

# Pseudo Training Data Generation for Unsupervised Cell Membrane Segmentation in Immunohistochemistry Images

Xi Long<sup>1†</sup>, Tianyang Wang<sup>1†</sup>, Yanjia Kan<sup>1</sup>, Yunze Wang<sup>1</sup>, Silin Chen<sup>2</sup>, Albert Zhou<sup>3</sup>, Xianxu Hou<sup>1</sup>, Jingxin Liu<sup>1\*</sup>

<sup>1</sup>Xi'an Jiaotong-Liverpool University, School of AI and Advanced Computing, Suzhou, China

<sup>2</sup>Beijing Jiaotong University, School of Mathematics and Statistics, Beijing, China

<sup>3</sup>University of Warwick, UK

**Abstract**—In the realm of clinical diagnostics and medical research, quantitative assessment of membrane activity in immunohistochemistry (IHC) images is standard practice. Despite a high demand for cell membrane segmentation, only a few algorithms have been developed, and there is a lack of open datasets in this field. In this paper, we propose a three-stage unsupervised framework to accurately segment positive cell membranes in IHC images. Our approach transforms the unsupervised segmentation task into a supervised one by generating pseudo-paired training data using Voronoi diagrams and CycleGAN. Additionally, we introduce a dual encoder segmentation model with domain adaptation modules to mitigate the domain shift between generated images and real images. To our best knowledge, this is the first work focusing on unsupervised learning for IHC cell membrane segmentation. Extensive experiments and ablation studies on our newly built IHC cell membrane segmentation dataset validate the effectiveness of our framework.

**Index Terms**—Histopathology, Immunohistochemistry, Membrane Segmentation, Unsupervised Semantic Segmentation

## I. INTRODUCTION

Immunohistochemistry (IHC) is a fundamental technique widely employed in both medical research and clinical diagnostics. By utilizing monoclonal and polyclonal antibodies, IHC facilitates the visualization and quantification of specific proteins within tissues, including those located on cell membranes, exemplified by the majority of CD antigens (i.e., HER2, PD-L1, CD30, etc.) [1], [2]. With the ongoing advancements in precision diagnosis and immunotherapy, there is a pressing need for precise interpretation of IHC, highlighting the crucial role of expertise in this domain.

Given the recent advances in the field of computer vision, an increasing number of clinicopathologic Computer-Aided Diagnosis (CAD) systems are being developed to assist pathologists in providing precise quantitative information [3]–[7]. The accurate segmentation of positively stained cell membranes represents a key step in analyzing membrane-stained IHC sections [8]–[11]. Despite the considerable demand for cell membrane segmentation, only a limited number of algorithms have been developed for this purpose.

Early algorithms primarily utilize image processing techniques due to the distinct appearance of the positively stained cell membrane, which typically exhibits a crisp brown colour, contrasting sharply with the achromatous negative cell membrane and the blue nucleus. [12] proposed employing Support Vector Machines (SVM) to classify pixels into brown staining and others, whereas colour deconvolution [13] followed by a series of post-processing steps proposed in [9] to segment the positive membrane. To recognize the entire cell membrane, [14] first detected the nucleus and then used nuclear membranes as a spatial reference to reconstruct the unstained tracts. Traditional methods, less reliant on precise annotations, struggle with complex situations like noise from overlapping and diffusing stains in pathology images. They are inefficient and time-consuming when using color thresholds for pixel classification or membrane segmentation in such noisy conditions.

Unsupervised learning (UL) methods achieved great success in the data dependency problem for medical imaging, which learns from the structural properties in the data without any manual annotations. Based on the character of these methods, UL can be generally divided into two classes: clustering based [15]–[18] and adversarial learning based [19]–[21]. MaskContrast [15] utilized unsupervised contrastive learning to create pixel embeddings for semantic segmentation, while STEGO [16] applied a similar contrastive approach, facilitating effective unsupervised segmentation by focusing on distinct object features. Both CRLC [17] and CC [18] improved image clustering by performing instance- and cluster-level contrastive learning, enhancing both representations and cluster assignments in an online and end-to-end manner. For adversarial based methods, Labels4Free [19] augmented StyleGAN [22] by integrating a segmentation branch, producing soft masks for foreground objects. ReDO [20] used an adversarial architecture to redraw objects within images, ensuring realistic outputs and high-quality masks. GANSeg [21] employed a hierarchical GAN to generate mask-image pairs, enhancing segmentation without requiring annotations. All methods leverage GANs to improve unsupervised segmentation quality. However, these

<sup>†</sup> These authors contributed equally to this work.

methods often struggle to generalize to new medical imaging domains, require substantial computational resources, and depend on high-quality unlabeled data. When the data quality is poor or noisy, their performance significantly degrades.

The segmentation of curvilinear objects in medical images [23]–[27] has traditionally relied on heuristic rules and specialized filters. For instance, the vesselness filter introduced in [23] utilized the Hessian matrix to enhance tube-like curvilinear objects. Further developments include the B-COSFIRE filters [24] were designed for denoising retinal images and segmenting retinal vessels. Additionally, contrast-limited adaptive histogram equalization techniques [25] enhanced image contrast before segmenting retinal vessels using hand-crafted filters. The optimally oriented flux (OOF) [26], [27] improved curvilinear tube-like objects and demonstrated improved performance in segmenting adjacent curvilinear structures, although it was sensitive to variations in their sizes. While these advancements have significantly improved the segmentation of curvilinear structures, traditional methods based on hand-crafted filters do not require training but demand meticulous parameter tuning for optimal performance. Furthermore, these optimized parameter settings are often data-dependent or even region-dependent, which limits their applicability across a wide variety of curvilinear objects. To overcome these challenges, Fractal-FDA synthesis methods [28] were developed to generate curvilinear structures by employing Fractal L-systems [29]. This innovative approach enhances the training of curvilinear object segmentation models by integrating these generated structures into unlabeled images, thereby improving adaptability and effectiveness across various imaging conditions. In contrast, our work introduces a unique application of Voronoi diagrams for generating pseudo-paired training data, significantly enhancing the segmentation of immunohistochemistry cell membranes through an innovative unsupervised learning framework.

Several studies have explored deep learning methods for cell membrane segmentation. For instance, [30] combined an SVM classifier to identify regions of interest (ROIs) with a U-net model for more accurate membrane region segmentation. [31] introduced a long short-term memory (LSTM) architecture for detecting cell membranes and nuclei, while an active learning-based technique was proposed by [32]. Further research, such as Her2Net [33], a deep learning framework that effectively segments and classifies cell membranes and nuclei in HER2-stained breast cancer images, leveraging a novel trapezoidal LSTM structure to enhance diagnostic accuracy. [34] also introduced a cell segmentation method using membrane pattern information and graph cuts, integrating directional cross-correlations with spatially constrained graph cuts to significantly enhance boundary detection accuracy in microscopy images. PS-Net [35] introduced a novel segmentation framework that leverages insights from human visual perception studies to enhance the accuracy of EM (Electron Microscopy) cell membrane segmentation. However, these methods typically require extensive annotations, making them labor-intensive and poorly suited for the complex visual context of IHC

pathology images, where distinguishing fine structures remains a significant challenge.

In this paper, we propose a three-stage unsupervised framework for accurately segmenting the positive cell membrane in IHC images, achieved by generating a pseudo-paired training set and domain adaptive segmentation model. Specifically, we first use random membrane structure images generated by Voronoi diagrams and randomly sampled IHC patch images to generate pseudo-paired data via CycleGAN [36]. Inspired by [37], we introduce *YNet* for membrane segmentation, which employs two encoders to handle low-frequency (LF) and high-frequency (HF) images through wavelet transformation, along with the dual domain adaptation module to address artificial perturbations from generated images. To the best of our knowledge, this is the first unsupervised semantic segmentation framework on immunohistochemical cell membrane images. The contribution of this work is three-fold:

- We present an unsupervised framework comprising three stages for cell membrane segmentation in IHC images;
- We introduce a novel segmentation model named *YNet*, which employs wavelet transformation to produce low-frequency (LF) and high-frequency (HF) images as inputs. Additionally, *YNet* incorporates a domain adaptive module for each frequency component to counteract the domain shift between generated data and real data.
- We evaluate the efficacy of our proposed framework through comprehensive experiments on a new IHC cell membrane segmentation dataset (to be released upon paper acceptance), demonstrating substantial improvement over existing methods.

## II. METHOD

The overview of our proposed framework in this study is structured into three interconnected stages, each critical to the segmentation of cell membranes images, as illustrated in Fig. 1. The first stage involves the generation of membrane curvilinear images using a geometric partitioning technique that simulates the intricate and varying structures of cell membranes. In the second stage, these generated images, along with real IHC image patches, are transformed into pseudo-paired training data employing CycleGAN. This pseudo-paired data serves as the input for the third stage, where a specialized dual-encoder segmentation model, incorporating wavelet transformation to handle different frequency components of images, is trained to differentiate and accurately segment the cell membranes. This model integrates domain adaptation techniques to address the variances between generated and real images, ensuring robust and precise segmentation performance.

### A. Membrane Curvilinear Image Generation

The Voronoi diagram is a geometric structure that divides space into regions, each region is determined by its nearest generated point, ensuring that any point within it is closer to its centre point than to other generated points [38]. Mathematically, the Voronoi diagram is defined by a set of generating points and the regions they create. Let  $P = \{p_1, p_2, \dots, p_n\}$

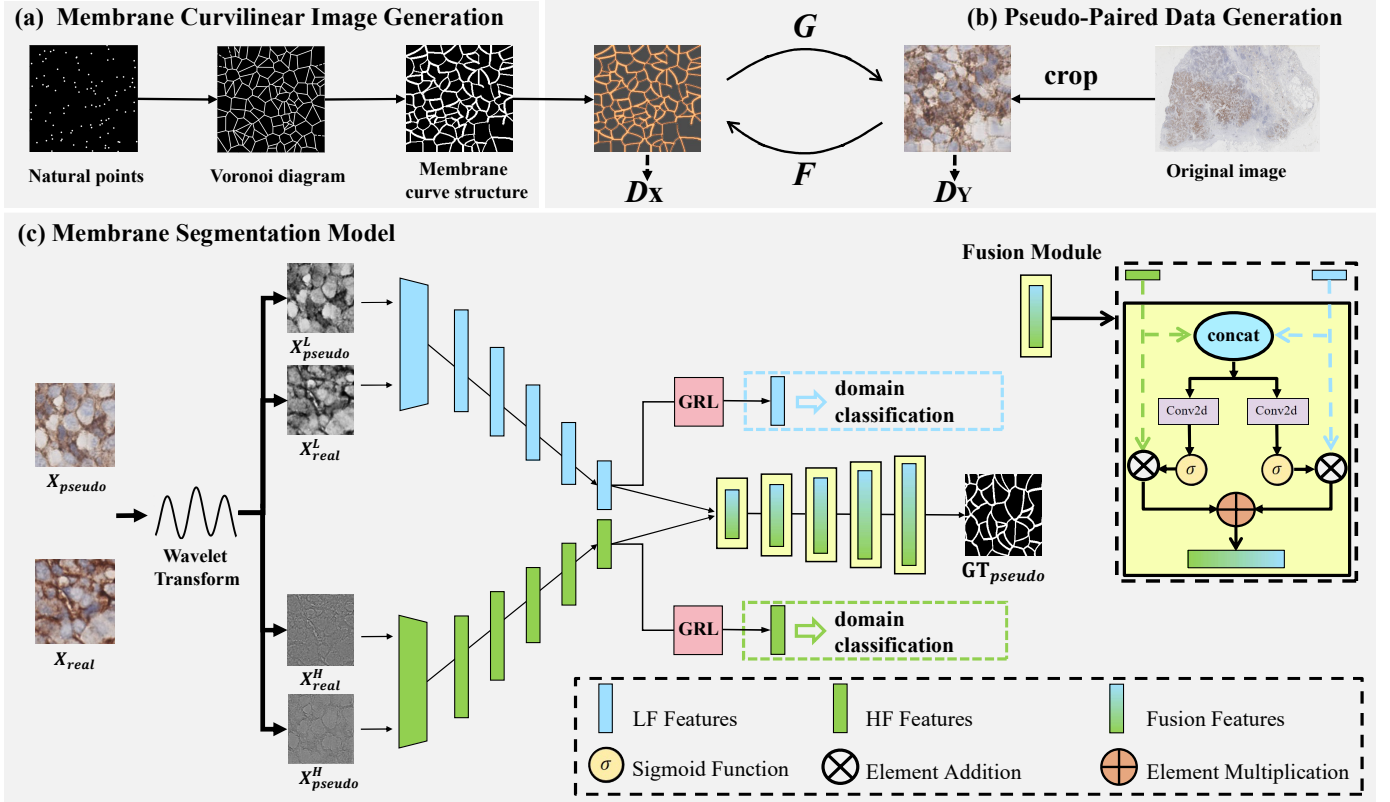


Fig. 1: Overview of our framework, which consists of three stages: (a) *Membrane Curvilinear Image Generation*: generating membrane structure image as the input of (b); (b) *Pseudo-paired Data Generation*: generating paired training data for (c); (c) *Membrane Segmentation Model*: the semantic segmentation model for the positive cell membrane segmentation.

be a set of  $n$  distinct points in a space  $S$ . The Voronoi region  $V(p_i)$  associated with each point  $p_i$  is defined as the set of all points  $x$  in  $S$  whose distance to  $p_i$  is less than or equal to its distance to any other point  $p_j$  in  $P$ . This can be formally expressed as:

$$V(p_i) = \{x \in S \mid \|x - p_i\| \leq \|x - p_j\| \text{ for all } p_j \in P, j \neq i\} \quad (1)$$

where  $\|x - p_i\|$  denotes the distance between the point  $x$  and the generated point  $p_i$ . The Voronoi diagram  $VD(P)$  of the set of points  $P$  is the collection of all Voronoi regions  $V(p_i)$  with each point  $p_i$  in  $P$ . This can be expressed as:

$$VD(P) = \{V(p_1), V(p_2), \dots, V(p_n)\} \quad (2)$$

The Voronoi diagram  $VD(P)$  of a point set  $P$  in a two-dimensional space  $S$  is partitioned by the vertical bisector of each point  $p_i$ . These bisectors define the region  $V$  of  $p_i$  centered at  $p_i$ , which contains all points that are at distance  $p_j$  or less than  $p_i$ . In essence, the Voronoi diagram consists of a series of equidistant regions centred at each point in the point set  $P$ .

**Natural perturbations.** To enhance the realism and complexity of biological cell membranes, we have augmented the traditional Voronoi diagram with natural perturbations. These perturbations introduce subtle irregularities into the Voronoi cell boundaries by incorporating controlled random noise into

the coordinates of the seed points. The perturbation is defined as follows:

$$p'_i = p_i + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \quad (3)$$

where  $p'_i$  denotes the coordinates of the seed points after perturbation. The noise  $\epsilon$  is sampled from a normal distribution  $\mathcal{N}(0, \sigma^2)$ . The standard deviation  $\sigma$  is chosen from the range  $(0.005, 0.01)$ . The number of seed points,  $n$ , is randomly chosen from the interval  $(90, 120)$ , primarily based on the density we observe in real IHC images to better match the real image distribution.

**Edge preservation.** To mimic the discontinuous and irregular edges of cell membranes, we propose a heuristic for selective edge preservation. Instead of keeping all Voronoi edges, we randomly delete a portion based on a specified edge retention rate within  $(0.6, 1.0)$ . This diversifies the Voronoi diagram and simulates the membrane's discontinuities and irregular edges.

**Smooth curve and random line width.** To simulate cell membrane contours, we integrate conic curves into the enhanced Voronoi diagram, creating smoother shapes between adjacent points. In addition, considering the non-uniform thickness of biological cell membranes, we randomly select curve widths ranging from  $(1, 5)$  to mimic thickness and morphology variations. Further, we enhance realism through colour mapping.

### B. Pseudo-Paired Data Generation

**Motivation.** In the previous section, we generate curvilinear structures resembling cell membranes using Voronoi diagrams. While these structures are morphologically similar to real cell membranes, they lack the realistic texture and colour style found in histopathology images. To address this limitation, we choose a classic and effective tool, CycleGAN [36], to migrate style from real pathology images to Voronoi diagram-generated ones (Fig. 1 (b)). CycleGAN is particularly well-suited for this task due to its ability to perform unpaired image-to-image translation, effectively learning the mapping between the generated and real images. This enables the creation of pseudo-pathology images, which can be used as paired data for subsequent semantic segmentation models.

**Style transfer process.** We leverage the CycleGAN framework to perform style transfer. CycleGAN operates by training two generators, denoted as  $G : X \rightarrow Y$  and  $F : Y \rightarrow X$ , where  $X$  represents the domain of Voronoi diagram-generated images and  $Y$  represents the domain of real pathology images. Additionally, it trains two discriminators,  $D_X$  and  $D_Y$ , which are responsible for distinguishing between source and translated images. To train CycleGAN, the adversarial loss is used to achieve the domain transfer as follows:

$$\begin{aligned} \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = & \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log D_Y(y)] \\ & + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log(1 - D_Y(G(x)))] \end{aligned} \quad (4)$$

$$\begin{aligned} \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) = & \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D_X(x)] \\ & + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log(1 - D_X(F(y)))] \end{aligned} \quad (5)$$

where  $\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y)$  represents the adversarial loss function for the generator  $G$  and discriminator  $D_Y$ , aiming to train  $G$  to generate images indistinguishable from real images in domain  $Y$ , and  $D_Y$  to correctly distinguish between real images from domain  $Y$  and those generated by  $G$  from domain  $X$ . Similarly,  $\mathcal{L}_{\text{GAN}}(F, D_X, Y, X)$  represents the adversarial loss function for the training from domain  $Y$  to domain  $X$ .

In addition to the adversarial losses, we also incorporate a cycle consistency loss as follows:

$$\begin{aligned} \mathcal{L}_{\text{cyc}}(G, F) = & \mathbb{E}_{x \sim p_{\text{data}}(x)} [|F(G(x)) - x|_1] \\ & + \mathbb{E}_{y \sim p_{\text{data}}(y)} [|G(F(y)) - y|_1] \end{aligned} \quad (6)$$

This loss facilitates bidirectional transformation, ensuring that images transformed between domains  $X$  and  $Y$  retain their original content. By integrating these loss functions, the model can effectively generate realistic cell membrane images suitable for unsupervised training of semantic segmentation networks.

### C. Membrane Segmentation Model

In this section, we introduce a novel *YNet* model, as depicted in Fig. 1 (c). Initially, we apply a wavelet transformation to decompose the input image into its high-frequency and low-frequency components. We then utilize two separate encoders

to extract features from these components. Furthermore, we employ a decoder to fuse the features from both frequencies, to produce the final segmentation. Finally, two domain adaptive modules are used to align the features extracted from pseudo (generated) and real cell membrane images.

**Wavelet transform.** We first employ wavelet transformation [39] to decompose raw images into low-frequency (LF) and high-frequency (HF) components of the IHC images, which helps in accurately identifying edges and textures. As illustrated in Fig. 1, we denote the low-frequency components for the pseudo and real membrane cell images as  $X_{\text{pseudo}}^L$  and  $X_{\text{real}}^L$ , respectively. Similarly,  $X_{\text{pseudo}}^H$  and  $X_{\text{real}}^H$  represent the corresponding high-frequency components. We represent LF images as  $L$  and HF images as  $H$ , where  $H$  is the sum of the horizontal, vertical, and diagonal HF components after normalization.  $L$  and  $H$  are defined as:

$$L = LL, \quad (7)$$

$$H = HL + LH + HH. \quad (8)$$

**Why choose wavelet transform?** Wavelet transform is particularly advantageous because it can effectively preserve both frequency and spatial location information of the image, which is essential for accurate segmentation. Compared to other methods like Fourier transform [40], wavelet transform is more efficient in generating LF and HF images. By utilizing wavelet transform, we can leverage the intrinsic *LF* and *HF* information of images to improve the segmentation performance. Specifically, for our cell membrane segmentation task, capturing the edges and details of curvilinear objects is a crucial. Processing high and low frequencies ensures that the model can accurately identify and segment the cell membranes. As illustrated in Fig. 2, the wavelet transform decomposes the raw image into low-frequency components ( $LL$ ), and high-frequency components in different directions ( $HL$ ,  $LH$ , and  $HH$ ), which helps retain both the overall structure and the intricate details necessary for precise segmentation.

**Encoder.** As shown in Fig. 1 (c), we utilize two distinct encoders to separately extract features from the low-frequency and high-frequency components of input images. These encoders share the same architecture but are trained with different weights. Each encoder comprises five convolutional blocks, with a reduction in feature size and a doubling of channel dimension after each block.

**Fusion module.** Inspired by the LSTM's integration of historical and current features, we adopt a similar gating mechanism to fuse low-frequency and high-frequency features:

$$O = \sigma(\text{Conv}(H, L); \theta_H) \cdot H + \sigma(\text{Conv}(H, L); \theta_L) \cdot L \quad (9)$$

where  $O$  is the merging feature used for decoder,  $\text{Conv}$  is convolution operation,  $\theta_H$  and  $\theta_L$  is trainable parameter for the convolution operation.

**Decoder.** Our decoder obtains the final segmentation by merging the features extracted from both the low-frequency

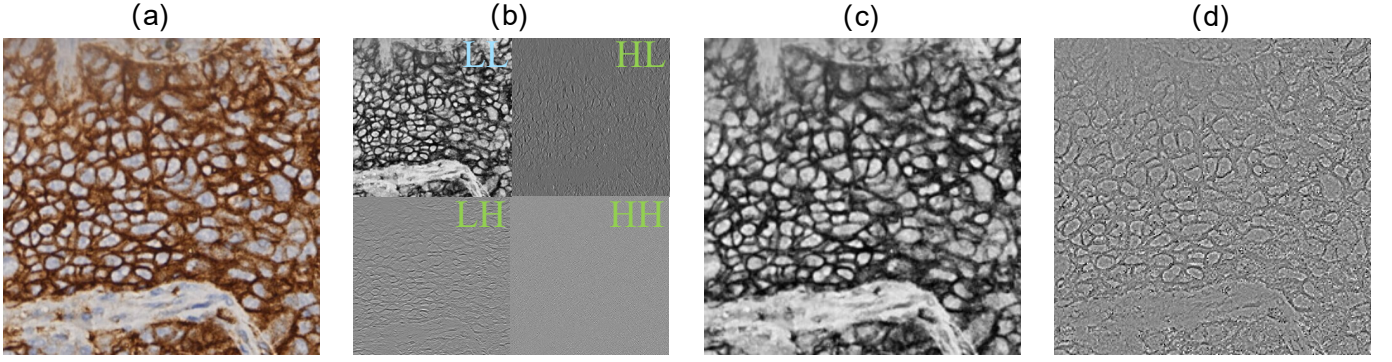


Fig. 2: **Visualization of Wavelet Transform.** (a) Original IHC image showing detailed cell membranes with intense staining; (b) Sub-band images of the wavelet transform: LL (low-frequency component), HL (horizontal high-frequency component), LH (vertical high-frequency component), and HH (diagonal high-frequency component); (c) LL image highlighting overall structural features; (d) Composite H image derived by summing normalized HL, LH, and HH components to emphasize fine details.

and high-frequency components. Following the U-Net framework [41], we concatenate the corresponding features from the encoder to the decoder. However, our method is distinct in that we concatenate both the encoded features from both low-frequency and high-frequency branches. During training, the cross-entropy loss is calculated as follows:

$$\mathcal{L}_{seg} = \mathcal{L}_{cls}(\hat{y}, y), \quad (10)$$

where  $\hat{y}$  represents the predicted logits and  $y$  is the labels from  $\mathbf{GT}_{pseudo}$ . These labels are derived from the first stage: (a) Membrane Curvilinear Image Generation. In this stage, we employ Voronoi diagram algorithm to create membrane structure images, which can be used as the pseudo-ground truth labels ( $\mathbf{GT}_{pseudo}$ ) for supervised training. These pseudo labels simulate real cell membrane structures, providing a basis for supervised training within an unsupervised learning setup.

**Feature alignment.** In our framework, pseudo-pair images are utilized for supervised segmentation training. The source domain samples represent pseudo-images generated by CycleGAN, while the target domain samples are the real cell images. Nevertheless, the domain gap between the generated pseudo-images and the real images remains a challenge. To align the encoded features from the pseudo and real images, we utilize a domain adaptive module equipped with a Gradient Reversal Layer (GRL) [42] at the end of each encoder. Note that we perform this feature alignment for both low-frequency and high-frequency branches to enhance the adaptability across frequency domain.

The GRL is a crucial component for domain adaptation. It works by reversing the gradient during the backpropagation step, effectively promoting the learning of domain-invariant features. During the forward propagation, the GRL acts as an identity function:

$$R_{\lambda}(x) = x, \quad (11)$$

Where  $x$  represents the features extracted from the generated pseudo-images and the real IHC images, which are input to

the GRL. This formula states that during forward propagation, the GRL acts as an identity function, passing the features through unchanged. However, during backpropagation, the GRL multiplies the gradient by a negative scalar  $\lambda$ , which dynamically adjusts during the training process:

$$\frac{dR_{\lambda}}{dx} = -\lambda I, \quad (12)$$

where  $I$  is the identity matrix, specifies that during backpropagation. It reverses the direction of the gradient by multiplying it with  $-\lambda$ . This action is strategically used to reduce the domain discrepancy: it aligns the feature distributions of the different domains by encouraging the model to minimize these discrepancies. Meanwhile, it maximizes the loss for the domain classifier and minimizes the loss for the label predictor.

The objective function for domain adaptation consists of two parts: the label prediction loss for the source domain samples and the domain classification loss for both source and target domain samples. The overall loss function of GRL is formulated as:

$$E(\theta_f, \theta_y, \theta_d) = L_y(G_y(G_f(x_i; \theta_f); \theta_y), y_i) - \lambda \sum_{i=1}^N L_d(G_d(G_f(x_i; \theta_f); \theta_d), y_i), \quad (13)$$

Here,  $L_y$  is the loss for label prediction,  $L_d$  is the loss for domain classification, and  $\theta_f$ ,  $\theta_y$ , and  $\theta_d$  are the parameters for the feature extractor, label predictor, and domain classifier, respectively.

The training objective is to find a saddle point for this function, where  $\theta_f$  represents the parameters of the feature extractor,  $\theta_y$  represents the parameters of the label predictor, and  $\theta_d$  represents the parameters of the domain classifier:

$$(\hat{\theta}_f, \hat{\theta}_y) = \arg \min_{\theta_f, \theta_y} E(\theta_f, \theta_y, \hat{\theta}_d), \quad (14)$$

$$\hat{\theta}_d = \arg \max_{\theta_d} E(\hat{\theta}_f, \hat{\theta}_y, \theta_d). \quad (15)$$

Where  $\hat{\theta}_f$  and  $\hat{\theta}_y$  are the optimal parameters that minimize the overall loss function  $E$  with respect to the feature extractor and label predictor, while  $\hat{\theta}_d$  are the parameters that maximize the domain classification loss. This ensures that the features learned are both discriminative for the main task and invariant to the domain gap. We formulate the binary classification loss as:

$$\mathcal{L}_{align} = \mathcal{L}_{cls}(\hat{y}^H, y^H) + \mathcal{L}_{cls}(\hat{y}^L, y^L), \quad (16)$$

where  $\hat{y}^H$  and  $\hat{y}^L$  represent the predicted logits of the domain classifier for the high-frequency and low-frequency branches, respectively.  $y^H$  and  $y^L$  are the corresponding domain labels.

**Overall training objective.** The overall loss can be formulated as:

$$\mathcal{L} = \mathcal{L}_{seg} + \alpha \mathcal{L}_{align}, \quad (17)$$

where  $\alpha$  is the hyper-parameter to balance the two losses, and we set  $\alpha = 0.5$  in our experiments, with further experiments conducted in Table III to explore the impact of different  $\alpha$  values on segmentation performance.

### III. EXPERIMENTS AND RESULTS

#### A. Dataset

We introduce the IHC Membrane Segmentation (IHC-MS) dataset, which consists of 205 image patches from Immunohistochemistry (IHC) staining. Each image with  $500 \times 500$  resolution was extracted from 11 lung squamous cell carcinoma Whole Slide Images (WSIs) stained for programmed cell death-1 (PD-1) at  $20\times$  optical magnification ( $0.475\mu\text{m}/\text{pixel}$ ). The positive stain membranes in these images were manually annotated by two trained students and verified by two experienced pathologists. As our proposed framework operates in an entirely unsupervised manner, all images in the IHC-MS dataset are designated for use in the test set. The IHC-MS dataset for testing consists of 205 images, with 155 positive and 50 negative samples. Despite the relatively small size, these images are effective for our purposes. We carefully choose 205 test images to represent various staining intensities and characteristics, as shown in Figure 3 with three distinct staining strengths. The dataset will be made publicly available upon acceptance.

#### B. Implementation Details

We generate 5000 pseudo-labeled masks using Voronoi Diagrams and train a CycleGAN using these masks along with an equal number of real IHC image patches. During the training of *YNet*, we also use these 5000 real images for domain adaptation. Both CycleGAN and *YNet* are trained using the AdamW optimizer [43] with a learning rate of  $1e^{-4}$ .

#### C. Results

We benchmark our framework against two traditional image processing-based methods such as Color Deconvolution (CD) [13] and the positive membrane detection module (PMD) introduced in [9], as well as two deep learning-based models, CycleGAN and Pix2Pix. For CycleGAN, we utilize our

trained model to generate segmentation results for the stained membrane images in our test dataset. For Pix2Pix [44], we train the model using our pseudo-paired dataset.

We report the segmentation results using the Dice Coefficient [45], Hausdorff Distance [46] and Accuracy in Table I. Our model achieved a Dice Coefficient of 52.15%, outperforming both traditional methods (CD at 43.40% and PMD at 48.09%) and generative adversarial network-based approaches (CycleGAN at 43.58% and Pix2Pix at 42.80%). Moreover, our approach significantly reduced the Hausdorff Distance to 46.69, suggesting that it not only accurately segments the membrane but also ensures a closer alignment to the true membrane boundaries compared to other methods. Our framework also excelled in overall accuracy, reaching 81.42%. Notably, the comparative methods require threshold settings for their segmentation process. To ensure fairness and optimize their performance, we conduct multiple rounds of testing, adjusting the thresholds accordingly. The results presented in Table I represent the best outcomes obtained for each method during our comprehensive assessment. It can be observed that our framework surpasses all other methods by a considerable margin on all evaluation metrics. The PMD achieves a higher Dice score compared to CycleGAN and Pix2Pix, but it exhibits notably poorer performance on the Hausdorff distance metric. Qualitative results in Figure 3 show that deep learning-based algorithms demonstrate more accurate membrane segmentation, and our approach exhibits superior performance in both positive and negative areas.

#### D. Ablation Study

In comparison to CycleGAN, we have validated the indispensability of the individual segmentation module. We further conduct an ablation study to underscore the necessity of the two modules within *YNet*, i.e., the wavelet transform and the domain adaptive module. As shown in Table II, adding the domain adaptive module improves the performance of all models, demonstrating its effectiveness. While utilizing either high-frequency or low-frequency input alone results in inferior performance compared to the original RGB image with a single encoder, the combination of both inputs through the Y structure yields superior results compared to the single-path model. Additionally, we explore the impact of the hyper-parameter  $\alpha$  as detailed in Table III, which balances the segmentation loss with the domain adaptation loss, on the overall performance of our model. We discover that setting  $\alpha = 0.5$  is optimal, yielding the highest performance in all metrics.

### IV. CONCLUSIONS

In this paper, we study the unsupervised cell membrane segmentation on IHC stained histopathology images. Our proposed framework tackles the unsupervised challenge by transforming it into a supervised task through the generation of Pseudo-paired training data, followed by the implementation of domain adaptive *YNet* to mitigate artificial perturbations inherent in the generated data. Our method demonstrates promis-



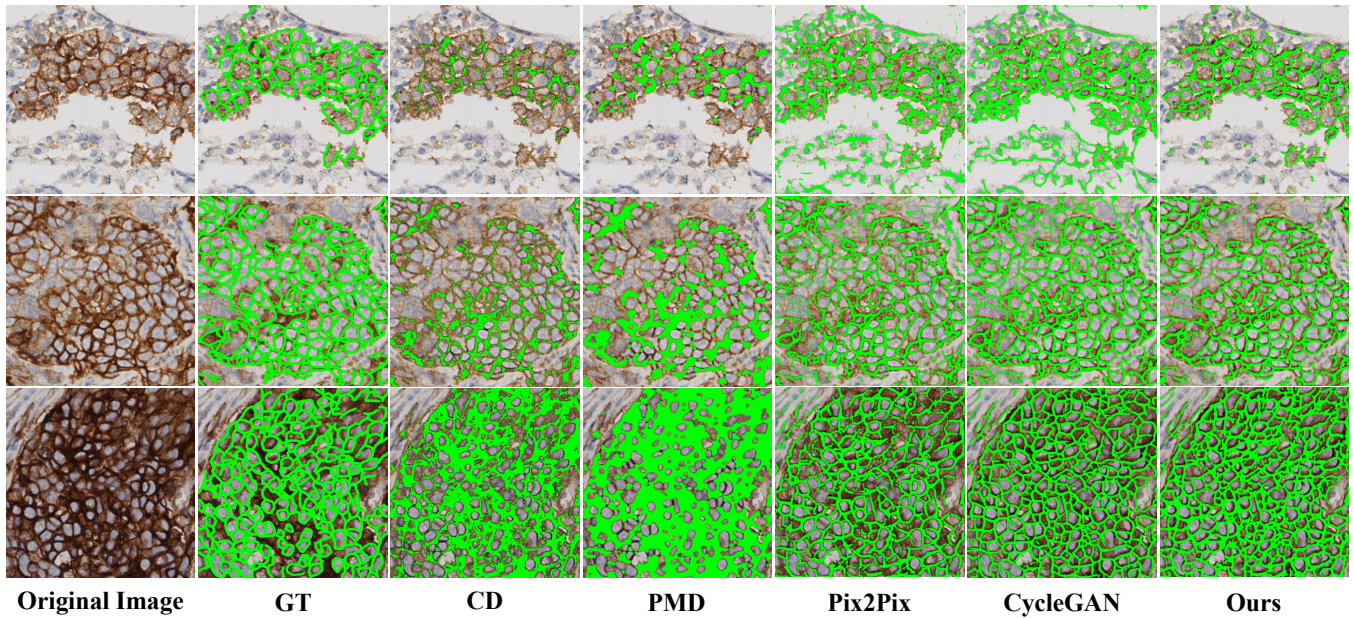


Fig. 3: Qualitative results of cell membrane segmentation on three different stain intensity images.

TABLE I: Comparison with state-of-the-art methods for unsupervised positive membrane segmentation on the new dataset IHC-MS.

	Method	DICE (%) $\uparrow$	Hausdorff Distance $\downarrow$	Accuracy (%) $\uparrow$
Traditional	CD [13]	43.40	61.42	80.63
	PMD [9]	48.09	62.02	80.94
Generative Adversarial	Pix2Pix [44]	42.80	54.35	74.38
	CycleGAN [36]	43.58	59.87	74.16
<b>Unsupervised</b>	<b>Ours</b>	<b>52.15</b>	<b>46.69</b>	<b>81.42</b>

TABLE II: Ablation on the effectiveness of various components, including LF images, HF images and domain adaptive module.

Raw	LF	HF	GRL	DICE (%) $\uparrow$	Hausdorff Distance $\downarrow$
✓				47.57	57.51
✓			✓	48.87	53.08
	✓			49.37	49.06
	✓		✓	51.49	49.00
		✓		43.53	116.96
		✓	✓	27.35	98.68
	✓	✓		51.10	55.07
	✓	✓	✓	<b>52.15</b>	<b>46.69</b>

TABLE III: Analysis of the Hyper-Parameter  $\alpha$  on Segmentation Performance Metrics

Hyper-Parameter $\alpha$	DICE (%) $\uparrow$	Hausdorff Distance $\downarrow$	Accuracy (%) $\uparrow$
0	51.10	55.07	80.87
0.1	46.25	58.32	80.19
<b>0.5</b>	<b>52.15</b>	<b>46.69</b>	<b>81.42</b>
1	51.95	49.49	81.15

ing results compared to existing unsupervised approaches. Our method offers a cost-effective, unsupervised solution, robust

across various staining, tissues, and noise levels. It ensures consistent performance under diverse imaging conditions (e.g., varying staining intensities, tissue types, and noise levels) and can be extended to other applications like Corneal Endothelial Cell Segmentation by changing the pseudo-paired data. We hope that our approach will contribute to advancements in unsupervised semantic segmentation and have the potential to improve the accuracy of clinical diagnoses in real-world scenarios.

## REFERENCES

- [1] S. A. Pileri, G. Roncador, C. Ceccarelli, M. Piccioli, A. Briskomatis, E. Sabattini, S. Ascani, D. Santini, P. P. Piccaluga, O. Leone *et al.*, “Antigen retrieval techniques in immunohistochemistry: comparison of different methods,” *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, vol. 183, no. 1, pp. 116–123, 1997.
- [2] M. S. Tuffaha, H. Guski, G. Kristiansen, M. S. Tuffaha, H. Guski, and G. Kristiansen, *Immunohistochemistry in tumor diagnostics*. Springer, 2018.
- [3] I. Kim, K. Kang, Y. Song, and T.-J. Kim, “Application of artificial intelligence in pathology: Trends and challenges,” *Diagnostics*, vol. 12, no. 11, p. 2794, 2022.
- [4] K. You, S. Lee, K. Jo, E. Park, T. Kooi, and H. Nam, “Intra-class contrastive learning improves computer aided diagnosis of breast cancer in mammography,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 55–64.

- [5] X. Feng, Q. Shen, C. Li, X. Wang, N. Xie, L. Xie, Y. Fang, and Z. Wu, "A privacy preserving computer-aided medical diagnosis framework with outsourced model," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 1080–1087.
- [6] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized medical imaging and graphics*, vol. 31, no. 4–5, pp. 198–211, 2007.
- [7] X. Zhang, S. Qiu, M. Geng, and H. He, "Enhancing detection of ssveps for high-speed brain-computer interface with a siamese architecture," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021, pp. 1623–1627.
- [8] F. D. Khameneh, S. Razavi, and M. Kamasak, "Automated segmentation of cell membranes to evaluate her2 status in whole slide images using a modified deep learning network," *Computers in biology and medicine*, vol. 110, pp. 164–174, 2019.
- [9] J. Liu, Q. Zheng, X. Mu, Y. Zuo, B. Xu, Y. Jin, Y. Wang, H. Tian, Y. Yang, Q. Xue *et al.*, "Automated tumor proportion score analysis for pd-11 (22c3) expression in lung squamous cell carcinoma," *Scientific Reports*, vol. 11, no. 1, p. 15907, 2021.
- [10] S. Tewary, I. Arun, R. Ahmed, S. Chatterjee, and S. Mukhopadhyay, "Autoihc-analyzer: computer-assisted microscopy for automated membrane extraction/scoring in her2 molecular markers," *Journal of Microscopy*, vol. 281, no. 1, pp. 87–96, 2021.
- [11] A. Brüggmann, M. Eld, G. Lelkaitis, S. Nielsen, M. Grunkin, J. D. Hansen, N. T. Foged, and M. Vyberg, "Digital image analysis of membrane connectivity is a robust measure of her2 immunostains," *Breast cancer research and treatment*, vol. 132, pp. 41–49, 2012.
- [12] Y.-L. Kuo, C.-C. Ko, and J.-Y. Lai, "Automated assessment in her2-neu immunohistochemical expression of breast cancer," in *2010 International Symposium on Computer, Communication, Control and Automation (3CA)*, vol. 2. IEEE, 2010, pp. 585–588.
- [13] A. C. Ruifrok, D. A. Johnston *et al.*, "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology*, vol. 23, no. 4, pp. 291–299, 2001.
- [14] E. Ficarra, S. Di Cataldo, A. Acquaviva, and E. Macii, "Automated segmentation of cells with ihc membrane staining," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1421–1429, 2011.
- [15] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, and L. Van Gool, "Unsupervised semantic segmentation by contrasting object mask proposals," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10052–10062.
- [16] M. Hamilton, Z. Zhang, B. Hariharan, N. Snavely, and W. T. Freeman, "Unsupervised semantic segmentation by distilling feature correspondences," *arXiv preprint arXiv:2203.08414*, 2022.
- [17] K. Do, T. Tran, and S. Venkatesh, "Clustering by maximizing mutual information across views," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 9928–9938.
- [18] Y. Li, P. Hu, Z. Liu, D. Peng, J. T. Zhou, and X. Peng, "Contrastive clustering," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 10, 2021, pp. 8547–8555.
- [19] R. Abdal, P. Zhu, N. J. Mitra, and P. Wonka, "Labels4free: Unsupervised segmentation using stylegan," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13970–13979.
- [20] M. Chen, T. Artières, and L. Denoyer, "Unsupervised object segmentation by redrawing," *Advances in neural information processing systems*, vol. 32, 2019.
- [21] X. He, B. Wandt, and H. Rhodin, "Ganseg: Learning to segment by unsupervised hierarchical image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1225–1235.
- [22] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8110–8119.
- [23] A. F. F. WJN, L. Vinken Koen, and A. Viergever Max, "Multiscale vessel enhancement filtering," 1998.
- [24] M. Alhussein, K. Aurangzeb, and S. I. Haider, "An unsupervised retinal vessel segmentation using hessian and intensity based approach," *IEEE Access*, vol. 8, pp. 165 056–165 070, 2020.
- [25] N. Memari, A. R. Ramli, M. I. B. Saripan, S. Mashohor, and M. Moghbel, "Retinal blood vessel segmentation by using matched filtering and fuzzy c-means clustering with integrated level set method for diabetic retinopathy assessment," *Journal of Medical and Biological Engineering*, vol. 39, pp. 713–731, 2019.
- [26] J. Wang and A. C. Chung, "Higher-order flux with spherical harmonics transform for vascular analysis," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23*. Springer, 2020, pp. 55–65.
- [27] M. W. Law and A. C. Chung, "Three dimensional curvilinear structure detection using optimally oriented flux," in *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12–18, 2008, Proceedings, Part IV 10*. Springer, 2008, pp. 368–382.
- [28] T. Shi, X. Ding, L. Zhang, and X. Yang, "Freecoc: self-supervised learning from fractals and unlabeled images for curvilinear object segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 876–886.
- [29] J. Mishra and S. Mishra, *L-system Fractals*. Elsevier, 2007.
- [30] F. D. Khameneh, S. Razavi, and M. Kamasak, "Automated segmentation of cell membranes to evaluate her2 status in whole slide images using a modified deep learning network," *Computers in biology and medicine*, vol. 110, pp. 164–174, 2019.
- [31] M. Saha and C. Chakraborty, "Her2net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2189–2200, 2018.
- [32] U. Gaur, M. Kourakis, E. Newman-Smith, W. Smith, and B. Manjunath, "Membrane segmentation via active learning with deep networks," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 1943–1947.
- [33] M. Saha and C. Chakraborty, "Her2net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2189–2200, 2018.
- [34] S. Dimopoulos, C. E. Mayer, F. Rudolf, and J. Stelling, "Accurate cell segmentation in microscopy images using membrane patterns," *Bioinformatics*, vol. 30, no. 18, pp. 2644–2651, 2014.
- [35] R. Shi, K. Bi, K. Du, L. Ma, F. Fang, L. Duan, T. Jiang, and T. Huang, "Ps-net: human perception-guided segmentation network for em cell membrane," *Bioinformatics*, vol. 39, no. 8, p. btad464, 2023.
- [36] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [37] Y. Zhou, J. Huang, C. Wang, L. Song, and G. Yang, "Xnet: Wavelet-based low and high frequency fusion networks for fully- and semi-supervised semantic segmentation of biomedical images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 21 085–21 096.
- [38] F. Aurenhammer and R. Klein, "Voronoi diagrams," *Handbook of computational geometry*, vol. 5, no. 10, pp. 201–290, 2000.
- [39] D. Zhang and D. Zhang, "Wavelet transform," *Fundamentals of image data mining: Analysis, Features, Classification and Retrieval*, pp. 35–44, 2019.
- [40] I. N. Sneddon, *Fourier transforms*. Courier Corporation, 1995.
- [41] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [42] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by back-propagation," in *International conference on machine learning*. PMLR, 2015, pp. 1180–1189.
- [43] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
- [44] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [45] S. Jha, R. Kumar, I. Priyadarshini, F. Smarandache, H. V. Long *et al.*, "Neutrosophic image segmentation with dice coefficients," *Measurement*, vol. 134, pp. 762–772, 2019.
- [46] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15, no. 9, pp. 850–863, 1993.