A Multi-view Deep Convolutional Neural Networks for Lung Nodule Segmentation

Shuo Wang, Mu Zhou, Olivier Gevaert, Zhenchao Tang, Di Dong, Zhenyu Liu, and Jie Tian, *Fellow, IEEE*

***Abstract*— We present a multi-view convolutional neural networks (MV-CNN) for lung nodule segmentation. The MV-CNN specialized in capturing a diverse set of nodule-sensitive features from axial, coronal and sagittal views in CT images simultaneously. The proposed network architecture consists of three CNN branches, where each branch includes seven stacked layers and takes multi-scale nodule patches as input. The three CNN branches are then integrated with a fully connected layer to predict whether the patch center voxel belongs to the nodule. The proposed method has been evaluated on 893 nodules from the public LIDC-IDRI dataset, where ground-truth annotations and CT imaging data were provided. We showed that MV-CNN demonstrated encouraging performance for segmenting various type of nodules including juxta-pleural, cavitary, and non-solid nodules, achieving an average dice similarity coefficient (DSC) of 77.67% and average surface distance (ASD) of 0.24, outperforming conventional image segmentation approaches.**

I. Introduction

Automated lung nodule segmentation from Computed Tomography (CT) images provides valuable information for lung cancer computer-aided diagnosis. Given the fact of growing volumes of lung nodule CT images, developing robust automated segmentation model is of great clinical importance to avoid tedious manual processing and reduce inter-observer variability from human experts.

Despite the development of different approaches for lung nodule segmentation in recent years [1]–[4], achieving accurate segmentation performance continues to require attention because of the following two primary challenges. First, intensity-based methods with morphology operations [1], [2] and region growing [3] perform well on isolated nodules but fail to segment nodules in challenging locations, especially for nodules attached to surroundings that usually appear in CT. Second, sophisticated model-based methods [4], [5] often involve shape hypothesis or user-interactive parameters that can be sensitive to different type of lung nodules.

Our study is motivated by recent success of applying convolutional neural networks (CNN) for medical image analysis and pattern recognition [6]–[9]. As opposed to traditional morphology operation and region growing methods, CNN learns discriminative features that are adaptive to specific tasks automatically [10]. For instance, Havaei *et al.* [9] applied a CNN model for brain tumor segmentation, showing improved performance over hand-crafted image features. In our study, instead of involving lung nodule shape hypothesis or tuning model-based parameters, we propose a multi-view convolutional neural networks (MV-CNN) [11] to distinguish nodule voxels from background voxels in CT imaging. Our model has learned nodule-sensitive features from 0.34 million voxel patches automatically and revealed appealing segmentation results for various type of lung nodules.

Overall, our contributions are as follows: 1) The proposed MV-CNN can segment lung nodules in CT images without any shape hypothesis or user-interactive parameter settings, and it learns discriminative nodule-sensitive features automatically from a large amount of image data; 2) We propose a multi-scale patch strategy as the input of the MV-CNN to capture both detailed textures and nodule shape information; 3) The MV-CNN integrates three branches that can learn deep features from three orthogonal image views in CT (Fig. 1).

This paper is organized as follows. A detailed description of the MV-CNN is presented in Section II. Experimental datasets and evaluation criteria are introduced in Section III. Section IV provides the quantitative performance. Finally, conclusion and further discussions are presented in Section V.

II. Method

Our MV-CNN is designed to be an efficient CNN-based architecture for lung nodule segmentation. It aims to convert lung nodule segmentation into CT voxel classification (Fig. 1). Given a voxel in CT image, we extract three multi-scale patches centered on this voxel as the input to the CNN model and predict if this voxel belongs to the nodule.

The proposed MV-CNN incorporates three branches that process voxel patches from axial, coronal and sagittal view CT images respectively. The three branches share the same structure that consists of six convolutional layers (C1 to C6), two max-pooling layers (Max pooling 1, 2), and one fully connected layer (F7). The six convolutional layers in each CNN branch are divided into three blocks, where each block shares the exact same structure including two convolutional layers of kernel size 3×3. Between each block, max pooling operation with pooling window 2×2 and pooling step 2 is applied for feature selection. At the end of the CNN model, the three branches are merged through a fully connected

This paper is supported by the National Natural Science Foundation of China under Grant No. 81227901, 81527805.

$^+$S. Wang (wangshuo2014@ia.ac.cn), *D. Dong, *Z. Liu and *J. Tian (tian@ieee.org) are with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences; Beijing Key Laboratory of Molecular Imaging; University of Chinese Academy of Sciences, Beijing 100190, China. * are corresponding authors. $^+$ are co-first authors.

$^+$M. Zhou and O. Gevaert are with the Stanford Center for Biomedical Informatics Research (BMIR), Department of Medicine, Stanford University, CA 94305, USA.

Z. Tang is with the School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, Shandong Province, 264209, China.

978-1-5090-2809-2/17/$31.00 ©2017 IEEE

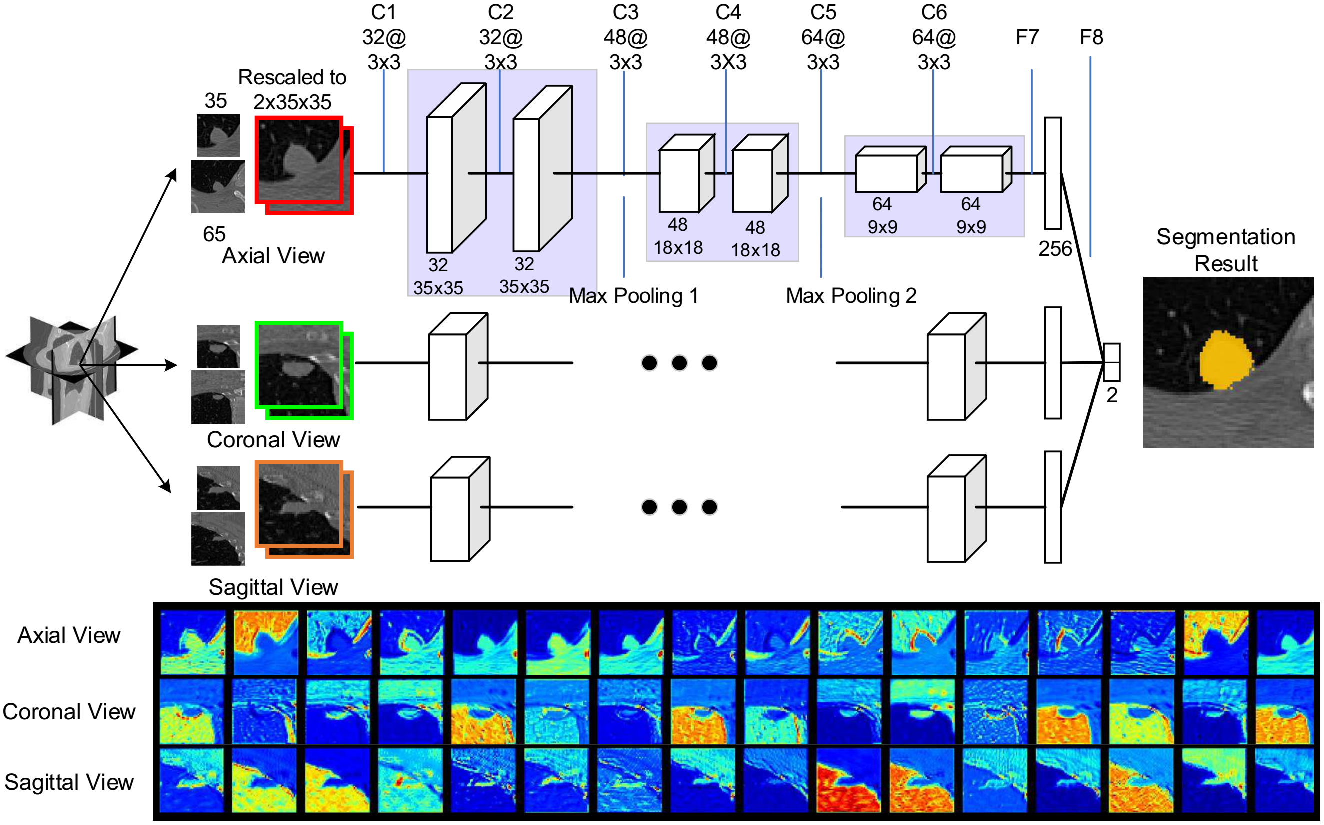

Fig. 1. Illustration of MV-CNN architecture. This network contains three branches aiming at capturing features from axial, coronal and sagittal image views and each branch takes a two-scale patch as input. Each CNN branch includes six convolutional layers (C1 to C6), two max-pooling layers, and a fully connected layer (F7). The three branches are finally merged in a fully connected layer (F8). The convolutional kernel size is denoted as $filter\ number@filter\ width \times filter\ height$ (i.e., 32@3 × 3 represents 32 filters of kernel size 3 × 3). The number below each layer indicates the feature map size after convolution. The bottom figure shows the feature maps of the C1 layer on three branches, indicating that the learned filters can capture different characteristics of nodules from input CT image (e.g., edge or solid part of a nodule).

layer (F8) to outcome the voxel label. A detailed architecture is illustrated in Fig 1. After each convolutional layer and the first fully connected layer (F7), a parametric rectified linear unit (PReLU) [12] is used as nonlinear activation function, and batch normalization is applied for training acceleration [13].

The input to each CNN branch is a two-channel multi-scale patch rather than a single-scale patch. The two scale patches are of size 65 × 65 and 35 × 35, and are scaled to 35 × 35 using third-order spline interpolation to form a two-channel patch. The small scale patch contains detailed texture information which is important for identifying voxel labels. Meanwhile, the large scale patch provides a broad scope of the nodule shape information. The multi-view structure takes 3-D information into consideration without involving much redundent image information compared with inputting a whole 3-D volume [14].

In the case of the output layer (F8) consisting of two units, the activation values are fed into a binary softmax function that are converted into probability distributions over the class labels. Namely, suppose that o_k is the k-th output of the network for a given input, the probability assigned to the k-th class is the output of the softmax function:

$$p_k = exp\left(o_k\right) / \sum_{h \subseteq \{0,1\}} exp\left(o_h\right) \qquad (1)$$

where $k = 0$ and $k = 1$ represent non-nodule and nodule voxels respectively.

The goal of network training is to maximize the probability of the correct class. This is achieved by minimizing the cross-entropy loss for each training sample. Suppose that y is the true label for a given input patch that belongs to {0,1}, the loss function is defined as:

$$L\left(W\right) = -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log \hat{y_n} + \left(1 - y_n\right)\log\left(1 - \hat{y_n}\right)\right] + \lambda |W| \qquad (2)$$

where $\hat{y_n}$ represents the predicted probability from MV-CNN and N is the number of samples. To avoid over fitting, the $1-norm$ regularization is used on the model weights W. λ controls the regularization strength, and is set to 5×10^{-4} in our model. The loss function is minimized during the model training process by computing the gradient of L over the network parameters W. During this process, the weights of the CNN are initialized with the Xavier algorithm [15], and they are updated by the stochastic gradient descent (SGD) algorithm using a momentum of 0.9, and a batch size of 128. The learning rate is set to 6×10^{-5} initially, and is reduced by a factor of 10 after every four epochs.

III. Experiment

A. Dataset

We used the public Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [16] for experimental evaluation. All the nodules in this dataset are annotated by up to four board-certified radiologists. Only the nodules with annotations from all four radiologists are used in our experiment (a total of 893 nodules). Nodule diameters in this dataset range from 2.03 mm to 38.12 mm, and the slice interval ranges from 0.45 mm to 5.0 mm. Because of the variability among four different radiologists, a 50% consensus criterion [1] is adopted to generate a single truth boundary.

To perform a rigorous evaluation of our method, we randomly partitioned the 893 nodules into three subsets including training, validation, and testing sets that are comprised of 450, 50, and 393 nodules respectively. We train the MV-CNN on the training set, and the validation set is used for determining the CNN training epoch number. Finally, the testing set is used for evaluating the model performance.

B. Evaluation criteria

Given the ground truth segmentation *Gt* and automated segmentation result *Auto*, the dice similarity coefficient (DSC) and average surface distance (ASD) are used as the primary evaluation criteria for assessing the automatic segmentation accuracy [17]. In addition, we also use the sensitivity (SEN) and positive predictive value (PPV) to demonstrate the voxel classification accuracy. Full definitions are listed in Eq. 3 to Eq. 5:

$$DSC = \frac{2 \cdot V\left(Gt \bigcap Auto\right)}{V\left(Gt\right) + V\left(Auto\right)} \tag{3}$$

$$ASD = \frac{1}{2}\left(\underset{i \in Gt}{\text{mean}} \min_{j \in Auto} d\left(i,j\right) + \underset{i \in Auto}{\text{mean}} \min_{j \in Gt} d\left(i,j\right)\right) \tag{4}$$

$$SEN = \frac{V\left(Gt \bigcap Auto\right)}{V\left(Gt\right)}, \; PPV = \frac{V\left(Gt \bigcap Auto\right)}{V\left(Auto\right)} \tag{5}$$

where *V* is the volume size counted in voxels and *d(i,j)* denotes the Euclidean distance between voxel *i* and voxel *j* measured in millimeters.

C. Model Training process

When generating training samples, we first identified the 3-D bounding cuboid for nodules in the training set, after which we extended the size of the cuboid by adding eight voxels along each axis to include additional non-nodule tissues inside. Afterwards, one quarter of the number of voxels in this expanded cubic were sampled uniformly. Finally, equal numbers of nodule and non-nodule samples were randomly selected to balance the training label. For each sampled voxel, multi-scale patches as presented in Section II were extracted from axial, coronal, and sagittal view images. As a result, 0.34 million patches were used for model training. We identified that DSC and ASD values determined on the validation set stabilized after 20 epochs of training, therefore we chose 20 epochs for MV-CNN training. After the model training was completed through *CAFFE* Toolkit [18], we reported segmentation results on the testing set.

IV. Results

A. Quantitative performance

To evaluate the performance of the proposed MV-CNN, two widly used methods: level set and graph cut were used for comparison which are provided in the public *Fiji* software [19] and the parameters were optimized through grid searching with this software. As listed in Table I, the proposed MV-CNN outperformed graph cut and level set methods. Moreover, Fig. 2 shows the DSC score distribution of the LIDC-IDRI testing set.

TABLE I

Mean and standard deviation of quantitative results for various segmentation methods. The best performance is indicated in bold font.

	DSC (%)	ASD (mm)	SEN (%)	PPV (%)
Level Set	60.09(16.83)	0.52(0.28)	65.55(21.61)	68.34(25.76)
Graph Cut	69.52(17.32)	0.47(0.31)	81.15(14.29)	65.43(25.18)
MV-CNN	**77.67(15.71)**	**0.24(0.33)**	**83.72(20.71)**	**77.58(15.83)**

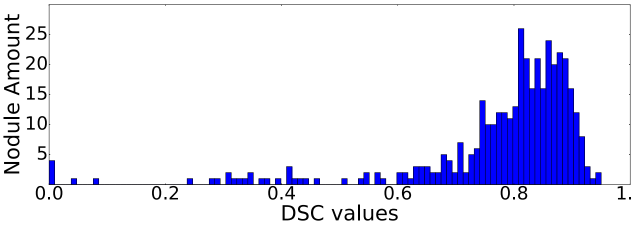

Fig. 2. DSC score distribution of the LIDC-IDRI testing set.

B. Visualization

The segmentation results are visualized to allow the comparison of different approaches. We demonstrate six representative nodules from the LIDC-IDRI testing set (Fig. 3). For isolated solid nodules (L1), both our method and the state-of-the-art methods perform well. However, when examining nodules attached to surrounding tissues, the level set and graph cut methods lead to reduced performance because they are unable to identify nodules from pleura (L2) or vessels (L3). In contrast, the proposed MV-CNN remains robust when segmenting such nodules, which can be probably attributed to the feature learning ability of the MV-CNN in capturing discriminative features from different image views. For cavitary (L4) and calcific (L5) nodules,

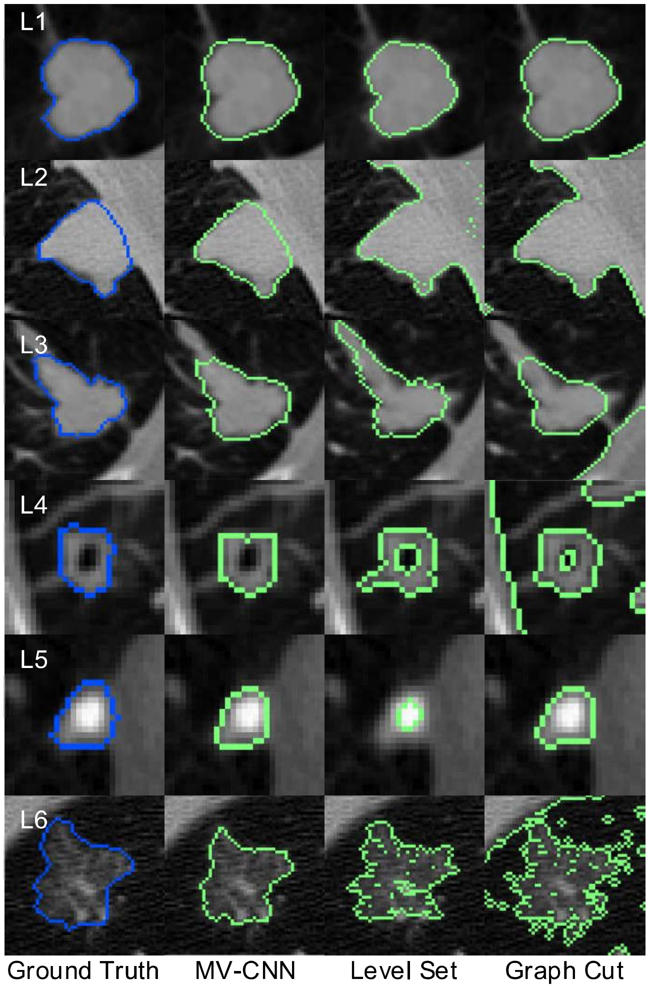

Fig. 3. Segmentation visualization. From left to right: nodule with ground truth, MV-CNN segmentation, level set segmentation, and graph cut segmentation. L1-L6 are nodules of different types from the LIDC-IDRI testing set.

level set and graph cut methods only identify part of the nodules. However, the MV-CNN is able to reserve the complete nodule shape. In addition, ground-glass opacity (GGO) nodules (L6) present another challenge for the level set and graph cut methods because they tend to show over-segmentation due to the low contrast between nodules and normal lung field, whereas the proposed method performs reasonably well in capturing the nodule shape with GGO.

V. Conclusion

In this paper, we presented a deep learning model MV-CNN for lung nodule segmentation, integrating three branches to extract features from three orthogonal image views in CT. An advantage of the proposed model is that it does not involve any nodule shape hypothesis or user-interactive parameter settings. After training on 0.34 million voxel patches, MV-CNN achieved encouraging performance on 393 nodules from the public LIDC-IDRI dataset (DSC = 77.67% and ASD = 0.24). In future work, we plan to train the model with a larger amount of dataset and explore whether the network depth will affect the model performance.

References

[1] T. Kubota, A. K. Jerebko, M. Dewan, M. Salganicoff, and A. Krishnan, "Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models," *Medical Image Analysis*, vol. 15, no. 1, pp. 133–154, 2011.

[2] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient cad system for pulmonary nodule detection in ct imagery," *Medical Image Analysis*, vol. 14, no. 3, pp. 390–406, 2010.

[3] J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic ct scans: a region growing approach," *IEEE transactions on medical imaging*, vol. 27, no. 4, pp. 467–480, 2008.

[4] A. A. Farag, H. E. A. El Munim, J. H. Graham, and A. A. Farag, "A novel approach for lung nodules segmentation in chest ct using level sets," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5202–5213, 2013.

[5] M. Keshani, Z. Azimifar, F. Tajeripour, and R. Boostani, "Lung nodule segmentation and recognition using svm classifier and active contour modeling: a complete intelligent system," *Computers in biology and medicine*, vol. 43, no. 4, pp. 287–300, 2013.

[6] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 588–599.

[7] W. Shen, M. Zhou, F. Yang, D. Dong, C. Yang, Y. Zang, and J. Tian, "Learning from experts: Developing transferable deep features for patient-level lung cancer prediction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 124–131.

[8] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[9] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, 2016.

[10] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, and J. Tian, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognition*, 2016.

[11] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. van Ginneken, "Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of The International Conference on Machine Learning*, 2015, pp. 448–456.

[14] M. Lai, "Deep learning for medical image segmentation," *arXiv preprint arXiv:1505.02000*, 2015.

[15] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *International conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[16] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves *et al.*, "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Medical physics*, vol. 38, no. 2, pp. 915–931, 2011.

[17] Y. Gao, Y. Shao, J. Lian, A. Z. Wang, R. C. Chen, and D. Shen, "Accurate segmentation of ct male pelvic organs via regression-based deformable models and multi-task random forests," *IEEE transactions on medical imaging*, vol. 35, no. 6, pp. 1532–1543, 2016.

[18] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[19] J. Schindelin, I. Arganda-Carreras, E. Frise *et al.*, "Fiji: an open-source platform for biological-image analysis," *Nature methods*, vol. 9, no. 7, pp. 676–682, 2012.