Fast ScanNet: Fast and Dense Analysis of Multi-Gigapixel Whole-Slide Images for Cancer Metastasis Detection

Huangjing Lin, *Student Member, IEEE*, Hao Chen, *Member, IEEE*,
Simon Graham, *Student Member, IEEE*, Qi Dou, *Member, IEEE*,
Nasir Rajpoot, *Senior Member, IEEE*, and
Pheng-Ann Heng, *Senior Member, IEEE*

Abstract—Lymph node metastasis is one of the most important indicators in breast cancer diagnosis, that is traditionally observed under the microscope by pathologists. In recent years, with the dramatic advance of high-throughput scanning and deep learning technology, automatic analysis of histology from whole-slide images has received a wealth of interest in the field of medical image computing, which aims to alleviate pathologists' workload and simultaneously reduce misdiagnosis rate. However, the automatic detection of lymph node metastases from whole-slide images remains a key challenge because such images are typically very large, where they can often be multiple gigabytes in size. Also, the presence of hard mimics may result in a large number of false positives. In this paper, we propose a novel method with anchor layers for model conversion, which not only leverages the efficiency of fully convolutional architectures to meet the speed requirement in clinical practice but also densely scans the whole-slide image to achieve accurate predictions on both micro- and macro-metastases. Incorporating the strategies of asynchronous sample prefetching and hard negative mining, the network can be effectively trained. The efficacy of our method is corroborated on the benchmark dataset of *2016 Camelyon Grand Challenge*. Our method achieved significant improvements in comparison with the state-of-the-art methods on tumor localization accuracy with a much faster speed and even surpassed human performance on both challenge tasks.

Index Terms—Histopathology image analysis, computational pathology, whole-slide image, deep learning, metastasis detection.

I. INTRODUCTION

A. Background

BREAST cancer is one of the leading causes of cancer related death within women [1]. One of the most important diagnostic criteria of breast cancer grading is identifying the metastases in the sentinel lymph nodes. According to the tumor, node and metastasis (TNM) breast cancer staging system [2], a higher staging of the patient indicates diagnosis of larger or more metastases, which affects therapeutic operation afterwards.

However, the pathologic diagnosis is extremely time-consuming, laborious and highly dependent on expertise which requires pathologists to carefully examine the biopsies under the microscope [3], [4]. Moreover, the population of pathologists cannot meet the sharply growing demand of diagnosis, with the cancer morbidity increasing [5]. The pathologists have to diagnose a large number of biopsy slides everyday, thus it is quite cumbersome, if not impossible, to conduct a thorough inspection of the whole slide. Also, the risk of misdiagnosis may increase if only small regions of interest are analyzed.

Over the last decade, automatic analysis of histology has become one of most rapidly expanding fields in medical image computing. Computer aided diagnostics can not only alleviate pathologists' workload, but also help to reduce the misdiagnosis rate. The advent of high-throughput scanning technology makes whole-slide image (WSI) analysis more essential in digital pathology.

Metastasis detection from WSIs plays a key role in breast cancer diagnosis. However, this task is very challenging due to several factors: 1) the large variations of textures and biological structures, as shown in Fig. 2(a); 2) the hard mimics with similar morphological appearance between normal and metastatic regions, as shown in Fig. 2(b); 3) as illustrated in Fig. 2(c), the image appearance is easily affected by the image acquisition process, e.g., sampling, staining and

H. Lin, Q. Dou, and P. A. Heng are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong.

H. Chen is with Imsight Medical Technology, Co., Ltd., Shenzhen 518063, China (e-mail: hchen@cse.cuhk.edu.hk).

N. Rajpoot is with the Department of Computer Science, University of Warwick, Coventry CV4 7AL, U.K.

S. Graham is with the Department of Computer Science, University of Warwick, Coventry CV4 7AL, U.K., and also with the Mathematics for Real-World Systems Centre for Doctoral Training, University of Warwick, Coventry CV4 7AL, U.K.
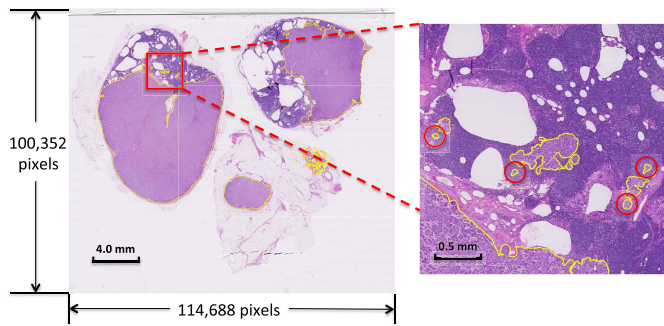
Fig. 1. Challenge in multi-gigapixel whole-slide image analysis. Detecting micro metastasis (denoted by red circles in right figure) is just like finding a needle in a haystack.
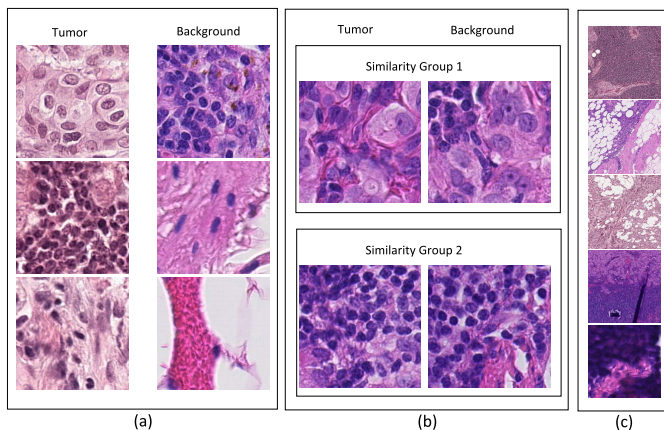

Fig. 2. Illustration of some key challenges for detecting tumor in the histology images. (a) Variations of biological structures and textures. (b) Hard mimics of tumor from normal tissues. (c) Appearance variations due to the acquisition process.

digitalizing [6]; 4) the significant size variance between micro- and macro-metastases, see Fig. 1. Tiny micro-metastases are easier to be neglected compared with the macro ones. Finally, the WSI is with size up to 200000×100000. How to efficiently process such a gigapixel image poses further challenges to automatic detection methods, especially in finding micro-metastases.

B. Related Work

The early researches of histology image analysis date back to 60s [7]. During last few decades, many works have been proposed for breast cancer diagnosis in a variety of applications including nuclei, tubules, and mitosis, etc.

1) Region of Interest Analysis: Previously due to the lack of whole-slide scanning techniques and computational power, most of the literature focused on regions of interest (ROI), e.g., image size 500 × 500, pre-selected by pathologists from WSIs [4], [8].

In the earlier years, most of the study focused on hand-crafted features. For nuclei detection and segmentation, a series of level set methods equipped hand-crafted features, such as Hough transform [9], concavity [10] and gradient [11], etc., were proposed as a fundamental prerequisite in many breast cancer histopathological applications. These features

were exquisitely considered with prior knowledge of boundary, region or shape. Besides, a series of wavelet filters equipped hand-crafted features, such as isotropic phase symmetry [12] and texture descriptor [13], were proposed for detection of beta cells, lymphocytes or glandular structure. Afterwards, researchers utilized machine learning methods, like Bayesian classifier [14] and Support Vector Machine (SVM) [15] to detect or segment nuclei. These kinds of hand-crafted methods were also applied in a wide range of other digital histopathological applications, such as level set for tubule segmentation in breast cancer [16], [17] and SVM for gland detection in prostate cancer [18], etc. However, hand-crafted features are limited in representation capability to solve such complex problems.

Recently, with the advance of computing power, deep learning methods have achieved remarkable success in many fields, including image classification [19]–[21], semantic segmentation [22]–[25], object detection [26], and natural language processing [27], etc. In the field of histology, Kumar *et al.* [28] released a dataset with ROIs extracted from whole-slide images and proposed boundary aware CNNs for nuclei segmentation. Xu *et al.* [29] proposed stacked sparse autoencoders (SSAE) with unsupervised pre-training for nuclei detection from breast cancer histopathological images. Su et al. [30] used deep CNNs as pixel classifiers for breast cancer region segmentation in histopathological images. Cireşan *et al.* [31] proposed a deep neural network for mitosis detection and achieved the best performance in two grand challenges [8], [32]. A deep cascaded framework was proposed in [33] to fast retrieve the mitoses and suppress false positives from breast cancer histology images. Xie *et al.* [34] proposed a structured regression approach for robust cell detection using convolutional neural network (CNN). A locality sensitive deep learning method was developed for detection and classification of nuclei in routine colon cancer histology images [35]. Chen *et al.* [36] and Sirinukunwattana *et al.* [37] exploited a deep contour-aware network for gland and nuclei instance segmentation from histopathological images, which significantly outperformed other methods in two recent challenges. A weakly supervised deep learning algorithm was developed to segment cancerous regions from histopathology images [38]. Ronneberger *et al.* [23] proposed the U-net for neuronal structure segmentation from microscopy images, which achieved good performance in a wide range of segmentation tasks. However, these methods only focus on a small number of pre-selected regions and therefore do not translate well to pathological practice.

2) Whole-Slide Image Analysis: There were limited works on WSI analysis in gigapixel until the appearance of high-throughput scanners in recent years, which has become a trend in the field of histopathological image analysis [39]. A series of challenges related to whole-slide image analysis have been presented for a variety of classification, detection or segmentation tasks. In the Her2 challenge contest [40], Qaiser et al. presented a benchmark for comparing the performance of automated algorithms for scoring of Her2 from digitized WSIs of invasive breast carcinoma, which demonstrated the enormous potential of automated algorithms in assisting

the pathologists with objective immunohistochemistry (IHC) scoring. In the challenge on cancer metastasis detection in lymph node (Camelyon16)[1] held in conjunction with ISBI 2016 [3], Wang *et al.* [41] presented an ensemble of two GoogLeNets, which achieved the best performance on metastasis detection. Kong *et al.* [42] proposed a spatially structured deep network containing appearance and spatial dependency information for cancer metastasis detection. In the breast histology image (BACH) classification challenge, Aresta *et al.* [43] held the challenge to find the best whole-slide level classifier for normal, benign, in situ carcinoma and invasive carcinoma classification. he breast tumor proliferation prediction (TUPAC16) challenge was held to predict tumor proliferation from whole-slide images [44]. Apart from the challenges, many works regarding whole-slide image analysis have been proposed recently. Graham *et al.* [45] classified whole-slide images of lung cancer based on morphological features extracted from probability maps produced by a patch-based ResNet32. Qaiser et al. proposed a patch-based k-NN classifier for colorectal tumor segmentation, with persistent homology profiles for exemplary training sample selection [46]. However, the above frameworks used patch-based classification for whole-slide image analysis, which significantly increases the computational cost at the highest resolution, and hence is suboptimal in real clinical practice. Then, Xu et al. proposed a sparse kernel technique to accelerate the pixel-wise predictions, which could alleviate the efficiency problem to some extent [47]. However, pixelwise predictions cost much more time and computational resources, which are not necessary in our underlying problem. Therefore, how to efficiently process the image while ensuring dense prediction is crucial for WSI analysis. A preliminary version of our work was presented in [48], which proposed a dense reconstruction method by leveraging the stitching process. Although, it could accelerate the detection speed, the time cost could be linearly increased with dense factors. In this paper, we propose a novel layer, i.e., Anchor Layer, which can equivalently remodel the offsetting based dense reconstruction mechanism [48] into a fully convolutional way. The proposed method can achieve more accurate results by densely scanning whole sliding images while demonstrating computational effectiveness by utilizing similar time cost.

C. Contributions

Aiming at tackling these challenges, we presented a new mechanism for accelerating the inference speed dramatically in this study while maintaining competitive accuracy. The main contributions of this paper is summarized as follows:

- We propose a novel framework, referred to as Fast ScanNet, for histopathology image inference. By leveraging the efficiency of the fully convolutional architecture to meet the speed requirement for clinical practice, the proposed Fast ScanNet is particularly fast in gigapixel level WSI analysis, generating results hundred times faster than patch-based frameworks by sharing the computations in the overlapping regions [49].

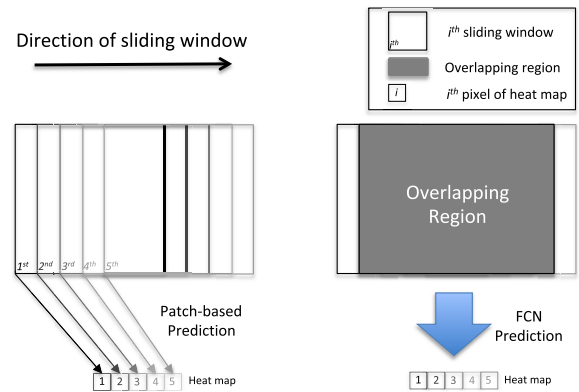[1]2016 ISBI Camelyon challenge: https://camelyon16.grand-challenge.org/

Fig. 3. The merit of using the fully convolutional architecture. It can speed up the inference by sharing computations in the overlapping regions.

- In order to further improve the performance, the idea of model conversion is exploited to ensure dense scanning, which makes the predictions more accurate on both micro- and macro-metastases. With the proposed anchor convolutional/pooling layers to remodel the offsetting scanning in our previous work [48], Fast ScanNet can improve the accuracy significantly by densely scanning while adding little time cost, i.e., only 10% extra time cost for 4 times dense predictions.
- Compared with the state-of-the-art methods [3] on the benchmark dataset of *Camelyon16*, our method achieved superior performance with a much faster speed of within one minute in tumor localization task and even surpassed human performance on both challenge tasks.

II. METHODOLOGY

A. Fast Prediction via Modified Fully Convolutional Network (FCN)

Patch-based frameworks are widely used in object detections tasks. One of the advantages is that data augmentation is very flexible when training a deep neural network. However, there exists a large amount of redundant computations in the overlapping regions when a window scans densely, as illustrated in Fig. 3, which results in dramatically increased computational cost when applied to large-scale gigapixel image analysis. In order to resolve this issue, we propose a modified fully convolutional network that can take arbitrary sized images as input for fast prediction, derived from the mechanism of sharing computation proposed in [49], but it is trained in a patch-based manner.

1) Modified FCN: Different from standard FCNs, our FCN removes upsampling path which is a must for segmentation but not necessary for detection tasks [36]. Furthermore, the upsampling path greatly slows down the detection process when considering the large size of WSIs but does not provide additional benefit to the quality of heat map for detection task, as shown in Fig. 4 (a). The fully convolutional network without the upsampling path can efficiently output a heat map with a much smaller size than the input image. We further construct a more powerful network with denser heat map as output by
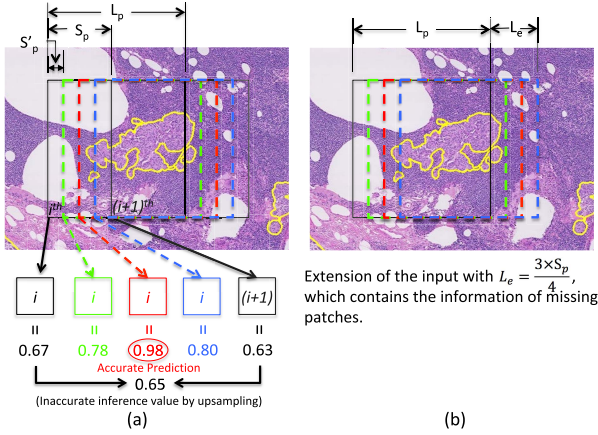
Fig. 4. (a) The different dense mechanism between our method and standard FCN with interpolation upsampling. The patches and pixels in black are the same as those in a standard FCN without dense model conversion. Green, red and blue ones are the patches that are missed by standard FCN. (b) The extension of input (the same goes for ROI) with dense coefficient $\alpha = 4$ before dense inference.

TABLE I
THE ARCHITECTURE OF OUR PROPOSED NETWORK

Layer	Features (Train)	Features (Inference)	Kernel size	Stride
Input	$244 \times 244 \times 3$	$2868 \times 2868 \times 3$	-	-
Conv1_1	$242 \times 242 \times 64$	$2866 \times 2866 \times 64$	3×3	1×1
Conv1_2	$240 \times 240 \times 64$	$2864 \times 2864 \times 64$	3×3	1×1
Pool1	$120 \times 120 \times 64$	$1432 \times 1432 \times 64$	2×2	2×2
Conv2_1	$118 \times 118 \times 128$	$1430 \times 1430 \times 128$	3×3	1×1
Conv2_2	$116 \times 116 \times 128$	$1428 \times 1428 \times 128$	3×3	1×1
Pool2	$58 \times 58 \times 128$	$714 \times 714 \times 128$	2×2	2×2
Conv3_1	$56 \times 56 \times 256$	$712 \times 712 \times 256$	3×3	1×1
Conv3_2	$54 \times 54 \times 256$	$710 \times 710 \times 256$	3×3	1×1
Conv3_3	$52 \times 52 \times 256$	$708 \times 708 \times 256$	3×3	1×1
Pool3	$26 \times 26 \times 256$	$354 \times 354 \times 256$	2×2	2×2
Conv4_1	$24 \times 24 \times 512$	$352 \times 352 \times 512$	3×3	1×1
Conv4_2	$22 \times 22 \times 512$	$350 \times 350 \times 512$	3×3	1×1
Conv4_3	$20 \times 20 \times 512$	$348 \times 348 \times 512$	3×3	1×1
Pool4	$10 \times 10 \times 512$	$174 \times 174 \times 512$	2×2	2×2
Conv5_1	$8 \times 8 \times 512$	$172 \times 172 \times 512$	3×3	1×1
Conv5_2	$6 \times 6 \times 512$	$170 \times 170 \times 512$	3×3	1×1
Conv5_3	$4 \times 4 \times 512$	$168 \times 168 \times 512$	3×3	1×1
Pool5	$2 \times 2 \times 512$	$84 \times 84 \times 512$	2×2	2×2
Conv6	$1 \times 1 \times 1024$	$83 \times 83 \times 1024$	2×2	1×1
Conv7	$1 \times 1 \times 1024$	$83 \times 83 \times 1024$	1×1	1×1
Conv8	$1 \times 1 \times 2$	$83 \times 83 \times 2$	1×1	1×1

converting the existing model with proposed anchor layers, which will be elaborated on in the next subsection.

Our FCN is designed based on VGG16 network [50] by replacing the last three fully connected layers with fully convolutional layers $1024 \times 1024 \times 2$ (i.e., kernel size 1×1). The padding operations are removed from our architecture to avoid the boundary effect and preserve the inference equivalence, as shown in the training phase of Table I. Based on this modification, our FCN can enjoy the transferred features learned from a large set of natural images [51], [52], which demonstrated consistent improvements over that without transfer learning. Note that the 2-D square sizes in this paper are briefly denoted by their length of side, such as L for length and S for stride. In the training phase, we employ patch samples with size as $L_p = 244$ randomly cropped from WSIs to train the modified VGG16 model. In the inference phase as shown in Table I, our FCN can take a large input, i.e., ROI with size up to $L_r = 2868$ (determined by the memory capacity

of GPU) and output a heat map with size $L_m = 83$. The proposed network inherently falls into the category of fully convolutional architecture, which is equivalent to a patch-based convolutional neural network with input size $L_p = 244$ and scanning stride $S_p = 32$, but the inference speed is much faster by removing the redundant computations of overlaps.

The relationship between the input size L_r and output size L_m is denoted as follows,

$$L_r = L_p + (L_m - 1) \times S_p \qquad (1)$$

Here, we use the L_p and S_p as patch size and scanning stride of FCN to represent the input size and scanning stride of equivalent patch-based model.

2) Effective Training Strategies: We further propose two effective training strategies to enhance the learning process of FCN.

a) Exhaustive sample augmentation: The samples from different classes are randomly fetched under certain probabilities to avoid sample imbalance problem. They are augmented with shift, rotation, flipping, scaling and color jittering on the fly before they are batched for training. In addition, we implement an asynchronous sample prefetching mechanism with multiple processors to maximize the utilization rate of GPUs, which can significant relieve the I/O bottleneck and accelerate the training and inference procedure.

b) Hard negative mining: We trained our model in two stages, i.e., general sample learning and hard negative sample mining. In the first learning stage, the model was initialized with VGG-16 model pre-trained from ImageNet and trained on randomly extracted training samples. While there exists lots of negative training samples from the WSIs, most of them can be easily distinguished from the true metastases. In the second learning stage, in order to enhance the discriminative capability of our model, we add the false positive samples, i.e., hard negative mining (HNM) examples, from the previously trained classifier back to the training data. This strategy makes the training process more effectively by focusing on hard cases, which contributes to greatly boost the recognition performance.

B. Fast and Dense ScanNet Inference With Anchor Layers

Densely scanning stride is important for tiny target retrieval. As Fig. 4 (a) shows, with a sparsely scanning stride, a framework might potentially miss the patches that can generate more accurate probabilities for tiny targets. Although such missing probabilities can be fabricated by upsampling path in standard FCN for an end-to-end result, it does not help to boost the performance for a detection task.

Modified FCN is fast, but the drawback of the fully convolutional architecture is that its scanning stride is determined by its downsampling layers. For example, if FCN downsamples the input by five 2×2 pooling layers, the scanning stride of the FCN should be $2^5 = 32$, i.e., this FCN is equivalent to a patch-based CNN network with a scanning stride of 32. Namely, the deeper network with the more downsampling layers, the sparser heat map is generated. This drawback makes
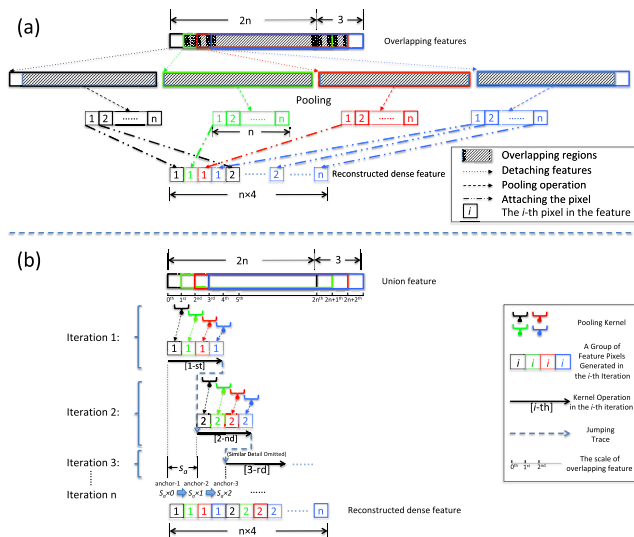
Fig. 5. Two methods of dense reconstruction. (a) ScanNet: Dense reconstruction by feature detaching and stitching [48]. (b) Fast ScanNet: dense reconstruction in convolutional version. The features in black represent the features without dense reconstruction.
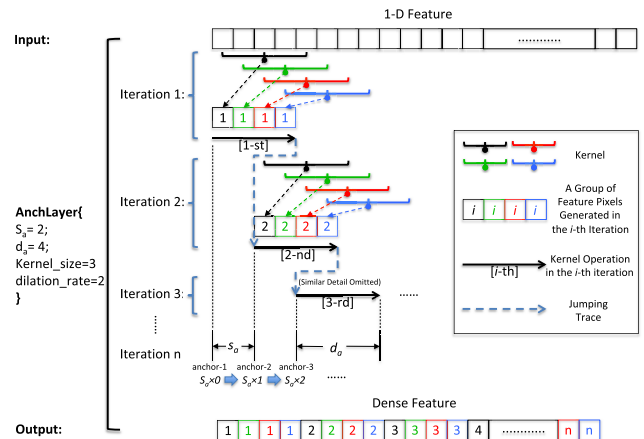

Fig. 6. A simple instance of anchor layer on 1-D case. It takes an 1-D feature as input, with a dense feature as output, the combination of the grouped feature pixels generated from all the iterations. The anchor stride and operating distance of instance are sized as 2 and 4; The convolutional/pooling kernel is sized as 3 with dilation rate as 2 and stride as constant 1 (this stride is constant 1 for Fast ScanNet because the feature pixels related to same group are neighbouring by 1 in all situations). Kernel operation is configured with operating distance $d_a = 4$, which displays the sliding process of kernel in a certain iteration. The jumping traces denote the movements of kernel from one iteration to the next.

it difficult to balance the tradeoff between the depth of network and the density of scanning.

Therefore, the key solution to an accurate and dense heat map lies in how to compute the missing patches in the gap of sliding window. For brevity, the samples are demonstrated in 1-D case in the following paragraphs, which can be easily generalized to 2-D or higher dimensions.

1) Anchor Layers: Fig. 5 (a) illustrates an intuitive solution for dense prediction from our preliminary work of Scan-Net [48]: detaching the overlapping features (input or output of layers of a neural network), computing the missing patches and reconstructing the dense feature by stitching the output of pooling alternately. After feature reconstruction, the size of output increase by 4 times from n to $4n$. Here, we denote the ratio of increment in a certain dimension as dense coefficient α. Namely, in a n-D case, the output will be enlarged α^n times. However, this solution is not optimal as the overlapping regions of features still exist redundant computations.

In order to leverage the efficiency of the fully convolutional architecture, we model the procedure of dense reconstruction by a new pooling layer, i.e., anchor pooling layer. As shown in Fig. 5 (b), we name the whole set of overlapping features as union feature. In iteration 1, as inputs are overlapped and shifted by 1 pixel, it is easy to compute a group of pixels (black-1, green-1, red-1 and blue-1) from the inputs respectively if we start from the position 0th to slide the kernel four times with stride 1. Afterward, in iteration 2, the kernel should jump back to the position 2nd for the black-2 which is the second pixel that should be pooled from the black feature. Here, we denote this jumping location as anchor. It is easily understood that the anchor moves forward 2 units iteratively, which is equal to the stride of original standard pooling layer.

For generalization, we propose the anchor convolutional layer (denoted as AnchConv) and anchor pooling layer (denoted as AnchPool) for the basis of our model conversion,

which helps to convert a sparse scanning framework to a dense one. As illustrated in Fig. 6, the anchor layers are different from standard convolutional/pooling layers in their settings of anchor stride s_a, operating distance d_a and dilation rate. The standard convolutional/pooling kernel operates on the feature from the beginning to the end continuously, but the kernel of anchor layer will stop and jump from one anchor position to the others once it finishes its operations in a certain distance. We define the length of anchor movement in each iteration as anchor stride, and call this certain distance as operation distance. Dilation rate is the same concept proposed in [53] to enlarge the receptive field of input, but in our model, we utilize it to ensure the pixels can be fetched and computed by kernel with certain intervals. Other parameters of anchor layers are inherited from standard convolution/pooling layers such as the kernel size and stride. To make it brief and understandable, a simple instance of anchor layer with kernel size 3 and dilation rate 2 is displayed on 1-D feature in Fig. 6. Here, we denote the anchor positions as "anchor-1, anchor-2, ..., anchor-i, etc." for "iteration 1, iteration 2, ..., iteration i, etc.", which start at the 0th of the feature and move forward s_a after each iteration. In iteration i, the kernel starts operations at the position of anchor-$i = s_a \times (i-1)$ and terminates in distance d_a. Afterwards, the kernel will jump to the anchor-$(i+1)$ for the next iteration.

2) Model Conversion: Taking the model proposed in Section II-A for instance, we convert the VGG16 modified FCN to Fast ScanNet for dense prediction. The conversion can be concluded in two Phases. In Phase 1, the architecture of Fast ScanNet is inherited from FCN without changes. In Phase 2, all the standard convolutional/pooling layers of FCN should be converted to the proposed anchor layers.
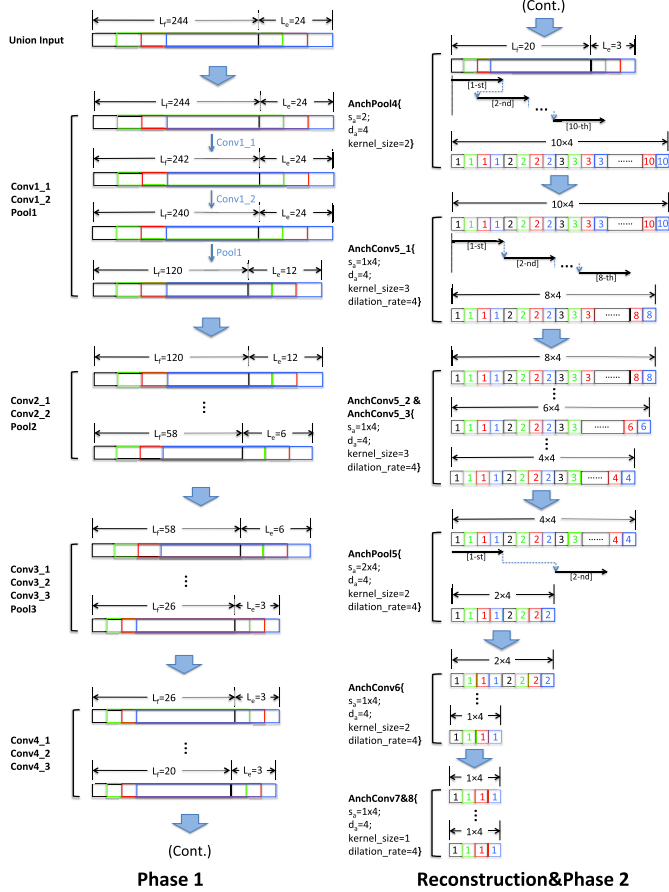
Fig. 7. The model conversion by anchor layers. Phase 1: standard convolutional/pooling operations on sparse features before L_e is divisible by pooling stride. Phase 2: AnchConv/AnchPool inference on dense features after L_e is indivisible by pooling stride. (Details of Phase 2 are illustrated in Fig. 8).

Fig. 8. The details of dense feature reconstruction and inference on dense features in Phase 2. The green, red, and blue squares represent the feature pixels from the missing patches. The black squares are the feature pixels same with the pixels from original version without dense prediction.

To ensure the conversion easily understandable, we simplify the illustration on 1-D case with configurations of dense coefficient $\alpha = 4$ and size as 244, as shown in the Fig. 7. First of all, the input (the same goes for large input ROI) should be cropped with an extra extension L_e. This ensures the input to be the union region of the existing and missing patches covering all of the needed spatial information. The L_e is formulated as $L_e = (\alpha - 1) \times S_p / \alpha$, as shown in Fig. 4 (b). The scanning stride of Fast ScanNet can be reduced from S_p to $S'_p = S_p / \alpha$. In our instance, i.e., the input are extended to $244 + 24$ and the scanning becomes dense with scanning stride reduced from 32 to 8.

At the beginning, without any changes, the inputs and features (outputs of layers) can be normally computed by standard convolutional/pooling layers of FCN, as is denoted as Phase 1 (as shown in Fig. 7). However, when it comes to Pool4, the operation fails in proceeding because it takes an odder sized union feature as $20 + 3$ for downsampling. The standard layers are unable to continue the further calculations, and hence we need an anchor layer to tackle this.

Then, the Pool4 should be replaced by AnchPool for feature reconstruction. As illustrated in Fig. 8 (a), the union feature consists of four overlapped features shifted by 1 and the stride
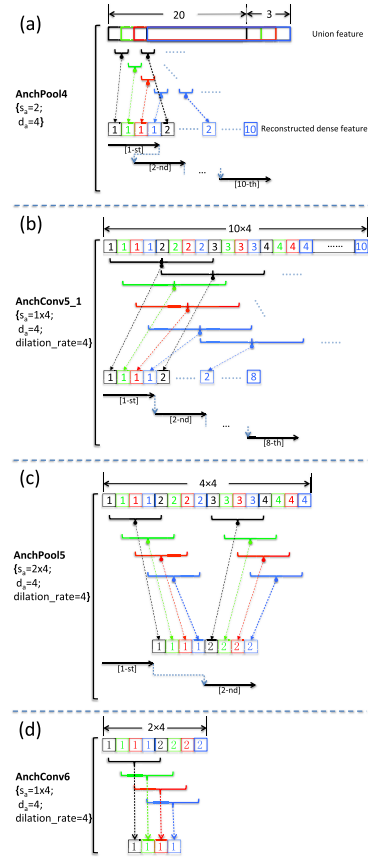
of original standard pooling layer is 2. The anchor stride s_a of AnchPool should be equal to the stride of original standard pooling layer as 2 and dilation rate should be equal to the dense coefficient α. Thus, we can reconstruct a dense feature for further proceeding.

Afterward, all of the standard layers of FCN should be converted to anchor layers, as they are no longer applicable to the dense features after feature reconstruction. We denoted this stage as Phase 2. From the Fig. 8 (b,c and d), it is observed that the dense features are staggered, i.e., the original neighbouring pixels (denoted in black) are separated by three missing ones (denoted in green, red and blue). Thus, dilation configurations should be employed as 4 (equal to dense coefficient α) to ensure the neighbouring pixels can be fetched correctly. Besides, the stride of anchor depends on the stride of original layer and dense coefficient.

In summary, the workflow of model conversion can be concluded as follow: first of all, the input should be extended with $L_e = (\alpha - 1) \times S_p / \alpha$ as union input. In Phase 1, the layers of Fast ScanNet are inherited from the original model without changes before the union features can be normally calculated. Then, the standard convolutional/pooling layer should be replaced with corresponding anchor layer for feature reconstruction. The setting of operating distance d_a
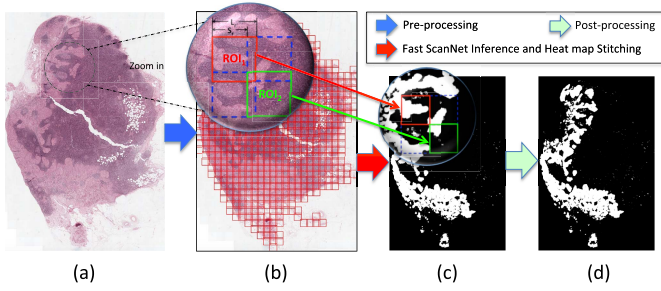
Fig. 9. The framework of whole-slide image processing. (a) The original WSI. (b) The pre-processed WSI, in which the squares are the ROIs selected from informative regions. (c) Whole-slide heat map, in which the red and green squares are the heat maps corresponding to the red and green denoted ROIs in the (b). (d) Final results of whole-slide heat map after post-processing.

should be equal to the dense coefficient α to reconstruct dense features. The setting of the anchor stride s_a should be equal to the original stride of standard convolutional/pooling layers. In Phase 2, all the convolutional/pooling operations should be replaced by AnchConv/AnchPool layers. The setting of operating distance d_a and dilation rate should be equal to the dense coefficient α to ensure the correct input fetching. The anchor stride should be formulated as $s_a = s_o \times \alpha$, where s_o is the original stride of standard convolution/pooling layers.

From the simple 1-D case in Fig. 7, the size of the heat map increases from 1 to 4. It should be noted that the Fast ScanNet can leverage the efficiency of the fully convolutional architecture in the all phases (including reconstruction and Phase 2) when fed with large sized inputs. We transform Fast ScanNet from the FCN with 2-D input (ROI) size as $L_r = 2868$, with the dense coefficient $\alpha = 2$. The input (ROI) of Fast ScanNet should be extended to $L'_r = 2884$ with $L_e = 16$ for dense inference accordingly. Finally, a 2^2 times denser heat map will be generated with size 166 compare to the original size 83

C. Whole-Slide Image Processing

Fig. 9 illustrates the pipeline for metastatic breast cancer detection from WSIs. We first employ a simple and effective OTSU method [54] for pre-processing to remove the non-informative regions of input WSI. Then we feed pre-processed ROIs into the Fast ScanNet to obtain the dense heat maps before they are stitched together for a whole-slide heat map (WSM). Finally, we post-process the image by morphology opening operations to remove the small outliers.

We crop the ROIs with extended size as $L'_r = 2884$ for ScanNet (with $\alpha = 2$, patch size $L_p = 244$ and scanning stride $S'_p = S_p/\alpha = 16$; equivalent to patch-based CNN with scanning stride $S'_p = 16$). The size of dense heat map in this configuration is $L'_m = 166$ (following rules in Eq.(1)). If the dense heat maps are stitched non-overlapping but conterminous, the inference on WSI level will be most efficient. To satisfy this point, the ROIs should be fetched under a certain interval $S_r = 2656$, denoted as ROI stride. The rules for ROI fetching are summarised as follows:

$$S_r = S'_p \times L'_m \qquad (2)$$

Sources	Train		Test	
	Normal	Tumor	Normal	Tumor
Radboud UMC	100	70	80	50
Utrecht UMC	60	40		
Total	160	110	80	50

The coordinates between the original WSI and WSM can be transformed by following:

$$H_I = H_M \times S'_p \qquad (3)$$

where H_I is the index of original WSI space, H_M is the index of WSM space.

The final results for *2016 Camelyon Grand Challenge* benchmarks are extracted from the WSM. For the localization task, each connected component in the binarized probability map (threshold was set as 0.5 emperically in our experiments) was considered as a detection, with score equal to the maximum probability within the region. For the WSI classification task, the prediction was simply computed as the maximal score within the slide without any sophisticated post-processing procedures.

III. EXPERIMENTS AND RESULTS

A. The Dataset

To evaluate our method, we used the dataset supplied as part of the Camelyon16 challenge in conjunction with ISBI 2016 [3], [55]. The overall aim of the challenge is to detect lymph node metastases in hematoxylin and eosin (H&E) stained whole-slide images of lymph node sections. This task is of high importance because micro-metastases are often missed and therefore WSIs may be falsely classified. Specifically, the two tasks of the challenge were to localize the tumor region and to perform whole-slide image classification. The Camelyon16 challenge dataset contains a total of 400 WSIs in TIF format that were collected from two independent centers: (i) Radboud University Medical Centre (Radboud UMC) and (ii) The University Medical Centre Utrecht (Utrecht UMC). The slides were acquired by 2 different scanners. Radboud UMC images were produced using a digital slide scanner (Pannoramic 250 Flash II; 3D Histach) with a 20x objective lens (level-0 pixel size, $0.243\mu m \times 0.243\mu m$). Utrecht UMC images were scanned by a digital slide scanner (NanoZoomer-XR Digital slide scanner C12000-01; Hamamatsu Photonics) with a 40x objective lens (level-0 pixel size, $0.226\mu m \times 0.226\mu m$). The dataset contains 160 WSIs that contain metastases and 110 normal WSIs. 40 abnormal and 28 normal WSIs were reserved as validation dataset to determine the hyper-parameters in post-processing including heat map threshold and morphological kernel size. The organizers held out 130 images for performance evaluation. Further details of the dataset can be seen in Table II.

B. Implementation Details

Our framework was implemented utilizing TensorFlow library on the workstation equipped with eight Geforce
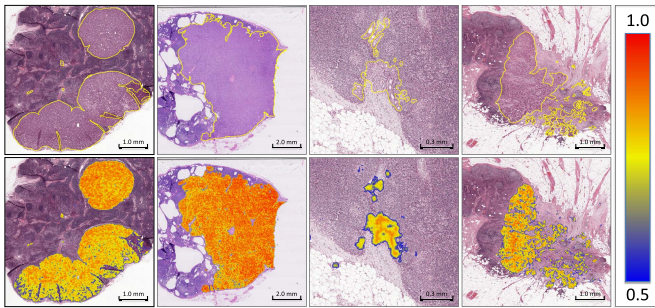
Fig. 10. Typical examples of metastasis detection results from our proposed method. Top: ground truth annotations from pathologists indicated by the yellow lines. Bottom: Our detection probability results overlaid on the original images with different colors.

GTX TITAN X GPU cards, Dual Inter Xeon(R) E5-2623 v4@ 2.60GHz CPUs, 512GB ECC Memory and 7TB SSD.

C. Evaluation Metrics

For performance evaluation, the Camelyon16 challenge employs two metrics to assess tumor region localization and WSI classification. For the task of tumor localization, the Free Response Operating Characteristic (FROC) curve [56] is used to determine the performance. The FROC curve is defined as the plot of sensitivity versus the average number of false-positives per image under different truncation of probability values. Specifically, the FROC score is defined as the average sensitivity at 6 predefined false positive rates: 0.25, 0.5, 1, 2, 4 and 8 false positives. For the task of WSI classification, the receiver operating curve (ROC) denotes the true positive rate versus false positive rate of image classification under different truncation of probability values. The area under the receiver operating curve [57], [58] is used for performance evaluation. This is also defined as the AUC score.

D. Qualitative Evaluation

To highlight the superior performance of our algorithm, we show some typical results of our framework in Fig. 10. Here, we can see a strong agreement between metastasis detection as predicted by our method and annotations from the experienced pathologists. We observe typical examples of macro-metasases within the first and second column in Fig. 10 with a diameter larger than $2mm$. We can see that the prediction of our proposed method on the bottom row of Fig. 10 has excellent agreement with the ground truth annotation on the top row of the same corresponding figure. We observe examples of cases with micro-metastases and isolated tumor cells (ITCs) with diameter smaller than $2mm$ in the third and fourth column in Fig. 10. Note that the ITCs were not considered for evaluation during this challenge. Similar to images with macro-metastases, we observe a strong agreement between the prediction of our method and the pathologist annotation. Visually, it is clear that our method is able to accurately detect both micro- and macro-metastases within the whole-slide images. This is of great clinical significance because the robustness of the algorithm to the size of the tumor

TABLE III
QUANTITATIVE COMPARISON WITH OTHER METHODS

Methods	FROC score	AUC score
Human performance	0.7325	0.9660
Fast ScanNet-16(ours)	**0.8533**	0.9875
HMS and MIT	0.8074	**0.9935**
HMS, Gordan Center, MGH	0.7600	0.9763
Radboud Uni. (DIAG)	0.5748	0.7786
EXB Research co.	0.5111	0.9156
Middle East Tech. Uni.	0.3889	0.8642
University of Toronto	0.3822	0.8149
DeepCare Inc	0.2439	0.8833
NLP LOGIX co. USA	0.3859	0.8298
ScanNet-32(w/o HNM)	0.7030	0.9415
ScanNet-32 [48]	0.8133	0.9669

reduces the likelihood that the algorithm will misdiagnose cases with only a small tumorous region.

E. Quantitative Evaluation and Comparison

To quantify the performance of our algorithm, we evaluate the method under various configurations. α is the parameter that controls how dense the output of the prediction will be. We set the value of α as 1 and 2 and name the corresponding networks "ScanNet-32" and "Fast ScanNet-16", respectively. We note from Table III that "Fast ScanNet-16" achieves a superior result to "ScanNet-32", particularly in the tumor localization task, because the algorithm is able to output a more dense prediction. This dense prediction is very important when detecting micro-metastases and therefore also contributes to a greater overall WSI classification performance.

Furthermore, we compared our method with several state-of-the-art methods, submitted as part of the challenge in Table III. We also show the FROC and ROC curves from the different methods in Fig. 11. We can be confident in the performance of our algorithm because there were many submissions to the Camelyon16 challenge from various top universities and companies. For the task of tumor localization, our proposed method achieved the best performance compared to all other methods, with the highest FROC score of 0.8533 outperforming the runner-up team [41] by a significantly large margin of 4.6%. This highlights the ability for our method to accurately detect various sized metastases in breast lymph node biosies. We also observe that the performance of the proposed method surpassed the pathologist performance of 0.7325 by more than 12%. As a result, our method not only provides an objective method, but also generates more accurate localization results. For the task of WSI classification, the AUC score of the proposed method was 0.9875 when we applied a simple post-processing technique on the tumor localization probability map. Similarly, this score also surpassed the pathologist performance of 0.9660 from the pathologists, and also compares favorably to the leading method of 0.9935 [41].

F. Runtime Comparison

It is clear that our proposed method achieves excellent performance in lymph node metastasis detection, but run-time
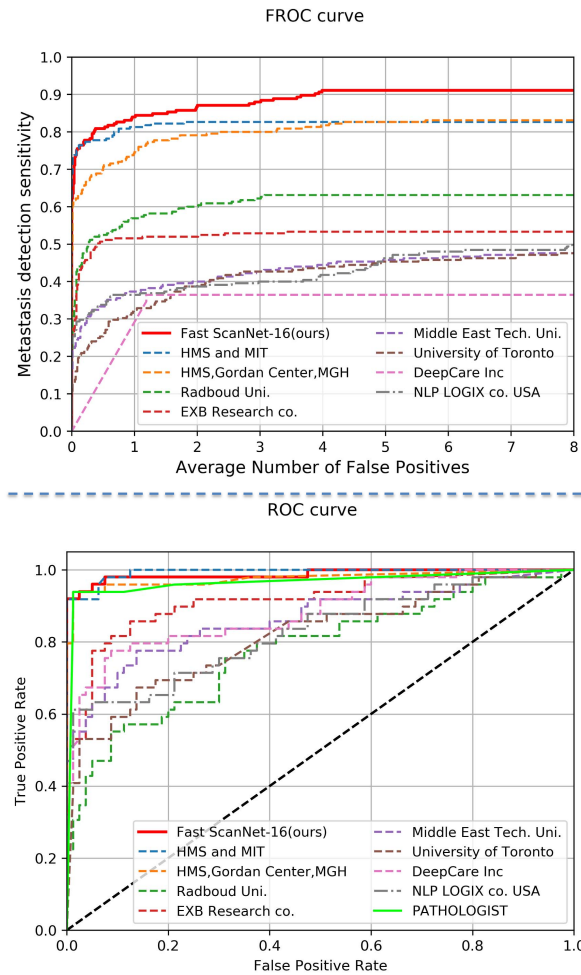
Fig. 11. Evaluation results. Top: FROC curves of tumor localization task of different methods. Bottom: ROC curves of WSI classification task of different methods.

TABLE IV
RUNTIME COMPARISON ON THE ROI (SIZE 2800×2800) (UNIT: MINUTE)

Network	Methods	stride 32	stride 16
Fast ScanNet	ours	**0.0182**	**0.0200**
ScanNet [48]	ours	**0.0182**	0.0730
GoogleNet	HMS and MIT	0.6683	2.6734
ResNet-34	EXB Research co.	0.7240	2.8962
AlexNet	NLP LOGIX co.	0.3342	1.3367
VGG16(patch-based)	-	0.9747	3.8987

the advantage of our Fast ScanNet is significant, especially on the densely scanning cases with stride as 16.

A typical WSI consists of 100 to 900 informative ROIs (400 in average) with size 2800×2800, i.e., it generally takes about one minute for our Fast ScanNet-16 to process a WSI by using the workstation equipped with eight Geforce GTX TITAN X GPU cards. The Fast ScanNet can be easily paralleled for acceleration if more GPU resources are available.

IV. DISCUSSION

Histopathological examination of tissue biopsies plays a fundamental role in cancer diagnosis and survival analysis [4]. Despite the importance of this task, providing a diagnosis can be extremely time-consuming because it is vital that pathologists thoroughly examine each case to ensure an accurate diagnosis. There are good reasons to believe that digital pathology with artificial intelligence for computer-aided diagnosis is a remedy for this dilemma because it can not only reduce the workload of the pathologists, but also give an objective and potentially more accurate diagnosis.

However, there is limited literature on deep learning applied to gigapixel whole-slide images before the *Camelyon16* challenge. In the past, most methods performed image analysis on pre-selected patches (size 500×500) manually chosen by experts. When using computational methods to analyze gigapixel whole-slide images, several challenges have to be addressed: the huge variance of tissue pattern; the severe sample imbalance of each class; and the demanding computational cost. The latter point is a major bottleneck and is one of the reasons why many current algorithms cannot be implemented in clinical practice. In this paper, we propose a method that, not only achieves state-of-the-art performance in lymph node metastasis detection, but also overcomes the major speed bottleneck within whole-slide image processing. Our framework can make fast predictions by leveraging the efficiency of FCN, while it is trained in a patch-based way for extensive augmentation. We propose the novel idea of using anchor layers, which allows dense scanning with little extra cost and achieves one hundred times acceleration in comparison with other methods.

Although our method achieved competitive performance in the underlying task, there are several directions that can be explored to further improve results. First, more powerful network architectures or sophisticated loss functions can be utilized to enhance the discrimination of models, such as

consideration is crucial if these methods are to be implemented in clinical practice. In order to evaluate the efficiency, we compared our Fast ScanNet with our previous work ScanNet by using the stitching mechanism for dense reconstruction [48] and several patch-based frameworks widely used by other teams. The efficiency is evaluated by measuring the time cost on a 2800×2800 sized ROI with scanning stride 32 and 16 (corresponding to dense coefficient $\alpha = 1$ or 2) under the setting of single GPU. From Table IV, comparing the time consumptions of Fast ScanNet under configurations of $\alpha = 1$ and 2, we observe that it only takes approximately 10% extra time cost to obtain a four times denser heat map for Fast ScanNet. In comparison with ScanNet when the stride is set as 16, the proposed Fast ScanNet can process images 3.65 times faster and therefore is suitable when processing multi gigapixel histology images, where timely diagnosis results are of crucial importance. Compared with the patch-based VGG16 network, the Fast ScanNet only takes 0.02 minutes to process the ROI with stride 16, about 195 times faster than its patch-based baseline. Actually, it amounts to predict 130 million patches in one minute per GPU (one pixel in heat map corresponds to one 244×244 patch). In contrast to other patch-based frameworks,

residual architecture [59], focal loss [60], etc. As well as this, we can explore further ways to increase the computational efficiency in our current method. A potential avenue would involve reducing the amount of convolutional operations using network pruning. In addition, in the whole-slide image classification task, the probabilities of image-level classification are simply computed as the maximal score in the slide without any sophisticated post-processing procedures, which might limit its performance in the classification task. Therefore, more sophisticated hand-crafted features derived from clinical insight in combination with deep learned features might potentially improve the performance on the classification task.

Deep learning for image analysis usually requires lots of data with high-quality annotations. However, the exhaustive pixel-wise annotation of WSIs is very laborious and time-consuming (about 1-2 hours for a pathologist to annotate one typical WSI), hence it is very difficult to acquire a very large dataset with precise annotations. One interesting research direction for whole-slide image analysis is how to make use of weakly supervised information. For example, the WSIs with rough annotations (i.e., some pixel-level boundaries are not exactly accurate) or only image-level labels are often easily accessible. Hence, exploring how we can capitalize on this information using neural networks is a very interesting path to follow within the field of histopathology image analysis.

V. CONCLUSIONS

In this paper, we present a novel framework that overcomes the major speed bottleneck in whole-slide image analysis, by leveraging fully convolutional networks for efficient inference. The framework reconstructs dense heat maps for ensuring accurate detection on both micro- and macro-metastases. By incorporating asynchronous sample prefetching and hard negative mining, the network can be trained with quite good discriminative ability. We demonstrate that the proposed method achieved superior performance compared to other state-of-the-art methods on the Camelyon 2016 Grand Challenge dataset and even surpassed human performance. Furthermore, the proposed method matched the speed requirements of clinical practice, where the framework can process whole-slide image within one minute. Future investigations include evaluating our method on more histology WSIs and integrating the framework into clinical practice. This may require a large amount of data from multi-centers to improve the generalization capability of the method.

REFERENCES

[1] *GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012*, Int. Agency Res. Cancer, World Health Org., Geneva, Switzerland, 2012, vol. 2.

[2] S. B. Edge and C. C. Compton, "The american joint committee on cancer: The 7th edition of the *AJCC cancer staging manual* and the future of TNM," *Ann. Surgical Oncol.*, vol. 17, no. 6, pp. 1471–1474, 2010.

[3] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017.

[4] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.

[5] G. Humphreys and A. Ghent, "World laments loss of pathology service," *Bull. World Health Org.*, vol. 88, no. 8, pp. 564–565, 2010.

[6] M. T. McCann, J. A. Ozolek, C. A. Castro, B. Parvin, and J. Kovacevic, "Automated histology analysis: Opportunities for signal processing," *IEEE Signal Process. Mag.*, vol. 32, no. 1, pp. 78–87, Jan. 2015.

[7] J. M. S. Prewitt and M. L. Mendelsohn, "The analysis of cell images," *Ann. New York Acad. Sci.*, vol. 128, no. 3, pp. 1035–1053, 1966.

[8] M. Veta *et al.*, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Med. Image Anal.*, vol. 20, no. 1, pp. 237–248, 2015.

[9] E. Cosatto, M. Miller, H. P. Graf, and J. S. Meyer, "Grading nuclear pleomorphism on histological micrographs," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[10] H. Fatakdawala *et al.*, "Expectation–maximization-driven geodesic active contour with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1676–1689, Jul. 2010.

[11] X. Qi, F. Xing, D. J. Foran, and L. Yang, "Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 754–765, Mar. 2012.

[12] M. Kuse, Y.-F. Wang, V. Kalasannavar, M. Khan, and N. Rajpoot, "Local isotropic phase symmetry measure for detection of beta cells and lymphocytes," *J. Pathol. Inform.*, vol. 2, Jan. 2012, doi: 10.4103/2153-3539.92028.

[13] K. Sirinukunwattana, D. R. J. Snead, and N. M. Rajpoot, "A novel texture descriptor for detection of glandular structures in colon histology images," *Proc. SPIE*, vol. 9420, p. 94200S, Mar. 2015, doi: 10.1117/12.2082010.

[14] C. Jung, C. Kim, S. W. Chae, and S. Oh, "Unsupervised segmentation of overlapped nuclei using Bayesian classification," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2825–2832, Dec. 2010.

[15] J. P. Vink, M. B. Van Leeuwen, C. H. M. van Deurzen, and G. De Haan, "Efficient nucleus detector in histopathology images," *J. Microsc.*, vol. 249, no. 2, pp. 124–135, 2012.

[16] J. Xu, A. Janowczyk, S. Chandran, and A. Madabhushi, "A weighted mean shift, normalized cuts initialized color gradient based geodesic active contour model: Applications to histopathology image segmentation," *Proc. SPIE*, vol. 7623, p. 76230Y, Mar. 2010, doi: 10.1117/12.845602.

[17] A. Basavanhally *et al.*, "Incorporating domain knowledge for tubule detection in breast histopathology using O'Callaghan neighborhoods," *Proc. SPIE*, vol. 7963, p. 796310, Mar. 2011, doi: 10.1117/12.878092.

[18] A. Tabesh *et al.*, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1366–1378, Oct. 2007.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.

[20] Z. Yu *et al.*, "Melanoma recognition in dermoscopy images via aggregated deep convolutional features," *IEEE Trans. Biomed. Eng.*, to be published, doi: 10.1109/TBME.2018.2866166.

[21] Z. Yu *et al.*, "A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 3, pp. 874–885, May 2018.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. New York, NY, USA: Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[24] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3D deeply supervised network for automatic liver segmentation from CT volumes," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. New York, NY, USA: Springer, 2016, pp. 149–157, doi: 10.1007/978-3-319-46723-8_18.

[25] H. Li *et al.*, "Dense deconvolutional network for skin lesion segmentation," *IEEE J. Biomed. Health Inform.*, to be published, doi: 10.1109/JBHI.2018.2859898.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[28] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[29] J. Xu *et al.*, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[30] H. Su, F. Liu, Y. Xie, F. Xing, S. Meyyappan, and L. Yang, "Region segmentation in histopathological breast cancer images using deep convolutional neural network," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 55–58.

[31] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. New York, NY, USA: Springer, 2013, pp. 411–418, doi: 10.1007/978-3-642-40763-5_51.

[32] L. Roux *et al.*, "Miitosis detection in breast cancer histology images via deep cascaded nest," *J. Pathol. Inform.*, vol. 4, p. 8, May 2013.

[33] H. Chen, Q. Dou, X. Wang, J. Qin, and P.-A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1160–1166.

[34] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, "Beyond classification: Structured regression for robust cell detection using convolutional neural network," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. New York, NY, USA: Springer, 2015, pp. 358–365, doi: 10.1007/978-3-319-24574-4_43.

[35] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.

[36] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Med. Image Anal.*, vol. 36, pp. 135–146, Feb. 2017.

[37] K. Sirinukunwattana *et al.*, "Gland segmentation in colon histology images: The glas challenge contest," *Med. Image Anal.*, vol. 35, pp. 489–502, Jan. 2017.

[38] Z. Jia, X. Huang, E. I-C. Chang, and Y. Xu, "Constrained deep weak supervision for histopathology image segmentation," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2376–2388, Nov. 2017.

[39] F. Xing, Y. Xie, H. Su, F. Liu, and L. Yang, "Deep learning in microscopy image analysis: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4550–4568, Oct. 2018.

[40] T. Qaiser *et al.*, "HER2 challenge contest: A detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues," *Histopathology*, vol. 72, no. 2, pp. 227–238, 2018.

[41] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck. (2016). "Deep learning for identifying metastatic breast cancer." [Online]. Available: https://arxiv.org/abs/1606.05718

[42] B. Kong, X. Wang, Z. Li, Q. Song, and S. Zhang, "Cancer metastasis detection via spatially structured deep network," in *Information Processing in Medical Imaging—IPMI*. New York, NY, USA: Springer, 2017, pp. 236–248, doi: 10.1007/978-3-319-59050-9_19.

[43] G. Aresta *et al.* (2018). "Bach: Grand challenge on breast cancer histology images." [Online]. Available: https://arxiv.org/abs/1808.04277

[44] M. Veta *et al.* (2018). "Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge." [Online]. Available: https://arxiv.org/abs/1807.08284

[45] S. Graham, M. Shaban, T. Qaiser, N. A. Koohbanani, S. A. Khurram, and N. Rajpoot, "Classification of lung cancer histology images using patch-level summary statistics," *Proc. SPIE*, vol. 10581, p. 1058119, Mar. 2018, doi: 10.1117/12.2293855.

[46] T. Qaiser, K. Sirinukunwattana, K. Nakane, Y.-W. Tsang, D. Epstein, and N. Rajpoot, "Persistent homology for fast tumor segmentation in whole slide histology images," *Procedia Comput. Sci.*, vol. 90, pp. 119–124, Jul. 2016, doi: 10.1016/j.procs.2016.07.033.

[47] Z. Xu and J. Huang, "Detecting 10,000 cells in one second," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. New York, NY, USA: Springer, 2016, pp. 676–684, doi: 10.1007/978-3-319-46723-8_78.

[48] H. Lin, H. Chen, Q. Dou, L. Wang, J. Qin, and P.-A. Heng, "ScanNet: A fast and dense scanning framework for metastastic breast cancer detection from whole-slide image," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 539–546.

[49] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. (2013). "OverFeat: Integrated recognition, localization and detection using convolutional networks." [Online]. Available: https://arxiv.org/abs/1312.6229

[50] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[51] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[52] H. Chen *et al.*, "Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. New York, NY, USA: Springer, 2015, pp. 507–514, doi: 10.1007/978-3-319-24553-9_62.

[53] F. Yu and V. Koltun. (2015). "Multi-scale context aggregation by dilated convolutions." [Online]. Available: https://arxiv.org/abs/1511.07122

[54] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, nos. 285–296, pp. 23–27, 1975.

[55] P. Bándi *et al.*, "From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge," *IEEE Trans. Med. Imag.*, to be published, doi: 10.1109/TMI.2018.2867350.

[56] D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Med. Phys.*, vol. 16, no. 4, pp. 561–568, Jul. 1989.

[57] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[58] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, 1993.

[59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[60] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2018.2858826.