Computer-Aided Pulmonary Embolism Detection Using a Novel Vessel-Aligned Multi-planar Image Representation and Convolutional Neural Networks

Nima Tajbakhsh[1], Michael B. Gotway[2], and Jianming Liang[1]

[1] Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ, USA
{Nima.Tajbakhsh,Jianming.Liang}@asu.edu
[2] Department of Radiology, Mayo Clinic, Scottsdale, AZ, USA
Gotway.Michael@mayo.edu

Abstract. Computer-aided detection (CAD) can play a major role in diagnosing pulmonary embolism (PE) at CT pulmonary angiography (CTPA). However, despite their demonstrated utility, to achieve a clinically acceptable sensitivity, existing PE CAD systems generate a high number of false positives, imposing extra burdens on radiologists to adjudicate these superfluous CAD findings. In this study, we investigate the feasibility of convolutional neural networks (CNNs) as an effective mechanism for eliminating false positives. A critical issue in successfully utilizing CNNs for detecting an object in 3D images is to develop a "right" image representation for the object. Toward this end, we have developed a vessel-aligned multi-planar image representation of emboli. Our image representation offers three advantages: (1) efficiency and compactness—concisely summarizing the 3D contextual information around an embolus in only 2 image channels, (2) consistency—automatically aligning the embolus in the 2-channel images according to the orientation of the affected vessel, and (3) expandability—naturally supporting data augmentation for training CNNs. We have evaluated our CAD approach using 121 CTPA datasets with a total of 326 emboli, achieving a sensitivity of 83% at 2 false positives per volume. This performance is superior to the best performing CAD system in the literature, which achieves a sensitivity of 71% at the same level of false positives. We have further evaluated our system using the entire 20 CTPA test datasets from the PE challenge. Our system outperforms the winning system from the challenge at 0mm localization error but is outperformed by it at 2mm and 5mm localization errors. In our view, the performance at 0mm localization error is more important than those at 2mm and 5mm localization errors.

Keywords: Computer-aided detection, pulmonary embolism, convolutional neural networks, vessel-aligned image representation.

1 Introduction

Pulmonary embolism (PE) is a thrombus, occasionally colloquially referred to as a blood clot, that travels from the legs, or rarely other parts of the body, to the lungs where it obstructs central, lobar, segmental, or subsegmental pulmonary arteries depending on the size of the embolus. The untreated mortality rate of PE may approach

N. Navab et al. (Eds.): MICCAI 2015, Part II, LNCS 9350, pp. 62–69, 2015.
DOI: 10.1007/978-3-319-24571-3_8

30%. However, with early diagnosis and treatment, the mortality rate decreases to as low as 2% to 11%. CT pulmonary angiography (CTPA) is the primary means for the evaluation of suspected PE. At CTPA, an embolus appears as a dark region surrounded by the brighter, contrast-enhanced vessel lumen. CTPA dataset interpretation demands a radiologist to carefully trace each branch of the pulmonary artery for any suspected PEs. Therefore, PE diagnosis often requires extensive reading time, and the accuracy of CTPA interpretation depends on the radiologists' experience, attention span, eye fatigue, and their sensitivity to visual characteristics of PEs.

Computer-aided detection (CAD) can play a major role in detecting and diagnosing PEs. Recent clinical studies have shown that CAD systems can help radiologists increase their sensitivity for PE detection [3]. However, despite their demonstrated utility, existing CAD systems still require a relatively high false positive rate in order to achieve a clinically acceptable PE sensitivity. The false positives generated by CAD systems prolong the reading time of CTPA studies, because each CAD finding must be examined by a radiologist and adjudicated. It is therefore highly desirable to develop a CAD system that can achieve higher sensitivity while maintaining a clinically acceptable false positive range (between 1 to 5 false positives per CTPA study).

This paper investigates the feasibility of convolutional neural networks (CNNs) as an effective tool for eliminating false positive detections. We have found that the effective utilization of CNNs for detecting PEs and removing false detections in 3D CTPA datasets is contingent on an effective image representation of PEs. As such, a key finding from our work is a vessel-aligned multi-planar image representation of emboli that offers three advantages: (1) our proposed image representation is *efficient* and *compact* because it concisely summarizes the 3D contextual information around an embolus in only 2 image channels; (2) our proposed image representation is *consistent* because it automatically aligns the embolus in the 2-channel images according to the orientation of the affected vessel; and (3) our proposed image representation is *expandable* because it naturally supports data augmentation for training a CNN. We have evaluated our CAD system using 121 CTPA datasets containing a total of 326 emboli, achieving a sensitivity of 83% at 2 false positives per volume. This performance is superior to the best performing CAD system in the literature, which achieves a sensitivity of 71% at the same level of false positives. We have further evaluated our system with the entire 20 CTPA test datasets from the PE challenge [1]. Our system outperforms MeVis', the best reported system, at 0mm localization error but is outperformed by MeVis' at 2mm and 5mm localization errors. In our view, the performance at 0mm localization error is more important than those at 2mm and 5mm localization errors.

2 Related Work

CAD systems for PE typically consist of four stages: 1) extracting a volume of interest (VOI) from the original dataset by performing lung segmentation [5,11,8] or vessel segmentation [7,11,2]; 2) generating a set of PE candidates within the VOI using algorithms such as tobogganing [5]; 3) extracting hand-crafted features from each PE candidate (e.g., [6]), and 4) computing a confidence score for each of the candidates using a rule based classifier [7], neural networks and a nearest neighbor classifier [11,8],

or a multi-instance classifier [5]. However, current CAD systems either produce many false positives to achieve a high detection sensitivity [7], or yield acceptable false positive rates but with only limited sensitivity levels [8,2,11] (see Table 1 for a detailed performance comparison). We hypothesize that inadequate modeling of PEs based on hand-crafted features results in suboptimal CAD performance, and therefore investigate the use of a new image representation for PEs, coupled with CNNs, to improve state-of-the-art performance.

3 Proposed Method

Given a CTPA dataset, our method first segments lungs and then generates a set of PE candidates within the lung area using the tobogganing algorithm [5]. Our method then uses our vessel-aligned multi-planar image representation to produce a 2-channel image representation for each PE candidate. The resulting 2-channel patches are then fed to a CNN to classify the underlying candidates into PE or non-PE categories. Please refer to [5] for the tobogganing algorithm and to [4] for the CNN. In the following, we shall focus on our suggested vessel-aligned multi-planar image representation.

3.1 Vessel-Aligned Multi-planar Image Representation

The success of CNNs for object detection in 3D volumetric datasets such as CT images heavily relies on the representation of the object of interest [9,10]. We have experimentally found that a suitable 3D image representation for CNNs must meet three requirements: (1) compactness and efficiency, (2) consistency across instances, and (3) expandability for data augmentation. With these requirements in mind, we propose an image representation, called vessel-aligned multi-planar image representation, for PE, which has these three critical properties. In the following, we first describe our unique image representation and then explain how it meets the above requirements.

To obtain our image representation, we first estimate the orientation of the vessel that contains the candidate. For this purpose, a 15x15x15mm neighborhood is extracted around the PE candidate. In the resulting subvolume, the PE appears as a filling defect, because PEs are relatively darker than the contrast-enhanced vessel. To minimize the influence of the filling defect on vessel orientation estimation, the vessel-like intensity value of 100 HU (Hounsfield units) is assigned to the PE voxels within the subvolume. Note that the tobogganing algorithm [4] has already labeled the PE voxels associated with each candidate. Next, a principle component analysis is performed in the connected component ($\geq$ 100 HU) that contains the PE. If v_1, v_2, v_3 denote the eigen vectors of the analyzed component ($\lambda_1 \geq \lambda_2 \geq \lambda_3$), then interpolating the volume along $\{v_1, v_2\}$ or $\{v_1, v_3\}$ results in the longitudinal view of the PE (the first channel of our image representation) and interpolating the volume along $\{v_2, v_3\}$ results in the cross-sectional view of the PE (the second channel of our image representation).

Our image representation is compact because it concisely summarizes the 3D contextual information around PEs in only 2 image channels. While it is theoretically possible to train a CNN using subvolumes with an arbitrary number of slices, the performance of such networks have been reported to be inferior to the CNNs that have been trained

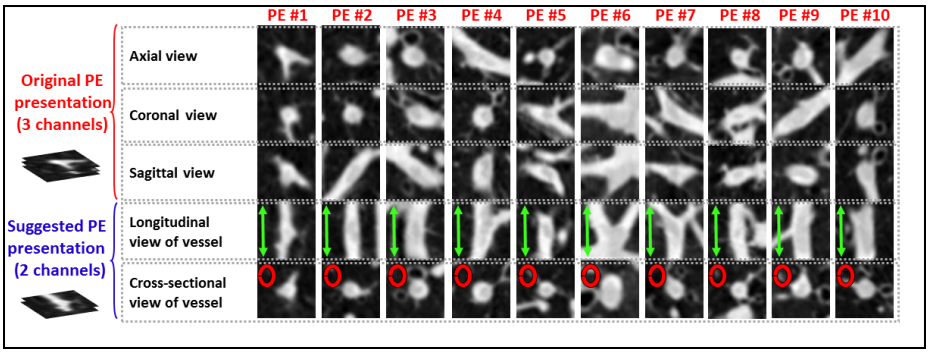

Fig. 1. The suggested 2-channel image representation characterizes emboli more consistently than the original axial, sagittal, and coronal views. As seen, in nearly all cases, the suggested scheme consistently captures PEs within the containing vessel as elongated and circular structures in the first and second channels, respectively. The three standard views do not provide this property given the varying orientation of the containing vessels. A consistent image appearance is the key to training an accurate image classifier.

using samples with a fewer number of slices [9]. In fact, the information embedded in the additional image slices has been shown to degrade classification performance [9]. This phenomenon is attributed to the curse of dimensionality, where a large number of image channels corresponds to learning a far larger number of network parameters, which in turn leads to over-fitting to the training samples and thus poor generalization performance. It is therefore desirable to efficiently represent the 3D context around the object of interest using a low dimensional image representation.

Our image representation consistently describes PEs and the containing vessels. In general, emboli can can affect pulmonary arteries in any orientation. As a result, images extracted from the axial, sagittal, coronal planes exhibit a significant variation in the appearance of emboli. This in turn complicates the classification task and hinders effective utilization of CNNs. With the benefit of vessel alignment, our image representation allows for a consistent image representation whereby emboli consistently appear as elongated structures in the longitudinal vessel view and as circular structures in the cross-sectional vessel view. Fig. 1 illustrates variations in PE appearances using the suggested vessel-aligned image representation and a standard image representation based on sagittal, coronal and axial views.

Our image representation amenably supports data augmentation, which is essential for effective training and testing of CNNs. In 2D applications, data augmentation is performed by applying arbitrary in-plane rotations and then collecting samples at multiple scales and translations. A 3D representation must also support the above operations to enable data augmentation. While it is straightforward to extend translation and scale to a 3D space, the rotation operation can be problematic. Our image representation is based on longitudinal and cross-sectionals planes; however, rotating such planes along a random axis will result in the arbitrary appearance of the same PE in the resulting 2-channel images (Fig. 2(a)). The major challenge is how to perform 3D rotation such that the PE representation remains consistent. Our image representation accommodates this need by rotating the planes around the vessel axis v_1. By doing so, we obtain two

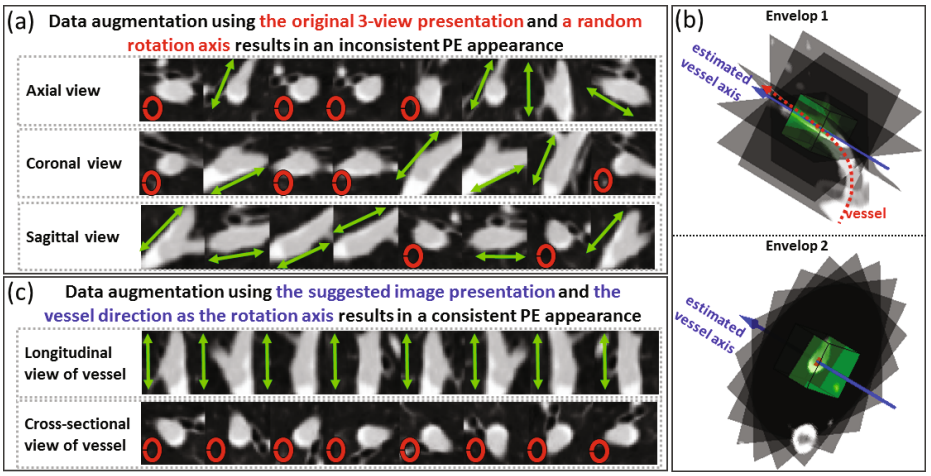

Fig. 2. (a) Data augmentation using random rotation axes, as suggested in [10], results in inconsistent PE appearance. (b) The suggested image representation uses two envelopes of planes to achieve consistency for data augmentation. (c) Consistent PE appearance after data augmentation using the suggested envelopes of planes. The green double arrows and red ellipses represent the shapes of PEs and the containing vessels.

envelopes of image planes (see Fig. 2(b)) where the first envelope contains the planes that all intersect at the vessel axis and the second envelope contains the image planes whose normals are the vessel axis. By selecting any pairs of planes from the two envelopes, one can generate a new PE instance while retaining the consistency. Fig. 2(c) illustrates consistency in appearance of PEs after data augmentation using the suggested envelopes of planes.

3.2 Convolutional Neural Networks (CNNs)

CNNs are deep learning machines that can potentially eliminate the need for designing hand-crafted features—they learn the features and train the classifier simultaneously. CNNs are so-named for their convolutional layers that learn discriminative patterns of the training samples at multiple scales. In this work, we employ the GPU-based open-source implementation of CNNs [4] and use the layout shown in Fig. 3. We have experimented with more sophisticated network architectures but observed no significant performance gain.

4 Experiments

We have evaluated our CAD system using 2 databases: (1) our private database consisting of 121 CTPA datasets with a total of 326 emboli, and (2) the test datasets from the PE challenge [1] consisting of 20 CTPA datasets with a total of 133 emboli.

Evaluations Using Our Database. The candidate generation module of our CAD system produces a total of 8585 PE candidates in the 121 CTPA datasets, of which 7722

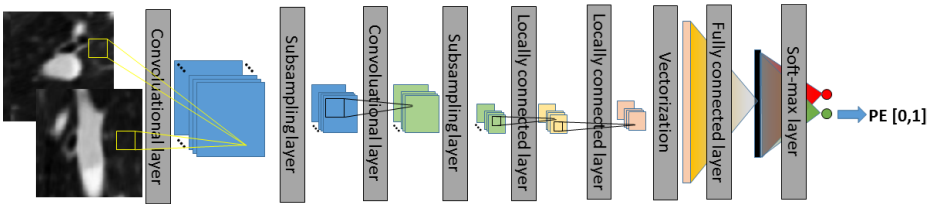

Fig. 3. The layout of the CNN used in our experiments.

are false positives and 863 are true positives. It is possible for a CAD system to produce multiple detections for a single large PE and that explains why the number of our true detections is greater than the number of PEs in the database. According to the available ground truth, the candidate generation module achieves a sensitivity of 93% for PE detection while producing, on average, 65.8 false positives per patient.

Our goal is to use CNNs to minimize the number of false positives while maintaining a high sensitivity for PE detection. To train CNNs, we randomly split the collected detections at the patient level into 3 groups, enabling a 3-fold cross validation of our CAD system. We then used the false positive detections as negative candidates and the true detections as positive candidates. Given the limited number of candidates, we formed the training set by performing data augmentation. For this purpose, we collected $N = N_r \times N_t \times N_s$ samples from each candidate location based on our vessel-aligned multi-planar PE representation, where N_r is the number of rotations, N_t is the number of translations, and N_s is the number of image scaling. To produce rotated patches, we rotated the longitudinal and cross-sectional vessel planes around the vessel axis $N_r = 5$ times. For scaling, we extracted patches at $N_s = 3$ different scales, resulting in 10mm, 15mm, and 20mm wide patches. In each scale, we have performed image interpolation so that the resulting patches are all 32x32 pixels. For translation, we shifted the candidate location along the vessel direction $N_t = 3$ times, up to 20% of the physical width of the patches. With data augmentation, we can increase the size of the training set by a factor of $N = 45$, which is sufficiently large to train CNNs. Given a test CTPA dataset, we first obtain a set of candidates, and then apply the trained CNN on N 2-channel image patches extracted from each candidate location. The confidence values for the underlying candidate is then computed as the average of the resulting N confidence values. Once all the test candidates are processed, we obtain an FROC curve by changing a threshold on the corresponding confidence values.

Fig. 4 shows the FROC curve of the suggested system. For comparison, we have computed the FROC curve of [5] using the prediction results provided by the corresponding author. We have chosen [5] for performance comparison because their suggested system has achieved the best performance reported in the literature on a reasonably large CTPA database (see Table 1). For further comparison, we have replaced our suggested image presentation with a 2.5D image representation as suggested in [10]. For fair comparisons, we have kept all the other stages the same. As seen in Fig. 4, our system outperforms [5], which is a CAD system based on a carefully designed set of hand-crafted features [6] and a multi-instance classifier. In addition, we observed that the CNN trained using a 2.5D image representation results in a performance which is not only inferior to our suggested image representation but also to the hand-crafted approach, demonstrating the

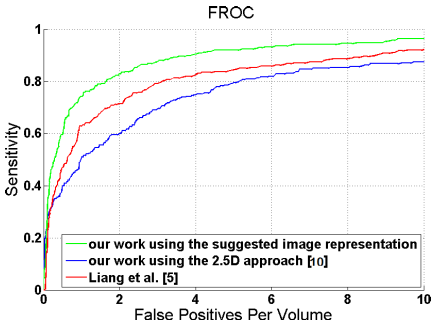

Fig. 4. Our CAD system using the suggested image representation outperforms the best hand-crafted approach [5] and also a CNN powered by a 2.5D approach [10].

Method	Sensitivity	FPs/vol	#datasets	#PEs
Liang et al. [5]	70.0%	2.0	132	716
Bouma et al. [2]	58%	4.0	19	116
Park et al. [8]	63.2	18.4	20	44
Ozkan et al. [7]	61%	8.2	33	450
Wang et al. [11]	62%	17.1	12	24
This work	83.4%	2.0	121	326
This work (2.5D)	60.4%	2.0	121	326
Liang et al. [5]	71.7%	2.0	121	326

Table 1. (top) Performance of the existing PE CAD systems obtained through different datasets. (bottom) Performance comparison based on our database of 121 CTPA datasets. Operating points are taken from Fig. 4.

significant contribution of our effective image representation in achieving the improved performance. Table 1 contrasts the performance of our proposed CAD system with that of the other CAD systems suggested in the literature.

Evaluations Using PE Challenge Database. We have further trained a CNN, powered by our unique image representation, using all 121 CTPA datasets from our database and then evaluated our CAD system using the test database from the PE challenge [1]. Since the ground-truth was not available on the website, our detection results were evaluated by the organizers. At 0mm localization error, our CAD system achieves a sensitivity of 34.6% at 2 FPs/vol, which outperforms the winning team (a commercial CAD system designed MeVis Medical Solutions) with a sensitivity of 28.4% at the same false positive rate. Our CAD system is, however, outperformed by MeVis' at 2mm and 5mm localization errors. For more detailed comparisons, please refer to [1]. Despite the demonstrated superiority at 0mm localization error, our CAD system exhibits a notable performance degradation compared to the results obtained using our database. We attribute this to faulty lung segmentation, which results in many PE candidates in the colon and diaphragm. Since such false positives had not been observed in our training sets, the trained CNN did not perform optimally in removing such false positives.

5 Conclusions and Discussions

In this work, we investigated the possibility of a unique PE representation, coupled with CNNs, to produce a more accurate PE CAD system. Our system contrasts with existing systems, wherein a traditional hand-crafted feature design is used for characterizing PEs. We evaluated our system in comparison with the most robust hand-crafted approach [5] and a learning-based approach using CNNs powered by a 2.5D PE representation, demonstrating a marked performance improvement. Our method was also tested using the test database from the PE challenge where it outperformed the academic systems at the three localization errors and also outperformed a commercial CAD system at 0mm localization error. Moving forward, we intend to improve the accuracy

of our CAD system using additional training cases to address the issue of faulty lung segmentation resulting from non-pulmonary candidates.

Acknowledgment. This project is supported by a seed grant awarded by Arizona State University and Mayo Clinic. We thank German Gonzalez, the organizer of the CAD PE challenge, for evaluating our results with the test datasets from the challenge.

References

1. http://www.cad-pe.org
2. Bouma, H., Sonnemans, J.J., Vilanova, A., Gerritsen, F.A.: Automatic detection of pulmonary embolism in cta images. IEEE Transactions on Medical Imaging 28(8), 1223–1230 (2009)
3. Das, M., Mühlenbruch, G., Helm, A., Bakai, A., Salganicoff, M., Stanzel, S., Liang, J., Wolf, M., Günther, R.W., Wildberger, J.E.: Computer-aided detection of pulmonary embolism: influence on radiologists detection performance with respect to vessel segments. European Radiology 18(7), 1350–1355 (2008)
4. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
5. Liang, J., Bi, J.: Computer aided detection of pulmonary embolism with tobogganing and mutiple instance classification in CT pulmonary angiography. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 630–641. Springer, Heidelberg (2007)
6. Liang, J., Bi, J.: Local characteristic features for computer aided detection of pulmonary embolism in ct angiography. In: Proceedings of the First MICCAI Workshop on Pulmonary Image Analysis, pp. 263–272 (2008)
7. Özkan, H., Osman, O., Şahin, S., Boz, A.F.: A novel method for pulmonary embolism detection in cta images. Computer Methods and Programs in Biomedicine 113(3), 757–766 (2014)
8. Park, S.C., Chapman, B.E., Zheng, B.: A multistage approach to improve performance of computer-aided detection of pulmonary embolisms depicted on CT images: Preliminary investigation. IEEE Transactions on Biomedical Engineering 58(6), 1519–1527 (2011)
9. Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 246–253. Springer, Heidelberg (2013)
10. Roth, H.R., et al.: A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part I. LNCS, vol. 8673, pp. 520–527. Springer, Heidelberg (2014)
11. Wang, X., Song, X., Chapman, B.E., Zheng, B.: Improving performance of computer-aided detection of pulmonary embolisms by incorporating a new pulmonary vascular-tree segmentation algorithm. In: SPIE Medical Imaging, pp. 83152U–83152U. International Society for Optics and Photonics (2012)