + \mathcal{L}_G^{NCE} + \lambda_{GSC} \mathcal{L}_{GSC} + \lambda_R \mathcal{L}_G^R \quad (9)$$

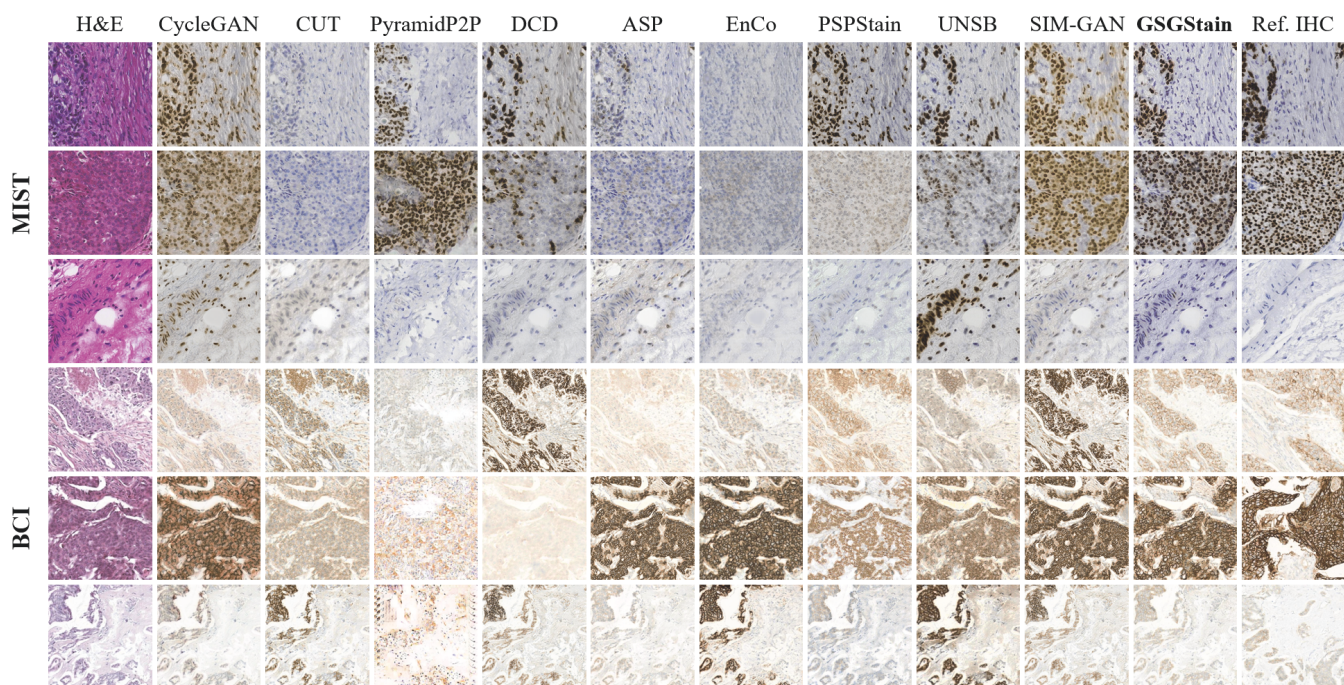


Figure 2: Qualitative comparison with state-of-the-art methods on the BCI and MIST datasets. From left to right: Input H&E, results from competing methods, our GSGStain (bolded), and the reference IHC. GSGStain produces results that are visually most consistent with the ground truth.

where adversarial and PatchNCE loss \mathcal{L}_G^{adv} , \mathcal{L}_G^{NCE} are defined in Park et al. (2020) and λ_{GSC} , λ_R are the hyperparameters. The generator then learns to render virtual IHC images that are both realistic and accurate under the fine-grained semantic guidance purified by the GSRM.

Experiments

Experimental Setup

Datasets. Our experiments were performed on two public benchmark datasets, BCI (Liu et al. 2022) and MIST (Li et al. 2023), both of which feature paired 1024×1024 pixel patches from adjacent tissue sections. Specifically, the BCI dataset provides 3,896 H&E-IHC pairs for training and 977 for testing. For the MIST dataset, we used the subsets containing the ER biomarkers, which consists of 4,153 training and 1,000 testing pairs.

Evaluation metrics. In the experiments, we quantitatively evaluate the proposed method using multiple key metrics, including: (i) Fréchet Inception Distance (FID) and (ii) Kernel Inception Distance (KID) to evaluate the overall realism and distributional quality (Mentzer et al. 2020), where lower scores signify a closer match to the real images; (iii) Structural Similarity Index Measure (SSIM) to evaluate structural preservation against the paired ground-truth image (Hore and Ziou 2010), where a higher score indicates better fidelity; and (iv) DAB_{KL} to calculate the Kullback-Leibler divergence of the DAB channel histograms for measuring the accuracy of protein expression, where a lower score indicates a closer match (Wang et al. 2025b).

Implementation Details. Our framework was implemented in PyTorch and trained for 100 epochs on an NVIDIA A100 Tensor Core GPU using the Adam optimizer ($\beta_1=0.5, \beta_2=0.999$). We used a batch size of 4 with an initial learning rate of 2e-4, which was linearly decayed to zero over the final 50 epochs. Input data was prepared by taking 512×512 random crops from whole-slide images, followed by standard data augmentation (random flips). The GSRM consists of a 4-layer PNA network operating on cell graphs constructed via a k=8 nearest neighbors algorithm. GSRM terms λ_{HPC} , λ_{IRC} were set to 1.0. When training the virtual staining network, λ_{GSC} , λ_R in Eq. (9) are set to 1.0, 0.1.

Comparative Results on BCI and MIST

To comprehensively evaluate the performance of our proposed GSGStain framework, we conducted a quantitative comparison against several advanced models on two public datasets, MIST and BCI. These comparison methods include both specialized virtual staining models, such as PyramidP2P (Liu et al. 2022), ASP (Li et al. 2023), PSPStain (Chen et al. 2024a), and SIM-GAN (Guan et al. 2025b), as well as state-of-the-art general-purpose models for natural image-to-image translation, such as CycleGAN (Zhu et al. 2017), CUT (Park et al. 2020), EnCo (Cai et al. 2024), DCD (Hu et al. 2023), and UNSB (Kim et al. 2024). As shown in Table 1, GSGStain outperforms all baseline methods, achieving the best results across most key evaluation metrics.

Our method exhibits significant advantages in image re-

Model	MIST				BCI			
	FID ↓	KID ↓	SSIM ↑	D _{KL} ↓	FID ↓	KID ↓	SSIM ↑	D _{KL} ↓
CycleGAN (Zhu et al. 2017)	49.179	1.240	0.243	2.745	45.813	1.159	0.391	2.324
CUT (Park et al. 2020)	57.608	2.101	0.243	3.935	50.701	1.951	0.455	2.969
PyramidP2P (Liu et al. 2022)	106.260	7.590	<u>0.259</u>	<u>2.207</u>	117.115	7.230	0.436	5.987
DCD (Hu et al. 2023)	58.035	1.809	0.257	2.245	54.985	1.629	0.439	2.031
ASP (Li et al. 2023)	<u>42.452</u>	<u>1.044</u>	0.243	2.213	39.542	<u>0.914</u>	0.462	2.015
EnCo (Cai et al. 2024)	78.811	3.301	0.253	3.437	51.123	2.101	0.431	4.197
PSPStain (Chen et al. 2024a)	52.531	2.139	0.250	2.362	<u>38.911</u>	0.939	0.418	<u>1.942</u>
UNSB (Kim et al. 2024)	50.922	1.461	0.243	2.894	47.122	1.301	0.424	<u>3.004</u>
SIM-GAN (Guan et al. 2025b)	47.918	1.851	0.261	2.365	44.918	1.681	0.382	2.165
GSGStain (Ours)	38.292	0.801	0.248	1.966	35.192	0.721	<u>0.457</u>	1.766

Table 1: Quantitative comparison on the MIST and BCI datasets. The best result for each metric on each dataset is in **bold**, and the second-best is underlined. Lower values are better for FID, KID, and D_{KL} (↓), while higher is better for SSIM (↑).

Settings		Metrics			
GSRM	DBD	FID ↓	KID ↓	SSIM ↑	D _{KL} ↓
✗	✗	57.608	2.101	0.243	3.935
✓	✗	45.372	1.419	0.237	2.128
✗	✓	42.326	1.109	0.251	3.438
✓	✓	38.292	0.801	0.248	1.966

Table 2: Ablation study of the GSRM and DBD components. Our full model includes both components. The best result in each column is highlighted in **bold**.

alism, distributional diversity, and pathological accuracy. Specifically, GSGStain achieves the lowest (best) FID and KID scores on both datasets. For instance, on the MIST dataset, our FID score (38.292) is substantially lower than that of the second-best method ASP (42.452). These results provide strong evidence that our generated images most closely resemble real IHC images in both visual quality and data distribution. Additionally, GSGStain achieves the best performance on the D_{KL} metric, which evaluates protein expression precision, a critical factor for clinical diagnosis accuracy. It is worth noting that in this weakly-paired evaluation context, the SSIM metric should be interpreted with caution, as a model that generates an image with fewer details might paradoxically achieve a higher SSIM score. Considering many baseline models use SSIM for evaluation, this metric is provided for reference only. In contrast, GSGStain’s superior performance on FID, KID, and D_{KL} confirms its ability to generate images with realistic pathological patterns and cell-level details. Together, these quantitative results demonstrate that GSGStain sets a new benchmark in virtual immunohistochemistry, offering a more reliable and precise solution with exceptional image realism and pathological fidelity.

Figure 2 presents the visual comparison results of all competing methods on both datasets. GSGStain demonstrates significant advantages in stain transfer tasks by accurately preserving histopathological morphology and staining fea-

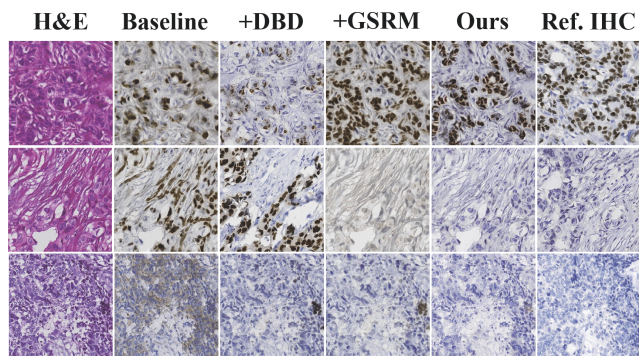


Figure 3: Visualization of our ablation study.

ture consistency. Experimental results show that compared to methods like CUT, EnCo and PSPStain which tend to produce overly homogenized staining effects, GSGStain better maintains the chemical properties and optical density characteristics of IHC staining. Both quantitative and qualitative analyses confirm that the images generated by GSGStain most closely approximate the reference standards in terms of staining accuracy, natural appearance, and pathological fidelity, exhibiting superior image realism and diagnostic value.

Ablation Study

Our ablation studies on the MIST dataset provide compelling evidence for the complementary roles of our core components as shown in Table 2 and Figure 3. The results show that the DBD module by itself enhances image distribution quality and realism, though it can generate false staining patterns due to a lack of precise semantic guidance. Conversely, incorporating the GSRM alone leads to substantial improvements in pathological fidelity by accurately identifying correct staining regions, but may lack the textural fidelity of the final model. Crucially, the complete model combining both GSRM and DBD modules achieves optimal performance across all core metrics (FID, KID, and

H&E	IHC	FID ↓	KID ↓	SSIM ↑	D _{KL} ↓
-	Handcrafted	68.315	2.854	0.218	3.341
ResNet50	Handcrafted	42.518	1.103	0.242	2.259
UNI2-h	Handcrafted	38.292	0.801	0.248	1.966
UNI2-h	UNI2-h	55.116	1.957	0.239	2.413
UNI2-h	-	59.842	2.316	0.225	4.588

Table 3: Ablation on graph node features. Best results are in **bold**.

D_{KL}). It successfully synthesizes the strengths of both components—rendering IHC images with both accurate pathological patterns guided by the GSRM and high textural realism enforced by the DBD, achieving the closest visual match to the reference IHC. These findings confirm that our framework successfully balances the critical objectives of generating realistic pathological details while maintaining the structural consistency necessary for diagnostic purposes.

Impact of Graph Feature Representation

Our systematic ablation study validates the design choices behind our final model architecture, as shown in Table 3. The combined use of a pathology foundation model UNI2-h for H&E features extraction with handcrafted IHC priors produces the strong performance demonstrated in our main experiments. Several key findings emerge from this investigation. First, employing the domain-specific UNI2-h encoder yields substantially better results than using a generic ResNet50 encoder. This demonstrates that higher-quality, domain-specific morphological features serve as a superior foundation for the GSRM, enabling it to accurately infer underlying pathological semantics. Interestingly, when examining IHC prior sources, we observed that replacing our simple handcrafted priors with complex deep features from UNI2-h actually degraded performance. This suggests that the simpler, low-dimensional handcrafted features offer a more direct and robust signal for GSRM correction, while abstract deep features may introduce structural noise from the misaligned slice. Finally, the dramatic performance drop in baselines lacking either deep H&E features or any IHC prior confirms that both morphological context and the weak supervision signal are essential for proper GSRM operation.

GSRM Performance Visualization

Figure 4 intuitively validates our GSRM on a representative tissue sample. The noisy semantics demonstrates the initial problem: due to spatial misalignment between the H&E and reference IHC slides, many cancerous cells are incorrectly assigned low-confidence staining semantics. The rectified semantics shows a dramatic improvement after GSRM processing. By leveraging morphological context from neighboring nodes, the GSRM corrects these erroneous assignments to high-confidence, accurately reflecting the ground-truth pathological status.

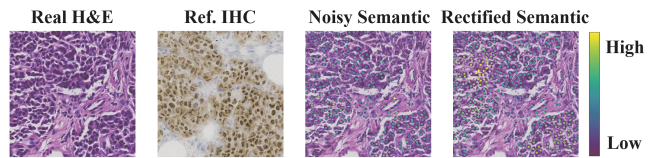


Figure 4: Visualization of GSRM semantic rectification.

Generator	FID ↓	KID ↓	D _{KL} ↓	Params(M)
Resnet-4blocks	42.172	0.953	2.158	5.48
Resnet-6blocks	38.292	0.801	1.966	7.84
Resnet-9blocks	39.481	0.868	2.019	11.38
UNet-128	40.049	0.915	2.087	41.83
UNet-256	39.266	0.882	1.993	54.41

Table 4: Discussion on the generator architecture. The best result for each metric is shown in **bold**.

Discussion on Generator Architecture

To determine the optimal generator architecture, we conducted a detailed ablation study considering both generative quality and parameter efficiency, with results shown in Table 4. The results demonstrate that our selected Resnet-6blocks backbone achieves optimal performance while maintaining high efficiency, as evidenced in our main experiments. From this analysis, we identify several key findings: First, architectural reduction to 4 blocks caused significant performance degradation, revealing insufficient model capacity for this task. Conversely, expanding to 9 blocks not only failed to improve results but actually degraded performance despite increased parameters, showing the expanded architecture offered no improvements. Notably, when compared to Unet variants, our Resnet-6blocks architecture achieves superior performance across all metrics while maintaining remarkable efficiency, using fewer than 20% of the parameters required by Unet-128. These findings robustly validate our architecture choice, as Resnet-6blocks achieves the ideal balance between model capability, generation quality, and computational efficiency.

Conclusion

We introduce GSGStain, a novel graph-semantic framework for virtual IHC staining. By re-framing the problem from pixel to graph space, we address semantic noise and spatial misalignment in adjacent tissue sections. The core GSRM leverages tissue microenvironment context to correct noisy biomarker features from misaligned IHC images. Optimized by GSC loss, this process ensures generated images align with rectified, high-fidelity semantics. A dual-branch discriminator further enhances realism by enforcing statistical alignment with real data. Extensive experiments demonstrate that GSGStain significantly outperforms state-of-the-art methods in image quality and pathological consistency. This work establishes a new paradigm for semantically robust virtual staining, paving the way for reliable computational pathology tools.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 82160345).

References

- Adnan, M.; Kalra, S.; and Tizhoosh, H. R. 2020. Representation learning of histopathology images using graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 988–989.
- Bilgin, C.; Demir, C.; Nagi, C.; and Yener, B. 2007. Cell-graph mining for breast tissue modeling and classification. In *2007 29th Annual international conference of the IEEE Engineering in Medicine and Biology Society*, 5311–5314. IEEE.
- Boyd, J.; Villa, I.; Mathieu, M.-C.; Deutsch, E.; Paragios, N.; Vakalopoulou, M.; and Christodoulidis, S. 2022. Region-guided cyclegans for stain transfer in whole slide images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 356–365. Springer.
- Cai, X.; Zhu, Y.; Miao, D.; Fu, L.; and Yao, Y. 2024. Rethinking the paradigm of content constraints in unpaired image-to-image translation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 891–899.
- Chen, F.; Zhang, R.; Zheng, B.; Sun, Y.; He, J.; and Qin, W. 2024a. Pathological semantics-preserving learning for H&E-to-IHC virtual staining. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 384–394. Springer.
- Chen, R. J.; Ding, T.; Lu, M. Y.; Williamson, D. F.; Jaume, G.; Chen, B.; Zhang, A.; Shao, D.; Song, A. H.; Shaban, M.; et al. 2024b. Towards a General-Purpose Foundation Model for Computational Pathology. *Nature Medicine*.
- Corso, G.; Cavalleri, L.; Beaini, D.; Liò, P.; and Veličković, P. 2020. Principal Neighbourhood Aggregation for Graph Nets. In *Advances in Neural Information Processing Systems*.
- de Haan, K.; Zhang, Y.; Zuckerman, J. E.; Liu, T.; Sisk, A. E.; Diaz, M. F.; Jen, K.-Y.; Nobori, A.; Liou, S.; Zhang, S.; et al. 2021. Deep learning-based transformation of H&E stained tissues into special stains. *Nature communications*, 12(1): 4884.
- Guan, X.; Zhang, Z.; Wang, Y.; Li, Y.; and Zhang, Y. 2025a. Supervised Information Mining from Weakly Paired Images for Breast IHC Virtual Staining. *IEEE Transactions on Medical Imaging*, 44(5): 2120–2130.
- Guan, X.; Zhang, Z.; Wang, Y.; Li, Y.; and Zhang, Y. 2025b. Supervised Information Mining from Weakly Paired Images for Breast IHC Virtual Staining. *IEEE Transactions on Medical Imaging*.
- Hore, A.; and Ziou, D. 2010. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition*, 2366–2369. IEEE.
- Hu, T.; Lin, M.; You, L.; Chao, F.; and Ji, R. 2023. Discriminator-cooperated feature map distillation for gan compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20351–20360.
- Huang, K.; Cheng, Y.; Gao, Q.; and Zhang, B. 2022. Weakly Unpaired Image Translation from Hematoxylin and Eosin Staining Image to Immunohistochemistry Staining Image. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1013–1019. IEEE.
- Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-Image Translation with Conditional Adversarial Networks. *CVPR*.
- Javed, S.; Mahmood, A.; Fraz, M. M.; Koohbanani, N. A.; Benes, K.; Tsang, Y.-W.; Hewitt, K.; Epstein, D.; Snead, D.; and Rajpoot, N. 2020. Cellular community detection for tissue phenotyping in colorectal cancer histology images. *Medical image analysis*, 63: 101696.
- Kim, B.; Kwon, G.; Kim, K.; and Ye, J. C. 2024. Unpaired Image-to-Image Translation via Neural Schrödinger Bridge. In *ICLR*.
- Li, F.; Hu, Z.; Chen, W.; and Kak, A. 2023. Adaptive supervised patchnce loss for learning h&e-to-ihc stain translation with inconsistent groundtruth image pairs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 632–641. Springer.
- Li, J.; Dong, J.; Huang, S.; Li, X.; Jiang, J.; Fan, X.; and Zhang, Y. 2024. Virtual Immunohistochemistry Staining for Histological Images Assisted by Weakly-supervised Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11259–11268.
- Liu, S.; Zhu, C.; Xu, F.; Jia, X.; Shi, Z.; and Jin, M. 2022. Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1815–1824.
- Ma, Y.; Zhang, Y.; Wang, Z.; Li, J.; Miao, Y.; Yang, F.; and Pan, W. 2024. DSFF-GAN: A novel stain transfer network for generating immunohistochemical image of endometrial cancer. *Computers in Biology and Medicine*, 170: 108046.
- Mentzer, F.; Toderici, G. D.; Tschannen, M.; and Agustsson, E. 2020. High-fidelity generative image compression. *Advances in neural information processing systems*, 33: 11913–11924.
- Park, T.; Efros, A. A.; Zhang, R.; and Zhu, J.-Y. 2020. Contrastive learning for unpaired image-to-image translation. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, 319–345. Springer.
- Pati, P.; Jaume, G.; Fernandes, L. A.; Foncubierta-Rodríguez, A.; Feroce, F.; Anniciello, A. M.; Scognamiglio, G.; Brancati, N.; Riccio, D.; Di Bonito, M.; et al. 2020. Hact-net: A hierarchical cell-to-tissue graph neural network for histopathological image classification. In *International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging*, 208–219. Springer.
- Pati, P.; Jaume, G.; Foncubierta-Rodríguez, A.; Feroce, F.; Anniciello, A. M.; Scognamiglio, G.; Brancati, N.; Fiche, M.; Dubruc, E.; Riccio, D.; et al. 2022. Hierarchical graph representations in digital pathology. *Medical image analysis*, 75: 102264.

- Pati, P.; Karkampouna, S.; Bonollo, F.; Comp erat, E.; Radi c, M.; Spahn, M.; Martinelli, A.; Wartenberg, M.; Kruithof-de Julio, M.; and Rapsomaniki, M. 2024. Accelerating histopathology workflows with generative AI-based virtually multiplexed tumour profiling. *Nature machine intelligence*, 6(9): 1077–1093.
- Peng, Q.; Lin, W.; Hu, Y.; Bao, A.; Lian, C.; Wei, W.; Yue, M.; Liu, J.; Yu, L.; and Wang, L. 2024a. Advancing H&E-to-IHC Virtual Staining with Task-Specific Domain Knowledge for HER2 Scoring. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 3–13. Springer.
- Peng, Q.; Lin, W.; Hu, Y.; Bao, A.; Lian, C.; Wei, W.; Yue, M.; Liu, J.; Yu, L.; and Wang, L. 2024b. Advancing H&E-to-IHC Virtual Staining with Task-Specific Domain Knowledge for HER2 Scoring. In Linguraru, M. G.; Dou, Q.; Feragen, A.; Giannarou, S.; Glocker, B.; Lekadir, K.; and Schnabel, J. A., eds., *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, 3–13. Cham: Springer Nature Switzerland. ISBN 978-3-031-72083-3.
- Peng, Y.; Xiong, B.; Chen, F.; Eybo, D.; Zhang, R.; Hu, W.; Cai, J.; and Qin, W. 2025. USIGAN: Unbalanced Self-Information Feature Transport for Weakly Paired Image IHC Virtual Staining. *arXiv preprint arXiv:2507.05843*.
- Stringer, C.; Michaelos, M.; and Pachitariu, M. 2020. Cellpose: a generalist algorithm for cellular segmentation. *bioRxiv*.
- Wang, S.; Zhang, Z.; Yan, H.; Xu, M.; and Wang, G. 2024. Mix-domain contrastive learning for unpaired h&e-to-ihc stain translation. In *2024 IEEE International Conference on Image Processing (ICIP)*, 2982–2988. IEEE.
- Wang, T.; Wang, M.; Wang, Z.; Wang, H.; Xu, Q.; Cong, F.; and Xu, H. 2025a. ODA-GAN: Orthogonal Decoupling Alignment GAN Assisted by Weakly-supervised Learning for Virtual Immunohistochemistry Staining. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, 25920–25929.
- Wang, T.; Wang, M.; Wang, Z.; Wang, H.; Xu, Q.; Cong, F.; and Xu, H. 2025b. ODA-GAN: Orthogonal Decoupling Alignment GAN Assisted by Weakly-supervised Learning for Virtual Immunohistochemistry Staining. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 25920–25929.
- Zaha, D. C. 2014. Significance of immunohistochemistry in breast cancer. *World journal of clinical oncology*, 5(3): 382.
- Zeng, B.; Lin, Y.; Wang, Y.; Chen, Y.; Dong, J.; Li, X.; and Zhang, Y. 2022. Semi-supervised pr virtual staining for breast histopathological images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 232–241. Springer.
- Zhang, R.; Cao, Y.; Li, Y.; Liu, Z.; Wang, J.; He, J.; Zhang, C.; Sui, X.; Zhang, P.; Cui, L.; et al. 2022. MVFStain: multiple virtual functional stain histopathology images generation based on specific domain mapping. *Medical Image Analysis*, 80: 102520.
- Zhou, Y.; Graham, S.; Alemi Koohbanani, N.; Shaban, M.; Heng, P.-A.; and Rajpoot, N. 2019. Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 0–0.
- Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.