This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Fuzzy Attention-based Border Rendering Network for Lung Organ Segmentation

Sheng Zhang[1]*, Yang Nan[1]*, Yingying Fang[1]*, Shiyi Wang[1], Xiaodan Xing[1], Zhifan Gao[2], and Guang Yang[1]✉

[1]Imperial College London, London, UK; [2]Sun Yat-sen University, Shenzhen, China
g.yang@imperial.ac.uk

Abstract. Automatic lung organ segmentation on CT images is crucial for lung disease diagnosis. However, the unlimited voxel values and class imbalance of lung organs can lead to false-negative/positive and leakage issues in advanced methods. Additionally, some slender lung organs are easily lost during the *recycled* down/up-sample procedure, e.g., bronchioles & arterioles, causing severe discontinuity issue. Inspired by these, this paper introduces an effective lung organ segmentation method called Fuzzy Attention-based Border Rendering (FABR) network. Since fuzzy logic can handle the uncertainty in feature extraction, hence the fusion of deep networks and fuzzy sets should be a viable solution for better performance. Meanwhile, unlike prior top-tier methods that operate on all regular dense points, our FABR depicts lung organ regions as cube-trees, focusing only on *recycle*-sampled border vulnerable points, rendering the severely discontinuous, false-negative/positive organ regions with a novel Global-Local Cube-tree Fusion (GLCF) module. All experimental results, on four challenging datasets of airway & artery, demonstrate that our method can achieve the favorable performance significantly.

Keywords: Lung organ segmentation · CT · Fuzzy logic · Border render.

1 Introduction

Automatic lung organ segmentation is one of the challenging tasks in the field of medical image analysis [13, 7]. Recently, this task has been extended to variously realistic applications, e.g., robotic surgery [3], lung disease diagnosis & prognosis [16, 2]. To achieve a superb segmentation performance, it is vital to learn a group of abundant and salient descriptions of lung image feature. However, current state-of-the-art methods of lung organ segmentation still face several challenges and aspects for improvement. Firstly, the unlimited voxel values, multi-site imaging discrepancy and class imbalance in lung organ images can lead to false-negative and leakage issues in prior segmentation methods, which badly influences the critical early diagnosis of imperceptible lung diseases, e.g.,

*Equal contribution.

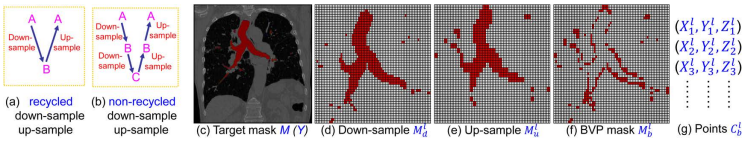

Fig. 1. The elaboration of **B**order **V**ulnerable **P**oints (BVP) caused by **recycled** down-sample and up-sample in the encoder-decoder backbone. Downsampling (c) gets (d), upsampling (d) gets (e), then (f) is the absolute difference of (c) & (e). In the test phase, (c) is binarized coarse prediction.

lung fibrosis, nodule and hypertension, etc. Secondly, the presence of numerous slender branches, e.g., bronchioles and arterioles, which are easily lost during the *recycled* down/up-sampling procedure in Fig. 1, can result in discontinuity, detail loss, and coarse mask predictions. Thirdly, most CNN-based medical segmentation methods treat all points equally during the mask rendering stage, overlooking the vulnerability of border points in Fig. 1 (f) and the importance of explicit border modeling. Lastly, while Vision Transformer (ViT) has shown promise in computer vision tasks [1, 4], its quadratic operation complexity limits its application in 3D high-resolution CT images due to hardware constraints. Meanwhile, most specific datasets for medical image analysis are small and scarce due to laborious manual annotation and privacy protection, which badly restricts the potential of transformer-based top-tier methods.

To address these limitations in this paper, we propose an effective lung organ segmentation method called FABR. Unlike prior approaches, the method FABR fuses fuzzy sets and deep network to diminish the uncertainty in feature representations, decouples and depicts medical image regions as cube-trees, specifically targeting the border vulnerable points illustrated in Fig. 2. To address the challenges of severe discontinuity and false-negative/positive bronchioles and arterioles, we propose one innovative module of global-local cube-tree fusion, which fuses the learnable global embedding and local lung organ features.

In summary, our main contributions are three-folds: (1) We seamlessly integrate efficient fuzzy attention theory and transformer-like expansion/compression convolutional network to diminish the uncertainty of lung organ feature representations; (2) We present an innovative global-local cube-tree fusion module, which explicitly models the border vulnerable points yielded by recycled down/up-sample for accurate lung organ segmentation; (3) We do extensive experiments on four challenging datasets to prove the efficacy of our method.

2 Methodology

The overview of our method FABR is detailed in Fig. 2. It mainly includes two modules, i.e., fuzzy attention-based transformer-like 3D U-shaped backbone and **G**lobal-**L**ocal **C**ube-tree **F**usion (GLCF) module. The fuzzy attention-based transformer-like backbone is inspired by the well-known ConvNeXt [19] and detailed in Fig. 3, which includes a preliminary stem, sequential transformer-like

regular/down/up-sample convolution blocks, a bottleneck and four efficient fuzzy attention modules, where each convolution block is constructed by applying a large kernel of $5 \times 5 \times 5$ 3D separable depth-wise convolution/deconvolution, group-normalization, transformer-like architecture (i.e., embedding 4× expansion/compression $1 \times 1 \times 1$ convolution layers like FFN module of transformer in our Fuzzy attention module) and GELU activation layer. The corresponding layers of the same scale between the encoder and decoder are linked by the efficient fuzzy attention layer. Besides, each-scale stage of the decoder is added by the $1 \times 1 \times 1$ 3D convolution and activation layers to predict the preliminary coarse masks of lung organ segmentation. Then, unlike the prior top-tier methods that operate on all regular dense points of the coarse masks to render the raw prediction, the proposed GLCF module decouples and depicts the medical image regions as cube-trees, which only focuses on the *recycle*-sampled BVP, and renders the severe discontinuity as well as false-negative/positive bronchioles or arterioles. We now elaborate the insights within the proposed method FABR for each innovative module in the following subsections.

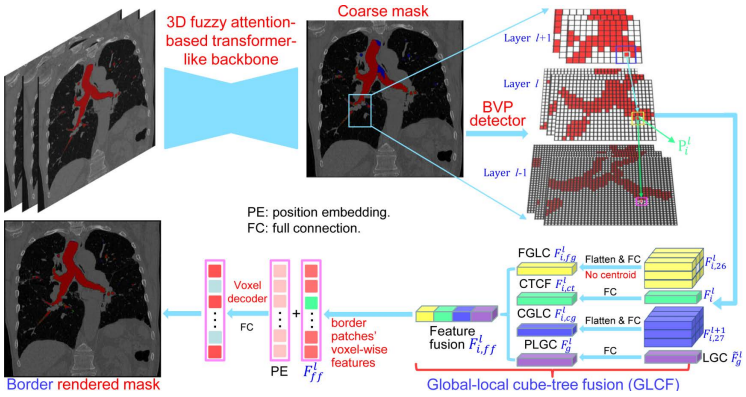

Fig. 2. The overview of our method FABR. FGLC: fine grain local context; CTCF: cube-tree centroid feature; CGLC: coarse grain local context; PLGC: projected learnable global context. BVP detector is shown in Fig. 1. Noting the matched relationship between top-right boxes' and bottom-right bars' colors.

2.1 Fuzzy Attention-based Transformer-like Backbone

One of the key challenges to design a robust lung organ segmentation module lies in the inherent uncertainty from the organ annotations and voxel values, e.g., bronchioles and arterioles. Various efforts have been done to enhance the network to focus on pertinent regions. Notably, Attention U-Net [11] introduces an attention gate to bolster accuracy by suppressing feature activations in irrelevant regions. However, we deem that the non-channel specifics of current attention map assign the same "attention" coefficient to all feature points along

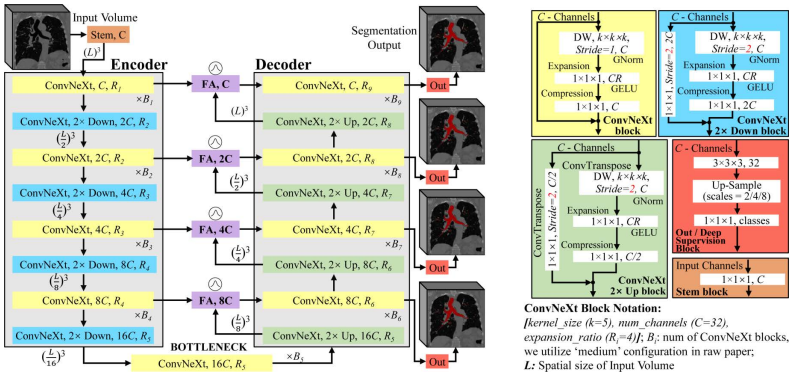

Fig. 3. Our FA-based transformer-like backbone design and coarse mask generation. FA: fuzzy attention module detailed in subsection 2.1. DW: depth-wise convolution.

the channel dimension. Specifically, given a feature map $F \in \mathbb{R}^{C\times H\times W\times D}$, the extant attention map is built as $\alpha \in \mathbb{R}^{H\times W\times D}$, while all features along the channel wise C share the same "importance". This mechanism is unreliable since the features in different channels are extracted by different convolution kernels; therefore, we advocate the attention map to be channel-specific.

Meanwhile, numerous studies have proved the efficacy of both fuzzy logic and neural networks in data representation [10]. Broadly speaking, neural networks strive to diminish noise in original data to extract meaningful feature representations, while fuzzy logic can derive fuzzy representations, mitigating the original data uncertainty. Hence, we fuse fuzzy logic with attention mechanism by utilizing trainable Gaussian membership functions (GMFs). This fusion serves to enhance the segmentation network's ability to focus on pertinent regions, concurrently diminishing uncertainty and variations in data representations.

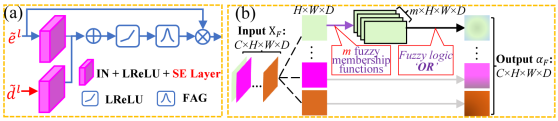

Fig. 4. The details of (a) our efficient fuzzy attention module and (b) fuzzy attention gate (FAG) in the subfigure (a). Zooming in for a better view.

As shown in Fig. 4(a), the proposed efficient fuzzy attention module is adopted within the skip connection, taking both feature maps $\{\tilde{e}^l, \tilde{d}^l\}$ from the *l-th* encoder and decoder layers as inputs, which are directly yielded by the transformer-like 4× expansion/compression layers in ConvNeXt [19] backbone, followed by an instance normalization and a Leaky-ReLU layers for feature reconstitution. Then, two very lightweight squeeze-excitation (SE) layers [5] are employed to further boost the channel-specificity. Next, a voxel-wise adding operation is adopted to fuse the information, followed by a Leaky-ReLU. Eventually, the feature representations are fed into the FAG to generate a voxel-wise attention map, shown in Fig. 4(b). Assume $X \in R^{C\times H\times W\times D}$ (regardless of batch size) as the input of

FAG. Due to the smoothness and concise notation of GMFs, learnable GMFs are proposed to specify the deep fuzzy sets. Each feature map (with size $H \times W \times D$) is filtered by m GMFs with the trainable centre $\mu_{i,j}$ and spread $\sigma_{i,j}$

$$f_{i,j}(X, \mu, \sigma) = e^{(-(X_j - \mu_{i,j})^2)/(2\sigma_{i,j}^2)}, \tag{1}$$

where $i \in \{1, \ldots, m\}, j \in \{1, \ldots, C\}$. Our goal is to use the m membership function to learn the "importance" of target fuzzy feature representations. Given the trade-off of model efficiency & efficacy, $m = 4$ GMFs are used. Thus, we assume that the information can be better preserved by applying the aggregation operator "OR" while suppressing irrelevant features. Given fuzzy sets $\tilde{A}$ and $\tilde{B}$, the operator "OR" is denoted as Equ. 2(a).

$$f_{\tilde{A} \cup \tilde{B}}(y) = f_{\tilde{A}}(y) \vee f_{\tilde{B}}(y), \ \forall y \in U, (a); \quad f_{\tilde{A} \cup \tilde{B}}(y) = \max(f_{\tilde{A}}(y), f_{\tilde{B}}(y)), (b) \tag{2}$$

where U is the universe of information and y is the element of U. To make the operator "OR" derivative, we modified it as Equ. 2(b). Then, the fuzzy degree $f_j(X, \mu, \sigma) \in \Theta^{H \times W \times D}, \Theta \in [0, 1]$ of the j-th channel can be obtained based on Equ. (1) and Equ. (2) as

$$f_j(X, \mu, \sigma) = \bigvee_{i=1}^{m} e^{\frac{-(X_j - \mu_{i,j})^2}{2\sigma_{i,j}^2}} = \max(e^{\frac{-(X_j - \mu_{i,j})^2}{2\sigma_{i,j}^2}}), \tag{3}$$

where $\bigvee$ indicates the union operation. Finally, the output tensor of proposed FAG has the same shape as input X, providing a voxel-wise attention map α^F.

2.2 Global-Local Cube-tree Fusion

To the best of our knowledge, most mask render-based two-stage semantic segmentation methods [6, 20] operate equally on all dense points of the coarse masks to improve the final performance, which is unnecessary to focus much on the already correctly predicted points. As shown in Fig. 1 and according to our statistical error analysis, most very vulnerable points occur on the object border due to the information loss caused by down-sample operation in the encoding process, especially for the innumerable bronchioles or arterioles in the tree-like structures. Thus, we only focus on the border vulnerable points and propose the novel global-local cube-tree fusion module. Specifically, (1) we "**recycle**" the down-sample and up-sample operations to produce masks M_d^l and M_u^l, and evaluate the absolute difference M_b^l of them in Fig. 1 to get the border vulnerable points C_b^l for the l-th layer; (2) as shown in the top-right side of Fig. 2, we build the cube-tree of the i-th point $P_i^l \in C_b^l$ by extracting the local contextual features $\{F_{i,26}^l, F_{i,27}^{l+1}\}$ of {26, 27}-neighbors of the {l, l+1}-th layers respectively, which are defined as the $3 \times 3 \times 3$ cube without and with centroid. For the last layer, it is of note that we extract the 27-neighbors' local contextual features $F_{i,27}^{l-1}$ in the adjacent layer l-1; (3) we flatten features $\{F_{i,26}^l, F_{i,27}^{l+1}\}$ in the spatial dimension and project them as well as centroid feature F_i^l into three vectors

$\{F^l_{i,fg}, F^l_{i,cg}, F^l_{i,ct}\}$, which are separately related to the fine grain, coarse grain local context information and cube-tree centroid feature; (4) global airway or artery features from the distribution of the whole dataset is also very important, hence, we introduce the learnable global features $\tilde{F}^l_g \in R^d$ to yield the projected global features F^l_g, where $d \in \{32, 64, 128, 256\}$ is the embedding dimension; (5) we fuse the four features into $F^l_{i,ff}$ as follows:

$$F^l_{i,ff} = \lambda_1 F^l_{i,cg} + \lambda_2 F^l_{i,ct} + \lambda_3 F^l_{i,fg} + \lambda_4 F^l_g, \tag{4}$$

where $\lambda_1 \sim \lambda_4 \in [0, 1]$ are the learnable coefficients to balance the importance of each feature; (6) we lastly add the feature $F^l_{i,ff}$ to the relative position embedding features $F^l_{i,pe} \in R^{C_1 \times H \times W \times D}$ (retaining the topology information for inductive bias) for the voxel-wise decoding and refined prediction. Obviously, our proposed global-local cube-tree fusion module focuses merely on all border vulnerable points in Fig. 1(f) rather than all regular dense points in Fig. 1(c), which is more related to the lung organ regions. Experimental results demonstrate the efficacy of this design.

2.3 Network Optimization

We define a total loss jointly optimizing the model in an end-to-end manner. The **ordinary loss** in Equ. 5 is employed to supervise the first stage training of the network and produce the coarse mask predictions.

$$L_{\texttt{ol}} = \sum\nolimits_{l=1}^{4} \{\lambda^l_o L_d(P^l, Y^l) + \lambda^l_o L_b(P^l, Y^l)\}, \tag{5}$$

where L_d, L_b are Dice loss and BCE loss separately. (P^l,Y^l) is the prediction and ground truth of the segmentation in the deep layer l. $\lambda^l_o \in \{0.5, 0.3, 0.1, 0.1\}$ are balance parameters. The **boundary rendering loss** in Equ. 6 will supervise the training of the second stage network and produce the fine mask predictions.

$$L_{\texttt{brl}} = \sum\nolimits_{l=1}^{4} \{\lambda^l_{br} L_d(P^l_{br}, Y^l_{br}) + \lambda^l_{br} L_b(P^l_{br}, Y^l_{br})\}, \tag{6}$$

where (P^l_{br},Y^l_{br}) is the voxel-wise border prediction and ground truth in the deep layer l. $\lambda^l_{br} \in \{0.5, 0.3, 0.1, 0.1\}$ are balance parameters. The **total loss** $L = L_{\texttt{ol}} + L_{\texttt{brl}}$ consists of the ordinary loss and boundary rendering loss.

3 Experiments

Datasets. We trained and compared our model with others using chest CT scans from the public BAS airway dataset and PARSE22 [9] artery dataset respectively. Besides, public AeroPath [14] and our in-house Lung fibrosis datasets are used for tests. BAS includes 90 cases, 20 cases from EXACT'09 and 70 cases from LIDC. (1) EXACT'09 [8] owns 20 cases for training and 20 cases for test (without labels), scanning from normal conditions to lung disease patients. LIDC has 70

cases with labels [12]. Lung fibrosis dataset has 25 labeled cases. AeroPath has 27 cases from patients with various pathologies. **Experiment setup:** We divide BAS dataset into 72/18 cases for train/test; Studies on PARSE2022 dataset follow official train/val/test split. The BAS and PARSE22 scans are both cropped as $128 \times 96 \times 144$ patches for training. All modules are trained by sample random flip for 120 epochs, an initial learning rate of 10^{-3}, an AdamW optimizer. The whole project is realized by Pytorch & MinkowskiEngine libraries.

3.1 Qualitative analysis

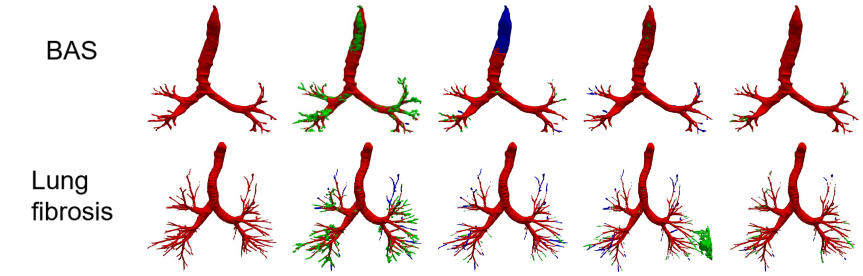

Fig. 5. Qualitative airway segmentation on BAS/Lung fibrosis datasets. GT: ground truth. Red color: true positive. Green color: false positive. Blue color: false negative.

We qualitatively analyze our method on four challenging lung organ datasets. In Fig. 5, SFCN [18] suffers from severe false positives and some false negatives, especially for the big green areas of airway leakages. WNet [21] is mainly influenced by false negatives on the main trachea. For the Fibrosis dataset at the third row, it also encounters the false negative problem in the terminal bronchioles moderately. FANN [10] bears the slight discontinuity issue of false negative in the terminal bronchioles of BAS dataset, and the severe discontinuity and airway leakage problems on the more challenging Fibrosis benchmark. Instead, due to the above two novel modules, our method can solve the defects of false negative, discontinuity, and leakages faced by past advanced methods. Besides, the results on PARSE22 artery dataset in supplementary Fig. 6 also proves this.

3.2 Quantitative analysis

We accurately compare our method with other advanced models in Tables 1-2.

Evaluation metrics. The metrics are referred to method FANN [10] and diverse, including IoU, precision, DLR, DBR, AMR, and an union metric CCFs that concurrently evaluates the core factors of continuity & completeness for airway & artery segmentation.

Table 1. Comparisons on the public BAS/Lung fibrosis datasets. All values are denoted by the percentage (%) of mean/std. Red font are the best results. DLR/DBR: detected length/branch ratio, AMR: airway missing ratio. "$\star$" depicts statistical significance (with Wilcoxon signed-rank test p-value < 0.05) compared with our method.

Methods	**BAS**					
	IoU ↑	Precision ↑	DLR ↑	DBR ↑	AMR ↓	CCFs ↑
nnUNet [6]	88.05/3.13	94.36/2.34*	86.84/7.00*	79.21/9.43*	6.96/4.02*	87.50/4.16*
NaviAir [17]	83.53/3.32*	86.76/4.01*	87.34/7.16*	81.01/9.52*	4.13/3.04*	85.01/3.57*
PSAR [15]	81.33/5.18	86.00/4.01	89.02/9.67	84.39/12.61	6.23/5.05	–/–
FANN [10]	87.38/4.45	91.87/3.20	92.71/7.93*	89.01/10.3*	5.22/4.50	89.69/5.54*
Ours	87.91/3.07	92.32/3.36	95.61/4.55	93.29/5.75	5.46/3.34	91.12/3.22
Methods	**Lung fibrosis**					
	IoU ↑	Precision ↑	DLR ↑	DBR ↑	AMR ↓	CCFs ↑
nnUNet [6]	83.12/4.95*	93.81/3.14*	58.15/6.80*	50.18/7.93*	11.74/2.93*	69.72/5.64*
NaviAir [17]	80.79/5.33*	92.51/1.61*	59.93/14.41*	51.47/14.89*	13.45/6.45*	69.08/11.60*
PSAR [15]	72.72/6.31	78.79/8.16	72.42/10.96	65.50/12.66	9.16/3.25	–/–
FANN [10]	82.69/4.02*	89.04/3.73	78.98/8.00*	73.44/9.54*	7.95/2.37*	80.99/5.17*
Ours	83.81/4.64	89.87/4.12	85.10/8.58	80.01/10.17	7.10/2.33	84.39/5.58

Table 2. Comparison on the public validation set of PARSE22. All values are from the official evaluation with the percentage (%) of multi-level dice coefficient.

Methods	**Main artery**				**Branch artery**				**Weighted Average**			
	25pc ↑	50pc ↑	75pc ↑	mean	25pc ↑	50pc ↑	75pc ↑	mean	25pc ↑	50pc ↑	75pc ↑	mean
NaviAir [17]	84.50	88.63	89.87	87.11	55.87	62.85	66.41	61.40	63.05	67.77	70.72	66.54
nnUNet [6]	89.51	92.63	94.96	91.33	79.77	85.48	87.71	82.54	81.82	86.69	88.88	84.29
FANN [10]	90.31	92.55	94.16	91.96	75.23	81.74	84.81	80.19	78.54	84.36	86.26	82.54
Ours	91.73	92.85	94.60	92.27	79.15	85.71	87.41	83.13	81.87	87.36	88.80	84.96

Comparison on BAS dataset. In the top of Table 1, our FABR obtains the best performance with a 91.12% CCFs, 95.61% DLR, and 93.29% DBR. NaviAir [17] has the lowest AMR (4.13%), while it performs poorly at the metrics of 83.53% IoU, 86.76% precision and 81.01% DBR. Even if nnUNet [6] acquires the best IoU and precision scores, its DLR and DBR metrics are unsatisfied. FANN achieves a suboptimal performance (89.69% CCFs, 92.71% DLR, 89.01%DBR).

Comparison on fibrosis dataset. Although it's the very challenging benchmark, our FABR still behaves robustly and exceeds the best method FANN by 3.4% CCFs with a total metrics of 84.39% CCFs, 83.81% IoU, 85.1% DLR, 80.01% DBR. The lowest AMR (7.1%) confirms that our method can solve the discontinuity issue well. Other methods also behave similarly to the BAS dataset. As seen in the two datasets, the main improvements of our method are consistently at the IoU, DLR and DBR metrics, which are mainly influenced by bronchioles and trachea borders that are easily lost due to network down/up-samples. Hence, our method can extract the robust bronchiole features and render border well via the two novel modules for the accurate lung organ segmentation.

Comparison on PARSE22 dataset. This dataset is more challenging due to more dense small bronchioles shown in supplementary Fig. 6. However, our method still reaches the best weighted average multi-level dice of 84.96% in Table 2 compared against some advanced methods via the official evaluation. As you can see, the remarkable gain comes from the "branch artery", which maintains the consistency with above airway segmentation.

Ablation studies. To verify the efficacy of each module, we perform the thorough ablation studies in supplementary Tables 3-5 and Figs. 7-8. In Table 3, the 2-*th* row on lung fibrosis dataset with the proposed FA-based transformer-like backbone achieves the largest 2.24% △CCFs, verifying the efficacy of fusing fuzzy sets and deep network to diminish the uncertainty in feature representations significantly. The 3-*th* row with GLCF module indicates 1.02% △CCFs, proving that we only need to focus much on the very hard BVP rather than all regular dense points, which provide the most important losing information of discontinuity or details in the network down-sample operation. Since we only extract the BVP to render, it can suppress the redundant background to further solve the severe class imbalance issue of foreground and background voxels. Supplementary Table 4 evidences the efficacy of GLCF module which improves the border accuracy obviously by 4.72%. In Table 5, the 2-*th* row with FA-based transformer-like backbone improves the DBR significantly on the terminal (1.8%), small (1.25%) and medium (1.65%) branches except the large trachea (-1.03%), for most uncertainty in the feature representations is from the terminal, small and medium branches that are too thin and hard to be discerned while annotating. The 3-*th* row with GLCF module realizes the significant promotion of DBR on the small (2.02%), medium (2.02%) and large (3.09%) branches, which is consistent with Fig. 8 to overcome the issue of detail loss in the network down-sample operation and render the BVP effectively. Supplementary Fig. 7 elucidates that our FA-based transformer-like backbone can enhance the feature representations of lung organs significantly.

4 Conclusion

Automated lung organ segmentation is vital to aid radiologists with lung disease diagnosis and prognosis. However, most prior top-tier methods suffer from the discontinuity, false-negative and leakage issues. Inspired by these, we proposed the innovative method FABR in the paper, which has two novel modules, i.e., (1) Fuzzy attention-based transformer-like backbone, diminishing the uncertainty of lung organ feature representations; (2) The global-local cube-tree feature fusion module, explicitly modeling the border vulnerable points yielded by recycled down/up-sample for accurate lung organ segmentation. Finally, extensive qualitative and quantitative experiments have proven the excellent performance of our method on four challenging lung organ segmentation datasets, involving CT scans of lung cancer, fibrosis, and mild lung diseases.

Acknowledgments. The study was supported in part by ERC IMI (101005122), H2020 (952172), MRC (MC/PC/21013), the Royal Society (IEC/NSFC/211235), NVIDIA Academic Hardware Grant Program, SABER project funded by Boehringer Ingelheim Ltd, NIHR Imperial Biomedical Research Centre (RDA01), Wellcome Leap Dynamic Resilience, UKRI Future Leaders Fellowship (MR/V023799/1), and UKRI Fellowship (EP/Z002206/1).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
2. Fang, Y., Wu, S., Zhang, S., Huang, C., Zeng, T., Xing, X., Walsh, S., Yang, G.: Dynamic multimodal information bottleneck for multimodality classification. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7696–7706 (2024)
3. Gao, X., Jin, Y., Zhao, Z., Dou, Q., Heng, P.A.: Future frame prediction for robot-assisted surgery. In: Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings 27. pp. 533–544. Springer (2021)
4. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
5. Hu, J., Shen, L., Albanie, S., Sun, G.: Squeeze-and-excitation networks. IEEE transactions on pattern analysis and machine intelligence **42**(8), 2011–2023 (2020)
6. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
7. Lin, Y., Liu, L., Ma, K., Zheng, Y.: Seg4reg+: Consistency learning between spine segmentation and cobb angle regression. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24. pp. 490–499. Springer (2021)
8. Lo, P., Van Ginneken, B., Reinhardt, J.M., Yavarna, T., De Jong, P.A., Irving, B., Fetita, C., Ortner, M., Pinho, R., Sijbers, J., et al.: Extraction of airways from ct (exact'09). IEEE Transactions on Medical Imaging **31**(11), 2093–2107 (2012)
9. Luo, G., Wang, K., Liu, J., Li, S., Liang, X., Li, X., Gan, S., Wang, W., Dong, S., Wang, W., et al.: Efficient automatic segmentation for multi-level pulmonary arteries: The parse challenge. arXiv preprint arXiv:2304.03708 (2023)
10. Nan, Y., Del Ser, J., Tang, Z., Tang, P., Xing, X., Herrera, F., Pedrycz, W., Walsh, S., Yang, G.: Fuzzy attention neural network to tackle discontinuity in airway segmentation. IEEE Transactions on Neural Networks and Learning Systems (2023)
11. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
12. Qin, Y., Gu, Y., Zheng, H., Chen, M., Yang, J., Zhu, Y.M.: Airwaynet-se: A simple-yet-effective approach to improve airway segmentation using context scale fusion. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). pp. 809–813. IEEE (2020)
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
14. Støverud, K.H., Bouget, D., Pedersen, A., Leira, H.O., Langø, T., Hofstad, E.F.: Aeropath: An airway segmentation benchmark dataset with challenging pathology. arXiv preprint arXiv:2311.01138 (2023)

15. Tang, Z., Nan, Y., Walsh, S., Yang, G.: Adversarial transformer for repairing human airway segmentation. IEEE Journal of Biomedical and Health Informatics (2023)
16. Tsay, J.C.J., Wu, B.G., Sulaiman, I., Gershner, K., Schluger, R., Li, Y., Yie, T.A., Meyn, P., Olsen, E., Perez, L., et al.: Lower airway dysbiosis affects lung cancer progression. Cancer discovery **11**(2), 293–307 (2021)
17. Wang, A., Tam, T.C.C., Poon, H.M., Yu, K.C., Lee, W.N.: Naviairway: a bronchiole-sensitive deep learning-based airway segmentation pipeline for planning of navigation bronchoscopy. Authorea Preprints (2023)
18. Wang, C., Hayashi, Y., Oda, M., Itoh, H., Kitasaka, T., Frangi, A.F., Mori, K.: Tubular structure segmentation using spatial fully connected network with radial distance loss for 3d medical images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. pp. 348–356. Springer (2019)
19. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., Xie, S.: Convnext v2: Co-designing and scaling convnets with masked autoencoders. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16133–16142 (2023)
20. Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y.: St++: Make self-training work better for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4268–4277 (2022)
21. Zheng, H., Qin, Y., Gu, Y., Xie, F., Yang, J., Sun, J., Yang, G.Z.: Alleviating class-wise gradient imbalance for pulmonary airway segmentation. IEEE transactions on medical imaging **40**(9), 2452–2462 (2021)