Fuzzy Attention Neural Network to Tackle Discontinuity in Airway Segmentation

Yang Nan, Javier Del Ser, *Senior Member, IEEE*, Zeyu Tang, Peng Tang, Xiaodan Xing, Yingying Fang, Francisco Herrera, *Senior Member, IEEE*, Witold Pedrycz, *Life Fellow, IEEE*, Simon Walsh, and Guang Yang, *Senior Member, IEEE*

***Abstract*— Airway segmentation is crucial for the examination, diagnosis, and prognosis of lung diseases, while its manual delineation is unduly burdensome. To alleviate this time-consuming and potentially subjective manual procedure, researchers have proposed methods to automatically segment airways from computerized tomography (CT) images. However, some small-sized airway branches (e.g., bronchus and terminal bronchioles) significantly aggravate the difficulty of automatic segmentation by machine learning models. In particular, the variance of voxel values and the severe data imbalance in airway branches make the computational module prone to discontinuous and false-negative predictions, especially for cohorts with different lung diseases. The attention mechanism has shown the capacity to segment complex structures, while fuzzy logic can reduce the uncertainty in feature representations. Therefore, the integration of deep attention networks and fuzzy theory, given by the fuzzy attention layer, should be an escalated solution for better generalization and robustness. This article presents an efficient method for airway segmentation, comprising a novel fuzzy attention neural network (FANN) and a comprehensive loss function to enhance the spatial continuity of airway segmentation. The deep fuzzy set is formulated by a set of voxels in the feature map and a learnable Gaussian membership function. Different from the existing attention mechanism, the proposed channel-specific fuzzy attention addresses the issue of heterogeneous features in different channels. Furthermore, a novel evaluation metric is proposed to assess both the continuity and completeness of airway structures. The efficiency, generalization, and robustness of the proposed method have been proved by training on normal lung disease while testing on datasets of lung cancer, COVID-19, and pulmonary fibrosis.**

***Index Terms*— Airway segmentation, COVID-19, fuzzy attention, fuzzy neural networks, pulmonary fibrosis.**

Manuscript received 9 September 2022; revised 29 November 2022 and 8 March 2023; accepted 19 April 2023. This work was supported in part by the Basque Government under Grant KK-2020/00049, in part by the Consolidated Research Group MATHMODE under Grant IT1294-19, in part by the National Natural Science Foundation of China under Grant 62076182, in part by the British Heart Foundation (BHF) under Grant TG/18/5/34111 and Grant PG/16/78/32402, in part by the European Research Council (ERC) Innovative Medicines Initiative (IMI) under Grant 101005122, in part by H2020 under Grant 952172, in part by the Medical Research Council (MRC) under Grant MC/PC/21013, in part by the Royal Society under Grant IEC/NSFC/211235, in part by the NVIDIA Academic Hardware Grant Program, in part by the SABER Project supported by Boehringer Ingelheim Ltd., in part by the National Institute for Health and Care Research (NIHR) Imperial Biomedical Research Center (RDA01), and in part by the UK Research and Innovation (UKRI) Future Leaders Fellowship under Grant MR/V023799/1. *(Corresponding authors: Simon Walsh; Guang Yang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Royal Brompton and Harefield Foundation Trust under Application No. 18/LO/1392.

Yang Nan, Zeyu Tang, Xiaodan Xing, and Yingying Fang are with the National Heart and Lung Institute, Imperial College London, SW7 2BX London, U.K.

Javier Del Ser is with the Department of Communications Engineering, University of the Basque Country UPV/EHU, 48940 Bilbao, Spain, and also with the TECNALIA, Basque Research and Technology Alliance (BRTA), 20850 Derio, Spain.

Peng Tang is with the Department of Informatics, Technical University of Munich, 80333 Munich, Germany.

Francisco Herrera is with the Department of Computer Sciences and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, Granada, Spain, and also with the Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada, also with the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abgudulaziz University, Jeddah, Saudi Arabia, and also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland.

Simon Walsh and Guang Yang are with the National Heart and Lung Institute, Imperial College London, SW7 2BX London, U.K., and also with the Royal Brompton Hospital, SW3 6NP London, U.K. (e-mail: s.walsh@imperial.ac.uk; g.yang@imperial.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2023.3269223.

Digital Object Identifier 10.1109/TNNLS.2023.3269223

I. Introduction

SEVERITY of traction bronchiectasis strongly predicts the outcome in suspected lung disease (e.g., progressive pulmonary fibrosis) [1], [2], [3]. However, visual grading of disease on high-resolution computed tomography (HRCT) is liable to significant interobserver variability, poor reproducibility, and is relatively insensitive to subtle but clinically important changes over short follow-up periods [4]. This provides the rationale for developing objective computer-based methods for disease quantification [5], [6], [7].

The feasibility of using conventional methods (e.g., morphology, intensity thresholding [8], region growing [9], [10], [11], fuzzy connectedness [12]) for airway extraction was reported in a previous challenge EXACT-09 [13]. However, these approaches were shown to suffer from weak robustness and coarse predictions due to the high complexity of bronchus patterns. In addition to conventional approaches, deep neural networks (e.g., nnUNet [14] and VNet [15]) have become a common solution for airway segmentation due to their high modeling capacity. Although great efforts have been made in airway segmentation [16], [17], [18], [19], [20], these methods still suffered from low continuity and completeness of airway predictions (as shown in Fig. 1).

Considering the airway segmentation, there mainly exist five major challenges (Fig. 1): 1) the imbalanced data

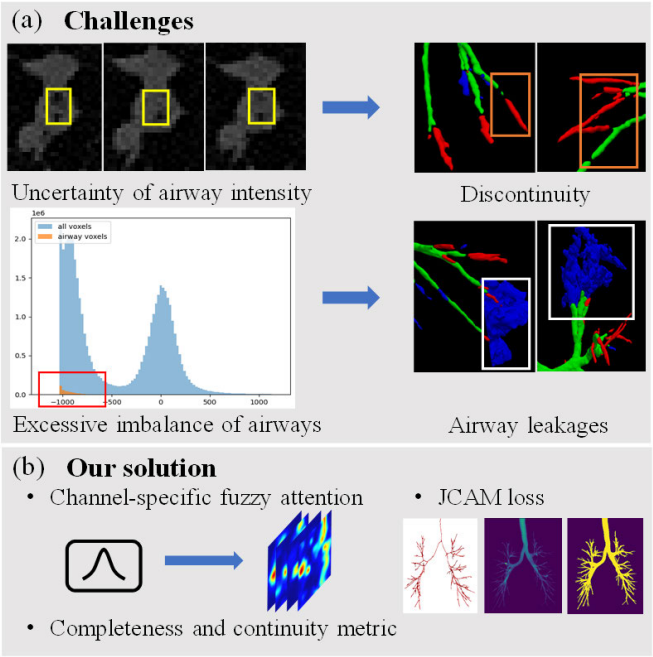

Fig. 1. Current challenges and our solution of airway segmentation. (a) Heterogeneity (yellow boxes) of the intensity of airway structures and the excessive data imbalance (red box), which leads to discontinuity predictions (orange boxes) and leakages (white box). The green, red, and blue voxels represent true positive predictions, missing airways, and airway leakages, respectively. (b) This article presents a FANN to help the module to better learn the robustness feature, a JCAM loss for optimization, and a novel metric that assesses the completeness and continuity of airway predictions.

issues between foreground and background samples, and that between the trachea and terminal branches. These imbalanced issues significantly aggravate the difficulty of training a data-driven-based module and make the terminal branches hard to be discovered. Although most neural networks were trained on patch datasets, the criteria for patch extraction were not explored. For instance, the previous studies simply extracted patches through sliding windows, while ignoring other sampling strategies; 2) the inter-class homogeneity of voxel values between airway trees and normal lung regions. The voxel values of lung regions and airway trees are similar, which requires the network to learn more spatial and structural information in addition to voxel intensities; 3) the intra-class heterogeneity of voxel values within airway trees caused by incorrect labeling and imaging issue. The inevitable incorrect annotations and the variances of voxel values in airway regions require a strong robustness module; 4) the weak reliability of evaluation metrics for module selection. Most researchers employed intersection over union (IoU) or dice coefficient score as the metric to save the best model [21]. Unfortunately, the imbalanced voxel volume of the main trachea and small branches makes the overlapped metrics unreliable; and 5) lack of generalization and robustness during clinical practice. Most modules suffer from weak generalization and robustness while requiring further fine-tuning for different cohorts of different diseases [21].

The attention mechanism [22] has shown a high capacity to segment complex structures by weighting the "importance" of different feature representations. However, the intra-class heterogeneity (caused by inaccurate annotation) and inter-class homogeneity (caused by similar physical patterns between lungs and airways) could introduce lots of "uncertainty" to the feature of interest. It has been reported that the uncertainty in feature representations can be effectively reduced by implementing fuzzy logic, and significant improvements have been observed [23], [24], [25]. Therefore, we hypothesize that the combination of deep attention mechanism and fuzzy theory is able to tackle the homogeneity and heterogeneity of airway voxel values. In addition, to enhance the continuity of the segmentation predictions, the center lines of airway branches were integrated with overlapped metrics in loss calculation. However, the overlapped centerline metrics mainly focused on completeness whereas could not well assess the continuity. As a result, we hypothesize the loss function should be modified by considering the continuity by evaluating the projection maps between 3-D masks and predictions.

This article presents an efficient scheme for accurate airway segmentation, including: 1) a smart patch sampling (SPS) strategy; 2) a novel channel-specific fuzzy attention layer; 3) a hybrid continuity and completeness loss function; and 4) a novel metric AF-score (airway f-score). Comprehensive experimental studies were conducted over an open-access dataset (90 cases) and our in-house test-only datasets (40 cases), including normal scans, patients with mild anomalies, lung cancer, fibrosis, and COVID-19. The performance was assessed by calculating the completeness (dice coefficient score) and continuity (detected length and branch ratio) of airway trees. In addition, the radius of each branch in the airway trees is estimated for a better understanding of the disease etiology, and the preview of segmentation results provides a straight comparison.

The main contributions of this article can be summarized as follows.

1) An SPS strategy is proposed to select appropriate patches for training deep models to address the data imbalanced issues between airway trees and lung tissues as well as the trachea and terminal branches.
2) A novel channel-specific fuzzy attention layer by incorporating attention mechanisms and fuzzy logic to alleviate the "uncertainty" caused by the intra-class heterogeneity and inter-class homogeneity. Different from the previous methods that integrated fuzzy logic with deep learning through sequential learning paradigms (e.g., multimodule embedding or feature representation fusion), this article incorporates fuzzy theory and attention mechanism as a fuzzy attention layer. The attention map in the fuzzy attention layer is formulated by a set of voxels in the feature map and a learnable Gaussian membership function, which is channel-specific, compared with the current channel-homogeneous attention mechanism. This fuzzy module can help the network better focus on the relevant feature representations while reducing their uncertainty to improve the module's generalization and robustness capacity. To the best of our knowledge, this is the first attempt to adopt fuzzy logic in the attention layer.

3) A hybrid loss function including airway continuity loss and accumulation mapping loss is proposed to enhance the continuity and completeness of terminal branches. The airway continuity loss assesses the airway center lines, whereas the accumulation mapping loss assesses the 3-D prediction by estimating the errors of mappings from different axes.
4) A novel evaluation metric continuity and completeness F-score (CCF score) to assess the continuity and completeness of module selection. This new metric can tackle the limitation that the branch continuity cannot be reflected through the commonly used dice coefficient score.
5) Comprehensive evaluations including out-of-distribution data assessments were given in terms of different lung diseases (pulmonary fibrosis cases, lung cancer, and COVID-19), which can better reflect the generalization ability of different approaches.

The rest of this article is organized as follows. The related works in airway segmentation and fuzzy logic are summarized in Section II. Details of the proposed method are illustrated in Section III. The experimental settings and results are described in Section IV. Sections V and VI present the discussion and conclusion of this study.

II. Related Works

A. Airway Segmentation in HRCT

Segmentation of airway trees has always been a challenging task, due to the stated issues (in the introduction). For example, the results in EXACT'09 demonstrated that no algorithm was capable to extract more than an average of 74% of the total length of all branches in the reference. To obtain a better performance, Charbonnier et al. [16] designed a 2-D CNN to detect airway leakages for post-processing. Yun et al. [17] presented a 2.5-D CNN that takes three adjacent slices in each orthogonal direction (axial, sagittal, coronal) as inputs to increase the spatial information. In addition to 2-D and 2.5-D architectures, 3-D CNNs [18], [19], [20] became a popular solution due to the better consistency and integrity of their prediction.

Some researchers regarded them as a class imbalance problem and proposed approaches to reduce interference of background regions. For instance, both Garcia-Uceda et al. [26] and Qin et al. [27] cropped the full-size volumes to a bounding box based on the pre-segmented lung region to remove the irrelevant background regions. Juarez et al. [18] applied weighted cross-entropy loss to balance the foreground and background samples. Zheng et al. [28] trained the network on sampled patches in the first stage and selected hard-segment regions based on false-negative predictions. Then, the network was fine-tuned in the second stage to further improve the performance.

Others aim to improve the prediction results by addressing discontinuous issues and airway leakages. For example, Qin et al. [19] proposed a 3-D UNet with voxel-connectivity awareness to segment airways. Meng et al. [29] and Reynisson et al. [30] improved the continuity of predictions by learning the airway centerlines. Nadeem et al. [31] proposed a 3-D UNet with freeze-and-grow propagation to reduce airway leakages and improve the accuracy of detected branches. Zheng et al. [28] analyzed and explained the inaccurate segmentation issues by gradient erosion and dilation of adjacent voxels. Wang et al. [32] presented a radial distance loss to enhance the completeness of tubular predictions.

However, even with these efforts, discontinuous predictions still exist, especially for those small airway branches. Moreover, most of these previous methods require multistage training protocols or complex post-processing procedures, which aggravate the computational and time costs in practical use.

B. Fuzzy Logic in Deep Semantic Segmentation

In addition to the well-known fuzzy clustering approaches [33], [34] for segmentation, there were many earlier attempts to combine deep neural networks and fuzzy logic through sequential learning protocol. For instance, using neural networks to extract low-dimensional feature representations and applying fuzzified representations for classification [35]. Unfortunately, these multistage approaches cannot be trained as an end-to-end scheme and did not aggregate fuzzy theory into the training of deep neural networks.

To address this issue, some researchers integrated fuzzy logic with deep neural networks [23] and proposed an end-to-end hierarchical scheme for deep fuzzy neural networks. With these efforts, researchers have presented approaches to combining fuzzy logic with pure CNNs for segmentation [36], [24], [37]. For instance, Price et al. [36] proposed a flexible and capable fuzzy layer to utilize the powerful aggregation of fuzzy integrals. Huang et al. [37] proposed a fuzzy fully convolutional network for breast ultrasound image segmentation by transforming the data into the fuzzy domain. Guan et al. [24] proposed a fuzzy CNN for lip segmentation, introducing the fuzzy logic module to enhance the robustness of segmentation results. Ding et al. [38] presented fuzzy-enabled multiscale learning to segment brain tumors from T1 and T2 scans simultaneously. However, most existing methods adopted fuzzy integral within a residual block for information fusion, which cannot address the main challenges of airway segmentation (shown in the experiment section). Besides, only the "AND" aggregation operator was applied to membership functions in most previous studies without exploring other operators.

C. Evaluation Metrics of Airway Predictions

Given the semantic prediction X and ground truth annotation Y, the main evaluation metrics for airway segmentation include overlapped metrics, detected branch ratio (DBR), detected length ratio (DLR), and airway leakage ratio (ALR), which are given as follows.

1) *Overlapped metrics:* Measure the percentage overlap between the prediction and ground truth, including dice coefficient score and Jaccard index score (or IoU score),

which are represented as

$$\text{Jaccard}(X, Y) = \frac{XY}{X + Y - XY} \tag{1}$$

$$\text{Dice}(X, Y) = \frac{2XY}{X + Y} \tag{2}$$

where XY are calculated by dot product of X and Y.

2) *DBR:* Measures the proportion of detected branches with respect to the ground truth annotations

$$\text{DBR} = \frac{N_x}{N_y} \times 100\% \tag{3}$$

where N_X is the number of branches that have been correctly recognized, N_Y is the number of branches in the ground truth annotation. In this study, branches with the IoU score greater than 0.8 are referred to be correctly identified.

3) *DLR:* Measures the proportion of detected branch length with respect to that of the ground truth annotations

$$\text{DLR} = \frac{L_x}{L_y} \times 100\% \tag{4}$$

where L_X is the total length of branches that have been correctly recognized, L_Y is the total length of branches in the ground truth annotation.

4) *ALR:* refers to the proportion of total false-positive volumes with respect to the ground truth annotations

$$\text{ALR} = \frac{V_x}{V_y} \times 100\% \tag{5}$$

where V_X is the volume of false-positive predictions, and V_Y is the volume of ground truth annotation.

Although the performance of airway segmentation can be evaluated through these metrics, none of them can be used independently for assessment.

III. Methodology

This section illustrates details of the proposed scheme for airway segmentation, comprising the overview of FANN, the SPS, a fuzzy attention layer, a Jaccard continuity and accumulation mapping (JCAM) loss, and a CCF score.

A. Overview

Given an input raw 3-D volume $V \in \mathbb{R}^{W \times H \times Z}$, an SPS strategy is first presented to extract appropriate patches $V_P \in \mathbb{R}^{w \times h \times c}$ for training and validation. Then, a fuzzy attention neural network (FANN) is proposed to segment the airway trees with the supervision of a comprehensive loss function (including dice loss, airway continuity loss, and accumulation mapping loss) on V_P. During the training, the AF-score is developed to assess the model performance and save the best values of the trainable network parameters.

The proposed FAAN is built based on the 3-D U-Net by adding a novel fuzzy attention layer, deep supervision, and airway continuity and accumulation mapping (ACAM) loss. The size of the volume patches V_P is set as the same ratio as the median shape of the ground truth annotations in the training data. FANN includes 3-D convolution layers, instance normalization (IN), LeakyReLu (LReLU), 3-D transpose convolution layers, fuzzy attention layers, and sigmoid activation layers, as shown in Fig. 2. It is of note that the proposed network has multiple outputs, including the main output and three low-level outputs. The final prediction is given from the main output, while the low-level outputs are collected for deep supervision through the auxiliary losses (losses are calculated in the same way as that of the main output with different weights). When calculating these auxiliary losses, the prediction is upsampled to align with the size of ground truth masks.

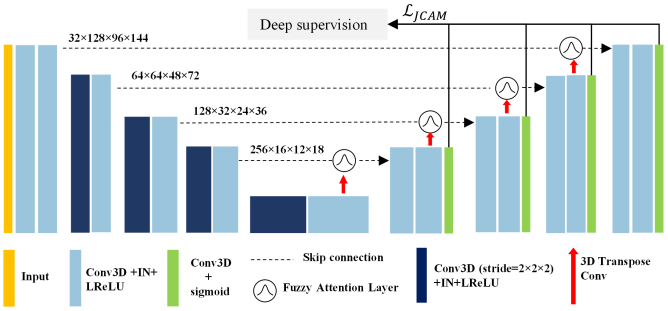

Fig. 2. FANN. The fuzzy attention layer is adopted within the skip connection between the encoders and decoders. Each decoder block produces a binary prediction for deep supervision. Sizes of feature maps are presented for a better understanding of the architecture of FANN.

B. Smart Patch Sampling

Deep learning is a data-driven method that highly relies on a large amount of data, thus a balanced class distribution is crucial for training a neural network. The huge imbalance of foreground and background samples are prone to gradient erosion issue [28]. Due to the large volume array in HRCT and GPU memory limitations, segmentation modules in HRCT are mainly based on patch training strategies. However, most studies simply extracted patches through sliding windows, which cannot completely address the class imbalanced issue. Although some researchers alleviated class imbalance by extracting hard samples [28], [39], these methods required a further training process to collect hard samples. In this study, we present SPS criteria for patch extraction that do not require further training/fine-tuning operations. Given the dataset that comprises 3-D images and their corresponding annotations, the workflow of SPS can be summarized as follows.

1) The average size S ($z \times y \times x$) of the 3-D minimum bounding volume of ground truth annotations is first calculated. The patch size is set as the same scale as S.
2) The centerline of manual annotation is extracted by skeletonization [40].
3) Overlapped sliding windows are adopted to extract image patches, mask patches, and centerline patches.
4) Patches with a centerline ratio (the proportion of the centerline voxels in a patch in terms of overall centerline voxels) larger than 15% or a voxel volume ratio (the proportion of the ground truth voxels in a patch in terms of overall ground truth voxels) larger than 10% are kept.

C. Fuzzy Attention Layer

One challenge toward obtaining a well-trained module for airway segmentation is the uncertainty of annotations and voxel values within the airway region. There have been efforts to make the network focus more on the relevant regions. For instance, the attention U-Net [22] proposes an attention gate to improve accuracy by suppressing feature activations in irrelevant regions. However, we deem that sigmoid activation may not be the best solution to organize the attention gate.

In addition to the "gradient vanishing" problem, one major concern with the sigmoid activation function is its sharpness, by which only a small interval can obtain the output value ranges between 0 and 1. Therefore, it is difficult to find a robust "boundary" in sigmoid activation that distinguishes whether the feature is relevant or not. Another issue is monotonicity: similar to the raw intensity distribution of airway structures that has both negative and positive variations, the feature representation of airways should also have two sides of variations. However, to align with the sigmoid activation function, the 1×1 convolution layers must learn to shift the values of all the features of interest to a single side to obtain a positive response. In particular, there must exist a certain "threshold" in the feature representation reconstitution (usually accomplished by the 1×1 convolutional layers in the attention layer) to determine whether the region is important or not. Moreover, nonchannel specifics of the current attention map assign the same "attention" coefficient to all the feature points along the channel-wise. Specifically, given a feature representation $F \in \mathbb{R}^{C\times W\times H\times D}$, the existing attention map is formulated as $\alpha \in \mathbb{R}^{W\times H\times D}$, while all the feature representations along the channel-wise C shares the same "importance." This mechanism is unreliable since the feature representations in different channels are extracted by different convolution kernels; therefore, we advocate the attention map to be channel-specific.

Studies have shown that both fuzzy logic and neural networks are efficient for data representation [23]. In general, the neural network aims to reduce noises in the original data to extract useful feature representations, whereas fuzzy logic is used to obtain fuzzy representations to reduce the uncertainty in the original data. Unlike the sigmoid activation function, which is fixed and sharp, we believe the Gaussian function is more effective to formulate the attention gate, due to its symmetry and flexibility (the mean and variance of the Gaussian function can be set as learnable parameters). Moreover, compared with other fuzzy membership functions, the Gaussian membership function benefits from its smooth and nonzero at all points, which avoids the gradient vanishing and collapse issues. Therefore, we combine fuzzy logic with the attention mechanism using trainable Gaussian membership functions to better help the segmentation network focus on the relevant region while reducing the uncertainty and variations of the data representations.

The proposed fuzzy attention layer is adopted within the skip connection, taking both feature representations from the l-th encoder layer space e^l and decoder layer space d^l as inputs [shown in Fig. 3(a)]. These two input feature vectors are first processed by a $1 \times 1 \times 1$ 3-D convolution layer,

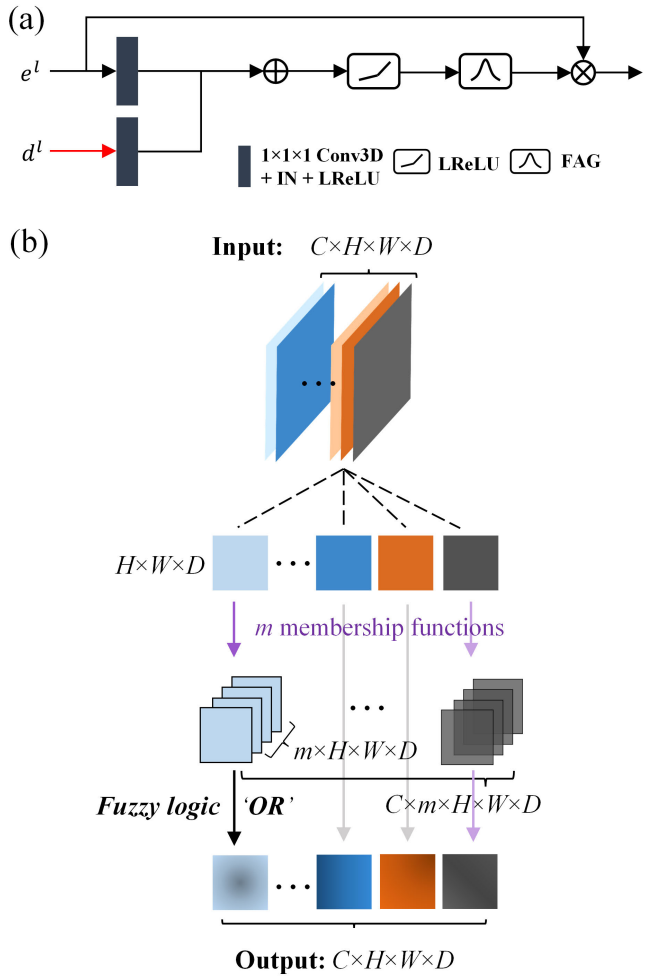

Fig. 3. Details of (a) fuzzy attention layer and (b) fuzzy attention gate (FAG). The fuzzy attention layer takes both feature representations from the lth encoder layer e^l and decoder layer d^l as inputs and outputs of an attention map that has the same dimension as the inputs. The learnable Gaussian membership function is adopted.

an instance normalization, and a LeakyReLU for feature reconstitution. Then, a voxelwise adding operation is adopted to fuse the information, followed by a LeakyReLU. Next, the feature representations are fed into the FAG to generate a voxelwise attention map, shown in Fig. 3(b). Assume X (with a shape of $C \times H \times W \times D$ regardless of batch size) as the input of the fuzzy attention gate. Due to the smoothness and concise notation of Gaussian membership functions, learnable Gaussian MFs are proposed to specify the deep fuzzy sets. Each feature map (with the height H, width W and depth D) is filtered by m Gaussian membership functions with the trainable center $\mu_{i,j}$ and spread $\sigma_{i,j}$

$$f_{i,j}(X,\mu,\sigma) = e^{\frac{-(x_j-\mu_{i,j})^2}{2\sigma_{i,j}^2}} \in \mathbb{R}^{C\times m\times H\times W\times D} \tag{6}$$

where $i = 1,\ldots,m$, $j = 1,\ldots,C$. Different from most previous studies that applied the operators "AND" to obtain fuzzy feature representations, our goal is to use the membership function to learn the "importance" of target feature representations. Therefore, we believe that the information can be better preserved by applying the aggregation operator "OR"

while suppressing irrelevant features. Having two fuzzy sets $\tilde{A}$ and $\tilde{B}$, the operator "OR" is described as

$$f_{\tilde{A}\cup\tilde{B}}(y) = f_{\tilde{A}}(y) \vee f_{\tilde{B}}(y) \quad \forall y \in U \tag{7}$$

where U is the universe of information and y is the element of the universe. To make the operator "OR" derivative, we modified it as

$$f_{\tilde{A}\cup\tilde{B}}(y) = \max(f_{\tilde{A}}(y), f_{\tilde{B}}(y)). \tag{8}$$

Therefore, the fuzzy degree $f_j(X, \mu, \sigma) \in \Theta^{H\times W\times D}, \Theta \in [0, 1]$ of $j-$th channel can be obtained based on (6) and (8) as

$$f_j(X, \mu, \sigma) = \bigvee_{i=1}^{m} e^{\frac{-\left(x_j-\mu_{i,j}\right)^2}{2\sigma_{i,j}^2}} = \max\left(e^{\frac{-\left(x_j-\mu_{i,j}\right)^2}{2\sigma_{i,j}^2}}\right) \tag{9}$$

where the large "V" indicates the union operation.

Eventually, the output tensor of the proposed FAG has the same shape as the input tensors, which can provide a voxelwise attention map. The pseudo-code of the fuzzy attention layer is shown in Algorithm 1.

Algorithm 1 Pseudo Code of Fuzzy Attention Layer

Input: feature representations $e^l, d^l \in \mathbb{R}^{C\times W\times H\times D}$,from lth encoder and decoder,, weights $w_e^{(l)}$, $w_d^{(l)}$ and bias $b_e^{(l)}, b_d^{(l)}$ connecting to lth encoder and decoder

Output: fused feature representation $y \in \mathbb{R}^{C\times W\times H\times D}$

1: randomly initialize parameters $\mu \in \mathbb{R}^{m\times C}$ and $\sigma \in \mathbb{R}^{m\times C}$ of membership functions $f(X, \mu, \sigma)$.
2: **compute** the input $X \in \mathbb{R}^{C\times H\times W\times D}$ of FAG, $X =$ LReLU[LReLU(IN($w_e^{(l)}e^l + b_e^{(l)}$)) + LReLU(IN($w_d^{(l)}d^l + b_d^{(l)}$))]
3: **for** i in C **do**
4: **Specify** the fuzzy sets $f_{i,j}(X, \mu, \sigma)$ through m Gaussian membership functions as (6)
5: **compute** fuzzy degrees $f_j(X, \mu, \sigma)$ of jth channel by adopting aggregation operator "OR" to fuzzy sets $f_{i,j}$ through (9).
6: **compute** output of jth channel $y_j = e^{(l)} f_j(X, \mu, \sigma)$
7: **end for**

D. JCAM Loss

Studies have shown no qualms about applying strategies such as cascade modules or cropping operations to better segment small organs or structures. By implementing these strategies, the inter-class imbalance issue can be well alleviated. However, another issue that the model needs to address is the intra-class imbalance, which is caused by different volumes of different branches (trachea, secondary/tertiary bronchus). The trachea and the secondary bronchus account for the majority of the total airway volume, leading to a discontinuous prediction for the small bronchus.

To bridge this gap, we propose a JCAM loss L_{JCAM} to force the network to pay more attention to the continuity of airway predictions. The L_{JCAM} assesses the error of different projections (through coronal, sagittal, and axial planes) and

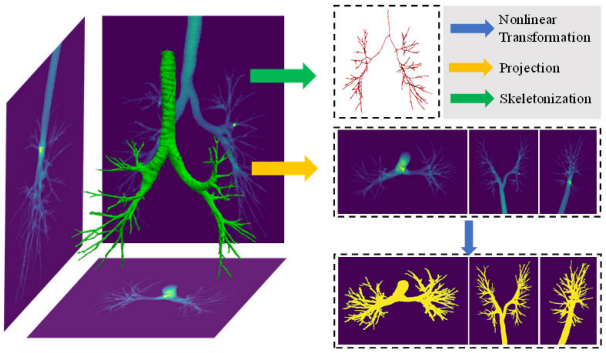

Fig. 4. Skeletonization and accumulation maps (acquired from the projection of coronal, sagittal and transverse planes, and nonlinear transformation) of the airway structure.

centerlines between the prediction and ground truth. Given the prediction array X, the corresponding ground truth annotation Y, the centerlines Y_{CL} of the ground truth masks are extracted by the skeletonization [40]. With the ground truth centerlines Y_{CL}, the continuity can be well evaluated by calculating the ratio of correct-predicted centerlines and ground truth centerlines

$$L_C = 1 - C = 1 - \frac{\sum X \cdot Y_{\text{CL}}}{\sum Y_{\text{CL}}} \tag{10}$$

without considering the airway branch size. In addition to the continuity, we also present two types of accumulation maps, including linear accumulation maps (LAMs) and nonlinear accumulation maps (nLAMs), as shown in Fig. 4. LAMs are calculated by summing the volume array along three different axes, while nLAMs are acquired by performing a nonlinear transformation on LAMs. By integrating LAM into the optimization function, the veracity of the 3-D prediction can be better assessed since each incorrect voxel leads to incorrect variations in different projection maps. Furthermore, to reduce the intra-class imbalance, a nonlinear transformation (tanh activation function) is introduced to LAMs to acquire nLAMs. Although nLAM (ranges from [0, 1]) discards the spatial information within LAM, it can better gauge the continuity and veracity between the predictions and ground truth masks.

Denote $A \oplus \tau$ as the summation operation that sums the array $A \in \mathbb{R}^{N\times C\times W\times H\times D}$ along τth channel. The LAM and nLAM are calculated by

$$L_{\text{LAM}} = \sum_{\tau\in[W,H,D]} l_1(X \oplus \tau, Y \oplus \tau) \tag{11}$$

$$L_{\text{nLAM}} = \sum_{\tau\in[W,H,D]} L_{\mathcal{J}}(\tanh(X \oplus \tau), \tanh(Y \oplus \tau)) \tag{12}$$

where $L_{\mathcal{J}}$ is the Jaccard loss which is defined as

$$L_{\mathcal{J}} = 1 - \mathcal{J} = 1 - \frac{XY + \varepsilon}{X + Y - XY + \varepsilon} \tag{13}$$

and ε is the smoothing factor $\varepsilon > 0$.

Overall, the L_{JCAM} can be summarized as

$$L_{\text{JCAM}} = \alpha L_{\mathcal{J}}(X, Y) + \beta L_C(X, Y_{\text{CL}}) + \varphi L_{\text{CE}}(X, Y) + \gamma L_{\text{LAM}}(X, Y) + \delta L_{\text{nLAM}}(X, Y) \tag{14}$$

where $\alpha, \beta, \gamma, \varphi, \delta$ are the weights of Jaccard loss $L_{\mathcal{J}}$, continuity loss L_C, cross-entropy loss L_{CE}, linear and nonlinear accumulation mapping loss L_{LAM}, L_{nLAM}, respectively.

E. Continuity and Completeness F-Score

In addition to developing novel layers or operations, module selection also plays a crucial role in machine learning. In airway segmentation, it is unreliable to find the best model by selecting the highest regional overlapping metrics, due to: 1) the intra-class imbalance of the airway structures and 2) the imperfection of manual annotations. A high dice or IoU score only reflects the overall overlapping ratio of the predictions voxels, which cannot assess the continuity and completeness. Thus, there is an urgent requirement to develop an effective metric to evaluate the continuity and completeness of tree structures (e.g., airways, blood vessels, neurons, etc.).

Here, we propose a CCF score for the aforementioned purpose

$$\text{CCF}_s = \left(1 + \omega^2\right) \times \frac{J \times C}{\omega^2 \times J + C} \tag{15}$$

where $\omega \in [0, 1]$ is the preference parameter, J and C are the Jaccard index and Continuity index as defined in (10) and (13), respectively. ω can be set larger (smaller) than 1 when $C(J)$ is more important. With this novel F-measure-based metric, the module that focuses on both continuity and completeness can be saved.

IV. Experiments and Results

This section details all the settings of experiments including datasets, implementation details, evaluation metrics, and results.

A. Datasets and Training Strategies

In this study, we trained our airway segmentation and comparison models using 90 clinical thoracic CT scans from the open-access BAS dataset. In particular, the in-house datasets collected from patients with fibrotic lung disease and COVID-19 are used for testing only. Details of all the data are shown in Table I.

1) Binary Airway Segmentation (BAS) Dataset: Overall, there are 90 cases in total for BASD with 20 cases from EXACT'09 and 70 cases from LIDC: 1) EXACT'09 [13] is a dataset that consists of 20 cases for training (with the corresponding annotations) and 20 cases for the test (without annotation publicly available), with the scanned patients ranging from normal conditions to severe pulmonary lung diseases and 2) LIDC dataset: The Lung Image Database Consortium image collection (LIDC-IDRI) includes 1018 cases of diagnostic and lung cancer screening thoracic CT scans [41]. Among the 1018 cases, 70 cases whose slice thickness ≤ 0.625 mm were randomly selected and annotated by experts [27], [28].

2) Our In-House Dataset: 1) *COVID-19 Dataset*: this subset includes 25 HRCT for patients with COVID-19, from Wuhan Renmin Hospital and 2) *Fibrosis Dataset*: this subset contains 25 cases from patients with fibrotic lung disease, collected from OSIC dataset.[1] The annotations of these cases were given by experts from Royal Brompton Hospital.

3) Training Strategies: Parameters of the proposed 3-D-UNet are initialized with He-normal initialization. Randomized rotation (rotation degree ranges from -10 to 10), and randomized flip (up, down, left, right) were implemented to augment the dataset during the training. All models were trained for 200 epochs, with an initial learning rate of $1e^{-3}$ and a decay of 0.5 at the 20th, 50th, 80th, 110th, and 150th epochs. For equal contributions of $L_{\mathcal{J}}, L_C, L_{\text{CE}}$, the value of the hyper-parameters α, β, γ in L_{JCAM} was set to 1, and that of φ, δ was set to 0.3 to constraint the accumulation mapping loss (since it is calculated through three dimensions). The ω in CCF score was set to 0.9 to prevent excessive leakages.

B. Experimental Settings and Evaluation Metrics

This section describes the experimental settings of airway segmentation. All the models were developed in the following environment.

1) Ubuntu 20.04.2 LTS with CPU i9-10980XE, NVIDIA RTX 3090 GPU, 128 GB memory.
2) CUDA 11.1.74 and cuDNN 8.0.5
3) Python 3.8.5 with numpy = 1.19.5, SimpleITK = 2.1.1, scikit-image = 0.18.2.

We randomly divided all 90 cases from the BAS dataset into 72 for training and validation, and 18 for the test. All the modules in ablation studies and comparison models were trained from scratch based on the same patch set given by SPS, using the same optimizer and training strategy.

To provide a fair comparison, all the models applied the same post-processing (extracting the largest connected component to remove noises) and inferencing criteria (nonoverlapping sliding window). The significance of the results between the proposed method and the comparison method was assessed by the Wilcoxon signed-rank test, with the p-value < 0.05 indicating significantly better(worse) results.

1) Ablation Studies: To better evaluate the effectiveness of the proposed sub-modules, we conducted ablation studies for airway segmentation including 3-D U-Net (BL), 3-D U-Net with fuzzy attention layer (BL + Fuzzy), 3-D U-Net with fuzzy attention layer and JCAM loss (BL + Fuzzy + JCAM), 3-D U-Net with fuzzy attention layer, JCAM loss and CCF score (BL + Fuzzy + JCAM + CCF_s), and other combinations (BL + JCAM, BL + Fuzzy + CCF_s). It is of note that the baseline 3-D UNet has the same structure as the FAAN without a fuzzy attention layer (shown in Fig. 2) for the sake of fair comparison.

2) Comparisons: To evaluate the effectiveness of the proposed method, we compared our method with common approaches, previous airway extraction solutions, and high-performance models for segmenting treelike structures. The common approaches include nnUNet [14], VNet [15], attention UNet [22], and the previous airway solutions are comprised of methods proposed by Juarez et al. [18], Wang et al. [32], WingsNet [28], NaviAirway [42], and BronchiNet [26]. Furthermore, we also introduced

[1] https://www.osicild.org/

TABLE I
DETAILS OF DATASETS IN THIS STUDY

Dataset	Composition	Volume size	Slice thickness	Disease
BAS	90 cases*	s×512×512, $s \in [157, 764]$	0.45mm-1.80mm	Healthy volunteers or patients with pulmonary disease
COVID-19	25 cases (test only)	s×512×512, $s \in [512, 1145]$	0.40mm-0.65mm	Patients with COVID-19
Fibrosis	25 cases (test only)	s×512×512, $s \in [516, 945]$	0.40mm-0.70mm	Patients with fibrotic lung disease

* 90 cases are split into 54,18 and 18 for training, validation and testing, respectively.

high-performance modules for human treelike structure segmentation, including ERNet [43], CENet [44], and ROSE [45]. In particular, a modified 3-D fuzzy network from [24] was introduced to explore the capacity of existing fuzzy theory-based studies. All the modules were trained on the same patch dataset, acquired based on SPS criteria. Besides, all the comparisons were performed on open BAS and our in-house datasets to assess the robustness and performance, respectively.

3) Evaluation Metrics: In this study, we applied 5 metrics to evaluate the performance of airway segmentation modules. In addition to the three metrics (IoU, DLR, DBR) in Section II-B, the precision and airway missing ratio (AMR) were also evaluated

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

$$\text{AMR} = \frac{\text{FN}}{V_Y} \times 100\% \tag{17}$$

where TP, FP, and FN are the number of true-positive volumes, false-positive volumes, and false-negative volumes.

C. Experimental Results

Results of all the experiments are calculated on the test cases shown as mean ± standard deviation in this section.

1) Ablation Studies: Results of the ablation experiments are presented in Table II, illustrating the mean and standard deviation of the different evaluation metrics. It can be observed that the baseline 3-D U-Net could achieve competitive airway segmentation performance, with a 0.8749 IoU score, 87.81% DLR, and 81.74% DBR. By integrating the fuzzy attention layer with 3-D-UNet, a considerable improvement in the DBR score was witnessed (roughly 2.2% average gain). Introducing the JCAM loss promotes the DBR ratio with 82.93%, while adopting the CCF score also helps to find a better module (with a 1.3% increment of DBR). The largest improvement is witnessed by adopting both fuzzy attention layer and JCAM loss, with 90.64% DBR and 92.95% DLR and the lowest AMR of 4.92%. By integrating all these novel strategies (BL + Fuzzy + JCAM + CCF_s), the proposed FANN achieves the highest CCF score 0.8969, and competitive DBR (89.01%), DLR (92.17%), as well as the AMR (5.22%), while a reduction of IoU and dice score is observed.

2) Comparison Experiments on BAS Data: The proposed FANN achieves state-of-the-art performance for airway segmentation on BAS (Table III), with a 0.8969 CCF score, 92.71% DLR, and 89.01% DBR. ERNet [43], LFSCNN [24], Rose [45], and V-Net [15] achieve high voxelwise precision (0.9587, 0.9837, 0.9499, and 0.9781, respectively), while the V-Net presents server discontinuity of airway predictions, with a high AMR (27.15%), low DLR and DBR (33.96% and 22.04%, respectively). Rose [45] achieved the highest IoU score (0.8825), with comparable DLR (84.84%) and DBR (77.31%). Attention UNet [22] showed competitive results with (88.61% DLR, 82.43% DBR, and 0.8762 IoU), followed by nnUNet [14] (with the IoU score of 0.8805, 86.84% DLR, and 79.21% DBR) and WingsNet [28] (88.00% DLR, 83.11% DBR, and 0.8445 IoU). The study proposed by Juarez et al. [18] also achieved good results on IoU (0.8371) and precision (0.9344) but lower DLR (73.46%) and DBR (64.25%). NaviAirway achieved the lowest AMR of 4.13%, whereas a relatively low precision of 0.8672. Unexpectedly, BronchiNet exerted poor capacity with a high AMR of 43.21%.

3) Comparison Experiments on COVID-19 Data: The proposed method achieved the best CCF score (0.9270) among all the comparisons on COVID-19 data, with 93.30% DLR, 90.18% DBR, and 2.37% AMR (shown in Table IV). The nnUNet [14] achieves the second-highest CCF score of 0.9148, with 91.36% DLR and 87.33% DBR. The attention UNet [22] also presents comparable results, with the highest IoU of 0.923, DLR of 88.53% and DBR of 82.97%. WingsNet [28] achieves a similar performance of DLR (92.50%) and DBR (90.67%) compared with the proposed FAAN, while its IoU (0.8282) is significantly lower than FANN. Unexpectedly, V-Net [15] still performs poorly on the airway segmentation task, with 34.96% DLR and 24.07% DBR. Interestingly, ERNet [43], Rose [45], and NaviAirway [42] achieved similar CCF scores (0.8801, 0.8801, and 0.8882, respectively), while NaviAirway presented better branch metrics than ERNet (87.29% DLR, 80.60% DBR compared with 84.97% DLR and 78.55% DBR).

4) Comparison Experiments on Fibrosis Data: The performance of all models witnesses a dramatic reduction in data acquired from patients with fibrotic lung disease (Table IV). The proposed FANN represents superior performance and generalization ability, with a 0.8099 CCF score, 78.98% DLR, 73.44% DBR, and 0.8904 precision. The attention UNet [22] obtained a good performance with 72.83 DLR, 66.88% DBR, and 0.8844 precision, followed by Wing Net [28] with 69.46% DLR, 63.01% DBR, 0.9401 precision, and 21.98% AMR. nnUNet [14] failed to obtain comparable results on fibrosis data as it did on the BAS and COVID data. Its low DLR (58.15%) and DBR (50.18%) indicate the poor continuity of airway prediction of the branches. Similarly, ERNet [43], Rose [45], LFSCNN [24], and NaviAirway [42] achieved competitive IoU scores (0.7840, 0.7898, 0.7998, and 0.8074, respectively) whereas their branch continuities are relatively

TABLE II
ABLATION STUDIES OF THE PROPOSED METHOD ON BAS DATASET

Model	IoU	Precision	DLR(%)	DBR(%)	AMR(%)	CCF_s
BL	**0.8749±0.0385**	0.9190±0.0302	87.81±8.74	81.74±11.15	5.11±4.10	0.8763±0.0514
BL+Fuzzy	0.8731±0.0493	**0.9205±0.0326**	89.20±10.73	83.92±12.48	5.45±5.41	0.8814±0.0650
BL+JCAM	0.8584±0.0448	0.9027±0.0388	88.71±7.83	82.93±10.90	5.33±4.09	0.8710±0.0554
BL+Fuzzy+JCAM	0.8625±0.0414	0.9039±0.0366	**93.95±7.96**	**90.64±10.42**	**4.92±4.22**	0.8913±0.0527
BL+Fuzzy+CCF_s	0.8701±0.0522	0.9112±0.0315	89.63±3.27	85.24±12.53	5.71±5.82	0.8816±0.0412
BL+Fuzzy+JCAM+CCF_s	0.8738±0.0445	0.9187±0.0320	92.71±7.93	89.01±10.3	5.22±4.50	**0.8969±0.0554**

% indicates the ratio metrics, **DLR**: detected length ratio, **DBR**: detected branch ratio, **AMR**: airway missing ratio. Model abbreviations **BL**: baseline 3D UNet with deep supervision, **JCAM**: Jaccard continuity and accumulation mapping loss, CCF_s: continuity and completeness f-score.

TABLE III
COMPARISON EXPERIMENTS ON THE OPEN DATASET BAS DATASET

Model	IoU	Precision	DLR(%)	DBR(%)	AMR(%)	CCF_s
ERNet [43]*	0.8639±0.0395‡	0.9587±0.0205‡	83.72±8.40‡	75.86±10.59‡	9.58±3.70‡	0.8517±0.0518‡
Rose [45]*	**0.8825±0.0457**‡	0.9499±0.0208‡	84.84±7.88‡	77.31±9.59‡	7.41±4.83‡	0.8669±0.0562‡
CENet [44]*	0.8601±0.0410‡	0.9498±0.0240‡	81.33±11.43‡	73.37±14.50‡	8.82±3.79‡	0.8385±0.0575‡
LFSCNN [24]*	0.8193±0.0676‡	0.9537±0.0224‡	69.09±15.08‡	58.26±16.92‡	12.58±5.03‡	0.7564±0.0898‡
BronchiNet [26]*	0.4984±0.1596‡	0.8588±0.0621‡	64.28±16.28‡	57.52±18.01‡	43.21±18.65‡	0.5541±0.1610‡
NaviAirway [42]*	0.8348±0.0335‡	0.8672±0.0406‡	87.34±7.15‡	80.99±9.50‡	**4.13±3.04**‡	0.8516±0.0439‡
Juarez et al. [18]†	0.8371±0.0752‡	0.9344±0.0314	73.46±14.19‡	64.25±15.55‡	11.17±7.62‡	0.7879±0.0952‡
Wang et al. [32]†	0.7330±0.0786‡	0.7636±0.0737‡	85.05±12.27‡	78.58±14.20‡	5.07±6.19	0.7813±0.0937‡
WingsNet [28]*	0.8445±0.0596‡	0.9373±0.0303‡	88.00±11.95‡‡	83.11±13.46‡	10.32±6.89‡	0.8600±0.0768‡
Attention UNet [22]†	0.8762±0.0414	0.9207±0.0309	88.61±8.32‡	82.43±10.65‡	5.16±4.18	0.8806±0.0534‡
nnUNet [14]*	0.8805±0.0313	0.9436±0.0234‡	86.84±7.00‡	79.21±9.43‡	6.96±4.02‡	0.8750±0.0416‡
V-Net [15]*	0.7140±0.1716‡	**0.9781±0.0150**‡	33.96±17.96‡	22.04±14.22‡	27.15±17.99‡	0.4779±0.1751‡
Proposed	0.8738±0.0445	0.9187±0.0320	**92.71±7.93**	**89.01±10.3**	5.22±4.50	**0.8969±0.0554**

* refers to results obtained by an open-source module (with trained weights) or modules trained from open-source codes. † indicates the reproduced results. ‡ represents statistical significance (with Wilcoxon signed-rank test p-value <0.05) compared with the proposed method.

low. Although the study proposed by Wang et al. [32] achieves a similar DBR and DLR compared with WingsNet, their predictions are prone to airway leakages (a low precision of 0.7468) and missing trachea (low IoU scores). Interestingly, BronchiNet achieved better performance compared with that on the BAS and COVID-19 datasets, with 76.38% DLR, 69.49% DBR, and a 0.6436 CCF score.

5) Visualization Results: Visualization results of airway leakages, missing airways, and correct predictions were shown in Fig. 5. Overall, the airways in scans from patients with COVID-19 were easier to extract, compared with the other two diseases. Due to extremely complex patterns (e.g., honeycombing) in fibrotic lung disease, most modules with missing airways witnessed a performance reduction. Interestingly, most of the airway leakages were caused by missing annotations. In particular, nnUNet achieved good results on COVID-19 data, while encountering missing airways in both BAS and fibrosis data. WingsNet achieved comparable results at small-sized branches, while the main trachea was prone to be ignored. Although Rose showed good performance on COVID-19 and BAS data, lots of severe missing airways were found in fibrosis data (e.g., there exists a large area of false-negative predictions in its fibrosis prediction in Fig. 5). The proposed method achieved the strongest robustness and performance, with the minimum missing airways and the highest airway extraction accuracy.

V. DISCUSSION

This section mainly discusses the overall performance, the importance of data sampling, airway leakage and neglect in the prediction, the importance of evaluation metrics, and the detection rate of different-sized branches.

A. Overall Performance Analysis

The experimental results on public and our in-house datasets have demonstrated the superior performance and robustness of the proposed method for airway segmentation. Among all the comparisons, V-Net and BronchiNet obtained the poorest performance, due to the discontinuity of airway predictions (all the predictions were post-processed by keeping the largest 3-D component). In particular, BronchiNet achieved the weakest performance among previous airway extraction models with low IoU scores and branch metrics, while methods proposed by Juarez et al. and Wang et al. exerted similar performance. The performance of WingsNet is relatively better than that of the NaviAirway, with a nearly 10% gain in branch metrics. For methods transferred from other image modalities, Rose presented the best robustness, compared with ERNet and CENet. Here, we mainly discuss four comparison models with the proposed method, including attention UNet, nnUNet, Rose, and WingsNet. In all three sub-datasets, nnUNet, Rose, and WingsNet represented competitive precision compared to attention models. However, the AMR score of WingsNet was

TABLE IV
COMPARISON EXPERIMENTS ON OUR IN-HOUSE DATASET

	Model	IoU	Precision	DLR(%)	DBR(%)	AMR(%)	CCF_s
COVID-19	ERNet [43]*	0.9064±0.0458‡	0.9655±0.0253‡	84.97±8.42‡	78.55±10.89‡	5.54±2.21‡	0.8801±0.0575‡
	Rose [45]*	0.9132±0.0338‡	0.9552±0.0157‡	85.91±8.12‡	78.89±9.83‡	4.09±3.57‡	0.8882±0.0457‡
	CENet [44]*	0.8883±0.0559‡	0.9611±0.0157‡	77.30±14.19‡	68.86±15.24‡	5.61±3.98‡	0.8328±0.0767‡
	LFSCNN [24]*	0.8883±0.0559‡	0.9611±0.0157‡	77.30±14.19‡	68.86±15.24‡	5.61±3.98‡	0.8328±0.0767‡
	BronchiNet [26]*	0.4200±0.1762‡	0.7253±0.0532‡	64.07±23.74‡	62.20±23.91‡	49.29±23.32‡	0.4965±0.1992‡
	NaviAirway [42]*	0.8861±0.0298‡	0.9100±0.0245‡	87.29±7.92‡	80.60±10.34‡	2.86±2.39‡	0.8801±0.0413‡
	Juarez et al. [18]†	0.8637±0.1016‡	0.9566±0.0134	71.73±15.07‡	63.02±15.07‡	10.08±10.79‡	0.7914±0.1189‡
	Wang et al. [32]†	0.7433±0.1010‡	0.7749±0.0736‡	84.87±13.20‡	79.93±14.09‡	5.41±10.35‡	0.7870±0.1129‡
	WingsNet [28]*	0.8282±0.0672‡	0.9537±0.0215‡	92.50±6.89	**90.67±8.95**	13.68±7.15‡	0.8689±0.0680‡
	Attention UNet [22]†	**0.9230±0.0279**	0.9464±0.0175	88.53±6.66‡	82.97±8.34‡	2.62±2.21	0.9057±0.0377‡
	nnUNet [14]*	0.9158±0.0381‡	0.9685±0.0407‡	91.36±6.72‡	87.33±8.89‡	2.61±1.73‡	0.9148±0.0473‡
	V-Net [15]*	0.7898±0.0871‡	**0.9893±0.0040**‡	34.96±9.66‡	24.07±8.80‡	20.33±8.86‡	0.5052±0.0911‡
	Proposed	0.9222±0.0261	0.9431±0.0166	**93.30±5.29**	90.18±7.59	**2.37±1.64**	**0.9270±0.0338**
Fibrosis	ERNet [43]*	0.7840±0.0948‡	0.9359±0.0404‡	55.87±14.92‡	49.09±15.35‡	12.46±3.38‡	0.6641±0.1133‡
	Rose [45]*	0.7898±0.0705‡	0.8979±0.0687	65.17±14.41‡	58.47±16.33‡	12.79±7.21‡	0.7213±0.0913‡
	CENet [44]*	0.7602±0.1020‡	0.9459±0.0252‡	50.42±16.40‡	43.68±16.85‡	14.67±4.89‡	0.6194±0.1228‡
	LFSCNN [24]*	0.7998±0.0527‡	0.9308±0.0169‡	57.95±13.05‡	50.43±14.12‡	11.85±3.65‡	0.6835±0.0719‡
	BronchiNet [26]*	0.5709±0.1146‡	0.8929±0.0432‡	76.38±15.51‡	69.49±15.09‡	34.22±7.71‡	0.6436±0.1298‡
	NaviAirway [42]*	0.8074±0.0533‡	0.9247±0.0165‡	59.94±14.40‡	51.48±14.90‡	13.45±6.45‡	0.6989±0.0742‡
	Juarez et al. [18]†	0.7911±0.0532‡	0.9296±0.0294‡	56.16±10.76‡	48.29±12.18‡	15.68±6.37‡	0.6688±0.0688‡
	Wang et al. [32]†	0.6979±0.0647‡	0.7468±0.0773‡	69.61±9.24‡	62.61±11.17‡	8.22±3.88	0.6971±0.0747‡
	WingsNet [28]*	0.7436±0.0805‡	0.9401±0.0156‡	69.46±9.71‡	63.01±11.39‡	21.98±8.48‡	0.7208±0.0872‡
	Attention UNet [22]†	0.8202±0.0553	0.8844±0.0640	72.83±9.41‡	66.88±10.84‡	7.98±2.97	0.7764±0.0678‡
	nnUNet [14]*	**0.8312±0.0495**‡	0.9381±0.0314‡	58.15±6.80‡	50.18±7.93‡	11.74±2.93‡	0.6972±0.0564‡
	V-Net [15]*	0.5002±0.0967‡	**0.9768±0.0006**‡	8.54±3.76‡	3.40±2.99‡	49.36±9.88‡	0.1576±0.0568‡
	Proposed	0.8269±0.0402	0.8904±0.0373	**78.98±8.00**	**73.44±9.54**	**7.95±2.37**	**0.8099±0.0517**

* refers to results obtained by an open-source module (with trained weights) or module trained from open-source codes. † indicates the reproduced results. ‡ represents statistical significance (with Wilcoxon signed-rank test p-value <0.05) compared with the proposed method.

roughly twice as large as that of nnUNet and Rose, while the DBR and DLR of the WingsNet were much better. This indicated that WingsNet obtained worse predictions on the main trachea, but better predictions on small/middle-sized branches. Although Rose did not perform as well as the other three models, it presented comparable performance to some well-designed airway extraction solutions, indicating its strong robustness and capacity. Compared with the other two models, the attention UNet presents better performance on all the datasets, showing the effectiveness of the attention mechanism.

In particular, FANN achieved significantly better results in DLR, DBR, and CCF scores than Rose, nnUNet, attention UNet, and WingsNet in the public BAS dataset. The IoU of FANN (0.8738) represented no significant differences compared with nnUNet (0.8805) and attention UNet (0.8762), which outperforms that of the WingsNet (0.8445). For COVID data, the proposed FANN gained better DLR and DBR compared to Rose, nnUNet, and attention UNet. Although WingsNet achieves a relatively low IoU (0.8282), it obtains a similar DLR and DBR to FANN (with Wilcoxon signed-rank test p-value > 0.05). On fibrosis data, although the nnUNet achieved a better IoU and precision score than FANN, it suffers from heavy discontinuity with poor DLR (58.15%) and DBR (50.18%) gained. WingsNet achieves the highest precision (0.9401) but the largest AMR (21.98%), indicating severe false-negative predictions. FANN achieves the best CCF score, DLR, DBR, and AMR among all the comparisons, with competitive IoU and precision scores.

B. Fuzzy Theory-Based Method

Although there have been some studies that integrated fuzzy theory with deep neural networks, only a few can be used for image segmentation. By replacing the 2-D layers in LFSCNN [24] with 3-D layers, LFSCNN achieved comparable results for airway extraction. In particular, the fuzzy module was arranged at the residual skip connection between the encoder block and decoder block, acting as the role of feature integration. Unfortunately, LFSCNN did not perform well on BAS, COVID-19, and Fibrosis data, with nearly 20% branch metrics gap to the proposed method. To better explore how fuzzy integration performs in [24], we conduct ablation experiments by only removing the fuzzy fusion module from LFSCNN. Unexpectedly, there witnessed an improvement in LFSCNN when removing the fuzzy integration on BAS and COVID-19 data, with 0.8462 IoU, 76.14% DLR, 66.36% DBR, and 0.8507 IoU, 69.63 DLR%, 59.60% DBR, respectively. Interestingly, that performance reduced on fibrosis data, with 0.7808 IoU, 51.43% DLR, and 42.46% DBR. This indicated that the module did not benefit from applying the fuzzy feature integration module [24] on scans acquired from patients with mild pathological changes of the lung. Interestingly, when the lung is invaded by complex pathological changes (e.g., honeycombings), there witnessed a performance improvement when applying the fuzzy integration module.

C. Data Sampling Strategy

The strategy of data sampling is prone to be overlooked for patch-based modules. In this article, we presented a smart data sampling strategy for patch extraction in the imbalanced dataset. To better explore the importance of data sampling, we conducted experiments on the open BAS dataset in terms of different data sampling strategies, including different

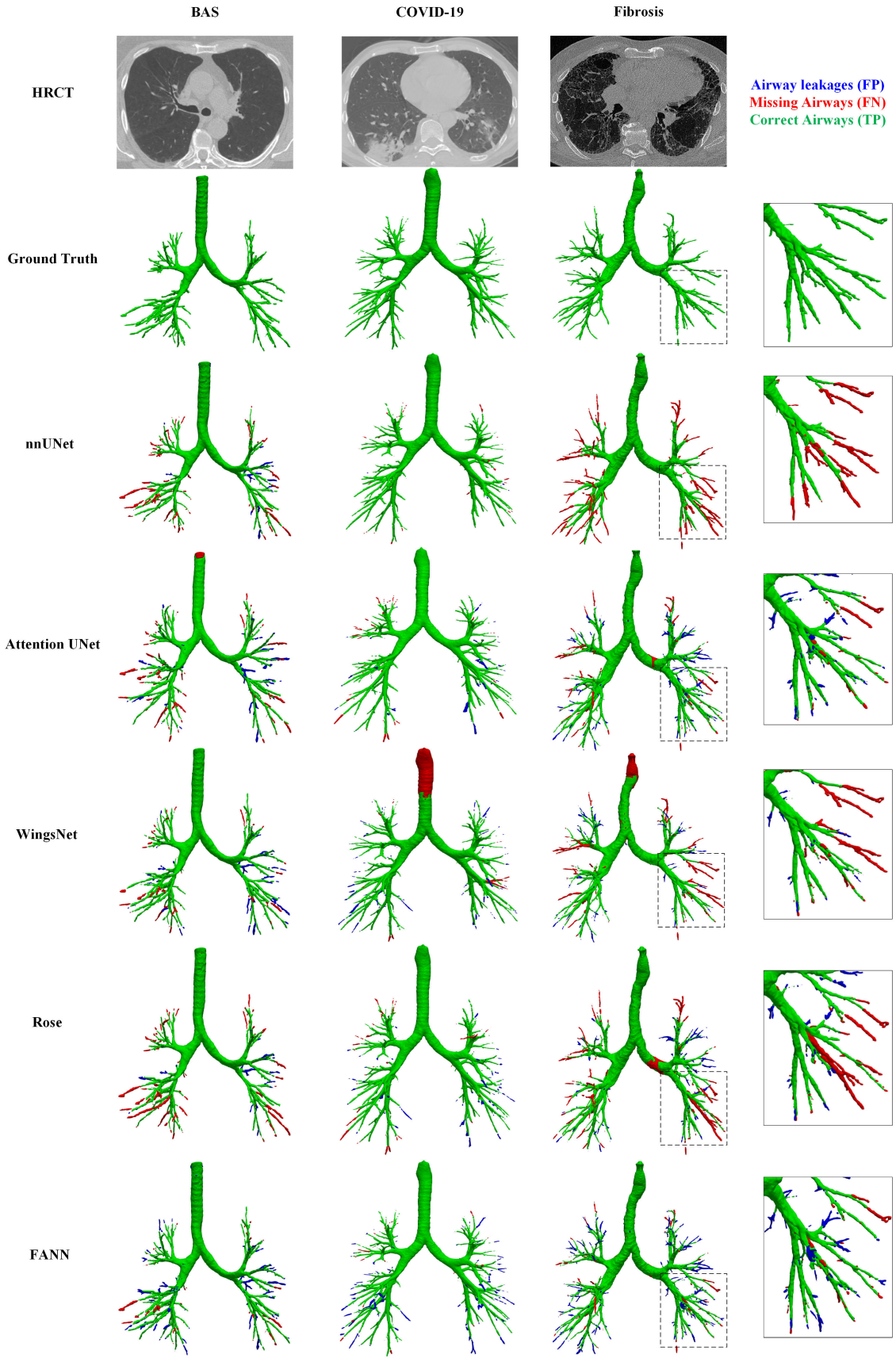

Fig. 5. Visualization results of the competitive module, including nnUNet, attention UNet, WingsNet, Rose, and the proposed FANN on fibrosis data (first row), COVID-19 data (second row), and BAS data (third row), with the green, red, blue representing the correct airways (TP), missing airways (FN), and airway leakages (FP), respectively.

TABLE V
COMPARISON EXPERIMENTS OF DATA SAMPLING STRATEGY

	Extraction	Patch size	IoU	Precision	DLR(%)	DBR(%)	AMR(%)
Cut-1	Sequential	uniform	0.8782±0.0641	0.9341±0.0409	85.17±14.28	76.41±15.27	7.64±7.22
Cut-2	Sequential	median	0.8854±0.0470	0.9408±0.0300	85.79±8.12	78.77±10.06	6.03±4.63
Cut-3	Sequential	mean	0.8862±0.0422	0.9394±0.0275	86.66±8.88	79.39±10.80	6.06±4.61
Cut-4	Drop-A	mean	0.8069±0.0826	0.9437±0.0316	84.64±10.00	77.57±12.70	14.87±8.13
Cut-5	Drop-B	mean	**0.8883±0.0407**	**0.9451±0.0286**	87.73±8.01	80.92±10.51	5.95±4.25
Cut-6	Smart	mean	0.8749±0.0385	0.9190±0.0302	**87.81±8.74**	**81.74±11.15**	**5.11±4.10**

The 'uniform' patch size refers to 128×128×128, while the 'median' and 'mean' are 160×96×160 and 128×96×144, respectively.

extraction criteria and patch sizes. Here, we define three extraction criteria, including: 1) *Sequential*: extracting the patches sequentially by sliding windows within the minimum bounding box volume of the 3-D ground truth area; 2) *Drop*: extracting the patches sequentially by sliding windows within the minimum bounding box volume of the 3-D ground truth area and discarding the main trachea (Drop-A) or pure negative patches (Drop-B) to alleviate data imbalance; 3) SPS as described in Section III-C. We also compare the patch size defined by three different criteria, with: 1) uniform of 128 × 128 × 128; 2) the roughly same ratio as the median size of the ground truth volumes (160 × 96 × 160; and 3) the roughly same ratio as the mean size of the ground truth volumes (128 × 96 × 144).

Although little increment of IoU is emerged from Cut-1, Cut-2, and Cut-3, a significant improvement in the DLR and DBR is noted. In addition, the results of Cut-3 and Cut-4 (Table V) indicate that the segmentation performance can be improved by discarding the pure negative patches (with 1.6% and 1% gain in DBR and DLR, respectively). However, discarding the main trachea may not be applicable since the module failed to predict the trachea by learning the features of small and medium-sized branches (shown in Cut-4). The model trained on Cut-5 presents a competitive DBR with a roughly 4.5% increment compared with Cut-1. The proposed SPS strategy achieves the best performance with the highest DLR, DBR, and lowest AMR.

D. Airway Leakages and Airway Neglect

Airway leakages are false-positive predictions that can be reflected by the precision score. Airway neglect refers to false-negative predictions that are assessed by AMR. As shown in Fig. 5, due to the unavoidable neglect of manual labeling, some leakages might be ascribed to the airway branches that are not annotated correctly. nnUNet presents a high precision score with limited airway leakages, while it suffers from relative AMR with many airway neglects (especially in fibrosis cases). WingsNet performs severe airway neglect of the main trachea, leading to a low IoU score with comparable DBR and DLR. Rose, WingsNet, attention UNet, and FANN all present some leakages at the terminal branches, indicating the incompleteness of airway annotations. Compared with all the studied models, FANN achieves the smallest AMR and most of its leakages belong to incomplete annotations. Therefore, we believe that a small number of airway leakages are encouraged, as they can help the model better predict terminal bronchioles.

TABLE VI
DETECTION RATE OF DIFFERENT-SIZED BRANCHES

	TB	SB	MB	LB
BAS	81.55%	95.76%	98.59%	99.38%
COVID-19	82.96%	97.05%	99.67%	100%
Fibrosis	47.15%	87.60%	98.32%	98.50%

E. Evaluation Metric

The conventional evaluation metrics such as IoU score or dice score no longer work well to find the best module during the training. As Tables II–IV illustrated, the module with the highest IoU score does not refer to the one with the best segmentation ability. For instance, the nnUNet achieved the best IoU score on the BAS dataset, but it obtained a weak performance on the fibrosis evaluation. Meanwhile, only focusing on the branch performance while ignoring the overlapped metric may also lead to airway leakages. For example, the BL + fuzzy + JCAM achieved better DLR and DBR but a lower CCF score than the final combination in the ablation studies of FAAN (Table I). To explore the most appropriate metric, we further evaluated the BL + fuzzy + JCAM on the fibrosis and COVID-19 datasets. It achieved 0.7961 (0.9099) IoU score, 74.66% (93.81%) DLR and 68.85% (91.87%) DBR, 10.29% (2.36%) AMR, and 0.7687 (0.9215) CCF score on fibrosis (COVID-19) data, with a considerable gap compared with the FAAN. This also proves the efficiency of the proposed CCF score, which can better help the module selection.

F. Detection Rate of Different-Sized Branches

To explore the segmentation performance of different-sized airway branches, the detection rate of the branches with different radii is calculated. Here, we define a branch is correctly detected when its IoU is larger than 0.8. We simply divided the airway branches into four sizes based on their estimated radius, including TB (terminal branches that range from 0 to 2 mm), SB (small branches 2–4 mm), MB (middle branches range from 4 to 8 mm), and LB (large branches that larger than 8 mm). The performance of the proposed FAAN on BAS, COVID-19 and fibrosis datasets is shown in Table VI. Table V illustrates that the proposed method

TABLE VII
COMPUTATIONAL COMPLEXITY OF DIFFERENT METHODS

	# Parameter (M)	FLOPs (G)	Avg speed (s)
WingsNet	1.47	259.49	20.39‡
ERNet	5.11	341.26	26.62‡
LFSCNN	29.36	649.25	31.93‡
FANN	14.12	308.86	35.17
3D U-Net	13.85	298.70	35.77
Attention U-Net	14.03	305.47	36.82‡
CENet	74.68	1530	50.08‡
Bronchinet	5.88	122.10	92.71‡
nnUNet*	31.19	532.96	169.8‡
Rose	20.13	10940	177.08‡

*The employed nnUNet was built followed by a "3d full resolution" configuration, where the details can be found at https://github.com/MIC-DKFZ/nnUNet. ‡ indicates significant differences (p-value <0.05). "# parameters" refers to the number of parameters.

has achieved superior performance of airway segmentation on BAS and COVID-19 data, while the terminal branch detection rate on the fibrosis data is relatively low. This is caused by the complex honeycombing in fibrotic lung disease which has similar structural similarity with the airway branches.

G. Computational Complexity

The computational complexities of all methods are reported and ranked in Table VII regarding the average inference time. The number of network parameters (# Parameter), floating point operations (FLOPs), and Average inference speed (Avg speed) were presented to show the model's comprehensive computational efficiency. The number of parameters affects the GPU memory cost during the inference, and the "FLOPs" affects model complexity and is important for inference speed. As Table VII shows, WingsNet is the lightest model among all comparison models, with 1.47 M parameters and a 20.39 s inference speed, followed by ERNet, with 5.11 M parameters and 26.62 s time cost. The average inference speed for FANN, and 3-D U-Net remains similar, with no significant differences (p-value > 0.05) between each other. In addition, the large number of parameters, FLOPs and time-consuming of nnUNet indicate a high computation complexity of image pre/postprocessing. Although BronchiNet has a relatively small number of parameters (5.88 M), it includes lots of pre/postprocessing which leads to more time-consuming (nearly 3 mins). In addition, CENet and Rose were initially designed for 2-D images and became complex when they were modified to 3-D architectures. Even though there were only 20.12 M trainable parameters in Rose, it required complicated computation cost with 10.94 T FLOPs which leads to an average inference speed of 177.08 s.

H. Potential Studies

This study has shown the feasibility of adopting fuzzy logic into deep attention neural networks. In particular, many potential research directions can be explored. For example, introducing a fuzzy cost function for better optimization, or bringing in other MFs such as generalized bell and sigmoid MFs.

VI. CONCLUSION

Airway segmentation is a crucial step that can help clinicians better assess disease and prognosis analysis. However, most existing methods heavily suffer from discontinuous prediction for small-sized airway branches. The proposed study illustrates the importance of data sampling strategy and evaluation metrics. The effectiveness of FANNs, JCAM loss, and the CCF score for 3-D airway segmentation has been demonstrated. The superior performance of the proposed method has been observed on both open datasets and our in-house datasets, including scans acquired from patients with cancer, COVID-19, fibrosis, and mild lung disease.

The imperfect terminal branches detection rate on fibrosis data indicates future research directions. Further studies such as novel fuzzy membership functions, adversarial learning, and diffusion modules should be considered for enhancing terminal branch extraction in fibrotic lung disease.

ACKNOWLEDGMENT

The authors would like to thank Prof. Jun Xia from the Department of Radiology, Shenzhen Second People's Hospital providing part of the data with ethics approval incorporated in this study.

REFERENCES

[1] S. L. F. Walsh, N. Sverzellati, A. Devaraj, A. U. Wells, and D. M. Hansell, "Chronic hypersensitivity pneumonitis: High resolution computed tomography patterns and pulmonary function indices as prognostic determinants," *Eur. Radiol.*, vol. 22, no. 8, pp. 1672–1679, Aug. 2012.

[2] S. L. F. Walsh, N. Sverzellati, A. Devaraj, G. J. Keir, A. U. Wells, and D. M. Hansell, "Connective tissue disease related fibrotic lung disease: High resolution computed tomographic and pulmonary function indices as prognostic determinants," *Thorax*, vol. 69, no. 3, pp. 216–222, Mar. 2014.

[3] A. J. Edey, A. A. Devaraj, R. P. Barker, A. G. Nicholson, A. U. Wells, and D. M. Hansell, "Fibrotic idiopathic interstitial pneumonias: HRCT findings that predict mortality," *Eur. Radiol.*, vol. 21, no. 8, pp. 1586–1593, Aug. 2011.

[4] S. L. F. Walsh, L. Calandriello, N. Sverzellati, A. U. Wells, and D. M. Hansell, "Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT," *Thorax*, vol. 71, no. 1, pp. 45–51, Jan. 2016.

[5] J. Jacob et al., "Mortality prediction in idiopathic pulmonary fibrosis: Evaluation of computer-based CT analysis with conventional severity measures," *Eur. Respiratory J.*, vol. 49, no. 1, Jan. 2017, Art. no. 1601011.

[6] S. M. Humphries et al., "Idiopathic pulmonary fibrosis: Data-driven textural analysis of extent of fibrosis at baseline and 15-month follow-up," *Radiology*, vol. 285, no. 1, pp. 270–278, Oct. 2017.

[7] D. M. Hansell, J. G. Goldin, T. E. King, D. A. Lynch, L. Richeldi, and A. U. Wells, "CT staging and monitoring of fibrotic interstitial lung diseases in clinical practice and treatment trials: A position paper from the Fleischner society," *Lancet Respiratory Med.*, vol. 3, no. 6, pp. 483–496, Jun. 2015.

[8] D. Aykac, E. A. Hoffman, G. McLennan, and J. M. Reinhardt, "Segmentation and analysis of the human airway tree from three-dimensional X-ray CT images," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 940–950, Aug. 2003.

[9] J. Tschirren, E. A. Hoffman, G. McLennan, and M. Sonka, "Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans," *IEEE Trans. Med. Imag.*, vol. 24, no. 12, pp. 1529–1539, Dec. 2005.

[10] A. P. Kiraly, W. E. Higgins, G. McLennan, E. A. Hoffman, and J. M. Reinhardt, "Three-dimensional human airway segmentation methods for clinical virtual bronchoscopy," *Academic Radiol.*, vol. 9, no. 10, pp. 1153–1168, Oct. 2002.

[11] M. W. Graham, J. D. Gibbs, D. C. Cornish, and W. E. Higgins, "Robust 3-D airway tree segmentation for image-guided peripheral bronchoscopy," *IEEE Trans. Med. Imag.*, vol. 29, no. 4, pp. 982–997, Apr. 2010.

[12] Z. Xu, U. Bagci, B. Foster, A. Mansoor, J. K. Udupa, and D. J. Mollura, "A hybrid method for airway segmentation and automated measurement of bronchial wall thickness on CT," *Med. Image Anal.*, vol. 24, no. 1, pp. 1–17, 2015.

[13] P. Lo et al., "Extraction of airways from CT (EXACT'09)," *IEEE Trans. Med. Imag.*, vol. 31, no. 11, pp. 2093–2107, Jul. 2012.

[14] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Dec. 2020.

[15] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[16] J.-P. Charbonnier, E. M. Van Rikxoort, A. A. Setio, C. M. Schaefer-Prokop, B. van Ginneken, and F. Ciompi, "Improving airway segmentation in computed tomography using leak detection with convolutional networks," *Med. Image Anal.*, vol. 36, pp. 52–60, Feb. 2017.

[17] J. Yun et al., "Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net," *Med. Image Anal.*, vol. 51, pp. 13–20, Jan. 2019.

[18] A. Garcia-Uceda Juarez, H. A. Tiddens, and M. D. Bruijne, "Automatic airway segmentation in chest ct using convolutional neural networks," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Springer, 2018, pp. 238–250.

[19] Y. Qin et al., "AirwayNet: A voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 212–220.

[20] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 424–432.

[21] H. Li, Z. Tang, Y. Nan, and G. Yang, "Human treelike tubular structure segmentation: A comprehensive review and future perspectives," *Comput. Biol. Med.*, vol. 151, Dec. 2022, Art. no. 106241.

[22] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," in *Proc. Int. Conf. Med. Imag. Deep Learn.*, 2018, p. 15.

[23] Y. Deng, Z. Ren, Y. Kong, F. Bao, and Q. Dai, "A hierarchical fused fuzzy deep neural network for data classification," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 1006–1012, Aug. 2017.

[24] C. Guan, S. Wang, and A. W.-C. Liew, "Lip image segmentation based on a fuzzy convolutional neural network," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 7, pp. 1242–1251, 2019.

[25] T. Shen, J. Wang, C. Gou, and F.-Y. Wang, "Hierarchical fused model with deep learning and type-2 fuzzy learning for breast cancer diagnosis," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 12, pp. 3204–3218, Dec. 2020.

[26] A. Garcia-Uceda, R. Selvan, Z. Saghir, H. A. W. M. Tiddens, and M. de Bruijne, "Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks," *Sci. Rep.*, vol. 11, no. 1, pp. 1–15, Aug. 2021.

[27] Y. Qin, Y. Gu, H. Zheng, M. Chen, J. Yang, and Y.-M. Zhu, "AirwayNet-SE: A simple-yet-effective approach to improve airway segmentation using context scale fusion," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 809–813.

[28] H. Zheng et al., "Alleviating class-wise gradient imbalance for pulmonary airway segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2452–2462, Sep. 2021.

[29] Q. Meng, H. R. Roth, T. Kitasaka, M. Oda, J. Ueno, and K. Mori, "Tracking and segmentation of the airways in chest ct using a fully convolutional network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 198–207.

[30] P. J. Reynisson et al., "Airway segmentation and centerline extraction from thoracic CT—Comparison of a new method to state of the art commercialized methods," *PLoS ONE*, vol. 10, no. 12, Dec. 2015, Art. no. e0144282.

[31] S. A. Nadeem et al., "A CT-based automated algorithm for airway segmentation using freeze-and-grow propagation and deep learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 405–418, Jan. 2021.

[32] C. Wang et al., "Tubular structure segmentation using spatial fully connected network with radial distance loss for 3D medical images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 348–356.

[33] T. Lei, P. Liu, X. Jia, X. Zhang, H. Meng, and A. K. Nandi, "Automatic fuzzy clustering framework for image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 2078–2092, Sep. 2020.

[34] T. Lei, X. Jia, Y. Zhang, S. Liu, H. Meng, and A. K. Nandi, "Superpixel-based fast fuzzy C-means clustering for color image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1753–1766, Sep. 2019.

[35] S. Zhou, Q. Chen, and X. Wang, "Fuzzy deep belief networks for semi-supervised sentiment classification," *Neurocomputing*, vol. 131, pp. 312–322, May 2014.

[36] S. R. Price, S. R. Price, and D. T. Anderson, "Introducing fuzzy layers for deep learning," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jun. 2019, pp. 1–6.

[37] K. Huang, Y. Zhang, H. D. Cheng, P. Xing, and B. Zhang, "Semantic segmentation of breast ultrasound image with fuzzy deep learning network and breast anatomy constraints," *Neurocomputing*, vol. 450, pp. 319–335, Aug. 2021.

[38] W. Ding, M. Abdel-Basset, H. Hawash, and W. Pedrycz, "Multimodal infant brain segmentation by fuzzy-informed deep learning," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 4, pp. 1088–1101, Apr. 2022.

[39] W. Xie, C. Jacobs, J.-P. Charbonnier, and B. van Ginneken, "Relational modeling for robust and efficient pulmonary lobe segmentation in CT scans," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2664–2675, Aug. 2020.

[40] T. C. Lee, R. L. Kashyap, and C. N. Chu, "Building skeleton models via 3-D medial surface axis thinning algorithms," *CVGIP, Graph. Models Image Process.*, vol. 56, no. 6, pp. 462–478, 1994.

[41] S. G. Armato 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.*, vol. 38, no. 2, pp. 915–931, Jan. 2011.

[42] A. Wang, T. Chi Chun Tam, H. Ming Poon, K.-C. Yu, and W.-N. Lee, "NaviAirway: A bronchiole-sensitive deep learning-based airway segmentation pipeline," 2022, *arXiv:2203.04294*.

[43] L. Xia et al., "3D vessel-like structure segmentation in medical images by an edge-reinforced network," *Med. Image Anal.*, vol. 82, Nov. 2022, Art. no. 102581.

[44] Z. Gu et al., "Ce-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[45] Y. Ma et al., "ROSE: A retinal OCT-angiography vessel segmentation dataset and new model," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 928–939, Mar. 2021.