Deep Regional Metastases Segmentation for Patient-Level Lymph Node Status Classification

LU WANG [1], **TIAN SONG** [1], (Member, IEEE), **TAKAFUMI KATAYAMA** [1],
XIANTAO JIANG [2], (Member, IEEE), **TAKASHI SHIMAMOTO** [1], (Member, IEEE),
AND JENQ-SHIOU LEU [3], (Senior Member, IEEE)

[1]Department of Electrical and Electrics Engineering, Tokushima University, Tokushima 770-8506, Japan
[2]Department of Information Engineering, Shanghai Maritime University, Shanghai 201306, China
[3]Department of Electrical and Computer Engineering, National Taiwan University of Science and Technology, Da'an District, Taipei 10607, Taiwan

Corresponding author: Tian Song (tiansong@ee.tokushima-u.ac.jp)

ABSTRACT Generally, automatic diagnosis of the presence of metastases in lymph nodes has therapeutic implications for breast cancer patients. Detection and classification of breast cancer metastases have high clinical relevance, especially in whole-slide images of histological lymph node sections. Fast early detection leads to huge improvement of patient's survival rate. However, currently pathologists mainly detect the metastases with microscopic assessments. This diagnosis procedure is extremely laborious and prone to inevitable missed diagnoses. Therefore, automated, accurate patient-level classification would hold great promise to reduce the pathologist's workload while also reduce the subjectivity of diagnosis. In this paper, we provide a novel deep regional metastases segmentation (DRMS) framework for the patient-level lymph node status classification. First, a deep segmentation network (DSNet) is proposed to detect the regional metastases in patch-level. Then, we adopt the density-based spatial clustering of applications with noise (DBSCAN) to predict the whole metastases from individual slides. Finally, we determine patient-level pN-stages by aggregating each individual slide-level prediction. In combination with the above techniques, the framework can make better use of the multi-grained information in histological lymph node section of whole-slice images. Experiments on large-scale clinical datasets (e.g., CAMELYON17) demonstrate that our method delivers advanced performance and provides consistent and accurate metastasis detection in clinical trials.

INDEX TERMS Breast cancer metastases, histological lymph node sections, patient level analysis.

I. INTRODUCTION

Breast cancer is the most frequent cancer among women worldwide. According to the statistics from International Cancer Research Center (ICRC) of the World Health Organization (WHO), breast cancer impacts 2.1 million each year, and causes the greatest number of cancer-related deaths among women. In 2018, it is estimated that 627,000 women died from breast cancer - that is approximately 15% of all cancer deaths among women. Globally, the incidence rates vary widely, and the age-standardized incidence rate is as high as 99.4 per 100,000 in North America. The inci-

dence rates in Eastern Europe, Southern America, Southern Africa, and Western Asia are slightly lower, however they are increasing at the same time. The improvement of survival rates can be mainly achieved through early detection and diagnosis, which require a large proportion of women to seek medical treatments, appropriate diagnosis, and treatment facilities.

Early detection and classification of breast cancer metastases in lymph nodes play an important role in appropriate diagnosis and treatment facilities. Lymph nodes are small glands that filter lymph, the fluid that circulates through the lymphatic system, as shown in Fig. 1.

The lymph nodes in the axilla are the first place for breast cancer to spread. Lymph node metastasis is one of the most

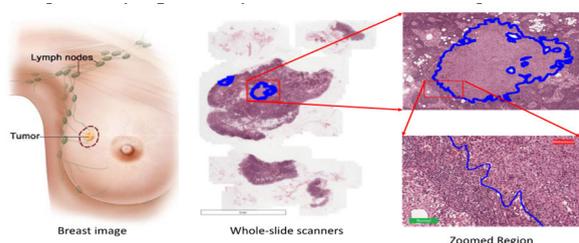The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

FIGURE 1. Example of a typical metastatic region. Whole-slide images (WSI) are generally stored in a multi-resolution pyramid structure. Each image in the pyramid is stored as a series of tiles to facilitate rapid retrieval of subregions of the image (https://camelyon17.grand-challenge.org/Background/). The regions of metastases have very similar appearances to the normal regions of scanned sections.

important prognostic factors in breast cancer [1], [2]. The prognosis is poor when cancer spreads to the lymph nodes. Therefore, lymph nodes are surgically removed and examined under endoscope. However, the diagnostic procedures are tedious and time-consuming. Most importantly, small metastases are very difficult to detect and are sometimes missed.

Currently, the UICC TNM classification system is an internationally recognized means of classifying the extent of cancer spread in patients with solid tumors. It is one of the most important tools to help clinicians select suitable treatment options and obtain prognostic indications. In breast cancer, UICC TNM staging takes into account the size of the tumor (T-stage), whether the cancer has spread to the regional lymph nodes (N-stage), and whether the tumor has metastasised to other parts of the body (M-stage). Since the histological assessment of lymph node metastases is an essential part of UICC TNM classification, in this paper we will focus on the pathologic N-stage (pN-stage).

Automatically detecting the regional metastases, especially from the whole-slide images, is a very challenging task due to the low contrast and similar structures between metastasised regions and various adjacent tissues or organs, as shown in Fig. 1. Moreover, the interference from other tumors makes the metastasised regions even more difficult to be separated from the background. In some cases, the metastasised regions are extremely small, which further complicates the task.

In this paper, we propose a deep regional metastases segmentation (DRMS) framework to predict pN-stage scores from patient's whole slide histopathology images. The proposed framework consists of four key modules: a region of interest (ROI) extraction module, a patch-level regional metastases localization module, a slide-level metastasis detection module and a patient-level lymph node classification module. First, the ROI extraction module generates candidate tissue regions from whole slide images (WSIs). Second, the patch-level regional metastases localization module adopts a deep segmentation network (DSNet) to predict cancer metastases within extracted ROIs. Third, the slide-level metastasis detection module adopts the DBSCAN

to fuse the predicted scores extracted from ROIs to build a slide-level lymph node prediction [3]. Finally, patient-level pN-stage is determined by aggregating slide-level predictions.

In summary, our contributions are four folds:

1. We propose a DRMS framework to predict pathologic N-stage (pN-stage) from patient's whole slide histopathology images.

2. We propose a multi-grained procedure for detecting cancer metastases on giga-pixel pathology images. With multi-grained analysis, the proposed method can capture rich information from pathology images.

3. We propose a deep segmentation network (DSNet) for patch-level metastases localization. The proposed network can aggregate the global-to-local context information of each patch.

4. Experiments on public benchmarks demonstrate that our approach achieves superior performance and obtains accurate results on metastases detection.

The rest of this paper is organized as follows. In Sect. 2, we give an overview of classical and deep learning methods for lymph node metastases detection, status classification and tumor segmentation. Then we introduce the proposed approach in Sect. 3. In Sect. 4, we evaluate and analyze the proposed method by extensive experiments and comparisons with other methods. Finally, we provide the conclusion and future work in Sect. 5.

II. RELATED WORKS

Our work aims to introduce deep learning-based segmentation techniques to detect and locate the metastases regions, and further predict the lymph node status. Some concerning works related to breast cancer metastases detection, lymph node status classification and deep learning-based tumor segmentation methods will be reviewed. For more details, we refer readers to comprehensive surveys [4], [5].

A. BREAST CANCER METASTASES DETECTION

Breast cancer has been studied in few decades due to its constructive effect for human health. Staging the breast cancer is a central component of treatment and management, which involves the microscopic examination of lymph nodes adjacent to the breast for evidence that the cancer has spread or metastasized [6], [7]. This process requires highly skilled pathologists and is quite time-consuming and error-prone, especially for lymph nodes with no or small tumors. Computer aided detection (CAD) of lymph node metastasis could improve the sensitivity, speed, and consistency of metastasis detection. However, only recently metastases detection has been introduced to the breast cancer diagnosis with power models. For example, Bejnordi first proposed a fully automated detection of DCIS in breast histopathology images [8]. Then, Bejnordi *et al.* proposed a context-aware convolutional neural network (CNN) for classification of breast carcinomas [9]. Bulten and Litjens [10] proposed an unsupervised prostate cancer detection on H&E images with deep

auto-encoders. Pinckaers and Litjens [11] trained CNNs with megapixel images for cancer localization. Though effectiveness, previous methods mainly consider low-resolution histopathology images. In modern hospitals, the scanning and digitization of whole histology slides are adopted to improve the tissue analysis for efficient cancer diagnosis and prognosis. How to handle the WSI for cancer metastases detection is still a challenging task.

B. LYMPH NODE STATUS CLASSIFICATION

For the lymph node status classification, well trained histopathologists can provide accurate classification results. However, due to the huge size of WSIs and the great number of potential cancer cases, fully automatic solutions for lymph node status classification is highly desirable. In the recent past, many related techniques have been proposed. The most impactful thing is organizing the challenges of CAMELYON16 and CAMELYON17. Several promising works have applied deep learning to lymph node status classification. For the CAMELYON16 challenge, the winner has shown an expressive performance for per slide and overall slide-level classification [12]. In their work, a modified GoogleNet is trained with a pre-sampled set of image patches to predict the coarse location of cancer metastases, then a random forest classifier is used to predict the slide label [13]. The CAMELYON organizers also trained CNNs on smaller datasets to detect breast cancer in lymph nodes and prostate cancer [14]. With new modules, several methods show state of the art results in lymph node status classification. However, they generally introduce a variety of image analysis tasks, resulting to high computation and memory requirements. Different from previous works, we observe that the granularity and variability of WSIs can help to improve the performances of lymph node status classification. Thus, we develop an integrated framework for WSI analysis. The proposed framework can make better use of the multi-grained information in histological lymph node section of whole-slice images.

C. HISTOPATHOLOGICAL TUMOR SEGMENTATION

Tumor segmentation in histopathology image can be coarsely categorized into hand-crafted feature-based methods and deep-learned feature-based methods. In the past decades, many hand-crafted features-based methods have been proposed to solve the tumor segmentation problems of histopathology image. For example, threshold-based methods, such as Otsu method classify foreground and background based on whether the intensity value is above a threshold [15]. Its segmentation performance has obvious bottlenecks for complex histopathology images. Due to the limited representation ability of the hand-crafted feature, the segmentation performance has not been effectively improved.

With the rapid development of deep learning technologies and large-scale datasets, several outstanding methods have been proposed in tumor segmentation to improve segmenta-

tion performance. Ronneberger *et al.* presented the U-shape network architecture consisting of a contracting path and a symmetric expanding path for biomedical image segmentation [16]. In general, U-Net has proven to be one of the most popular architectures for medical image segmentation tasks. Zhou *et al.* presented a novel U-Net++ architecture which combined a series of nested dense skip pathways [17]. Its segmentation performance is superior to other methods in medical image segmentation tasks. Qaiser *et al.* utilized the concept of persistent homology profiles (PHPs) for adaptive tumor segmentation [18]. Because the various appearances of tumors, hard samples are also one of the common challenges in tumor segmentation. The works in Lin *et al.* and Wang *et al.* adopted hard negative mining to obtain better performance in different tumor segmentation [12], [19], [20]. In this work, we propose a novel network for improving the ability of localizing the cancer regions. Besides, we also introduce a new multi-grained procedure for detecting cancer metastases on giga-pixel pathology images.

III. THE PROPOSED FRAMEWORK

Fig. 2 illustrates the overall procedure of our proposed framework. The whole framework consists of four key modules: a region of interest (ROI) extraction module, a patch-level regional metastases location module, a slide-level metastasis detection module and a patient-level lymph node classification module. We start with identifying the ROIs and removing non-information regions on the WSIs. Subsequently, we feed image patches extracted from the ROIs into our deep segmentation network (DSNet) for coarse metastases prediction. After that, we adopt DBSCAN to fuse the patch-level prediction and achieve the slide-level lymph node metastasis prediction [3]. Meanwhile, we extract corresponding features from the predicted metastases area. We finally train a XGBoost [21] classifier with these features to predict the pN-stage of each patient. In the following section, we will describe the key components in detail.

A. ROI EXTRACTION

For histopathology images, a typical whole slide image is approximately $200,000 \times 100.000$ pixels on the highest resolution level. If we deal with such large images directly, enormous computation is required because of the huge size of the slide. Besides, most of the regions have no useful information for metastases detection. Thus, directly dividing WSIs into patches is not appropriate for model training, as shown in Fig. 3. The Otsu threshold [15] method is widely used in recent studies for ROI extractions from the WSI because of its simple calculation and the segmentation is not affected by image brightness and contrast. In this work, to speed up the processing, we also adopt Otsu's adaptive thresholding technique on the low-resolution version of the WSIs to remove most of the non-tissue background quickly. Following [22], we convert RGB to gray from 32-times down-sampled WSI and then extract tissue regions by Otsu
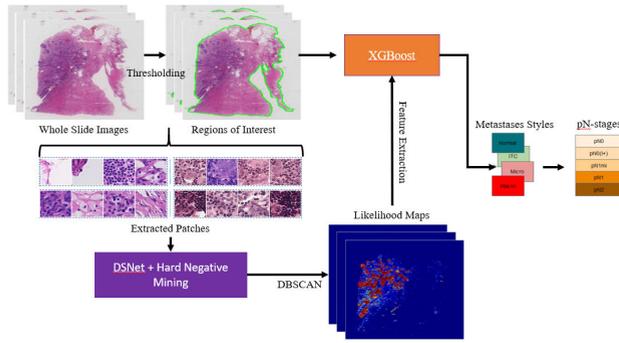
FIGURE 2. The overall procedure of our proposed framework. 1) Split the whole slide image into smaller patches; 2) Classify the extract patch-level images with our proposed DSNet and hard negative mining; 3) Generate likelihood maps based on DBSCAN cluster for slide level feature extraction and prediction; 4) Determine Patient-level pN-stages by each individual slide-level prediction.
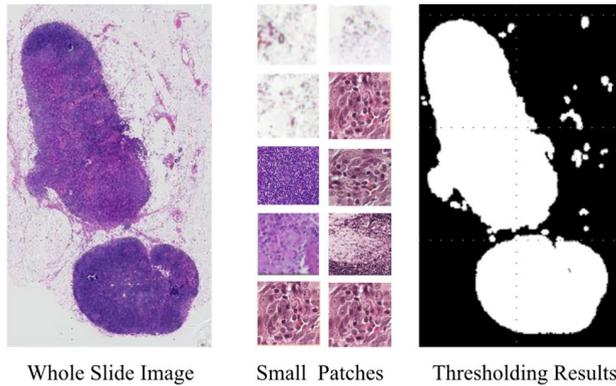

FIGURE 3. From whole slide images to patches.

FIGURE 4. Left: metastatic patches. Right: challenging normal patches.

FIGURE 5. The proposed deep segmentation network (DSNet) for patch-level metastases localization. The encoder part is the dilated ResNet-101 with an ASPP-like contextual structure. Besides, we introduce a pyramid attention module (PAM) to generate attention maps for feature enhancement. We also provide the scale-aware selection (SAS) module to select the contextual information. Finally, we use a skip-connection of low-layers for metastases region localization.

threshold method. We observed that metastasis regions are usually located at the edge of the tissue regions. Therefore, careful tissue region extraction method is needed. After the operation is done, holes and small points, which will influence feature extraction operation, still exist. Following previous works, morphological algorithms are operated to fill holes and clear isolated points. Then, we generate the patches through random grid sampling. Examples of the resulting patches are shown in Fig. 4. We can see that although most examples have texture patterns, it is difficult to distinguish between normal or metastatic patches.

B. PATCH-LEVEL METASTASES LOCALIZATION

The patch-level localization aims to predict the metastases regions at coarse granularities. To achieve this goal, we opt to a deep segmentation network, inspired by [23], as shown in Fig. 5. The DSNet takes in the patch images, extract multi-level features with dilated ResNet-101 [24]. (Dilation size = 2 at the last residual module), resulting in 1/16 resolution of input images and extracts locally weighted, multiscale contextual information with the proposed pyramidal attentional ASPP (PA-ASPP).
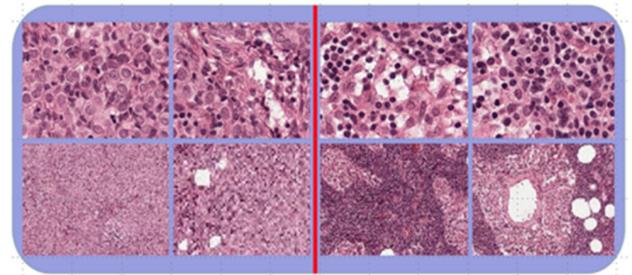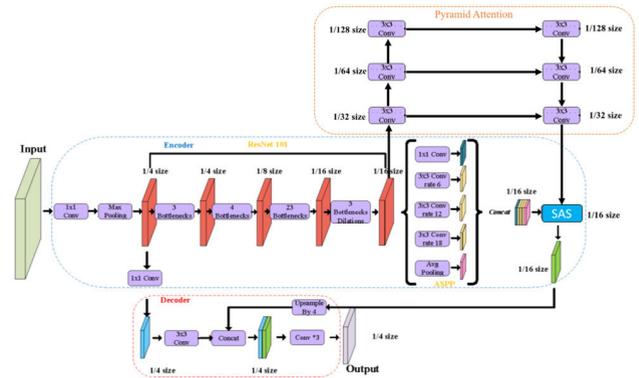
The network's ability to understand and utilize contextual information is critical to achieving accurate image segmentation. Increasing the network receptive field by consecutive down sampling is a straightforward way to capture contextual clues for dense prediction. However, the receptive field is restricted by the number of down-sampling operations performed, as performing down-sampling can cause spatial information loss. Inspired by the methods of SPP [25] and ASPP [23], [26] and the approach used by radiologists to locate regional metastases by observing neighboring contexts, we propose the PA-ASPP for extracting locally weighted, multiscale contextual information. The PA-ASPP feeds the input into two branches, one is to the pyramid feature aggregation (top branch) for providing pixel-level attention and the other one is to the ASPP (bottom branch) for providing rich multi-scale features and large receptive fields. The pyramid feature aggregation encodes the multi-level features from three different scales like U-Net [16]. This feature fusion incorporates neighboring scales of context features more precisely to produce better pixel-level attention.

The obtained attention is up-sampled to the same size of output of ASPP and used to weight the multi-scale features in pixel-level for selecting contextual information. The ASPP enriches contextual information by concatenating features

obtained from one 1×1 convolution, three 3×3 convolutions with dilation rates equal to (6, 12, 18), and one global average pooling layer and then passing through another 1×1 convolution to modify channel dimensions of multi-scale features before multiplying the attention maps. Finally, we use a scale-aware selection (SAS) module to locally weight multi-scale contextual features. More specifically, we first use a squeeze-and-excitation module to select the key channels from multi-scale features [27]. Then, the selected features are highlighted pixel-wisely by attention maps from Pyramid Attention Module (PAM). The SAS can capture important contexts of ASPP for the accurate and consistent dense prediction.

To train the proposed network, we adopt the input resolution of 512×512 to capture more detailed information. To increase the variation in the training set, we perform extensive data augmentation including random flipping, scaling, rotation over angles between 0 and 360, and cropping. Besides, our model is trained with hard negative mining to improve the discrimination. It means that for each epoch, model inferences whole patches of the slide and chooses patches whose intersection over union (IOU) with the annotated mask are less than 0.95 as the training set.

C. SLIDE-LEVEL METASTASES DETECTION
To capture the slide-level information, we should integrate the patch information to improve the spatial consistency. Following the work in [28], we introduce the DBSCAN algorithm to group together small metastases areas which were in close proximity.

D. SPATIENT-LEVEL CLASSIFICATION
After slide-level metastases detection, we can realize the patient-level classification based on the corresponding metastases regions. More specifically, each whole slide probability map is first converted into a feature vector after being regulated with 5 threshold values, i.e., 0.5, 0.6, 0.7, 0.8, 0.9, which is then used to train a pN-stage classifier. According to the morphological and geometrical information, we extract 13 types of features from each probability map (65 features in total). The task is to determine a pN-stage for each patient in the test dataset. To compose a pN-stage, the number of positive lymph nodes (i.e., nodes with a metastasis) need to be counted. According to the size of the tumor, there are three types of positive lymph nodes in clinic:

Macro-metastases: Metastases greater than 2.0 mm.

Micro-metastases: Metastases greater than 0.2 mm or more than 200 cells, but smaller than 2.0 mm.

Isolated Tumor Cell (ITC): Single tumor cells or a cluster of tumor cells smaller than 0.2 mm or less than 200 cells.

Although lymph nodes containing only ITCs are not counted as positive lymph nodes, if no macro-metastases or micro-metastases are found in the patient's lymph nodes, the pathologist must report ITC.

In this work, we follow previous works and adopt a simplified version of the pN-staging system in breast cancer. The task is to automatically determine the following applied pN-stages for each patient:

-pN0: No micro-metastases or macro-metastases or ITCs found.

-pN0(i+): Only ITCs found.

-pN1mi: Micro-metastases found, but no macro-metastases found.

-pN1: Metastases found in 1–3 lymph nodes, of which at least one is a macro-metastasis.

-pN2: Metastases found in 4–9 lymph nodes, of which at least one is a macro-metastasis.

After feature vectors are extracted, we use XGBoost to classify lymph nodes into four classes (Normal, ITC, Micro, Macro). Since the total number of WSIs is rather small, it is important to prevent overfitting. Thus, K-fold cross validation and other hyper-parameters, such as subsample or max depth, are adopted for preventing overfitting. Firstly, grid-search algorithm is used to search for the best hyper-parameters. Then, the XGBoost model is trained with ten-fold cross validation. Finally, the best model is selected and the 100 patients in the CAMELYON17 test set are predicted. The patient's pN-stage is graded according to the given rules.

E. LOSS FUNCTION
Tumor regions cover a very small proportion of pixels in WSIs, thereby leading to class imbalance. This issue was circumvented by training the network to minimize a hybrid loss function. In this work, the hybrid cost function comprised of a class-weighted cross-entropy loss and a loss function based on the Dice overlap coefficient. The Dice coefficient is an overlap metric used for assessing the quality of segmentation maps. The Dice loss is a differentiable function that approximates to Dice coefficient, it uses the predicted posterior probability map and ground truth binary image, as defined in Eq. 1. The weighted cross-entropy loss is defined in Eq. 2. In the equations, p_{ln} is the predicted posterior probability map, and g_{ln} is the ground truth image.

$$\mathcal{L}_{dice} = 1 - 2 \frac{\sum_l^L w_l \sum_n^N g_{ln} p_{ln}}{\sum_l^L w_l \sum_n^N g_{ln} + p_{ln}} \tag{1}$$

$$\mathcal{L}_{wce} = -\frac{1}{N} \sum_{n=1}^N (w r_n \log(p_n) + (1 - r_n) \log(1 - p_n)) \tag{2}$$

The total loss is defined as a linear combination of the two loss components as defined in Eq. 3.

$$\mathcal{L} = \alpha \mathcal{L}_{dice} + \beta \mathcal{L}_{wce} \tag{3}$$

The proposed networks are trained by minimizing the total loss. α and β are empirically assigned to the individual loss components. In this work we set $\alpha = \beta = 0.5$.

IV. EXPERIMENTS
In this section, we first introduce our experimental settings such as public benchmarks, training details, and evaluation

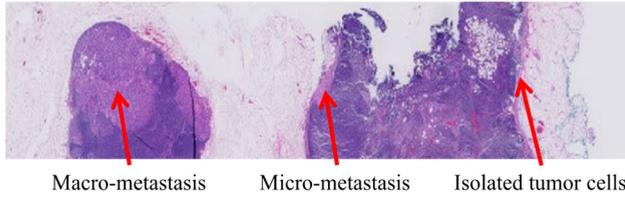Macro-metastasis   Micro-metastasis   Isolated tumor cells

FIGURE 6. Representative samples of the different sizes of breast cancer metastases in sentinel lymph nodes.

metrics. Then we report the experimental results on public benchmarks. Finally, we analyze the key components with extensive ablation experiments.

A. EXPERIMENTAL SETTINGS

1) DATASETS
In this work, we evaluate our framework on CAMELYON17 dataset [29]. However, following previous works, we also used the WSIs from CAMELYON16 as the training dataset. The CAMELYON16 dataset contains 400 WSIs with region annotations for all its metastasis slides [30]. The WSIs are collected from two different medical centers. The CAMELYON17 dataset contains 1000 WSIs with 5 slides per patient: 500 slides for the train set, 500 slides for the test set. The WSIs are collected from five different medical centers. Typically, different sizes of breast cancer metastases are in sentinel lymph nodes, as shown in Fig. 6. Since the CAMELYON17 dataset provides only 50 slides with region annotations, we split 100 patients (500 slides) into 43 patients for the train set, 57 patients for the validation set for hyper-parameter tuning. In detail, if a patient's any slide includes region annotation, we allocate that patient as a train set. For training of the patch-level localization, 400 WSIs from CAMELYON16 dataset and 160 WSIs (50 WSIs with region annotation and 110 negative WSIs) from CAMELYON17 train set are used. For training of the slide-level lymph node classifier, we use 285 WSIs (57 patients) from CAMELYON17 validation set.

2) EVALUATION METRICS
To evaluate the performance of different methods, we adopt the quadratic weighted Cohen's kappa as the evaluation metric [31]. Given n test patients and m categories (pN-stages), let n_{ij} denote the number of patients with the pN_i-stage that were categorized to pN_j-stage. Let r_i denote the total number of patients with the pN_i-stage and s_j the total number of patients categorized to the pN_j-stage. Finally, let w_{ij} denote the disagreement weight associated with the pN_i and pN_j. The weight matrix is

The weight matrix is

$$w_{ij} = (i - j)^2, \quad i, j \in [1, 2, \ldots m] \quad (4)$$

The mean observed degree of disagreement is

$$D_0 = \frac{1}{n} \sum_i^m \sum_j^n n_{ij} w_{ij} \quad (5)$$

The mean degree of disagreement expected by chance is

$$D_e = \frac{1}{n^2} \sum_i^m \sum_j^n r_i s_j w_{ij} \quad (6)$$

The weighted kappa is then defined by

$$k_w = \frac{D_e - D_0}{D_e} \quad (7)$$

The k_w metric ranges from -1 to $+1$: a negative value indicates lower than chance agreement, zero indicates exact chance agreement, and a positive value indicates better than chance agreement. As pN-stages are ordinal, choosing a quadratic weighted kappa could penalize misclassification more severely when errors are more than one stage apart.

3) TRAINING DETAILS
During training and inference, we first perform the thresholding for localizing the ROIs, then extracted 512×512 patches from WSIs at the highest magnification level resolution. We adopt the ImageNet pre-trained ResNet101 (He et al. 2016) as the backbone with dilation convolutions. New layers are initialized with the MSRA method. The deep learning toolbox TensorFlow was used to train the model with a NVIDIA GTX TITAN X GPU [32]. We used the Adam optimization method with a learning rate 1e-4. The network was trained for 4 epochs with a batch size 32 [33]. For the post-processing, we regulated tumor probability heat map with 5 threshold values, i.e., 0.5, 0.6, 0.7, 0.8, 0.9. Given a heat map, we extracted 13 features including the major axis length of the tumor region (Feature1), maximum probability score (Feature2), average probability score (Feature3), a total area of the tumor region (Feature4), max, mean, variance of eccentricity (Feature5-7), orientation (Feature8-10) and solidity (Feature11-13) from the tumor regions. We built a XGBoost classifier to discriminate lymph node classes using extracted features. Finally, each patient's pN-stage was determined by the given rule together with the lymph node slide prediction result.

B. COMPARISONS WITH OTHER STATE-OF-THE-ART METHODS
Using a single segmentation model, our slide-level lymph node classification model and patient-level pN-stage prediction achieved 0.9351 accuracy and 0.9017 quadratic weighted kappa score, respectively. We trained additional models with different data augmentation. Finally, four models were ensembled by averaging probability heat map and reached 0.9410 slide-level accuracy and 0.9632 patient-level quadratic weighted kappa score. Tab. 1 shows the published results on CAMELYON17 testing dataset. Our best performing model gave a Cohen's kappa score of 0.9473, which is better than the second in Open Leaderboard (out of 120 valid submission entries). According to the brief description of the top-ranked model, it also used DeepLabv3+ as the basic framework for patch level segmentation. However, the

TABLE 1. Performance comparison with other approaches for automated pN-Staging in CAMELYON17 challenge. The score of Cohen Kapa is from the open public leader board.

Methods	KappaScore
PSPNet (PPM) (Zhao et al. 2017)	0.9012
PSPNet (PPM→ ASPP)	0.9188
PSPNet (PPM→PA-ASPP)	0.9341
DeepLabv2 (L.-C. Chen, Papandreou, et al. 2018)	0.9002
DeepLabv3+ (L.-C. Chen, Zhu, et al. 2018)	0.9213
Ours w/o PA	0.8422
Ours w/o SAS	0.9327
Ours	0.9473

(a) (b) (c) (d)

FIGURE 8. Visual comparisons with different contextual modules. From left to right: Ground truth, PPM \cite{zhao} ASPP \cite{chen2} and our PA-ASPP. Our approach shows better results in coherence.
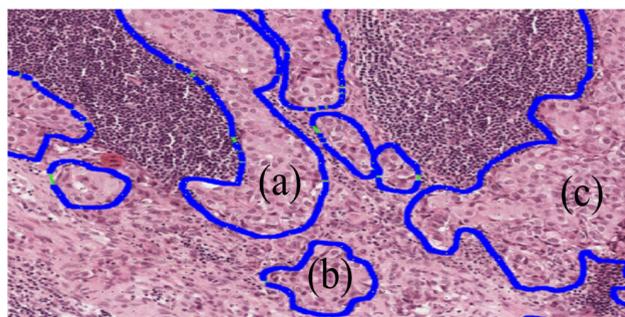

FIGURE 7. Slide-level predictions of our proposed framework in different cases. Our methods can show good results in continuous regions (a), isolated regions (b) and low-contrast regions (c).

top-ranked model used the DenseNet with high generalization ability as the backbone of the Encoder. Our proposed DSNet+ adopted Resnet as the backbone and improve the DeepLabv3+ framework by introducing the PA and SAS attentions to enhance the model generalization ability. In the future, we will explore more advanced segmentation frameworks and backbone networks for the dedicated task.

Fig. 7 shows an example for slide-level predictions. From the results, we can see that our method can achieve outstanding performance in different cases, including continuous regions (a), isolated regions (b) and low-contrast regions (c). In (a) the predicted whole tumor segmentation was similar to pathologist provided ground truth of whole tumor region and most of the samples in the dataset fell into this category.

C. ABLATIONS STUDIES

Comprehensive experiments are conducted to demonstrate the effectiveness of the proposed method. The 5-fold cross-validation is applied to validate the stability and generalization of networks. It is worth mentioning that all models are trained following the same experimental settings.

1) EFFECTS OF THE PA-ASPP AND SAS

The proposed PA-ASPP is used to aggregate the global-to-local context information. The PPM [25] and ASPP [23] also aggregate multi-scale contextual features for semantic labeling. To verify the effects of our PA-ASPP, we adopt
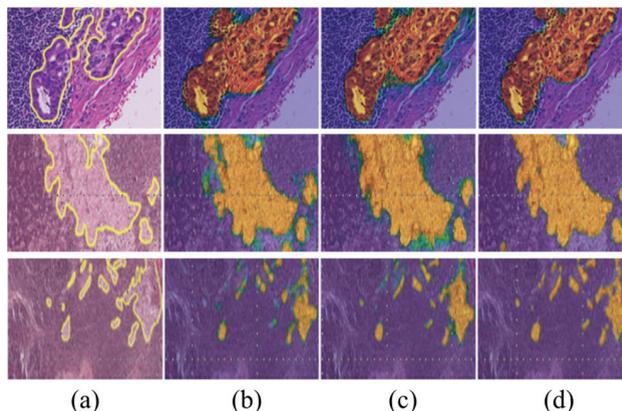
the PSPNet [25] with the ResNet-101 backbone for a fair comparison. We replace the PPM of PSPNet with our PA-ASPP and ASPP modules. The 2-4 rows of Tab. 1 show the quantitative results. With our PA-ASSP module, the Cohen Kappa score improves about 3.3% and 1.6% over the PPM and ASPP, respectively. Visual comparisons of PPM, ASPP and our PA-ASPP are shown in Fig. 8. It can be observed that our segmentation results constantly contain more accurate and coherent structures. These results demonstrate that our proposed PA-ASPP is more effective than previous methods and can generate more coherent results. Besides, with the SAS, the model can adaptively capture the regions with different scales, which shows the effectiveness.

2) EFFECTS OF HARD EXAMPLE MINING

Previous patch extraction scheme from the whole-slide images is based on random uniform sampling. However, this leads to some hard examples being excluded from the training set which resulted in the model's poor performance on such regions of WSI. Therefore, we attempt to solve this issue by hard example mining the poorly performing regions in WSI and fine-tuning the trained model with this hard-mined set. To cancel the effects of the PA-ASPP module, we apply them to the DeepLabv3+ for direct performance comparison. Tab. 2 shows the quantitative performances. Compared with the original DeepLabv3+, the HEM can consistently improve the classification results. Note that the HEM is a general proposal. It can be used in varied settings for training segmentation networks.
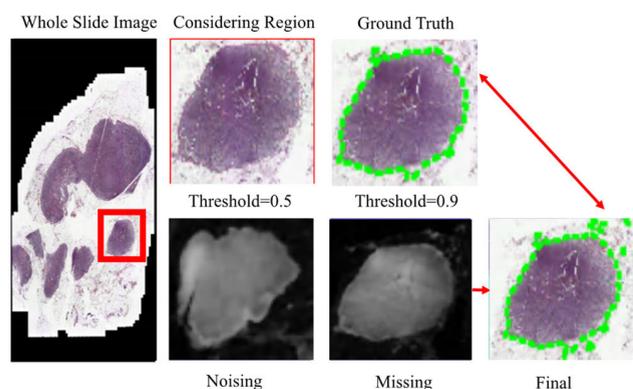
3) EFFECTS OF THRESHOLDING REGIONS

For predicting patient-level results, we convert each whole-slide probability map into a feature vector with 5 threshold values, i.e., 0.5, 0.6, 0.7, 0.8, 0.9. Based on the regions, we train a lymph node classifier. To verify the effect, we show the visual results in Tab. 3. As we can see, at a low threshold, the final prediction can include noise for the

TABLE 2. Ablation results on the camelyon17 validation dataset.

Rank	Team	Cohen Kapa	Date
1	shlee	0.9570	12 Feb. 2019
2	nicolas.pinchaud1	0.9386	6 Dec. 2019
3	Jarveelee3	0.9243	28 Feb. 2020
4	mahendrakhened	0.9090	6 April 2020
5	ozymandias.watchman	0.9085	25 March 2019
6	GigapixelAI	0.9058	12 Sep. 2019
7	kvougas	0.9052	27 Nov. 2019
8	Erizz	0.9046	10 March 2019
9	Jaylee	0.8993	31 July 2017
	ours	0.9473	25 June. 2021

TABLE 3. Effects of hard example mining (HEM).

Methods	KappaScore
DeepLabv3+ without HEM	0.9365
DeepLabv3+ with HEM	0.9473

FIGURE 9. Slide-level predictions of our proposed framework with different thresholding values.

cancer localization. While at a high threshold, it will miss some key regions. Our multiple thresholds can effectively highlight the most of regions. Fig. 9 shows the effects with different thresholding values.

4) LIMITATIONS

We address the patient-level lymph node status classification problem by regional metastases segmentation of gigapixel whole-slide images using the divide and conquer strategy. The proposed approach shows superior performance when compared to its competitors. However, our method still shows low accuracy for ITC, and the model size is still very large for real applications. Besides, the patch extraction is based on the Otsu thresholding. In some cases of CAMELYON2017, the complex background is still be included.

V. CONCLUSION

In this work, we present a fully automatic metastases detection framework to predict pN-stage of patient-level lymph node statuses from the whole-slide histopathology images.

Several useful strategies are proposed to improve the regional segmentation accuracy and integrate multi-granularity representations for the final classification. Our method adopts a divide and conquer strategy and allows the model to process the whole-slide images. More specifically, our method first divides the whole-slide images into tumor-related patches, then performs tumor segmentation on regional patches, and integrates segmented patches to generate the segmentation of the entire WSI. More importantly, we propose a new deep segmentation network (DSNet) with pyramid attentions and scale-aware selection so that the expressive multi-level features can be emphasized to facilitate the feature refinement. Extensive comparative experiments with 5-fold cross-validation on CAMELYON17 demonstrate that our proposed method can significantly outperform several state-of-the-art methods in typical metrics. Due to the characteristics of digital pathological images, it's impossible to manually annotated large quantity qualified datasets for model training, [34] has proposed a weakly supervised deep learning framework to solve the issue. To enable clinical application, we will improve the patch sampling method to generate more valuable training examples; and we will improve the system framework to enable weakly supervised learning and semi-supervised learning. On the other side, according to the review article [35], the applications for DL based pathology analytics for Breast Cancer are still focused on tumor detection, the basic image analysis and at internal validation stage. To enable the advanced clinical applications, we will explore more advanced segmentation frameworks and backbone networks for the dedicated task, and we will build an end-to-end learning framework to predict the pN phase of the WSI and enable the accurate grading and subtyping clinical applications.

REFERENCES

[1] B. E. Bejnordi, M. Veta, P. J. V. Diest, B. V. Ginneken, N. Karssemeijer, G. Litjens, A. W. M. Jeroen van der Laak, and T. C. Consortium, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, pp. 2199–2210, Dec. 2017, doi: 10.1001/jama.2017.14585.

[2] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *J. Cancer Stem Cell Res.*, vol. 4, no. 3, p. 1, 2016.

[3] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. Trans. Knowl. Discovery Data, (ACM)*, 1996, pp. 226–231.

[4] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[5] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017, *arXiv:1704.06857*. [Online]. Available: http://arxiv.org/abs/1704.06857

[6] S. K. Apple, "Sentinel lymph node in breast cancer: Review article from a Pathologist's point of view," *J. Pathol. Transl. Med.* vol. 50, no. 2, p. 83, 2016.

[7] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Med.*, vol. 25, no. 1, pp. 24–29, Jan. 2019.

[8] B. E. Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer, and J. A. W. M. van der Laak, "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.

[9] B. E. Bejnordi, G. Zuidhof, M. Balkenhol, M. Hermsen, P. Bult, B. van Ginneken, N. Karssemeijer, G. Litjens, and J. van der Laak, "Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images," *J. Med. Imag.*, vol. 4, no. 4, 2017, Art. no. 044504.

[10] W. Bulten and G. Litjens, "Unsupervised prostate cancer detection on H&E using convolutional adversarial autoencoders," 2018, *arXiv:1804.07098*. [Online]. Available: http://arxiv.org/abs/1804.07098

[11] H. Pinckaers and G. Litjens, "Training convolutional neural networks with megapixel images," 2018, *arXiv:1804.05712*. [Online]. Available: http://arxiv.org/abs/1804.05712

[12] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," 2016, *arXiv:1606.05718*. [Online]. Available: http://arxiv.org/abs/1606.05718

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[14] G. Litjens, C. I. Sanchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-Van De Kaa, P. Bult, B. Van Ginneken, and J. Van Der Laak, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Sci. Rep.*, vol. 6, May 2016, Art. no. 26286.

[15] 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.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI* in Lecture Notes in Computer Science, 2015, pp. 234–241.

[17] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Lecture Notes in Computer Science), vol. 11045, D. Stoyanov *et al.*, Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-030-00889-5_1.

[18] T. Qaiser, Y.-W. Tsang, D. Taniyama, N. Sakamoto, K. Nakane, D. Epstein, and N. Rajpoot, "Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features," *Med. Image Anal.*, vol. 55, pp. 1–14, Jul. 2019.

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

[20] 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.

[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[22] B. Lee and K. Paeng, "Breast cancer stage classification in histopathology images," in *Submission Results Camelyon17 Challenge*, Nov. 2019. [Online]. Available: https://grand-challenge-public-prod.s3.amazonaws.com/evaluation-supplementary/80/dd7ab09f-dcdf-4bf6-be6d-2b070c9bfbd4/lunit_camelyon17_paper.pdf

[23] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.

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

[25] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[26] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. ECCV*, Sep. 2018, pp. 801–818.

[27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[28] S. Lee, S. Oh, K. Choi, and S. Kim, "Automatic classification on patient-level breast cancer metastases," in *Submission Results Camelyon17 Challenge*. Nov. 2019. [Online]. Available: https://grand-challenge-public-prod.s3.amazonaws.com/evaluation-supplementary/80/46fc579c-51f0-40c4-bd1a-7c28e8033f33/Camelyon17_.pdf

[29] P. Bandi, O. Geessink, Q. Manson, M. Van Dijk, M. Balkenhol, M. Hermsen, B. E. Bejnordi, B. Lee, K. Paeng, A. Zhong, and Q. Li, "From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 550–560, Feb. 2019.

[30] G. Litjens, P. Bandi, B. E. Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. van de Loo, R. Vogels, Q. F. Manson, N. Stathonikos, A. Baidoshvili, P. van Diest, C. Wauters, M. van Dijk, and J. van der Laak, "1399 H&E-stained sentinel lymph node sections of breast cancer patients: The CAMELYON dataset," *GigaScience*, vol. 7, no. 6, Jun. 2018, Art. no. giy065.

[31] J. L. Fleiss and J. Cohen, "The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability," *Educ. Psychol. Meas.*, vol. 33, no. 3, pp. 613–619, 1973.

[32] M. Abadi *et al.*, "TensorFlow: Large–scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv-export-lb.library.cornell.edu/abs/1603.04467

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[34] G. Campanella, G. M. Hanna, L. Geneslaw, A. Miraflor, V. W. K. Silva, J. K. Busam, E. Brogi, E. V. Reuter, S. D. Klimstra, and J. T. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Med.*, vol. 25, no. 8, pp. 1301–1309, Aug. 2020.

[35] A. Echle, N. T. Rindtorff, T. J. Brinker, T. Luedde, A. T. Pearson, and J. N. Kather, "Deep learning in cancer pathology: A new generation of clinical biomarkers," *Brit. J. Cancer*, vol. 124, no. 4, pp. 686–696, Feb. 2021.

LU WANG received the B.E. and M.E. degrees from Dalian University of Technology, China, in 1995 and 1998, respectively. She is currently pursuing the Ph.D. degree with the Department of Electrical and Electrics Engineering, Tokushima University. She has more than 20 years of working experience in ICT industry. Her current research interests include medical image processing, video coding algorithms, and hardware design.

TIAN SONG (Member, IEEE) received the B.E. degree from Dalian University of Technology, China, in 1995, and the M.E. and Dr.Eng. degrees from Osaka University, in 2001 and 2004, respectively. In 2004, he joined Tokushima University as an Assistant Professor. He is currently an Associate Professor with the Department of Electrical and Electronic Engineering, Graduate School of Advanced Technology and Science, Tokushima University. His current research interests include video coding algorithms, VLSI architectures, and system design methodology. He is a member of IEICE and RISP.
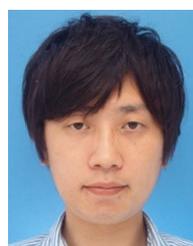
TAKAFUMI KATAYAMA received the B.E. and M.E. degrees in electrical engineering from Tokushima University, where he is currently pursuing the Ph.D. degree in electrical engineering. He belonged to Renesas Electronics Corporation, from 2012 to 2014. His current research interests include video coding algorithms and hardware design.

XIANTAO JIANG (Member, IEEE) received the M.E. degree from Shanghai Maritime University (SMU), in 2012, and the dual Ph.D. degree from Tongji University, China, and Tokushima University of Japan, in 2016. After that, he joined the College of Information Engineering, Shanghai Maritime University, as an Assistant Professor. He held a postdoctoral position at Carleton University, Ottawa, Canada, from 2019 to 2021. His research interests include blockchain, the Internet of Vehicles, and deep learning.

TAKASHI SHIMAMOTO (Member, IEEE) received the B.E. and M.E. degrees in electrical engineering from Tokushima University, in 1982 and 1984, respectively, and the Dr.Eng. degree from Osaka University, in 1992. In 1984, he joined Tokushima University as an Assistant Professor. He is currently a Professor with the Department of Electrical and Electronic Engineering, Graduate School of Advanced Technology and Science, Tokushima University. His research interest includes heuristic algorithms for VLSI CAD. He is a member of IEICE and RISP.

JENQ-SHIOU LEU (Senior Member, IEEE) received the B.S. degree in mathematics and the M.S. degree in computer science and information engineering from the National Taiwan University, Taipei, Taiwan, and the Ph.D. degree on a part-time basis in computer science from the National Tsing Hua University, Hsinchu, Taiwan. He was with Rising Star Technology, Taiwan, as an Research and Development Engineer, from 1995 to 1997, and in the telecommunication industry (Mobitai Communications and Taiwan Mobile), from 1997 to 2007, as an Assistant Manager. In February 2007, he joined the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology (NTUST), as an Assistant Professor. From February 2011 to January 2014, he was an Associate Professor. Since February 2014, he has been a Professor. He served as the Department Chairman, from August 2017 to July 2020. His research interests include heterogeneous network integration, mobile service and platform design, cybersecurity, distributed computing, and green and orange technology integration. He has published extensively in these areas, with 78 SCI indexed journal articles, 65 conference papers or book chapters, and led 19 MOST projects, 28 industry-academia projects, and ten cross-university projects after he joined NTUST.

● ● ●