

ODA-GAN: Orthogonal Decoupling Alignment GAN Assisted by Weakly-supervised Learning for Virtual Immunohistochemistry Staining

Tong Wang¹ Mingkang Wang¹ Zhongze Wang²
 Hongkai Wang¹ Qi Xu^{1*} Fengyu Cong¹ Hongming Xu^{1*}

¹Dalian University of Technology, Dalian, China

²East China University of Science and Technology, Shanghai, China

{wt, wangmingkang}@mail.dlut.edu.cn, zzwang@mail.ecust.edu.cn

{wang.hongkai, xuqi, cong, mxu}@dlut.edu.cn

Abstract

Recently, virtual staining has emerged as a promising alternative to revolutionize histological staining by digitally generating stains. However, most existing methods suffer from the curse of staining unreality and unreliability. In this paper, we propose the Orthogonal Decoupling Alignment Generative Adversarial Network (ODA-GAN) for unpaired virtual immunohistochemistry (IHC) staining. Our approach is based on the assumption that an image consists of IHC staining-related features, which influence staining distribution and intensity, and staining-unrelated features, such as tissue morphology. Leveraging a pathology foundation model, we first develop a weakly-supervised segmentation pipeline as an alternative to expert annotations. We introduce an Orthogonal MLP (O-MLP) module to project image features into an orthogonal space, decoupling them into staining-related and unrelated components. Additionally, we propose a Dual-stream PatchNCE (DPNCE) loss to resolve contrastive learning contradictions in the staining-related space, thereby enhancing staining accuracy. To further improve realism, we introduce a Multi-layer Domain Alignment (MDA) module to bridge the domain gap between generated and real IHC images. Evaluations on three benchmark datasets show that our ODA-GAN reaches state-of-the-art (SOTA) performance. Our source code is available at <https://github.com/ittong/ODA-GAN>.

1. Introduction

Histological staining is essential in clinical pathology, enabling visualization of tissue and cellular structures for detailed microscopic assessment [2]. Among various staining techniques, hematoxylin and eosin (H&E) is the undisputed

workhorse, routinely employed to delineate tissue structures and cell morphology [7]. Although H&E staining is widely available for all cancer patients, it may not always provide sufficient diagnostic information. Complementarily, IHC staining [10] is extensively used to highlight the distribution and localization of specific markers, offering critical insights for tumor subtyping, prognosis, and personalized treatment. In clinical practice, generating these H&E-IHC paired images often involves repeated serial sectioning and staining with different antibodies, a process that is both time-consuming and tissue-exhaustive. Moreover, serial IHC staining can occasionally produce unaligned images and suboptimal quality due to artifacts, while limited tissue availability may result in missing stains.

Fortunately, virtual staining techniques [8, 11, 20, 22, 31], which can convert H&E images into IHC images using generative networks, have opened new opportunities to revolutionize these decade-old staining methods. Initial virtual IHC staining models were primarily relied on precisely aligned H&E and IHC images, allowing for pixel-wise loss optimization between the virtual and real images [17]. However, the inherent discrepancies in these image pairs result in detrimental pixel-wise optimization and severe square effect [34]. Besides, these methods fail to address situations when adjacent layers are not available.

In recent years, research on training virtual staining models under unpaired settings has gained momentum. Most unpaired virtual IHC staining models draw inspiration from CycleGAN [43], which depends on cycle consistency loss to preserve content consistency across two domains. While effective, the cycle consistency assumes a bijective mapping between two domains [43], which is overly restrictive and often unsuitable for many virtual IHC staining tasks. A notable problem is the incorrect cross-domain mapping, where positive signals in the H&E domain are misaligned with negative signals in the IHC domain.

*Corresponding authors: Qi Xu, Hongming Xu

To bypass the bijection limitation, CUT-like methods [20, 26, 41] have emerged as appealing alternatives to CycleGAN, achieving content preservation by maximizing the mutual information between target and source domains [30]. However, despite their practical successes, we find that these methods incur a contradiction between maximizing mutual information and ensuring pathological consistency, since subpatches sharing the same staining labels are treated as negative samples and thus pushed away (see Fig. 3 (a)). One feasible solution to this problem is to decouple image features into staining-related components influencing the distribution and intensity of IHC staining, and staining-unrelated components determining the tissue morphology. By processing these components separately, we can maximize mutual information between the query subpatch and the key subpatch at the same spatial location in the staining-unrelated space, while also maximizing mutual information with the key subpatches that share the same staining labels in the staining-related space. This solution poses two challenges. The first challenge is **how to effectively decouple image features into staining-related and unrelated components**. The inherent blending of these features creates a conflict between maximizing mutual information and ensuring pathological consistency. The second challenge is **how to segment H&E images into regions exhibiting different IHC staining signals**. Acquiring extensive expert annotations is challenging; however, obtaining staining labels for subpatches is essential in guiding the mutual information maximization process.

In this paper, we propose the ODA-GAN for unpaired virtual IHC staining that addresses the above challenges. First, we introduce an orthogonal MLP (O-MLP) module, which employs an MLP with orthogonal constraints to project image features into an orthogonal space, combined with a learnable mask to decouple them into staining-related and unrelated features. Second, we utilize multi-instance learning (MIL) [27] to effectively acquire pseudo-labels of H&E images, and apply Grad-CAM [33] to visualize pathological patterns for weakly-supervised semantic segmentation (WSSS), such that the aforementioned objective can be realized by our proposed Dual-stream PatchNCE (DPNCE). Lastly, to generate more realistic staining results, we propose a Multi-layer Domain Alignment (MDA) module to bridge virtual and real IHC domains. In summary, our main contributions are as follows:

- We propose a weakly-supervised segmentation pipeline based on a pathology foundation model to transfer slide-level labels to patch-level semantic segmentation masks, assisting in virtual IHC staining.
- We propose a DPNCE loss to resolve contrastive learning contradictions by decoupling image features into staining-related and unrelated components, processing them separately to ensure pathological consistency.

- We propose an MDA module to minimize the distribution discrepancy between virtual and real image features, aligning generated and real IHC images to enable more realistic staining.

2. Related work

Virtual IHC Staining. Virtual IHC staining methods can be broadly categorized into supervised and unsupervised approaches based on different GAN variations [13]. Early methods primarily follow a supervised paradigm, relying on precisely aligned image pairs for training. For example, Liu et al. [23] developed the PyramidP2P network using a registered paired dataset with a weakened pixel-wise constraint [17]. To address inherent inconsistencies between image pairs, Li et al. [20] introduced a novel weighted objective function to mitigate the negative effects of noisy supervision. However, curating such paired datasets is challenging due to the error-prone nature of the registration process, a difficulty that becomes more pronounced when substantial discrepancies exist between adjacent tissue layers.

In the absence of paired data, unsupervised methods routinely adopt cycle consistency losses to preserve H&E content while translating IHC styles, yet they fall short in ensuring pathological consistency between the generated IHC and source H&E images. Some approaches incorporate additional constraints, such as perceptual embedding consistency [19] or patch-level pathological consistency [21, 22, 39]. However, the bijective mapping assumption on which they rely limits their staining performance. To bypass these limitations, we adopt the more appropriate CUT method [30] as a baseline, with significant extensions to enhance its applicability for virtual IHC staining.

Weakly-supervised Learning. MIL is a widely studied weakly-supervised learning approach in whole slide image (WSI) analysis [6]. In the MIL paradigm for WSI analysis [5, 16, 24], the gigapixel WSI is treated as a bag containing numerous cropped patches, or instances, typically generated by a sliding window. Most approaches adhere to the standard workflow: cropped patches are embedded into feature vectors, and these vectors are then aggregated to make a bag-level prediction, utilizing only bag-level supervision. Although only bag labels serve as ground truths for training, MIL models can often obtain accurate patch-level prediction within regions of high attention scores. Based on this capability, Li et al. [21] proposed a patch-level pathology information extractor (PPIE) for H&E image classification to guide Cycle-GAN to improve IHC staining.

WSSS aims to generate pixel-level predictions from weak, cost-effective annotations, such as image-level labels [1]. Most prevalent WSSS methods using image-level labels typically rely on Class Activation Map (CAM) [42] or its variants [33] to achieve semantic segmentation. With

recent advances in pathology foundation models [9, 25, 35] and their ability to generalize across complex pathology patterns, accurate segmentation of regions of interest in H&E images via WSSS is now an achievable goal.

3. Methods

As shown in Fig. 1 (a), our ODA-GAN is trained using two unpaired sets of image patches extracted from WSI: source H&E images, denoted as $X = \{x_i\}_{i=0}^{N_X}$, and target IHC images, denoted as $Y = \{y_i\}_{i=0}^{N_Y}$. The backbone of the ODA-GAN is a cascaded generator G with two sequential components: an encoder G_{enc} and a decoder G_{dec} . The goal is to achieve virtual staining using the generator, which takes an H&E image x_i as input and converts it into a virtual IHC image $\hat{y}_i$, i.e., $\hat{y}_i = G(x_i) = G_{dec}(G_{enc}(x_i))$.

3.1. Weakly-supervised Segmentation.

In this section, we propose a weakly-supervised segmentation pipeline that effectively segments likely antigen-positive regions in H&E images, assisting contrastive learning within our ODA-GAN framework. This pipeline consists of two main tasks: (1) training an MLP for classification of feature vectors embedded by the UNI [9] into positive or negative categories, and (2) cascading the MLP with the UNI to generate Grad-CAM [33] masks.

Labeling. As shown in Fig. 2, given a set of H&E WSI bags and their corresponding patches, we begin by embedding all patches into feature vectors using the SOTA vision encoder, UNI [9]. We then train a representative MIL method, ABMIL [16], to identify high-likelihood positive patches with high attention scores from each positive WSI, i.e., top- k selection. When k is relatively small, it is ensured that all selected instances are positive; however, with larger k values, the selected collection may include an increasing number of negative instances. Correspondingly, we also randomly select a comparable number of patches from negative WSI bags, ensuring that each selected instance is indeed negative. We use this curated set of high-likelihood positive and confirmed negative instances to train an MLP, which learns to classify these patches as positive or negative. After that, we label all remaining WSI patches using our initial MLP classifier, then retrain the MLP on this expanded set with the generated pseudo-labels to further refine its ability to capture staining-related patterns. The MLP classifier consists of two hidden layers with dimensions of 512 and 256, with dropout applied for regularization.

Masking. After cascading the previously trained MLP with the UNI, the complete network processes all patches to generate pseudo-labels and corresponding confidence

scores c . Simultaneously, we apply Grad-CAM [33] to visualize the regions within each patch that contribute to predicting positive staining signals. To generate these visualizations, we compute the gradients of the positive class with respect to the ViT Norm layer within UNI, capturing feature maps and gradients via hooks during forward and backward passes. These gradients are then summed over the feature maps to create a heatmap for each patch, highlighting regions relevant to the positive prediction. Finally, the heatmap is binarized by a simple thresholding to generate masks for antigen-positive regions. Negative patches, as predicted by the classifier, are not assigned a mask. Note that only high-confidence patches (e.g., $c > 0.9$) are selected to assist in our ODA-GAN training. For IHC images, we directly obtain masks by thresholding their DAB channels, and assign pseudo-labels based on the presence or absence of positive staining regions in the mask, as per [21].

3.2. Sample Repartition.

The CUT-like methods adopt a multilayer, patch-wise contrastive loss function, known as PatchNCE [30]. As illustrated in Fig. 3 (a), this approach pairs a query subpatch p_q from the generated IHC image with a corresponding subpatch p_+ from the same spatial location in the input H&E image as a positive sample. Other subpatches in the input image serve as negative samples p_- . However, this partition incurs a conflict: subpatches with identical staining labels are incorrectly treated as negative samples and pushed away, confusing contrastive learning about staining signals that are crucial for accurate virtual IHC staining.

To address the limitations of CUT-like methods, we propose a refined sample partitioning strategy to select positive and negative subpatches based on staining-positive masks generated by the Grad-CAM. As shown in Fig. 3 (b), each subpatch p_s is assigned with a staining label $l_s \in \{0, 1\}$ based on the segmented mask. In the IHC staining-related space, a query subpatch p_q is paired with all subpatches sharing the same staining label, designed as positive samples p_{++} and p_{-+} , while subpatches with differing staining labels are considered negative samples p_{--} . In the IHC staining-unrelated space, the partitioning aligns with that of the CUT [30]. The dual-subscript notation is adopted to indicate each sample's properties, as outlined below:

- p_{++} : the patch in X located at the same position as p_q , sharing the same staining label;
- p_{-+} : patches in X that share the same staining label as p_q , but are at different spatial locations;
- p_{--} : patches in X with a different staining label and located at different spatial positions from p_q ;

3.3. ODA-GAN.

As shown in Fig. 1 (b), our ODA-GAN is trained using four objective functions: 1) orthogonal decoupling loss, 2) ad-

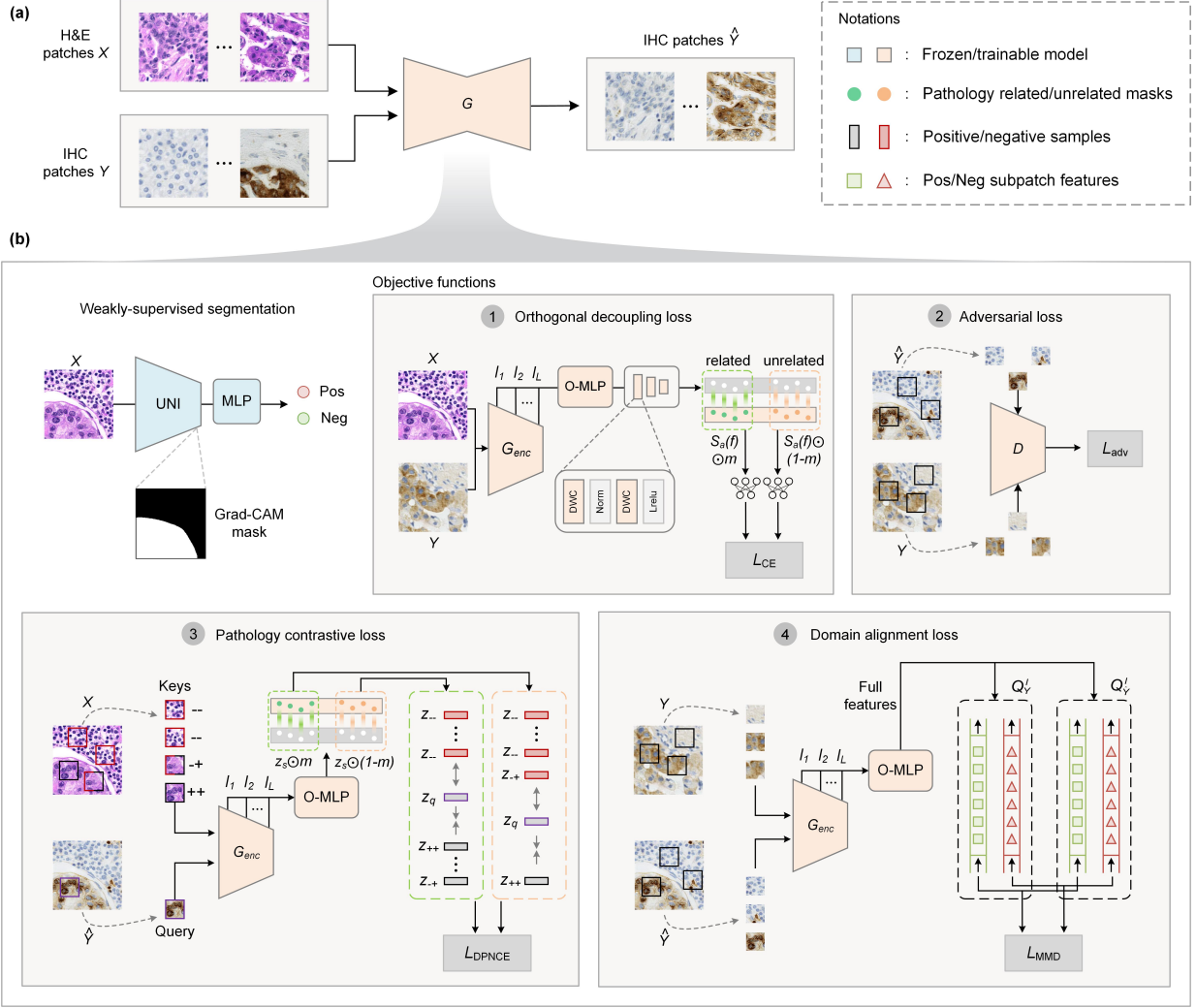


Figure 1. Overview of our ODA-GAN. (a) ODA-GAN consists of a generator G that takes unpaired H&E and IHC images as input and is trained to convert H&E images into IHC images; (b) Based on weakly-supervised segmentation masks, the ODA-GAN optimizes several losses during training: orthogonal decoupling loss, adversarial loss, pathology contrastive loss, and domain alignment loss.

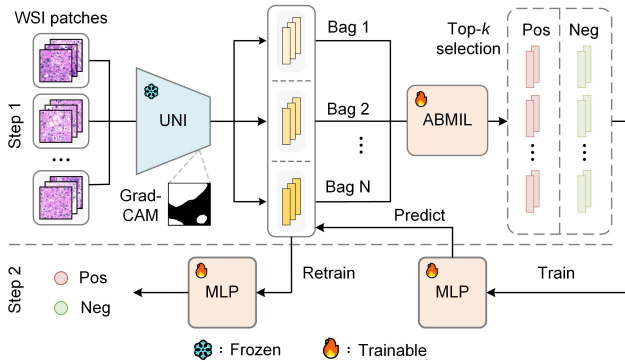


Figure 2. Overview of our weakly-supervised segmentation.

versarial loss, 3) pathology contrastive loss, and 4) domain alignment loss. The first loss ensures effective decoupling of image features, while the remaining three losses enforce effective staining translation, encouraging realistic outputs while maintaining content consistency.

Orthogonal decoupling loss. To decouple staining-related and unrelated features, the relevance between the two types of features should be reduced. Therefore, we introduce an Orthogonal MLP (O-MLP) that is constrained on the Stiefel manifold [37] to project image features into an orthogonal space to minimize their mutual relevance. The projection of each input subpatch p_s sampled from a source image x_i is defined as:

$$z_s = H_{\Theta} (G_{enc}^l(p_s)), \text{ s.t. } \Theta^T \Theta = I \quad (1)$$

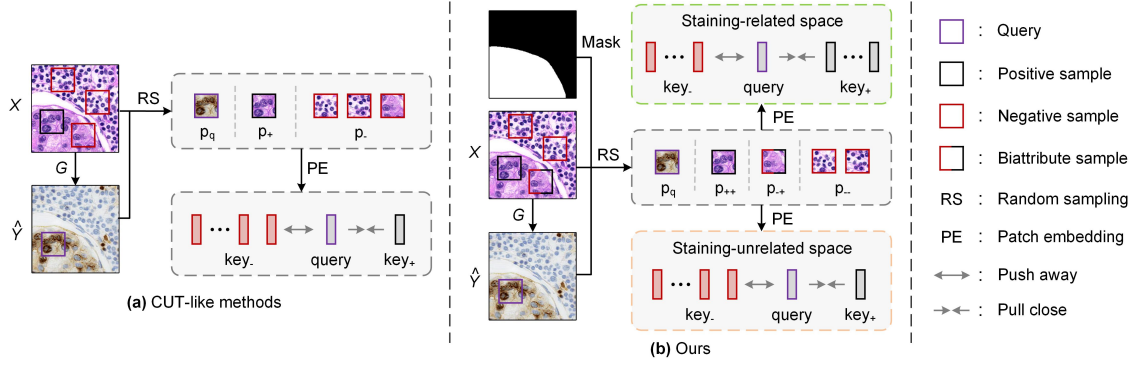


Figure 3. Illustration of sample repartition using (a) CUT-like methods and (b) our method. CUT-like methods aim to maximize the resemblance between H&E and virtually synthesized IHC images by pulling the query and its corresponding key subpatch at the same location closer, while pushing other subpatches away. In contrast, our method treats all keys sharing the same staining labels to the query as positive samples in the staining-related space.

where $H_{\Theta}(\cdot)$ denotes the O-MLP with parameter matrix Θ that processes feature maps extracted from the l -th layer of G_{enc} , $\forall l \in \{1, 2, \dots, L\}$.

Similarly, feature maps of x_i are projected into orthogonal space using the O-MLP, denoted as f , and then fed into a deep-wise encoder $E(\cdot)$, i.e., $E(f) \triangleq S_b \circ (S_a(f) \odot \mathbf{m})$, where $\odot$ denotes function composition. The encoder $E(\cdot)$ is organized into three components: deep-wise convolutional (DWC) layers S_a , a learnable masking layer $\mathbf{m}$, and fully-connected layers S_b . A globally average pooling (GAP) operation is included in S_a to condense feature maps into a feature vector. The learnable mask $\mathbf{m}$, with dimensions matching the feature vector, has elements constrained within the range $[0, 1]$. This mask is used to selectively retain features associated with antigen staining: the closer an element is to 1, the more likely the feature is related to antigen staining, aiming to decouple these features from unrelated components. Intuitively, a model should rely exclusively on staining-related features to identify staining labels of input images; whereas using staining-unrelated features will lead to inaccurate predictions. The optimization of our feature decoupling is based on the following loss $\mathcal{L}_{dec}$:

$$\mathcal{L}_{dec}(H_{\Theta}, E, f) = \mathcal{L}_{CE}(S_b \circ (S_a(f) \odot \mathbf{m}), l_f) - \mathcal{L}_{CE}(S_b \circ (S_a(f) \odot (1 - \mathbf{m})), l_f), \quad (2)$$

where $\mathcal{L}_{CE}$ is the cross-entropy loss function; $\odot$ is the element-wise product; $l_f \in \{0, 1\}$ is the pseudo-label of a feature map f for the image x_i .

Adversarial loss. The adversarial loss $\mathcal{L}_{adv}(\cdot)$ is a standard GAN min-max loss [13], implemented with a PatchGAN discriminator D [17] and calculated as follows:

$$\mathcal{L}_{adv}(G, D, X, Y) = \mathbb{E}_{y \sim Y} \log D(y_i) + \mathbb{E}_{x \sim X} \log(1 - D(G(x_i))). \quad (3)$$

Pathology contrastive loss. Given mask $\mathbf{m}$ that can decouple feature maps into staining-related and unrelated components, we introduce a Dual-stream PatchNCE (DPNCE) loss, as illustrated in the third panel of Fig. 1 (b). Unlike the PatchNCE loss [30], where subpatches sharing the same staining labels are incorrectly treated as negative samples, DPNCE loss aims to circumvent the contradiction between maximizing mutual information and ensuring pathological consistency. Here, let us first define the masked mutual information $l(\cdot)$ between two feature vectors z_1, z_2 as:

$$l(\mathbf{m}, z_1, z_2) = \exp(\mathbf{m} \odot z_2^T \times z_1 / \tau), \quad (4)$$

where τ is a temperature scaling parameter.

In the staining-related stream, mutual information between the query subpatch and all key subpatches sharing the same staining labels should be maximized. This involves maximizing the mutual information $\mathcal{P}$ of p_q with p_{++} and the feature centroid of all subpatches p_{-+} :

$$\mathcal{P} = l(\mathbf{m}, z_q, z_{++}) + l\left(\mathbf{m}, z_q, \frac{1}{N_{-+}} \sum_{n=1}^{N_{-+}} z_{-+}^n\right), \quad (5)$$

where z_q, z_{++} , and z_{-+} separately represent feature vectors of p_q, p_{++}, p_{-+} , and N_{-+} is the number of patches p_{-+} . Simultaneously, mutual information $\mathcal{N}$ between the query p_q and key patches with different staining labels should be minimized:

$$\mathcal{N} = \sum_{n=1}^{N_{--}} l(\mathbf{m}, z_q, z_{--}^n). \quad (6)$$

Thus, the contrastive loss for positive staining regions $\mathcal{L}_{pos}$ is calculated as:

$$\mathcal{L}_{pos} = -\log\left(\frac{\mathcal{P}}{\mathcal{P} + \mathcal{N}}\right). \quad (7)$$

The $\mathcal{L}_{neg}$ for negative staining regions is calculated in the same manner. The loss of this stream $\mathcal{L}_{related}$ is a combination of these two terms:

$$\mathcal{L}_{related} = \frac{1}{2} (\mathcal{L}_{pos} + \mathcal{L}_{neg}). \quad (8)$$

In the staining-unrelated stream, only the key subpatch at the same spatial location as the query subpatch is treated as a positive sample. Thus, the $\mathcal{L}_{unrelated}$ loss follows the PatchNCE [30] loss, except that the masked mutual information between two feature vectors is calculated as:

$$l((\mathbf{1} - \mathbf{m}), z_1, z_2) = \exp((\mathbf{1} - \mathbf{m}) \odot z_2^T \times z_1 / \tau). \quad (9)$$

Our DPNCE loss is then computed as the sum of dual-stream losses :

$$\mathcal{L}_{DPNCE}(G, H_\Theta, X) = \mathcal{L}_{related} + \mathcal{L}_{unrelated}. \quad (10)$$

Additionally, we compute the DPNCE loss using Y , $\mathcal{L}_{DPNCE}(G, H_\Theta, Y)$, as an identity loss, to prevent the generator G from making unnecessary changes based on the DAB mask. Finally, the pathology contrastive loss $\mathcal{L}_{con}$ is computed as the sum of two DPNCE losses:

$$\mathcal{L}_{con} = \mathcal{L}_{DPNCE}(G, H_\Theta, X) + \mathcal{L}_{DPNCE}(G, H_\Theta, Y). \quad (11)$$

Domain alignment loss. To achieve more realistic image generation, we propose a Multi-layer Domain Alignment (MDA) loss $\mathcal{L}_{ada}$ to bridge real (i.e., Y) and virtual (i.e., $\hat{Y}$) IHC image domains, as illustrated in the fourth panel of Fig. 1 (b). During training, full features of subpatches (without decoupling) are pushed into corresponding dynamic queue list (i.e., Q_Y or $Q_{\hat{Y}}$) determined by their staining labels for updating. Each layer of G_{enc} has queue lists, and we minimize the Maximum Mean Discrepancy (MMD) distance [4] between feature vectors in Q_Y^l and $Q_{\hat{Y}}^l$ to encourage the generated IHC images align the distribution with real IHC images, which is defined as:

$$\text{MMD}_l(Q_Y^l, Q_{\hat{Y}}^l) = \left\| \frac{1}{|Q_Y^l|} \sum_{z_y \in Q_Y^l} \phi(z_y) - \frac{1}{|Q_{\hat{Y}}^l|} \sum_{z_{\hat{y}} \in Q_{\hat{Y}}^l} \phi(z_{\hat{y}}) \right\|^2, \quad (12)$$

where $\phi(\cdot)$ is a kernel function [4], operating on source queue features, $z_y \in Q_Y^l$, and target queue features, $z_{\hat{y}} \in Q_{\hat{Y}}^l$. The $\mathcal{L}_{ada}$ loss is then defined as the sum of the MMD distances at each layer:

$$\mathcal{L}_{ada} = \sum_{l=1}^L \text{MMD}_l(Q_Y^l, Q_{\hat{Y}}^l). \quad (13)$$

Taken together, the total loss $\mathcal{L}$ for optimizing our ODA-GAN is given by:

$$\mathcal{L} = \mathcal{L}_{dec} + \mathcal{L}_{adv} + \mathcal{L}_{con} + \mathcal{L}_{ada}. \quad (14)$$

4. Experiments

4.1. Setup

Datasets. We conduct experiments on three benchmark datasets, namely BCI [23], MIST_{HER2} [20], and GPC3. Both BCI and MIST_{HER2} are paired datasets, each consisting of patches sized 1024×1024 pixels. The BCI dataset consists of 3396 H&E-HER2 training patches and 500 testing patches from 51 WSI pairs. The MIST_{HER2} consists of 4642 H&E-HER2 training patches and 1000 testing patches from 64 WSI pairs. The GPC3 is a private unpaired dataset, consisting of patches sized 256×256 pixels. It comprises about 1.0 million patches derived from 50 H&E and 50 GPC3 stained WSIs.

Evaluation metrics. We evaluate the similarity between real and virtual IHC images using several metrics, including FID [28], KID [3], and LPIPS [40] for comparing image distributions, and SSIM [36] for measuring structural similarity. Additionally, we introduce the DAB_{KL} metric, which assesses staining distribution similarity by calculating the Kullback-Leibler (KL) divergence between 256-dimensional gray-level histogram vectors of the DAB channel for paired images. Since real and virtual IHC images are not perfectly aligned, calculating KL divergence between feature vectors of the DAB channel offers a more accurate evaluation of staining distribution consistency.

4.2. Performance Evaluation

BCI and MIST_{HER2}. We compared our method against five representative fully supervised methods: Pix2pix [17], PyramidP2P [23], TDKStain [32], PSPStain [8], and ASP [20], on two paired datasets. Fig. 4 shows visual comparisons of generated IHC images between ours and SOTA methods on four randomly selected images. In Fig. 4, red continuous rectangles indicate areas where tissue morphology errors occur in the generated IHC images, while blue dashed rectangles highlight regions with staining errors. It is observed in Fig. 4 that the Pix2pix, PyramidP2P, and TDKStain methods often fail to preserve the tissue morphology of input H&E images, leading to noticeable tissue morphological distortions and artifacts in the generated IHC images. This issue arises from intrinsic content discrepancies between the input H&E images and IHC references used as ground truths, which negatively impact model optimization. The CUT-like methods, such as ASP and PSPStain, which leverage contrastive learning, demonstrate strong performance in preserving tissue morphology of input H&E images. However, the misalignment of supervised staining-related information may introduce errors in virtual staining results. In contrast, our proposed ODA-GAN achieves superior visual results than these comparative methods.

The quantitative comparison of different methods on

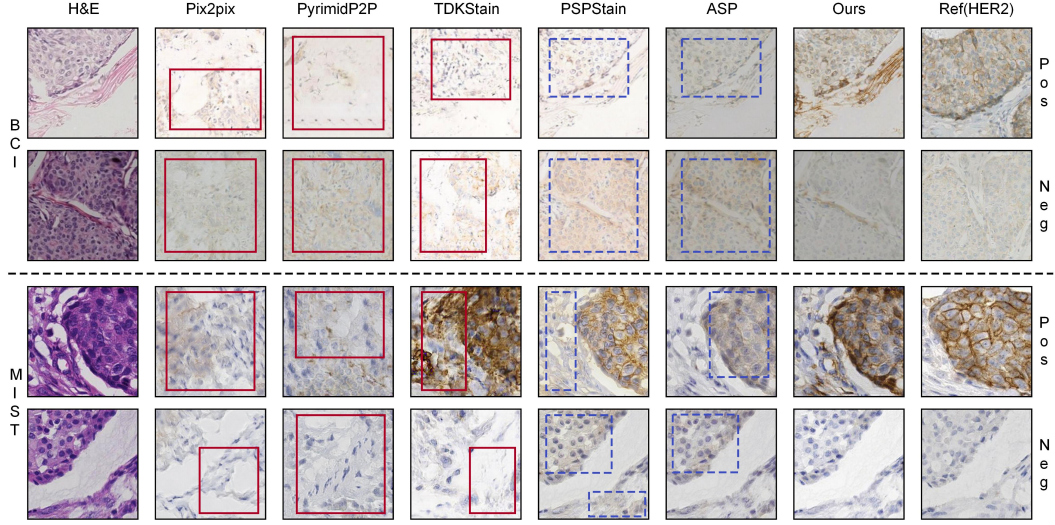


Figure 4. Visual comparison of SOTA virtual staining methods and ours. Red continuous rectangles indicate generated regions with texture errors, while blue dashed rectangles highlight areas with pathological information errors. Note that the referenced IHC images are not perfectly aligned with the H&E images, as they are generated from adjacent tissue sections.

Table 1. Quantitative comparison of virtual staining on BCI. Note that KID values multiplied by 1000 are shown.

Methods	KID↓	FID↓	LPIPS↓	SSIM↑	DAB _{KL} ↓
Pix2pix [20]	56.1	112.5	0.358	0.4857	3.1203
PyramidP2P [23]	36.6	124.8	0.398	0.4672	7.4862
TDKStain [32]	35.4	108.6	0.352	0.4504	4.2043
PSPStain [8]	15.1	99.0	0.375	0.4506	3.2129
ASP [20]	12.7	68.7	0.389	0.4917	3.0166
Ours	10.1	59.3	0.350	0.4733	2.9329

Table 2. Quantitative comparison of virtual staining on MIST_{HER2}.

Methods	KID↓	FID↓	LPIPS↓	SSIM↑	DAB _{KL} ↓
Pix2pix [17]	102.0	147.6	0.470	0.1981	2.6126
PyramidP2P [23]	80.3	180.5	0.467	0.1979	1.8471
TDKStain [32]	43.0	138.3	0.460	0.1617	1.9276
PSPStain [8]	34.0	148.1	0.446	0.1738	2.1065
ASP [20]	23.9	89.3	0.455	0.2004	2.9478
Ours	8.6	68.0	0.420	0.1893	1.2832

BCI and MIST_{HER2} is presented in Table 1 and Table 2. It is observed that our ODA-GAN achieves the best results in KID, FID, LPIPS and DAB_{KL} metrics across both datasets, as well as the second-best results in SSIM, demonstrating the effectiveness of our approach in IHC virtual staining. This superior performance is primarily attributed to the capacity of our ODA-GAN to preserve input texture fidelity and to generate realistic stains in virtual IHC images.

GPC3. We also compared our method against two representative and three latest unsupervised methods: CycleGAN [43], CUT [30], DCD [15], Confusion-GAN [21], and StegoGAN [38], on the unpaired GPC3 dataset. Fig. 5

shows a visual comparison of randomly selected positive and negative samples. Notably, all methods successfully preserve the tissue morphology of input H&E images, primarily due to the cycle consistency constraints, PatchNCE loss, or their respective variants. However, existing methods deliver inaccurate positive staining results due to inherent limitations, such as strict bijection constraints or conflicts in contrastive learning. In contrast, our ODA-GAN effectively generates staining results that closely match the distribution and patterns of real IHC images. Table 3 provides quantitative comparisons of virtual staining on the GPC3 dataset. Our ODA-GAN outperforms all existing methods across evaluated metrics, demonstrating the efficacy of our approach.

Table 3. Quantitative comparison of virtual staining on GPC3.

Methods	KID↓	FID↓	LPIPS↓	SSIM↑	DAB _{KL} ↓
CycleGAN [43]	24.1	91.7	0.502	0.2311	4.2925
CUT [30]	25.3	89.3	0.509	0.2214	3.4831
DCD [15]	20.8	79.4	0.495	0.2630	3.5784
Confusion-GAN [21]	19.4	65.0	0.491	0.2649	3.6612
StegoGAN [38]	4.7	39.4	0.489	0.2816	2.9114
Ours	3.2	31.4	0.472	0.2944	2.6553

4.3. Ablation Study and Analysis

Effect of UNI and top- k selection. Since the weakly-supervised segmentation pipeline plays a critical role in assisting ODA-GAN model training, we conducted an ablation study to investigate key factors in the pipeline, including the UNI backbone, the MLP training strategy, and the parameter k in top- k selection. We evaluated patch-level

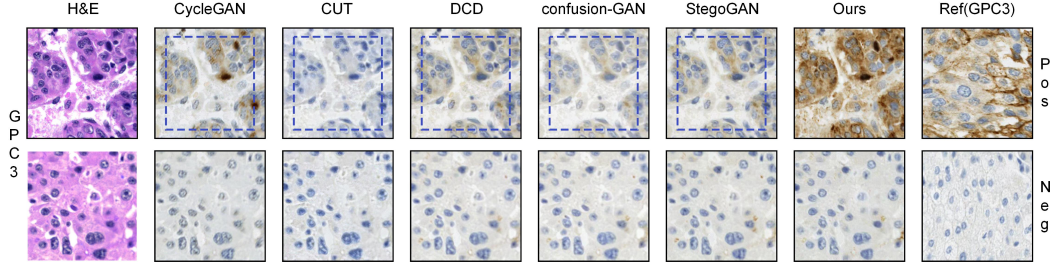


Figure 5. Visual comparison of SOTA virtual staining methods and ours on the GPC3 dataset. Blue dashed rectangles highlight areas with pathological information errors.

classification on the GPC3 test set, using AUC [14], FPR, TPR, and G-mean [18] metrics. Additionally, we recruited a pathologist to label 200 H&E images by referencing adjacent IHC images in the GPC3 dataset. We evaluated patch-level segmentation using IOU [12] and DICE [29] metrics.

As shown in Table 4, our UNI-based pseudo-labeling method significantly outperforms both the PPIE method [21], which trains a small CNN on a synthesized augmented dataset that leverages a diffusion model, and the Fine-tuning method, which only trains an MLP on top- k selected instances embedded by UNI. This phenomenon is primarily attributed to the easy-to-separate feature embedding of UNI and the effective data exploitation in our pipeline. It is also found in Table 4 that setting k to 30 yields optimal results in this study. A smaller k provides insufficient training data for the MLP, whereas a larger k or without (wo) top- k selection may introduce incorrect labels, compromising training accuracy.

Table 4. Patch-level classification and segmentation results obtained from different top- k patches on the GPC3 dataset.

Methods	k	AUC $\uparrow$	FPR $\downarrow$	TPR $\uparrow$	G-mean $\uparrow$	DICE $\uparrow$	IOU $\uparrow$
PPIE [21]	10	0.713	0.277	0.641	0.680	0.65	0.49
	30	0.782	0.226	0.726	0.750	0.71	0.55
	100	0.637	0.408	0.589	0.591	0.57	0.41
	wo	0.732	0.323	0.686	0.681	0.70	0.52
Fine-tuning	10	0.798	0.055	0.344	0.571	0.62	0.46
	30	0.829	0.098	0.576	0.720	0.65	0.49
	100	0.825	0.084	0.523	0.692	0.65	0.50
	wo	0.813	0.118	0.589	0.721	0.69	0.54
Our pseudo-labeling	10	0.819	0.100	0.453	0.639	0.74	0.60
	30	0.933	0.204	0.935	0.863	0.83	0.71
	100	0.896	0.204	0.841	0.818	0.78	0.64
	wo	0.911	0.162	0.859	0.848	0.79	0.66

Effect of DPNCCE and MDA. To further evaluate the contributions of DPNCCE and MDA modules to our ODA-GAN model, we conducted an ablation study on the GPC3 dataset. As shown in Table 5, both DPNCCE and MDA significantly enhance performance, with their combination yielding the best results. Specifically, DPNCCE effectively mitigates the adverse effects of contradictions within the CUT baseline, thereby promoting pathological consistency.

Meanwhile, the MDA module bridges the domain gap between virtual and real IHC images, enabling ODA-GAN to generate more realistic staining results that closely match the distribution of real IHC images.

Table 5. Exploring the impact of DPNCCE and MDA on GPC3.

DPNCCE	$\times$	$\checkmark$	$\times$	$\checkmark$
MDA	$\times$	$\times$	$\checkmark$	$\checkmark$
KID $\downarrow$	25.3	17.8	12.0	3.2
FID $\downarrow$	89.3	60.2	65.8	31.4
LPIPS $\downarrow$	0.509	0.488	0.495	0.472
SSIM $\uparrow$	0.2214	0.2640	0.2737	0.2944
DAB _{KL} $\downarrow$	3.4831	3.3321	3.3396	2.6553

Subjective Evaluation. We conducted a final visual Turing test to assess the indistinguishability between real and virtual IHC images based on pathologists’ visual judgments. The results showed that our ODA-GAN generated realistic staining, successfully deceiving the experts. More details are available in the supplementary material.

5. Conclusion

In this paper, we propose a framework called ODA-GAN, assisted by weakly-supervised learning, for virtual IHC staining in an unpaired setting. This method is motivated by our analysis of limitations in existing virtual IHC staining methods, which often produce unrealistic or unreliable results. We address existing issues in an ingenious manner based on recent technical advances. Results across three benchmark datasets validate the effectiveness of our approach. Considering that existing unpaired methods perform well in preserving tissue morphology, we believe that future work should still focus on generating more realistic and accurate stains for virtual IHC images in this domain.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant 82102135, and the Fundamental Research Funds for Central Universities under Grant DUT24YG201.

References

- [1] Jiwoon Ahn and Suha Kwak. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4981–4990, 2018. [2](#)
- [2] Bijie Bai, Xilin Yang, Yuzhu Li, Yijie Zhang, Nir Pillar, and Aydogan Ozcan. Deep learning-enabled virtual histological staining of biological samples. *Light: Science & Applications*, 12(1):57, 2023. [1](#)
- [3] Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*, 2018. [6](#)
- [4] Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006. [6](#)
- [5] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019. [2](#)
- [6] Marc-André Carboneau, Veronika Cheplygina, Eric Granger, and Ghyslaine Gagnon. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, 77:329–353, 2018. [2](#)
- [7] John KC Chan. The wonderful colors of the hematoxylin-eosin stain in diagnostic surgical pathology. *International journal of surgical pathology*, 22(1):12–32, 2014. [1](#)
- [8] Fuqiang Chen, Ranran Zhang, Boyun Zheng, Yiwen Sun, Jiahui He, and Wenjian Qin. Pathological semantics-preserving learning for h&e-to-ihc virtual staining. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 384–394. Springer, 2024. [1](#), [6](#), [7](#)
- [9] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, 2024. [3](#)
- [10] Leandro Luongo De Matos, Damila Cristina Trufelli, Maria Graciela Luongo De Matos, and Maria Aparecida da Silva Pinhal. Immunohistochemistry as an important tool in biomarkers detection and clinical practice. *Biomarker insights*, 5:BMI-S2185, 2010. [1](#)
- [11] B DoanNgan, DarrowMorgan Angus, LeeHan Sung, et al. Label-free virtual her2 immunohistochemical staining of breast tissue using deep learning. *BME frontiers*, 2022. [1](#)
- [12] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010. [8](#)
- [13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. [2](#), [5](#)
- [14] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982. [8](#)
- [15] Tie Hu, Mingbao Lin, Lizhou You, Fei Chao, and Rongrong Ji. Discriminator-cooperated feature map distillation for gan compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20351–20360, 2023. [7](#)
- [16] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018. [2](#), [3](#)
- [17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. [1](#), [2](#), [5](#), [6](#), [7](#)
- [18] Miroslav Kubat, Stan Matwin, et al. Addressing the curse of imbalanced training sets: one-sided selection. In *ICML*, page 179. Citeseer, 1997. [8](#)
- [19] Amal Lahiani, Irina Klamann, Nassir Navab, Shadi Albarqouni, and Eldad Klaiman. Seamless virtual whole slide image synthesis and validation using perceptual embedding consistency. *IEEE Journal of Biomedical and Health Informatics*, 25(2):403–411, 2020. [2](#)
- [20] Fangda Li, Zhiqiang Hu, Wen Chen, and Avinash Kak. 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*, pages 632–641. Springer, 2023. [1](#), [2](#), [6](#), [7](#)
- [21] Jiahua Li, Jiuyang Dong, Shenjin Huang, Xi Li, Junjun Jiang, Xiaopeng Fan, and Yongbing Zhang. Virtual immunohistochemistry staining for histological images assisted by weakly-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11259–11268, 2024. [2](#), [3](#), [7](#), [8](#)
- [22] Shuting Liu, Baochang Zhang, Yiqing Liu, Anjia Han, Huijuan Shi, Tian Guan, and Yonghong He. Unpaired stain transfer using pathology-consistent constrained generative adversarial networks. *IEEE transactions on medical imaging*, 40(8):1977–1989, 2021. [1](#), [2](#)
- [23] Shengjie Liu, Chuang Zhu, Feng Xu, Xinyu Jia, Zhongyue Shi, and Mulan Jin. Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1815–1824, 2022. [2](#), [6](#), [7](#)
- [24] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021. [2](#)
- [25] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-

- language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, 2024. 3
- [26] Xiaotong Luo, Wenjin Yang, Yuan Xie, and Yanyun Qu. Farewell to cyclegan: Single gan with decoupled constraint for unpaired image dehazing. Available at SSRN 4620491. 2
- [27] Oded Maron and Tomás Lozano-Pérez. A framework for multiple-instance learning. *Advances in neural information processing systems*, 10, 1997. 2
- [28] Fabian Mentzer, George D Toderici, Michael Tschannen, and Eirikur Agustsson. High-fidelity generative image compression. *Advances in Neural Information Processing Systems*, 33:11913–11924, 2020. 6
- [29] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016. 8
- [30] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. 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*, pages 319–345. Springer, 2020. 2, 3, 5, 6, 7
- [31] Pushpak Pati, Sofia Karkampouna, Francesco Bonollo, Eva Comp  rat, Martina Radi  , Martin Spahn, Adriano Martinelli, Martin Wartenberg, Marianna Kruithof-de Julio, and Marianna Rapsomaniki. Accelerating histopathology workflows with generative ai-based virtually multiplexed tumour profiling. *Nature Machine Intelligence*, pages 1–17, 2024. 1
- [32] Qiong Peng, Weiping Lin, Yihuang Hu, Ailisi Bao, Chenyu Lian, Weiwei Wei, Meng Yue, Jingxin Liu, Lequan Yu, and Liansheng Wang. 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*, pages 3–13. Springer, 2024. 6, 7
- [33] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 2, 3
- [34] Kexin Sun, Zhineng Chen, Gongwei Wang, Jun Liu, Xiongjun Ye, and Yu-Gang Jiang. Bi-directional feature fusion generative adversarial network for ultra-high resolution pathological image virtual re-staining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3904–3913, 2023. 1
- [35] Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis*, 81:102559, 2022. 3
- [36] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6
- [37] Zhongze Wang, Haitao Zhao, Jingchao Peng, Lujian Yao, and Kaijie Zhao. Odc: Orthogonal decoupling contrastive regularization for unpaired image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25479–25489, 2024. 4
- [38] Sidi Wu, Yizi Chen, Samuel Mermet, Lorenz Hurni, Konrad Schindler, Nicolas Gonthier, and Loic Landrieu. Stegogan: Leveraging steganography for non-bijective image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7922–7931, 2024. 7
- [39] Bowei Zeng, Yiyang Lin, Yifeng Wang, Yang Chen, Jiuyang Dong, Xi Li, and Yongbing Zhang. Semi-supervised pr virtual staining for breast histopathological images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–241. Springer, 2022. 2
- [40] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6
- [41] Wei Zhang, Tik Ho Hui, Pui Ying Tse, Fraser Hill, Condon Lau, and Xinyue Li. High-resolution medical image translation via patch alignment-based bidirectional contrastive learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 178–188. Springer, 2024. 2
- [42] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016. 2
- [43] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 1, 7