IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 8, 2020, accepted June 20, 2020, date of publication June 26, 2020, date of current version July 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005180

Automated Gleason Grading and Gleason Pattern Region Segmentation Based on Deep Learning for Pathological Images of Prostate Cancer

YUCHUN LI[1], MENGXING HUANG[1], (Member, IEEE), YU ZHANG[1], (Member, IEEE), JING CHEN[2,3], HAIXIA XU[2,3], GANG WANG[2,3], AND WENLONG FENG[1]

[1]State Key Laboratory of Marine Resource Utilization in South China, College of Information Science and Technology, Hainan University, Haikou 570228, China
[2]Haikou Municipal People's Hospital, Haikou 570100, China
[3]Xiangya Medical College Affiliated Hospital, Central South University, Haikou 570288, China

Corresponding authors: Mengxing Huang (huangmx09@163.com) and Yu Zhang (yuzhang_nwpu@163.com)

This work was supported in part by the Foundation Item: Key Research and Development Project of Hainan Province under Grant ZDYF2019020, in part by the National Key Research and Development Program of China under Grant 2018YFB1404400, and in part by the Education Department of Hainan Province under Grant Hnky201922.

ABSTRACT Prostate cancer is the second-deadliest cancer in men in the United States, seriously affecting people's life and health. The Gleason grading system is one of the most reliable methods to quantify the invasiveness of prostate cancer, which is of great significance for risk assessment and treatment planning for patients. However, the task of automating Gleason grading is difficult because of the complexity of pathological images of prostate cancer. This paper presents an automated Gleason grading and Gleason pattern region segmentation method based on deep learning for pathological images of prostate cancer. An architecture combining the atrous spatial pyramid pooling and the multiscale standard convolution is proposed for the segmentation of the Gleason pattern region to get accurate Gleason grading. In addition, the postprocessing procedure based on conditional random fields is applied to the prediction. The quantitative experiments on 1211 prostate cancer tissue microarrays demonstrate that our results have a high correlation with the manual segmentations. The mean intersection over union and the overall pixel accuracy for the Gleason pattern region are 77.29% and 89.51%, respectively. Furthermore, the results of the automatic Gleason grading were comparable to the results of experienced pathologists. The inter-annotator agreements between the model and the pathologists, quantified via Cohen's quadratic kappa statistic, was 0.77 on average. Our study shows that the method of combining different deep neural network architectures is suitable for more objective and reproducible Gleason grading of prostate cancer.

INDEX TERMS Prostate cancer, gleason grading, image segmentation, deep learning, atrous spatial pyramid pooling, computer-aided diagnosis.

I. INTRODUCTION

Prostate cancer is the second-deadliest cancer in men in the U.S., seriously affecting people's life and health [1]. Prostate cancer mainly refers to a malignant tumor of an epithelial tissue that occurs in the prostate. In recent years, prostate cancer has gradually increased. At present, pathologists take out a small amount of prostate tissue using ultrasound-guided prostate biopsy technology [2]. By a series of advanced methods such as those using microscope, histochemistry, and immunofluorescence, they observe and analyze the pathology in the obtained prostate tissue to judge on the nature, type, differentiation degree, and classification of prostate tissue. Normal prostate tissue consists of stroma and regularly arranged glands. However, cancer tissue has epithelial cells that will replicate and destroy the regular arrangement of glandular units. In high-grade cancers, the stroma and lumen are generally replaced by epithelial cells.

The Gleason grading system is one of the most reliable methods to quantify the invasiveness of prostate cancer [3]. Specifically, prostate cancer is classified according to the system established by Donald Gleason in 1966 [4]. The Gleason grading system was revised by the International Society for Urological Pathology (ISUP) in 2014 and was

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

acknowledged by the World Health Organization (WHO) [5]. It remains the most powerful prognostic tool, although the histological Gleason grading system has been modified according to the clinical diagnosis of prostate cancer [6]. Prostate cancer is divided into five grades according to the Gleason grading system [7]. The histology of grade 1 is characterized by a dense arrangement of well-differentiated glands, forming well-defined nodules (Figure 1(a)); Grade 2 shows that the well-differentiated glands are more loosely arranged, forming clearer nodules (Figure 1(b)); Grade 3 appears as scattered, independent, well-differentiated glands (Figure 1(c)); Grade 4 manifests as poorly differentiated, fused, or sieve-like (including glomerular-like structures) glands (Figure 1(d)); Grade 5 is characterized by a lack of glandular differentiation and necrosis (Figure 1(e)). The higher the grade, the worse the prognosis. Gleason grading is the sum of the tumor's major component and minor component ($> 5\%$). If there is no minor component, the Gleason score would be twice the major component pattern. Otherwise, the sum of the primary and secondary pattern is the Gleason score, such as 7 $(3 + 4)$ or $3 + 4 = 7$. Special situations exist: if the composition of the tumor is in two hierarchical forms and the proportion of tumors of the secondary component is $\leq 5\%$, and the secondary component has a lower pattern, then Gleason score = primary component pattern + primary component pattern; otherwise, if the secondary component's pattern is higher, then Gleason score = primary component pattern + secondary component pattern; if the tumor component has more than 2 hierarchical forms, then Gleason score = primary component pattern + highest level pattern [8]. Although the Gleason grading system is gradually applied, up to now, most Gleason pattern have been manually assigned by pathologists. This process is very time-consuming and affected by inter- and intra-observer variability [9]. This problem is especially prominent when distinguishing between Gleason pattern 3 and Gleason pattern 4, which will directly affect the follow-up treatment because $3 + 4$ is assigned to Gleason grading 2, but $4 + 3$ is Gleason grading 3 [10], [11].

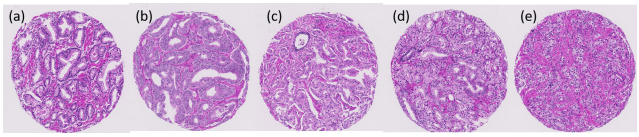

FIGURE 1. Gleason grading example of the pathological image. (a): Gleason grading 1, (b): Gleason grading 2, (c): Gleason grading 3, (d): Gleason grading 4, (e): Gleason grading 5.

With the rapid development and application of automatic computer-aided diagnosis technology and the corresponding progress in the available computing power of whole-slide microscopic imaging, it is feasible to develop an automatic Gleason grading system for prostate cancer. Huang *et al.* [12] proposed a clinical decision support framework to transform heterogeneous health data from different sources. Li *et al.* [13] used machine learning to detect and segment medical images and achieved good results. Therefore, automatic Gleason grading of prostate cancer with computer-aided design tools will help pathologists save time and identify more accurate treatment.

In this paper, we propose a novel method for automatic Gleason grading and Gleason pattern region segmentation of images with prostate cancer pathologies based on a convolutional neural network (CNN). An architecture that combines the atrous spatial pyramid pooling (ASPP) from Deeplab-V3 [14] and the multiscale standard convolution inspired by a multiscale parallel branch convolutional neural network (MPB-CNN) [15] is proposed for the segmentation of the Gleason pattern region (not gland segmentation) to get accurate Gleason grading. In addition, we improved the network by combining the standard convolution with atrous convolution, and postprocessing based on a conditional random field (CRF) is applied to the prediction. Our approach yields higher performance on Gleason grading than previous work. Furthermore, the Gleason pattern assignment of the model achieved stratification of pathologists and divided patients into groups with different prognosis.

II. RELATED WORK

This section introduces the related work on traditional Gleason grading algorithms and applications of convolutional neural networks in medical image segmentation.

A. TRADITIONAL GLEASON GRADING ALGORITHMS

There have been many studies on computer-aided Gleason scoring for prostate cancer diagnosis. A common approach to machine learning is to extract features and apply classifiers to the selected features, such as Bayesian classifier [16], support vector machines (SVM) [17], or random forests [18]. Smith *et al.* [19] proposed a similarity measurement method that extracted the texture features and assigned the input image to Gleason levels 1 to 3 and the combined grades of 4 and 5 by the nearest neighbor classifier. Wetzel *et al.* [20] studied whether computational geometry abstractions, such as spanning trees, which are part of feature sets, can help accurately retrieve the matching levels. Moreover, they proposed to represent the tissue image of each level by using spanning trees in the tumor image to connect the nuclei. Farjam *et al.* [21] used the texture features and k-means clustering to extract the structural features of the gland region and used the tree structure algorithm to classify the images into 1 to 5 levels. Nguyen *et al.* [22] described how prostate cancer can be classified into three major categories (benign, grade 3, and grade 4) by detecting the basic components of prostate tissue using various texture features and the color space of the tissue image. Gorelick *et al.* [23] used the AdaBoost classifier to obtain high-level tissue information about the locations and grades of tumors. Waliszewski *et al.* [24] proposed a method based on fractal analysis when classifying adjacent Gleason groups (such as 3+3 and 3+4), with an average sensitivity of 81% and an average specificity of 75%. Mosquera-Lopez *et al.* [25] described an

improved local binary mode of tissue description with new features and an automated prostate cancer biopsy image classification system that integrates a new multistage learning framework with multiple classes and binary classifiers. Bhattacharjee *et al.* [26] judged the benign and malignant of prostate cancer based on SVM in histological pathological images and predicted the Gleason grading. Although the above methods have achieved good results, all systems need to accurately locate small image areas to extract features. The difference in Gleason pattern in microarray images of prostate cancer tissue is not obvious, and it is difficult to directly classify features in a small area. This problem is very difficult; thus, the reproducibility is limited and the methods cannot be quickly applied in medical systems.

B. APPLICATIONS OF CONVOLUTIONAL NEURAL NETWORKS IN MEDICAL IMAGE SEGMENTATION

In recent years, medical image classification and segmentation with convolution neural networks [27] have been widely applied and practiced, as it achieved superior performance. Fully convolutional networks (FCNs) [28] are semantic segmentation networks successful in biomedical image analysis [29], [30]. Scholars have improved neural networks to adapt them to biomedical image processing, proposing U-Net [31] and multiscale U-Net [32]. Ravì *et al.* [33] critically analyzed the relative advantages, potential drawbacks, and prospects of health informatics research based on deep learning, including translation bioinformatics, medical imaging, and other key applications. Shi *et al.* [34] proposed a multimode superposition depth polynomial network (mm-SDPN) algorithm, which is composed of two-stage SDPN, to integrate and learn the feature representation of multimode neuroimaging data for the diagnosis of Alzheimer's disease. Jiang *et al.* [35] proposed an effective lung nodule detection scheme based on multigroup patches of lung images. The method achieved 80.06% sensitivity in detecting four levels of pulmonary nodules by training a four-channel CNN using the radiologist's knowledge. Huang *et al.* [36] proposed a blood cell classification framework using medical hyperspectral images, which transforms the convolution kernel into the frequency domain to learn features and combines them with the characteristics of Gabor wavelet to improve classification of cells. Yan *et al.* [37] proposed a three-stage deep learning model-thick vessel segmentation, thin vessel segmentation, and vessel fusion-as vascular segmentation tasks. The three-stage deep learning model is superior to the latest vascular segmentation methods. Certainly, there are many applications of deep learning in the Gleason grading of pathological images of prostate cancer. Ing *et al.* [38] collected 513 high-resolution images of primary prostate cancer and tested the performance of four CNNs in the semantic segmentation of high and low-grade tumors. Among them, the precision of U-Net [31] was 0.885. Arvaniti *et al.* [39] proposed a MobileNet [40] neural network for training the prostate cancer image annotated by pathologists. The consistency between the model and the two pathologists was 0.75 and 0.71, respectively. The experimental results were similar to those between pathologists. Ren *et al.* [41] used computer-aided analysis to classify Gleason pattern 3 and Gleason pattern 4. First, deep CNNs automatically divided the boundaries of each gland region. Next, color, shape, and texture features were extracted and forwarded to the random forest classifier. Kwak and Hewitt [42] proposed a deep learning method to detect prostate cancer. First, tissue was segmented to identify the lumen in a digitized image of the prostate tissue sample. Then, CNN was used to automatically extract higher-order image features of the lumen and predict cancer in 5 different scales. Gurcan *et al.* [43] proposed an automatic visual inspection of the whole slide images detection method based on deep learning frameworks for high grade Gleason score. Li *et al.* [44] presented a new region-based convolutional neural network (R-CNN) [45] for multitask prediction using both epithelial and hierarchical network headers. After five cross-validations, their model achieved an average intersection of 79.56% and pixel accuracy of 89.40% for Gleason grading. Lucas *et al.* [46] realized the automatic detection of Gleason pattern 3 and Gleason pattern $\geq$4 in digital prostate biopsy by using a CNN to automatically detect Gleason patterns and through the retraining of Inception-v3 CNN to determine the grade group. Karimi *et al.* [47] demonstrated classification and data enhancement that combined three independent CNNs that worked on three different image blocks, respectively. After training of CNN, a trained logistic regression model was separately used for prediction. The authors proposed new data enhancement methods to effectively train the model and perform accurate Gleason grading of prostate cancer in histopathological images with limited datasets. Nagpal *et al.* [48] presented a deep learning system (DLS) that combined the data collected in the laboratory for Gleason scoring whole-slide images of prostatectomies. The DLS achieved a significantly higher diagnostic accuracy of 0.70. Bulten *et al.* [49] developed a fully automatic deep learning system to grade prostate biopsies according to the Gleason grading criteria. The deep learning system outperformed 10 of 15 pathologists in a separate observation experiment. The system was developed to map individual glands, assign Gleason growth patterns, and determine biopsy levels [50]. Madabhushi *et al.* [51] proposed deep learning as an automated method of training networks using labeled images without other assumptions (which has proven to be very useful in many similar areas of digital pathology) and discussed two deep learning methods for Gleason grading. Although the application of the above-mentioned deep learning methods has achieved good results, most studies have only performed a few grades, such as benign and malignant prostate cancer, Gleason 3 and Gleason 4, but our method classification is relatively complete. Some studies used classification networks, and the results obtained are judgments on the entire image, and the subdivision of the Gleason pattern cannot be clearly obtained. There are still some studies based on gland segmentation to obtain Gleason grading, but the distribution

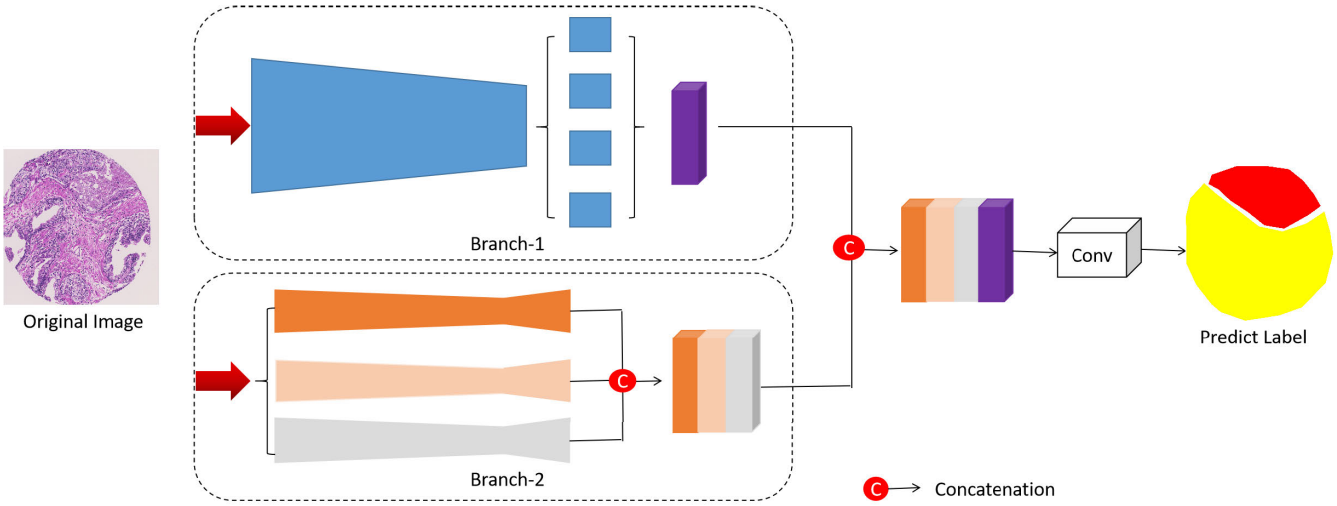

FIGURE 2. **Overview of the proposed model architecture. The original prostate cancer tissue microarray image is fed into two different branches. For Branch-1, the atrous spatial pyramid pooling architecture is used to capture image multiscale information based on atrous convolution. For Branch-2, the parallel branch convolutional neural network is used to capture image multiscale information based on standard convolution. Finally, the CRF based post-processing procedure is applied to the prediction.**

of glands in the image is relatively dense, and the required gold standard will consume a lot of labeling time. In contrast, our method can segment the region of Gleason pattern to gain more complete Gleason grading results, and also save a lot of annotating time. Moreover, the convolutional neural networks currently used for Gleason grading are mostly of a single structure. The network we proposed is a fusion of two structures, and uses the atrous convolution and standard convolution together to exert their respective advantages.

III. PROPOSED APPROACH

In this section, we first introduce the network architecture proposed in this study. Then, we describe our model: details of implementation and training, loss function, and postprocessing. Finally, we define the metrics used to evaluate our method and the proposed data augmentation.

A. NETWORK STRUCTURE

The main framework of our method proposed in this study (which has two branches) is illustrated in Figure 2.

For Branch-1, the ASPP architecture is used to capture multiscale image information using the atrous convolution. This method can obtain multiscale information, and the amount of calculations is relatively small. It also uses the method of atrous convolution. For Branch-2, the parallel-branch CNN is used to capture multiscale image information using the standard convolution. These three branches use convolution kernels of different sizes to capture multiscale information. The features of Branch-1 and the three maps obtained by Branch-2 are cascaded together as the input of the main network to obtain the final segmentation results.

The fully connected CRF-based postprocessing is applied to the prediction. This step is only for postprocessing; CRF is not used during the training.

B. IMPLEMENTATION AND TRAINING

Figure 3 shows pathological images and the corresponding annotation using the Gleason grading of prostate cancer. The texture structure of prostate cancer pathology image is extremely complicated. In particular, there is a tiny difference between Gleason pattern 3 and Gleason pattern 4. Even professional and experienced pathologists need to magnify the image several times to describe the Gleason score region. Because the image segmentation target has different scales, the ASPP and parallel branch structure that can obtain multiscale information of the image are combined, and atrous convolution and standard convolution are used, respectively. Therefore, in this paper, our method is divided into two branches, and both branches are implemented using the open-source deep-learning library TensorFlow [52].

Branch 1 sends a large number of prostate cancer tissue microarray (TMA) images to the network model for training. For multiscale object segmentation, it uses serial and parallel perforated convolution modules, using multiple atrous rates to obtain multiscale information. Branch-1 first uses convolution to reduce the number of channels (and reduce calculations) and then adds a pyramid model based on atrous convolution to capture multiscale information. The idea of network architecture comes from the ASPP module, which excavates convolution features at different scales and describes the image information in more detail. However, for the prostate cancer TMA images, which contain many small targets, atrous convolution is not as large as possible. The selection of rates in this study is shown in Figure 4: rates are set to 6, 12, 18, respectively.

We are limited by the lack of precisely annotated training data (because of the difficulty and cost of generating high-quality data). Data augmentation is applied to the prostate cancer TMA images, including training sets and

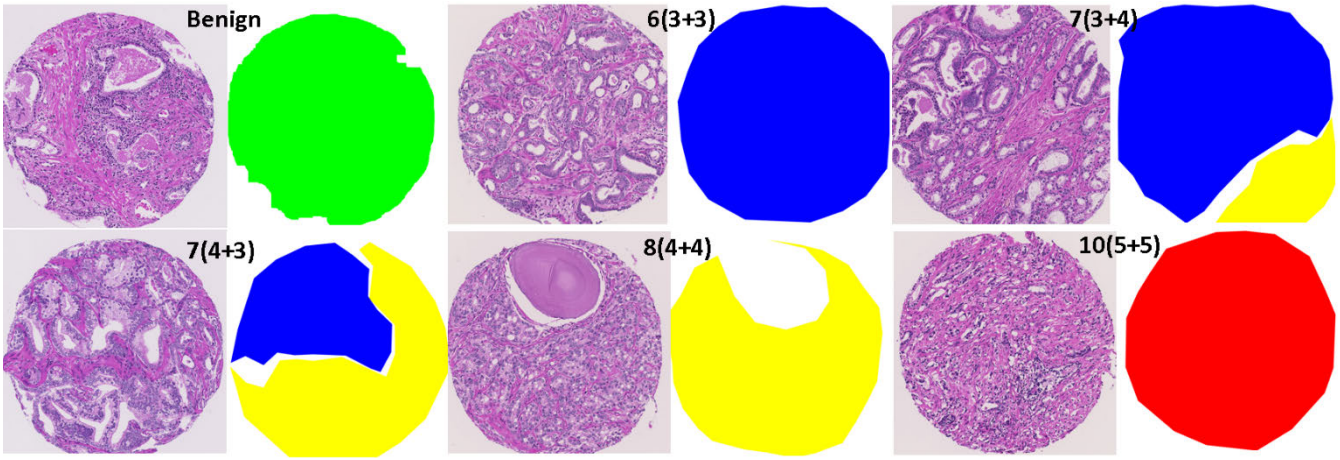

FIGURE 3. Pathological image and annotation of prostate cancer Gleason grading. Green indicates "benign," blue indicates "Gleason pattern 3," yellow indicates "Gleason pattern 4," and red indicates "Gleason pattern 5." The prognosis is better in the first row and worse in the second row.

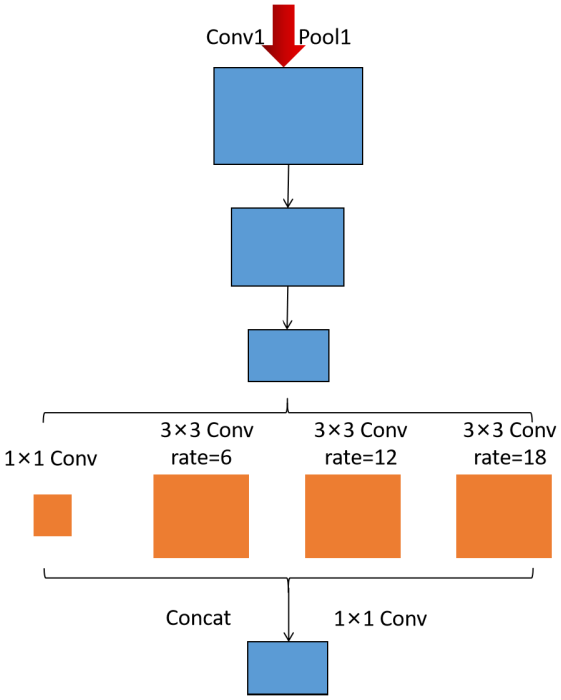

FIGURE 4. Branch-1 Infrastructure Network Structure. Rates are set to 6, 12, 18 respectively.

validation sets. Augmentation includes horizontal and vertical reversals and 90° random integer multiplier rotations. Even so, our dataset cannot support the complete training of the network. We overcome this limitation by using natural image data (called transfer learning). Transfer learning helps to reduce overfitting on finite medical datasets and allows us to leverage networks with more parameters. Therefore, we used the implementation of Deeplab-V3 [14], which was trained on the PASCAL VOC2012 dataset [53], as a pretrained model to initialize the network.

Branch 2 sends the prostate cancer TMA images to the parallel-branch CNN to training. The parallel-branch CNN architecture includes three parallel branches. All branches share the same network architecture, including an encoder and a decoder. Input training image blocks are fed into three branches, respectively. In our approach, the base architecture of each branch is FCN [28]. All convolutions are standard; however, these three branches use convolution kernels of different sizes to capture multiscale information. Other network parameters are fine-tuned according to the characteristics of our images. The detailed network architecture is shown in Figure 5.

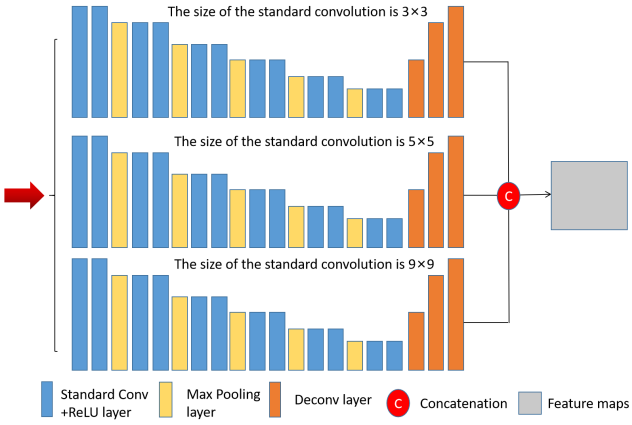

FIGURE 5. Branch-2 Infrastructure Network Structure.

Every input image was sent to two branches respectively acquiring two features with different scales. For branch 1, the selection of rates in this study is shown in Figure 4: rates are set to 6, 12, 18, respectively. For branch 2, the sizes of the standard convolution were set to 3×3, 5×5 and 9×9, respectively. Then we cascaded the output of the two branches as the input to predict the segmentation label map followed by stacking convolutional layers in the backbone. In the encoder part of each branch, after several convolutional layers, max-pooling with a window of 2×2 and a

stride of 2 was performed. The decoder section includes three upsampling operations. The first two upsampling operations are performed by a 4×4 deconvolution operation, and the last upsampling restores the size of the feature map to the size of the original image. We optimized it using the gradient descent algorithm with the batch size of 20. The learning rate was set to 0.0001. The number of iterations in this study was set to 50000. The most important constants are selected through hyperparameter adjustment, including the constant value in the loss function, learning rate, kernel size, and the number of iterations. In addition, the remaining parameters were selected based on experience, including batch size and partial loss function parameters. Table 1 summarizes the network parameters.

TABLE 1. Parameter settings for our network.

Parameter	Value
Kernel size	3×3
Rate	6,12,18
Pooling	2×2 window
Learning rate	0.0001
Batch size	20
Iterations	50000
w_0	10
σ	5

C. LOSS FUNCTIONS

During the training, the main network is a combination of different network architectures of Branch-1 and Branch-2. The pixel-wise cross-entropy loss was applied to predictions. The output of the network is pixel-wise softmax:

$$p_k(x) = \exp\left(a_k(x)\right) / \left(\sum_{k=1}^{K} \exp\left(a_k(x)\right)\right) \tag{1}$$

where x is the pixel position on the two-dimensional plane, $a_k(x)$ is the value of the k^{th} channel corresponding to pixel x in the final output layer of the network; $p_k(x)$ is the probability that pixel x belongs to class k. For the i^{th} branch with a pixel-wise cross-entropy loss F, the loss was computed as follows:

$$F = \sum_{x} w(x) \log\left(p_{I(x)}(x)\right) \tag{2}$$

where $p_{I(x)}(x)$ is the output probability of x on the channel where the real label is located; $w(x)$ is the weight associated with pixel x. It is defined as follows

$$w(x) = w_c(x) + w_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right) \tag{3}$$

where w_c is a weight graph representing the frequency of the balance class, $d_1(x)$ is the distance from pixel x to the area closest to it, $d_2(x)$ is the second closest distance from the pixel x to its region, based on experience, set $w_0 = 10$ and $\sigma \approx 5$.

D. POSTPROCESSING

The CRF-based postprocessing is applied to the prediction and is not used during the training. Li *et al.* [44] and Chen *et al.* [54] later used this method as a postprocessing step. The CRF model employs the energy function [54]:

$$E(x) = \sum_{i} \theta_i\left(x_i\right) + \sum_{ij} \theta_{ij}\left(x_i, x_j\right) \tag{4}$$

where x is defined in the entire image and x_i is the label assignment for i^{th} pixel. We refer to the first term as unary potential, which is defined as $\theta_i x_i = -\log P\left(x_i\right)$, where Px_i is the label assignment probability at pixel i as computed by the segmentation. We refer to the second term as the pairwise potential, which is $\theta_{ij}\left(x_i, x_j\right) = \mu\left(x_i, x_j\right) W$, where $\mu\left(x_i, x_j\right) = 1$ if $x_i \neq x_j$, and zero otherwise; W are the bilateral position and color terms in the kernels.

$$w_1 \exp\left(-\frac{\left\|P_i - P_j\right\|^2}{2\sigma_\alpha^2} - \frac{\left\|I_i - I_j\right\|^2}{2\sigma_\beta^2}\right) + w_2 \exp\left(-\frac{\left\|P_i - P_j\right\|^2}{2\sigma_\gamma^2}\right) \tag{5}$$

Two Gaussian kernels are used by W in different feature spaces. Here, P is the pixel position and I is the pixel RGB color. The first kernel uses P and I to keep adjacent pixels with similar colors in the same category, and the second one only uses P to remove small isolated areas. Hyperparameters $\sigma_\alpha, \sigma_\beta$, and σ_γ control the scale of Gaussian kernels and were obtained experimentally.

E. EVALUATION METRICS

Some evaluation metrics were used to access the performance of Gleason grading and region segmentation. To make our method comparable with other methods, we use the standard metrics [44]: Jaccard coefficient (J), mean intersection over union (mIOU), and overall pixel accuracy (OPA) to evaluate the performance of the segmentation results. Jaccard coefficient is defined as the ratio of the size of the intersection of A and B to the size of the union of A and B. Here, Jaccard coefficient is defined as

$$\mathrm{J} = \frac{U_{ii}}{T_i + P_i - U_{ii}} \tag{6}$$

where U_{ii} is the number of pixels labeled as i and also predicted as i, T_i is the number of pixels labeled, P_i is the number of pixels predicted. The mIOU is defined as

$$\mathrm{mIOU} = \sum_{i=1}^{N} J_i \tag{7}$$

where N is the number of classes. OPA is defined as

$$\mathrm{OPA} = \frac{\sum_i U_{ii}}{\sum_i \sum_i C_{ii}} \tag{8}$$

where C_{ii} is the sum of pixels labeled as i and predicted as i.

Cohen's kappa [55] statistics are widely used to assess consistency among evaluators. Cohen's kappa considers the

possibility of accidental agreement: its value is 0 for accidental agreement and 1 for perfect agreement. For ordered classes, weighted Cohen's kappa is more appropriate, as it penalizes more serious disagreements between annotators. Here, we use a quadratic weighted kappa statistic defined as follows:

$$\text{kappa} = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}, w_{i,j} = \frac{(i-j)^2}{(M-1)^2} \quad (9)$$

where M is the total number of considered classes; indices i, j refer to the ordered classes $1 \leq i, j \leq M$, $O_{i,j}$ is the number of images that were assigned to class i by the first expert and class j by the second, and $E_{i,j}$ denotes the expected number of images labeled as class i by the first expert and class j by the second expert, assuming no correlation between classes.

F. DATA AUGMENTATION

To train an acceptable deep learning model, enough training samples are necessary. For data augmentation in our experiment, we chose to flip each prostate cancer TMA image from left to right, from top to bottom, and rotate the image by 90 degrees, 180 degrees and 270 degrees respectively; this flipping preserves the visual structure. While the data augmentation method expands the amount of data, it can also select approximately equal numbers of different classes of data to train to solve the class imbalance problem.

IV. EXPERIMENTS

In this section, we first introduce the dataset and the experimental setup used in this study. Then, we analyze our results quantitatively and qualitatively. We calculate the number of pixels in the corresponding region in the output of our model as the area of the Gleason pattern region for comparison with other works.

A. DATASETS

Table 2 demonstrates the dataset used in this study, which includes 7 groups of prostate cancer TMAs, each containing 200-300 images. These five TMAs have been used and published in many other studies, such as TMAs 76, 80 [39], [56], [57], TMAs 111 [39], [58], TMAs 204 [39], [59], and TMAs 199 [39]. The total number of prostate cancer tissue microarrays is 1211 (325 from a local hospital and 886 from [39]). The classes of Gleason grading of TMA_A, TMA 111, TMA 199, and TMA 204 were unevenly distributed in each group; thus, it was more reasonable to combine them for training. Instead, TMAs 76 is used as a verification set because it has the most balanced distribution. TMA_B and TMAs 80 contain the largest number of cases and were selected as the test set. According to the latest revision of the Gleason grading by ISUP in 2014, Table 2 divides the TMAs data except for "Benign" into five categories (Gleason score=6, 3+4, 4+3, 8, and 9-10, respectively). Figure 3 shows the samples of prostate cancer TMA images and Gleason annotations provided by the pathologists (green: benign; blue: Gleason pattern 3 regions; yellow: Gleason pattern 4 regions; red: Gleason pattern 5 regions). We use TMARKER41 software to describe the cancerous area and mark it with the corresponding Gleason pattern. The prostate TMA images [39] were assigned to each region by 3, 4, or 5 Gleason patterns for annotation by the first pathologist (K.S.F). TMA images from the test set [39] were annotated independently by the second pathologist (J.H.R.) to quantify the variability between pathologists. Similarly, our data is also labeled by two senior pathologists. In the process, the pathologist carefully depicts the cancerous region. In addition, pathologists labeled TMA images containing only benign prostate tissue as "Benign." From Figure 3, the shape of the individual region of the Gleason pattern in the TMAs is grotesque, and the inter-class differences are also minimal. The tissue microarray images of [39] used in this study has been made publicly available on Github (for specific link reference [39]). Unfortunately, for reasons of ethics and patient privacy we are not able to provide the data of local hospital into a public database.

TABLE 2. Dataset gleason grading summary.

	Benign	6	7(3+4)	7(4+3)	8	9,10	Total
TMA_A	3	76	26	19	16	32	172
TMA_B	2	48	28	23	25	27	153
TMA 76	42	35	11	14	15	16	133
TMA 80	12	88	21	17	91	16	245
TMA 111	0	95	31	8	69	24	227
TMA 199	61	69	13	4	26	3	176
TMA 204	0	1	0	2	25	77	105

B. EXPERIMENTAL SETUP

The experiment was performed on a hardware setup with Intel Xeon CPU, one 11 GB NVIDIA Tesla V100 PCIe GPU, and 16 GB RAM. The software was Python 3.7 and TensorFlow. The total number of prostate cancer tissue microarrays is 1211 (325 from a local hospital and 886 from [39]; the detailed data allocation is described in Section IV-A). The experiment was designed considering the data of each patient independently.

C. QUANTITATIVE EVALUATION

Table 3 compares the performance of automated and manual segmentation. From Table 3, our model achieves better performance (mIOU: 77.29%; OPA: 89.51%) than other methods for Gleason patterns (Gleason 3, 4, or 5, and benign). We compared our model with several baseline models using the same data and the same post-processing operations. Semantic segmentation uses FCN [28], and most neural networks are based on it, but FCN has a big problem with segmenting images of prostate cancer TMAs. U-Net [31] and multiscale U-Net [32] are based on FCN and are suitable for medical image segmentation. However, these networks are not ideal for the prostate cancer TMA images. The results were not satisfactory for Deeplab-V3 [14], using atrous

TABLE 3. Model performance on segmenting prostate histological images as "benign" (B), "gleason pattern 3" (3), "gleason pattern 4" (4), and "GLEASON PATTERN 5" (5).

	J_B	J_3	J_4	J_5	mIOU	OPA
FCN [28]	70.47%	62.48%	65.84%	61.46%	65.06%	70.47%
U-Net [31]	77.36%	70.37%	67.38%	71.84%	71.73%	80.46%
Multiscale U-Net [32]	74.37%	66.87%	64.93%	73.39%	69.86%	79.34%
Depplab-V3 [14]	75.65%	71.14%	67.35%	80.56%	73.67%	84.95%
MPB-CNN [15]	78.34%	70.52%	69.29%	79.65%	74.45%	85.73%
Path R-cnn [44]	78.92%	71.09%	68.45%	80.18%	76.87%	86.92%
Our method	**80.57%**	**74.51%**	**71.73%**	**82.35%**	**77.29%**	**89.51%**

convolution and ASPP, and MPB-CNN [15], using atrous convolution and multiscale parallel CNN. Compared with the Path R-cnn network [44] that also uses the segmentation network to solve the Gleason grading problem, our method still works better.

For test data, each prostate cancer TMA is annotated with the detected Gleason patterns (Gleason 3, 4, or 5) by the model and the pathologists. A final Gleason score (Gleason 6, 7, 8, 9, or 10) is assigned as the sum of the two Gleason patterns. If no cancer is detected, the prostate cancer TMA is classified as benign. The confusion matrix for Gleason score allocation is demonstrated in Figure 6, where (a) (Kappa=0.79) and (b) (Kappa=0.74) are confusion matrices for Gleason score allocation used to compare data from [39]; (c) (Kappa=0.77) and (d) (Kappa=0.79) are used to compare data from the local hospital; (e) (Kappa=0.75) and (f) (Kappa=0.71) are confusion matrices for Gleason score allocation in [39]. In public data, our method achieves better results compared to the original article [39].

D. QUALITATIVE EVALUATION

The Gleason grading system is one of the most reliable methods to quantify the invasiveness of prostate cancer [3]. However, it is very difficult for pathologists to observe Gleason grading through multiple prostate cancer TMAs. Most of the previous studies [21], [38], [41], [44], [46] used an accurate segmentation of glands to achieve the Gleason grading of images with prostate cancer pathologies. Although the above methods have achieved good results, it is a time-consuming and difficult task to annotate glands in prostate cancer TMAs. In our method, the pathologist only needs to mark the region of Gleason patterns (Gleason 3, 4, 5) in the prostate cancer TMAs. A final Gleason score is assigned as the sum of the two Gleason patterns. Figure 7 shows the segmentation results of Gleason patterns in prostate cancer TMAs with our method. In Figure 7, the first column is the original histological image of the prostate cancer tissue microarray. The second column and third column show the annotated image of the prostate cancer tissue microarray by the first pathologist and the second pathologist (as the ground truth), respectively. The fourth column shows the results of our method. In "Benign," the benign region in the image is not continuous, and there are many holes in the middle. The segmentation performance of our method is still considerable.

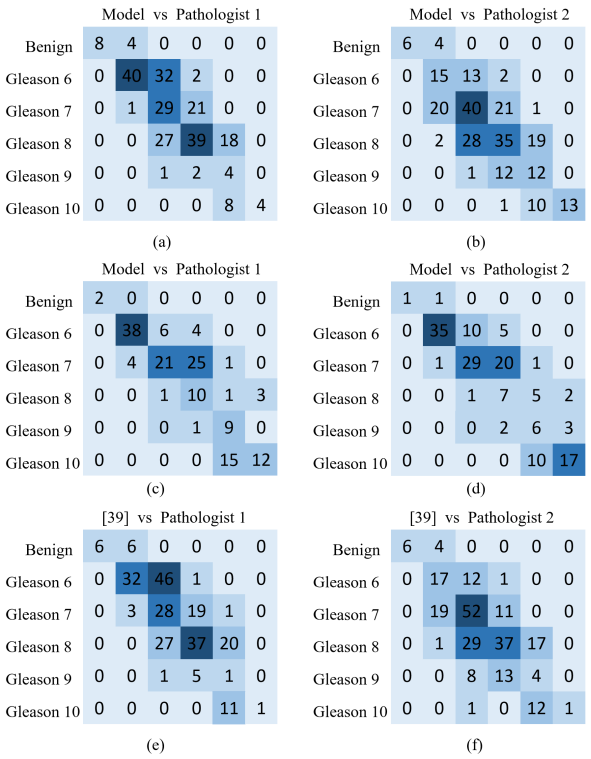

FIGURE 6. Model evaluation on test data. Each prostate cancer tissue microarray is annotated with detected Gleason patterns (Gleason 3, 4 or 5) by the model and pathologists. A final Gleason score is assigned as the sum of the two Gleason patterns. If no cancer is detected, the prostate cancer tissue microarray is classifed as benign. (a) and (b): Confusion matrix for Gleason score allocation used to compare data from [39]. (c) and (d): Confusion matrix for Gleason score allocation used to compare data from the local hospital. (e) and (f): Confusion matrix for Gleason score allocation in [39].

For the segmentation of Gleason score 6, Gleason score 8, and Gleason score 10, the segmentation results are in high agreement with the annotations of two pathologists. However, it is very difficult to segment multiple Gleason patterns on one prostate cancer TMA, such as Gleason socre 7 and Gleason score 9. Even the descriptions of the two pathologists are different.

V. DISCUSSION

A. CONTRIBUTIONS

The first novel idea proposed in our study is an accurate Gleason grading by segmentation of the Gleason pattern

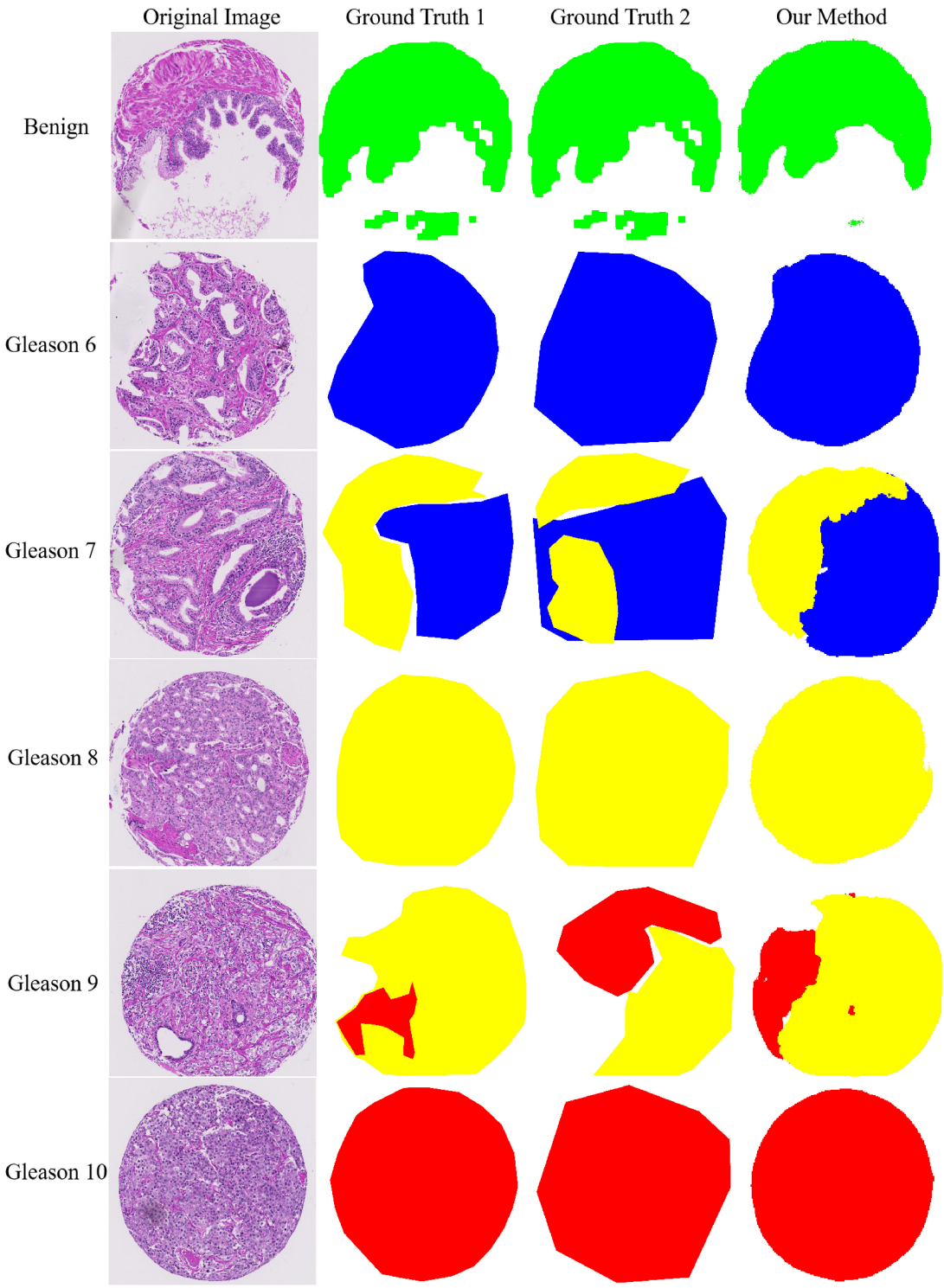

FIGURE 7. Segmentation results of low-level Gleason pattern region of our method. First column: Original histological image of prostate cancer tissue microarray. Second column and third column: The annotated image of the prostate cancer tissue microarray by the first pathologist and the second pathologist (as the ground truth), respectively. The green region represents "benign," the blue region represents "Gleason pattern 3," the yellow region represents "Gleason pattern 4," and the red region represents "Gleason pattern 5." Fourth column: The results of our method.

region, including glands and other tissues. Most of the previous studies [21], [38], [41], [44], [46] used an accurate segmentation of glands to produce the Gleason grading of images with prostate cancer pathologies. Although the above methods have achieved good results, it is a time-consuming and difficult task to annotate glands in prostate cancer TMAs. In our method, the pathologist only needs to mark the region of Gleason pattern in the prostate cancer TMA.

Furthermore, tissue relationship information outside the gland region may be lost during feature extraction.

The second novel idea proposed in our work is to combine two different network models with two different convolution methods based on deep learning for Gleason grading of the prostate cancer TMA. One is ASPP, the other is multiscale standard convolution. ASPP is a powerful network architecture for intensive prediction tasks, which highlights convolution with an upsampling filter or atrous convolution. Atrous convolution effectively enlarges the filter's receptive field to mix richer context information without increasing the number of parameters and computational complexity. Adding a simple and effective decoding module and fine-grained segmentation improve the segmentation of the Gleason pattern regions for prostate cancer TMAs with complex textures and structures. For Branch-1, because of the lack of image samples, we use transfer learning with Deeplab-V3. Transfer learning helps to reduce overfitting on smaller medical datasets and allows us to leverage networks with more parameters. The multiscale parallel-branch CNN uses standard convolution. The feature mapping of three different branches is cascaded by several convolutions and low-level feature calculations to achieve the goal of accurately segmenting images of different scales. For Branch-2, we augmented the data to provide more detailed information for the network when extracting features and to increase the training dataset. Deep learning has been used in recent work [44], [46], [47] on automatic Gleason grading. Li *et al.* [44] presented a new region-based convolutional neural network (R-CNN) [45] for multitask prediction using both epithelial and hierarchical network headers. They use the same network architecture to solve two branch tasks that are different from what we do. Our study is also different from work by Lucas *et al.* [46], who used CNN to automatically detect Gleason patterns and retrained Inception-v3 CNN to determine the grade group. Karimi *et al.* [47] used classification and data augmentation that combined three independent CNNs that worked in three different image blocks, respectively. However, our proposed method is the first study that combines networks with different architectures.

Our last contribution is the CRF-based postprocessing applied to the prediction. Li *et al.* [44] and Chen *et al.* [54] used CRF as a postprocessing step, which helps in removing unnatural boundaries created by the stitching.

B. LIMITATIONS

Our method has limitations. First, considering the distribution of prostate cancer TMAs, we fixed the training set, verification set, and testing set. However, we think it is fair to compare the models in our study because we use the same train-test data split as in [39] (apart from the data we collected). In the future, we hope to do cross-validation by increasing the amount of data. Second, we only used rigid deformation for data augmentation. Principal-component-analysis (PCA)-based intensity jitter method [60], generative adversarial networks (GANs) method [61], and data augmentation in the feature space [47] can be considered in the future. Finally, for the multiple Gleason patterns on one prostate cancer TMAs (such as Gleason score 7 and Gleason score 9), the result of segmentation is not precise enough. We will improve the network to make the segmentation results more accurate.

VI. CONCLUSION

In this paper, we propose a novel method for automatic Gleason grading and Gleason pattern region segmentation of images with prostate cancer pathologies based on a convolutional neural network (CNN). An architecture that combines the atrous spatial pyramid pooling (ASPP) from Deeplab-V3 and the multiscale standard convolution inspired by a multiscale parallel branch convolutional neural network (MPB-CNN) is proposed for the segmentation of the Gleason pattern region (not gland segmentation) to get accurate Gleason grading. In addition, we improved the network by combining the standard convolution with atrous convolution, and postprocessing based on a conditional random field (CRF) is applied to the prediction. Our study shows that the method of combining different deep neural network architecture is suitable for more objective and reproducible prostate cancer Gleason grading. Our approach yields higher performance on Gleason grading than previous work. Furthermore, the Gleason pattern assignment of the model achieved stratification of pathologists and divided patients into groups with different prognosis.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA, Cancer J. Clin.*, vol. 66, no. 1, pp. 7–30, Jan. 2016.

[2] M. Seitz, A. Shukla-Dave, A. Bjartell, K. Touijer, A. Sciarra, P. J. Bastian, C. Stief, H. Hricak, and A. Graser, "Functional magnetic resonance imaging in prostate cancer," *Eur. Urol.*, vol. 55, no. 4, pp. 801–814, 2009.

[3] D. F. Gleason, "Histologic grading of prostate cancer: A perspective," *Hum. Pathol.*, vol. 23, no. 3, pp. 273–279, Mar. 1992.

[4] D. F. Gleason and G. T. Mellinger, "Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging," *J. Urol.*, vol. 111, no. 1, pp. 58–64, 1974.

[5] S. F. Faraj, S. M. Bezerra, K. Yousefi, H. Fedor, S. Glavaris, M. Han, A. W. Partin, E. Humphreys, J. Tosoian, M. H. Johnson, E. Davicioni, B. J. Trock, E. M. Schaeffer, A. E. Ross, and G. J. Netto, "Clinical validation of the 2005 ISUP Gleason grading system in a cohort of intermediate and high risk men undergoing radical prostatectomy," *PLoS ONE*, vol. 11, no. 1, Jan. 2016, Art. no. e0146189.

[6] J. Gordetsky and J. Epstein, "Grading of prostatic adenocarcinoma: Current state and prognostic implications," *Diagnostic Pathol.*, vol. 11, no. 1, p. 25, Dec. 2016.

[7] J. I. Epstein, "An update of the Gleason grading system," *J. Urol.*, vol. 183, no. 2, pp. 433–440, Feb. 2010.

[8] O. N. Kryvenko and J. I. Epstein, "Prostate cancer grading: A decade after the 2005 modified Gleason grading system," *Arch. Pathol. Lab. Med.*, vol. 140, no. 10, pp. 1140–1152, Oct. 2016.

[9] P. A. Humphrey, "Gleason grading and prognostic factors in carcinoma of the prostate," *Mod. Pathol.*, vol. 17, no. 3, p. 292, 2004.

[10] H. J. Lavery and M. J. Droller, "Do Gleason patterns 3 and 4 prostate cancer represent separate disease states?" *J. Urol.*, vol. 188, no. 5, pp. 1667–1675, Nov. 2012.

[11] C. C. Huang, M. X. Kong, M. Zhou, A. B. Rosenkrantz, S. S. Taneja, J. Melamed, and F.-M. Deng, "Gleason score 3+4=7 prostate cancer with minimal quantity of Gleason pattern 4 on needle biopsy is associated with low-risk tumor in radical prostatectomy specimen," *Amer. J. Surgical Pathol.*, vol. 38, no. 8, pp. 1096–1101, Aug. 2014.

[12] M. Huang, H. Han, H. Wang, L. Li, Y. Zhang, and U. A. Bhatti, "A clinical decision support framework for heterogeneous data sources," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 6, pp. 1824–1833, Nov. 2018.

[13] Y. Li, S. Niu, Z. Ji, W. Fan, S. Yuan, and Q. Chen, "Automated choroidal neovascularization detection for time series SD-OCT images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 381–388.

[14] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

[15] Y. Zhang, Z. Ji, Y. Wang, S. Niu, W. Fan, S. Yuan, and Q. Chen, "MPB-CNN: A multi-scale parallel branch CNN for choroidal neovascularization segmentation in SD-OCT images," *OSA Continuum*, vol. 2, no. 3, pp. 1011–1027, 2019.

[16] P.-W. Huang and C.-H. Lee, "Automatic classification for pathological prostate images based on fractal analysis," *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 1037–1050, Jul. 2009.

[17] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *Proc. 5th IEEE Int. Symp. Biomed. Imag., Nano Macro*, May 2008, pp. 284–287.

[18] G. Nir, S. Hor, D. Karimi, L. Fazli, B. F. Skinnider, P. Tavassoli, D. Turbin, C. F. Villamil, G. Wang, R. S. Wilson, K. A. Iczkowski, M. S. Lucia, P. C. Black, P. Abolmaesumi, S. L. Goldenberg, and S. E. Salcudean, "Automatic grading of prostate cancer in digitized histopathology images: Learning from multiple experts," *Med. Image Anal.*, vol. 50, pp. 167–180, Dec. 2018.

[19] Y. Smith, G. Zajicek, M. Werman, G. Pizov, and Y. Sherman, "Similarity measurement method for the classification of architecturally differentiated images," *Comput. Biomed. Res.*, vol. 32, no. 1, pp. 1–12, Feb. 1999.

[20] A. W. Wetzel, R. Crowley, S. Kim, R. Dawson, L. Zheng, Y. M. Joo, Y. Yagi, J. Gilbertson, C. Gadd, D. W. Deerfield, and M. J. Becich "Evaluation of prostate tumor grades by content-based image retrieval," *Proc. SPIE*, vol. 3584, pp. 244–252, Jan. 1999.

[21] R. Farjam, H. Soltanian-Zadeh, R. A. Zoroofi, and K. Jafari-Khouzani, "Tree-structured grading of pathological images of prostate," *Proc. SPIE*, vol. 5747, pp. 840–851, Apr. 2005.

[22] K. Nguyen, B. Sabata, and A. K. Jain, "Prostate cancer grading: Gland segmentation and structural features," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 951–961, May 2012.

[23] L. Gorelick, O. Veksler, M. Gaed, J. A. Gomez, M. Moussa, G. Bauman, A. Fenster, and A. D. Ward, "Prostate histopathology: Learning tissue component histograms for cancer detection and classification," *IEEE Trans. Med. Imag.*, vol. 32, no. 10, pp. 1804–1818, Oct. 2013.

[24] P. Waliszewski, F. Wagenlehner, S. Gattenlöhner, and W. Weidner, "On the relationship between tumor structure and complexity of the spatial distribution of cancer cell nuclei: A fractal geometrical model of prostate carcinoma," *Prostate*, vol. 75, no. 4, pp. 399–414, Mar. 2015.

[25] C. Mosquera-Lopez, S. Agaian, and A. Velez-Hoyos, "The development of a multi-stage learning scheme using new tissue descriptors for automatic grading of prostatic carcinoma," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 3586–3590.

[26] S. Bhattacharjee, H.-G. Park, C.-H. Kim, D. Prakash, N. Madusanka, J.-H. So, N.-H. Cho, and H.-K. Choi, "Quantitative analysis of benign and malignant tumors in histopathology: Predicting prostate cancer grading using SVM," *Appl. Sci.*, vol. 9, no. 15, p. 2969, Jul. 2019.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

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

[29] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 294–297.

[30] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[32] J. Li, K. V. Sarma, K. C. Ho, A. Gertych, B. S. Knudsen, and C. W. Arnold, "A multi-scale U-Net for semantic segmentation of histological images from radical prostatectomies," in *Proc. AMIA Annu. Symp.* Bethesda, MD, USA: American Medical Informatics Association, 2017, p. 1140.

[33] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

[34] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 173–183, Jan. 2018.

[35] H. Jiang, H. Ma, W. Qian, M. Gao, and Y. Li, "An automatic detection system of lung nodule based on multigroup patch-based deep learning network," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1227–1237, Jul. 2018.

[36] Q. Huang, W. Li, B. Zhang, Q. Li, R. Tao, and N. H. Lovell, "Blood cell classification based on hyperspectral imaging with modulated Gabor and CNN," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 160–170, Jan. 2020.

[37] Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1427–1436, Jul. 2019.

[38] N. Ing, Z. Ma, J. Li, H. Salemi, C. Arnold, B. S. Knudsen, and A. Gertych, "Semantic segmentation for prostate cancer grading by convolutional neural networks," *Proc. SPIE*, vol. 10581, Mar. 2018, Art. no. 105811B.

[39] E. Arvaniti, K. S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P. J. Wild, J. H. Rüschoff, and M. Claassen, "Automated Gleason grading of prostate cancer tissue microarrays via deep learning," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, Art. no. 12054.

[40] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[41] J. Ren, E. Sadimin, D. J. Foran, and X. Qi, "Computer aided analysis of prostate histopathology images to support a refined Gleason grading system," *Proc. SPIE*, vol. 10133, Feb. 2017, Art. no. 101331V.

[42] J. T. Kwak and S. M. Hewitt, "Lumen-based detection of prostate cancer via convolutional neural networks," *Proc. SPIE*, vol. 10140, Mar. 2017, Art. no. 1014008.

[43] O. J. del Toro, M. Atzori, S. Otálora, M. Andersson, K. Eurén, M. Hedlund, P. Rönnquist, and H. Müller, "Convolutional neural networks for an automatic classification of prostate tissue slides with high-grade Gleason score," *Proc. SPIE*, vol. 10140, Mar. 2017, Art. no. 101400O.

[44] W. Li, J. Li, K. V. Sarma, K. C. Ho, S. Shen, B. S. Knudsen, A. Gertych, and C. W. Arnold, "Path R-CNN for prostate cancer diagnosis and Gleason grading of histological images," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 945–954, Apr. 2019.

[45] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[46] M. Lucas, I. Jansen, C. D. Savci-Heijink, S. L. Meijer, O. J. de Boer, T. G. van Leeuwen, D. M. de Bruin, and H. A. Marquering, "Deep learning for automatic Gleason pattern classification for grade group determination of prostate biopsies," *Virchows Archiv*, vol. 475, pp. 77–83, May 2019.

[47] 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 Informat.*, vol. 24, no. 5, pp. 1413–1426, May 2020.

[48] K. Nagpal, D. Foote, Y. Liu, P.-H.-C. Chen, E. Wulczyn, F. Tan, N. Olson, J. L. Smith, A. Mohtashamian, J. H. Wren, G. S. Corrado, R. MacDonald, L. H. Peng, M. B. Amin, A. J. Evans, A. R. Sangoi, C. H. Mermel, J. D. Hipp, and M. C. Stumpe, "Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer," *npj Digit. Med.*, vol. 2, no. 1, pp. 1–10, Dec. 2019.

[49] W. Bulten, H. Pinckaers, H. van Boven, R. Vink, T. de Bel, B. van Ginneken, J. van der Laak, C. Hulsbergen-van de Kaa, and G. Litjens, "Automated Gleason grading of prostate biopsies using deep learning," 2019, *arXiv:1907.07980*. [Online]. Available: http://arxiv.org/abs/1907.07980

[50] W. Bulten, H. Pinckaers, H. van Boven, R. Vink, T. de Bel, B. van Ginneken, J. van der Laak, C. Hulsbergen-van de Kaa, and G. Litjens, "Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study," *Lancet Oncol.*, vol. 21, no. 2, pp. 233–241, Feb. 2020.

[51] A. Madabhushi, M. D. Feldman, and P. Leo, "Deep-learning approaches for Gleason grading of prostate biopsies," *Lancet Oncol.*, vol. 21, no. 2, pp. 187–189, Feb. 2020.
[52] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2015, *arXiv:1603.04467*. [Online]. Available: https://arxiv.org/abs/1603.04467
[53] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, Jan. 2015.
[54] 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. 2018.
[55] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.
[56] Q. Zhong, T. Guo, M. Rechsteiner, J. H. Rüschoff, N. Rupp, C. Fankhauser, K. Saba, A. Mortezavi, C. Poyet, T. Hermanns, Y. Zhu, H. Moch, R. Aebersold, and P. J. Wild, "A curated collection of tissue microarray images and clinical outcome data of prostate cancer patients," *Sci. Data*, vol. 4, no. 1, Dec. 2017, Art. no. 170014.
[57] Q. Zhong *et al.*, "Image-based computational quantification and visualization of genetic alterations and tumour heterogeneity," *Sci. Rep.*, vol. 6, no. 1, Apr. 2016, Art. no. 24146.
[58] C. D. Fankhauser, P. J. Schüffler, S. Gillessen, A. Omlin, N. J. Rupp, J. H. Rueschoff, T. Hermanns, C. Poyet, T. Sulser, H. Moch, and P. J. Wild, "Comprehensive immunohistochemical analysis of PD-L1 shows scarce expression in castration-resistant prostate cancer," *Oncotarget*, vol. 9, no. 12, p. 10284, 2018.
[59] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 109–117.
[60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[61] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

YUCHUN LI received the B.S. degree from the University of Jinan, Jinan, China, in 2016, and the M.S. degree from the Nanjing University of Science and Technology, Nanjing, China, in 2019. She is currently pursuing the Ph.D. degree with Hainan University, China. Her current research interests are computer-aided diagnostic medical image processing and artificial intelligence.

MENGXING HUANG (Member, IEEE) received the Ph.D. degree from Northwestern Polytechnical University, in 2007. He then joined staff with the Research Institute of Information Technology, Tsinghua University as a Postdoctoral Researcher. In 2009, he joined Hainan University. He is currently a Professor and a Ph.D. Supervisor of computer science and technology, and the Dean of the College of Information Science and Technology. He is also the Executive Vice-President of the Hainan Province Institute of Smart City, and the Leader of the Service Science and Technology Team, Hainan University. He has published more than 60 academic articles as the first or corresponding author. He has reported 12 patents of invention, owns three software copyright, and published two monographs and two translations. He has been awarded one Second Class and one Third Class Prizes of The Hainan Provincial Scientific and Technological Progress. His current research interests include signal processing for sensor systems, big data, and intelligent information processing.

YU ZHANG (Member, IEEE) received the Ph.D. degree from Northwestern Polytechnical University, in 2016. He is currently a Lecture with the College of Information Science and Technology, Hainan University. His research interests include embedded systems, cyber-physical systems, and reliable software.

JING CHEN received the Ph.D. degree from Central South University. She is currently an Excellent Expert of Hainan Province, the Director of the Radiology Department of Haikou People's Hospital, and also a Supervisor of master's degree student and has been working in medical imaging technology for a long term.

HAIXIA XU is currently with the Department of Pathology, Haikou People's Hospital, the Deputy Chief Physician, and the Master of Medicine. She has been involved in clinical pathological diagnosis for 20 years and has accumulated rich diagnostic experience. She is a member of the Urology and Male Fertility Group, Chinese Medical Association Pathology Branch, the Deputy Chairman of the Chinese Immunopathology Professional Committee, a member of the Expert Committee of the Chinese Medical Education Association, and a Standing Member and the Secretary of the Pathological Committee of the Hainan Medical Association. She is also good at diagnosis and differentiation of male urinary reproductive systems, respiratory system tumors, and non-tumorous lesions.

GANG WANG received the bachelor's and master's degrees from the Xiangya Medical College, Central South University. He completed his residency and advanced training at Urology at third Xiangya Hospital as well. In 2009, he joined the staff at Central South University He has an Active Clinical Practice with Urologic Oncology, Prostatic disease, and Andrology. He is currently attending of the Urology Department.

WENLONG FENG received the B.S. degree from Jilin University, in 1990, and the M.S. degree from the South China University of Technology, in 2004. He is currently a Professor with the College of Information Science and Technology, Hainan University. His research interests include computer networks, cloud computing, and data mining.