Early Pulmonary Embolism Detection from Computed Tomography Pulmonary Angiography Using Convolutional Neural Networks

Ching-Yuan Yu[1], Yun-Chien Cheng[#]
Department of Mechanical Engineering
National Chiao-Tung University
Hsinchu, Taiwan.
chingyuanyu.me07g@nctu.edu.tw, yccheng@gmail.com
[#]Corresponding author

Chin Kuo[#]
Department of Radiation Oncology
National Cheng Kung University Hospital
College of Medicine, National Cheng Kung University
Tainan, Taiwan.
tiffa663@gmail.com

***Abstract*—In this study, we developed the first computer-aided detection (CAD) system aimed at triage patients with pulmonary embolism (PE) to reduce the death rate during the waiting period. Computed tomography pulmonary angiography (CTPA) is used for definite diagnosis of PE, and CTPA imaging reports are read by radiologists who suggest further management, which requires time and hence a waiting period to obtain a diagnosis. Patients may die during this waiting period, and a CAD method can triage patients with PE from those without PE. In this study, we proposed a CAD system to achieve the aforementioned purpose. Our purpose is different from related studies and CAD systems that were aimed at identifying key PE lesion images in images of patients with PE to expedite PE diagnosis. Our CAD system consists of a novel classification-model ensemble for PE detection and a segmentation model to label PE lesion on each image. We utilized data from the National Cheng Kung University Hospital and open resource to construct models. In the classification model, the algorithm achieved an area under the receiver operating characteristic curve of 0.88 (accuracy = 0.85). In the segmentation model, the mean intersection over union was 0.689. Overall, our CAD system successfully distinguished patients with PE from those without PE and automatically labeled the PE lesion to expedite PE diagnosis.**

***Contribution*—This is the first CAD system aimed at triage patients with PE that uses the multiple convolutional neural network architecture.**

Keywords—classification; computed tomography pulmonary angiography; computer-aided detection system; convolutional neural network; deep learning; pulmonary embolism; segmentation.

I. INTRODUCTION

Pulmonary embolism (PE) is the obstruction of pulmonary arteries caused by blood clots, tumors, fat, air, and tissue. Because of extensive inducing factors and lack of specific clinical signs and symptoms, PE diagnosis is often delayed [1, 2]. Nevertheless, early diagnosis and management of PE are associated with decreased mortality [3]. In clinical practice, once a differential diagnosis of PE is obtained, computed tomography pulmonary angiography (CTPA) is performed for definitive diagnosis [4]. Generally, there is a waiting period when radiologists analyze CTPA reports and suggest further management; however, the patient may die during this waiting period. Approximately 34% of patients with PE suddenly died even before therapy could be initiated or before the therapy showed any effect [5]. Therefore, a real time PE alarm solution to reduce the waiting period is necessary. A computer-aided detection (CAD) model for PE detection that can read CTPA images immediately may solve the waiting-time problem, thereby facilitating early detection and treatment of patients with PE. Various CAD models for PE detection have been suggested. For example, Bouma *et al.* [6] selected features based on the shape of the vessel and lumen, intensity of voxel, and shape and size of textures by using different machine learning classifiers and then compared results. Ö zkan *et al.* [7] extracted 3D lung vessel tree from 2D PE images and then used median filtering, thereby eliminating candidates with only one image or an unexpected shape to reduce the false-positive rate. Park *et al.* [8] selected features depending on the intensity, shape, and boundary and reduced the false-positive rate by using an artificial neural network classifier, k-NN detection scores, and a 3D-grouping-based scoring method. Tajbakhsh *et al.* [9] adopted a convolutional neural network (CNN) as the classifier to reduce the false-positive rate. In addition, they proposed vessel-aligned multiplanar image representation of PE, which employed two image channels to improve classification by providing more details of the region of the candidate images. In 2019, Yang *et al.* [10] proposed a novel two-step method comprising two CNN structures. The first step employed a 3D-object-detection model to distinguish the PE candidate images from the entire CTA scan, and the second step implemented ResNet-18 as the classifier to remove false positives. All the aforementioned models were aimed at expediting PE diagnosis. Thus, these models were created to identify key PE lesion images in images of patients with PE. These models helped radiologists to reduce the time required to

manually trace numerous images. However, these models could not triage patients with PE from those without PE. Therefore, in this study, the aim was to develop a two-step process explicitly focused on triage PE diagnosis. We employed CNN to develop a two-step CAD system consisting of the following two components: (1) a classification model for PE detection and (2) a semantic segmentation model to label the PE region.

Notably, the major contributions of this study can be summarized as follows: (1) we proposed a classification model that can provide the preliminary impression of PE or non-PE on a CTPA image immediately after obtaining it, thereby facilitating the management of patients with PE by radiologists and clinicians without delay; (2) our semantic segmentation model could label the location of the lesion and reduce the labor and time required to manually trace hundreds of images; and (3) using our method, we achieved a precision of 85.4% and a recall of 84.9% in per-image evaluation and a sensitivity of 82% and a specificity of 90% in per-patient evaluation.

TABLE I METHOD COMPARISON

	Characteristics	**Pros**	**Cons**
Machine Learing (ML)	image processing & machine learing classifier	Easy to implement	Time consuming Information missing
CNNs	Image processing & CNNs classifier	Easy to implement Increase precision from ML method	Time consuming Information missing
Object Detection & CNNS	Object detection & CNNs classifier	Candidate proposal improvement Precision improvement	Only detect PE in PE patient
Ours	Multiple CNNs classifier	PE triage without delay Reduce labor with lesion labeling	Still miss small number of positive cases in triage (NPV: 0.91)

II. METHOD

This study was approved by the Institutional Review Board of National Cheng Kung University Hospital (NCKUH) and was granted an exemption and waiver of informed consent.

A. Classification model: PE detection

- **Model architecture:** The architecture design was aimed at distinguishing PE images with high sensitivity. Physicians considered local and global information in CTPA images when diagnosing PE symptoms. Therefore, dilated residual network (DRN) [7] was the first architecture we selected as the backbone of the classification model. DRN spread the receptive field the same computation consumes through in dilated convolution to present more information and to improve feature representation of images (**Figure 1**). The second architecture is illustrated in **Figure 2**. MixNet [8] was proposed in 2019 and achieved a similar effect by using a mixed depth-wise convolutional mechanism. The mixed depth-wise convolutional layer broadened the receptive field by applying different kernel sizes to each feature channel. To achieve higher sensitivity, we applied model ensemble to balance the strengths and weaknesses of models. The images were used as input for each model, and average predictions were the final output. By using the multiple ensemble strategy illustrated in **Figure 3**, we improved the per-image prediction for sensitivity by 3% and for specificity by 2%.
- **Per-patient classification:** Every individual who underwent CTPA examination would be given a preliminary diagnosis of PE or non-PE. Once any image of patient was predicted as PE image, our system would classify this patient as a PE patient. Otherwise, the patient was predicted as non-PE patient. We input all the images from every individual into aforementioned classification model. The result showed that approaching high sensitivity increased false-positive prediction of patients without PE in per patient tests. Moreover, most false-positive predictions were images in the lung region. Thus, we trained another DRN by using a modified PE100 dataset that consisted of only PE images of patients with PE and images from the lung region of patients without PE to refine our classification prediction and to reduce the false-positive rate in patients without PE.

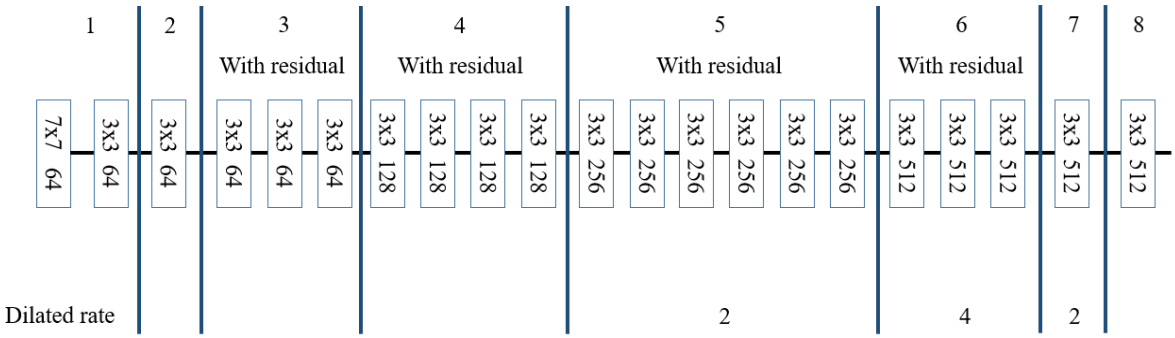

Figure 1. DRN architecture.

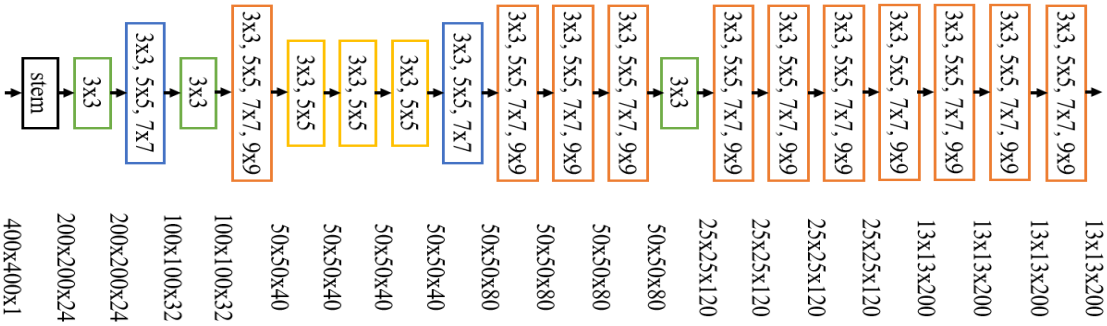

Figure 2. MixNet architecture.

B. *Semantic segmentation model: PE lesion pixel-wise segmentation*

The aim of this semantic segmentation model was to label the PE region in detected images. Normally, PE only occupies a small region in each image; therefore, labeling the region precisely was challenging. To address this problem, acquiring precise location information was a priority compared with presenting complicated high-level features. Therefore, we proposed using ARUX-Net based on the ARU-Net [9] architecture, which exhibited excellent performance in detecting minute historical text lines. ARU-Net [9] was proposed based on U-Net [10]. Using residual block and attention mechanism, the gradient vanish problem was solved and the model could focus on appropriate spots. We used the squeeze-and-excitation layer [11] and joint pyramid upsampling (JPU) mechanism [12] on the model to enhance feature extraction and feature decoding, respectively. As illustrated in **Figure 5**, ARUX-Net consists of two CNN structures with RUX-Net as the encoder-decoder architecture and A-Net as the attention mechanism architecture. In A-Net, we utilized a four-layer basic CNN structure to lead the model to focus on appropriate spots. In RUX-Net, we implemented an encoder-decoder architecture by using four downsampling layers, a JPU mechanism, and three upsampling layers. In each downsampling layer, we applied two convolution sets and a squeeze-and-excitation layer. Each convolution set consisted of a 2D convolution layer, batch normalization layer, and an activation layer. We concatenated the feature maps of Down3 and Down4 in **Figure 5** and refined feature representation by using the JPU mechanism. Then, we upsampled the feature map by using three upsampling layers and combined the features extracted in Down1 and Down2 by using a shortcut as was originally designed in U-Net.

C. *Implementation details*

The network was coded and implemented using Python v3.6 with Pytorch. The algorithm was trained on a system with following features: 2 GTX 1080 Ti graphics processing unit (GPU) (12G per GPU) (Nvidia; Santa Clara, CA, USA). The method proposed by Ranger [14] was utilized as optimizer of classification and segmentation models with the original learning rate at 1e-3, which could automatically tune the learning rate and the momentum in each epoch. We selected binary cross-entropy (BCE) with focal loss design [15] as the loss function of the classification model and BCE combined with dice loss [16] as the loss function of the segmentation model.

D. *Model evaluation and performance statistics*

The performance of the classification model was evaluated using the test set by area under the curve (AUC) of the receiver operating characteristic, precision and recall per image. The precision and recall are defined as (1) and (2). Per patient test was evaluated with sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). We used mean intersection-over-union (mean IoU) to evaluate the performance of the semantic segmentation model by using the test set.

$$\frac{TP_0}{TP_0 + FP_0} \times W_0 + \frac{TP_1}{TP_1 + FP_1} \times W_1 \quad (1)$$

$$\frac{TP_0}{TP_0 + FN_0} \times W_0 + \frac{TP_1}{TP_1 + FN_1} \times W_1 \quad (2)$$

TP_1, TP_0 are true positive in PE class and non-PE class, FP_1, FP_0 are false positive in PE class and non-PE class,FN_1, FN0, are false negative in PE class and non-PE class, W_1,W_0 are image amount ratio of PE class and non-PE class.

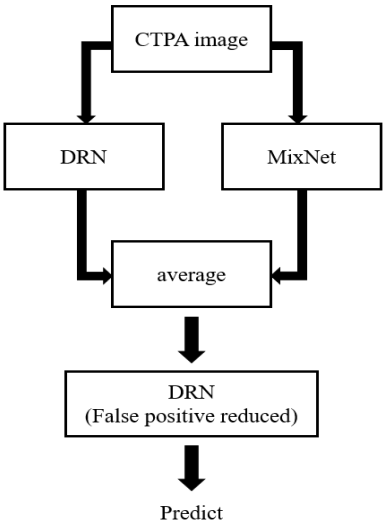

Figure 3. Classification process.

III. EXPERIMENTS AND RESULTS

A. Data and pre-processing

We conducted a retrospective review of consecutive patients who underwent CTPA examination. The dataset utilized to construct the PE detection model composed of data from NCKUH and open resource [12]. The dataset from NCKUH consisted of any patient who underwent the CTPA examination between Jan. 01, 2018, and Sep. 01, 2019. CTPA digital imaging and communications in medicine (DICOM) images were reviewed and labeled by two physicians, namely a senior radiologist and a radiation oncologist. Radiologists gave a diagnosis of PE based on CTPA radiographic features. The CTPA of PE would show filling detect within pulmonary vasculature. When the pulmonary artery was reviewed on its axial plane, the central filling detect from the thrombus was surrounded by a thin rim of contrast or a thin stream of contrast adjacent to the embolus. The affected vessel might also enlarge.

In the dataset used to construct the classification model, 200 patients were included, of whom 100 were patients with PE and the remaining 100 were patients without PE. Of 200 patients, 165 were selected from the NCKUH dataset and 35 from open source. We obtained 3892 images with PE and 20211 images without PE. Scans were performed on Siemens CT scanners. Scanner specifications and image protocol details are presented in **Table 1**. CT scans were of diagnostic quality and were obtained using 100 kVp energy, image thickness 8 mm, pixel space 0.49 mm, and Optiray contrast administered intravenously. The datasets were shuffled and divided by patient into training, validation, and testing sets in the ratio of 7:2:1, respectively. For a similar number of patients, the number of non-PE images was more than four times the number of PE images. Thus, we upsampled PE images as five times to address the data imbalance problem. We then center-cropped a region of 400×400 in CTPA images, adjusted the contrast by setting Hounsfield Unit values of >600 as 600, and linearly transformed all pixel values to [−1, 1].

The dataset utilized to develop the PE lesion pixel-wise segmentation was an open resource [12]. It contained 2301 PE images. We applied random horizontal, vertical flip, and center-crop on a 400×400 region of CTPA images for data augmentation. The dataset was divided by patient into training, validation, and testing sets in the ratio of 7:2:1, respectively.

B. Results of classification model: PE detection

We trained the DRN [6] to achieve a precision of 85.4%, a recall of 86.2%, and an AUC of 0.707. MixNet-L [7] achieved a precision of 82.4%, a recall of 84.1%, and an AUC of 0.644. After the ensemble, performance of the classification model improved and precision of 85.8% and recall of 87.2% were achieved. With the false-positive rate reduced, the performance improved and a precision of 92.3%, a recall of 92.0%, and an AUC of 0.878 were achieved. We compared our results with ResNet 50 as benchmark. (**Table 3**). Moreover, we plotted the receiver operating characteristic (ROC) curve showed as Figure 4.

In the per-patient test, the ensemble with the false-positive reduction architecture achieved a sensitivity of 82% and a specificity of 90%. Among 21 patients within test dataset, 9 out 11 PE patients and 9 out of 10 non-PE patients were predicted correctly. One PE patient was predicted wrong because of poor image quality, and the other PE patient was misdiagnosed due to solitary small lesion. Because the dataset contains relatively few extremely poor-quality data, it is difficult for the model to identify poor-quality images such as the presence of artifacts, overexposure, and under-exposure, field-size misplacement. The non-PE who had prominent mediastinal lymph nodes was mistaken as PE. Overall, the performance of classification per patient was clinically acceptable [17]. Besides, we achieved a PPV of 0.90 and a NPV of 0.91 in per patient test. Clinical alert fatigue is less concerned with high PPV.

C. Semantic segmentation model: PE lesion pixel-wise segmentation

In current clinical practice, physicians are required to manually track hundreds of slices. By using the semantic segmentation algorithm, physicians can directly focus on the colored areas marked by the model without intersecting with hundreds of non-lesion slices. Further make the diagnosis more effective (**Figure 5**).

In the segmentation test set, we obtained 108 images with mask, and the model achieved a mean IoU of 0.689. We compared the results with U-Net which only achieved a mean IoU of 0.613.

D. Clinical workflow incorporating the CAD system

The clinical workflow incorporating the CAD system proposed in this study is illustrated in **Figure 6**. Immediately after CTPA examinations were obtained, the CTPA DICOM images were sent to the classification model for analysis. The classification model was a multiple CNN-structure ensemble aimed at analyzing each image to detect PE, and the output was whether the patient had PE, which facilitated early detection and treatment. Once it was revealed that the patient had PE, CTPA images were further analyzed using the semantic segmentation model that labeled the PE region in the detected

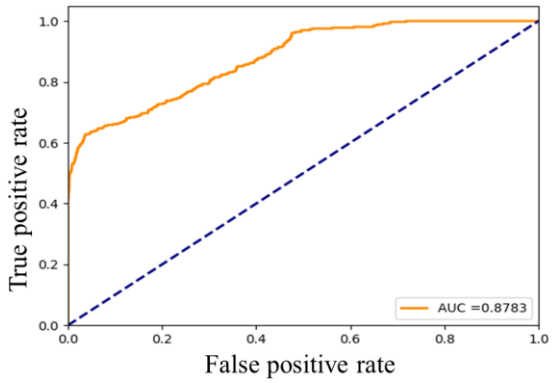

Figure 4. Receiver operating characteristic (ROC) curve

images to reduce the effort of manually tracing the lesions.

TABLE II. DATA CHARACTERISTICS

Category	Data from National Cheng Kung University Hospital	Open source
Patient number (PE vs no PE)	100 vs 100	33 vs 2
Image number (PE vs no PE)	1591 vs 9828	2301 vs 6491
Manufacturer	Siemens	Neusoft
Slice thickness (cm)	0.8	0.1-0.2
Tube voltage (kVp)	100	120
Axial spatial resolution (pixels)	512 x 642	512 x 512

TABLE III. PER-IMAGE TEST RESULT

Model	Precision	Recall	AUC
ResNet 50	0.814	0.746	0.774
DRN	0.854	0.862	0.707
MixNet	0.824	0.841	0.644
Ensemble	0.858	0.872	0.687
Final CAD	0.923	0.920	0.878

TABLE IV. PER-PATIENT TEST RESULT

Sensitivity	Specificity	PPV	NPV
0.82	0.90	0.90	0.91

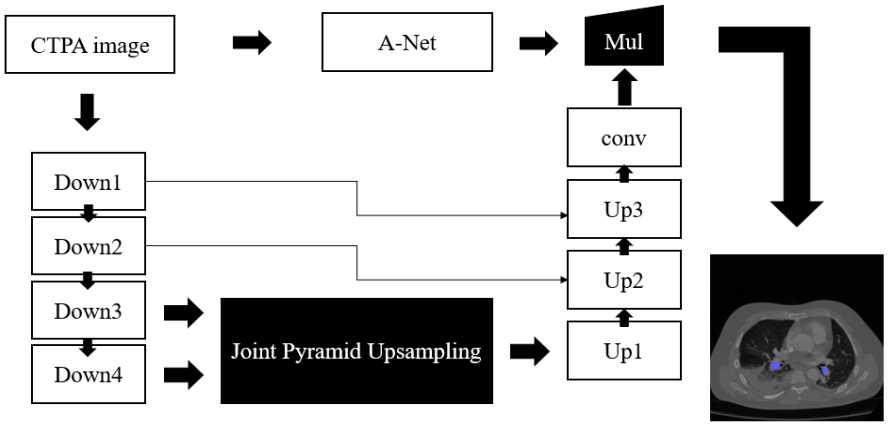

Figure 5. ARUX-Net architecture

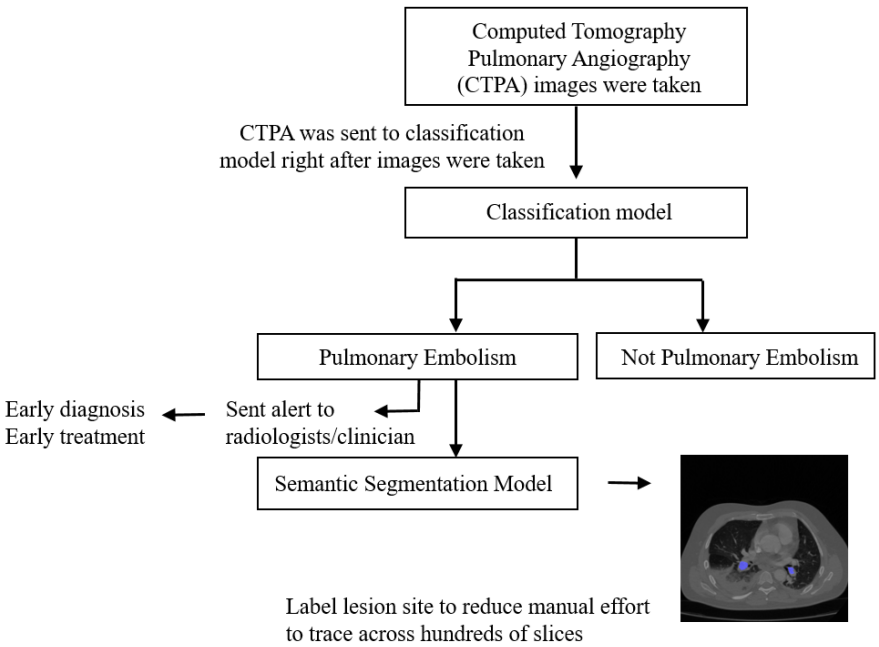

Figure 6. Proposed CAD system.

IV. CONCLUSIONS

This work presented a CAD triage system to early detect PE patients among patients who underwent CTPA examination. In classification step, we proposed a model-ensemble architecture to achieve a better precision and applied another DRN to reduce the false positivity in non-PE patients. In segmentation step, our model was composed of multiple novel mechanism to precisely label the PE lesions with higher mIoU compared with previous studies. With the detection precision of 0.858 and the segmentation mean IoU of 0.689, our models were capable of detecting PE patients, labeling images with PE lesions and fulfilling the clinical demands. The model is fundamentally limited by the dataset. The dataset came from two institutions. Further external verification is needed. We plan to externally validate this model with another dataset in the future work.

ACKNOWLEDGMENT

We are grateful for kindly assistance from the Department of Diagnostic Radiology, NCKUH for offering medical imaging and image labels.

REFERENCES

[1] H. Bouma, J. J. Sonnemans, A. Vilanova, and F. A. Gerritsen, "Automatic detection of pulmonary embolism in CTA images," IEEE transactions on medical imaging, vol. 28, no. 8, pp. 1223-1230, 2009.

[2] H. Özkan, O. Osman, S. Şahin, and A. F. Boz, "A novel method for pulmonary embolism detection in CTA images," Computer methods and programs in biomedicine, vol. 113, no. 3, pp. 757-766, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169260713004045?via%3Dihub.

[3] S. C. Park, B. E. Chapman, and B. Zheng, "A multistage approach to improve performance of computer-aided detection of pulmonary embolisms depicted on CT images: Preliminary investigation," IEEE transactions on biomedical engineering, vol. 58, no. 6, pp. 1519-1527, 2010.

[4] N. Tajbakhsh, M. B. Gotway, and J. Liang, "Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: Springer, pp. 62-69.

[5] Cohen AT, Agnelli G, Anderson FA, "Venous thromboembolism (VTE) in Europe. The number of VTE events and associated morbidity and mortality, " Thromb Haemost. 2007, 98(4), pp. 756-64.

[6] X. Yang et al., "A Two-Stage Convolutional Neural Network for Pulmonary Embolism Detection From CTPA Images," IEEE Access, vol. 7, pp. 84849-84857, 2019.

[7] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 472-480.

[8] M. Tan and Q. V. Le, "Mixnet: Mixed depthwise convolutional kernels," arXiv preprint arXiv:1907.09595, 2019.

[9] T. Grüning, G. Leifert, T. Strauß, J. Michael, and R. Labahn, "A two-stage method for text line detection in historical documents," International Journal on Document Analysis and Recognition (IJDAR), vol. 22, no. 3, pp. 285-302, 2019.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, 2015: Springer, pp. 234-241.

[11] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132-7141.

[12] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation," arXiv preprint arXiv:1903.11816, 2019.

[13] M. Masoudi, H.-R. Pourreza, M. Saadatmand-Tarzjan, N. Eftekhari, F. S. Zargar, and M. P. Rad, "A new dataset of computed-tomography angiography images for computer-aided detection of pulmonary embolism," Scientific data, vol. 5, 2018.

[14] Q. Tong, G. Liang, and J. Bi, "Calibrating the Learning Rate for Adaptive Gradient Methods to Improve Generalization Performance," arXiv preprint arXiv:1908.00700, 2019.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980-2988.

[16] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 Fourth International Conference on 3D Vision (3DV), 2016: IEEE, pp. 565-571.

[17] H. Bouma, J. J. Sonnemans, A. Vilanova and F. A. Gerritsen, "Automatic Detection of Pulmonary Embolism in CTA Images," in IEEE Transactions on Medical Imaging, vol. 28, no. 8, pp. 1223-1230, Aug. 2009, doi: 10.1109/TMI.2009.2013618.