Heliyon 9 (2023) e16060

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Feature-enhanced adversarial semi-supervised semantic segmentation network for pulmonary embolism annotation

Ting-Wei Cheng [a], Yi Wei Chua [a], Ching-Chun Huang [b], Jerry Chang [a], Chin Kuo [c,d,**,1], Yun-Chien Cheng [a,*,1]

[a] Department of Mechanical Engineering, College of Engineering, National Yang Ming Chiao Tung University, Hsin-Chu, Taiwan
[b] Department of Computer Science, College of Computer Science, National Yang Ming Chiao Tung University, Hsin-Chu, Taiwan
[c] Department of Oncology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[d] College of Artificial Intelligence, National Yang Ming Chiao Tung University, Hsin-Chu, Taiwan

ARTICLE INFO

Keywords:
Pulmonary embolism
Computed tomography pulmonary angiogram
Semantic segmentation
Semi-supervised learning
Unlabeled images

ABSTRACT

This study established a feature-enhanced adversarial semi-supervised semantic segmentation model to automatically annotate pulmonary embolism (PE) lesion areas in computed tomography pulmonary angiogram (CTPA) images. In the current study, all of the PE CTPA image segmentation methods were trained by supervised learning. However, when CTPA images come from different hospitals, the supervised learning models need to be retrained and the images need to be relabeled. Therefore, this study proposed a semi-supervised learning method to make the model applicable to different datasets by the addition of a small number of unlabeled images. By training the model with both labeled and unlabeled images, the accuracy of unlabeled images was improved and the labeling cost was reduced. Our proposed semi-supervised segmentation model included a segmentation network and a discriminator network. We added feature information generated from the encoder of the segmentation network to the discriminator so that it could learn the similarities between the prediction label and ground truth label. The HRNet-based architecture was modified and used as the segmentation network. This HRNet-based architecture could maintain a higher resolution for convolutional operations to improve the prediction of small PE lesion areas. We used a labeled open-source dataset and an unlabeled National Cheng Kung University Hospital (NCKUH) (IRB number: B-ER-108-380) dataset to train the semi-supervised learning model, and the resulting mean intersection over union (mIOU), dice score, and sensitivity reached 0.3510, 0.4854, and 0.4253, respectively, on the NCKUH dataset. Then we fine-tuned and tested the model with a small number of unlabeled PE CTPA images in a dataset from China Medical University Hospital (CMUH) (IRB number: CMUH110-REC3-173). Comparing the results of our semi-supervised model with those of the supervised model, the mIOU, dice score, and sensitivity improved from 0.2344, 0.3325, and 0.3151 to 0.3721, 0.5113, and 0.4967, respectively. In conclusion, our semi-supervised model can improve the accuracy on other datasets and reduce the labor cost of labeling with the use of only a small number of unlabeled images for fine-tuning.

* Corresponding author. Department of Mechanical Engineering, College of Engineering, National Yang Ming Chiao Tung University, Hsin-Chu, Taiwan.
** Corresponding author. Department of Oncology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan.
E-mail addresses: tiffa663@gmail.com (C. Kuo), yccheng@nycu.edu.tw (Y.-C. Cheng).
[1] The authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2023.e16060
Received 26 March 2023; Received in revised form 28 April 2023; Accepted 4 May 2023
Available online 6 May 2023
2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Pulmonary embolism (PE) is a disease in which the blood vessels in the lungs are blocked by foreign bodies, resulting in hypoxia in the lung tissue. Acute PE can even lead to death. Since the symptoms are similar to those of a cold, doctors need to arrange a variety of diagnostic methods to confirm the PE condition. The diagnostic methods include electrocardiogram, chest X-ray, computed tomography (CT) scan, blood test, and computed tomography pulmonary angiogram (CTPA). Among these methods, CTPA is the main criterion for the final diagnosis of PE. Therefore, the detection of PE in CTPA images is clinically important. However, in PE CTPA detection, doctors need to find the lesion area from a large amount of images, which makes diagnosis difficult. To accelerate the diagnostic process and improve the PE CTPA detection accuracy, the use of deep learning to classify and annotate PE has become a popular research topic in recent years.

The deep learning research on PE CTPA images can be divided into two categories: classification and segmentation. PE CTPA image classification can identify patients with PE early to reduce mortality. Many studies have examined PE CTPA image classification [1–3], and the use of deep learning can achieve rapid classification. Another category is PE lesion segmentation, which enables doctors to find the lesion area of PE quickly and easily provide follow-up treatment. Most of the current studies on PE lesion segmentation with supervised learning [4,5] have encountered two challenges: (1) A great amount of manual labeling is required, resulting in huge labor costs, and (2) different hospitals use different CTPA imaging settings, leading to different PE CTPA image characteristics, such as contrast and vessel detail. As a result, the accuracy may decrease when datasets from different hospitals are used for testing.

To overcome these two challenges, we built a model with a semi-supervised learning method; that is, the model is trained with both labeled images and unlabeled images. As shown in Fig. 1, we used labeled open CTPA datasets and unlabeled CTPA datasets from one hospital. By using the unlabeled data for training, the amount of manual labeling can be reduced. Furthermore, once the model is trained, the model can be further trained and hence fine-tuned with other unlabeled datasets from different hospital imaging settings. As a result, the model can be fine-tuned and applied to different hospital images, and fewer data are needed.

We established a feature-enhanced semi-supervised segmentation model for annotating PE CTPA images. We trained the model with both a labeled open dataset and an unlabeled hospital dataset so that the model could adapt better to the unlabeled dataset. Then we used an unlabeled dataset from another hospital to fine-tune and test the model. Moreover, we also strengthened the ability of our model to perform PE feature extraction. The improvements and advantages of our proposed model included the following.

- We strengthened the ability of the discriminator by adding feature maps to the predicted label and then using it as the input of the discriminator during training. The results showed that the boundaries and shapes predicted were improved.
- The HRNet-based architecture was used to enhance the PE feature extraction. We reduced a downsampling layer and removed the 4th stage of HRNet [6] to reduce the calculation and extract PE features at a higher resolution. In addition, the bilinear upsampling layer in the decoder was replaced with the CARAFE [7] upsampling module to better restore the PE shape and position.
- This is the first study to establish semi-supervised segmentation for PE CTPA datasets. The segmentation model can first be trained with one data set, and, with our semi-supervised method, the model can be trained for application to datasets from other hospitals with the use of only a few images.

The use of the semi-supervised learning method can reduce the demand for labeled images and reduce labor costs. In addition, due to the use of unlabeled images for training and fine-tuning of the model, our model is more adaptable to the images obtained from different hospitals than are models trained by supervised learning. Therefore, our method is more widely applicable to different datasets.

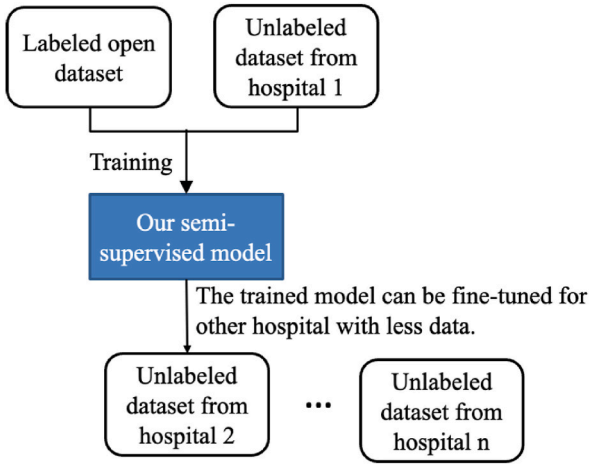

Fig. 1. Our semi-supervised model was trained with an open dataset and an unlabeled dataset from one hospital. The trained model can be fine-tuned with other unlabeled datasets for apply to other hospitals.

2. Related work

Previous studies have used only supervised learning methods for PE lesion segmentation. Carlos et al. used the 2D, 2.5D, and 3D U-Net models to train PE CTPA images and compared the differences between those models [5]. The sensitivity of 3D U-Net, which performed the best, was 0.55. Kun et al. used mask RCNN for object detection in PE CTPA images [4] and achieved a sensitivity of 0.747. However, the above-mentioned studies were all tested on a single dataset, and the model performed well on only a single dataset. CTPA images vary according to the machine settings of different hospitals, making it difficult to apply a model directly to other hospital datasets. In this case, some of the current research using the domain adaptation method to map the source domain and target domain into common feature space to minimize the dataset bias and negative influence. However, domain adaptation requires a pre-existing concept of the source and target domain to avoid distributional matching cannot guarantee good generalization on the target domain [8]. In addition, in some high-dimensional practical settings, multiple transformations can align the data distributions but not the posterior or class-conditional distributions [9]. Therefore, this study proposes the use of semi-supervised learning methods which can be better capture the underlying structure of the data distribution and generalize to new domains. This can be particularly important in medical image analysis where different imaging modalities or acquisition protocols can lead to significant domain shifts. With semi-supervised method the adaptability will improve while solving different datasets so that the model can be applied to CTPA images obtained from different hospitals with few data needed to get outperformance accuracy.

Semi-supervised learning method involves training a model with both labeled and unlabeled data. Semi-supervised learning methods have also been used on magnetic resonance imaging (MRI) and microscope images. Considering the various common semi-supervised medical segmentation methods, they can be categorized into three distinct strategies: pseudo labels, unsupervised regularization, and semi-supervised learning with knowledge priors. Firstly, assigning pseudo annotations for unlabeled images and integrating them with labeled images to update the segmentation model is a direct and intuitive method for utilizing unlabeled data. However, when using an under-trained segmentation model with limited labeled data, the output may be noisy, potentially leading to unstable training and negatively impacting performance [10]. Recent advances in semi-supervised medical image segmentation have shifted the focus towards incorporating unlabeled data in the training procedure through unsupervised regularization, such as unsupervised loss functions, which generate a supervision signal without the need for iterative pseudo label generation and segmentation model updates. Secondly, Knowledge priors refer to the information that a learning model already possesses before learning new information, which can be useful in addressing new tasks. Medical images contain numerous anatomical priors, such as the shapes and positions of organs, and incorporating this prior knowledge into semi-supervised models can lead to improved performance in medical image segmentation or representation ability for the new task [11]. However, the effectiveness of this approach relies heavily on the availability of prior knowledge, which can be challenging to obtain or may not be accurate in certain scenarios. Different from generating pseudo labels and prior knowledge method, unsupervised regularization with adversarial learning use adversarial methods to encourage the segmentation of unlabeled images to be more similar to that of labeled images. Such methods typically involve a discriminator that distinguishes between inputs from labeled annotations and unlabeled predictions. Dong et al. used an adversarial semi-supervised learning architecture to train MRI images for segmenting prostate, bladder, and rectum [12]. By adding unlabeled images during training, the model could revise the predicted boundary. Zhou et al. applied semi-supervised learning to microscope images for annotating diabetic retinopathy [13] It was verified that semi-supervised learning can also perform well in small-object segmentation with complex shapes. This is another reason why we proposed the application of semi-supervised learning to the PE dataset, which contains small objects, so as to enable our model to better adapt to different datasets and improve the accuracy.

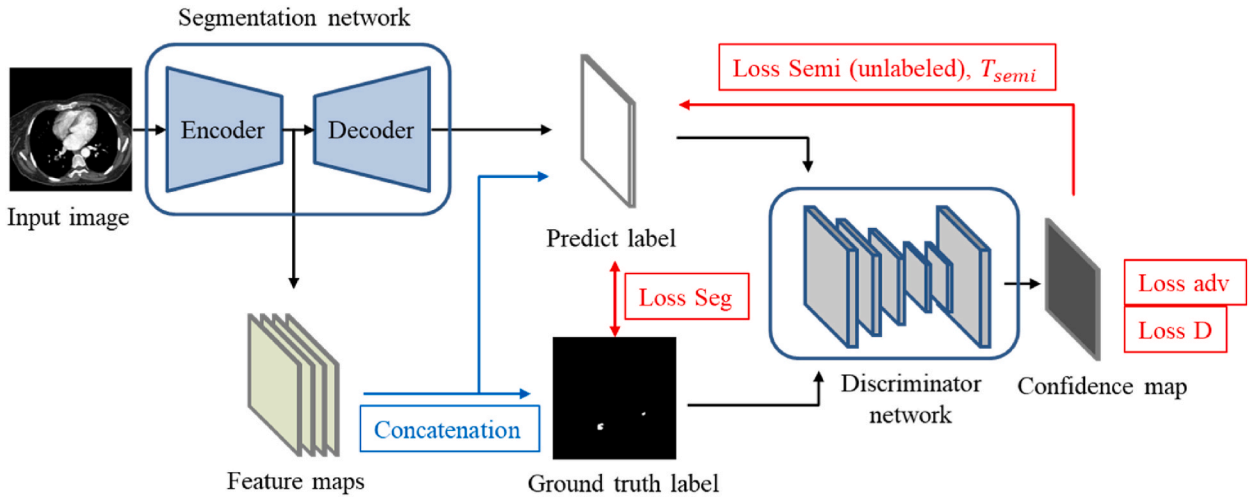

Fig. 2. Feature-enhanced adversarial semi-supervised segmentation architecture. The discriminator network trained using $\mathscr{L}_D$. We employ three loss functions to optimize the segmentation network: cross-entropy loss $\mathscr{L}_{Seg}$ based on the segmentation ground truth, adversarial loss $\mathscr{L}_{adv}$ to deceive the discriminator, and semi-supervised loss $\mathscr{L}_{semi}$ that depends on the confidence map.

3. Methods

3.1. Feature-enhanced adversarial semi-supervised segmentation

Our semi-supervised semantic segmentation architecture refers to the adversarial learning architecture by Hung et al. [14]. The concept is to train a segmentation network and a discriminator with labeled data so that the discriminator can be used to provide a prediction with positional features when predicting unlabeled data. The model used by Hung et al. was mainly for panoptic segmentation. The input of the discriminator was the masks of different classes, which enabled the discriminator to learn features for differentiating classes. However, there is only one class in the PE task, making it difficult for the discriminator to learn features through a single mask. To better leverage the discriminator in the PE task, we proposed a novel design by adding a feature map to the predicted label. The feature map was extracted from the semantic segmentation model encoder and used as the input of the discriminator, as shown in Fig. 2. Therefore, the discriminator not only learned from the masks but also from the mask position and PE-related features. Our semi-supervised segmentation network contained a segmentation network and a discriminator. The segmentation network was a revised version of HRNet with enhanced PE features to increase the computational efficiency while maintaining high-resolution convolutional operations. The added discriminator trained by adversarial learning methods can further make predictions of unlabeled data that are close to the real data annotation.

Our training process involves both labeled and unlabeled images in a semi-supervised setting. Labeled data is used to supervise the segmentation network with both the standard cross-entropy loss $\mathscr{L}_{Seg}$ and the adversarial loss $\mathscr{L}_{adv}$ through the discriminator network. It is important to note that we train the discriminator network exclusively with labeled data. On the other hand, we train the segmentation network with our proposed semi-supervised method for the unlabeled data. The initial segmentation prediction of the unlabeled image is computed by the segmentation network, and a confidence map is obtained by passing the prediction through the discriminator network. This confidence map serves as a supervisory signal, and we use a self-taught scheme to train the segmentation network with a masked cross-entropy loss $\mathscr{L}_{semi}$. The confidence map indicates the quality of the predicted segmented regions that the segmentation network can trust during training. The feature maps generated from the encoder are used as the input of the discriminator to improve the ability of the discriminator.

3.1.1. Segmentation network

Although HRNet can extract features with high resolution, the resolution is still too low for segmentation of the small-sized PE features. Therefore, we modified our HRNet-based segmentation model to extract the features of PE at an even higher resolution. To increase computation efficiency while detecting small objects at this higher resolution, we made three modifications to HRNet to save computing resources, as shown in Fig. 3. The purpose of the modification was to maintain the PE features at a high resolution during the training. Moreover, we used the CARAFE upsampling module to replace the bilinear upsampling module of the encoder to increase the accuracy of the predicted shapes and boundaries. The first modification was to remove the second convolutional layer of HRNet to reduce the downsampling rate before entering the network from 1/4 to only 1/2. This allowed the images to preserve a larger feature map, which would be suitable for PE feature detection. However, this removal significantly increased the computing cost. Therefore, in the second modification, we removed the 4th stage of the original HRNet to reduce the computation complexity. Since the feature map

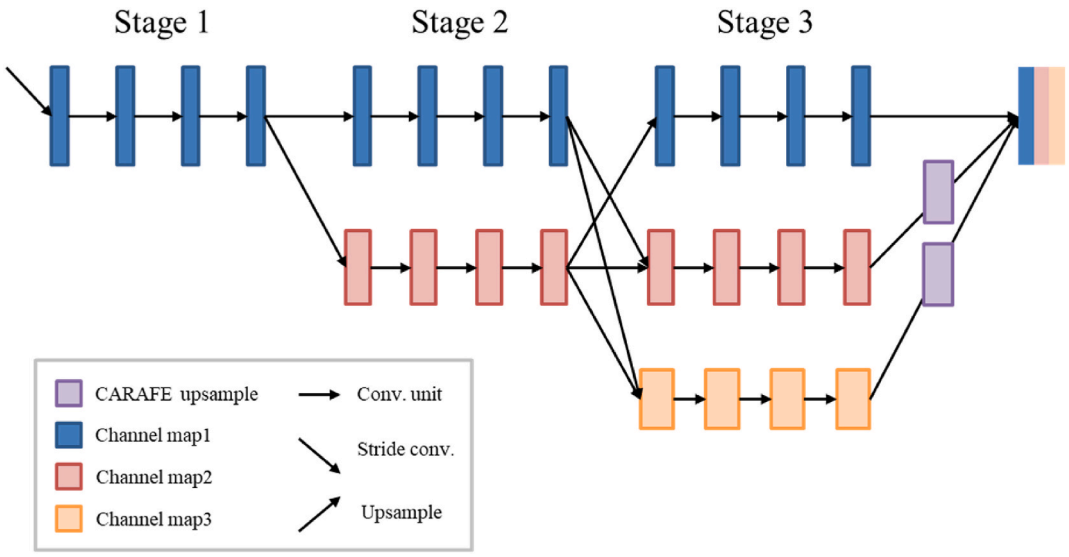

Fig. 3. Segmentation network structure. We used HRNet as our backbone. Unlike the original structure of HRNet, we only downsampled the image 3 times to increase the clarity of the feature map. The CARAFE module was used for upsampling in the decoder to help restore the predicted position and shape.

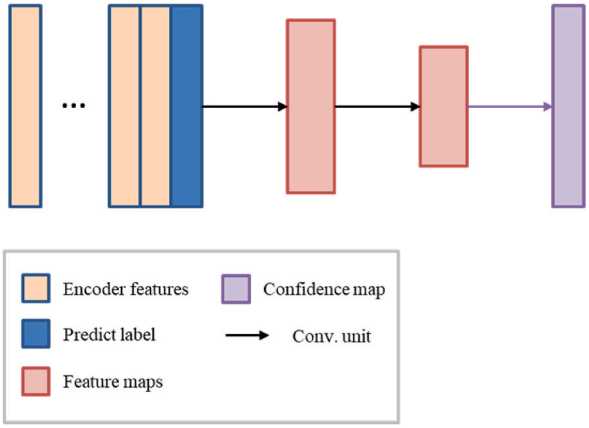

Fig. 4. Discriminator network structure. Five convolutional layers were used. The output was a confidence map, which could help to determine whether the predicted mask of an unlabeled PE image was correct.

of the smallest scale size in the 4th stage was only 25 × 25 (pixels), the boundary and location features of PE were found to be blurred by the visualized feature map, so the decoder could not easily improve the prediction accuracy effectively. Therefore, when the 4th stage was removed, the computing cost was greatly reduced but the accuracy remained. The third modification was to build the CARAFE module as the upsampling layer of the decoder. Compared with bilinear interpolation, the CARAFE module predicted the upsampling kernel for each pixel in the feature map so that it could better restore the shape and position of the PE.

3.1.2. Discriminator network

The discriminator network was the core of adversarial semi-supervised learning. The discriminator network was used to generate the confidence maps from the predicted label and ground truth label for adversarial learning. Two loss functions were used to train the discriminator. Loss_D was used to differentiate the ground truth confidence map from the predicted confidence map. Loss_adv was used to calculate the similarity of the ground truth confidence map with the predicted confidence map. By minimizing the sum of loss functions, the discriminator was able to make the predicted label closer to the ground truth map. Thus, the discriminator could revise the predicted mask during training with unlabeled PE images. As shown in Fig. 4, the discriminator consisted of three convolutional layers and an upsampling layer. As the input of the discriminator, the feature map generated by the encoder of the segmentation network was concatenated with the prediction mask, the weight of which was set to 5. Therefore, the discriminator focused on both the positional information in the mask and PE features. Due to our proposed discriminator, the mask of unlabeled PE images could achieve higher accuracy.

3.2. Post-processing

To ensure that the prediction results were inside the lung, we added two further image processes after the output of our semi-supervised segmentation model: a lung area mask and boundary lines. For obtaining the lung mask, the CTPA image HU threshold was set to −160 to binarize the image. Then, in the middle of the cropped area, a pixel with a value of 0 was randomly selected as the starting point for region growing. Finally, the lung mask resulting from the region growing was used as a filter for the predicted lung mask, with which we could focus on PE lesion detection in the lung area. In addition, since the CTPA images were obtained from different hospitals, the PE images had different sizes, brightness values, and displacements. Thus, the boundary lines method was also used. We first binarized the image by setting the HU threshold to −160 and found the top and bottom boundaries of the lungs. Then the areas outside of the boundary lines were removed from the predicted mask. The boundary line method was used after the lung area mask methods. By adding the two correction processes, we were able to ensure that most of our predicted masks were inside the lung area.

4. Experiments and results

Our semi-supervised segmentation model was trained with a labeled dataset and an unlabeled one. The prediction accuracy of the unlabeled dataset was increased after the training. Our model was also applied to other PE datasets. The results showed that only a small amount of unlabeled data (tens of patients) was required for fine-tuning and that the prediction accuracy of the unlabeled data was greatly improved by our method. In this study, we used two different datasets to train the model and one external dataset to evaluate the model. The mIOU, dice score, and sensitivity were used for evaluation. Furthermore, our model was compared with the 2D and 2.5D U-Net models proposed by Carlos et al., which used only supervised learning.

4.1. Training and evaluation datasets

This study used three different CTPA datasets: the open source dataset, the National Cheng Kung University Hospital (NCKUH) dataset, and the China Medical University Hospital (CMUH) dataset. (I) The open source dataset was provided by Mojtaba et al. [15]. In all, 33 patients were included in the dataset; 27 patients, in the training set; and 6 patients, in the validation set. All the data were labeled images. This dataset was used to train the supervised segmentation model in the pre-training stage. (II) The second dataset, with 70 patients, was provided by NCKUH. The NCKUH dataset included 63 patients in the training set (unlabeled images) and 7 patients in the validation set (labeled images). The NCKUH dataset and the open source dataset were jointly used to train our semi-supervised segmentation model, and the two validation sets were used to evaluate the trained model. (III) The third dataset, with 27 patients, was provided by CMUH and was used as the external testing set in this study. We used a small number of CMUH unlabeled images (from 5 patients) to fine-tune our model and then used the remaining images from 22 patients (labeled images) as a testing set. The results were compared with ground truth labels manually annotated by a physician. Both the NCKUH dataset and the CMUH dataset were composed of axial images from chest CT scans performed according to a pulmonary angiography protocol. The annotations of the two datasets were reviewed by board-certificated radiologists and contoured by a board-certificated radiation oncologist. This study was approved by the institutional review boards of NCKUH and CMUH.

4.2. Implementation details

In this study, the training process of the semi-supervised segmentation model was divided into two steps. First, we trained the supervised segmentation model. Then we used the parameters of the trained model as the pre-trained parameters for the semi-supervised model. When training the supervised segmentation model, we used only the open source dataset, which was labeled images. In total, 100 epochs were used in the training. The loss functions were binary cross entropy (BCE) loss and dice loss. The optimizer was stochastic gradient descent (SGD) [16] with momentum of 0.9. The initial learning rate was set to 1e-4. For semi-supervised segmentation, both the open source dataset and the NCKUH dataset were used during training. In the segmentation network, BCE loss and dice loss with SGD optimization were also used. The initial learning rate was set to 2e-4. In the discriminator network, BCE loss was used, and the initial learning rate was set to 1e-4 due to its simple architecture. The code and model are available at https://github.com/zach0306/PE_avgSemiseg.git.

4.3. Open source dataset supervised model results

The pre-trained model for semi-supervised learning was trained with the open source dataset. To demonstrate the ability of our segmentation model, we compared our proposed segmentation model with other baseline models, which used only supervised learning. In this study, we tested the basic segmentation models, including 2D U-Net [17], 2.5D U-Net, DeepLabV3+ [18], and HRNet, on the open source dataset. The results are listed in Table 1. The 2D U-Net model performed better than the 2.5D U-Net model on our dataset, which was consistent with the results of Carlos et al. The HRNet performed better on PE images because it retained a high resolution architecture for convolutional operations and integrated feature maps of various scales. We further improved our HRNet-based model by increasing the input size (HRNet Large), adding the CARAFE module (HRNet Large/CARAFE), and removing the 4th stage of HRNet (HRNet Large/CARAFE/without stage4). The resulting mIOU, dice score, and sensitivity achieved 0.4801, 0.6018, and 0.6801, respectively. Since the 4th stage was removed from HRNet, the performance of our model was slightly lower than that of HRNet with the 4th stage in supervised learning. However, the HRNet without the 4th stage achieved higher computing efficiency during the semi-supervised learning described in section 3.4. The visualization of the predicted results is shown in Fig. 5. When the lesion area of PE was small, the prediction was usually larger than the ground truth. Hence, the false positives lowered the accuracy significantly. Nevertheless, false positives are not a major concern in clinical diagnosis. Thus, we also evaluated our model by dice score and sensitivity, which focused on the true positive area. The high true positives showed that our model correctly annotated the PE lesions most of the time.

Table 1
Supervised segmentation results of open source dataset. HRNet (Large/CARAFE) achieved the best mIOU, but HRNet (Large/CARAFE/without stage4) had fewer parameters and was more suitable for semi-supervised learning.

Supervised segmentation (Open source dataset)			
Segmentation model	Validation set MIOU	Dice score	Sensitivity
2D U-Net	0.3433	0.4168	0.4831
2.5D U-Net	0.3378	0.4061	0.4712
DeepLav3+	0.3044	0.3707	0.4455
HRNet	0.4218	0.5168	0.5750
Our (HRNet large)	0.4611	0.5877	0.6735
Our (HRNet large/CARAFE)	**0.4801**	**0.6018**	0.6810
Our (HRNet large/CARAFE/without stage4)	0.4683	0.5938	**0.6856**

4.4. Semi-supervised results

We tested the supervised learning model on the NCKUH and CMUH datasets. The results showed that the supervised learning model overly relied on the PE features of the training dataset, so the performances on these two new datasets were not ideal. To adapt our model to different datasets, we added the unlabeled PE images from NCKUH for semi-supervised learning. Furthermore, we also used an additional CMUH dataset to fine-tune and test our semi-supervised model. In this section, we used three datasets to evaluate our semi-supervised segmentation model and compared the differences between supervised learning and semi-supervised learning in predicting unlabeled images. Through semi-supervised learning, our model achieved great improvements in accuracy on different datasets by the use of fine-tuning with a small number of unlabeled images.

4.4.1. Open source dataset semi-supervised model results

In the present study, we used HRNet (Large/CARAFE/without stage4) as the backbone of the semi-supervised segmentation network. We also tested the semi-supervised models with different segmentation backbones. As shown in Table 2, the HRNet-based models achieved higher accuracy than the U-Net and DeepLabV3+ models. Since the PE lesions were small, a higher-resolution convolutional operation was needed to extract the PE features. Compared with the original HRNet, our proposed model enhanced the feature extraction capability for small objects such that the accuracy was greatly improved. To increase the PE mask identification ability of the discriminator, we concatenated feature maps to the predicted labels and ground truth labels and used them as the input of the discriminator. Although the mIOU was not improved significantly, the dice score and the sensitivity were. As shown in Fig. 6, after the addition of feature maps, the model was better able to predict the shapes and boundaries. These results showed that our feature-enhanced methods could increase true positives and improve both the dice score and sensitivity.

The semi-supervised learning results and supervised learning results for the open source dataset are compared in Table 3. The mIOU and the dice score of the semi-supervised learning method slightly decreased, but the sensitivity significantly improved. The result showed that semi-supervised learning increased the number of true positives. However, the semi-supervised model tended to predict larger PE lesion areas, which led to an increase in false positives. Therefore, the sensitivity was much higher than the mIOU and dice score. From comparing the results of supervised learning and semi-supervised learning on the open dataset, we found that semi-supervised learning provided limited improvement on the training dataset, but it improved the boundary prediction and thereby improved the true positive rate.

4.4.2. NCKUH dataset semi-supervised model results

We used the NCKUH dataset as the unlabeled training dataset. To evaluate our semi-supervised model on unlabeled images, we segmented a validation set from the NCKUH dataset to evaluate both the supervised model and the semi-supervised model. In addition, the NCKUH dataset was not used for fine-tuning during the testing of the supervised model. As shown in Table 4, if the supervised model was not fine-tuned for other datasets, the results were not ideal. However, through semi-supervised training, the accuracy on unlabeled images was significantly improved by only a small number of additional unlabeled images. Using our feature-enhanced adversarial semi-supervised segmentation model, the resulting mIOU, dice score, and sensitivity reached 0.3510, 0.4854 and 0.4253, respectively. The results showed that semi-supervised learning improved the accuracy on unlabeled images due to the addition of the unlabeled dataset to the training. As shown in Fig. 7, the vascular imaging of the NCKUH dataset was more obvious than that of the open source dataset. Therefore, the supervised model trained on the open source dataset was unable to perform well on the NCKUH dataset, while the semi-supervised learning solved the problem of over-fitting to a specific dataset in supervised learning.

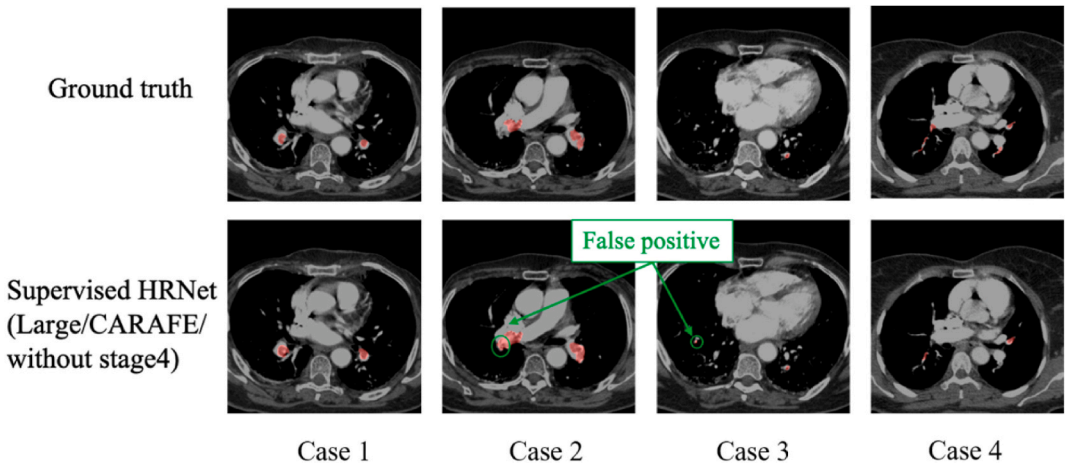

Fig. 5. The predicted images of the supervised segmentation in the open-source dataset are presented in the first and second columns. The red regions in the first column represent the ground truth, while those in the second column represent our predicted label. While cases 1 and 4 demonstrate accurate predictions, cases 2 and 3 show false positive results.

Table 2
Semi-supervised results of the open source dataset. The mIOU, dice score, and sensitivity were improved after the addition of feature maps to the predicted labels and ground truth labels.

Semi-Supervised segmentation (Open source dataset)			
Segmentation model	Validation set MIOU	Dice score	Sensitivity
U-Net	0.2810	0.3418	0.4927
DeepLav3+	0.2513	0.3188	0.4205
HRNet	0.3142	0.4323	0.5764
Our (HRNet large/CARAFE/without stage4)	0.4548	0.5588	0.7056
Our (HRNet large/CARAFE/without stage4) + feature maps	**0.4549**	**0.5611**	**0.7180**

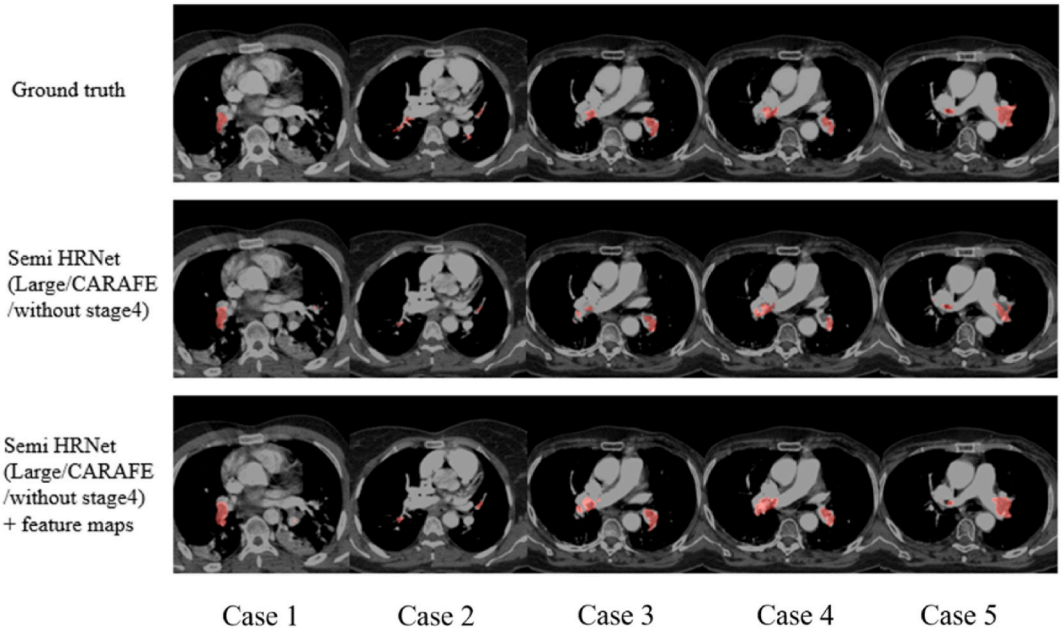

Fig. 6. Predicted images of semi-supervised segmentation in the open-source dataset are shown in the first row of the figure. The cases displayed in the column represent a selection of results to illustrate the performance of different models design. The second row shows the results obtained without the addition of feature maps before the discriminator. The third row displays the results obtained with the addition of feature maps before the discriminator, which reduced the number of false negatives.

Table 3
Comparison between supervised and semi-supervised methods on the open source dataset. The higher sensitivity showed that the semi-supervised method yielded more true positives.

Open source dataset				
Segmentation backbone		Validation set mIOU	Dice score	Sensitivity
Supervised segmentation	HRNet (large/CARAFE/without stage4)	**0.4683**	**0.5938**	0.6856
Semi-Supervised segmentation	HRNet (large/CARAFE/without stage4)	0.4548	0.5588	0.7056
	HRNet (large/CARAFE/without stage4) + feature maps	0.4549	0.5611	**0.7180**

Table 4
Comparison between supervised and semi-supervised models on the NCKUH dataset. Through semi-supervised learning, the accuracy on unlabeled data was significantly improved.

NCKUH dataset				
Segmentation backbone		Testing set mIOU	Dice score	Sensitivity
Supervised segmentation	HRNet (large/CARAFE/without stage4)	0.2128	0.3488	0.3169
Semi-Supervised segmentation	HRNet (large/CARAFE/without stage4)	0.3386	0.4597	0.4088
	HRNet (large/CARAFE/without stage4) + feature maps	**0.3510**	**0.4854**	**0.4253**

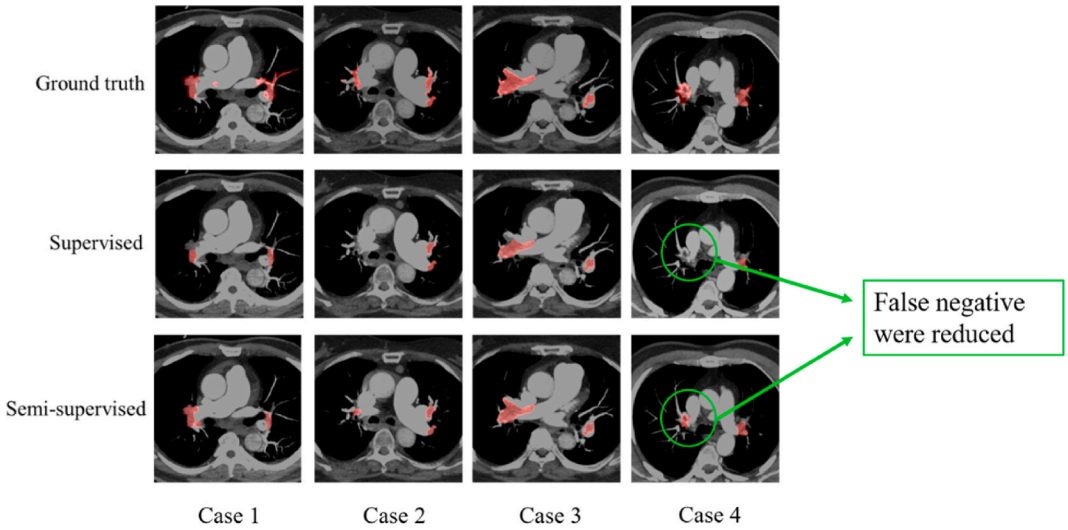

Fig. 7. Predicted images of segmentation results in the NCKUH dataset. The results of semi-supervised training on the NCKUH dataset show a reduction in false negatives compare with supervised training, as observed in the various cases presented in the columns.

Table 5
Comparison between supervised and semi-supervised methods on the CMUH dataset. The accuracy was improved through fine-tuning the semi-supervised model with a small number of CMUH images.

CMUH dataset				
Segmentation backbone		Testing set mIOU	Dice score	Sensitivity
Supervised segmentation	HRNet (large/CARAFE/without stage4)	0.2344	0.3325	0.3151
Semi-Supervised segmentation	HRNet (large/CARAFE/without stage4) + feature maps	**0.3721**	**0.5113**	**0.4967**

4.4.3. CMUH dataset semi-supervised model results

To verify whether our semi-supervised model could be applied to different datasets, we used the CMUH dataset for additional testing. We added only a small number of CMUH unlabeled images (from 5 patients) to fine-tune the model. As shown in Table 5, through fine-tuning of the semi-supervised learning, the resulting mIOU, dice score, and sensitivity improved from 0.2344, 0.3325, and 0.3151 to 0.3721, 0.5113, and 0.4967, respectively. As shown in Fig. 8, the semi-supervised model labeled the PE lesion area more completely than the supervised model did. This result showed that our semi-supervised model can work well on other unlabeled datasets.

5. Discussion

In this work, we first designed a supervised segmentation network which was suitable for the PE dataset. Unlike the models of Carlos et al., our HRNet-based model was more suitable for small object detection. Therefore, our supervised model was able to achieve better dice scores and sensitivity. Then, to better fit our model to different PE datasets, we designed a feature-enhanced adversarial semi-supervised segmentation model. We found that the semi-supervised model did not significantly improve the accuracy on the testing set of the open source dataset, as the improvement was limited due to the use of the same training set in the supervised learning. Therefore, our semi-supervised model only slightly revised the predicted boundaries and shapes. However, our semi-supervised learning provided a significant improvement in the accuracy of the NCKUH and CMUH datasets. Because of the addition of feature maps to the predicted label and ground truth label as input to the discriminator for training, our model was able to mark the shapes of the PE lesion areas more accurately and thus increased the accuracy.

The sensitivity was higher than the dice score only on the open source dataset. The reason was that the predicted label of the open source dataset contained more false positives, which are unfavorable for the calculation of a dice score. However, the predicted labels of other datasets rarely contained false positives, so the dice score was slightly higher than the sensitivity. This difference could be ascribed mainly to the open source dataset and other datasets being sourced from different labeling personnel with dissimilar labeling standards. The open source dataset had stricter standards for the PE lesion area, and only the obstructed area was annotated. In the other datasets, all areas around the obstructed area were annotated. Therefore, the true positives were increased and the false positives were decreased in the NCKUH and CMUH datasets.

The prediction results of unlabeled images (NCKUH and CMUH datasets) are shown in Fig. 9. The complex shapes and the large number of PE lesion areas resulted in the decrease in accuracy. Our model was able to predict only the approximate locations in these images and was unable to fully present the details, resulting in a large number of false negative pixels. However, the approximate

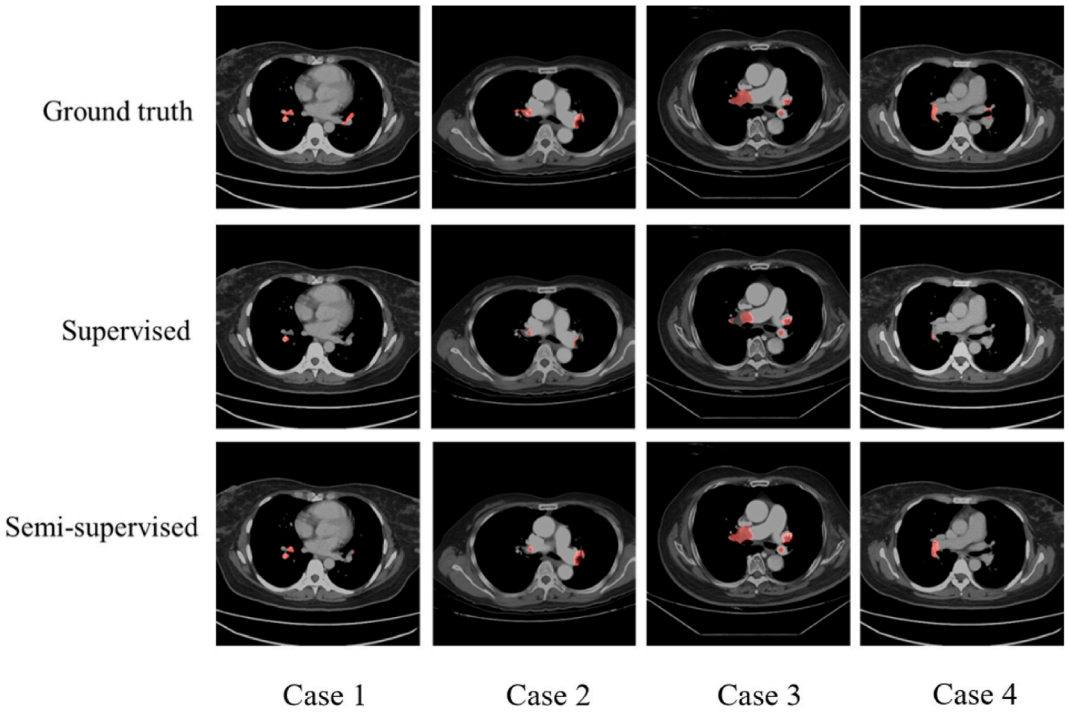

Fig. 8. Predicted images of segmentation results in the CMUH dataset show that fine-tuning the semi-supervised model with images from five patients improved its ability to fit the shapes of the pulmonary embolism (PE) lesion area, as illustrated in the different cases presented in the columns.

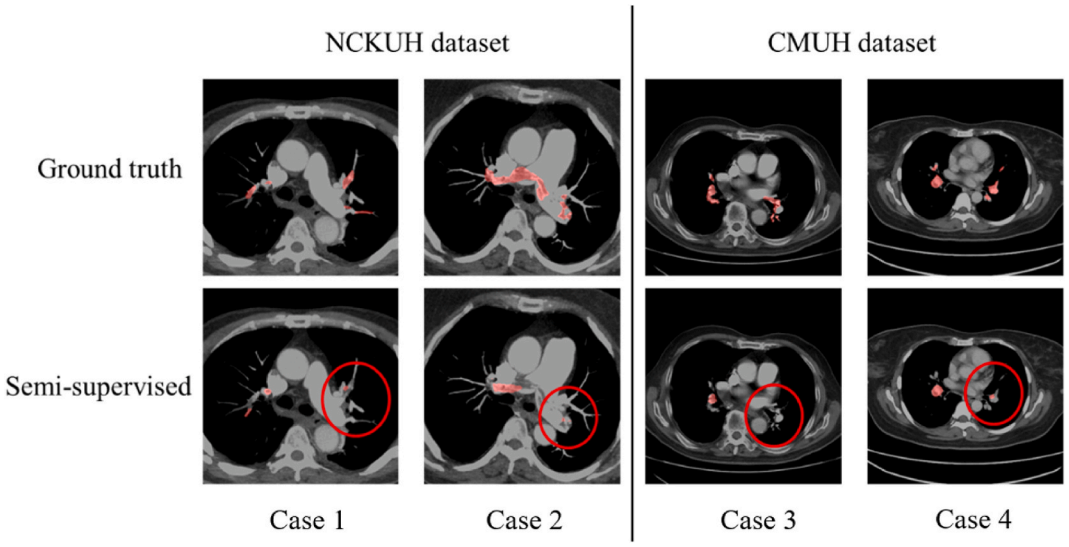

Fig. 9. We compared the segmentation results of our semi-supervised model on the NCKUH and CMUH datasets, and selected representative cases to highlight the presence of false negative pixels in the predicted labels. These pixels, indicative of missed PE lesions, can reduce the accuracy of the model.

locations we predicted were sufficiently helpful for doctors to check the PE lesion areas. Therefore, despite the large number of false negative pixels, our model can still assist doctors in quickly finding PE lesion areas.

The clinical limitations of CT angiography are impacted by selection bias, as the cases that currently undergo this test usually involve high-risk patients with high d-dimer levels and a strong suspicion of pulmonary embolism. Consequently, the rate of positive cases among those who undergo CTA is high. However, there may be a considerable number of hidden cases of asymptomatic pulmonary embolism, in which patients show no symptoms and therefore do not undergo CTA. Unfortunately, there is no data available

on the percentage of cases with asymptomatic pulmonary embolism. This lack of data limits the model's ability to make accurate judgments, which can lead to false negative readings in mild cases.

6. Conclusion

In this work, we proposed a semi-supervised semantic segmentation model for PE lesion annotation. In the semantic segmentation network, we increased the size of the input feature map to preserve the features of PE. We also modified the oversized HRNet model to reduce the computational load while maintaining high accuracy. The other novel aspect of our model is the addition the feature map information to the discriminator by using the feature map of the encoder to improve the accuracy.

In the open source dataset, our supervised model outperformed the model of Carlos et al. The resulting mIOU, dice score, and sensitivity reached 0.4683, 0.5938, and 0.6856, respectively. Furthermore, the addition of a small number of unlabeled PE images for fine-tuning of the model improved the mIOU, dice score, and sensitivity of the NCKUH dataset to 0.3510, 0.4854, and 0.4253. The resulting mIOU, dice score, and sensitivity of the CMUH dataset increased to 0.3721, 0.5113, and 0.4967, respectively. Superior to the supervised learning models, our semi-supervised model achieves higher accuracy in predicting unlabeled images. The results also showed that our semi-supervised model can be applied to other PE datasets.

For CTPA images obtained from different hospitals, we can fine-tune our model by adding a small number of additional unlabeled images from another hospital and achieve an accuracy equivalent to the accuracy of the training dataset from the original hospital. Although the accuracy of our semi-supervised learning model may still be improved, we have demonstrated a promising training method that can be easily applied to multiple datasets. While our model demonstrated promising results, we recognize the need for further improvements. We plan to implement image pre-processing techniques to enhance the detail features of blood vessels and incorporate attention modules to improve the model's ability to distinguish between the background and target.

Author contribution statement

Ting-Wei Cheng: Performed the experiments.

Yi Wei Chua, Jerry Chang: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Ching-Chun Huang: Conceived and designed the experiments; Analyzed and interpreted the data.

Chin Kuo: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data.

Yun-Chien Cheng: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

Prof. YunChien Chien Cheng was supported by Ministry of Science and Technology, Taiwan {111-2221-E-A49 -045 -MY3}

Data availability statement

The authors do not have permission to share data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S.C. Huang, T. Kothari, I. Banerjee, C. Chute, R.L. Ball, N. Borus, M.P. Lungren, PENet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging, NPJ digital medicine 3 (1) (2020) 61, https://doi.org/10.1038/s41746-020-0266-y.

[2] X. Yang, Y. Lin, J. Su, X. Wang, X. Li, J. Lin, K.T. Cheng, A two-stage convolutional neural network for pulmonary embolism detection from CTPA images, IEEE Access 7 (2019) 84849–84857.

[3] N. Tajbakhsh, M.B. Gotway, J. Liang, Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Springer International Publishing, Munich, Germany, 2015, pp. 62–69.

[4] K. Long, L. Tang, X. Pu, Y. Ren, M. Zheng, L. Gao, F. Deng, Probability-based Mask R-CNN for pulmonary embolism detection, Neurocomputing 422 (2021) 345–353.

[5] C. Cano-Espinosa, M. Cazorla, G. González, Computer aided detection of pulmonary embolism using multi-slice multi-axial segmentation, Appl. Sci. 10 (8) (2020) 2945.

[6] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, B. Xiao, Deep high-resolution representation learning for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 43 (10) (2020) 3349–3364.

[7] J. Wang, K. Chen, R. Xu, Z. Liu, C.C. Loy, D. Lin, Carafe: content-aware reassembly of features, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3007–3016.

[8] B. Li, Y. Wang, T. Che, S. Zhang, S. Zhao, P. Xu, K. Keutzer, Rethinking Distributional Matching Based Domain Adaptation, 2020 arXiv preprint arXiv: 2006.13352.

[9] W.M. Kouw, M. Loog, A review of domain adaptation without target labels, IEEE Trans. Pattern Anal. Mach. Intell. 43 (3) (2019) 766–785.

[10] D.H. Lee, Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks, in: Workshop on Challenges in Representation learning(ICML) 3, 2013, p. 896.
[11] H. Zheng, L. Lin, H. Hu, Q. Zhang, Q. Chen, Y. Iwamoto, J. Wu, Semi-supervised segmentation of liver using adversarial learning with deep atlas prior, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22, Springer International Publishing, 2019, pp. 148–156.
[12] D. Nie, Y. Gao, L. Wang, D. Shen, ASDNet: attention based semi-supervised deep networks for medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11, Springer International Publishing, 2018, pp. 370–378.
[13] Y. Zhou, X. He, L. Huang, L. Liu, F. Zhu, S. Cui, L. Shao, Collaborative learning of semi-supervised segmentation and classification for medical images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2079–2088.
[14] W.C. Hung, Y.H. Tsai, Y.T. Liou, Y.Y. Lin, M.H. Yang, Adversarial Learning for Semi-supervised Semantic Segmentation, 2018 arXiv preprint arXiv:1802.07934.
[15] M. Masoudi, H.R. Pourreza, M. Saadatmand-Tarzjan, N. Eftekhari, F.S. Zargar, M.P. Rad, A new dataset of computed-tomography angiography images for computer-aided detection of pulmonary embolism, Sci. Data 5 (1) (2018) 1–9.
[16] S. Ruder, An Overview of Gradient Descent Optimization Algorithms, 2016 arXiv preprint arXiv:1609.04747.
[17] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Springer International Publishing, Munich, Germany, 2015, pp. 234–241.
[18] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV), 2018, pp. 801–818.