HHS Public Access

Author manuscript

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2017 August 23.

Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2017 ; 10134: . doi:10.1117/12.2255795.

3D Convolutional Neural Network for Automatic Detection of Lung Nodules in Chest CT

Sardar Hamidian[a], **Berkman Sahiner**[b], **Nicholas Petrick**[b], and **Aria Pezeshk**[b]

[a]Department of Computer Science, George Washington University, Washington, DC

[b]Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD

Abstract

Deep convolutional neural networks (CNNs) form the backbone of many state-of-the-art computer vision systems for classification and segmentation of 2D images. The same principles and architectures can be extended to three dimensions to obtain 3D CNNs that are suitable for volumetric data such as CT scans. In this work, we train a 3D CNN for automatic detection of pulmonary nodules in chest CT images using volumes of interest extracted from the LIDC dataset. We then convert the 3D CNN which has a fixed field of view to a 3D fully convolutional network (FCN) which can generate the score map for the entire volume efficiently in a single pass. Compared to the sliding window approach for applying a CNN across the entire input volume, the FCN leads to a nearly 800-fold speed-up, and thereby fast generation of output scores for a single case. This screening FCN is used to generate difficult negative examples that are used to train a new discriminant CNN. The overall system consists of the screening FCN for fast generation of candidate regions of interest, followed by the discrimination CNN.

Keywords

Deep learning; convolutional neural networks; computer-aided diagnosis; chest CT

1. INTRODUCTION

Convolutional neural networks (CNNs) form the backbone of many state-of-the-art computer vision systems for detection and segmentation of objects in 2D images.[1–5] These networks can be stacked to form deep networks that can automatically learn useful representations directly from the data without the need for manual feature engineering, or extensive pre-processing pipelines. As with other deep learning architectures, the performance of deep CNNs is largely dependent on the availability of large datasets. Otherwise, the large number of parameters in these systems cannot be properly trained. While most systems reported in the literature have focused on 2D images, the same principles used in 2D CNNs can be extended to 3D by using 3D convolution kernels and 3D pooling in order to obtain systems that can be applied to volumetric data such as computed

Send correspondence to Aria Pezeshk. aria.pezeshk@fda.hhs.gov, WO62-4110, 10903 New Hampshire Ave., Silver Spring, MD, 20993.

tomography (CT) images. Introduction of 3D convolution kernels increases the number of parameters in the architecture, the training time, as well as the need for more data. Since medical image datasets are often small, training of 3D CNNs on data from volumetric modalities is not always straightforward.

Traditional systems for automatic detection of pulmonary nodules often consist of long complex pipelines that apply a sequence of algorithms such as lung region detection, candidate object selection, feature extraction, false positive reduction, and classification.[6–8] Each of these procedures take a long time to develop and may include additional subcomponents with individual sets of parameters controlling their performance, thereby further complicating the design of such systems.

Deep learning approaches to lung nodule detection in chest CT have so far relied on 2D views of nodules and CT slices.[9] In this paper we report a computer-aided detection (CADe) system based on 3D CNNs currently under development for lung CT that circumvents the aforementioned pre-processing pipelines. The system works in two steps: screening and discrimination. First, we train a 3D CNN to classify volumes of interest (VOIs) extracted from the Lung Image Database Consortium (LIDC) dataset[10] as containing a nodule or not. The training set for this CNN consists of 3D patches containing nodules (positive samples), and randomly selected 3D patches without nodules (negative samples). The CNN is then converted to a 3D fully convolutional neural network (FCN)[11] that can process an entire CT volume efficiently in a single pass to produce a score map for the entire volume. We use this FCN as a screening tool to generate a set of hard negatives (negative samples that are difficult to distinguish from positive samples), and train a second more specialized CNN. The second CNN is in turn converted into a new FCN, which can be used for screening each CT exam and to generate candidate regions of interest. A third discrimination CNN can then be trained based on false positives generated from the screening stage. This CNN is applied to the 3D patches found during screening in order to reduce the number of false positives and to classify each candidate region of interest as containing a nodule or not.

The two-step design of our system that uses FCNs for both screening and discrimination steps has two major advantages compared to the traditional CADe architectures described earlier. First, this system avoids the complex pre-processing pipelines and can be directly applied to entire CT volumes. Second, the FCN screening stage allows for fast processing of each volume, so that the discrimination CNN stage can be applied only to a small subset of regions. As the FCN is very fast, the overall processing time is significantly reduced. The contributions of this paper are twofold. First, we have extended the highly successful fully convolutional networks to 3D and developed an efficient system for detection of pulmonary nodules. Second, to our knowledge this is the first time that either 3D CNNs or 3D FCNs are applied for detection of pulmonary nodules.

2. METHODS

2.1 Screening Stage

The screening stage is a common component of a CAD system. This step reduces the size of the initial search space and identifies a subset of most likely candidates that should be

analyzed further. The screening CNN in our system is first trained using 3D convolution kernels to classify 3D patches extracted from each CT case. The training set for this CNN was constructed as follows: the positive samples consisted of 3D VOIs that were centered over each nodule in the dataset, while the negative samples were selected randomly by extracting VOIs with the same size as the positive samples from a random location within the CT scan (including both the inside and outside of the lungs). The negative patches were selected such that they did not have any overlap with a nodule. The number of negative samples extracted in such a way can be almost as large as desired since most of the area within a chest CT is nodule-free. The number of positive samples on the other hand is limited. To improve the invariance of the system to trivial differences in appearance of nodules, and in order to reduce the aforementioned class imbalance problem, the positive samples are augmented by including ipped and rotated copies of each extracted positive patch in the training set.

Our deep network consists of three successive layers of convolution and max-pooling, followed by a fully connected layer, and a final fully connected softmax layer as shown in Table 1. The convolution layers were zero-padded so that any reduction in size would be due to the max-pooling layers. We used Nesterov momentum[12] and dropout[13] to improve the convergence rate and generalization performance of the network.

CNNs have a fixed field of view and can only process images or volumes of fixed size (for simplicity, we will base the following discussion on images but the same principle holds for volumetric data). For images that are larger than the expected input size, a CNN can be applied in a sliding fashion across the entire image to obtain classification scores over the whole image. This brute force way of applying CNNs to whole images is very slow even in two dimensions. One way to overcome this shortcoming of CNNs is to convert the fully connected layers into convolutional layers in order to obtain fully convolutional networks (FCNs). This conversion involves careful rearrangement of learned weights and biases in the fully connected layers into equivalent convolution kernels, and results in the same output scores as the traditional CNN. Once this conversion is done, the entire 2D image can be passed through the FCN in a single pass to obtain a score map across the whole image. As convolutions are highly optimized, each image can be processed much faster compared to the sliding window approach. In three dimensions (as in e.g. CT exams), each volume contains numerous sub-volumes over which the sliding CNN window should pass, thereby rendering the sliding window approach ineffective for volumetric data. Here we convert our 3D CNN into a 3D FCN. While this conversion theoretically makes it possible to process an entire CT volume in a single pass, the large size of a whole CT case means that the volume has to be split into several smaller sub-volumes due to memory limitations in existing graphical processing units (GPUs). Nevertheless, this process is still far more computationally efficient compared to the brute force sliding window method.

The random selection of negative patches is a crude way of selecting the negative training examples, resulting in many negative patches that contain areas of the chest CT that are too simple to classify. Training the initial CNN on this data results in the generation of a large number of false positives when the corresponding FCN is applied to the whole CT exams (compared to when the FCN is applied to training set patches). In order to improve the

performance of the screening stage, the initial training set is augmented by adding false positives found from applying the initial FCN to the whole CT exams in training. These false positives are essentially a set of hard negatives, i.e. negative samples that are difficult to distinguish from positive samples, that help the classifier better distinguish nodules from areas in the lung with similar appearance. A second CNN is subsequently trained on the new training set, and converted to a new FCN that is used for screening the CT cases.

2.2 Discrimination stage

The screening stage described in the previous section still generates a large number of false positives. The purpose of the discrimination stage is to reduce this number so that the clinician is provided with an output that has high sensitivity for detection of nodules while producing a manageable number of false positives per case.

The output of the the final screening FCN is a score volume, where the intensity of each voxel is the probability of that voxel being a nodule. A threshold can be applied to these probabilities to obtain a smaller number of candidate segments. The false positives with high probabilities in this score map are negative patches that have a similar appearance to actual nodules. We therefore construct a third training set consisting only of false positives from the screening stage. A new CNN is then trained using this data with the same architecture as that shown in Table 1, and applied only to the screened candidate patches for final discrimination.

2.3 Data used in the study

We used a subset of 509 cases from (LIDC) dataset with slice thickness between 1.5mm and 3mm to train our models, with an additional 25 cases used for testing our models. The location of each nodule in the LIDC dataset is marked by between one and four radiologists, and a segmentation is provided for each detected ≥3mm nodule by the radiologist. Due to the inherent variability in radiologists' assessment of nodules, many of the nodules in the dataset were not detected by one or more radiologists. We therefore used the following set of rules to unify the decisions: we assigned ≥3mm nodules that were detected by at least two radiologists and that had more than 20% overlap in their segmentations as a true positive (percent overlap was defined as the intersection of volumes delineated by individual radiologists divided by the union of the volumes from all radiologists for that nodule). Using the scheme described in the previous section, positive patches (i.e. those containing nodules) and negative patches were extracted from the training cases. The training and test sets constructed in this fashion contained 833 and 104 nodules, respectively. The positive samples in training were further augmented according to the procedure described in the previous section.

3. RESULTS

Figure 1 shows examples of the score map generated by the screening FCN corresponding to a single slice of several different CT exams. It can be seen that the FCN can detect the areas containing nodules with high probability, but at the same time the score map also contains other points with mid to high probability. These points correspond to false positive (FP)

areas that need to be suppressed by the discrimination stage. Figure 2a shows the free-response operating characteristic (FROC) curve of the screening FCN. The model reaches 80% sensitivity at 22.4 FPs per case, and 95% sensitivity at 563 FPs per case. Any nodules that are not detected by screening will be missed by the discrimination stage. We can therefore reduce the number of FPs presented to the discrimination stage by removing low probability FPs at the level that provides 95% sensitivity.

The discrimination CNN is only evaluated at screening candidate points that pass the threshold discussed earlier. Figure 2b shows the FROC curve for the discrimination stage. This model reaches a sensitivity of 80% at 15.28 FPs per case.

We also applied the original screening CNN in the sliding window fashion across the entire volume of multiple CT cases, and compared the elapsed time to do so with the time for processing the same set of cases using the FCN. On average, our tests showed an 800-fold improvement in time to process each CT exam for the FCN compared to the brute-force sliding window method for the CNN. For instance, for a 120 slice CT case this translates to FCN generating a score map for the entire CT exam in only roughly 5 seconds compared to nearly 4200 seconds for the sliding window approach using the original CNN.

4. CONCLUSION

We are developing a CADe system for lung nodule detection using 3D convolutional neural networks that works in two steps: screening and discrimination. A 3D fully convolutional network is first used for efficient processing of the entire CT volume and to generate candidate regions of interest. A more specialized CNN subsequently classifies each candidate region as nodule or background. This screening architecture leads to a 800-fold speed-up compared to using the brute-force method of sliding the 3D CNN across the volume to obtain the classification scores for the whole CT exam.

Pulmonary nodules exhibit a wide range of sizes and shapes. Additional work is under way to improve the performance of the CAD system, and to incorporate a multi-scale approach so that nodules with varying sizes can be detected at similar sensitivities compared to one another.

Acknowledgments

S.H. was supported by an appointment to the Research Participation Program at the Center for Devices and Radiological Health administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

References

1. Krizhevsky, A., Sutskever, I., Hinton, GE. ImageNet classification with deep convolutional neural networks. In: Bartlett, P.Pereira, F.Burges, C.Bottou, L., Weinberger, K., editors. Advances in Neural Information Processing Systems (NIPS). Curran Associates, Inc; Red Hook, NY: 2012. p. 1106-1114.
2. Albarqouni S, Baur C, Achilles F, Belagiannis V, Demirci S, Navab N. Aggnet: Deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Trans. Med. Imaging. 2016; 35(5):1313–1321. [PubMed: 26891484]

3. Dai, J., He, K., Sun, J. Convolutional feature masking for joint object and stuff segmentation; IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015. p. 3992-4000.
4. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556. 2014
5. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. May.2015 521:436–444. [PubMed: 26017442]
6. Cascio D, Magro R, Fauci F, Iacomi M, Raso G. Automatic detection of lung nodules in ct datasets based on stable 3d mass-spring models. Comput. Biol. Med. Nov.2012 42:1098–1109. [PubMed: 23020972]
7. Ge Z, Sahiner B, Chan H-P, Hadjiiski LM, Cascade PN, Bogot N, Kazerooni EA, Wei J, Zhou C. Computer-aided detection of lung nodules: False positive reduction using a 3D gradient field method and 3D ellipsoid fitting. Medical Physics. 2005; 32(8):2443–2454. [PubMed: 16193773]
8. Guo W, Li Q. High performance lung nodule detection schemes in CT using local and global information. Medical Physics. 2012; 39(8):5157–5168. [PubMed: 22894441]
9. Setio AAA, Ciompi F, Litjens GJS, Gerke PK, Jacobs C, van Riel SJ, Wille MMW, Naqibullah M, Sánchez CI, van Ginneken B. Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. IEEE Trans. Med. Imaging. 2016; 35(5):1160–1169. [PubMed: 26955024]
10. Armato SG III, 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. 2011; 38(2):915–931. [PubMed: 21452728]
11. Long, J., Shelhamer, E., Darrell, T. Fully convolutional networks for semantic segmentation; IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015. p. 3431-3440.
12. Sutskever, I., Martens, J., Dahl, G., Hinton, G. On the importance of initialization and momentum in deep learning. In: Dasgupta, S., Mcallester, D., editors. Proceedings of the 30th International Conference on Machine Learning (ICML-13); JMLR Workshop and Conference Proceedings; May. 2013 p. 1139-1147.
13. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. Jan.2014 15:1929–1958.

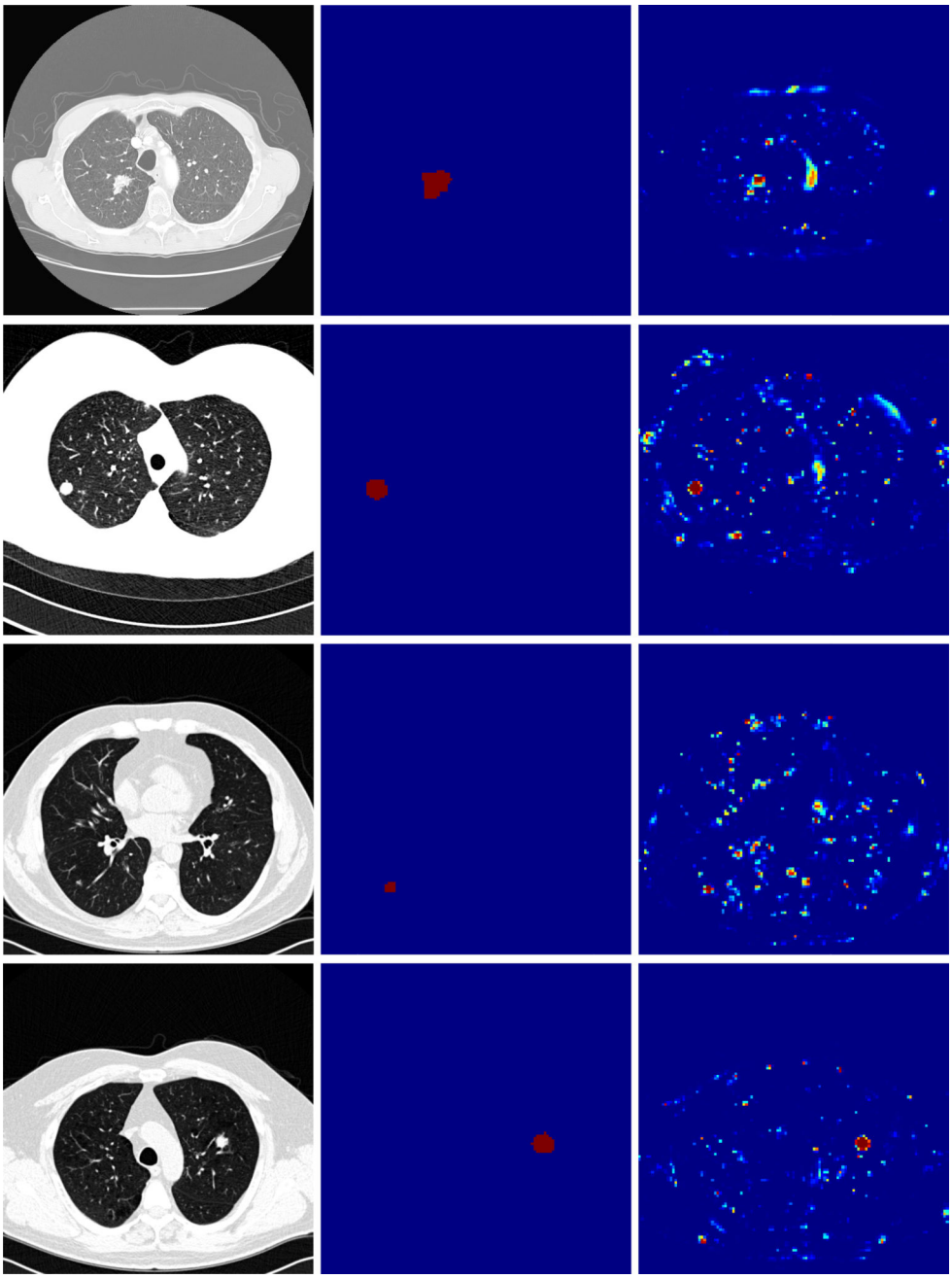

Figure 1.

Applying the screening FCN to whole CT exams. In each row from left to right: one slice of a CT exam containing a nodule, corresponding nodule segmentation from radiologists, corresponding score map generated by the screening FCN. Blue to red indicate low to high probabilities of a pixel being a nodule, respectively.

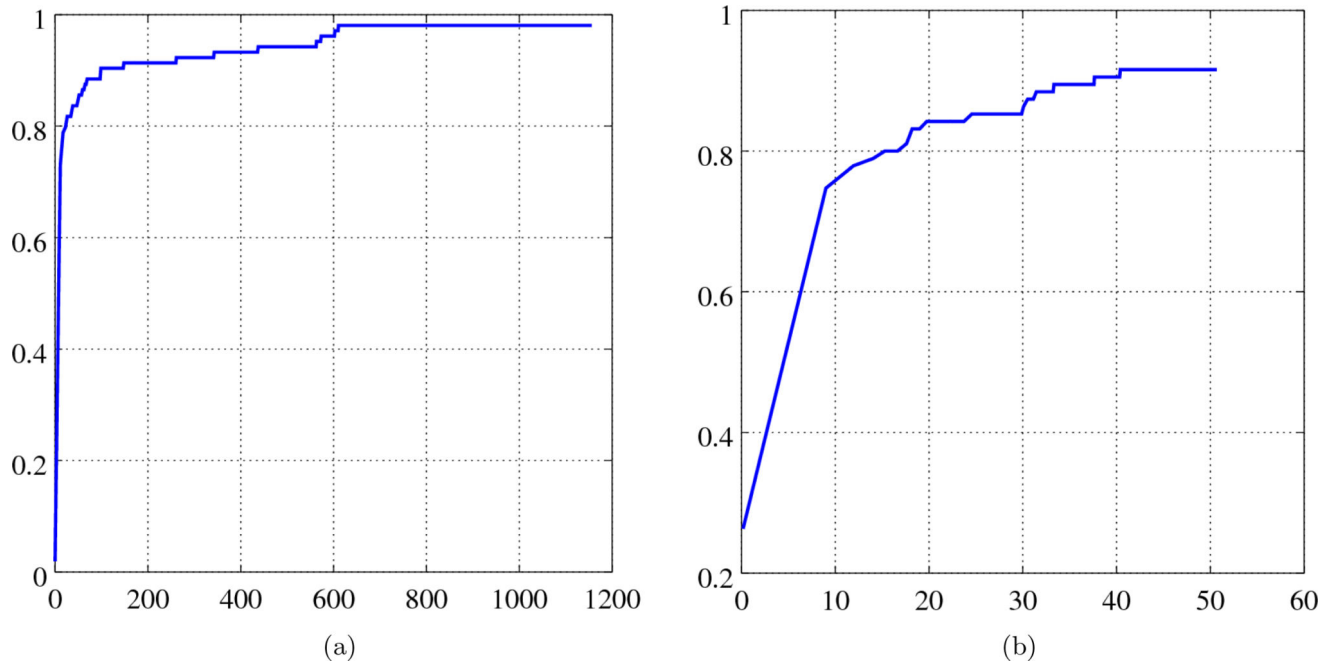

Figure 2.

(a) FROC curve for the screening FCN. (b) FROC curve for the discrimination CNN. Note that the x axis limits are different in the two curves.

Table 1

Architecture of the 3D screening CNN; C, M, and FC denote convolution, max-pooling, and fully connected layers, respectively.

Layer	Kernel Size	Stride	Output Size	Feature Volumes
Input	—	—	36 × 36 × 8	—
C1	5 × 5 × 3	1	36 × 36 × 8	32
M1	2 × 2 × 1	1	18 × 18 × 8	32
C2	5 × 5 × 3	1	18 × 18 × 8	32
M2	2 × 2 × 2	1	9 × 9 × 4	32
C3	3 × 3 × 1	1	9 × 9 × 4	32
FC1	—	1	1 × 1 × 1	900
FC2	—	1	1 × 1 × 1	2