



FDA: Feature Decomposition and Aggregation for Robust Airway Segmentation

Minghui Zhang¹, Xin Yu³, Hanxiao Zhang¹, Hao Zheng¹, Weihao Yu¹,
Hong Pan², Xiangran Cai³, and Yun Gu^{1,4}(✉)

¹ Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China
geron762@sjtu.edu.cn

² Department of Computer Science and Software Engineering,
Swinburne University of Technology, Victoria, Australia

³ Medical Image Centre, The First Affiliated Hospital of Jinan University,
Guangzhou, China

⁴ Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai, China

Abstract. 3D Convolutional Neural Networks (CNNs) have been widely adopted for airway segmentation. The performance of 3D CNNs is greatly influenced by the dataset while the public airway datasets are mainly clean CT scans with coarse annotation, thus difficult to be generalized to noisy CT scans (e.g. COVID-19 CT scans). In this work, we proposed a new dual-stream network to address the variability between the clean domain and noisy domain, which utilizes the clean CT scans and a small amount of labeled noisy CT scans for airway segmentation. We designed two different encoders to extract the transferable clean features and the unique noisy features separately, followed by two independent decoders. Further on, the transferable features are refined by the channel-wise feature recalibration and Signed Distance Map (SDM) regression. The feature recalibration module emphasizes critical features and the SDM pays more attention to the bronchi, which is beneficial to extracting the transferable topological features robust to the coarse labels. Extensive experimental results demonstrated the obvious improvement brought by our proposed method. Compared to other state-of-the-art transfer learning methods, our method accurately segmented more bronchi in the noisy CT scans.

Keywords: Clean and noisy domain · Decomposition and aggregation · Airway segmentation

1 Introduction

The novel coronavirus 2019 (COVID-19) has turned into a pandemic, infecting humans all over the world. To relieve the burden of clinicians, many researchers take the advantage of deep learning methods for automated COVID-19 diagnosis and infection measurement from imaging data (e.g., CT scans, Chest X-ray).

© Springer Nature Switzerland AG 2021

S. Albarqouni et al. (Eds.): DART 2021/FAIR 2021, LNCS 12968, pp. 25–34, 2021.

https://doi.org/10.1007/978-3-030-87722-4_3

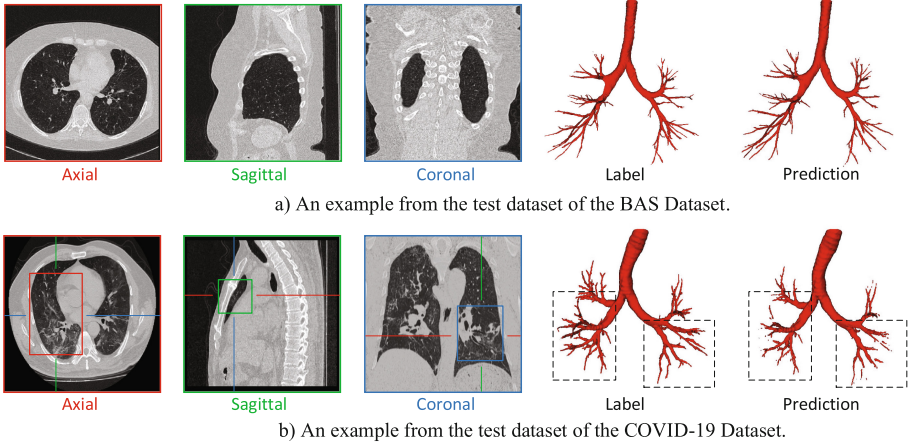


Fig. 1. Different patterns exist between the BAS dataset and the COVID-19 dataset. One well-trained segmentation model in the clean domain leads to low accuracy when testing in the noisy domain.

Current studies mainly focus on designing a discriminative or robust model to distinguish COVID-19 from other patients with pneumonia [11, 17], lesion localization [18], and segmentation [16]. In this work, we tackle another challenging problem, airway segmentation of COVID-19 CT scans. The accurate segmentation enables the quantitative measurements of airway dimensions and wall thickness which can reveal the abnormality of patients with pulmonary disease and the extraction of patient-specific airway model from CT image is required for navigation in bronchoscopic-assisted surgery. It helps the sputum suction for novel COVID-19 patients.

However, due to the fine-grained pulmonary airway structure, manual annotation is time-consuming, error-prone, and highly relies on the expertise of clinicians. Moreover, COVID-19 CT scans share ground-glass opacities in the early stage and pulmonary consolidation in the late stage [3] that adds additional difficulty for annotation. Even though the fully convolutional networks (FCNs) could automatically segment the airway, there remain the following challenges. First, FCNs are data-driven methods, while there are few public airway datasets with annotation and the data size is also limited. The public airway datasets, including EXACT’09 dataset [8] and the Binary Airway Segmentation (BAS) dataset [12], focus on the cases with the abnormality of airway structures mainly caused by chronic obstructive pulmonary disease (COPD). These cases are relatively clean and we term their distribution as **Clean Domain**, on the contrary, we term the distribution of COVID-19 CT scans as **Noisy Domain**. Figure 1 shows that fully convolutional networks (FCNs) methods [5, 12] trained on the clean domain cannot be perfectly generalized to the noisy domain. Although this challenge can be addressed via the collection and labeling of new cases, it is impractical for novel diseases, e.g. COVID-19, which cannot guarantee the scale of datasets and

the quality of annotation. Second, transfer learning methods (e.g. domain adaptation [2, 13], feature alignment [2, 14]) can improve the performance on target domains by transferring the knowledge contained in source domains or learning domain-invariant features. However, these methods are inadequate to apply in our scenario because this target noisy domain contains specific features (e.g. patterns of shadow patches) which cannot be learned from the source domain. Third, the annotation of the airway is extremely hard as they are elongated fine structures with plentiful peripheral bronchi of quite different sizes and orientations. The annotation in the EXACT’09 dataset [8] and the BAS dataset [12] are overall coarse and unsatisfactory. However, the deep learning methods are intended to fit the coarse labels, and thereby they are difficult to learn the robust features for airway representation.

To alleviate such challenges, we propose a dual-stream network to extract the robust and transferable features from the clean CT scans (clean domain) and a few labeled COVID-19 CT scans (noisy domain). Our contributions are threefold:

- We hypothesize that the COVID-19 CT scans own the general features and specific features for airway segmentation. The general features (e.g. the topological structure) are likely to learn from the other clean CT scans, while the specific features (e.g. patterns of shadow patches) should be extracted independently. Therefore, we designed a dual-stream network, which takes both the clean CT scans and a few labeled COVID-19 CT scans as input to synergistically learn general features and independently learn specific features for airway segmentation.
- We introduce the feature calibration module and the Signed Distance Map (SDM) for the clean CT scans with coarse labels, and through this way, robust features can be obtained for the extraction of general features.
- With extensive experiments on the clean CT scans and the COVID-19 CT scans, our method revealed the superiority in the extraction of transferable and robust features and achieved improvement compared to other methods under the evaluation of tree length detected rate and the branch detected rate.

2 Method

A new dual-stream network is proposed, which simultaneously processes the clean CT scans and a few noisy COVID-19 CT scans to learn robust and transferable features for airway segmentation. In this section, we detail the architecture of the proposed dual-stream network, which is illustrated in Fig. 2.

2.1 Learning Transferable Features

COVID-19 CT scans share a similar airway topological structure with the clean CT scans, meanwhile introduce unique patterns, e.g., multifocal patchy shadowing and ground-glass opacities, which are not common in clean CT scans. Since

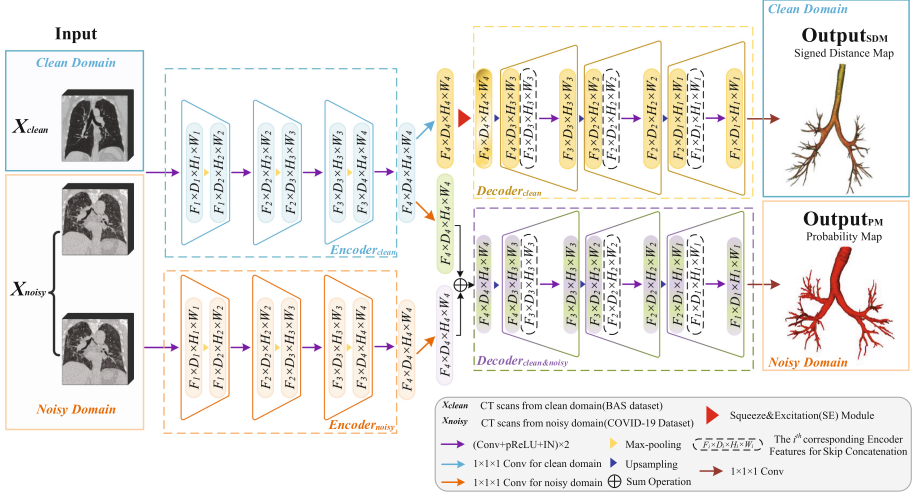


Fig. 2. Detailed structure and workflow of the proposed dual-stream network. $Encoder_{clean}$ aims to synergistically learn transferable features from both X_{clean} and X_{noisy} . $Encoder_{noisy}$ extracts specific features of X_{noisy} . The encoded features of X_{clean} are fed into $Decoder_{clean}$, where SE and SDM modules refine the features. For X_{noisy} , the decomposed features are then aggregated again via the channel-wise summation operation and fed into $Decoder_{clean\&noisy}$.

the number of clean CT scans is relatively large and the airway structure is also clearer, we aim to adapt the knowledge from clean CT scans to improve the airway segmentation of COVID-19 CT scans. Therefore, a dual-stream network is designed to synergistically learn transferable features from both COVID-19 CT scans and clean CT scans, and independently learn specific features only from COVID-19 CT scans. As illustrated in Fig. 2, let X_{clean} denote the input of sub-volume CT scans from the clean CT scans, X_{noisy} from the noisy COVID-19 CT scans. $Encoder_{clean}$ and $Encoder_{noisy}$ are encoder blocks for feature extraction. $Decoder_{clean}$ and $Decoder_{clean\&noisy}$ are decoder blocks to generate the segmentation results based on the features from encoders. The $Output_{SDM}$ represents the output of clean CT scans, and the $Output_{PM}$ represents the output of COVID-19 CT scans, they can be briefly defined as follows:

$$Output_{SDM} = Decoder_{clean}(Encoder_{clean}(X_{clean}))$$

$$Output_{PM} = Decoder_{clean\&noisy}(Encoder_{clean}(X_{noisy}) + Encoder_{noisy}(X_{noisy}))$$

where we omit the detail of $1 \times 1 \times 1$ convolution and the Squeeze&Excitation (SE) module for a straightforward explanation of the overall workflow. In this case, the features of X_{noisy} are decomposed into two parts: $Encoder_{clean}$ aims to extract high-level, semantic and transferable features from both clean CT scans and COVID-19 CT scans; $Encoder_{noisy}$ is designed to obtain the specific features which belongs to the COVID-19 samples. The features of clean CT

scans extracted by $Encoder_{clean}$ are fed into $Decoder_{clean}$. For COVID-19 CT images, the decomposed features are then aggregated again via the channel-wise summation operation and fed into $Decoder_{clean\&noisy}$ to reconstruct the volumetric airway structures.

2.2 Refinement of Transferable Features

As is mentioned before, the annotation of the public airway dataset is overall coarse and unsatisfactory. Since we have determined which features to transfer, then the transferable features can be further refined to be more robust through feature recalibration and introducing signed distance map.

Feature Recalibration: 3D Channel SE (cSE) module [20] is designed to investigate the channel-wise attention. We embed this module between $Encoder_{clean}$ and $Decoder_{clean}$, aiming to refine the transferable features. Take \mathbf{U} as input and $\tilde{\mathbf{U}}$ as output, $\mathbf{U}, \tilde{\mathbf{U}} \in \mathbb{R}^{F \times D \times H \times W}$ with the number of channels F , depth D , height H , width W . 3D cSE firstly compresses the spatial domain then obtains channel-wise dependencies $\tilde{\mathbf{Z}}$, which are formulated as follows:

$$Z_i = \frac{1}{D} \frac{1}{H} \frac{1}{W} \sum_{d=1}^D \sum_{h=1}^H \sum_{w=1}^W U_i(d, h, w), \quad i = 1, 2, \dots, F \quad (1)$$

$$\tilde{\mathbf{Z}} = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{Z})), \quad (2)$$

where $\delta(\cdot)$ denotes the ReLU function and $\sigma(\cdot)$ refers to sigmoid activation, $\mathbf{W}_1 \in \mathbb{R}^{\frac{F}{r_c} \times F}$, and $\mathbf{W}_2 \in \mathbb{R}^{F \times \frac{F}{r_c}}$. The r_c represents the reduction factor in the channel domain, similar to [4]. The output of 3D cSE is obtained by: $\tilde{\mathbf{U}} = \mathbf{U} \odot \tilde{\mathbf{Z}}$.

Signed Distance Map: In recent years, introducing the distance transformed map into CNNs have proven effectivity in medical image segmentation task [6, 10, 19] due to its superiority of paying attention to the global structural information. The manual annotation of the plentiful tenuous bronchi is error-prone and often be labeled thinner or thicker. The 3D FCNs cooperating with common loss function treat the labeled foreground equally and intend to fit such coarse labels, which are difficult to extract robust features. Even though the thickness of the annotated bronchi is uncertain, the phenomenon of breakage or leakage in the annotation can be avoided by experienced radiologists. Therefore, the overall topology is correctly delineated, and we can use the topological structure instead of the coarse label as a supervised signal. Besides, the intra-class imbalance problem in airway segmentation is severe. Distance transform map is used to rebalance the distribution of trachea, main bronchi, lobar bronchi, and distal segmental bronchi. We use the signed distance map transform as a voxel-wise reweighting method, incorporating with the regression loss that focuses on the relatively small values (such as the lobar bronchi and distal segmental bronchi) by having larger gradient magnitudes.

Given the airway as target structure and each voxel x in the volume set X , we construct the Signed Distance Map (SDM) function termed as $\phi(x)$, defined as:

$$\phi(x) = \begin{cases} 0, & x \in \text{airway and } x \in \mathcal{C}, \\ -\inf_{z \in \mathcal{C}} \|x - z\|_2, & x \in \text{airway and } x \notin \mathcal{C}, \\ +\inf_{z \in \mathcal{C}} \|x - z\|_2, & x \notin \text{airway}, \end{cases} \quad (3)$$

where the \mathcal{C} represents the surface of the airway, we further normalize the SDM into $[-1, +1]$. We then transformed the segmentation task on clean CT scans to an SDM regression problem and introduce the loss function that penalizes the prediction SDM for having the wrong sign and forces the 3D CNNs to learn more robust features that contain topological features for airway. Denote the y_x as the ground truth of SDM and f_x as the prediction of the SDM, the loss function for the regression problem can be defined as follows:

$$L_{reg} = \sum_{\forall x} \|f_x - y_x\|_1 - \sum_{\forall x} \frac{f_x y_x}{f_x y_x + f_x^2 + y_x^2}, \quad (4)$$

where $\|\cdot\|_1$ denotes the L_1 norm.

2.3 Training Loss Functions

The training loss functions consist of two parts, The first part is the L_{reg} for the clean CT scans, and the second part is the L_{seg} for the noisy CT scans, we combine the Dice [9] and Focal loss [7] to construct the L_{seg} :

$$L_{seg} = -\frac{2 \sum_{\forall x} p_x g_x}{\sum_{\forall x} (p_x + g_x)} - \frac{1}{|X|} \left(\sum_{\forall x} (1 - p_x)^2 \log(p_x) \right), \quad (5)$$

where g_x is the binary ground truth and p_x is the prediction. The total loss is defined as $L_{total} = L_{seg} + L_{reg}$.

3 Experiments and Results

Dataset: We used two datasets to evaluate our method.

- Clean Domain: Binary Airway Segmentation (BAS) dataset [12]. It contains 90 CT scans (70 CT scans from LIDC [1]) and 20 CT scans from the training set of the EXACT'09 dataset [8]. The spatial resolution ranges from 0.5 to 0.82 mm and the slice thickness ranges from 0.5 to 1.0 mm. We randomly split the 90 CT scans into the training set (50 scans), validation set (20 scans), and test set (20 scans).
- Noisy Domain: COVID-19 dataset. We collected 58 COVID-19 patients from three hospitals and the airway ground truth of each COVID-19 CT scan was corrected by three experienced radiologists. The spatial resolution of the COVID-19 dataset ranges from 0.58 to 0.84 mm and slice thickness varies from 0.5 to 1.0 mm. The COVID-19 dataset is randomly divided into 10 scans for training and 48 scans for testing.

Table 1. Results (%) on the test set of the COVID-19 dataset. Values are shown as mean \pm standard deviation. ‘B’ indicates the training set of the BAS dataset and ‘C’ indicates the training set of the COVID-19 dataset.

Method		Length	Branch	DSC
3D UNet	Train on B only	72.4 \pm 4.8	62.1 \pm 4.5	93.2 \pm 1.5
	Train on B + C	82.8 \pm 4.8	83.8 \pm 3.8	95.2 \pm 1.3
	Train on C only	85.7 \pm 5.1	84.9 \pm 3.5	95.9 \pm 1.2
	Train on B, finetuned on C	86.8 \pm 5.3	85.0 \pm 4.1	95.7 \pm 1.1
3D UNet + cSE (Medical Physics,2019) [20]		86.2 \pm 5.3	84.6 \pm 4.2	95.8 \pm 1.2
Feature Alignment (TMI,2020) [2]		87.9 \pm 4.9	85.5 \pm 4.8	95.5 \pm 1.6
Domain Adaptation (TPAMI,2018) [13]		87.0 \pm 4.6	84.9 \pm 4.0	96.0 \pm 1.3
Proposed w/o cSE&SDM		90.2 \pm 5.3	87.6 \pm 4.2	96.5 \pm 1.2
Proposed w/o cSE		91.1 \pm 4.3	86.8 \pm 3.7	96.8 \pm 1.0
Proposed w/o SDM		91.0 \pm 4.7	86.3 \pm 4.1	96.6 \pm 1.0
Proposed		92.1 \pm 4.3	87.8 \pm 3.7	96.8 \pm 1.1

Network Configuration and Implementation Details: As shown in Fig. 2, each block in the encoder or decoder contains two convolutional layers followed by pReLU and Instance Normalization [15]. The initial number of channel is set to 32, thus $\{F_1, F_2, F_3, F_4\} = \{32, 64, 128, 256\}$. During the preprocessing procedure, we clamped the voxel values to $[-1200, 600]$ HU, normalized them into $[0, 255]$, and cropped the lung field to remove unrelated background regions. We adopted a large input size of $128 \times 224 \times 304$ CT cubes densely cropped near airways and chose a batch size of 1 (randomly chose a clean CT scan and noisy COVID-19 CT scan) in the training phase. On-the-fly data augmentation included the random horizontal flipping and random rotation between $[-10^\circ, 10^\circ]$. All models were trained by Adam optimizer with the initial learning rate of 0.002. The total epoch is set to 60 and the learning rate was divide by 10 in the 50th epoch, the hyperparameter of r_c used in 3D cSE module is set to 2. Preliminary experiments confirmed training procedures converged under this setup. In testing phase, we performed the sliding window prediction with stride 48. All the models were implemented in PyTorch framework with a single NVIDIA Geforce RTX 3090 GPU (24 GB graphical memory).

Evaluation Metrics: We adopted three metrics to evaluate methods, including the a) tree length detected rate (Length) [8], b) branch detected rate (Branch) [8], and c) Dice score coefficient (DSC). All metrics are evaluated on the largest component of each airway segmentation result.

Quantitative Results: Experimental results showed that the way of training on the BAS dataset then evaluating on the COVID-19 dataset performed worst, as expected. Training merely on the COVID-19 dataset performed better than training on both the BAS dataset and the COVID-19 dataset, which implied the necessity of transfer learning rather than merely together different datasets.

3D UNet with cSE [20] was trained on the COVID-19 dataset and the results showed no significant improvement. For comparison, three commonly used transfer learning methods, Fine-tuned (pre-trained on BAS dataset, fine-tuned on COVID-19 dataset), Feature Alignment (FA) [2] through adversarial training, and Domain Adaptation (DA) by sharing weights [13] were reimplemented to be applied in our task, the results in Table 1 demonstrated our proposed method is superior to these methods, the proposed method achieved the highest performance on all metrics of Length (92.1%), Branch (87.8%), and DSC (96.8%). We also conducted the ablation study to investigate the effectiveness of each component of the proposed method. In Table 1, we observed that the original dual-stream network had outperformed the other methods, with the achievement of 90.2% Length, 87.6% Branch, and 96.5% DSC. The improvement confirmed the validity of our proposed dual-stream network. Furthermore, cSE module and SDM could boost performance independently and the combination of cSE and SDM brings the highest performance gain, which demonstrated the necessity of refinement for transferable features.

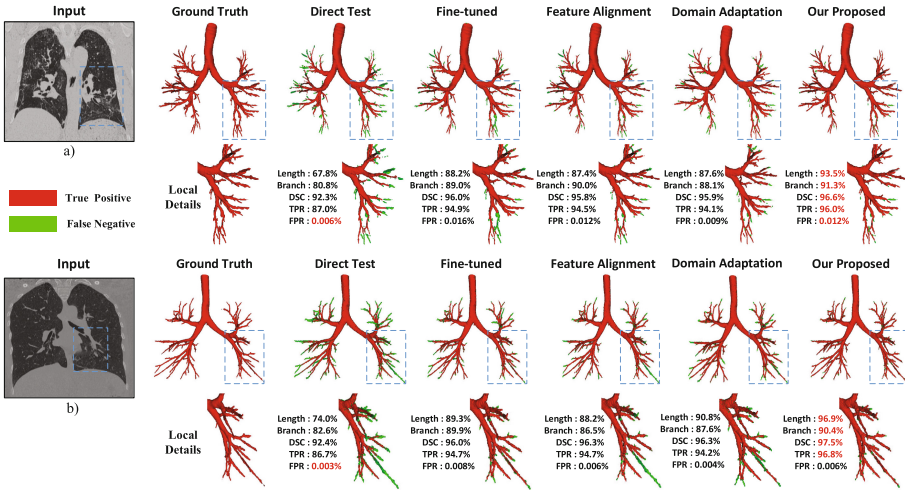


Fig. 3. Visualization of segmentation results. a) is a severe case and b) is a mild case in the test set of the COVID-19 dataset. The blue dotted boxes indicate the local regions full of shadow patches and are zoomed in for better observation.

Qualitative Results: The visualization of segmentation results is presented in Fig. 3. Compared to other methods, the proposed method gains improvement on both the severe and mild cases of the COVID-19 dataset, which accurately detected more bronchi surrounded by multifocal patchy shadowing of COVID-19.

4 Conclusion

This paper proposed a novel dual-stream network to learn transferable and robust features from clean CT scans to noisy CT for airway segmentation. Our proposed method not only extracted the transferable clean features but also extract unique noisy features separately, transferable features were further refined by the cSE module and SDM. Extensive experimental results showed our proposed method accurately segmented more bronchi than other methods.

Acknowledgements. This work was partly supported by National Natural Science Foundation of China (No. 62003208), National Key R&D Program of China (No. 2019YFB1311503), Shanghai Sailing Program (No. 20YF1420800), Shanghai Municipal of Science and Technology Project (Grant No. 20JC1419500), and Science and Technology Commission of Shanghai Municipality under Grant 20DZ2220400.

References

1. Armato, S.G., II., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on ct scans. *Med. Phys.* **38**(2), 915–931 (2011)
2. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Trans. Med. Imag.* **39**(7), 2494–2505 (2020)
3. Chung, M., et al.: Ct imaging features of 2019 novel coronavirus (2019-ncov). *Radiology* **295**(1), 202–207 (2020)
4. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141 (2018)
5. Juarez, A.G.U., Tiddens, H.A., de Bruijne, M.: Automatic airway segmentation in chest CT using convolutional neural networks. In: *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pp. 238–250. Springer (2018)
6. Karimi, D., Salcudean, S.E.: Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Trans. Med. Imag.* **39**(2), 499–513 (2019)
7. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988 (2017)
8. Lo, P., et al.: Extraction of airways from CT (exact’09). *IEEE Trans. Med. Imag.* **31**(11), 2093–2107 (2012)
9. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571. IEEE (2016)
10. Navarro, F., et al.: Shape-aware complementary-task learning for multi-organ segmentation. In: Suk, H.-I., Liu, M., Yan, P., Lian, C. (eds.) *MLMI 2019*. LNCS, vol. 11861, pp. 620–627. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32692-0_71
11. Ouyang, X., et al.: Dual-sampling attention network for diagnosis of Covid-19 from community acquired pneumonia. *IEEE Trans. Med. Imag.* **39**(8), 2595–2605 (2020)

12. Qin, Y., et al.: Airwaynet: a voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 212–220. Springer (2019)
13. Rozantsev, A., Salzmann, M., Fua, P.: Beyond sharing weights for deep domain adaptation. *IEEE Trans. Patt. Analy. Mach. Intell.* **41**(4), 801–814 (2018)
14. Sun, B., Saenko, K.: Deep CORAL: correlation alignment for deep domain adaptation. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9915, pp. 443–450. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49409-8_35
15. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: the missing ingredient for fast stylization. arXiv preprint [arXiv:1607.08022](https://arxiv.org/abs/1607.08022) (2016)
16. Wang, G., et al.: A noise-robust framework for automatic segmentation of Covid-19 pneumonia lesions from CT images. *IEEE Trans. Med. Imag.* **39**(8), 2653–2663 (2020)
17. Wang, J., et al.: Prior-attention residual learning for more discriminative Covid-19 screening in CT images. *IEEE Trans. Med. Imag.* **39**(8), 2572–2583 (2020)
18. Wang, X., et al.: A weakly-supervised framework for Covid-19 classification and lesion localization from chest CT. *IEEE Trans. Med. Imag.* **39**(8), 2615–2625 (2020)
19. Xue, Y., et al.: Shape-aware organ segmentation by predicting signed distance maps. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 12565–12572 (2020)
20. Zhu, W., et al.: Anatomynet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med. Phys.* **46**(2), 576–589 (2019)