The American Journal of Pathology, Vol. 192, No. 3, March 2022

ELSEVIER

The American Journal of PATHOLOGY

ajp.amjpathol.org

MACHINE LEARNING, COMPUTATIONAL PATHOLOGY, AND BIOPHYSICAL IMAGING

Weakly Supervised Framework for Cancer Region Detection of Hepatocellular Carcinoma in Whole-Slide Pathologic Images Based on Multiscale Attention Convolutional Neural Network

Songhui Diao,*† Yinli Tian,*‡ Wanming Hu,§ Jiaxin Hou,*† Ricardo Lambo,* Zhicheng Zhang,¶ Yaoqin Xie,*† Xiu Nie,‖ Fa Zhang,** Daniel Racoceanu,†† and Wenjian Qin*†

From the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China; the Shenzhen College of Advanced Technology,† University of Chinese Academy of Science, Shenzhen, China; the School of Microelectronics and Communication Engineering,‡ Chongqing University, Chongqing, China; the Department of Pathology,§ Sun Yat-sen University Cancer Center, Guangzhou, China; the Department of Radiation Oncology,¶ Stanford University, Stanford, California; the Department of Pathology,‖ Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; the Institute of Computing Technology,** Chinese Academy of Sciences, Beijing, China; and the Sorbonne Université,†† Paris Brain Institute—Institut du Cerveau—ICM, Institut National de Santé et en Recherche Médicale, Centre National de Recherche Scientifique, Assistance Publique Hôpitaux de Paris, Hôpital de la Pitié-Salpêtrière, Paris, France*

Accepted for publication November 17, 2021.

Address correspondence to Wenjian Qin, Ph.D., Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Ave., Shenzhen, 518055, P.R China. E-mail: wj.qin@siat.ac.cn.

Visual inspection of hepatocellular carcinoma cancer regions by experienced pathologists in whole-slide images (WSIs) is a challenging, labor-intensive, and time-consuming task because of the large scale and high resolution of WSIs. Therefore, a weakly supervised framework based on a multiscale attention convolutional neural network (MSAN-CNN) was introduced into this process. Herein, patch-based images with image-level normal/tumor annotation (rather than images with pixel-level annotation) were fed into a classification neural network. To further improve the performances of cancer region detection, multiscale attention was introduced into the classification neural network. A total of 100 cases were obtained from The Cancer Genome Atlas and divided into 70 training and 30 testing data sets that were fed into the MSAN-CNN framework. The experimental results showed that this framework significantly outperforms the single-scale detection method according to the area under the curve and accuracy, sensitivity, and specificity metrics. When compared with the diagnoses made by three pathologists, MSAN-CNN performed better than a junior- and an intermediate-level pathologist, and slightly worse than a senior pathologist. Furthermore, MSAN-CNN provided a very fast detection time compared with the pathologists. Therefore, a weakly supervised framework based on MSAN-CNN has great potential to assist pathologists in the fast and accurate detection of cancer regions of hepatocellular carcinoma on WSIs. *(Am J Pathol 2022, 192: 553—563; https://doi.org/10.1016/j.ajpath.2021.11.009)*

Hepatocellular carcinoma (HCC), a main subtype of primary malignant liver cancer, is usually diagnosed in terms of months of survival and leads to a high mortality rate.[1] A range of imaging techniques can be used to diagnose HCC, including magnetic resonance imaging, computed tomography, ultrasound, and histopathologic imaging. However, histopathologic imaging is still the gold standard for HCC diagnosis.[2] Assessing the histopathologic grade of HCC requires visual inspection of cancer regions by experienced pathologists.[3] Nevertheless, visual inspection of cancer regions by pathologists in whole-slide pathologic images is labor-intensive and time consuming because such images usually are in the gigapixel range. It also is highly reliant on

Supported by the Shenzhen Science and Technology Program of China grant JCYJ20200109115420720 (W.Q.); National Natural Science Foundation of China grants 61901463 (W.Q.), 62001464 (Z.Z.), and U20A20373 (Y.X.); and Guangdong province key research and development areas grant 2020B1111140001 (W.Q.).

S.D., Y.T., and W.H. contributed equally to this work.

Disclosures: None declared.

https://doi.org/10.1016/j.ajpath.2021.11.009

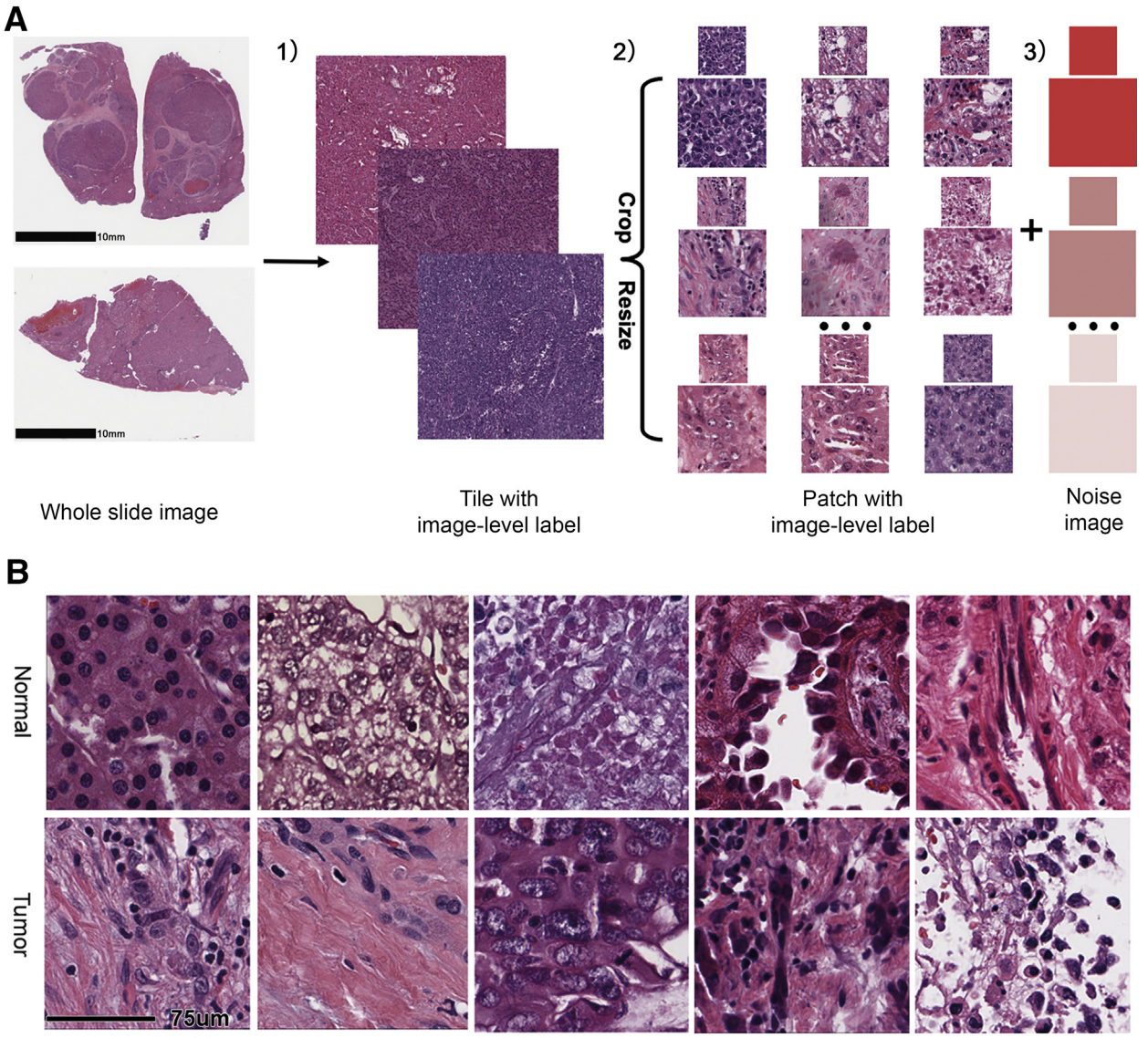

Figure 1 Illustration of data preprocessing. **A:** The data preprocessing included the following three steps. **(1)** The whole-slide images (WSIs) were cut into multiple tiles, and these tiles were classified as normal or tumor by the pathologists. **(2)** The classified tiles were cropped into the patches, each patch had an image-level label that was the same as the label of the tile from which it came. The patch sizes were resized as required by the network. **(3)** Some noisy images produced by the Red-Green-Blue value of the blood vessels or staining impurities during the generation of patches were added. **B:** Examples of some patches generated by preprocessing. **Top row:** Some normal patches that were selected randomly from the normal patch data sets. **Bottom row:** Tumor patches that were selected randomly from the tumor patch data sets. Each **column** represents different samples. +, noise images were to patches with an image-level label; …, many patches with an image-level label and noise images were omitted in the figures. Scale bars: 10 mm (**A**); 75 μm (**B**).

expert knowledge because of the varied appearance of HCC lesions across patients.[4] There is therefore a strong demand for an automatic method for quickly and accurately detecting cancer regions of HCC.

Over the years, numerous methods have been proposed for the automatic detection of liver cancer regions.[5,6] For instance, Atupelage et al[6] introduced an automated method based on cell nuclei classification in which the nuclear segmentation was performed by a random forest classifier with pixel-based classification. A drawback of this method is that the features of each pixel need to be calculated and selected manually. With the rapid development of deep learning, there has been a surge of interest in the automatic detection of liver cancer regions focused on using convolutional neural networks (CNN).[7–9] For instance, Schmitz et al[7] introduced a family of multi-encoder, fully convolutional neural networks with deep fusion for HCC segmentation. Wang et al[8] proposed a hybrid neural network based on multitask learning and ensemble learning techniques for automatic HCC segmentation in hematoxylin and eosin–stained whole-slide images (WSIs). Although these detection methods provided promising results, they use pixel-level detection. Pixel-level annotation is typically generated by experienced pathologists and is therefore time consuming and labor-intensive.

Table 1 Case Characteristics in Training, Validation, and Testing Sets

	Training examinations			Testing examinations		
Data type	Data set	Normal	Cancer	Data set	Normal	Cancer
Cases	70	—	—	30	—	—
Patches (A)	5.9×10^5	3.1×10^5	2.8×10^5	2.9×10^5	1.8×10^5	1.7×10^5
Patches (B)	2.7×10^5	1.0×10^5	1.6×10^5	1.6×10^5	7.5×10^4	7.9×10^4

Row A presents patches with a magnification of ×40, ×15, and ×20; row B presents the patches with a magnification of ×32 and ×16.

Furthermore, pathologist labeling is subjective, and inconsistent annotations inevitably are likely to affect the training process.

Weakly supervised detection is an emerging field that addresses the annotation limitation of pixel-level detection, and uses patch-level with class labels instead of pixel-level labels. Several studies about weakly supervised detection have been published recently.[8,10–14] For instance, Priego-Torres et al[10] performed a weakly supervised automatic segmentation of stained breast cancer images and obtained 0.956 segmentation accuracy and 0.925 frequency weighted intersection over union. Motivated by the success of the weakly supervised detection method, it was used for cancer region detection of HCC,[14] in which the model's performance was compared under ×15 and ×20 magnification, and a segmentation accuracy of 0.880 and 0.872 was achieved, respectively. That work focused only on single-scale detection. However, in the clinical diagnosis and grading of malignancy, pathologists often combine multimagnification information, varying in spatial scale from the subnuclear [≈O (0.1 μm)] through cellular [≈O (10 μm)] and intercellular [≈O (100 μm)], to glandular and other higher organizational features [≈O (1 mm)].[7] Hence, the performance of cancer region detection based on the information of a single scale is artificially limiting.

To address the limitation of the single-scale detection method, in this study, a multiscale attention learning strategy was developed that was inspired by the action of the pathologist (ie, the diagnosis and grading of malignancy involving a range of different scales). Some studies have adopted multiscale learning for liver cancer detection,[7,15] and their results show its superiority compared with single-scale detection methods. However, earlier works typically used a simple method to fuse the features of different scales, such as straightforward concatenation.[7] They regarded all scales as equally important, which is inconsistent with the pathologists' process of clinical diagnosis. With this drawback in mind, an attention strategy was introduced to dynamically learn the relative weight to attach to different scales (ie, using the network to determine the scales of information to focus on). This mechanism, termed attention mechanism, has been introduced into deep learning for medical image analysis tasks.[16–22] For instance, Liu et al[16] proposed a deep residual-attention CNN to segment ischemic stroke and white matter hyperintensity lesions simultaneously in magnetic resonance images, in which they introduced an attention branch that included a trunk branch and a dilated soft mask branch for generating (detecting) high-quality features of the input images. Wang et al[17] designed a voxel-wise weight map to allow the generator to pay more attention to the lesion region. Lei et al[19] presented a self–co-attention network for automatic breast anatomy segmentation, in which three attention mechanisms—channel-wise attention, spatial-wise attention, and the co-attention mechanism—were used to improve the segmentation performance. These studies have proven that the attention mechanism is effective in medical image analysis. However, there are still few studies on the cancer region detection of HCC in whole-slide pathologic images.

Here, an attention strategy was used in cancer region detection and validated through a comparison of the different performance metrics of a multiscale attention detection model and a single-scale detection model. The results of the former model were also compared against those of three pathologists: a junior pathologist, an intermediate-level pathologist, and a senior pathologist. To further verify how multiscale attention affects the classification performance, the attention maps that compared changes in the feature maps of single-scale and multiscale inputs were visualized.

Materials and Methods

Data Set

One hundred liver WSIs were collected from a publicly available web-based resource for cancer researchers (The Cancer Genome Atlas, *https://portal.gdc.cancer.gov*, last accessed October 28, 2019). The available 100 liver WSIs were selected randomly from the database, and all selected slides contained tumor tissues. These WSIs were acquired by scanning formalin-fixed, paraffin-embedded hematoxylin and eosin–stained tissues using an Aperio AT Turbo (Leica, Wetzlar, Germany) at the maximum available resolution of ×40 (0.25 um/pixel). The image-level annotation (normal or tumor) was performed by two board-certified pathologists with at least 15 years of clinical experience. In addition, to compare the detection result produced by the proposed method with pathologists, 37 tiles from the testing sets were selected randomly and the other three pathologists with 2 years of experience (termed a junior pathologist), 8 years of experience (termed an intermediate-level pathologist), and more than 20 years of experience (termed a senior

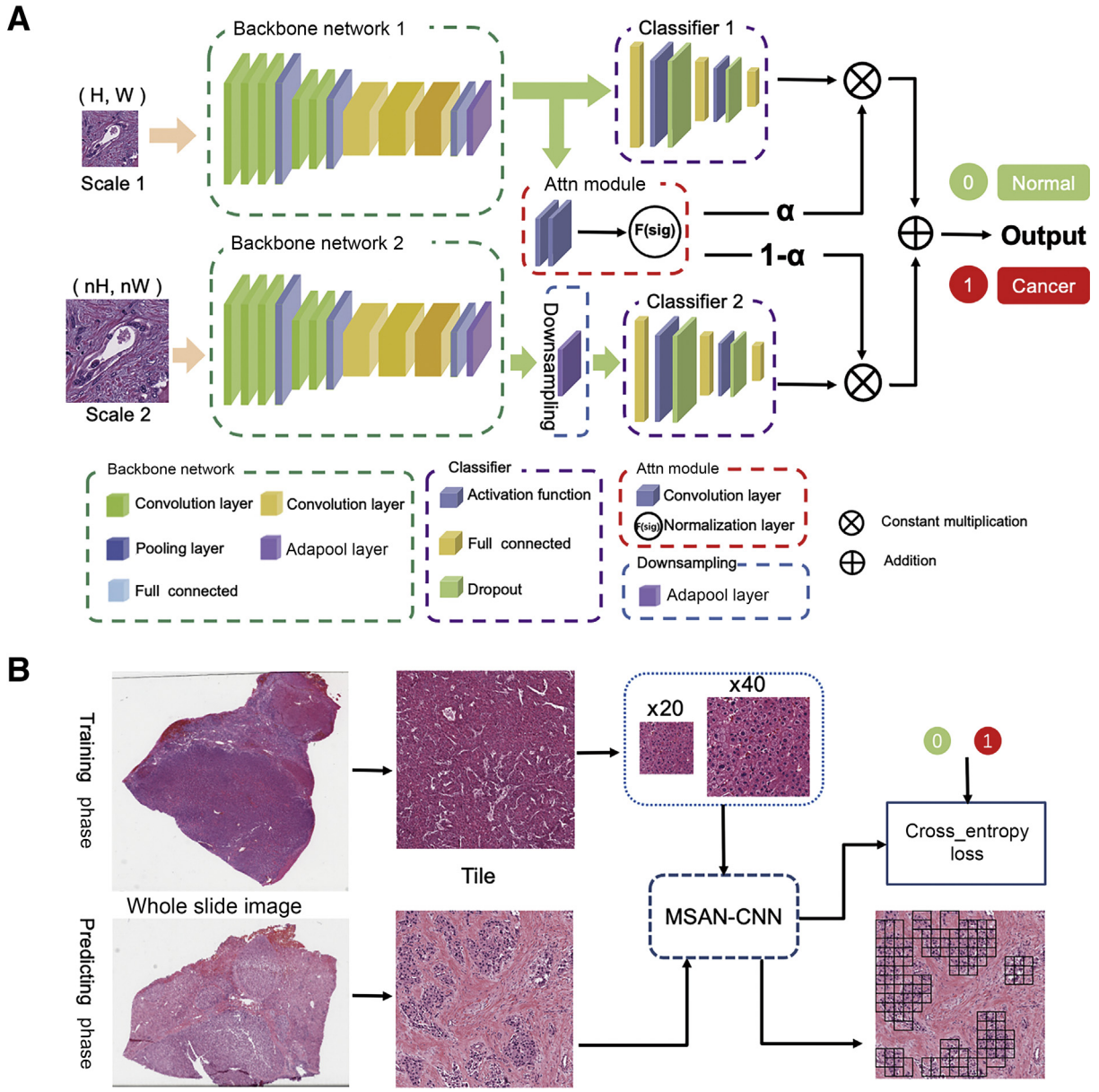

Figure 2 Illustration of the framework for automatic cancer region detection of hepatocellular carcinoma using multiscale attention convolutional neural network architecture. **A:** Details of the multiscale attention convolutional neural network (MSAN-CNN) framework are indicated. **B:** The flowchart of experimental process is indicated. The process was divided into a training phase and a predicting phase. (H, W) indicates height and width of the patch; n indicates value of proportion. AdaPool, adaptive pooling; Attn, attention.

pathologist) were invited to manually outline the pixel-level cancer region of these selected tiles by using the Automated Slide Analysis Platform (version 1.9; *https://github.com/computationalpathologygroup/ASAP*). These five pathologists were from three different institutions.

Methods

Image Preprocessing

The preprocessing methods used for these different magnification data sets were as follows: WSIs were first cut into multiple tiles with 4096 × 4096 pixels using Python (Beaverton, OR). Next, these tiles were categorized (as normal or tumor) independently by two board-certified pathologists with at least 15 years of clinical experience. Then, the marked tiles were cropped into small patches with 448 × 448 pixels. Each patch was given an image-level label, which was the same as the label of the tile from which it came. The cropped patches labeled normal were cut from the normal tiles, and no tumor tissue was included; the patches labeled tumor were cut from the tumor tiles and no normal tissue was included. Finally, the cropped patches were resized as required by the network. Furthermore, to improve robustness, some noisy images produced by the

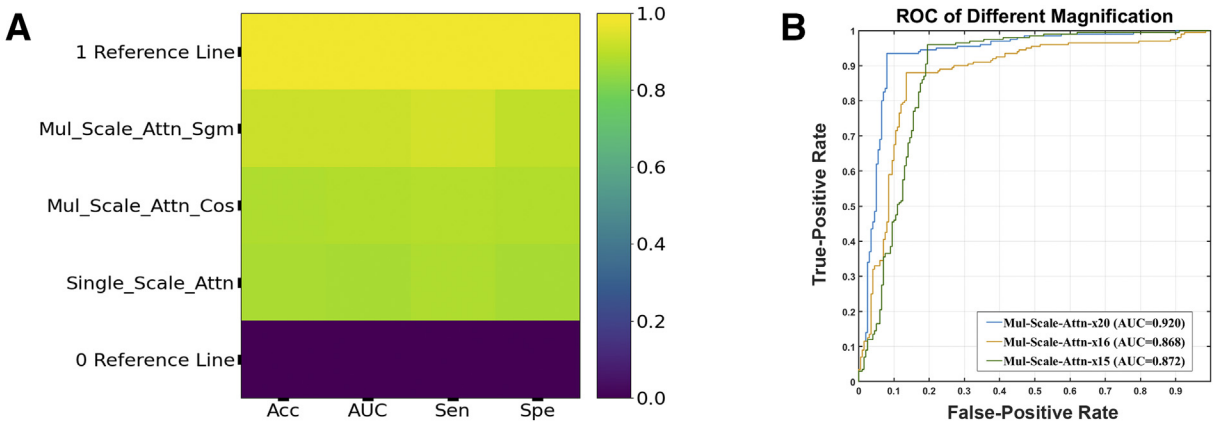

Figure 3 Classification results of the patches. **A:** Classification result comparison of accuracy, area under curve (AUC), sensitivity, and specificity of the three model configurations: Single_Scale, Mul_Scale_Attn_Cos, and Mul_Scale_Attn_Sgm. **B:** Receiver operating characteristic curves (ROCs) for the prediction of three different combinations of multiscale inputs. Acc, accuracy; Mul-Scale-Attn-×15, combination of ×15 + ×40 magnification; Mul-Scale-Attn-×16, combination of ×16 + ×32 magnification; Mul-Scale-Attn-×20, combination of ×20 + ×40 magnification; Mul_Scale_Attn_Cos, multiscale attention detection with the absolute value of cosine as the normalization function; Mul_Scale_Attn_Sgm, multiscale attention detection with sigmoid as the normalization function; Sen, sensitivity; Single_Scale, single-scale detection; Spe, specificity.

Red-Green-Blue value of the blood vessels or staining impurities were added during the generation of patches. The details of the preprocessing method are illustrated in Figure 1A. Randomly selected examples of patches generated by preprocessing are shown in Figure 1B, in which the examples of normal patches were selected from the normal patch data sets, and the examples of tumor patches were selected from the tumor patch data sets.

In the method proposed here, multiscale attention CNN (MSAN-CNN), five different scale images (ie, ×40, ×32, ×20, ×16, and ×15) were used. The ×40 magnification data used the full resolution of a scan, and the ×32 magnification data were resized from the ×40 data at the image level. For the ×20, ×16, and ×15 magnification data, images with ×40 magnification were scaled down to ×20 and ×15, and from ×32 to ×16 at the patch level. The corresponding patch sizes were as follows: 448 × 448 pixels for images of ×40 magnification, 560 × 560 pixels for images of ×32 magnification, 224 × 224 pixels for images of ×20 and ×16 magnification, and 168 × 168 pixels for images of ×15 magnification. Furthermore, to increase the training sample, data augmentation by horizontal flipping, vertical flipping, and rotation was used. Details of the data set are shown in Table 1.

Multiscale Attention Mechanism

The core idea of the proposed multiscale attention mechanism was that using attention weights to determine which scale of the image contributes more to the classification result. Specifically, two differently scaled images (scale 1 and scale 2) were fed into the backbone to extract features. Then, the scale 1 image was fed into an attention module to produce the attention weight: α. Then, α was used to focus the scale 1 image and 1-α was used to focus the scale 2 image. If $\alpha > 0.5$, it meant that scale 1 images made a greater contribution to the final classification result, and vice versa.

Table 2 Results of Training and Validation Sets for Three Different Combinations of Multiscale Inputs

Data set	Evaluation indicator	Accuracy	AUC	Sensitivity	Specificity
Validation set	×16 + ×32	0.986 ± 0.003 (S)	0.986 ± 0.003 (S)	0.989 ± 0.002 (S)	0.984 ± 0.005 (S)
		0.933 ± 0.028 (C)	0.929 ± 0.028 (C)	0.949 ± 0.030 (C)	0.908 ± 0.035 (C)
	×15 + ×40	0.986 ± 0.001 (S)	0.986 ± 0.001 (S)	0.985 ± 0.004 (S)	0.988 ± 0.003 (S)
		0.896 ± 0.106 (C)	0.892 ± 0.109 (C)	0.925 ± 0.150 (C)	0.899 ± 0.080 (C)
	×20 + ×40	0.988 ± 0.002 (S)	0.988 ± 0.002 (S)	0.991 ± 0.004 (S)	0.985 ± 0.001 (S)
		0.905 ± 0.096 (C)	0.906 ± 0.095 (C)	0.887 ± 0.119 (C)	0.924 ± 0.085 (C)
Testing set	×16 + ×32	0.868 ± 0.013 (S)	0.868 ± 0.013 (S)	0.927 ± 0.010 (S)	0.793 ± 0.025 (S)
		0.826 ± 0.034 (C)	0.827 ± 0.033 (C)	0.908 ± 0.046 (C)	0.746 ± 0.076 (C)
	×15 + ×40	0.875 ± 0.011 (S)	0.872 ± 0.010 (S)	0.923 ± 0.013 (S)	0.831 ± 0.026 (S)
		0.824 ± 0.102 (C)	0.823 ± 0.099 (C)	0.842 ± 0.053 (C)	0.790 ± 0.149 (C)
	×20 + ×40	**0.921 ± 0.014 (S)**	**0.920 ± 0.014 (S)**	**0.933 ± 0.014 (S)**	**0.906 ± 0.016 (S)**
		0.885 ± 0.096 (C)	0.886 ± 0.097 (C)	0.890 ± 0.083 (C)	0.883 ± 0.127 (C)

The best results on validation sets are shown in bold. Values are reported as means ± SD using fivefold cross-validation.
AUC, area under the curve; C, cosine absolute value; S, sigmoid function.

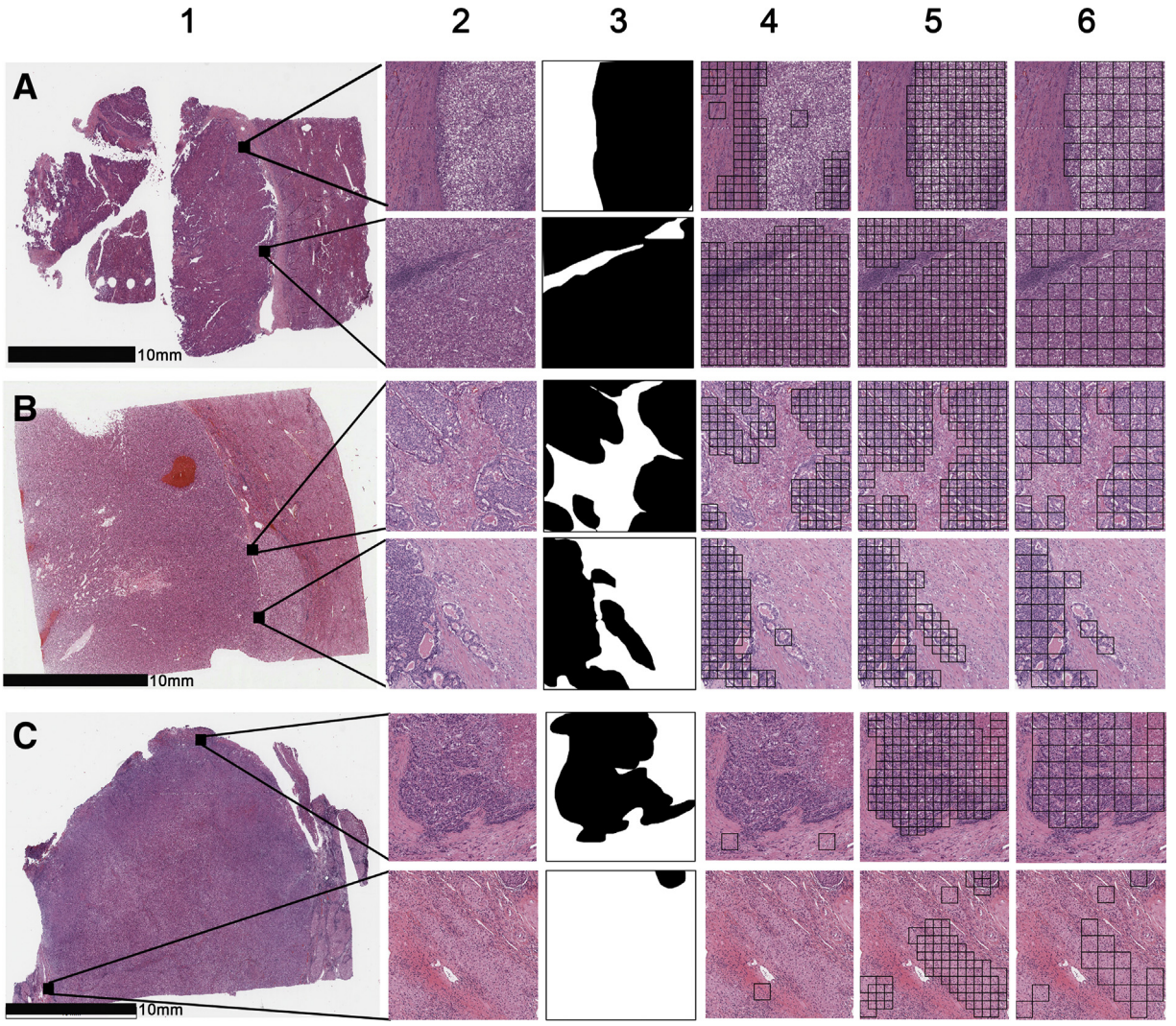

Figure 4 Qualitative results of the single-scale detection, multiscale attention with overlap detection and multiscale attention without overlap detection methods. **A**–**C**: **Column 1** shows three representative samples from different whole-slide images (WSIs), from each of which two tiles are selected for display and analysis in **columns 2 to 6**. In the second to third columns are images of tiles (**2**), ground-truth (**3**); and in the fourth to the sixth columns are the results of their analysis by single-scale (**4**), multiscale with overlap (**5**), and multiscale without overlap (**6**) methods, respectively. The white parts represent normal regions, and the black parts represent tumor regions in the ground-truth and the predicted result. Scale bars = 10 mm.

MSAN-CNN Architecture

In this section, the details of the proposed MSAN-CNN are presented. The implementation details are shown in Figure 2A. First, images of low and high scale, scale 1 and scale 2, respectively, were fed into two separate backbones to extract features, in which each backbone used the VGG-19[23] network. VGG is one of the state-of-the-art deep-learning models for classification that won second place in the ImageNet Large-Scale Visual Recognition Challenge 2014 competition. In a previous study,[14] VGG-19, used as the base network, yielded encouraging detection results. Then, the features of scale 1 were fed into two branches. The first branch was a classifier (termed *Classifier 1*) combined with a set of connection layers, the activation functions and dropout functions. The second branch was an attention module that includes a set of convolution layers and a normalization layer. The weights of attention were produced by a normalized function:

$$\alpha = sig(\mathrm{F}) \tag{1}$$

where F presents the features of scale 1, *sig* is the logistic sigmoid function $f(x) = 1/[1 + e(-x)]$, and α is the learned attention weight [(S) will be used to indicate the equation hereafter].

At the same time, to make the length and width of features consistent with scale 1, scale 2 was down-sampled by an AdaPool operation (*https://pytorch.org/docs/stable/nn.html?Highlight=adaptive avgpool#pooling-layers*, last accessed September 26, 2021). Afterward, the down-sampled features were fed into a classifier (termed *Classifier 2*) that was

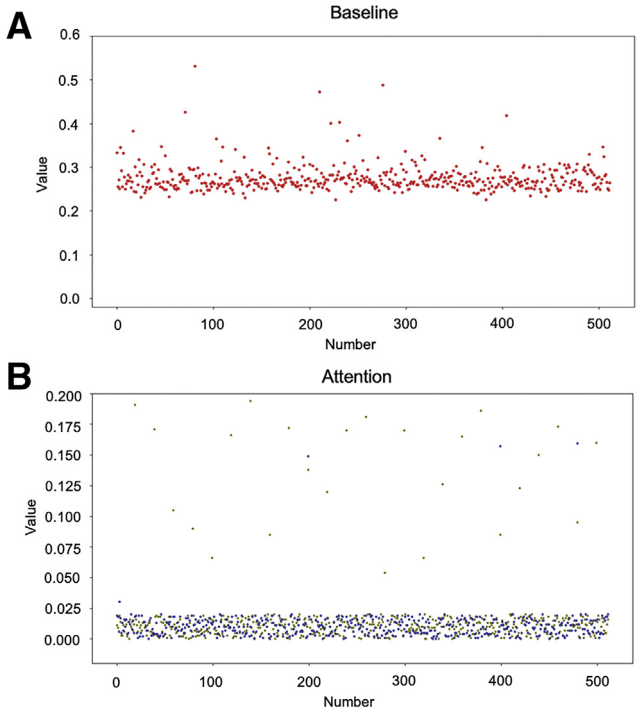

Figure 5 Visualization of the attention maps. **A:** Attention maps of the baseline; ie, single-scale detection method of an input image that has a magnification of ×20. **B:** Attention maps of multiscale attention convolutional neural network (MSAN-CNN); the **blue dots** represent the input image that is a multiscale combination of ×20 + ×40 and the **brown dots** represent the input image that is a multiscale combination of ×15 + ×40.

identical to Classifier 1. Subsequently, α was multiplied by the result of *Classifier 1*, and 1-α was multiplied by the result of *Classifier 2*. Hence, the learned attention weights α and 1-α were used to determine the importance of these different scales. Finally, a fused classifier result was obtained according to the following formula:

$$\text{output} = \text{c\_1} + \text{c\_2} \quad (2)$$

where c_1 represents the weighted results of *Classifier 1* and c_2 represents the weighted results of *Classifier 2*.

Experiments

The proposed MSAN-CNN model was trained on extracted patches with image-level labels from the proposed data set by PyTorch on a NVIDIA GPU Tesla V100 (32G) (Santa Clara, CA). The data sets were split randomly into 2 parts: 70 (70%) WSIs were training data, which were used to train and validate the model, and 30 (30%) WSIs were used to test the trained model. Training and testing sets were kept independent of each other (ie, the cases in the training set never appeared in the testing set). In our experiment, fivefold cross-validation was performed. The illustration of the workflow is shown in Figure 2B, which is divided into the training process and the testing process. During the training process, the differently scaled patches were fed into the proposed MSAN-CNN for classification. Three different combinations of patches as multiscale inputs were used (×20, ×40; ×15, ×40; and ×16, ×32). In which ×20, ×16, and ×15 were used as low-scale images, and ×40 and ×32 were used as high-scale images. For the attention mechanism, two normalization functions were used to generate attention weight: the sigmoid and the absolute value of cosine. The architecture and the hyperparameters included 32 mini-batches and 80 epochs of this MSAN-CNN, which was optimized on the training set using manual hyperparameter tuning. The network was trained with Adam[24] and, to avoid overfitting, the model parameters that performed best on the validation set were saved. During the testing process, the small patches from the testing data set were classified by the trained classifier. Our code is available at (GitHub, *https://github.com/SH-Diao123/MSAN-CNN*, last accessed November 10, 2021).

Post-Processing

After detection, the results undergo processing in the form of aggregation operations based on the classification of the patches. Herein, the probability of each patch was combined independently for aggregation, and the probability of overlapping regions was excluded.

Evaluating MSAN-CNN Model Performance on The Cancer Genome Atlas Data

To quantitatively evaluate the accuracy of cancer region detection of HCC, the accuracy, sensitivity, specificity, receiver operator characteristic curve, and area under the curve (AUC) were used as metrics. A qualitative analysis of HCC detection also was performed and an attention map was visualized to better illustrate the advantages of the proposed MSAN-CNN. Finally, the results were compared with those of the three pathologists.

Results

Classification Results

Figure 3A compares the patch-level classification results of normal and tumor classifications for three different modeling configurations: Single_Scale, Mul_Scale_Attn_Cos, and Mul_Scale_Attn_Sgm. Single_Scale, which is a single-scale detection method using the ×20 magnification patches as the input of the classification neural network, was proposed in a previous work.[14] Mul_Scale_Attn_Cos and Mul_Scale_Attn_Sgm were multiscale attention methods that used a multiscale combination of input patches at magnifications of ×20 and ×40. The difference between Mul_Scale_Attn_Cos and Mul_Scale_Attn_Sgm is that they use different normalization functions. The former uses the absolute value of cosine as the normalization function and the latter uses the sigmoid as the normalization function for realizing attention to different multiscale combinations. Both of them used

multiscale combinations of input patches at magnifications of ×20 and ×40. Figure 3A shows that the Mul_Scale_Attn_Sgm model obtained the best results on all evaluation metrics, and that the Mul_Scale_Attn_Cos module obtained the second best detection performance. This shows the effectiveness of the multiscale attention module in the classification task. It also shows that the normalization function affects the classification results (ie, a sigmoid is preferable to the absolute value of cosine).

The MSAN-CNN model's performance was tested further under three additional multiscale combinations of input patches: magnification patches of ×20, ×40; ×15, ×40; and ×16, ×32. The results are listed in Table 2. Based on the test data set, for accuracy, AUC, sensitivity, and specificity, the best results were 0.921, 0.920, 0.933, and 0.906, respectively. All of them were obtained at the magnification combinations of ×20, ×40 (S). Moreover, the sigmoid normalization function performed better than the absolute value of the cosine normalization function. Furthermore, the significance of the differences in accuracy, AUC, sensitivity, and specificity among magnification patches of ×20, ×40; ×15, ×40; and ×16, ×32 were evaluated using the *t*-test. The results show all *P* values are less than 0.01, suggesting that the improvement of ×20, ×40 (S) magnification is statistically significant compared with other multiscale inputs. The receiver operator characteristic curves of three different combinations of multiscale inputs also were shown, in which all of them used the sigmoid as the normalization function. The results are shown in Figure 3B, indicating that the greatest AUC occurs using the multiscale combination of ×20, ×40 magnification. The multiscale combination of ×15, ×40 magnification gets the second AUC score.

Qualitative Analysis

Figure 4 shows the results of a qualitative comparison in which the single-scale detection method was used on magnification patches of ×20, and the multiscale detection method was used on multiscale combination magnification patches of ×20, ×40. Single-scale detection for the first tile of WSI A (first row of column 4) failed to detect most cancer regions of HCC and mistakenly detected the normal regions as tumor regions. However, the proposed MSAN-CNN, used for the same tile, both with and without overlap (first row of columns 5 and 6, respectively), had considerably fewer false-positive and false-negative results. The second tile of WSI A is largely covered with tumor regions and has few normal regions. In this case, the single-scale detection method detected most tumor regions but mistakenly detected the normal regions as tumor regions (second row of column 4), while the proposed MSAN-CNN correctly detected both the normal and tumor regions (second row of columns 5 and 6). This indicates that the proposed MSAN-CNN has the ability to address class-imbalance cases. For WSI B, the single-scale detection method failed to detect some small liver cancer

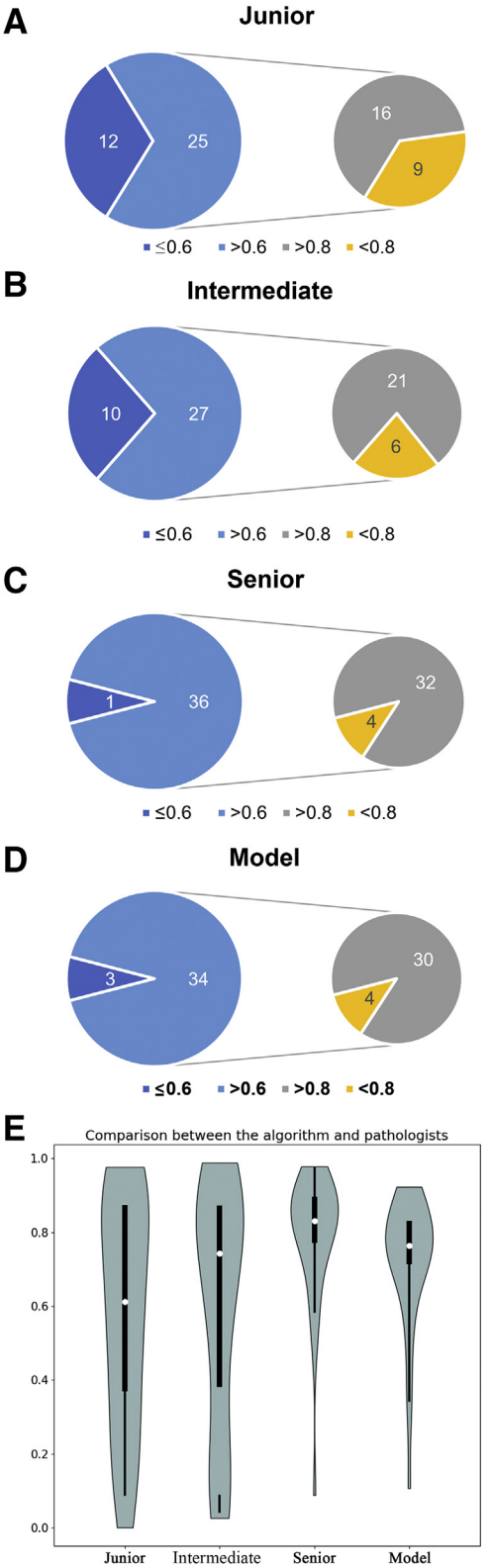

Figure 6 Detection performance measured with dice coefficient and intersection over union (IOU) for comparison with a junior and an intermediate-level pathologist on 37 different tiles. **A**—**D:** Detection performance measured with dice coefficient on a junior, an intermediate-level, and a senior pathologist, and the proposed model. **E:** Detection performance measured with IOU on a junior, an intermediate-level, and a senior pathologist, and the proposed model based on a violin plot.

Table 3 Performance Comparison in Terms of Dice Coefficients between Our Presented Approach and State-of-the-Art Approaches

Model	Dice
Unet++[25]	0.807 ± 0.016
Deeplabv3[26]	0.798 ± 0.019
Lite R-ASPP[27]	0.849 ± 0.003
Ours	0.861 ± 0.024

regions. However, most liver cancer regions were detected correctly with the proposed MSAN-CNN. In case of WSI C, the results of the single-scale detection method failed to detect cancer regions of HCC, while MSAN-CNN detected them successfully. However, MSAN-CNN also incorrectly identified some normal tissue as tumor tissue.

Visualization of the Attention Maps

To better understand how the multiscale attention module improves the final classification result, the attention maps were visualized (ie, scatter diagrams produced by averaging the features of all channels from the last layer before the classifier). The attention maps of the single-scale and MSAN-CNN detection methods are shown in Figure 5, A and B, respectively. The feature distribution of the single-scale detection method is concentrated and single, so almost all of the features play important roles in the classification results. However, in the MSAN-CNN, only some features are essential for the classification results and unimportant features have values close to zero. This indicates that the MSAN-CNN learns to discriminate between features better than the single-scale detection method (ie, the MSAN-CNN model focuses on the features that can distinguish the normal and the tumoral parts of the tissue).

The Results Compared with Those of Pathologists

Figure 6 shows the results of a comparison between cancer region detection of HCC using the proposed method against that of three pathologists (a junior, an intermediate level, and a senior pathologist) for 37 tiles. The dice coefficient was used as the evaluation metric and recorded the numbers of dice values greater than 0.6, less than 0.6, greater than 0.8, and less than 0.8, respectively. As reported in Figure 6, A-D, for the proposed model, the number of dice values greater than 0.6 and 0.8 was higher than in the case of both junior and intermediate-level pathologists, and slightly lower than in the case of a senior pathologist. To clearly show the advantages of our model, the violin plot of the intersection over union was compared for the junior-level, the intermediate-level, and senior-level pathologists, and the MSAN-CNN, as illustrated in Figure 6E. The median of the latter was higher than that of either the junior pathologist or the intermediate-level pathologist, and slightly lower than that of the senior pathologist. The proposed model had a more concentrated statistical distribution than that of the junior or intermediate-level pathologists, which shows that it generated more stable diagnostic results compared with those by these two pathologists. Overall, this indicates that our model performs better than the junior and intermediate-level pathologists, and has the potential to generate detection results comparable with those of a senior pathologist.

The 37 tiles were selected randomly from the testing data set and processed on a computer with the following experiment setting: 32 core Intel I Xeon I CPU, E5-2620@2.10 GHz (Santa Clara, CA); Supermicro SYS-7048GR-TR 512.0G RAM (San Jose, CA); NVIDIA Tesla V100 GPU with 64 G; Ubuntu 16.04 (London, UK); and Python 3.6.4 with PyTorch 1.0 platform (Menlo Park, CA). It should be noted that our method generated detection results for every tile and that it only took an average of 7 seconds for processing of each tile, while the average time for the pathologist's annotation is 10 minutes.

Comparison with State-of-the-Art Methods

The proposed MSAN-CNN was compared with three state-of-the-art supervised segmentation methods: Unet++,[25] Deeplabv3,[26] and Lite R-ASPP.[27] As reported in Table 3, the proposed MSAN-CNN obtained an average dice coefficient of 0.861, which beat all other supervised segmentation methods.

Discussion

Recently, deep learning has shown great potential in medical image analysis in general,[28] and in the analysis of liver cancer regions based on computed tomography, magnetic resonance imaging, and ultrasound in particular.[29–31] However, the pixel-level detection method of HCC based on deep learning in WSI is a challenging task owing to the need of vast amounts of ground-truth data to train the network. To address this issue, many weakly supervised methods have been reported. However, they focused mainly on the single-scale detection method, while clinical diagnosis and grading of malignancy often rely on information that varies in spatial scale. Some tasks also have used multiscale information to detect the liver cancer region, but failed to consider the importance of various scales in different cases.

In the present work, MSAN-CNN was introduced as an extension of a former, initial study.[14] Compared with previous multiscale methods, a multiscale attention classifier was used here, which is an easy way of mimicking the action of pathologists who combine multimagnification information for diagnoses. The current experiments show that the classification performance of our MSAN-CNN outperformed the singe-scale detection method, as shown in Figure 3A.

To explore the effect of different multiscale combinations, three different multiscale combinations as the input of the network were tested. The multiscale combination of

magnifications of ×20, ×40 provided the best results in terms of accuracy, AUC, sensitivity, and specificity, as shown in Table 2 and Figure 3B.

Ablation studies were performed to study the effect of the normalization functions on the attention mechanism. The sigmoid function outperformed the absolute value of cosine by showing better accuracy, sensitivity, specificity, and AUC, as shown in Table 2 and Figure 3A. One explanation is that the value of the sigmoid function is unique over large intervals, while the absolute value of cosine is repeated. Therefore, repeated values may give the same attention to images of different scales, whereas, in fact, their contributions to the diagnosis result are different.

To test the performance of the proposed method in cancer region detection, the study qualitatively evaluated the results predicted by the different models, as shown in Figure 4. The proposed MSAN-CNN was able to effectively reduce false-positive and false-negative results compared with the single-scale detection method.

To further illustrate how the multiscale attention model affects the final classification results, the attention maps were visualized, as shown in Figure 5. The MSAN-CNN studied more discriminative features than the single-scale detection method. The discriminative features make the classification network better at distinguishing the normal and the tumoral parts of the tissue.

In the present study, the results produced by MSAN-CNN were also compared with those of the pathologists, using dice value statistics and the medians of a violin plot. Our MSAN-CNN performed better than the junior pathologist and the intermediate-level pathologist. More importantly, the proposed model obtained more stable diagnostic results compared with the ones generated by these two pathologists. The statistical distribution of the MSAN-CNN violin plot was more concentrated than in the case of the junior pathologist and the intermediate-level pathologist, as shown in Figure 6E. Interestingly, the results of the proposed model were slightly inferior to those of the senior pathologist. However, the time spent by the MSAN-CNN method was dramatically less than that required by the pathologists.

In conclusion, this study addressed the challenge of cancer region detection of HCC based on a single scale by using a multiscale attention convolutional neural network. Therefore, the pipeline proposed here has good potential as a tool for assisting pathologists to delineate cancer regions. The proposed MSAN-CNN framework has shown promising results. However, in this study it only distinguished between two categories: normal and tumor. Future work will focus on the detection of multiple tumor types simultaneously.

References

1. Bialecki ES, Di Bisceglie AM: Diagnosis of hepatocellular carcinoma. HPB (Oxford) 2005, 7:26–34
2. Irshad H, Veillard A, Roux L, Racoceanu D: Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. IEEE Rev Biomed Eng 2013, 7:97–114
3. Chen MY, Zhang B, Topatana W, Cao JS, Zhu HP, Juengpanich S, Mao QJ, Yu H, Cai XJ: Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. NPJ Precis Oncol 2020, 4:14
4. Schlageter M, Terracciano LM, D'Angelo S, Sorrentino P: Histopathology of hepatocellular carcinoma. World J Gastroenterol 2014, 20: 15955–15964
5. Aziz MA, Kanazawa H, Murakami Y, Kimura F, Yamaguchi M, Kiyuna T, Yamashita Y, Saito A, Ishikawa M, Kobayashi N: Enhancing automatic classification of hepatocellular carcinoma images through image masking, tissue changes and trabecular features. J Pathol Inform 2015, 6:26
6. Atupelage C, Nagahashi H, Kimura F, Yamaguchi M, Tokiya A, Hashiguchi A, Sakamoto M: Computational hepatocellular carcinoma tumor grading based on cell nuclei classification. J Med Imaging 2014, 1:034501
7. Schmitz R, Madesta F, Nielsen M, Krause J, Steurer S, Werner R, Rösch T: Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture. Med Image Anal 2021, 70:101996
8. Wang X, Fang Y, Yang S, Zhu D, Wang M, Zhang J, Tong K-y, Han X: A hybrid network for automatic hepatocellular carcinoma segmentation in H&E-stained whole slide images. Med Image Anal 2021, 68:101914
9. Wang R, He Y, Yao C, Wang S, Xue Y, Zhang Z, Wang J, Liu X: Classification and segmentation of hyperspectral data of hepatocellular carcinoma samples using 1-D convolutional neural network. Cytometry A 2020, 97:31–38
10. Priego-Torres BM, Sanchez-Morillo D, Fernandez-Granero MA, Garcia-Rojo M: Automatic segmentation of whole-slide H&E stained breast histopathology images using a deep convolutional neural network architecture. Expert Syst Appl 2020, 151:113387
11. Song ZG, Zou SM, Zhou WX, Huang Y, Shao LW, Yuan J, Gou XN, Jin W, Wang ZB, Chen X, Ding XH, Liu JH, Yu CK, Ku C, Liu CC, Sun Z, Xu G, Wang YF, Zhang XQ, Wang DD, Wang SH, Xu W, Davis RC, Shi HY: Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. Nat Commun 2020, 11:4294
12. Ciga O, Martel AL: Learning to segment images with classification labels. Med Image Anal 2021, 68:101912
13. Thomas SM, Lefevre JG, Baxter G, Hamilton NA: Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer. Med Image Anal 2021, 68:101915
14. Diao S, Luo W, Hou J, Yu H, Chen Y, Xiong J, Xie Y, Qin W: Computer aided cancer regions detection of hepatocellular carcinoma in whole-slide pathological images based on deep learning. International Conference on Medical Imaging Physics and Engineering (ICMIPE). IEEE 2019:1–6
15. Huang W-C, Chung P-C, Tsai H-W, Chow N-H, Juang Y-Z, Tsai H-H, Lin S-H, Wang C-H: Automatic HCC detection using convolutional network with multi-magnification input images. International Conference on Artificial Intelligence Circuits and Systems (AICAS). IEEE 2019:194–198
16. Liu LL, Kurgan L, Wu FX, Wang JX: Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease. Med Image Anal 2020, 65:101791
17. Wang GT, Song T, Dong Q, Cui M, Huang N, Zhang ST: Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks. Med Image Anal 2020, 65:101787
18. Sekuboyina A, Kukačka J, Kirschke JS, Menze BH, Valentinitsch A: Attention-driven deep learning for pathological spine segmentation. International Workshop on Computational Methods and Clinical Applications in Musculoskeletal Imaging 2017:108–119
19. Lei BY, Huang S, Li H, Li R, Bian C, Chou YH, Qin J, Zhou P, Gong XH, Cheng JZ: Self-co-attention neural network for anatomy

segmentation in whole breast ultrasound. Med Image Anal 2020, 64: 101753
20. Dou H, Karimi D, Rollins CK, Ortinau CM, Vasung L, Velasco-Annis C, Ouaalam A, Yang X, Ni D, Gholipour A: A deep attentive convolutional neural network for automatic cortical plate segmentation in fetal MRI. IEEE Trans Med Imaging 2021, 40:1123−1133
21. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B: Attention u-net: learning where to look for the pancreas. arXiv 2018, [Preprint], arXiv:1804.03999
22. Roy AG, Navab N, Wachinger C: Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. IEEE Trans Med Imaging 2018, 38:540−549
23. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. arXiv 2014, [Preprint], arXiv:1409.1556
24. Kingma DP, Ba J: Adam: a method for stochastic optimization. arXiv 2014 [Preprint], arXiv:1412.6980
25. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J: Unet++: A Nested U-net Architecture for Medical Image Segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. New York, NY, Springer, 2018. pp. 3−11
26. Chen LC, Papandreou G, Schroff F, Adam H: Rethinking atrous convolution for semantic image segmentation. arXiv 2017, [Preprint], arXiv:1706.05587
27. Howard A, Sandler M, Chu G, Chen L-C, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V: Searching for mobilenetv3. IEEE International Conference on Computer Vision 2019:1314−1324
28. Zemouri R, Zerhouni N, Racoceanu D: Deep learning in the biomedical applications: recent and future status. Appl Sci (Basel) 2019, 9: 1526
29. Lange A, Muniraj T, Aslanian HR: Endoscopic ultrasound for the diagnosis and staging of liver tumors. Gastrointest Endosc Clin 2019, 29:339−350
30. Li XM, Chen H, Qi XJ, Dou Q, Fu CW, Heng PA, H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. Ieee Trans Med Imaging 2018, 37:2663−2674
31. Christ PF, Ettlinger F, Grün F, Elshaera MEA, Lipkova J, Schlecht S, Ahmaddy F, Tatavarty S, Bickel M, Bilic P: Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv 2017, [Preprint], arXiv:1702.05970