Mutual-Complementing Framework for Nuclei Detection and Segmentation in Pathology Image

Zunlei Feng[1,5,6,#], Zhonghua Wang[1,#], Xinchao Wang[2], Yining Mao[1], Thomas Li[3], Jie Lei[4], Yuexuan Wang[1], Mingli Song[1,5,6*]
[1]Zhejiang University, [2]National University of Singapore, [3]The University of Hong Kong,
[4]Zhejiang University of Technology, [5]Zhejiang Provincial Key Laboratory of Service Robot,
[6]Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies

Abstract

Detection and segmentation of nuclei are fundamental analysis operations in pathology images, the assessments derived from which serve as the gold standard for cancer diagnosis. Manual segmenting nuclei is expensive and time-consuming. What's more, accurate segmentation detection of nuclei can be challenging due to the large appearance variation, conjoined and overlapping nuclei, and serious degeneration of histological structures. Supervised methods highly rely on massive annotated samples. The existing two unsupervised methods are prone to failure on degenerated samples. This paper proposes a Mutual-Complementing Framework (MCF) for nuclei detection and segmentation in pathology images. Two branches of MCF are trained in the mutual-complementing manner, where the detection branch complements the pseudo mask of the segmentation branch, while the progressive trained segmentation branch complements the missing nucleus templates through calculating the mask residual between the predicted mask and detected result. In the detection branch, two response map fusion strategies and gradient direction based postprocessing are devised to obtain the optimal detection response. Furthermore, the confidence loss combined with the synthetic samples and self-finetuning is adopted to train the segmentation network with only high confidence areas. Extensive experiments demonstrate that MCF achieves comparable performance with only a few nucleus patches as supervision. Especially, MCF possesses good robustness (only dropping by about 6%) on degenerated samples, which are critical and common cases in clinical diagnosis.

1. Introduction

With the advent of whole slide imaging scanners, the histopathological analysis performed on these digital pathology images has been demonstrated as an effective and reliable tool for cancer diagnosis and prognosis [9]. The assessments derived from the pathology image serve as the gold standard for cancer diagnosis in many clinical protocols. The shape and distribution of cell nuclei in pathology images are used to determine cell, tissue, and cancer types, and are critical in cancer identification, grading, and prognosis [9]. Accurate detection and segmentation of these nuclei from histology images is an essential prerequisite to obtain reliable morphological statistics for quantitative diagnosis, prognosis, and analysis.

However, manual searching and segmenting nuclei from large-scale histopathology images in a conventional way can be expensive, error-prone, and time-consuming. Furthermore, it often suffers from a high inter and intra-observer variability [4, 15], which results in limited reproducibility. Therefore, automatic detection and segmentation methods are highly demanded in clinical practice to improve the efficiency, reliability, and scalability of large-scale histopathological image analysis.

Automated detection and segmentation of histological nuclei from histology images can be quite challenging for several reasons. First, there is a huge variation of object appearance depending on the histologic grade as well as the type of the disease. Second, there exist conjoined and overlapping nuclei. Third, the structures of nuclei have seriously degenerated in the malignant cases£¬ which lead to weak boundaries. In addition, the variation of tissue preparation procedures such as sectioning and staining can cause deformation, artifacts and inconsistency of tissue appearance, which can impede the segmentation process as well.

Recently, deep learning approaches have achieved promising results in the medical image analysis area [16]. Some CNN based method has been applied into nuclei detection [3, 10, 13, 24, 31] and segmentation [1, 3, 6, 16, 20, 33, 21, 22, 25, 29, 38] . However, those supervised methods highly rely on a large amount of annotation. Careful annota-

*Corresponding author. #Equal contribution to this work.

tion of nucleus boundaries is time-consuming, error-prone, and may also suffer from subjective interpretation errors. As shown in [4, 15], there is a larger interoberver disparity among pathologists identifying nuclei in pathology images.

To relieve the rely on the massive annotated samples, some weakly supervised methods [18, 31] are proposed. However, the satisfactory performance of those methods still requires plenty of annotated samples. Furthermore, Le *et al.* [13] proposed a crosswise sparse convolutional autoencoder (CAE) for unsupervised nucleus detection based on the local sparsity assumption. What's more, Hou *et al.* [12] synthesized training samples image patches with some predefined nuclei texture and color, refined those samples with GAN, finally trained a segmentation CNN with those generated sample pairs. However, the above two unsupervised methods [12, 13] usually failed on complex and degenerated cases caused by canceration, which are critical and common cases in clinical diagnosis.

In this paper, we proposed a Mutual-Complementing Framework (MCF) for nuclei detection and segmentation in pathology images. MCF can achieve good detection and segmentation results with only a few nucleus patches. As described above, different pathology images have variational appearances. However, those nuclei can also be distributed into some similar templates according to structure and color. So, by initially selecting some general nucleus patches, a detection branch with the corresponding correlation filters can locate most nuclei. For obtaining the optimal response, Cropped Filter Fusion (CFF) and Multi-scale Filter Fusion (MFF) are devised to fuse top T high responses, which can avoid disturbance of some unmatched correlation filters. Gradient direction based postprocessing is also proposed to eliminate some incorrect responses, such as connected regions of nuclei. With a dynamic binarization technique, those detected areas with high confidence can serve as the pseudo mask of the segmentation branch. For the segmentation branch, a confidence loss is devised to train the segmentation network with only high confidence areas in the pseudo mask. What's more, we synthesize some samples with high confidence areas to enhance the segmentation performance of the nucleus boundary. Finally, the trained segmentation network is finetuned with the high confidence areas predicted by itself.

For the same input image, the mask residual between the detected and segmented results is calculated for finding the missing nuclei, which can be used to optimize the detection branch. Based on the mutual-complementing mechanism, the detection and segmentation branches are trained iteratively. Extensive experiments show that MCF achieves promising results with only a few nuclei patch templates, and possesses good robustness on degenerated samples.

Our main contribution can be summarized as follows: 1) We propose a Mutual-Complementing Framework (MCF) for nuclei detection and segmentation, where the mutual-complementing mechanism enables MCF to achieve iteratively optimized segmentation and detection results with only a few nucleus patches. 2) Two response map fusion strategies and gradient direction postprocessing are devised to obtain the optimal responses and eliminate some incorrect responses. 3) The confidence loss combined with synthetic samples and self-finetuning is designed to train the segmentation network with only high confidence areas to improve the overall segmentation performance. 4) Extensive experiments show that, with only a few nucleus patches, MCF achieves comparable detection and segmentation performance on par with fully supervised methods, and possesses excellent robustness on degenerated samples.

1.1. Related Works

In the field of nuclei segmentation and detection, some hand-crafted feature based methods can be found in [1, 30]. Recently, deep learning approaches have achieved promising results in nuclei segmentation and detection, among many other related domains [14, 32, 34, 35, 36, 37]. Here, we give a brief survey of recent progress in deep learning based nuclei segmentation and detection. Another related technique is the correlation filter, which has been widely used in the object tracking area [5, 11, 28]. To our knowledge, there is only one work [2] that applies the standard correlation filters for nuclei detection.

For the fully supervised methods, Zhou *et al.* [38] proposed a sparsity constrained convolutional regression network that addresses the nuclei segmentation as a pixel classification problem. Naylor *et al.* [20] employed FCN to discriminate the nuclei from the background and then applies the watershed method to split the nuclei. The classic Unet is also modified for the nuclei segmentation methods [1, 6, 21, 25, 26]. The difference is that annotated boundary and inside masks are adopted to supervise the training of the segmentation network in [1, 6, 16]. Xing *et al.* [29] proposed a CNN for generating initial segmentation results, and thereafter alternately performed bottom-up shape deformation and top-down shape inference to achieve correct nucleus segmentation. Naylor *et al.* [19] formulated nuclei segmentation as a regression task of the distance map with a fully convolutional network. Sirinukunwattana *et al.* [24] presented a spatially constrained CNN sensitive to the local neighborhood for nucleus detection and classification in routine stained histology images of colorectal adenocarcinomas. The performance of those fully supervised methods highly relies on a large amount of annotated samples.

To alleviate the requirement of massive training samples and annotations, Xu *et al.* [31] first pretrained a Stacked Sparse Autoencoder (SSAE) with original image reconstruction, then finetuned the SSAE through classifying each nuclei patch for automated nuclei detection on breast cancer

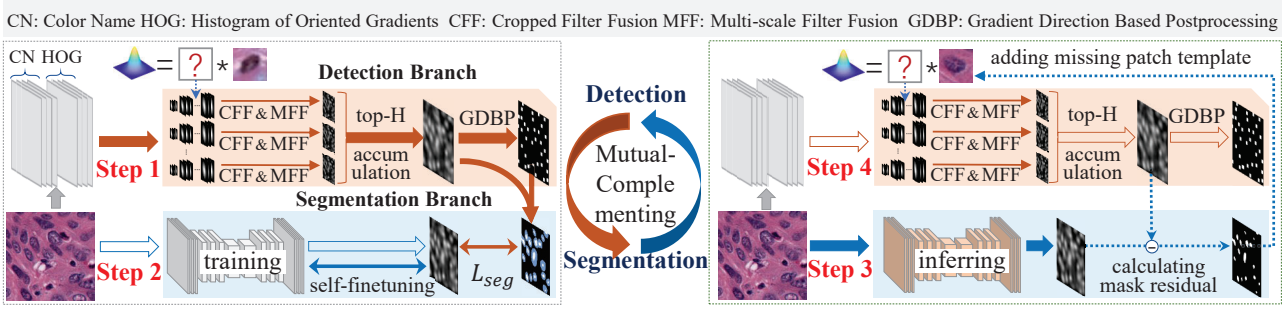

Figure 1. The flow diagram of the Mutual-Complementing Framework (MCF), which is composed of detection and segmentation branches. With several correlation filters as initial parameters, the detection branch predicts the nuclei response maps (Step 1). Then, with the detection response map as the pseudo mask, the confidence loss L_{seg} that only calculates the high response area is devised for supervising the training of the segmentation network (Step 2). Next, the mask residual is calculated between the predicted segmentation mask and the detection response map (Step 3). The missing nucleus patch templates derived from the mask residual are added to the detection branch (Step 4). The two branches of MCF are trained in a mutual-complementing manner, where the detection branch complements the pseudo mask of the segmentation branch, while the progressive trained segmentation network complements the missing nucleus patch templates.

histopathology. Mahmood *et al.* [18] adopted the CycleGAN [39] to synthesize some pathology images and corresponding masks, then trained a condition GAN with the synthetic samples and original annotated samples for nuclei segmentation. Qu *et al.* proposed a weakly-supervised method for nuclei detection and segmentation with only partial points annotations.

Furthermore, Le *et al.* [13] proposed a crosswise sparse convolutional autoencoder (CAE) for unsupervised nucleus detection and feature extraction simultaneously. CAE is based on the local sparsity assumption, which will fail on degenerated cases caused by cancerization. Some researchers [8, 12, 17] adopted GAN to synthesize histopathology images with some random masks for training the nuclei segmentation network in an unsupervised manner. However, [8, 12, 17] inherited the drawbacks of GAN [23] (unstable and time-consuming training). The predefined nuclei texture and color usually lead to failures in hard and degenerated samples.

2. Mutual-Complementing Framework

This paper proposes a Mutual-Complementing Framework (MCF) for nuclei detection and segmentation, which combines the advantage of correlation filter and deep learning. The proposed MCF is composed of a multiple correlation filter fusion detection branch and a CNN based segmentation network, which is shown in Fig. 1. Two branches of MCF work in a mutual-complementing manner. The detection branch supplements the pseudo mask of the segmentation network, while the segmentation network supplements the missing nuclei patch templates. The proposed MCF can achieve promising detection and segmentation results with only a few nuclei patch templates through iterative training.

2.1. Multiple Correlation Filter Fusion Detection

The correlation filter has been successfully applied in the object tracking task [5], where the correlation between the correlation filter and the corresponding input image patch will have the highest response. Most nuclei have similar shapes and appearance in pathology images, indicating that an appropriate correlation filter can detect most nuclei in the pathology image. A bank of correlation filters that correspond to diverse nuclei patches can detect all the nuclei in the pathology image. For an input nucleus patch I_p, the optimal correlation filter w_i can maximize the response $v = I_{patch} \odot w^i$, where $\otimes$ denotes the correlation operation. Following [11], when dealing with the image patch I_{patch}, the correlation filter w^1 in the Fourier domain can be calculated efficiently with cyclic shift and convolution theorem as follows:

$$\hat{w}^1 = \frac{\hat{\mathbf{x}}^* \odot \hat{\mathbf{y}}}{\hat{\mathbf{x}}^* \odot \hat{\mathbf{x}} + \lambda}, \tag{1}$$

where $\hat{\mathbf{x}} = DFT(\mathbf{x})$, DFT denotes discrete Fourier transform, $\mathbf{x}$ denotes the cyclic shift feature vector extracted from the image patch I^p, $\hat{}$ is a shorthand for the DFT of a vector, λ is a regularization parameter. The extracted features x contain two kinds of features: Color Name (CN) features [27] and Histogram of Oriented Gradients (HOG) [7] features, which are two main and frequently-used features in the object tracking area.

There are various shapes, scales, and crowded nuclei in the pathology images, which pose challenges for accurate nuclei detection. For the various shape nuclei, we adopt a bank of representative nucleus templates to detect as many nuclei as possible. For the different scale and crowded nuclei, we devise the correlation response map fusion strategy (Multi-scale Filter Fusion and Cropped Filter Fusion) to obtain optimal responses. In Multi-scale Filter Fusion (MFF),

each nucleus patch template is resized into different scale then the corresponding correlation filters are calculated with Eqn.(1), which can handle the scale variation. Next, in Cropped Filter Fusion (CFF), multiple center patches with different sizes are cropped from a correlation filter, which takes advantage of the color features to detect crowded nuclei. After MFF and CFF, the gradient direction based post-processing is proposed to solve the ambiguous boundary of crowded nuclei.

2.1.1 Correlation Response Map Fusion Strategy

For an input image I, a response map I_m can be calculated by correlating with a correlation filter in a sliding manner. For handling different sizes of nuclei, a nucleus patch I_{patch} will be resized into K scales. For the n-th nucleus patch, K scale correlation filters $\{w_1^n, w_2^n, ..., w_k^n, ..., w_K^n\}$ can be calculated with Eqn.(1). Then, for handling crowd nuclei, U center patches with different sizes are cropped from each correlation filter w_k^n. For the n-th nucleus patch, we finally get $K * U$ correlation filters $\{(w_{1,1}^n, w_{1,2}^n, ..., w_{1,U}^n), (w_{2,1}^n, w_{2,2}^n, ..., w_{2,U}^n), ..., (w_{K,1}^n, w_{K,2}^n, ..., w_{K,U}^n)\}$. The response map $S_{K,U}^n$ for the input image I can be calculated with the correlation filter $w_{t,u}^n$ using $S_{K,U}^n = w_{K,U}^n(I)$, where $w_{K,U}^n(\cdot)$ denotes the correlation operation with the correlation filter $w_{K,U}^n$ in a sliding manner.

The fusion of multiple response maps $\{S_{K,U}^n\}$ is a critical step. The optimal response map S for each nucleus in I should reflect the response of the most matched correlation filter. In MCF, we first fuse the response maps of cropped filters of the same scale, then fuse the multi-scale response maps, and finally fuse those processed response maps of different correlation filters.

CFF & MFF. With response maps $\{(S_{1,1}^n, S_{1,2}^n, ..., S_{1,U}^n), ..., (S_{K,1}^n, S_{K,2}^n, .., S_{K,U}^n)\}$, MFF map S^n and CFF map S_k^n are calculated as follows:

$$S^n(i,j) = \mathbb{A}^T\{G_k \odot S_k^n[i,j], .., G_k \odot S_k^n[i,j]\}, \quad (2)$$
$$S_k^n(i,j) = \mathbb{A}^T\{G_{K,1} \odot S_{K,1}^n[i,j], .., G_{K,U} \odot S_{K,U}^n[i,j]\},$$

where $\mathbb{A}^T\{\}$ denotes accumulating the top T values, $S_{K,U}^n[i,j]$ and $S_k^n[i,j]$ denote the patch area of response map $S_{K,U}^n$ and CFF map S_k^n centered at position (i,j), respectively. $S^n(i,j)$ and $S_k^n(i,j)$ denote the response value of MFF map S^n and CFF map S_k^n centered at position (i,j), respectively. G_k is the discrete weight template for the k-th scale filter. The sum value of G_k equal to zero, the shape of the template G_k is same as the 2D Gaussian shape, and the size of G_k is same with the k-th scale filter. $G_{K,U}$ denotes the u-th cropped patch area from G_k. The size of $S_{K,U}^n[i,j]$ and $S_k^n[i,j]$ is the same as the size of the u-th cropped filter and the k-th scale filter, respectively.

Futhermore, the final response map S is calculated by accumulating the top H values as follows:

$$S(i,j) = \mathbb{A}^H\{S^1(i,j), S^2(i,j), , .., S^N(i,j)\}. \quad (3)$$

For each image I, the final response map S will be normalized into $[0,1]$ as follows $\tilde{S} = norm(S)$.

2.1.2 Gradient Direction Based Postprocessing

The conjoined and overlapping nuclei in pathology images are the common situation, which will lead to adhesive high response values. In the response map, nuclei centers have high responses, around areas of nuclei have median responses, and the background areas have low responses. So, we put forward a gradient direction based postprocessing to get the accurate nuclei position. The gradient direction based postprocessing is based on the fact that nuclei centers have positive response gradients in all directions. For each position (i,j), we calculate the response gradient d_{ij}^k in eight directions (left, right, up, down, left-up, left-down, right-up and right-down), which is formulated as follows:

$$D(i,j) = \begin{cases} \sum_{r=1}^{8} d_{ij}^r, & if\ d_{ij}^r > 0,\ r \in \{1,...,8\}, \\ 0, & else \end{cases}$$
$$d_{ij}^r = \sum_{p=1}^{P}\sum_{q=1}^{Q}\{\tilde{S}(\bar{i},\bar{j}) - \tilde{S}(\bar{\bar{i}},\bar{\bar{j}})\},\ s.t.\ p < q, \quad (4)$$
$$when\ r = 1, \bar{i} = i, \bar{j} = j + p, \bar{\bar{i}} = i, \bar{\bar{j}} = j + q;$$
$$when\ r = 2, \bar{i} = i, \bar{j} = j - p, \bar{\bar{i}} = i, \bar{\bar{j}} = j - q;$$
$$when\ r = 3, \bar{i} = i + p, \bar{j} = j, \bar{\bar{i}} = i + q, \bar{\bar{j}} = j;$$
$$when\ r = 4, \bar{i} = i - p, \bar{j} = j, \bar{\bar{i}} = i - p, \bar{\bar{j}} = j;$$
$$when\ r = 5, \bar{i} = i + p, \bar{j} = j + p, \bar{\bar{i}} = i + q, \bar{\bar{j}} = j + q;$$
$$when\ r = 6, \bar{i} = i - p, \bar{j} = j - p, \bar{\bar{i}} = i - q, \bar{\bar{j}} = j - q;$$
$$when\ r = 7, \bar{i} = i + p, \bar{j} = j - p, \bar{\bar{i}} = i + q, \bar{\bar{j}} = j - q;$$
$$when\ r = 8, \bar{i} = i - p, \bar{j} = j + p, \bar{\bar{i}} = i - q, \bar{\bar{j}} = j + q;$$

where P denotes the radius of the nucleus center, Q denotes the radius of the nucleus area, P is small than Q. The high response of the accumulative gradient map D denotes the position of nucleus center.

The big and small values of S indicate the high confidence position for nucleus and background, respectively. Those high confidence areas combined with the accumulative gradient map D can serve as the pseudo mask for training the segmentation network.

2.2. Confidence Segmentation Network

2.2.1 Confidence Segmentation Loss

The core part of the segmentation network is how to train the segmentation network with pseudo masks generated

with the above detection branch. In MCF, the high confidence areas of the response map are used for training the segmentation network. For the input image I, the confidence segmentation loss $\mathcal{L}_{\mathbf{seg}}$ is defined as follows:

$$\mathcal{L}_{\mathbf{seg}} = I_{bg}^m * \log(\mathbf{p_0}) + I_{nc}^m * \log(\mathbf{p_1}), \tag{5}$$

$$\mathbf{p}_c(i,j) = exp(\mathbf{a}_c(i,j)) / \sum_{c=0}^{1} exp(\mathbf{a}_c(i,j)), c = 0 \; or \; 1,$$

$$I_{bg}^m = \mathbb{T}(\tilde{S} < t_1), I_{nc}^m = \mathbb{DLT}(\tilde{S}[D(i,j)] > t_{D(i,j)}),$$

where I_{bg}^m and I_{nc}^m denote the confident mask of nuclei and background, $\mathbf{a}_c(i,j)$ denotes the activation value in channel c at the pixel position (i,j), $\mathbf{p_1}$ and $\mathbf{p_0}$ denote the predicted probabilities of true label foreground nuclei and background, $\mathbb{T}(\tilde{S} < t_1)$ denotes binarizing the final response map $\tilde{S}$ with the thresholding t_1. $\mathbb{DLT}(\tilde{S}[D(i,j)] > t_{D(i,j)})$ denotes the dynamic local binarization. For each nucleus center position (i,j) indicated by the accumulative gradient map D, each local patch $\tilde{S}[D(i,j)]$ of response map $\tilde{S}$ is binarized with the dynamic thresholding $t_{D(i,j)} = 0.8 * \tilde{S}(i,j)$, where $\tilde{S}(i,j)$ is the response value at position (i,j). The size of the local patch $\tilde{S}[D(i,j)]$ is the same as the size of the middle-size correlation filter.

With the confidence segmentation loss L_{seg}, the segmentation network will learn the distinguishing ability for features of nuclei and background, which will progressively predict the category of uncertain areas.

2.2.2 Sample Synthesis & Self-finetuning

The segmentation network trained with pseudo masks provided by the detection branch has two flaws: 1) the ambiguous boundary areas of nuclei are not involved in the training of segmentation network, which leads to poor performance on ambiguous boundary area; 2) some high confidence area predicted by trained segmentation networks not detected by the detection branch does not involve the training stage of segmentation network. For the former flaw, we enlarge the confident nuclei areas in I_{nc}^m and paste those areas on the pure background image synthesized with the confident background area. Then, the segmentation network is trained with the standard Cross-Entropy (CE) loss function on those synthetic samples.

For the latter flaw, the trained segmentation network predicts the foreground probability map $\overline{\mathbf{p_1}}$ and background probability $\overline{\mathbf{p_0}}$, which contain some high confidence areas not detected by the detection branch. Then, the segmentation network is finetuned with the predicted high confidence foreground mask $\overline{I_{nc}^m} = \mathbb{T}(\overline{\mathbf{p_1}} > t2)$ and background mask $\overline{I_{bg}^m} = \mathbb{T}(\overline{\mathbf{p_0}} < t3)$ using

$$\mathcal{L}_{\mathbf{self}} = \overline{I_{bg}^m} * \log(\mathbf{p_0}) + \overline{I_{nc}^m} * \log(\mathbf{p_1}). \tag{6}$$

2.2.3 Mask Residual Calculation

With the nuclei probability map $\overline{\overline{\mathbf{p_1}}}$ inferred by the self-finetuned segmentation network, the mask residual I_{rs}^m is calculated for finding the missing nuclei patches as follows:

$$I_{rs}^m = |\mathbb{T}(\overline{\overline{\mathbf{p_1}}} > t4) - I_{nc}^m|, \tag{7}$$

where $|\cdot|$ denotes the absolute value function. Large contiguous nonzero areas in I_{rs}^m indicate the missing nuclei, which can be used as the nucleus templates for the detection branch.

2.3. Iterative Optimization

The multiple correlation filter fusion detection branch and confidence segmentation branch are optimized iteratively. The detection branch complements the pseudo mask for the segmentation network, which is trained with the confidence segmentation loss L_{seg} in Eqn.(5), CE loss on synthetic samples, and self-finetuning segmentation loss L_{self} in Eqn.(6). Meanwhile, the trained segmentation network complements the missing nuclei patches with the mask residual I_{rs}^m in Eqn.(7). After several rounds of iterative optimization, MCF converges to the optimal stage for detection and segmentation.

3. Experiment

Dataset. The datasets we adopted contain TNBC [19], MICCAI18 [16], and NuCLS [4]. The MICCAI18 [16] contains 30 training and 14 testing tissue images extracted from whole slide images of two cancer types. The TNBC [19] contains 50 samples with a total of 4022 annotated nuclei. In the experiment, 40 and 10 samples are set as training and testing samples. For NuCLS [4], we adopt the corrected single-rater dataset containing $1,744$ field-of-views with $59,485$ nuclei. Furthermore, to verify the effectiveness of each method on degenerated samples, we collect a new dataset containing 50 cancerous samples extracted from whole slide images of 10 patients. For all samples, the color normalization tool VahadaneNormalizer[1] is adopted to normalize the color variations caused by different staining and scanning of pathology images.

Framework architecture and parameters. In the experiment, we adopt the Unet [21] as the segmentation network, which could also be replaced by other SOTA segmentation networks. For MCF, the common parameters are set as follows: $K = 5, T = 3, H = 3, P = 3, Q = 10, t_1 = 0.4, t_2 = 0.6, t_3 = 0.4, t_4 = 0.65$. The appropriate number of nucleus patch templates for TNBC [19], MICCAI18 [16] and NuCLS [4] are 12, 10 and 16, respectively.

[1] https://staintools.readthedocs.io/en/latest/normalization.html

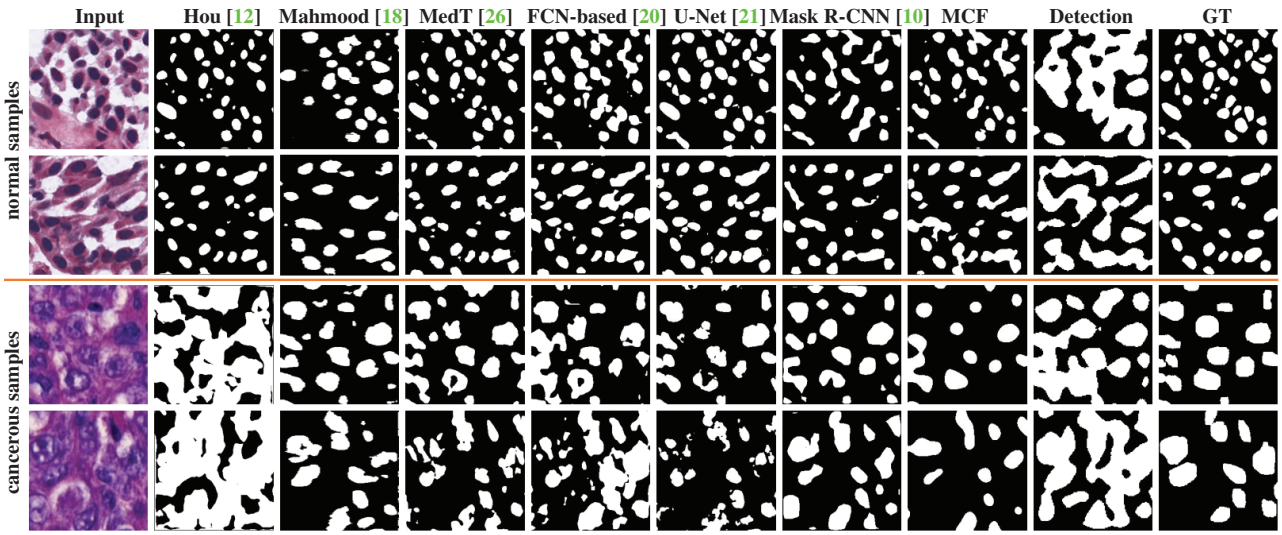

Figure 2. The visualized segmentation results of normal testing samples and cancerous samples.

	Type	*Unsupervised*	*Weakly-supervised*	*Fully Supervised*				*Ours*	
	Method	**Hou [12]**	**Mahmood [18]**	**MedT [26]**	**FCN-based [20]**	**U-Net [21]**	**Mask R-CNN [10]**	**MCF**	**Detection**
TNBC	**PA**	93.33 / 61.34	90.32 / 93.53	94.14 / 83.69	92.39 / 87.50	95.56 / 85.32	93.79 / 81.90	89.95 / 81.78	62.83 / 57.95
	MPA	79.87 / 71.64	67.36 / 66.72	85.46 / 78.03	84.29 / 72.71	89.55 / 81.39	82.81 / 71.03	80.51 / 77.49	72.52 / 69.61
	IOU	55.43 / 36.90	33.37 / 29.44	62.89 / 47.48	56.51 / 44.70	71.20 / 52.37	59.56 / 41.60	57.94 / 45.42	24.63 / 31.74
	FWIoU	87.69 / 46.82	82.20 / 88.43	89.36 / 73.62	86.82 / 77.12	91.79 / 75.96	88.62 / 68.87	81.75 / 71.34	53.03 / 44.27
	DICE	71.32 / 53.91	50.04 / 45.48	77.22 / 64.39	72.21 / 61.78	83.18 / 68.74	74.66 / 58.76	73.37 / 62.46	39.53 / 48.19
MICCAI18	**PA**	89.14 / 64.03	86.27 / 84.83	91.05 / 83.51	85.41 / 82.52	91.72 / 84.26	89.78 / 80.36	86.39 / 82.38	72.50 / 64.92
	MPA	76.95 / 73.14	75.88 / 72.71	85.25 / 75.61	81.83 / 72.57	87.04 / 78.41	80.85 / 74.08	81.43 / 74.94	76.70 / 72.68
	IOU	52.51 / 34.97	47.61 / 44.94	64.50 / 44.78	52.87 / 43.65	67.35 / 52.42	57.98 / 46.39	53.66 / 46.89	41.05 / 41.22
	FWIoU	80.05 / 50.96	76.35 / 72.79	83.93 / 73.12	76.12 / 70.30	85.10 / 73.47	81.57 / 71.88	77.30 / 70.67	60.33 / 50.11
	DICE	68.87 / 51.82	64.51 / 62.02	78.42 / 61.86	69.17 / 60.77	80.49 / 68.78	73.40 / 63.37	69.84 / 63.85	58.21 / 58.38

Table 1. The segmentation results of different methods on TNBC [19] dataset and MICCAI18 [16] dataset. 'score 1 / score 2' denotes the segmentation results of original testing samples and newly collected cancerous samples.

Metric. For segmentation, the metrics we adopted include Pixel Accuracy (PA), Mean Pixel Accuracy (MPA), Intersection over Union (IoU), Frequency Weighted Intersection over Union (FWIoU)£¬ and Dice. For detection, the metrics contain Precision, Recall, F-score.

3.1. Quantitative Evaluation

The proposed mutual-complementing framework can predict detection and segmentation results simultaneously. So, we compare the MCF with the SOTA nuclei detection and segmentation methods. The segmentation methods include *unsupervised methods* (Hou[12]), *weakly-supervised methods* (Mahmood [18]), *fully supervised methods* (MedT [26], FCN-based [20], U-Net [21] and Mask R-CNN [10]). The detection methods contain *unsupervised methods* (CAE [13]), *fully supervised methods* (SSAE [31], Mask R-CNN [10]). For all methods, the same data augmentation strategies, including flipping, mirroring, and cropping, are adopted. Other experiment settings are kept as the default in the original paper. Table 1 & 2 show the quantitative segmentation and detection results, where all the scores are the average of three runs. The 'score 1 / score 2' denotes the results of original testing samples and newly collected cancerous samples, respectively. It should be noted that all the cancerous samples of the new collected dataset are only used in the testing stage.

From Table 1, we can see that the fully supervised method U-Net [21] achieves the best segmentation performance on original testing samples and cancerous samples of two datasets. All fully supervised methods drop by about 7% on cancerous samples, which indicates that accurate segmentation of cancerous samples is a major challenge. The unsupervised method Hou [12] achieves closed PA performance on par with the fully supervised methods, while it drops by about 20% on cancerous samples. The weakly-supervised method Mahmood [18] combines all annotated samples and some synthetic samples with GAN to train the whole framework. However, [18] achieves lower scores than other fully supervised methods, which indicates that when supplying sufficient synthetic samples, the synthetic samples only have a slight impact even bring some negative effects. With only 10 nucleus patch templates, MCF achieves closed performance on par with fully supervised methods. Of particular note is that MCF only drops

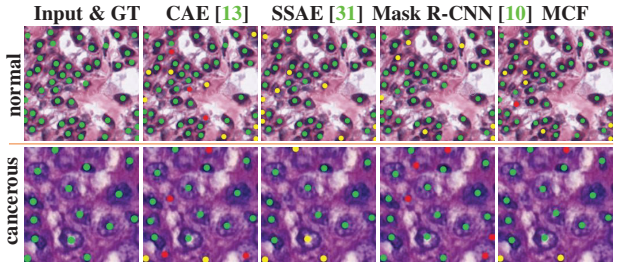

Figure 3. The detection visual results of different methods.

	Type	*Unsupervised*	*fully-supervised*		*Ours*
	Method	**CAE [13]**	**SSAE [31]**	**Mask R-CNN [10]**	**MCF**
TNBC	**Precision**	50.68 / 44.26	63.43 / 59.88	83.84 / 62.59	64.02 / 60.77
	Recall	69.37 / 61.98	59.96 / 53.59	81.74 / 73.78	71.43 / 65.32
	F-score	60.22 / 51.64	61.65 / 56.56	82.77 / 67.73	67.52 / 62.96
MICCAI18	**Precision**	43.71 / 40.02	62.54 / 58.74	82.73 / 66.83	65.40 / 61.89
	Recall	64.14 / 58.15	57.80 / 51.56	79.31 / 70.98	78.99 / 70.28
	F-score	51.99 / 47.41	60.08 / 54.92	80.98 / 68.84	71.56 / 65.82

Table 2. The detection results of different methods on TNBC [19] dataset and MICCAI18 [16] dataset.

by about 6% on cancerous samples compared to results on normal samples, which indicates that MCF has good robustness on degenerated samples. The proposed MCF has significant potential for analyzing cancerous cases in clinical diagnosis. What's more, for showing the advantage of mutual-complementing mechanism, we also give the segmentation results of the detection branch (denoted by 'Detection' in Table 1 and Fig. 2), which are obtained by binarizing the detection response map with a fixed thresholding value. It's obvious that 'Detection' drops about 25% and 12% on TNBC and MICCAI18, respectively.

For the detection results in Table 2, the fully-supervised method Mask R-CNN [10] achieves the best performance, and the unsupervised method CAE [13] achieves the worst detection results. With only 10 nucleus patch templates as supervision, MCF only drops by about 7% and 2% on TNBC [19] dataset and MICCAI18 [16] dataset compared with fully supervised methods. What's more, MCF maintains closed detection performance on cancerous samples compared with results on normal samples, which verifies the excellent robustness of MCF again.

3.2. Qualitative Evaluation

The visual segmentation results of different methods are shown in Fig. 2, where the first two rows are the normal samples from TNBC [19] and MICCAI18 [16], respectively. The last two rows are cancerous samples from the newly collected cancerous samples. We can see that most of the fully supervised methods achieve similar and promising results. Deep learning-based methods indeed have the potential to be applied to pathological nuclei segmentation. The unsupervised method Hou [12] also achieves promising segmentation results on the normal samples but poor segmentation results on cancerous samples. The root reason is that Hou [12] is based on the predefined nuclei texture and color. The degenerated sample usually has different and infrequent color and texture features. What's more, the style transfer is adopted by Hou [12], which leads to the overfitting on target domain and failure on new cancerous samples. On the contrary, MCF still achieves promising results on cancerous samples, which verifies the excellent robustness of MCF on degenerated samples.

Meanwhile, the visual detection results are given in Fig. 3. where the True Positives (TP), False Positives (FP), and False Negatives (FN) are marked with green, yellow, and red color points. Mask R-CNN [10] achieves the most accurate detection results than other methods. Compared with the unsupervised method CAE [13], MCF has higher TP and lower FP and FN. What's more, MCF still achieves not bad detection results on cancerous samples, which demonstrates the significant potential for nuclei detection in degenerated samples.

3.3. Ablation Study

3.3.1 The Effective of Each Component

In this section, we conduct an ablation study on the detection branch including Cropped Filter Fusion (*-CFF*), Multi-scale Filter Fusion (*-MFF*), Top-H accumulation, and Gradient Direction Based Postprocessing (*-GDBP*), and segmentation branch including removing segmentation network (*-seg.*), synthetic samples (*-syn.*), self-finetuning (*-self.*), training with standard Cross-Entropy (CE) and weighted CE (*+weig.*). '*+weig.*' denotes training the segmentation network with weighted loss $\mathcal{L}_{\mathbf{weight}} = weight_{nc} * \log(\mathbf{p_1}) + weight_{bg} * \log(\mathbf{p_0})$, where $weight_{nc}$ and $weight_{bg}$ denote the confidence values for foreground nuclei and background. Table 3 gives the ablation study on the above components. For the detection branch, we can see that precision scores of '*-CFF*', '*-MFF*' and '$-$*GDBP*' drop by about 5%, 15% and 3% compared with MCF, which indicates that CFF, MFF, and GDBP have significant influence on the final detection performance. What's more, '*-MFF*' achieves the highest Recall score. The reason is that detected results of the method without MFF contain massive positive and negative results. For the segmentation branch, the precision score of '*-seg.*' drop 14%, which indicates that the segmentation branch indeed detects some missing nuclei and refines the nuclei boundaries. '*-self.*', '$-$*syn.*', 'CE', and '*+weig.*' drop by about 2%, 3%, 15% and 10% compared with MCF, which verifies the necessity of self-finetuning, synthetic samples, and confidence loss. Fig. 4 visualize the response map with and w/o above components, which directly demonstrates the effectiveness of each component. More visual results of different cropped filters and different scale filters are given in *supplementary materials*.

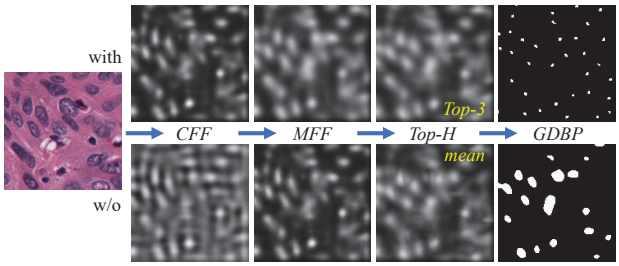

Figure 4. The response maps with and w/o some components.

	Index	-*CFF*	-*MFF*	H=1	mean	-*GDBP*	**MCF**
Detection	**Precision**	59.92	50.27	55.12	53.84	64.36	65.40
	Recall	79.57	89.68	77.16	76.33	71.82	78.99
	F-score	68.36	64.43	64.30	63.14	67.89	71.56
	Index	-*seg.*	-*self.*	-*syn.*	CE	+*weig.*	**MCF**
Segmentation	**PA**	72.50	85.56	84.38	86.32	73.69	86.39
	MPA	76.70	78.71	79.62	69.47	77.52	81.43
	IOU	41.05	50.01	49.71	38.63	40.99	53.66
	FWIoU	60.33	75.95	74.63	61.39	61.98	77.30
	DICE	58.21	66.68	66.41	55.73	58.14	69.84

Table 3. The ablation study results on MICCAI18 [16] dataset.

3.3.2 The Number of Nucleus Patch Templates

The number of nucleus patch templates has a significant impact on the detection branch and the whole framework. So, we study the detection and segmentation performance variation with different numbers of nucleus patch templates on the MICCAI18 [16] dataset, which is shown in Fig. 5. For the segmentation results with n nucleus patch templates, the segmentation network is trained on the pseudo masks generated by the detection branch with n nucleus patch templates. From Fig. 5, we can see that the detection and segmentation performance both increase with the increase of the template number. For MICCAI18 [16] dataset, 10 nucleus templates are the critical point, after which the detection and segmentation performance have a slight increase. More nucleus patch templates can increase the overall performance but also increase the computation overhead of the detection branch. So, limited nucleus patch templates that cover most types of nuclei are sufficient for MCF.

3.3.3 Iteration Times

The detection and segmentation branches of MCF are trained iteratively. The two branches' iteration times are closely related to the initial template number and the incremental template number at each iteration step. Taking experiments in Fig. 5 as an example, the initial template number and the incremental template number are both 1. The whole framework needs 10 iterations. If the initial template bank contains 10 nucleus templates, the whole framework will only needs $1 \sim 3$ iterations.

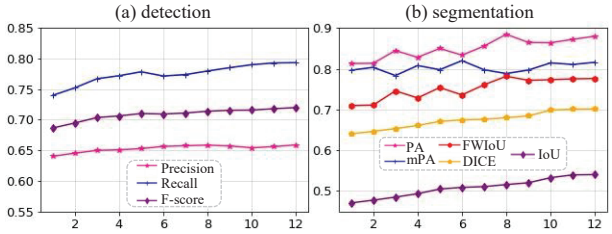

Figure 5. The ablation study on the (a) detection and (b) segmentation performance with different number of nucleus templates.

4. Conclusion

In this paper, we propose a Mutual-Complementing Framework (MCF) for nuclei detection and segmentation in pathology images, which are fundamental analysis operations in pathological diagnosis. In MCF, the detection and segmentation branch are trained in a mutual-complementing manner where the detection branch complements the pseudo mask of the segmentation branch, while the progressive trained segmentation branch complements the missing nucleus templates for the detection branch. In the detection branch, the cropped filter fusion and multi-scale filter fusion are devised for obtaining the optimal responses of crowded nuclei and different scale nuclei, respectively. Meanwhile, we put forward the gradient direction based postprocessing for locating the nuclei centers, which can handle the location problem of conjoined and overlapping nuclei. In the segmentation branch, the confidence loss is proposed to train the segmentation network only with high confidence areas. What's more, some synthetic nucleus samples and self-finetuning are used to enhance the segmentation performance on the boundary of nuclei and missing nuclei, respectively. Experiments on three datasets demonstrate that, with only a few nucleus patch templates, the proposed MCF achieves closed performance on par with fully supervised methods. In particular, MCF has excellent robustness (only dropping by about 6%) on degenerated samples, which are common and crucial cases in clinical diagnosis. In the future, we will focus on reducing the performance drop on degenerated samples.

5. Acknowledgments

This work is supported by National Natural Science Foundation of China (No.62002318), Key Research and Development Program of Zhejiang Province (2020C01023), Zhejiang Provincial Science and Technology Project for Public Welfare (LGF21F020020), the Major Scientific Research Project of Zhejiang Lab (No. 2019KD0AC01), Ningbo Natural Science Foundation (202003N4318), and Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies.

References

[1] Hao Chen A, Xiaojuan Qi A, Lequan Yu A, Qi Dou A, Jing Qin B, and Pheng Ann Heng A. Dcan: Deep contour-aware networks for object instance segmentation from histology images - sciencedirect. *Medical Image Analysis*, 36:135–146, 2017. 1, 2

[2] Asif Ahmad, Amina Asif, Nasir Rajpoot, Muhammad Arif, and Fayyaz Ul Minhas. Correlation filters for detection of cellular nuclei in histopathology images. *Journal of Medical Systems*, 42(1):1–8, 2018. 2

[3] Saad Ullah Akram, Juho Kannala, Lauri Eklund, and Janne Heikkila. Cell segmentation proposal network for microscopy image analysis. In *IEEE International Conference on Image Processing*, pages 3199–3203, 2016. 1

[4] Mohamed Amgad et al. Nucls: A scalable crowdsourcing, deep learning approach and dataset for nucleus classification, localization and segmentation. *Journal of Medical Internet Research*, 2021. 1, 2, 5

[5] David S. Bolme, J. Ross Beveridge, Bruce A. Draper, and Yui Man Lui. Visual object tracking using adaptive correlation filters. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 2, 3

[6] Yuxin Cui, Guiying Zhang, Zhonghao Liu, Zheng Xiong, and Jianjun Hu. A deep learning algorithm for one-step contour aware nuclei segmentation of histopathological images. *Medical Biological Engineering Computing*, 2018. 1, 2

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. 3

[8] Xuan Gong, Shuyan Chen, Baochang Zhang, and David Doermann. Style consistent image generation for nuclei instance segmentation. In *IEEE Winter Conference on Applications of Computer Vision*, pages 3993–4002, 2021. 3

[9] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, and Bulent Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147–171, 2009. 1

[10] Kaiming He, Georgia Gkioxari, Piotr Doll¨¢r, and Ross Girshick. Mask r-cnn. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 1, 6, 7

[11] Jo?o F. Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014. 2, 3

[12] Le Hou, Ayush Agarwal, Dimitris Samaras, Tahsin M. Kurc, and Joel H. Saltz. Robust histopathology image analysis: To label or to synthesize? In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2, 3, 6, 7

[13] Le Hou, Vu Nguyen, Ariel B. Kanevsky, Dimitris Samaras, Tahsin M. Kurc, Tianhao Zhao, Rajarsi R. Gupta, Yi Gao, Wenjin Chen, and David and Foran. Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. *Pattern recognition*, 2019. 1, 2, 3, 6, 7

[14] Shaoli Huang, Xinchao Wang, and Dacheng Tao. SnapMix: Semantically Proportional Mixing for Augmenting Fine-grained Data. In *AAAI Conference on Artificial Intelligence*, 2021. 2

[15] H. Irshad, L. Montaser-Kouhsari, G. Waltz, O. Bucur, J. A. Nowak, F. Dong, N. W. Knoblauch, and A. H. Beck. Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*, pages 294–305, 2015. 1, 2

[16] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, pages 1–1, 2017. 1, 2, 5, 6, 7, 8

[17] Dongnan Liu, Donghao Zhang, Yang Song, Fan Zhang, Lauren ODonnell, Heng Huang, Mei Chen, and Weidong Cai. Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4243–4252, 2020. 3

[18] Faisal Mahmood, Daniel Borders, Richard J. Chen, Gregory N. Mckay, and Nicholas J. Durr. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Transactions on Medical Imaging*, PP(99):1–1, 2019. 2, 3, 6

[19] Naylor, Peter, Lae, Marick, Reyal, Fabien, Walter, and Thomas. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 38(2):448–459, 2019. 2, 5, 6, 7

[20] Peter Naylor, Marick La¨¦, Fabien Reyal, and Thomas Walter. Nuclei segmentation in histopathology images using deep neural networks. In *IEEE International Symposium on Biomedical Imaging*, 2017. 1, 2, 6

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. 1, 2, 5, 6

[22] Massimo Salvi and Filippo Molinari. Multi-tissue and multi-scale approach for nuclei segmentation in h&e stained images. *Biomedical Engineering Online*, 17(1), 2018. 1

[23] Chengchao Shen, Youtan Yin, Xinchao Wang, Xubin Li, Jie Song, and Mingli Song. Training Generative Adversarial Networks in One Stage. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 3

[24] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee Wah Tsang, David R. J. Snead, Ian A. Cree, and Nasir M. Rajpoot. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging*, 35(5):1196–1206, 2016. 1, 2

[25] Mahmood Tahir, Owais Muhammad, Noh Kyoung Jun, Yoon Hyo Sik, Haider Adnan, Sultan Haseeb, and Park Kang Ryoung. Artificial intelligence-based segmentation of nuclei in multi-organ histopathology images: Model development and validation. *JMIR Medical Informantics*, 2021. 1, 2

[26] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel. Medical transformer: Gated axial-

attention for medical image segmentation. *arXiv preprint arXiv:2102.10662*, 2021. 2, 6

[27] J. Van, de Weijer, C. Schmid, J. Verbeek, and D. Larlus. Learning color names for real-world applications. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 18(7):1512–23, 2009. 3

[28] Xinchao Wang, Engin Turetken, Francois Fleuret, and Pascal Fua. Tracking interacting objects using intertwined flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38:2312–2326, 2016. 2

[29] Fuyong Xing, Yuanpu Xie, and Lin Yang. An automatic learning-based framework for robust nucleus segmentation. *IEEE Transactions on Medical Imaging*, 35(2):550–566, 2016. 1, 2

[30] Fuyong Xing and Lin Yang. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. *IEEE Rev Biomed Eng*, pages 234–263, 2016. 2

[31] Jun Xu, Lei Xiang, Renlong Hang, and Jianzhong Wu. Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. In *IEEE International Symposium on Biomedical Imaging*, 2014. 1, 2, 6, 7

[32] Yiding Yang, Zunlei Feng, Mingli Song, and Xinchao Wang. Factorizable graph convolutional networks. In *Advances in Neural Information Processing Systems*, 2020. 2

[33] Yiding Yang, Jiayan Qiu, Mingli Song, Dacheng Tao, and Xinchao Wang. Distilling knowledge from graph convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7074–7083, 2020. 1

[34] Yiding Yang, Zhou Ren, Haoxiang Li, Chunluan Zhou, Xinchao Wang, and Gang Hua. Learning Dynamics via Graph Neural Networks for Human Pose Estimation and Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 2

[35] Jingwen Ye, Yixin Ji, Xinchao Wang, Kairi Ou, Dapeng Tao, and Mingli Song. Student Becoming the Master: Knowledge Amalgamation for Joint Scene Parsing, Depth Estimation, and More. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2

[36] Xiaoqing Yin, Xinchao Wang, Jun Yu, Maojun Zhang, Pascal Fua, and Dacheng Tao. FishEyeRecNet: A Multi-Context Collaborative Deep Network for Fisheye Image Rectification. In *European Conference on Computer Vision*, 2018. 2

[37] Haichao Yu, Ding Liu, Honghui Shi, Hanchao Yu, Zhangyang Wang, Xinchao Wang, Brent Cross, Matthew Bramlet, and Thomas S Huang. Computed Tomography Super-Resolution Using Convolutional Neural Networks. In *IEEE International Conference on Image Processing*, 2017. 2

[38] Yin Zhou, Hang Chang, Kenneth E. Barner, and Bahram Parvin. Nuclei segmentation via sparsity constrained convolutional regression. In *IEEE International Symposium on Biomedical Imaging*, 2015. 1, 2

[39] Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *IEEE International Conference on Computer Vision*, 2017. 3