

# FastNet: A Lightweight Convolutional Neural Network for Tumors Fast Identification in Mobile-Computer-Assisted Devices

Peng Xiao<sup>1</sup>, Zhen Qin<sup>1</sup>, *Member, IEEE*, Dajiang Chen<sup>1</sup>, *Member, IEEE*, Ning Zhang<sup>1</sup>, *Senior Member, IEEE*, Yi Ding<sup>1</sup>, *Member, IEEE*, Fuhu Deng<sup>1</sup>, Zhiguang Qin<sup>1</sup>, and Minghui Pang

**Abstract**—Histopathology diagnosis is an important standard for breast tumors identifying. However, histopathology image analysis is complex, tedious, and error-prone, due to the super-resolution image. In recent years, deep learning technology has been successfully applied to histopathology image analysis and made great progress. The well-known deep neural networks usually have tens of million parameters, which consume much memory to deploy the state-of-the-art model. In addition, deep neural networks rely on high-performance hardware resources, which impede the deployment of the state-of-the-art model on portable equipment. In this work, a novel framework which consists of a weight accumulation method and a lightweight fast neural network (FastNet) was proposed for tumor fast identification (TFI) in mobile-computer-assisted devices. The weight accumulation method was designed to obtain the tissue mask regions of interest and remove the useless background area in histopathology images, which greatly reduces the redundant computation cost. Furthermore, we proposed the lightweight FastNet to improve the computational efficiency on mobile devices. A novel attention loss (AttLoss) function was designed and applied in FastNet. The AttLoss function pays more attention on the positive samples and the indistinguishable samples, which greatly improves the performance. The proposed FastNet was compared with three state-of-the-art methods commonly used for image classification and object detection. Experimental results indicated that FastNet achieves the highest recall of 96.94%, the highest

$F_1$  score of 97.33%, and the highest accuracy of 97.34%, besides least trainable parameters of 0.22M and smallest floating point operations of 210M FLOPs.

**Index Terms**—Histopathology, Internet of Things (IoT), lightweight, mobile-computer-assisted devices, tumor detection.

## I. INTRODUCTION

### A. Background

CANCER is the leading cause of deaths all over the world [1], [2]. Breast cancer occurrence rates continue to rise approximately 0.5% year by year. The breast cancer, accounting for 30% in female cancers, causes the most deaths among the female in the United States [3]. Histopathology diagnosis is the gold standard to confirm carcinoma [4], [5]. Deep learning technology has been successfully applied on histopathology images diagnosis and surpassed pathologists in efficiency and accuracy [6], [7], [8], [9]. The histopathology image usually contains gigapixel, which make it difficult training a deep neural network to detect the tumor region on mobile-edge devices [10], [11], [12].

With the rapid advancement of the Internet of Things (IoT), lightweight research has become significant for resource-constrained IoT devices. The application of IoT-based deep learning in the healthcare service has grown substantially in recent years, resulting in the advancement of diagnostic equipment and opening up new avenues for medical treatment [13]. When IoT devices share sensitive information between patients and centralized cloud servers, privacy security is a serious problem. Some lightweight smart healthcare frameworks and lightweight authentication protocol are proposed to protect the privacy of patient data for mobile-edge computing [14], [15]. In addition, deep neural networks rely on high-performance hardware resources, which impede the deployment of state-of-the-art model on limited memory mobile devices. Therefore, the lightweight of the model is particularly crucial for deployment on resource-constrained devices. Quantization, pruning, and knowledge distillation are common techniques for lightweight models. Depthwise separable convolutions [16], group convolution [17], and parallel 1-D depthwise convolutions [18] are usually used to instead of standard convolutions to reduce parameters. Some lightweight convolutional networks, combining cloud-edge

Manuscript received 7 October 2022; revised 28 December 2022; accepted 3 January 2023. Date of publication 9 January 2023; date of current version 23 May 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62072074, Grant 62076054, Grant 62027827, Grant 61902054, and Grant 62002047; in part by the Frontier Science and Technology Innovation Projects of National Key Research and Development Program under Grant 2019QY1405; in part by the Sichuan Science and Technology Innovation Platform and Talent Plan under Grant 2020JDJQ0020 and Grant 2022JDJQ0039; in part by the Sichuan Science and Technology Support Plan under Grant 2020YFSY0010, Grant 2022YFQ0045, Grant 2022YFS0220, Grant 2019YJ0636, and Grant 2021YFG0131; and in part by the Medico-Engineering Cooperation Funds From University of Electronic Science and Technology of China under Grant ZYGX2021YGLH212 and Grant ZYGX2022YGRH012. (*Corresponding author: Zhen Qin.*)

Peng Xiao, Zhen Qin, Dajiang Chen, Yi Ding, Fuhu Deng, and Zhiguang Qin are with the Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: xp@std.uestc.edu.cn; qinzh@uestc.edu.cn; djchen@uestc.edu.cn; yi.ding@uestc.edu.cn; fuhu.deng@uestc.edu.cn; qinzg@uestc.edu.cn).

Ning Zhang is with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada (e-mail: ning.zhang@uwindsor.ca).

Minghui Pang is with the Department of Geriatric General Surgery, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: mhfang@uestc.edu.cn).

Digital Object Identifier 10.1109/JIOT.2023.3235651

computing with AI technology, are proposed to meet the real-time requirement [19], [20].

The previous research concentrates on the region of interest (ROI) in the histopathology images [21]. Related algorithm processed the histopathology image at low resolution and performed more complicated analysis at high magnifications in the way the pathologists analyze the whole slide image [22]. Concentrating on ROI can avoid many unnecessary regions from the background area, save much computational time, and gain more accuracy [23]. However, only using the ROI data, they lost tremendous amount of organizational structure and spatial information. Facing the extremely high resolution of the histopathology images, most researchers first removed the background region from the histopathology images, and then extracted small patches from tissue region. Finally, they train the deep convolutional neural network to classify these patches as normal or tumor patches [24], [25], [26]. Some researchers screened more discriminant patches as more effective training data to improve the accuracy [27]. However, they only took patches arbitrarily as training data and did not consider the spatial relationship of patches. Some researchers applied the vision transformers to medical images [28], but these vision transformers were often computationally and model size-demanding due to the exponential complexity of self-attention, despite stronger variations with better recognition performance. Some researchers found that texture information and spatial connections between patches are extremely important [29], [30], [31]. Considering the spatial relationship of adjacent patches, these researchers achieved more accurate tumor identification. However, it was time consuming to identify tumor with all the patches and spatial relationship among adjacent patches in super-resolution images. Their model cannot produce the diagnosis result in the expected time due to super-resolution images and complex model. Their models were not suitable for being transferred to tiny portable equipment. Some researchers focused on the portable device [16], [32], such as smart phone, unmanned aerial vehicle, and medical equipment. But their methods achieved a little worse result than the state-of-the-art methods.

In this article, a tumor fast identification (TFI) framework was proposed to address these difficulties. In the TFI framework, we removed the unnecessary background regions and obtained the tissue mask ROIs with weight accumulation method. The tissue mask was obtained from combining the result of the RGB color model, the HSV color model and threshold segmentation, which reduces largely the workload of neural network without reducing the performance. A novel attention loss (AttLoss) function was proposed to increase the weights of positive samples and the indistinguishable samples. The AttLoss function improved classification performance. A lightweight fast neural network (FastNet) was proposed for patch classification in the TFI framework. Compared with the well-known lightweight neural networks (MobileNet V2 [33], ShuffleNet V2 [34], and EfficientNet [35]), the FastNet proposed by us achieved the highest recall and  $F_1$  score with the least trainable parameters and smallest computational complexity.

## B. Related Work

The early researches concentrated on the ROI in the histopathology images. A super pixel clustering method based on segmentation and ROI search was proposed by Li and Huang [23]. This method combined coarse-to-fine super pixel clustering and boundary update. A novel method was proposed to automatically detect the ROIs in histopathology images by Bejnordi et al. [22]. The proposed method targeted ROIs at low resolution and then performs further analysis in high resolution, which simulates the way that a pathologist analyzes the histopathology image.

Facing the extremely high resolution of the histopathology images, most researchers extracted small patches from histopathology images and trained deep convolutional neural network to predict patches. Vang et al. [36] used Inception V3 to classify patches level and predicted the patches by utilizing logistic regression, gradient elevator (GBM), and majority voting in image level. Their experiments achieved a 12.5% higher performance than the state-of-the-art model. Li et al. [27] proposed a patch filtering approach grounded on convolutional neural network and clustering method to screen more discriminant patches. Their method got accuracy of 88.89% on the whole test data and 95% accuracy on the partial test data.

Some researchers found that texture information and spatial connections between patches are extremely important. An NCRF framework was designed for considering spatial correlation among adjacent patches [30]. Li and Ping connected the NCRF framework directly to a fully connected conditional random field to detect cancer metastasis in histopathology images. Karimi et al. [37] used three different convolutional neural networks to predict tumor in multiple patch resolution. Combining the prediction results, their methods extracted more contextual information in different scales. Ding et al. [38] proposed a deep residual network to detect the small target with spatial structure information. Le Trinh et al. [39] achieved an excellent tumor classification by training deep neural network on multiscale histopathology images.

A dilated fully convolutional network was applied to pixel-wise prediction by Dou et al. [40] in biomedical image segmentation. Chiang et al. [41] proposed an independent sliding window detector for volumes of interest extraction. They used candidate aggregation algorithm, 3-D CNN, and sliding window detector to expedite the tumor detection. Yuan et al. [42] proposed a novel end-to-end deep learning method which can extract more significative features for a small target. Ale et al. [43] proposed a novel edge computing scheme to proactive cache data, which makes it practical for the rural people to receive remote histopathology analysis. Chen et al. [44] proposed an event-driven attack method, which exploit users' behavior to collect their private information. Inspired by this, remote histopathology analysis must ensure the privacy of users. An outstanding deep CNN was proposed by Chen et al. [45] to segment medical images. In their method, the model performance was improved by using prior knowledge. Yang et al. [11] sampled multiple scales patches from histopathology image and then fine-tuned the three pretrained deep convolutional neural network. They

combined the three fine-tuned models to form an ensemble model and achieved  $91.75 \pm 2.32\%$  accuracy. Lin et al. [46] designed a novel approach that can leverage the potency of the fully convolutional framework to accelerate the prediction and improve accuracy.

Some researchers focused on the lightweight neural network. SqueezeNet [32] had  $50 \times$  fewer parameters than AlexNet while maintaining accuracy. Howard et al. built lightweight neural networks named MobileNet [16] by using depthwise separable convolutions. They demonstrated that their MobileNets achieved strong performance compared to other well-known methods in object detection. ShuffleNet [47] was designed for mobile devices with limited computation resource. Group convolution was applied in the ShuffleNet to reduce parameters while maintaining the performance. The EfficientNet [35] studied impact on model with different network depth, width, and resolution. Using a compound coefficient, EfficientNet can scale all dimensions of depth/width/resolution. Kumar et al. [13] presented an effective and portable CNN model for histopathology image categorization, and demonstrated its performance on the Raspberry Pi and three mobile devices. MobileViTs [48] were developed specifically for mobile devices and combine the advantages of CNNs with vision transformers. With the same amount of parameters, MobileViT outperformed MobileNets in image classification. Chen et al. proposed the Mobile-Former [49], which enables bidirectional fusion of local and global features and outperforms MobileNetV3 at low FLOPs. It was a parallel architecture of MobileNet and transformer with a two-way bridge in between. Pan et al. [50] introduced a novel class of lightweight ViTs (EdgeViTs) that compared with the best lightweight CNNs in the tradeoff between accuracy and on-device efficiency. They directly took into account latency and energy consumption of various models as opposed to depending on floating point operations or parameters. Despite stronger variations with superior recognition performance, these vision transformers required frequently computation and huge model size due to the exponential complexity of self-attention.

### C. Contributions

To address these key challenges, we propose a TFI framework which consists of a weight accumulation method and a lightweight FastNet to largely reduce the computation cost and accelerate tumor prediction. The main contributions of our work are summarized as follows.

- 1) A weight accumulation method consisting of ROIs locating and patch generating from ROIs of histopathology images is proposed to greatly reduce the redundant computation cost and speeds up the prediction of tumor regions.
- 2) A novel AttLoss function is proposed to increase the weights of positive samples and the indistinguishable samples. The AttLoss function greatly improves classification performance.
- 3) A lightweight FastNet is designed for patch classification. Compared with the common deep neural networks (MobileNet V2, ShuffleNet V2, and EfficientNet-B0),

TABLE I  
EXPERIMENT DATA SET CLASS COMPOSITION

Institution	Train Normal	Train Metastases	Test
Nijmegen	100	70	80
Utrecht	60	40	50
Total	160	110	130

the FastNet proposed by us achieves the highest recall and  $F_1$  score with the least trainable parameters and smallest computational complexity.

The remainder of this article is organized as follows. The data sets used in our experiments is describe in detail in Section II. The histopathology images are visualized in this section. The proposed TFI framework for tumor detection is introduced in Section III. In Section IV, we show all experimental implementation details and results. The conclusion of this article is given in Section V.

## II. DATA SETS AND HISTOPATHOLOGY IMAGES

Our experiments were implemented on the data sets from CAMELYON16 competition, which consist of two different subset data sets from the medical center of Radboud University and Utrecht in The Netherlands. There were totally 400 histopathology images obtained from breast cancer patients. The details are described as follows.

### A. Training Data Sets and Test Data Sets

The training data sets contain two subsets. One data sets, consisting of 170 histopathology images (100 healthy histopathology images and 70 lesion histopathology images), was collected by Radboud University Medical Center. The other data sets, consisting of 100 histopathology images (including 60 healthy histopathology images and 40 lesion histopathology images), was collected by the University Medical Center Utrecht. The test data sets, consisting of 130 histopathology images, were collected from both of medical centers. The data set class compositions are as shown in Table I.

The data sets were annotated by experienced pathologists. The ground-truth labels contained the expert's annotations of tumor area on histopathology images. There were two kinds of ground-truth label being provided.

- 1) The vertex of the annotated outlines of the regions from tumor metastases is provided in xml format.
- 2) Histopathology images binary masks of the regions from the cancer metastasis.

### B. Visualization of Histopathology Images

Ordinarily, histopathology images were scanned and saved as a multilevel pyramid framework. The multilevel pyramid image files consisted of original image and nine downsampling versions of the original image, which is shown in Fig. 1. Each downsampling image in the pyramid consisted of lots of interest of regions. Since the huge size of the histopathology images, it was laboursome to train a convolutional neural

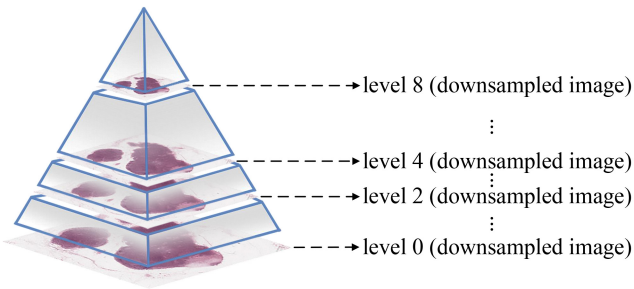


Fig. 1. Multilevel pyramid structure in histopathology images.

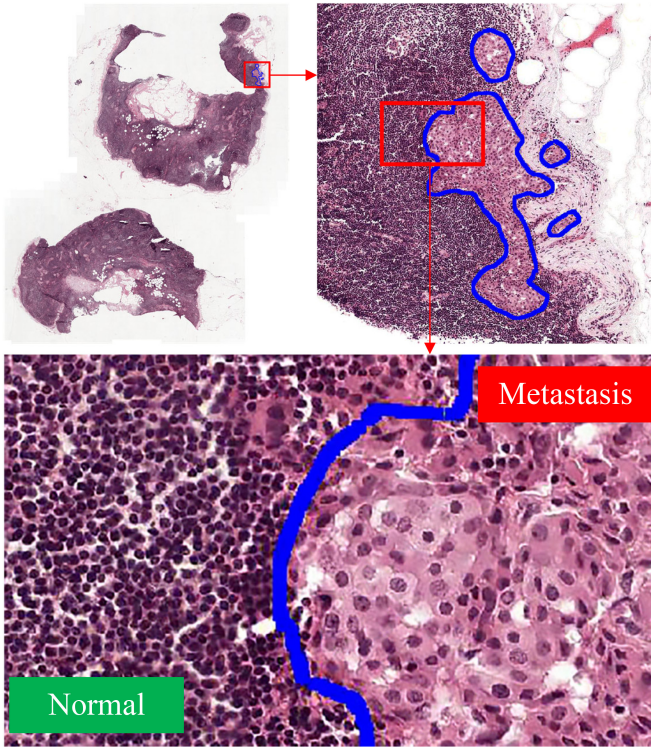


Fig. 2. Visualization of histopathology image and example of a metastatic region. The tumor region takes up only extremely small part of the histopathology image.

network model with gigabyte pixel image. So we extracted patches from the ROI from downsampling images for training.

Automated slide analysis platform (ASAP) is a famous and excellent tools to visualize, analyze, and annotate the histopathology images. We visualized the histopathology images with the ASAP platform and analyze the annotations. The visualization result is as shown in Fig. 2.

### III. TFI FRAMEWORK FOR TUMORS FAST IDENTIFICATION

#### A. Overview of the Tumor Fast Identification Framework

To achieve TFI on mobile-computer-assisted devices, we proposed a TFI framework presented in Fig. 3. Our proposed TFI framework consists of three main parts: 1) weight accumulation method; 2) AttLoss function; and 3) a lightweight FastNet. The weight accumulation method has two main steps. First step is obtaining ROIs in histopathology images, and the second step is patch generating from ROIs with weight accumulation formula. The AttLoss function is used to solve

the problem of extremely imbalanced data distribution. The FastNet has two main components: 1) the multiscale block and 2) the multibranch block. The multiscale block is designed for different scales features extraction, and the multibranch block is applied to extract more details from the previous layer.

As shown in Fig. 3, in the training step, first, we removed the above 70% background area of histopathology images with the weight accumulation method, which greatly reduce the redundant computation cost. Many patches are extracted from the ROIs as the training data sets. Additionally, an AttLoss function was also designed to increase the weights of positive samples and the indistinguishable samples for better classification performance in the FastNet. Next, the FastNet model was trained with the training data sets and AttLoss function. In the prediction step, the test data sets were removed the background area with the same weight accumulation method and split into many patches for tumor probability prediction. Finally, all the tumor prediction probabilities were embedded on the histopathology image in the order of patches, which form a whole tumor probability heatmap. In the probability heatmap, the red region shows the tumor area with extremely high probability, and the green region represents the normal tissue area. In the following section, we introduce the weight accumulation method, AttLoss function (AttLoss), and FastNet in detail.

#### B. Weight Accumulation Method for Image Processing

1) *Regions of Interest in Histopathology Images:* In histopathology images, haematoxylin and eosin (H&E) staining are the most commonly used method of tissue staining. H&E is the abbreviations of hematoxylin and eosin. The hematoxylin stains the nucleus dark blue, the eosin stains the cytoplasm and extracellular matrix pink, and other structures show different outlines and similar colors. The pathologist can distinguish between the nucleus and cytoplasm easily in the histopathology images stained by H&E.

There are large amounts of useless background areas in histopathology images. Locating the ROIs and removing the background areas in the image can reduce vast redundant computation cost. Otsu's method [51] is an efficient algorithm for image binaryzation, which calculates the threshold to divide the original image into foreground and background images, proposed by Japanese scholar OTSU in 1979. Therefore, we apply the Otsu's method to obtain the tissue mask ROIs in histopathology images.

Typically, 70% of histopathology images are background region, as shown in Fig. 4(a). To obtain the tissue mask ROIs, Otsu's method was automatically applied in the RGB color model shown in Fig. 4(b) and the HSV color model shown in Fig. 4(c), respectively, to binarize the images, and we adjusted the threshold of the RGB color value to obtain the threshold tissue mask, as shown in Fig. 4(d). Finally, we combined the RGB tissue mask, HSV tissue mask, and threshold tissue mask to obtain the final Fig. 4(e). More details are described as follows.

First, Otsu's method was leveraged in the RGB color model, in which three mask images were obtained from  $R$  channel,

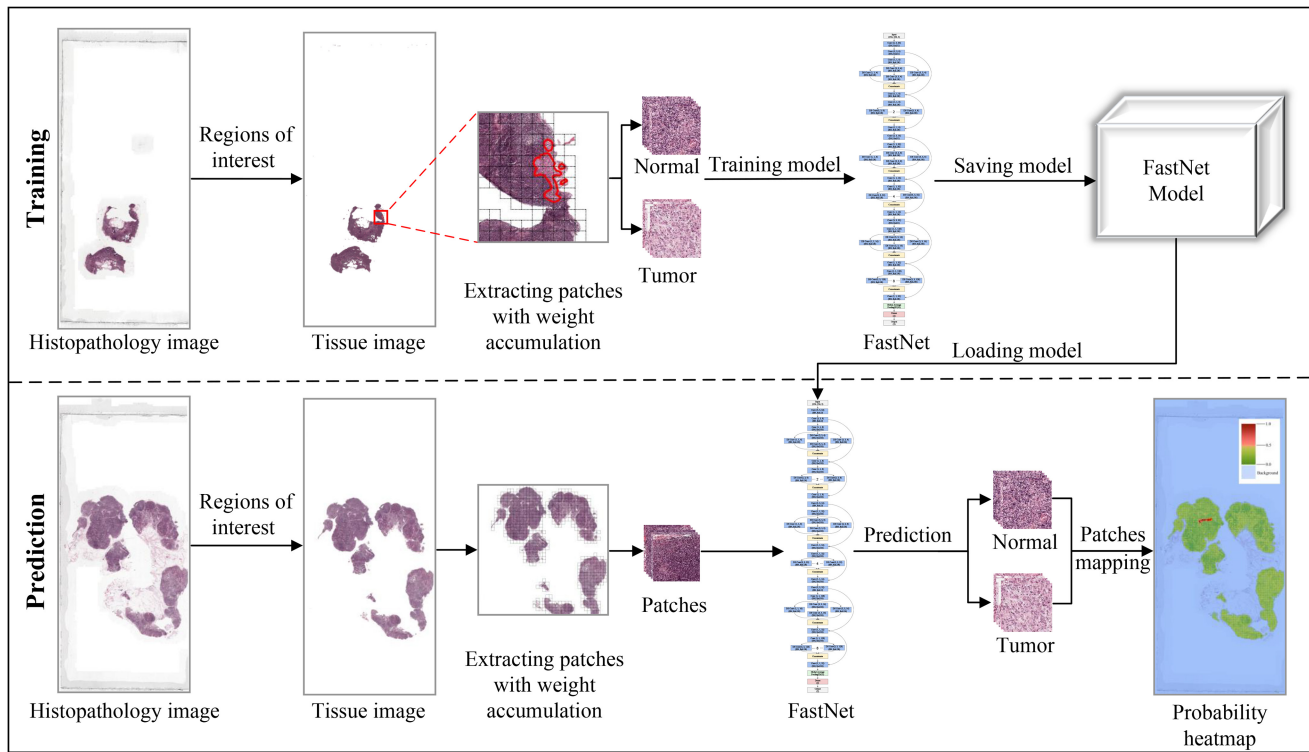


Fig. 3. Overview of our proposed TFI framework, where “Training” displays the procedures of ROIs locating, patch generating from ROIs and model training, and “Prediction” displays the procedures of model prediction and the output of probability heatmap.

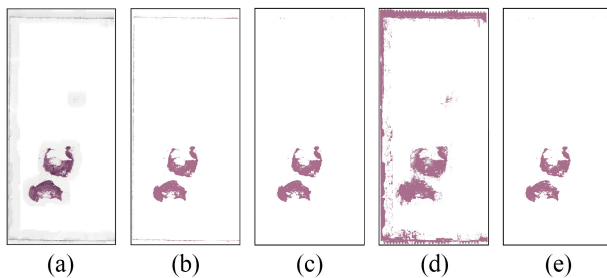


Fig. 4. (a) Original image. (b) Tissue mask image in the RGB color model. (c) Tissue mask image in the HSV color model. (d) Tissue mask image with threshold segmentation (50–240). (e) Tissue mask ROIs combined with (b)–(d).

$G$  channel and  $B$  channel. Then, ANDing the three mask images, the tissue mask was obtained in the RGB color model, marked as RGB\_MASK and as shown in Fig. 4(b). According to the previous research, Otsu’s method performs better in the HSV color model. As a result, we obtained one more new tissue mask in the HSV color model, marked as HSV\_MASK and as shown in Fig. 4(c). An experimental analysis was designed to find optimum background segmentation threshold. It was found that, when the RGB value is between 50–240, the background segmentation works pretty well. Taking the threshold as 50 and 240, we obtained the third tissue mask, marked as THRESHOLD\_MASK and as shown in Fig. 4(d). Finally, we ANDed the RGB\_MASK, HSV\_MASK and THRESHOLD\_MASK to obtain the tissue mask ROIs, as shown in Fig. 4(e). We calculated all the tissue mask ROIs, the average percentage of ROIs is approximately

30%. Obviously, removing the useless background areas can reduce an ocean of redundant computation cost.

2) *Patch Generating From Regions of Interest With Weight Accumulation Formula*: Removing the numerous invalid background regions, we had obtained the tissue mask ROIs. Considering the huge size of the histopathology image, it is difficult feeding the whole image into the convolutional neural network for training. We extracted a multitude of patches from the ROIs weight accumulation method, and trained the neural network with patches. The main procedure is as shown in Fig. 5.

*Ground Truth*: Each histopathology image was annotated by expert pathologists, and the annotations were stored in.xml files consisting of vertex in the annotated outlines. We converted the.xml files to.json files for further processing.

*Tumor Regions of Interest (TROI)*: With the vertices of the annotated contours in.json file, we draw a new polygon gray scale image, and generate TROI. The tumor regions of interest (NROI) is stored as tumor mask image.

*Normal Regions of Interest (NROI)*: Normal region of interest come from both normal images and tumor images. In the normal images, the whole tissue mask is the normal region of interest. Conversely, in the tumor images, the normal region of interest is the rest region of removing the TROI from the whole tissue mask. The NROI is stored as normal mask image.

*Extract Patches From TROI and NROI With Weight Accumulation Formula*: For this part, we describe the method extracting patch from TROI and NROI. The patches extracted from TROI and NROI were considered as positive samples and

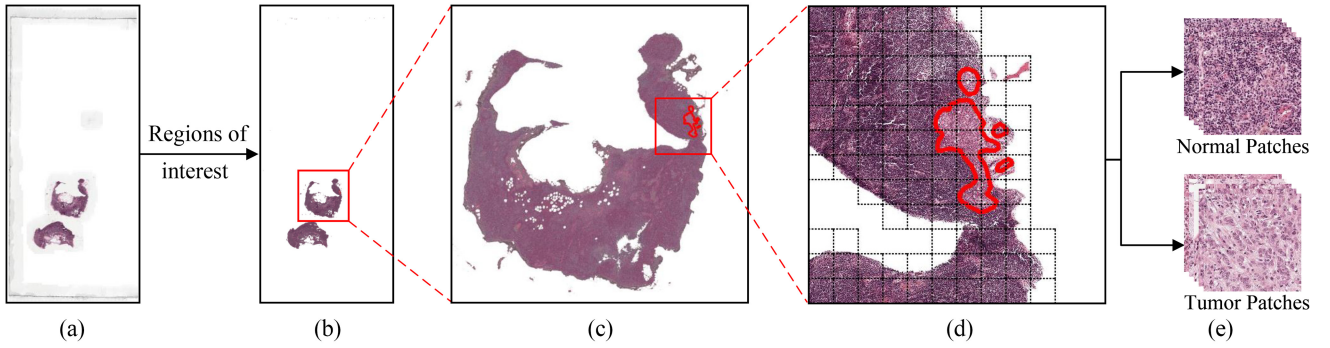


Fig. 5. Methods of extracting patches from histopathology image. (a) Original image. (b) Original image removed background. (c) Partial region of the magnifying image (b). (d) Splitting the images into patches. (e) Data set class composition.

negative samples for training, respectively. The mask image is binary image, in which each pixel value represents true (1) or false (0). First, we calculated the maximum x coordinate (defined as MAX\_X), the maximum y coordinate (defined as MAX\_Y), the minimum x coordinate (defined as MIN\_X) and the minimum y coordinate (defined as MIN\_Y) in each tumor mask image or normal mask image. Combining the four coordinates in pairs, we obtained four points: 1) P1(MIN\_X, MIN\_Y); 2) P2(MIN\_X, MAX\_Y); 3) P3(MAX\_X, MIN\_Y); and 4) P4(MAX\_X, MAX\_Y). The four points formed a rectangle that containing the whole TROI or NROI. Second, we defined the patch's resolution as 256 pixels  $\times$  256 pixels (WIDTH = 256 pixels). We split the rectangle region into patches with padding in the original image.

*Definition 1:*  $I$  means there are  $I$  rows patches in total, defined as follows:

$$I = \frac{\text{MAX\_Y} - \text{MIN\_Y}}{\text{WIDTH}} + 1. \quad (1)$$

*Definition 2:*  $J$  means there are  $J$  columns patches in total, defined as follows:

$$J = \frac{\text{MAX\_X} - \text{MIN\_X}}{\text{WIDTH}} + 1. \quad (2)$$

Hence, we extracted  $I \times J$  patches from the rectangle region in the TROI or NROI. Nevertheless, not all patches meet the true labels. Each patch's mask image contained  $256 \times 256$  pixel values, which is 0 or 1. We proposed a weight accumulation method, in which each mask image pixel value is accumulated as a weighted value  $W$ .

*Definition 3:*  $W$  of weight accumulation formula is defined as

$$W_{i,j} = \sum_{m=0}^{255} \sum_{n=0}^{255} V_{m,n}^{i,j} \quad (3)$$

where,  $W_{i,j}$  denotes the weighted value of the  $i$ th row,  $j$ th column patch  $(i, j)$ .  $V_{m,n}^{i,j}$  represents the pixel value at row  $m$  and column  $n$  in the patch  $(i, j)$ .

*Definition 4:* The patch  $(i, j)$  is considered as training data following the rule below:

$$\text{Patch}_{i,j} \begin{cases} 1, & W_{i,j} \geq \frac{1}{2} \times 256 \times 256 \\ 0, & W_{i,j} < \frac{1}{2} \times 256 \times 256. \end{cases} \quad (4)$$

If  $W_{i,j} \geq (1/2) \times 256 \times 256$ , the patch  $(i, j)$  is considered as the training data. Contrarily,  $W_{i,j} < (1/2) \times 256 \times 256$ , the patch  $(i, j)$  is discarded. The main procedure of extracting patches from the tissue mask ROIs in histopathology images is shown in Fig. 5.

### C. Imbalanced Categories and Loss Function

For extremely imbalanced categories, neural network will tend to predict negative samples. The prediction probability of negative samples will be very high, and the gradient of negative samples will be very large. Usually, the tumor regions takes up only a extremely small part of the histopathology image. We just extracted a few positive samples from the tumor region. In our data, the negative patches was far more than the positive patches. The cross-entropy loss function is usually applied in the binary classification. Equation (5) is the standard formula of cross-entropy loss function for binary classification

$$\text{CE Loss} = -\frac{1}{N} \sum_{i=1}^N (y_i \log p_i + (1 - y_i) \log(1 - p_i)). \quad (5)$$

To solve the problem of extremely imbalanced data distribution, we designed our AttLoss function with reference to the outstanding focal loss function [52]. We added a hyperparameter  $\alpha$  base on cross-entropy loss function to increase weight from the positive samples. In generally, each sample contributes the same weight loss as others when trained. In order to further improve the train accuracy, we added a hyperparameter  $\gamma$  to pay more attention on the indistinguishable samples and less attention on the distinguishable samples. Equations (6) and (7) are the focal loss function and our AttLoss function, respectively

$$\text{FocalLoss} = -\frac{1}{N} \sum_{i=1}^N (\alpha y_i (1 - p_i)^\gamma \log p_i + (1 - \alpha) (1 - y_i) p_i^\gamma \log(1 - p_i)). \quad (6)$$

The AttLoss function is as defined as

$$\text{AttLoss} = -\frac{1}{N} \sum_{i=1}^N (\alpha^{y_i} y_i (1 - p_i)^\gamma \log p_i + \alpha^{y_i} (1 - y_i) (1 - (1 - p_i)^\gamma) \log(1 - p_i)). \quad (7)$$

TABLE II

$(1 - p_i)^\gamma$  IS THE WEIGHTS OF INDISTINGUISHABLE SAMPLES ( $P = 0.5$ ) AND DISTINGUISHABLE SAMPLES ( $P = 0.9$ ) UNDER DIFFERENT  $\gamma$ ;  
 $[(1 - P_{(0.5)})^\gamma / (1 - P_{(0.9)})^\gamma]$  IS THE RATIO OF WEIGHTS ( $P = 0.5$ ) TO WEIGHTS ( $P = 0.9$ )

$(1 - p_i)^\gamma$	0.1	0.2	0.5	1	2	3	4	$\gamma$ 5	6	7	8	9	10
0.5	0.933	0.871	0.707	0.500	0.250	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.000977
0.6	0.912	0.833	0.632	0.400	0.160	0.064	0.026	0.010	0.004	0.002	0.001	0.000262	0.000105
0.7	0.887	0.786	0.548	0.300	0.090	0.027	0.008	0.002	0.001	0.000219	6.56E-05	1.97E-05	5.9E-06
0.8	0.851	0.725	0.447	0.200	0.040	0.008	0.002	0.0003	6.4E-05	1.28E-05	2.56E-06	5.12E-07	1.02E-07
0.9	0.794	0.631	0.316	0.100	0.010	0.001	1E-04	1E-05	1E-06	1E-07	1E-08	1E-09	1E-10
$\frac{(1 - P_{(0.5)})^\gamma}{(1 - P_{(0.9)})^\gamma}$	1.2	1.4	2.2	5	25	125	625	3125	15625	78125	390625	1953125	9765625

TABLE III

$1 - p_i^\gamma$  IS THE WEIGHTS OF INDISTINGUISHABLE SAMPLES ( $P = 0.5$ ) AND DISTINGUISHABLE SAMPLES ( $P = 0.9$ ) UNDER DIFFERENT  $\gamma$ ;  
 $[(1 - P_{(0.5)})^\gamma / (1 - P_{(0.9)})^\gamma]$  IS THE RATIO OF WEIGHTS ( $P = 0.5$ ) TO WEIGHTS ( $P = 0.9$ )

$1 - p_i^\gamma$	0.1	0.2	0.5	1	2	3	$\gamma$ 4	5	6	7	8	9	10
0.5	0.067	0.129	0.293	0.500	0.750	0.875	0.938	0.969	0.984	0.992	0.996	0.998	0.999
0.6	0.050	0.097	0.225	0.400	0.640	0.784	0.870	0.922	0.953	0.972	0.983	0.990	0.994
0.7	0.035	0.069	0.163	0.300	0.510	0.657	0.760	0.832	0.882	0.918	0.942	0.960	0.972
0.8	0.022	0.044	0.106	0.200	0.360	0.488	0.590	0.672	0.738	0.790	0.832	0.866	0.893
0.9	0.010	0.021	0.051	0.100	0.190	0.271	0.344	0.410	0.469	0.522	0.570	0.613	0.651
$\frac{1 - P_{(0.5)}^\gamma}{1 - P_{(0.9)}^\gamma}$	6.39	6.21	5.71	5	3.95	3.23	2.73	2.37	2.10	1.90	1.75	1.63	1.53

In (7),  $p_i$  represents the probability that the  $i$ th sample is predicted as positive class.  $y_i$  is the labels (0 or 1) of  $i$ th patch.  $\alpha$  and  $\gamma$  are the hyperparameters. We increased weight of the positive samples by setting a bigger alpha ( $\alpha > 1$ ), and decreased the weight of the positive samples by setting a smaller alpha ( $0 < \alpha < 1$ ).

This AttLoss function evolves from the basis of the standard cross-entropy loss.  $\alpha^{y_i}$  decide the loss weight of the positive samples and the negative samples. When  $y_i$  is the label 1,  $\alpha^{y_i}$  is equal to  $\alpha$ , the positive sample has the  $\alpha$  times loss compared to the original loss. When the  $y_i$  is the label 0,  $\alpha^{y_i}$  is equal to 1, the negative sample loss remain the original loss.

To decrease the weights of the distinguishable samples, we used a factor  $1 - p_i^\gamma$  for positive sample and a factor  $1 - (1 - p_i)^\gamma$  for negative samples, respectively. For positive samples, the bigger  $p_i$  is, the smaller the factor  $1 - p_i^\gamma$  is. For the negative samples, the bigger the  $1 - p_i$  is, the smaller the factor  $1 - (1 - p_i)^\gamma$  is. As a result, the more easily the samples are classified, the less weights the samples have, which means more weights on the indistinguishable samples.

Compared with the outstanding focal loss function, we replaced the factor  $\alpha$  and  $1 - \alpha$  with  $\alpha^{y_i}$ , which only increase the weights of positive samples. We also replaced the factor  $(1 - p_i)^\gamma$  and  $p_i^\gamma$  with  $1 - p_i^\gamma$  and  $1 - (1 - p_i)^\gamma$ , respectively. In order to more clearly compare the difference between focal loss and AttLoss, we made two tables to investigate the impact of gamma on loss weight.

In the focal loss function, when  $P = 0.5$ , the  $(1 - p_i)^\gamma$  represents the weights of indistinguishable samples; when  $P = 0.9$ , the  $(1 - p_i)^\gamma$  represents the weights of distinguishable samples. As shown in the Table II, when  $\gamma = 1$ , the weights of indistinguishable samples ( $P = 0.5$ ) are five times the weights of distinguishable samples ( $P = 0.9$ ). When  $\gamma = 2$ ,

the weights ( $P = 0.5$ ) are 25 times the weights ( $P = 0.9$ ). When  $\gamma = 10$ , the weights ( $P = 0.5$ ) are 9.76 million times the weights ( $P = 0.9$ ). The factor  $(1 - p_i)^\gamma$  lower the weights of the distinguishable samples extremely fast, which cause gradual deterioration of the performance of the distinguishable samples.

In our AttLoss function, as can be seen from the Table III, when  $\gamma = 1$ , the weights of indistinguishable samples ( $P = 0.5$ ) are 5 times the weights of distinguishable samples ( $P = 0.9$ ). When  $\gamma = 2$ , the weights ( $P = 0.5$ ) are 3.95 times the weights ( $P = 0.9$ ). When  $\gamma = 10$ , the weights ( $P = 0.5$ ) are 1.53 times the weights ( $P = 0.9$ ). In contrast, the factor  $1 - p_i^\gamma$  in the AttLoss decreased the weight of distinguishable samples at a more reasonable rate than  $(1 - p_i)^\gamma$ . The  $\gamma$  parameters are easier to tune in our experiments. The experiments demonstrated that our AttLoss function outperforms the focal loss function in our experiments.

#### D. FastNet Details

Deep convolutional neural network is trained for classifying patches to normal patch or tumor patch. However, it cost a large amount of computation to train and identify the tumor regions due to large-scale gigapixel histopathology image. It is demonstrated that increasing depth with  $3 \times 3$  convolution filters has significant improvements on performance [53]. Shortcut connections can solve the degradation problem in deeper network [54]. Multibranch in the inception module can obtain different receptive field and extract multiscale features. It is proved that factoring convolutions and using aggressive regularization is efficient to improve computational efficiency and reduce parameter count when scale up the networks [55]. Depthwise separable convolution [56] is confirmed to be

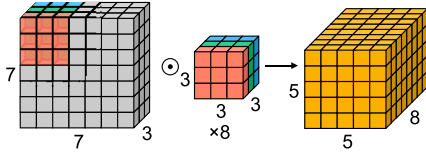


Fig. 6. Standard convolution illustration. Input shape:  $7 \times 7 \times 3$ , output shape:  $5 \times 5 \times 8$ , parameters: 216, and FLOPs: 10 800.

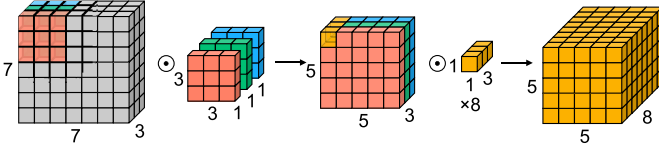


Fig. 7. Depthwise separable convolution illustration. Input shape:  $7 \times 7 \times 3$ , output shape:  $5 \times 5 \times 8$ , parameters: 51, and FLOPs: 2550.

smaller and faster in MobileNets [16]. Inspired by the former studies, we proposed a FastNet (fast convolutional neural network) to train a model for identify tumor region rapidly. The FastNet was stacked alternately by multiscale blocks and multibranch blocks. Multiscale block was designed for different scales features extraction, and multibranch block was applied to extract more details from the previous layer. Both multiscale blocks and multibranch blocks added shortcut connections to solve the degradation problem when stacked deeper. ReLU6 was used as the nonlinearity activation function in multiscale blocks and multibranch blocks because of its robustness. Depthwise separable convolutions were applied in multiscale blocks and multibranch blocks to reduce parameters and computation cost. It was demonstrated that expanding the channels before depthwise separable convolutions can decrease the case where the parameter is zero [33]. The depthwise separable convolution, multiscale blocks, multibranch blocks, and FastNet architecture were described in detail in the following section.

1) *Depthwise Separable Convolution*: Depthwise separable convolution is applied in many well-known neural network [16], [33], [57] to reduce parameters and computation cost. Depthwise separable convolution consists of depthwise convolution and pointwise convolution. Each input channel is extracted feature by a single filter in depthwise convolution. Pointwise convolution uses standard  $1 \times 1$  convolution to concatenate each channel. The depthwise separable convolutions has  $(1/N) + (1/D_K^2)$  times less parameters and computation cost than standard convolution [16]. The depthwise separable convolutions in Fig. 7 has 51 parameters and 2550 FLOPs, which is approximately four times less parameters and FLOPs than the standard convolution of 216 parameters and 10 800 FLOPs in Fig. 6. The standard convolution process is shown in Fig. 6, and the depthwise separable convolution process is shown in Fig. 7.

2) *Multiscale Block*: The multiscale block architecture is shown in Fig. 8(a). First, we used standard  $1 \times 1$  convolution to expand the input channels from  $M$  to  $M \times N$ . Then, we used one  $1 \times 1$  depthwise separable convolutions, one  $3 \times 3$  depthwise separable convolutions, and one  $5 \times 5$  depthwise separable convolutions (Factorization into two  $3 \times 3$

depthwise separable convolutions [55]) to extract multiscale features. Next we concatenated  $1 \times 1$  branch,  $3 \times 3$  branch and  $5 \times 5$  branch. Finally, we add standard  $1 \times 1$  convolution to reduce the channels and added residuals on the result. The output shape was the same as the input shape.  $W$  represents the length and width of the input.  $M$  is the input channels,  $N$  is the channel expansion factor, and  $B$  is the channels in each scale branch. Increasing the  $N$  and  $B$ , the multiscale block has more robustness and multiscale feature extraction capability but more parameters and computation cost.

3) *Multibranch Block*: In the multibranch block architecture is shown in Fig. 8(b). First, we used same standard  $1 \times 1$  convolution to expand the input channels from  $M$  to  $M \times N$ . Then,  $B$   $3 \times 3$  depthwise separable convolutions branches are applied to extract more features. Next, we concatenated  $B$  depthwise separable convolutions branches and also added residuals on the result. The output shape is also the same as the input shape.  $W$  also represents the length and width of the input.  $M$  is also the input channels,  $N$  is also the channel expansion factor, but  $B$  is the branch number used in the multibranch blocks. Increasing the  $N$  and  $B$ , the multibranch block has more robustness and detailed feature extraction capability but more parameters and computation cost. We adjust the channel expansion factor and branch number  $B$  to tradeoff the robustness, efficiency and computation cost.

#### E. FastNet Architecture

In the previous section, we investigated the effects of different scales and branch on parameters, computation cost and experimental performance. Then, we build the FastNet after comprehensively considering the parameters and computation cost of the multiscale and multibranch blocks. The network architecture details is presented in Table IV. First, One convolutional layer was used to extract the low-dimensional features and reduce the size of the original image with  $2 \times 2$  stride. Second, we stacked convolutional layer, multiscale block and multibranch block alternately three times to downsample the feature maps, extract multiscale feature and more details. The details of convolutional layer, multiscale block and multibranch block is shown in Table IV. In the multiscale block,  $M$  represents the input channels,  $N$  represents the channel expansion factor, and  $B$  represents the channels in each multiscale branch. In the multibranch block,  $M$  is also the input channels,  $N$  is also the channel expansion factor, but  $B$  is the branch number used in each multibranch block. Third, a global average pooling was used to aggregate spatial dimensions features. Finally, the fully connected layer was used to output the binary classification results. The FastNet architecture is shown in Fig. 9.

## IV. EXPERIMENTS

### A. Data Sets and Experimental Environment

We used the CAMELYON16 challenge data sets as our experiment data sets. The data sets consisted of two different subset data sets from the medical centers of Radboud University and Utrecht in The Netherlands. There were altogether 400 histopathology images obtained from breast cancer



TABLE IV

OVERALL ARCHITECTURE OF FASTNET.  $M$  IS THE INPUT CHANNEL NUMBERS OF THE MULTISCALE BLOCKS AND MULTIBRANCH BLOCKS;  $N$  IS THE CHANNEL EXPANSION FACTOR OF THE MULTISCALE BLOCKS AND MULTIBRANCH BLOCKS; IN THE MULTISCALE BLOCKS,  $B$  REPRESENTS THE CHANNEL NUMBER IN EACH SCALE BRANCH; IN MULTIBRANCH BLOCKS,  $B$  REPRESENTS BRANCH NUMBER

Layers	Operator	Input	Output	Filters	Kernel size	Stride	Padding	Activation	M	N	B
1	Input	(256, 256, 3)	(256, 256, 3)	-	-	-	-	-	-	-	-
2	Conv2D	(256, 256, 3)	(127, 127, 64)	64	(3, 3)	(2, 2)	valid	relu	-	-	-
3	BatchNormalization	(127, 127, 64)	(127, 127, 64)	-	-	-	-	-	-	-	-
4	Conv2D	(127, 127, 64)	(63, 63, 8)	8	(3, 3)	(2, 2)	valid	relu	-	-	-
5	BatchNormalization	(63, 63, 8)	(63, 63, 8)	-	-	-	-	-	-	-	-
6	Multi-scale block	(63, 63, 8)	(63, 63, 8)	-	-	-	-	-	8	1	4
7	Multi-branch block	(63, 63, 8)	(63, 63, 8)	-	-	-	-	-	8	1	2
8	Conv2D	(63, 63, 8)	(31, 31, 16)	16	(3, 3)	(2, 2)	valid	relu	-	-	-
9	BatchNormalization	(31, 31, 16)	(31, 31, 16)	-	-	-	-	-	-	-	-
10	Multi-scale block	(31, 31, 16)	(31, 31, 16)	-	-	-	-	-	16	2	8
11	Multi-branch block	(31, 31, 16)	(31, 31, 16)	-	-	-	-	-	16	2	4
12	Conv2D	(31, 31, 16)	(15, 15, 32)	32	(3, 3)	(2, 2)	valid	relu	-	-	-
13	BatchNormalization	(15, 15, 32)	(15, 15, 32)	-	-	-	-	-	-	-	-
14	Multi-scale block	(15, 15, 32)	(15, 15, 32)	-	-	-	-	-	32	4	16
15	Multi-branch block	(15, 15, 32)	(15, 15, 32)	-	-	-	-	-	32	4	8
16	GlobalAveragePooling2D	(15, 15, 32)	(, 32)	-	-	-	-	-	-	-	-
17	Dense	(, 32)	(, 2)	-	-	-	-	softmax	-	-	-

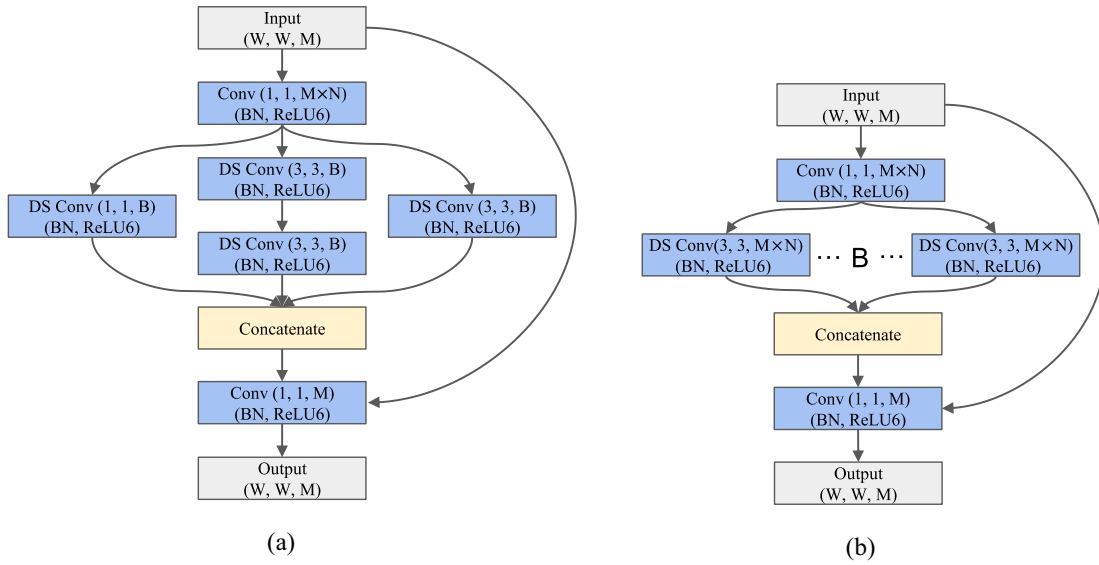


Fig. 8. Multiscale block and Multibranch block architectures. (a) Multiscale block for different scales feature extraction. (b) Multibranch block for more details feature extraction.

patients. The data sets from University Medical Center Utrecht consisted of 100 healthy histopathology images and 70 lesion histopathology images. The data sets from Radboud University Medical Center consisted of 60 healthy histopathology images and 40 lesion histopathology images. The test data sets, collected from both medical center, consist of 130 histopathology images.

All experiments were implemented on a workstation with Intel Core i7-8700k CPU, a GPU of Nvidia GTX 3090 and 64 GB of system RAM. Ubuntu 20.04 LTS was installed on the workstation. The project was implemented with deep learning framework Keras. Cross-entropy, focal loss and AttLoss are compared as the loss function, and Adam was used as the optimizer for our experiment models.

### B. Evaluation Metrics

We used params, FLOPs, recall, precision,  $F_1$  score, and accuracy to evaluate our model. Params is the trainable

parameter of model. FLOPs is floating point operations, which is a measure of algorithm or model complexity and represents the total number of floating-point computations for the CPU and GPU. The recall, precision,  $F_1$  score, and accuracy are defined as (8)–(11), respectively. The recall is a percentage of how many positive samples in the positive samples are correctly predicted. The precision represents the percentage of how many true-positive samples that predicted as positive sample in the all samples. In our experiments, we preferred high recall as well as high precision. We introduced  $F_1$  score which is the harmonic mean of precision and recall. In extremely imbalanced categories classifications,  $F_1$  score is more important than accuracy

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

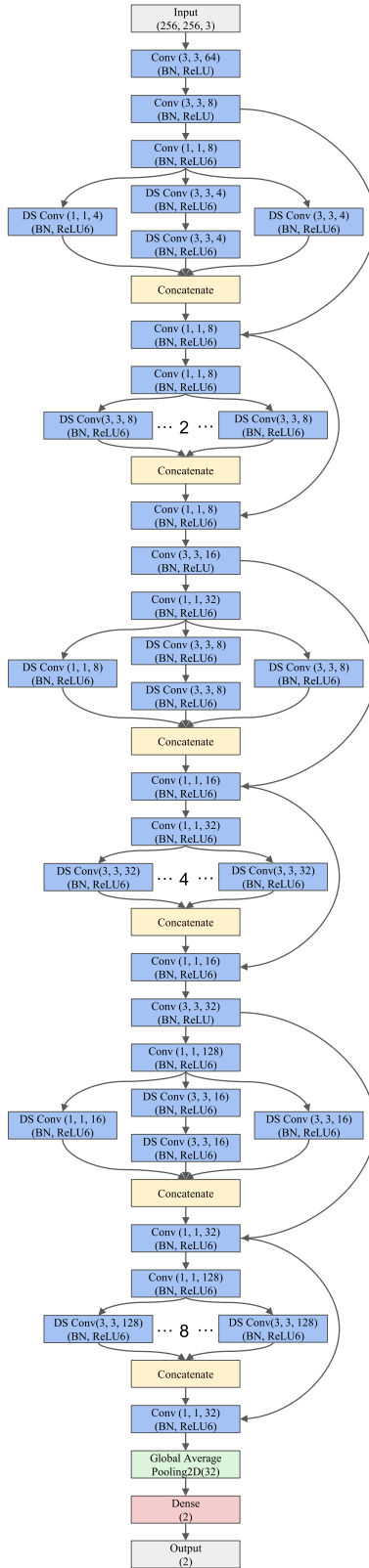


Fig. 9. FastNet architecture. Numbers 2, 4, and 8 are the branch numbers used in each multibranch block.

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (11)$$

TABLE V

EXPERIMENTAL RESULTS OF DIFFERENT KINDS MULTISCALE NETWORK. (“ID” MEANS THE NUMBER OF EXPERIMENTS; “FLOPs” IS MEASURED BY FLOATING POINT OPERATIONS; “TRAINING ACC” IS THE TRAINING ACCURACY; “VAL ACC” IS THE VALIDATION ACCURACY; AND THE BOLD FONT INDICATES THE STATE-OF-THE-ART RESULT)

ID	FLOPs	Params	Training Acc	Val Acc
1	<b>13310</b>	<b>6650</b>	0.8978	0.8972
2	13905	6946	0.9294	0.9151
3	14807	7394	0.9344	0.9236
4	16016	7994	<b>0.9412</b>	<b>0.9293</b>

TABLE VI

EXPERIMENTAL RESULTS OF DIFFERENT NUMBER OF MULTIBRANCH NETWORK. (“BRANCHES” MEANS THE NUMBER OF BRANCHES IN THE MULTIBRANCH BLOCK; “FLOPs” IS MEASURED BY FLOATING POINT OPERATIONS; “TRAINING ACC” IS THE TRAINING ACCURACY; “VAL ACC” IS THE VALIDATION ACCURACY; AND THE BOLD FONT INDICATES THE STATE-OF-THE-ART RESULT)

Branches	FLOPs	Params	Train Acc	Val Acc
1	<b>14478</b>	<b>7234</b>	0.8981	0.8989
2	15793	7890	0.9296	0.9141
4	18423	9202	0.9322	0.9188
8	23683	11826	0.9345	0.9246
16	34203	17074	<b>0.9365</b>	<b>0.9266</b>

### C. Experimental Setup

We used weight accumulation method to remove the background from the training data. Tissue images were split into  $256 \times 256$  patches in two classes at high-resolution level. The data sets were divided into training data sets and testing data sets randomly. All the experiments are trained on training data sets and evaluated on testing data sets. First, we investigate the effect of different scales and branches on the experimental performance. Considering the parameters and computation cost, we used the suitable multiple scales and different branches based on the experimental results in FastNet. Second, we first evaluated the experiment performance performed on depthwise separable convolution and standard convolution. Additionally, we investigated the impacts of shortcut using in the multiscale block and multibranch block on the network performance. Third, we compared our AttLoss with cross-entropy loss and focal loss on four popular lightweight networks, respectively. Finally, we compared our FastNet with MobileNetV2, ShuffleNet V2, and EfficientNet-B0 in three loss functions, respectively.

### D. Performance Comparison of Different Scales and Different Branches

We designed a network with single multiscale block to investigate the effects of different scales on experimental results. In the experiment 1, we used only one  $1 \times 1$  convolution branch in the multiscale block as the baseline. In the experiment 2,  $1 \times 1$  convolution branch, and  $3 \times 3$  convolution branch were concatenated in the multiscale block.  $1 \times 1$  convolution branch,  $3 \times 3$  convolution branch, and  $5 \times 5$  convolution branch were concatenated in the experiment 3. In the experiment 4, we concatenated  $1 \times 1$  convolution branch,  $3 \times 3$  convolution branch,  $5 \times 5$  convolution branch, and  $7 \times 7$  convolution branch in the multiscale block. The

TABLE VII

PARAMETERS, FLOPS AND PERFORMANCE COMPARISON-BASED CAMELYON16 OF FASTNET WITH STANDARD CONVOLUTION, DEPTHWISE SEPARABLE CONVOLUTION AND SHORTCUT, RESPECTIVELY. (INPUT SHAPE IS  $256 \times 256 \times 3$ ; LOSS FUNCTION IS CROSS-ENTROPY LOSS FUNCTION; "FLOPS" IS MEASURED BY FLOATING POINT OPERATIONS; AND THE BOLD FONT INDICATES THE STATE-OF-THE-ART RESULT)

ID	Methods	Convolution Layer	Shortcut	Params	FLOPs	Accuracy	Recall	Precision	$F_1$ Score
1	FastNet	Standard	no	1.32M	773.64M	0.9533	0.9607	0.9467	0.9536
2	FastNet	Standard	yes	1.32M	773.75M	0.9715	<b>0.9647</b>	0.9781	0.9713
3	FastNet	Depth Separable	no	0.22M	210.31M	0.9628	0.9586	0.9668	0.9627
4	FastNet	Depth Separable	yes	0.22M	210.42M	<b>0.9721</b>	0.9638	<b>0.9795</b>	<b>0.9716</b>

TABLE VIII

PERFORMANCE COMPARISONS FOR ATTLOSS WITH CROSS-ENTROPY LOSS AND FOCAL LOSS USING SAME MODELS. (INPUT SHAPE IS  $256 \times 256 \times 3$ ; "ID" MEANS ID-TH GROUP EXPERIMENTS; "FLOPS" IS MEASURED BY FLOATING POINT OPERATIONS; AND THE BOLD FONT INDICATES THE STATE-OF-THE-ART RESULT)

ID	Methods	Loss	Shortcut	Params	FLOPs	Accuracy	Recall	Precision	F1 Score
1	MobileNet V2 [33]	Cross Entropy	yes	0.69M	251M	0.9479	0.9362	0.9587	0.9473
		Focal Loss				0.9311	<b>0.9634</b>	0.9449	0.9541
		Attention Loss				<b>0.9587</b>	0.9475	<b>0.9638</b>	<b>0.9556</b>
2	ShuffleNet V2 [34]	Cross Entropy	yes	1.26M	381M	0.9614	0.9457	0.9763	0.9608
		Focal Loss				0.9659	<b>0.9549</b>	0.9763	0.9655
		Attention Loss				<b>0.9668</b>	0.9528	<b>0.9788</b>	<b>0.9656</b>
3	EfficientNet-B0 [35]	Cross Entropy	yes	4.01M	1027M	0.9574	<b>0.9689</b>	0.9563	0.9626
		Focal Loss				0.9606	0.9514	<b>0.9752</b>	0.9632
		Attention Loss				<b>0.9631</b>	0.9639	0.9689	<b>0.9664</b>
4	FastNet (ours)	Cross Entropy	yes	0.22M	210M	0.9686	0.9531	0.9837	0.9681
		Focal Loss				0.9695	0.9533	<b>0.9853</b>	0.9690
		Attention Loss				<b>0.9734</b>	<b>0.9694</b>	0.9771	<b>0.9733</b>

experimental results are shown in Table V, which demonstrate that more multiscale branches in the multiscale block can significantly improve the network performance but more parameters and computation cost. Considering model parameters and complexity, we made a tradeoff between performance and complexity. Finally, we adopted the  $1 \times 1$  convolution branch,  $3 \times 3$  convolution branch and  $5 \times 5$  convolution branch to extract multiscale feature in the multiscale block.

Similarly, we also designed a network with single multi-branch block to investigate the effects of different branches on experimental results. We conducted five experiments with 1, 2, 4, 8, and 16 branches, respectively. The experimental results are shown in Table VI, which demonstrate that more branches can significantly improve network performance but bring more parameters and computation cost. Consequently, we appropriately adjust the number of branches in FastNet as required.

#### E. Comparison Among Standard Convolution, Depthwise Separable Convolution and Shortcut

First, we evaluated our FastNet architecture using standard convolution and depthwise separable convolution, respectively. As shown in Table VII, comparing experiments 1 with 3 and experiments 2 with 4, the FastNet using depthwise separable convolution achieve a higher accuracy, precision,  $F_1$  score, and a competitive recall than FastNet using standard convolution with extremely few parameters and small computational complexity. Second, we investigated the impacts of shortcut using in the multiscale block and multibranch block. Comparing the experiments 1 with 2 and experiments 3 with 4,

we found that FastNet with shortcut outperform the FastNet without shortcut in accuracy, recall, precision and  $F_1$  score. Considering efficiency of depthwise separable convolution and excellent performance of shortcut, in the next experiments, the FastNet utilized the depthwise separable convolution instead of standard convolution and included the shortcut by default.

#### F. Performance Comparison for Attention Loss Function With Cross-Entropy Loss Function and Focal Loss Function

In the tumor detection scene, both recall and precision are extremely important.  $F_1$  score is the harmonic mean of precision and recall, so we considered the  $F_1$  score as the first evaluation metrics. Second, we considered the accuracy as the second evaluation metric. We evaluated the performance of the AttLoss function and compared it with the cross-entropy loss function and focal loss function. The loss function comparison experiments were performed on the MobileNet V2, ShuffleNet V2, EfficientNet-B0, and FastNet, respectively. Experimental results are shown in Table VIII. In the 4th group experiments, we can clearly see that the AttLoss function proposed by us achieved the best accuracy, recall,  $F_1$  score and competitive precision. The 1st, 2nd, and 3rd group experiments are the ablation experiments. In the ablation experiments, our AttLoss function also achieved the best  $F_1$  score and accuracy at a competitive recall and precision in different neural networks.

#### G. Performance Comparison for FastNet With MobileNet V2, ShuffleNet V2, and EfficientNet-B0

We evaluated and compared the performance of FastNet with MobileNet V2, ShuffleNet V2, and EfficientNet-B0 using

TABLE IX  
PERFORMANCE COMPARISONS FOR FASTNET WITH MOBILENET V2, SHUFFLENET V2, AND EFFICIENTNET-B0 USING SAME LOSS FUNCTION. (INPUT SHAPE IS  $256 \times 256 \times 3$ ; "ID" MEANS ID-TH GROUP EXPERIMENTS; "FLOPS" IS MEASURED BY FLOATING POINT OPERATIONS; AND THE BOLD FONT INDICATES THE STATE-OF-THE-ART RESULT)

ID	Methods	Loss	Shortcut	Params	FLOPs	Accuracy	Recall	Precision	F1 Score
1	MobileNet V2 [33]	Cross Entropy	yes	0.69M	251M	0.9479	0.9362	0.9587	0.9473
	ShuffleNet V2 [34]			1.26M	381M	0.9614	0.9457	0.9763	0.9608
	EfficientNet-B0 [35]			4.01M	1027M	0.9574	<b>0.9689</b>	0.9563	0.9626
	FastNet (ours)			<b>0.22M</b>	<b>210M</b>	<b>0.9686</b>	0.9531	<b>0.9837</b>	<b>0.9681</b>
2	MobileNet V2 [33]	Focal Loss	yes	0.69M	251M	0.9311	<b>0.9634</b>	0.9449	0.9541
	ShuffleNet V2 [34]			1.26M	381M	0.9659	0.9549	0.9763	0.9655
	EfficientNet-B0 [35]			4.01M	1027M	0.9606	0.9514	0.9752	0.9632
	FastNet (ours)			<b>0.22M</b>	<b>210M</b>	<b>0.9695</b>	0.9533	<b>0.9853</b>	<b>0.9690</b>
3	MobileNet V2 [33]	Attention Loss	yes	0.69M	251M	0.9587	0.9475	0.9638	0.9556
	ShuffleNet V2 [34]			1.26M	381M	0.9668	0.9528	<b>0.9788</b>	0.9656
	EfficientNet-B0 [35]			4.01M	1027M	0.9631	0.9639	0.9689	0.9664
	FastNet (ours)			<b>0.22M</b>	<b>210M</b>	<b>0.9734</b>	<b>0.9694</b>	0.9771	<b>0.9733</b>

different loss function. The experimental results are shown in Table IX. In the 1st group experiments, using cross-entropy loss function, our FastNet achieved the best accuracy, precision,  $F_1$  score, and suboptimal recall with the least trainable parameters and smallest computational complexity. In the 2nd group experiments, using focal loss function, our FastNet achieved the best accuracy, precision,  $F_1$  score, and competitive recall with the least trainable parameters and smallest computational complexity. In the 3rd group experiments, using AttLoss function proposed by us, our FastNet achieved the best accuracy, recall,  $F_1$  score, and suboptimal precision with the least trainable parameters and smallest computational complexity. Compared with the 1st group and the 2nd group experiments, the FastNet, with AttLoss function proposed by us in the 3rd group experiments, achieved the best accuracy, recall, and  $F_1$  score with the least trainable parameters and smallest computational complexity. Therefore, the experimental results demonstrated that the FastNet proposed by us outperforms MobileNet V2, ShuffleNet V2, and EfficientNet-B0 in our medical images classification. The least trainable parameters and smallest computational complexity make our FastNet higher efficiency in portable machines.

## V. CONCLUSION

In this article, a TFI framework was presented to detect tumor regions fast. In the TFI framework, a weight accumulation method for histopathology images was designed to greatly reduce the redundant computation cost. We removed the background area in the RGB color model and the HSV color model, respectively. We combined tissue masks from different color model, which increases the accuracy of tissue mask. A novel AttLoss function was proposed to increase the weights of positive samples and the indistinguishable samples. In the AttLoss function, we added a hyperparameter  $\alpha$  based on cross-entropy loss function to decrease weight from the negative samples and added a hyperparameter  $\gamma$  to pay more attention on the indistinguishable samples and less attention on the distinguishable samples. A lightweight FastNet was proposed for patch classification in the TFI framework. Compared with the MobileNetV2, ShuffleNet V2, and EfficientNet-B0, the FastNet proposed by us achieved the highest recall, accuracy,

and  $F_1$  score with the least trainable parameters and smallest computational cost. While there are some drawbacks. It is not precise enough to regard the whole patch as tumor or normal. At the edge of tumor regions, the patch contains both tumor pixels and normal tissue pixels. The future work will focus on pixel-level tumor segmentation, which will achieve more precise tumor boundary.

## REFERENCES

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA Cancer J. Clin.*, vol. 69, no. 1, pp. 7–34, 2019.
- [2] R. L. Siegel et al., "Colorectal cancer statistics, 2020," *CA Cancer J. Clin.*, vol. 70, no. 3, pp. 145–164, 2020.
- [3] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021," *CA Cancer J. Clin.*, vol. 71, no. 1, pp. 7–33, 2021.
- [4] N. Coudray et al., "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *Nat. Med.*, vol. 24, no. 10, pp. 1559–1567, 2018.
- [5] M. Kitar et al., "To determine validity of ultrasound in predicting acute appendicitis among children keeping histopathology as gold standard," *Ann. Med. Surg.*, vol. 38, pp. 22–27, Feb. 2019.
- [6] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [7] 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.
- [8] R. Poplin et al., "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nat. Biomed. Eng.*, vol. 2, no. 3, pp. 158–164, 2018.
- [9] Q. Ye et al., "Method of tumor pathological micronecrosis quantification via deep learning from label fuzzy proportions," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 9, pp. 3288–3299, Sep. 2021.
- [10] P. Filipczuk, T. Fevens, A. Krzyżak, and R. Monczak, "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies," *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 2169–2178, Dec. 2013.
- [11] Z. Yang, L. Ran, S. Zhang, Y. Xia, and Y. Zhang, "EMS-net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images," *Neurocomputing*, vol. 366, pp. 46–53, Nov. 2019.
- [12] K. Stacke, G. Eilertsen, J. Unger, and C. Lundström, "Measuring domain shift for deep learning in histopathology," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 2, pp. 325–336, Feb. 2021.
- [13] A. Kumar, A. Sharma, V. Bharti, A. K. Singh, S. K. Singh, and S. Saxena, "MobiHisNet: A lightweight CNN in mobile edge computing for histopathological image classification," *IEEE Internet Things J.*, vol. 8, no. 24, pp. 17778–17789, Dec. 2021.
- [14] W. Wang et al., "Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8883–8891, Jun. 2022.

- [15] A. Singh, K. Chatterjee, A. K. Singh, and N. Kumar, "Secure smart healthcare framework using lightweight DNA sequence and chaos for mobile edge computing," *IEEE Internet Things J.*, early access, Nov. 3, 2020, doi: [10.1109/JIOT.2022.3219113](https://doi.org/10.1109/JIOT.2022.3219113).
- [16] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [17] Y. Ioannou, D. Robertson, R. Cipolla, and A. Criminisi, "Deep roots: Improving CNN efficiency with hierarchical filter groups," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1231–1240.
- [18] Z. Ou and W. Ye, "Lightweight deep learning model for radar-based fall detection with metric learning," *IEEE Internet Things J.*, early access, Dec. 15, 2022, doi: [10.1109/JIOT.2022.3229462](https://doi.org/10.1109/JIOT.2022.3229462).
- [19] L. Ren, Y. Liu, X. Wang, J. Lü, and M. J. Deen, "Cloud-edge-based lightweight temporal convolutional networks for remaining useful life prediction in IIoT," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12578–12587, Aug. 2021.
- [20] X. Liu, Y. Chen, J. Li, and A. Cangelosi, "Real-time robotic mirrored behavior of facial expressions and head motions based on lightweight networks," *IEEE Internet Things J.*, vol. 10, no. 2, pp. 1401–1413, Jan. 2023.
- [21] Y. Zheng et al., "Size-scalable content-based histopathological image retrieval from database that consists of WSIs," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1278–1287, Jul. 2018.
- [22] B. E. Bejnordi, G. Litjens, M. Hermsen, N. Karssemeijer, and J. A. W. M. van der Laak, "A multi-scale superpixel classification approach to the detection of regions of interest in whole slide histopathology images," in *Proc. Med. Imag. Digit. Pathol.*, 2015, Art. no. 94200H.
- [23] R. Li and J. Huang, "Fast regions-of-interest detection in whole slide histopathology images," in *Proc. Int. Workshop Patch-Based Techn. Med. Imag.*, 2015, pp. 120–127.
- [24] Y. Xu, J.-Y. Zhu, E. I.-C. Chang, M. Lai, and Z. Tu, "Weakly supervised histopathology cancer image segmentation and classification," *Med. Image Anal.*, vol. 18, no. 3, pp. 591–604, 2014.
- [25] J. Lotz et al., "Patch-based nonlinear image registration for gigapixel whole slide images," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 9, pp. 1812–1819, Sep. 2016.
- [26] B. E. Bejnordi et al., "Automated detection of DCIS in whole-slide H&E stained breast histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 9, pp. 2141–2150, Sep. 2016.
- [27] Y. Li, J. Wu, and Q. Wu, "Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning," *IEEE Access*, vol. 7, pp. 21400–21408, 2019.
- [28] Y. Li et al., "AGMB-transformer: Anatomy-guided multi-branch transformer network for automated evaluation of root canal therapy," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 4, pp. 1684–1695, Apr. 2022.
- [29] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2424–2433.
- [30] Y. Li and W. Ping, "Cancer metastasis detection with neural conditional random field," 2018, *arXiv:1806.07064*.
- [31] Y. Zheng et al., "Histopathological whole slide image analysis using context-based CBIR," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1641–1652, Jul. 2018.
- [32] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*.
- [33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
- [34] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 122–138.
- [35] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
- [36] Y. S. Vang, Z. Chen, and X. Xie, "Deep learning framework for multi-class breast cancer histology image classification," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 914–922.
- [37] D. Karimi, G. Nir, L. Fazli, P. C. Black, L. Goldenberg, and S. E. Salcudean, "Deep learning-based Gleason grading of prostate cancer from histopathology images-role of multiscale decision aggregation and data augmentation," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 5, pp. 1413–1426, May 2020.
- [38] Y. Ding et al., "DeepEDN: A deep-learning-based image encryption and decryption network for Internet of Medical Things," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1504–1518, Feb. 2021.
- [39] T. T. Le Trinh, B. Song, K. Kim, Y. M. Cho, and J. T. Kwak, "Multi-scale binary pattern encoding network for cancer classification in pathology images," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 3, pp. 1152–1163, Mar. 2022.
- [40] Q. Dou, C. Ouyang, C. Chen, H. Chen, and P.-A. Heng, "Unsupervised cross-modality domain adaptation of ConvNets for biomedical image segmentations with adversarial loss," 2018, *arXiv:1804.10916*.
- [41] T.-C. Chiang, Y.-S. Huang, R.-T. Chen, C.-S. Huang, and R.-F. Chang, "Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 240–249, Jan. 2019.
- [42] Y. Yuan, Z. Xiong, and Q. Wang, "VSSA-NET: Vertical spatial sequence attention network for traffic sign detection," *IEEE Trans. Image Process.*, vol. 28, pp. 3423–3434, 2019.
- [43] L. Ale, N. Zhang, H. Wu, D. Chen, and T. Han, "Online proactive caching in mobile edge computing using bidirectional deep recurrent neural network," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5520–5530, Jun. 2019.
- [44] D. Chen et al., "MAGLeak: A learning-based side-channel attack for password recognition with multiple sensors in IIoT environment," *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 467–476, Jan. 2022.
- [45] H. Chen, Z. Qin, Y. Ding, L. Tian, and Z. Qin, "Brain tumor segmentation with deep convolutional symmetric neural network," *Neurocomputing*, vol. 392, pp. 305–313, Jun. 2020.
- [46] H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, and P.-A. Heng, "Fast ScanNet: Fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1948–1958, Aug. 2019.
- [47] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6848–6856.
- [48] S. Mehta and M. Rastegari, "MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer," 2021, *arXiv:2110.02178*.
- [49] Y. Chen et al., "Mobile-former: Bridging MobileNet and transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5270–5279.
- [50] J. Pan et al., "EdgeViTs: Competing light-weight CNNs on mobile devices with vision transformers," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 294–311.
- [51] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
- [52] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
- [53] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [55] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
- [56] L. Sifre and S. Mallat, "Rigid-motion scattering for texture classification," 2014, *arXiv:1403.1687*.
- [57] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1800–1807.



**Peng Xiao** received the M.Sc. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2018, where he is currently pursuing the Ph.D. degree with the School of Information and Software Engineering.

His research interests include medical image analysis, computer vision, and model compression.

Mr. Xiao is a Reviewer of the *Journal of Biomedical and Health Informatics*.



**Zhen Qin** (Member, IEEE) received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2012.

He is currently a Professor with the School of Information and Software Engineering, UESTC. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. His research interests include data fusion analysis, mobile social networks, wireless sensor networks, and image processing.



**Yi Ding** (Member, IEEE) received the B.S. degree in software engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2008, and the Ph.D. degree in computer science from Dublin Institute of Technology, Dublin, Ireland, in 2012.

He is currently a Full Professor with the School of Information and Software Engineering, and a Researcher with the Network and Data Security Key Laboratory of Sichuan Province, UESTC. His research interests include machine learning, deep learning, information security, medical image processing, and computer-aided diagnosis.



**Dajiang Chen** (Member, IEEE) received the Ph.D. degree in information and communication engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2014.

He is currently an Associate Professor with the School of Information and Software Engineering, UESTC. He was a Postdoctoral Fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, from 2015 to 2017.

Dr. Chen served as the Workshop Chair for BDEC-SmartCity'19 (in conjunction with IEEE WiMob 2019). He also serves/served as a Technical Program Committee Member for IEEE Globecom, IEEE ICC, IEEE VTC, IEEE WPMC, and IEEE WF-5G. His current research interests include physical-layer security, secure channel coding, and machine learning and its applications in wireless network security and wireless communications.



**Fuhu Deng** received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2006 and 2009, respectively, and the Ph.D. degree from Dublin Institute of Technology, Dublin, Ireland, in 2014.

He is currently an Associate Professor with the School of Information and Software Engineering, UESTC. His research interests include wireless communications, WLAN, big data analysis, artificial intelligence, and image processing.



**Ning Zhang** (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2015.

He is an Associate Professor with the Department of Electrical and Computer Engineering at University of Windsor, Canada. After that, he was a Postdoctoral Research Fellow with the University of Waterloo and University of Toronto, Toronto, ON, Canada, respectively. His research interests include connected vehicles, mobile-edge computing,

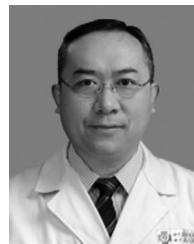
wireless networking, and machine learning.

Dr. Zhang received the NSERC PDF Award in 2015 and six Best Paper Awards from IEEE Globecom in 2014, IEEE WCSP in 2015, IEEE ICC in 2019, IEEE ICC in 2019, the IEEE Technical Committee on Transmission Access and Optical Systems in 2019, and *Journal of Communications and Information Networks* in 2018, respectively. He is a Highly Cited Researcher (Web of Science). He serves as an Associate Editor for the IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, and IEEE SYSTEMS JOURNAL; and a Guest Editor for several international journals, such as the IEEE WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He also serves/served as a TPC chair for IEEE SAGC 2020, a Track Chair for several international conferences, including IEEE VTC 2020, IEEE ICC 2022, AICON 2020, and CollaborateCom 2020, and a Co-Chair for numerous international workshops.



**Zhiguang Qin** received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1996.

He is the Full Professor with the School of Information and Software Engineering, UESTC, where he is also the Director of the Key Laboratory of New Computer Application Technology and UESTC-IBM Technology Center. His research interests include medical image processing, computer networking, information security, cryptography, information management, intelligent traffic, electronic commerce, distribution, and middleware.



**Minghui Pang** received the Ph.D. degree from Southern Medical University, Guangzhou, China, in 2012.

He is currently a Chief Physician with the Department of Geriatric General Surgery, Sichuan Provincial People's Hospital, Chengdu, China. He is also a Full Professor with the Medical School, University of Electronic Science and Technology of China, Chengdu. He was a Visiting Scholar with the Department of Harvard Medical School, Boston, MA, USA. His research interests include rectal cancer, colon cancer, gastric cancer, gastrointestinal stromal tumor, and abdominal wall hernia.