PFA-ScanNet: Pyramidal Feature Aggregation with Synergistic Learning for Breast Cancer Metastasis Analysis

Zixu Zhao¹, Huangjing Lin¹, Hao Chen², and Pheng-Ann Heng¹

¹ Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China
² Imsight Medical Technology, Co., Ltd. Hong Kong SAR, China

Abstract. Automatic detection of cancer metastasis from whole slide images (WSIs) is a crucial step for following patient staging as well as prognosis. However, recent convolutional neural network (CNN) based approaches are struggling with the trade-off between accuracy and computation cost due to the difficulty in processing large-scale gigapixel images. To address this challenge, we propose a novel deep neural network, namely Pyramidal Feature Aggregation ScanNet (PFA-ScanNet) with pyramidal feature aggregation in both top-down and bottom-up paths. The discrimination capability of our detector is increased by leveraging the merit of contextual and spatial information from multi-scale features with larger receptive fields and less parameters. We also develop an extra decoder branch to synergistically learn the semantic information along with the detector, significantly improving the performance in recognizing the metastasis. Furthermore, a high-efficiency inference mechanism is designed with dense pooling layers, which allows dense and fast scanning for gigapixel WSI analysis. Our approach achieved the state-of-the-art FROC score of 89.1% on the Camelyon16 dataset, as well as competitive kappa score of 0.905 on the Camelyon17 leaderboard. In addition, our proposed method shows leading speed advantage over the state-of-the-art methods, which makes automatic analysis of breast cancer metastasis more applicable in the clinical usage.

1 Introduction

The prognosis of breast cancer mainly focuses on grading the stage of cancer, which is measured by the tumor, node, and distant metastasis (TNM) staging system [1]. With the boosting progress in high-throughput scanning and artificial intelligence technology, automatic detection of breast cancer metastasis in sentinel lymph nodes has great potential in cancer staging to assist clinical management. The algorithm is expected to detect the presence of metastases in five slides with lymphatic tissues dissected from a patient, and measure their extent to four metastasis categories and finally grade the pathologic N stage (pN-stage)
following the TNM staging system. However, the task is challenging due to several factors: (1) the difficulty in handling large-scale gigapixel images (e.g., 1-3 GB per slide); (2) the existence of hard mimics between normal and cancerous region; (3) the significant size variance among different metastasis categories.

Recently, many deep learning based methods adopt patch-based models to directly analyze whole slide images (WSIs) \cite{2,3,4,5}. The most common way is to extract small patches in a sliding window manner and feed them to the model for inference. For example, ResNet-101 and Inception-v3 are leveraged as the backbone of detectors in \cite{2} and \cite{3}, bringing the detection results to 85.5% and 88.5% with regard to FROC on Camelyon16\cite{1} dataset, respectively. However, the patch-based inference leads to dramatically increased computational cost when applied to gigapixel WSI analysis, which are not applicable in clinical usage. To reduce the computational burden, Kong et al. \cite{6} utilized a lightweight network (student network) supervised by a large capacity network (teacher network) with transfer learning. Also, Lin et al. proposed a modified fully convolutional network (FCN), namely Fast ScanNet \cite{7} to overcome the speed bottleneck by allowing dense scanning in anchor layers. These scan-based models are yet relatively simple in structure so that encoding multi-scale features is beyond attainment. Therefore, the discrimination capability of scan-based models is still limited.

Another challenging problem of lymph node classification lies in how to effectively retrieve tiny metastasis, i.e., ITC (< 0.2mm) and micro-metastasis (< 2mm), while rejecting most of the hard mimics. Several methods \cite{5,7} circumvent false positives via hard negative mining, which focus on the most challenging negative patches. This seems to benefit the performance overall but decreases the sensitivity on small ITC lesions remarkably. Furthermore, it may disintegrate the prediction into pieces due to mimic patches existed in metastatic regions, leading to inaccurate evaluation on metastasis size. To tackle this issue, Li et al. proposed a neural conditional random field (NCRF) deep learning framework \cite{4} combining with hard negative mining. Although spatial correlations is considered, it still achieved limited performance on metastasis detection. In terms of this problem, extra guidance is introduced in our method to make full use of the semantic information that helps the model distinguish hard mimics.

Aiming at developing a detection system as accurate as possible while maintaining the efficiency, we propose a novel Pyramidial Feature Aggregation ScanNet (PFA-ScanNet). Our contributions are threefold: (1) We propose to aggregate pyramidal features in a top-down path to increase the discrimination capability of the detector by leveraging the contextual and spatial information. Parameter-efficient Feature Extraction (PFE) module is raised to refine pyramidal features by extracting informative regions with larger receptive fields and less parameters. (2) A high-efficiency inference mechanism is carefully designed with dense pooling layers. It allows the detector to take large sized images as input for inference while being trained in a flexible patch-based fashion. (3) We develop an extra decoder branch to synergistically learn the semantic information from pyramidal features aggregated in a bottom-up path. The semantic guidance with

\footnote{http://camelyon16.grand-challenge.org/}
synergistic learning improves the ability to retrieve metastasis with significantly different size. The effectiveness of our PFA-ScanNet is validated on both the public Camelyon16 and Camelyon17 dataset.

2 Method

The proposed PFA-ScanNet is a scan-based fully convolutional network consisting of a main detector for classification and an extra decoder for segmentation. As shown in Fig. 1 Parameter-efficient Feature Extraction (PFE) modules are integrated into our network at each feature level in order to extract informative features with large receptive fields and less parameters. Our network leverages the merit of contextual and spatial information from pyramidal features by aggregating them in both top-down and bottom-up paths. Besides, our method conducts synergistic learning between the detector and decoder to collaboratively learn the semantic information from the aggregated features.

2.1 Pyramidal Feature Aggregation for Accurate Classification

Inspired by the feature pyramid network [8], we propose to make full use of pyramidal features by hierarchically aggregating them in a top-down path for accuracy classification. We firstly raise the Parameter-efficient Feature Extraction (PFE) module to extract informative regions with larger receptive fields and less computation cost from pyramidal features. It also benefits the fast inference in section 2.2. Fig. 2(a) shows the detailed structure of PFE. Let \( \{ X_i \} \) denotes the pyramidal feature generated by the detector at feature level \( i \) (\( i = 2, 3, 4, 5 \)).
$X_i$ is firstly passed through a global convolution with a large kernel to enlarge the receptive fields and reduce the feature map number. To further reduce the computation burden and number of parameters, we employ symmetric and separable large filters, which is a combination of $1 \times 15 + 15 \times 1$ and $15 \times 1 + 1 \times 15$ convolutions instead of directly using larger kernel of size $15 \times 15$. The most informative area is then cropped from the refined feature $X_i'$. Specifically, as for $\{X_3', X_4', X_5'\}$ created in three PFEs, regions with $\{1/4, 1/2, 1\}$ size of them are cropped to capture different receptive fields before average pooling layer. The generated informative feature $M_i$ is added with $M_{i+1}$ from a higher feature level and finally passed through a $1 \times 1$ convolution with softmax activation to predict the probability. In this way, our detector can efficiently encode pyramidal features and thus present a strong discrimination capability.

![Diagram](image-url)

**Fig. 2.** Detail structure of (a) Parameter-efficient Feature Extraction Module (PFE) and (b) Boundary-aware Module (BM) in proposed PFA-ScanNet.

### 2.2 WSI Processing with High-efficiency Inference

To meet the speed requirement, we inherit the FCN-structured network in Fast ScanNet as the trunk of our detector but remove the last three fully convolutional layers which take large computation costs in the inference phase. Unlike the anchor layer in which the convolutional kernel jumps among different anchor positions [7], we propose a dense pooling layer in PFE which allows dense scanning with little extra cost. A dense coefficient $\alpha$ is introduced in the dense pooling layer to control the pooling strides of average pooling operation. The pooling strides are $\{128, 64, 32\}$ associated with feature level $\{3, 4, 5\}$ in the training phase and will be converted to $\{128/\alpha, 64/\alpha, 32/\alpha\}$ in the inference phase. In this way, it allows dense and fast predictions when $\alpha$ increases in the form of $\alpha = 2^n \times 16$ ($n = 0, 1, 2, \ldots$). Accordingly, our network can take region of interests (ROIs) with a size of $L_R$ as input for inference while being trained with small patches with a size of $L_p$ for extensive augmentation. In other words, our network inherently falls into the category of FCN architecture, which is equivalent to a patch-based CNN with input size $L_p$ and scanning stride $S_p$, but the inference speed becomes much faster by removing redundant computations of overlaps. To better understand this mechanism, we denote the scanning stride for refetching ROIs as $S_R$ and size of the predicted probability tile as $L_m$, and
summarize the rules for high-efficiency inference as follows:

\[
\begin{align*}
L_R &= L_p + (L_m - 1) \times (S_p/\alpha), \\
S_R &= (S_p/\alpha) \times L_m,
\end{align*}
\] (1)

2.3 Semantic Guidance with Synergistic Learning

Given that the surrounding tissue region is helpful to determine whether the small patch is metastasis or not, we develop our network with an extra decoder branch to synergistically learn the semantic information along with the detector. In the decoder, feature map \(X'_i\) generated in PFE is firstly passed through a Boundary-aware Module (BM) to refine the boundary of the metastatic region. As shown in Fig. 2(b), BM models the boundary alignment in a residual structure [9], which merges the feature map \(X'_i\) after two \(3 \times 3\) convolutions with itself by summation operation to take advantage of the local contextual information and localization cue. Afterwards, the generated feature is upsampled with a deconvolution layer and then added with feature map \(X'_{i-1}\) of higher resolution to generate new score maps in a bottom-up path. Deep supervision is injected to specific layers to learn the multi-level semantic information, which can also speed up the convergence rate.

Synergistic learning is conducted by training the detector and decoder simultaneously. Nevertheless, it is hard to minimize the classification loss and segmentation loss simultaneously in one iteration caused by the misalignment between two specific tasks. Specifically, the mislabelled region and zigzag boundaries in WSI annotations have the tendency to overwhelm other informative regions in segmentation loss calculation and thus dominate the gradients. To solve the problem, we modify the binary cross-entropy loss into a truncated form [10] that can reduce the contribution of outliers with high confidence prediction. Our segmentation loss is shown as follows:

\[
\mathcal{L}_{seg}(\mathcal{X}; W) = \begin{cases} 
\sum_{x \in \mathcal{X}} \sum_{t \in \{0, 1\}} \left( -\log(\gamma) + \frac{1}{2} \left( 1 - \frac{p(t|x; W)}{\gamma} \right) \right), & p(t|x; W) < \gamma \\
\sum_{x \in \mathcal{X}} \sum_{t \in \{0, 1\}} -\log(p(t|x; W)), & p(t|x; W) \geq \gamma
\end{cases}
\] (2)

where \(W\) denotes parameters of our model, \(\mathcal{X}\) denotes the training patches, and \(p(t|x; W)\) is the predicted probability for the ground truth label \(t\) given the input pixel \(x\). The segmentation loss will clip outliers at the truncated point \(\gamma \in [0, 0.5]\) when \(p(t|x; W) < \gamma\), while preserving the loss value for others. Therefore, it can ease the gradient domination and benefit the learning of informative regions. When \(\gamma = 0\), it will degrade into binary cross-entropy. Meanwhile, we directly employ the binary cross-entropy loss as our classification loss to train the detector. Let \(W_d\) denote parameters in the detector and \(\lambda\) be the trade-off hyperparameter, the overall loss function for synergistic learning is defined as:

\[
\mathcal{L}_{total}(\mathcal{X}; W) = \mathcal{L}_{cla}(\mathcal{X}; W_d) + \lambda \mathcal{L}_{seg}(\mathcal{X}; W)
\] (3)
2.4 Overall Framework for pN-stage Classification

The overall pipeline of our framework contains: (1) Data Preprocessing. We first extract informative tissue regions from WSIs with Otsu algorithm \[11\]. Training patches and corresponding mask patches are augmented with random flipping, scaling, rotation, and cropping together. Color jittering and HSV augmentation are applied to training patches to overcome color variance. (2) Slide-level Metastasis Detection. Unlike hard negative mining that focus on false negative samples with multi-stage inference, we choose to extract ITC and boundary patches at first and add them to the original training set to better train the PFA-ScanNet with synergistic learning. This method helps the segmentation model converge faster and learn valid information rather than focusing on simple cancerous patches. (3) Patient-level pN-stage Classification. Morphological features such as major axis length and region area are extracted from the probability maps to formulate feature vectors, which are utilized to train a random forest classifier. The lymph node category is then classified into four types, i.e., normal, ITC, micro, and macro. The patient’s pN-stage is finally determined by the given rules in Chamelyon17 Challenge\[2\].

3 Experimental Results

Datasets. We evaluate our method on Camelyon16 and Camelyon17 datasets provided in Camelyon17 Challenge. The Camelyon16 dataset contains a total of 400 WSIs with 160 cancerous WSIs and 110 normal WSIs. The Camelyon17 dataset contains 1000 WSIs with 5 slides per patient (500 training and 500 testing), providing only 50 slides with lesion-level annotations. For Camelyon16 Challenge, our network is trained with only Camelyon16 training set (270 WSIs) for fair comparison with other methods. For Camelyon17 Challenge, we use the whole Camelyon16 dataset and 215 slides (43 patients) including 50 slides with lesion-level annotations from Camelyon17 training set to train the network.

Implementation Details. We implement our network using TensorFlow library on the workstation equipped with four NVIDIA TITAN Xp GPUs. The sizes of training patches and mask patches are 692 × 692 and 512 × 512, respectively. Our model takes ROIs with a size up to 2708 × 2708 (determined by the memory capacity of GPU) for inference and outputs a 64 × 64 sized probability tile. To maximize the performance of synergistic learning, we set the truncated point γ in Equation (2) as 0.04. The hyperparameter λ is set to 0.5 in Equation (3). SGD optimizer is used to optimize the whole network with momentum of 0.9 and learning rate is initialized as 0.0001.

Evaluation Metrics. We adopt two metrics provided in Camelyon16 Challenge to evaluate slide-level metastasis detection, including AUC and average FROC. The latter is an average sensitivity at 6 false positive rates: 1/4, 1/2, 1, 2, 4, and 8 per WSI. For pN-stage classification, we utilize quadratic weighted Cohen’s kappa provided in Camelyon17 Challenge as the evaluation metric.

Quantitative Evaluation and Comparison. We validate our method on

http://camelyon17.grand-challenge.org/
Camelyon16 testing set with ground-truths held out and Camelyon17 testing set. Results of Camelyon17 Challenge are provided by organizers. Table 1 compares our method with top-ranking teams as well as the state-of-the-art methods. It is observed that our method without synergistic learning (PFA-ScanNet w/o SL) achieves striking improvements (14% in kappa score and 3% in FROC) compared with the advanced method (Fast ScanNet), demonstrating the superiority of aggregating pyramidal features with distinctive receptive fields. After introducing the synergistic learning, our PFA-ScanNet boosts the results to 89.1% with regard to FROC (1st) and 99.2% in terms of AUC on Camelyon16 testing set, outperforming the state-of-the-art methods. We also achieve competitive kappa score of 0.905 on Camelyon17 testing set without model ensembles, surpassing the Camelyon17 winner (Lunit Inc.). For the speed performance, we measure the time cost of each method on a 2708 × 2708 sized ROI with scanning stride of 32 (corresponding to the dense coefficient \(\alpha = 16\)) using one single GPU. As illustrated in Table 1, our method shows leading speed advantages over the state-of-the-art methods on Camelyon16 and Camelyon17 Challenges. It takes only 2% time of the Camelyon17 winner (Lunit Inc.) to obtain probability maps and achieves higher kappa score. Besides, our method (a large capacity network) is on par with B.Kong’s method (a lightweight network) [6] in terms of speed performance while achieving notably higher accuracy on detection results.

### Qualitative Analysis

Fig. 3 visualizes metastasis detection results of five typical cases. As we can observe, the proposed PFA-ScanNet can generate more accurate results compared with other two methods (PFA-ScanNet w/o SL and Fast ScanNet), which increase the ability to detect macro- and micro-metastases. Besides, the challenging ITC case can be detected by our method with few false positives, highlighting the advantage of the proposed PFA-ScanNet with synergistic learning.

### 4 Conclusions

Automatic cancer metastasis analysis is essential for cancer staging and following patient’s treatment. In this paper, we propose the PFA-ScanNet with synergistic learning for metastasis detection and pN-stage classification to improve the accuracy close to clinical usage while maintaining the computational efficiency.

### Table 1. Comparison with different approaches on Camelyon16 and Camelyon17 testing sets. Note that runtime (unit: minute) is compared on the ROI (size: 2708 × 2708).

| Method                          | Camelyon16 testing set | Camelyon17 testing set |
|--------------------------------|------------------------|------------------------|
|                                | Runtime                | AUC                    | FROC                  | Team                  | Runtime | Kappa Score |
| Harvard & MIT [5]              | 0.668                  | 0.994                  | 0.807                 | MIL-GPAT              | 0.247   | 0.857       |
| Y.Li et al. (NCRF) [11]        | 0.743                  | 0.910                  | DeepBio               | 1.583                 | 0.864   |
| Lunit Inc. [2]                 | 1.136                  | 0.985                  | 0.855                 | VCA-TUe               | 1.162   | 0.873       |
| B.Kong et al. [7]              | 0.014                  | -                      | 0.856                 | HMS-MGH-CCDS          | 0.067   | 0.881       |
| Y.Liu et al. (LYNA) [12]       | 1.155                  | 0.993                  | 0.861                 | ContextVision         | 1.636   | 0.883       |
| Y.Liu et al. [3]               | 1.155                  | 0.977                  | 0.885                 | Lunit Inc. [2]        | 1.136   | 0.899       |
| Fast ScanNet [7]               | 0.020                  | 0.987                  | 0.853                 | Fast ScanNet          | 0.020   | 0.718       |
| PFA-ScanNet w/o SL             | 0.018                  | 0.983                  | 0.878                 | PFA-ScanNet w/o SL    | 0.018   | 0.887       |
| PFA-ScanNet (Ours)             | 0.018                  | 0.992                  | 0.891                 | PFA-ScanNet (Ours)    | 0.018   | 0.905       |
Fig. 3. Typical examples of probability maps. The colors ranging from blue to red denote low to high probability. The lesion-level annotation is shown in black.

Competitive results have been demonstrated on the Camelyon16 and Camelyon17 datasets with a much faster speed. Inherently our method can be applied to a wide range of medical image classification tasks to boost the analysis of gigapixel WSIs.

References

1. Leslie H Sobin and Irvin D Fleming, “Tum classification of malignant tumors, (1997),” Cancer: Interdisciplinary International Journal of the American Cancer Society, vol. 80, no. 9, pp. 1803–1804, 1997.
2. Byungjae Lee and Kyunghyun Paeng, “A robust and effective approach towards accurate metastasis detection and pn-stage classification in breast cancer,” in MICCAI, 2018.
3. Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E Dahl, Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q Nelson, Greg S Corrado, et al., “Detecting cancer metastases on gigapixel pathology images,” arXiv preprint arXiv:1703.02442, 2017.
4. Yi Li and Wei Ping, “Cancer metastasis detection with neural conditional random field,” arXiv preprint arXiv:1806.07064, 2018.
5. Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck, “Deep learning for identifying metastatic breast cancer,” arXiv preprint arXiv:1606.05718, 2016.
6. Bin Kong, Shanhui Sun, Xin Wang, Qi Song, and Shaoting Zhang, “Invasive cancer detection utilizing compressed convolutional neural network and transfer learning,” in MICCAI, 2018.
7. H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, and P. Heng, “Fast scannet: Fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection,” IEEE Transactions on Medical Imaging, pp. 1–1, 2019.
8. Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, “Feature pyramid networks for object detection,” in CVPR, 2017.
9. Chao Peng, Xiangyu Zhang, Gang Yu, Guiming Luo, and Jian Sun, “Large kernel matters – improve semantic segmentation by global convolutional network,” in CVPR, 2017.
10. Yanning Zhou, Omer Fahri Onder, Qi Dou, Efstratios Tsougenis, Hao Chen, and Pheng-Ann Heng, “Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation,” arXiv preprint arXiv:1903.05358, 2019.

11. Nobuyuki Otsu, “A threshold selection method from gray-level histograms,” IEEE transactions on systems, man, and cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

12. Y Liu, T Kohlberger, M Norouzi, GE Dahl, JL Smith, A Mohtashamian, N Olson, LH Peng, JD Hipp, and MC Stumpe, “Artificial intelligence-based breast cancer nodal metastasis detection,” Archives of pathology & laboratory medicine, 2018.