LUNG NODULE DETECTION IN CT USING 3D CONVOLUTIONAL NEURAL NETWORKS

Xiaojie Huang, Junjie Shan*, and Vivek Vaidya*

GE Global Research, Niskayuna, NY

ABSTRACT

We propose a new computer-aided detection system that uses 3D convolutional neural networks (CNN) for detecting lung nodules in low dose computed tomography. The system leverages both a priori knowledge about lung nodules and confounding anatomical structures and data-driven machine-learned features and classifier. Specifically, we generate nodule candidates using a local geometric-model-based filter and further reduce the structure variability by estimating the local orientation. The nodule candidates in the form of 3D cubes are fed into a deep 3D convolutional neural network that is trained to differentiate nodule and non-nodule inputs. We use data augmentation techniques to generate a large number of training examples and apply regularization to avoid overfitting. On a set of 99 CT scans, the proposed system achieved state-of-the-art performance and significantly outperformed a similar hybrid system that uses conventional shallow learning. The experimental results showed benefits of using a priori models to reduce the problem space for data-driven machine learning of complex deep neural networks. The results also showed the advantages of 3D CNN over 2D CNN in volumetric medical image analysis.

***Index Terms*—** Lung nodule, computer-aided detection, CT, deep learning, 3D convolutional neural networks

1. INTRODUCTION

Lung cancer is the second largest cause of cancer deaths in the US. The survival rate of lung cancer can be substantially improved if it is detected and treated in the early stage. Low dose computed tomography (CT) chest scans have been shown effective in screening lung cancer [1], however, reading the large CT volumes and detecting lung nodules accurately and repeatably demand enormous amount of radiologist's effort. An accurate computer-aided detection (CAD) system is essential for an efficient and cost-effective lung cancer screening workflow.

To this end, a variety of approaches have been proposed for lung nodule detection in CT images. A typical nodule detection system starts by generating candidate nodule objects, and then classifies them based on certain predetermined features designed to differentiate true nodules from non-nodule structures (e.g., [2–7]). Many of these features are designed to model the intensity distribution and geometry of nodules in CT scans. One major limitation of these features is the requirement of nodule candidate segmentation, which may not be accurate. One class of successful CAD systems use geometrical-model-based features to classify each voxel for its membership in a nodule [2,8,9]. While these systems have proven effective, a considerable number of nodules remain undetected at a high specificity operating point.

Machine learning has proven very effective in discovering discriminative features and boosting the accuracy of many visual recognition systems. To utilize the full power of machine learning, we usually need very big training data which may not be realistic in the medical imaging domain in the near future. Given such constraints, a hybrid system (e.g., [10]) that leverages as much available a priori domain knowledge as possible to reduce the problem space and regularize the data-driven machine learning process turns out to be advantageous. Recently, deep convolutional neural networks (CNN) have been shown to be very successful in solving a variety of visual recognition problems and keep breaking the performance records in respective challenges (e.g., [11, 12]). CNN has also been applied to several medical image analysis problems (e.g., [13–21]). While most medical images are 3D tensors, conventional CNN methods are based on 2D kernels. Existing deep learning methods for medical image analysis typically convert 3D data into 2D multi-channel representations, such as the tri-planar representation [13], and then feed them into 2D CNN. While 2D CNN has been shown useful in solving these problems, it intrinsically loses the 3D context of the original image, which limits the performance of the overall system.

In this paper, we propose a new CAD system that uses 3D CNN techniques for detecting lung nodules in low dose CT. The system combines a priori intensity and geometrical knowledge about lung nodules and confounding anatomical structures with data-driven machine learning of features and classifiers. Specifically, the system has two steps: 1) generating nodule candidates using a local geometric-model-based filter and 2) classifying candidates using 3D CNN. A priori knowledge is not only used for candidate generation, but also for reducing the structure variability of the input to 3D CNN through candidate orientation estimation using intensity-weighted principal component analysis (PCA). This

*These authors contributed equally to this work.

978-1-5090-1172-8/17/$31.00 ©2017 IEEE

design reduces the problem complexity and the difficulty in training deep 3D CNN with limited data. We propose a deep 3D CNN architecture specifically for this problem inspired by the recent successful deep CNN architectures in visual recognition (e.g., [11, 12]). We use data augmentation techniques to generate sufficiently large training data and use regularization techniques, such as dropout, to avoid overfitting. We also study the effects of principal direction alignment prior to CNN, the advantages of 3D CNN over 2D CNN, and the effects of dense evaluation in the prediction step. To our best knowledge, the only existing 3D CNN CAD system for lung nodule is [15]. Our system is different in the following aspects: 1) candidate generation with curvature-based filter and bayesian models, 2) principal direction alignment of candidate cubes, 3) data augmentation, 4) isotropic 3D convolutional kernels that work across any image spacing, 5) network architecture, and 6) much better performance as shown in experimental results.

2. METHODS

2.1. Model-based Candidate Generation

Geometric model-based metrics computed locally from CT images have been shown to be effective in pulmonary nodule detection [8, 9]. In particular, studies suggest that curvature-based metrics lead to reliable pulmonary nodule detection [8]. Explicit local shape modeling of nodules, vessels, and vessel junctions in a Bayesian framework has been proposed and shown to be computationally efficient and to outperform other model-based methods [9]. We use this approach to efficiently generate nodule candidates while filtering out large areas free of true nodules. Model parameters are set for the system to operate with high sensitivity to retain most of the true nodules. Each cluster of highlighted voxels that is $\geq$ 3 mm is treated as a nodule candidate for the next step. This step uses a priori knowledge to greatly reduces the computational cost and problem complexity for the following 3D CNN classification step.

In addition, another important a priori knowledge that can be utilized to largely simplify the problem is that confounding anatomical structures and even nodules have large orientation-induced variability. While such variability may be modeled and the machine (particularly with neural networks) may be able to learn the orientation-invariant features, this becomes a challenge when the neural network and training data are not very big. We use this a priori knowledge by estimating the orientation of the structure, encoded by a rotation matrix, at a candidate voxel using an intensity-weighted PCA method [10]. With the rotation matrix, a region of interest (ROI) in the form of a $32 \times 32 \times 32$ (30mm$\times$30mm$\times$30mm) cube is extracted at the candidate voxel with a 3D sampling grid that is aligned to the principal directions. The intensity is clipped to the range of [-1000 HU, 1000 HU] and rescaled to the range of [0, 1].

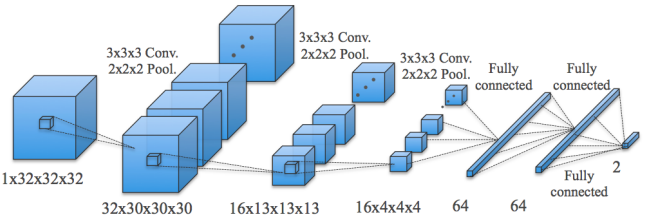

Fig. 1: The neural network architecture of our CAD system.

2.2. 3D CNN Architectures

There have been a series of very successful CNN models applied to visual recognition in 2D natural images with very large training datasets consisting of millions of different images (e.g., [11, 12]). We are inspired by these models in designing our 3D CNN architecture. Given the nature of problem, the cube dimension, the size of our training data, and available GPU computational power, we use a relatively simple, but still deep compared to conventional systems, 3D CNN architecture shown in Fig. 1. We use a $32\times32\times32$ input layer. We use three convolutional layers with 32, 16, and 16 small $3 \times 3 \times 3$ kernels respectively. Each convolutional layer is followed by a max-pooling layer with overlapping $2 \times 2 \times 2$ windows. We use three fully connected layers with 64, 64, and 2 neurons respectively. Rectified linear units (ReLU) are used in each convolutional and fully connected layers. To regularize the model, we apply ℓ_2 weight regularization with a hyper-parameter of 0.0005 to each layer, and dropout with a rate of 0.5 to the first two fully connected layers. Our model has about 34K parameters.

2.3. Training

The following design choices we made in generating the training examples and augmenting the training data are crucial for the performance of our CAD system. Given the huge number of model parameters and fairly limited amount of training data, we perform data augmentation to achieve a sufficiently large set of training examples. We treat each voxel in the ground truth nodules as a nodule example. The benefits of this choice are two-fold. Firstly, this largely increases the number of nodule training examples. Secondly, this forces the neural network to learn translation-invariant capability, which is very important as the candidates generated by the first step are not necessarily centered at the local anatomical structures. To further augment the nodule training examples, as illustrated in Fig. 2, we randomly perturb the estimated principal direction of the candidate structure by up to 18 degrees along each of the axises and randomly flip the first principal direction. This design further augments the nodule examples by 8 times and introduces robustness to errors in PCA-based orientation es-

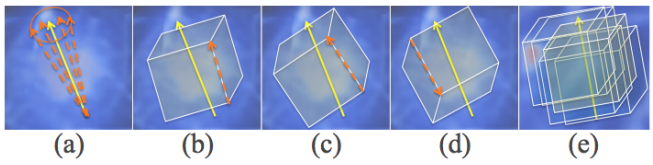

Fig. 2: (a-d): Extraction of positive cubes with principal direction alignment and random perturbation of principal direction. (e): Extraction of negative cubes with principal direction alignment and dense sampling. The yellow arrows point to the local principal directions, and dashed orange arrows represent perturbed or flipped directions.

timation. There are a very large number of non-nodule training examples naturally generated by the high-sensitivity low-specificity candidate generation step. For each candidate clusters that are not true nodules, as illustrated in Fig. 2, we apply a sampling grid with certain spacing to sample a sufficiently large set of non-nodule training examples that are also aligned to the local principal direction.

With data augmentation, we generate thousands of examples from each training scan. The neural network is trained using a stochastic gradient descent algorithm with an adaptive learning rate scheme–Adadelta [22]. We initialize the weights using a normalized initialization scheme by Glorot and Bengio [23]. We use a mini-batch size of 256 examples. We train the network up to 10 epochs with early termination and select the best model based on the loss on a validation dataset.

2.4. Prediction

As observed in [12], densely evaluating and pooling the predictions for different versions of the same object improves recognition performance. To classify each candidate cluster, multiple cubes were densely sampled inside the cluster and fed into the trained 3D CNN classifier and the average of multiple probabilistic predictions is treated as the predicted probability of the cluster being a lung nodule. Dense evaluation is expected to reduce the variance of classification and boost the final detection accuracy of our CAD system. As in the training phase, each candidate cube is aligned to the principal direction.

3. EXPERIMENTAL RESULTS

3.1. Data

We validated the proposed CAD system using scans from the Lung Image Database Consortium (LIDC) [24]. In accordance to the recommended National Lung Screening Trial (NLST) [1] screening protocols, we used 99 scans with $\leq$ 1.25 mm slice thickness. The in-plane voxel spacing within each slice ranged from 0.55 mm $\times$ 0.55 mm to 0.96 mm $\times$ 0.96 mm. Four different experts contoured nodules that were greater than 3 mm in each scan. If multiple experts contoured the same nodule, the contours were merged together based on a consensus rule to obtain the ground-truth nodule detections. If less than half of the experts labeled a certain region as inside a nodule, we excluded this region from the ground-truth. In this preliminary study, we focused on developing a better candidate classifier rather than improving candidate generation. We excluded Ground Glass Opacity (GGO) and juxta-pleural nodules attached to the lung boundary, as the employed candidate generator [9] was not developed to handle these nodules.

3.2. Experiments

We implemented our neural network using Keras and trained and tested our system on a HP 840 workstation with a GeForce GTX 1080 GPU. We performed 10-fold cross-validation of the proposed method. For each fold, we generated about 600K examples from the training scans. It took about one hour to train the network for each fold. We compared the proposed method to several alternative designs of a lung nodule CAD system: 1) pure geometric modeling [9], 2) conventional random forests with predefined features [10], 3) 2D CNN with tri-planar representation, 4) 3D CNN without principal direction alignment, and 5) 3D CNN without dense evaluation. Figure 3 compares the FROC curves obtained using these alternative lung nodule CAD methods.

The proposed method (yellow curve) achieved a high detection sensitivity of 90% at 5 false positives per scan. Compared to the reported performance of recent published studies on different datasets, this result was comparable to the numbers reported in the recent multi-view convolutional network study [14] and was much better (5 vs 33 false positives per scan at 90% sensitivity) than the recently published 3D CNN lung nodule detection work [15]. Comparing our result to the pure geometric-model-based system [9] (blue curve), we saw that the proposed neural network was very effective in eliminating false positives generated by the conventional model-based detector. The lung nodule CAD system (the baseline), published in [10], had a similar hybrid design of combining model-based candidate generation with learning-based false positive reduction where it used predefined contextual features and random forests. At 90% sensitivity, the proposed method reduced the number of false positives per scan of the baseline system (green curve) from 35 to 5. This significant improvement showed the benefits of employing a deep neural network framework to learn features for the specific lung nodule detection purpose.

We validated the benefits of the major design choices we made in this work. While 2D CNN with multi-slice representation has been popular in medical image analysis applications, it unavoidably leads to considerable loss of information. Here, we compared the proposed system to an alternative 2D CNN method with the popular tri-planar 2D representation. In this 2D CNN implementation, we kept everything the same

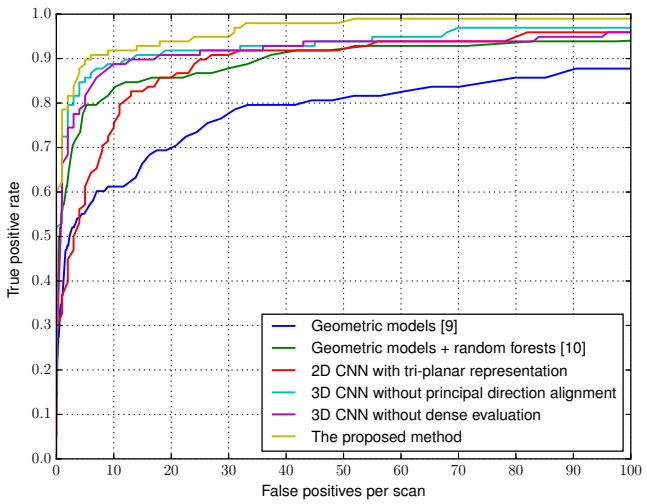

Fig. 3: FROC curves achieved by the six detection systems.

as in the proposed method expect that a tri-planar 3-channel 2D representation was used. The 2D CNN architecture also had the same structure as the proposed 3D CNN architecture expect that 3D convolutional kernels were replaced with 2D convolutional kernels. As the red curve showed, 2D CNN did not even perform as well as the baseline random forest method. This was likely due to the loss of 3D context when the 3D volume of interest was reduced to a few 2D slices. The proposed 3D approach significantly outperformed the 2D approach. These results showed the importance of 3D context for differentiating nodules from non-nodule structures.

We tested the system without aligning the principal direction of local structures to study its effect. The result (cyan curve) showed that without such a priori model the system had a noticeable drop in detection performance. This difference was largely attributed to that principal direction alignment helped reduce the problem complexity and generalization errors of the neural network. In theory, the neural network should be able to learn orientation-invariant features. However, that requires larger, more complex neural networks and more training examples which may be generated by further rotational data augmentation.

We also tested the system without dense evaluation to study the effect of dense evaluation. We kept everything the same except that we evaluated each cluster only once by generating a candidate cube at the center of the cluster. The result (purple curve) showed that dense evaluation outperformed single evaluation substantially. This difference can be explained by that dense evaluation improved robustness to errors made at candidate generation and cube extraction steps.

4. CONCLUSION AND DISCUSSION

This paper presented a new CAD system using 3D CNN for lung nodule detection in low dose CT. The system combined the strength of a priori knowledge about lung nodules and confounding anatomical structures and data-driven deep neural network learning. It consisted of two steps: 1) nodule candidate generation using a local geometric-model-based filter and 2) candidate classification using a specifically configured deep 3D CNN. A priori knowledge was not only used for candidate generation, but also for reducing the variability of the input to 3D CNN through principal direction alignment of candidate cubes based on intensity-weighted principal component analysis (PCA). We augmented training data by voxel-wise sampling and random rotation, which greatly enriched our training examples and introduced translational invariance and robustness to orientation estimation errors. In classification, the system performed dense grid-sampled evaluation of candidate clusters to reduce classification variance.

These designs greatly reduced the problem complexity, enabled us to train an effective deep 3D CNN with fairly limited amount of data, and boosted the performance of the proposed CAD system. In experiment, our system achieved state-of-the-art performance and significantly outperformed the baseline system using conventional shallow learning. The results also showed the advantages of 3D CNN over 2D CNN, and the advantages of candidate principal direction alignment and dense evaluation. Our results demonstrated that with limited data it is promising to train large, deep 3D CNN to significantly improve the performance of CAD system with the help of a priori knowledge, data augmentation, and regularization.

The main limitations of this work are: 1) GGO and juxtapleural nodules were not addressed; 2) the experiment was limited to cross validation. In the future, we will test the proposed method on separate test data. We will explore more versatile candidate generators for developing a more complete system that applies to all nodule types. We will also explore training more complex neural networks using larger datasets.

5. REFERENCES

[1] D.R. Aberle, A.M. Adams, et al., "Reduced lung-cancer mortality with low-dose computed tomographic screening.," *The New England journal of medicine*, vol. 365, no. 5, pp. 395–409, 2011.

[2] K. Murphy, B. van Ginneken, et al., "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification.," *Medical image analysis*, vol. 13, no. 5, pp. 757–70, Oct. 2009.

[3] T.W. Way, B. Sahiner, et al., "Computer-aided diagnosis of pulmonary nodules on ct scans: improvement of classification performance with nodule surface features," *Medical physics*, vol. 36, no. 7, pp. 3086–3098, 2009.

[4] B. van Ginneken, S.G. Armato III, et al., "Comparing

and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the anode09 study," *Medical image analysis*, vol. 14, no. 6, pp. 707–722, 2010.

[5] T. Messay, R.C. Hardie, et al., "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery.," *Medical image analysis*, vol. 14, no. 3, pp. 390–406, 2010.

[6] M. Liu, L. Lu, et al., "Sparse classification for computer aided diagnosis using learned dictionaries.," in *MICCAI*, vol. 14, pp. 41–8. Jan. 2011.

[7] M. Tan, R. Deklerck, et al., "A novel computer-aided lung nodule detection system for ct images," *Medical physics*, vol. 38, no. 10, pp. 5630–5645, 2011.

[8] P.R.S. Mendonça, R. Bhotika, et al., "Model-based analysis of local shape for lesion detection in ct scans," in *MICCAI*, pp. 688–695. 2005.

[9] P.R.S. Mendonça, R. Bhotika, et al., "Lung nodule detection via Bayesian voxel labeling.," in *IPMI*, vol. 20, pp. 134–46. Jan. 2007.

[10] J. Bai, X. Huang, et al., "Learning orientation invariant contextual features for nodule detection in lung CT scans," in *ISBI*, 2015, pp. 1135–1138.

[11] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] H. C. Shin, H. R. Roth, et al., "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE TMI*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[14] A. A. A. Setio, F. Ciompi, et al., "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, May 2016.

[15] R. Anirudh, J.J. Thiagarajan, et al., "Lung nodule detection using 3d convolutional neural networks trained on weakly labeled data," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2016, pp. 978532–978532.

[16] N. Tajbakhsh, J. Y. Shin, et al., "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[17] K. Kamnitsas, L. Chen, et al., "Multi-scale 3d convolutional neural networks for lesion segmentation in brain mri," *Ischemic Stroke Lesion Segmentation*, p. 13, 2015.

[18] H.R. Roth, L. Lu, et al., "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 556–564.

[19] M. Havaei, A. Davy, et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, 2016.

[20] B. van Ginneken, A. A. A. Setio, et al., "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, April 2015, pp. 286–289.

[21] A. Teramoto, H. Fujita, et al., "Automated detection of pulmonary nodules in pet/ct images: Ensemble false-positive reduction using a convolutional neural network technique," *Medical Physics*, vol. 43, no. 6, pp. 2821–2827, 2016.

[22] M.D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.

[23] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks.," in *Aistats*, 2010, vol. 9, pp. 249–256.

[24] S.G. Armato III, G. McLennan, 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.