Airway Anomaly Detection by Prototype-based Graph Neural Network

Tianyi Zhao and Zhaozheng Yin

Stonybrook University, Stonybrook NY, USA

Abstract. Detecting the airway anomaly can be an essential part to aid the lung disease diagnosis. Since normal human airways share an anatomical structure, we design a graph prototype whose structure follows the normal airway anatomy. Then, we learn the prototype and a graph neural network from a weakly-supervised airway dataset, i.e., only the holistic label is available, indicating if the airway has anomaly or not, but which bronchus node has the anomaly is unknown. During inference, the graph neural network predicts the anomaly score at both the holistic level and node-level of an airway. We initialize the airway anomaly detection problem by creating a large airway dataset with 2589 samples, and our prototype-based graph neural network shows high sensitivity and specificity on this new benchmark dataset.

Keywords: Anomaly detection · Bronchus classification · Graph neural network.

1 Introduction

Computer-aided diagnosis (CAD) becomes more and more important to assist doctors in the interpretation of medical images, especially in the pandemic situation. Many lung-related image analysis tasks have been investigated such as lung segmentation [1], nodule detection [2] and airway segmentation [3–7]. A complete lung-related CAD system should not only detect if the disease exists, but also provide detailed analysis on the disease including localizing which lobe the region-of-interest (ROI) belongs to. The clinical definition of different lobes are directly related to the bronchus hierarchy. Recently, a classification algorithm is proposed to label the bronchus, which can help segment the lobe [7]. However, this algorithm has an assumption that the bronchi to be classified follow the anatomical structure. It is unaware of the anomaly appearing in the airway structure. In fact, detecting anomaly in an airway structure can aid the lung disease diagnosis. If a CAD system can provide an anomaly score for every bronchus, this could be considered as a new digital bio-marker (i.e., the anomaly bronchi deserve special attention or treatment from the doctor).

A normal human airway follows an anatomical structure that has 5 lobes and 18 segments with their specific characteristics, as shown in Fig.1(a). An anomaly airway does not fully follow the pre-defined structure due to its disease or surgery (e.g., Fig.1(b)). Inspired by these two observations, we propose a

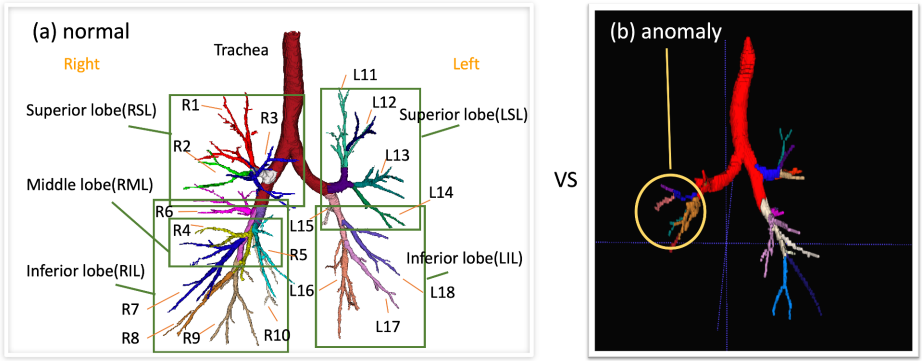

Fig. 1. (a) Normal human airway follows a common anatomical structure. (b) An airway with anomaly exhibits large airway tree variations compared to the anatomical structure, indicating the possible disease area or related surgery.

prototype-based airway anomaly detection algorithm, where the prototype is a learned graph representation of the normal airway and a graph neural network is learned to estimate the anomaly score for each bronchus node of an airway.

Though detecting airway anomaly is valuable to aid lung disease diagnosis, unfortunately there is no dataset with related labels in the community yet. Particularly, labeling the bronchus as anomaly or not one-by-one in CT image stacks is very tedious. For example, the three widely-used public datasets [8–10] only have segmentation labels on lung regions or the disease ROIs in CT images, without the detailed disease/anomaly/normal label on individual bronchus. To initialize the airway anomaly detection problem in our community, we explore a weakly-supervised anomaly detection approach using the existing datasets where only the holistic label is given, indicating if the airway contains anomaly or not, but which bronchus node in the airway has anomaly is unknown. We deploy the multi-instance setting, in which if all the bronchus nodes are normal, then the airway is normal, otherwise, if at least one node is abnormal, the airway has anomaly. In summary, this paper has three-fold contributions:

- We propose an airway anomaly detection algorithm to predict the anomaly score for the holistic airway and for each node of the airway tree. The anomaly score is calculated by a graph convolutional neural network (GCN).
- We propose a virtual prototype whose graph structure follows the anatomical structure of normal airways. The GCN and prototype are alternatively trained such that the GCN generates low anomaly scores on normal airways and the prototype and high anomaly scores on anomaly samples.
- We initialize the airway anomaly detection problem in our lung disease diagnose community by creating a large airway dataset and formulating it in a weakly-supervised way, i.e., given only the holistic level label for training, we train a graph neural network able to predict the anomaly score of an airway at both the holistic and node levels during inference.

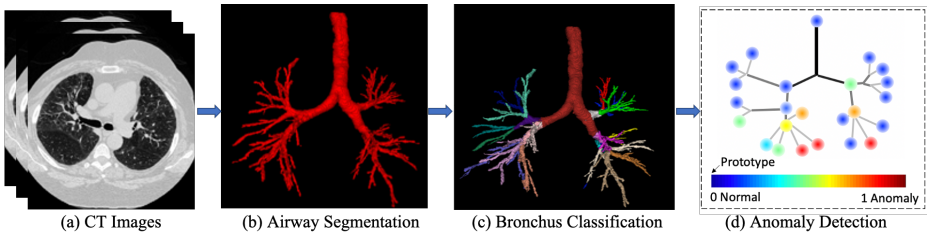

Fig. 2. A typical process of the airway-related CAD system. The airway is first segmented from the CT images. Then the bronchus are classified according to the anatomical structure. Finally, our proposed anomaly detection algorithm processes the graph-based classification result and estimates the anomaly at each node.

2 Related Work

2.1 Airway-related CAD System

The airway tree can provide valuable information for airway disease diagnosis (e.g., Chronic Obstructive Pulmonary Disease (COPD)). A typical process of the airway-related CAD system is shown in Fig. 2. Several Convolutional Neural Network (CNN) based methods have been proposed to segment the airway tree volume from the lung CT images, such as 3D U-Net based methods [3–5], 2.5D CNN [6], and a 2-stage 2D+3D CNN [7]. Based on the airway segmentation, a classification algorithm is proposed in [7] to classify each airway segment to one of the 5 lobes (and one of the 18 segmental bronchus). However, the bronchus classification method is unaware of the anomaly.

2.2 Anomaly detection

Anomaly detection or novelty detection, is to detect an incoming sample with a new unknown class. Novelty detection is also related to the domain adaption problem, where the known class belongs to the source domain, and the novel class belongs to the new target domain. The novelty detection method can be distribution based [11, 12], or probabilistic based [13]. For example, a softmax and threshold baseline is proposed in [11] for out-of-distribution detection. A novel margin-based loss term by self-supervised leave-out classifier is proposed in [12]. In this paper, we invent a prototype-based algorithm, where the anomaly of an airway including its nodes is predicted by a graph neural network.

2.3 Graph Neural Network

Graph cnvolutioal neural network (GCN) is widely used for graph-structure data processing, such as the social network and biological data. The graph construction and layer-wise propagation rules have been described in [14, 15] Recently, GCN has been enhanced in many ways such as Graph Attention Networks [16]

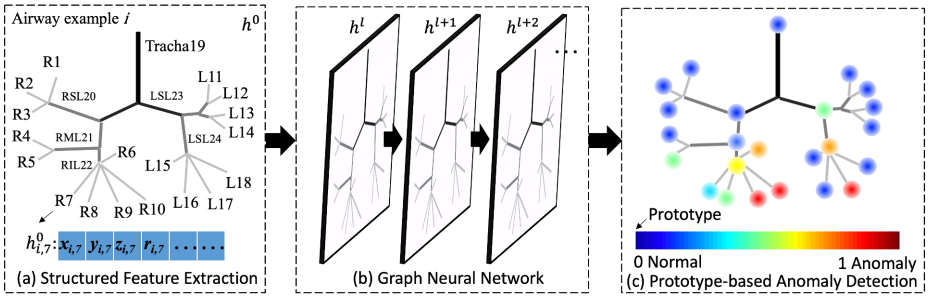

Fig. 3. The overview of our anomaly detection algorithm with three components: feature extraction, graph neural network and prototype-based anomaly prediction.

with self-attentional layers; Graph-ResNet [17] with the residual block; GraphSAGE [18] replacing the embedding vector of each graph node with a set of aggregator functions, and hierarchical deep representations [19, 20] to learn the graph-level representation. In this paper, we integrate an anatomical-based graph module to GCN for the multi-level anomaly detection.

3 Method

The overview of our airway anomaly detection is shown in Fig.3 with three components: (1) The feature extraction component extracts airway features from the airway segmentation results for every node, which composes the input of the graph neural network; (2) then, the graph neural network process the initial graph input; and (3) a prototype based anomaly detection algorithm computes the anomaly score for each node and the entire graph.

3.1 Structured Feature Extraction

Given the segmented and classified airway results shared by [7] (Fig. 2(b-c)), we encode the anatomical structure of the airways, as well as their image properties and graph properties into feature vectors. The feature vectors can be viewed as an anatomical representation of the airway, which will be used in the graph neural network for anomaly detection.

Formally, we define $h^0_{i,k} \in R^D$ as the initial feature vector for node k of the ith airway sample (e.g., $h^0_{i,7}$ in Fig.3(a) represents the initial feature vector of the R7 bronchus node in the right lobe of the ith sample). Based on the anatomical airway structure, we design 7 attributes in the feature representation:
(1) The coordinates of node k of sample i is represented as $(x_{i,k}, y_{i,k}, z_{i,k})$. The coordinates are centralized by the coordinates of the trachea and normalized by the spacing of the CT scan and the bounding box of the airway tree.
(2) The direction of node k of sample i is represented as $(r_{i,k})$ in 3D dimension. The direction is calculated from the coordinates of the current node to the

average coordinates of the leaf nodes of the current node.
(3) The level of the node. The level of the root node (the trachea node) is 0. If the level of a node is l, then the children of this node are at level $l+1$.
(4) The bottom-up level of the node. The level of the leaf node is 0. If the level of a node is l, then the parent of this node is at level $l+1$.
(5) The length of the node is calculated from the coordinates of the current node to the coordinates of the furthest leaf node of the current node.
(6) The number of the descendants of the node.
(7) The image feature is extracted from the last layer of segmentation network based on the coordinates of the node. The dimension of the image feature is 64. The dimension of the entire feature vector is $D = 74$.

3.2 Graph Neural Network

After obtaining feature vectors $h^0_{i,k}$ for all individual nodes, we feed them into a graph convolutional neural network (GCN) modeling the dependencies of neighboring nodes in the tree graph. We use a shallow graph neural network which contains six graph convolutional layers and performs the anomaly detection task under multiple levels (i.e., the holistic level and the node level). Each graph-based convolutional layer is defined as:

$$h^{l+1}_{i,k} = w_0 h^l_{i,k} + \sum_{k' \in N(k)} w_1 h^l_{i,k'}, \tag{1}$$

where $h^l_{i,k}$ is the feature vector of node k of sample i at layer l of the GCN. $N(k)$ represents the neighbor set of the node k. w_0 and w_1 are learned parameters of the graph-based convolutional kernels, with w_1 being shared by all edges. The graph convolutional layer accounts for the way in which nodes neighboring a given node regularize the node-to-neighbor exchange of information.

3.3 Prototype-based Anomaly Detection

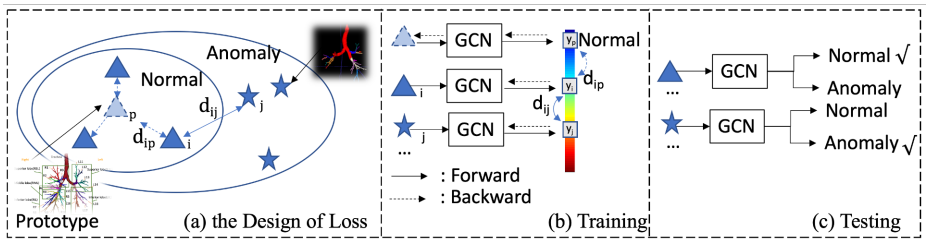

Fig. 4. (a) The design of loss to learn the prototype and GCN. (b) The training process on GCN and prototype. (c) Prediction during testing.

Our graph network (GCN) generates an anomaly score ($\hat{y}_{i,k} \in [0,1]$) for every node k of airway sample i (i.e., $G(h_i^0) : R^{K \times D} \rightarrow [0,1]^K$, where G represents the GCN function, h_i^0 is the initial segmented airway graph of sample i, and K is the number of graph nodes, $K = 24$). In our study, we only have the holistic label, indicating if the airway is normal or anomaly, but we do not have the node-level label. During the training, we label all the nodes based on the holistic label, i.e., $y_{i,k} = y_i$. During the testing, if at least one node is anomaly, the airway is anomaly (i.e., if any $\hat{y}_{i,k} = 1$, then $\hat{y}_i = 1$), otherwise, the airway is normal.

We define the prototype p of normal airways by its node feature matrix $h_p \in R^{K \times D}$. Our objective is to learn both the graph neural network function G and the prototype h_p, but G and h_p are intertwined together, so we propose to learn them in an alternative way:

First, suppose an initial prototype h_p^0 is given and the anomaly score for each node of the prototype is 0. The GCN function G is expected to generate a low anomaly score close to 0 for every node of a normal sample i (i.e., $G(h_p^0) \approx G(h_i^0)$). Meanwhile, G is expected to generate a large anomaly score for anomaly sample j. Motivated by these expectations, illustrated in Fig. 4(a) as well, we define the following loss function to train the GCN:

$$L_{GCN} = \sum_{i,y_i=0} d_{ip}(\bar{G}(h_p^0), G(h_i^0)) - \sum_{i,y_i=0} \sum_{j,y_j=1} d_{ij}(G(h_i^0), G(h_j^0)), \tag{2}$$

where d_{ip} is the distance between the prototype and the normal sample i. By minimizing d_{ip}, the normal training sample i is close to the prototype, leading to a low anomaly score on sample i. d_{ij} is the distance between a normal sample i and an anomaly sample j. By maximizing d_{ij}, we maintain the discrimination between normal and anomaly samples. We use the Euclidean distance for d_{ip} and d_{ij}. Note that, the GCN output on the prototype is represented by $\bar{G}(h_p^0)$, which means the prototype is fixed when training G.

Secondly, we assume G is given, represented by $\bar{G}$. Then, we learn the prototype h_p by minimizing the following loss function:

$$L_{prototype} = \sum_{k}^{K} f(\bar{G}(h_p)_k, 0), \tag{3}$$

where $f(\cdot)$ is the binary cross entropy function. The alternative GCN and prototype learning is summarized in Alg.1 below.

Algorithm 1: The training procedure for GCN and prototype

Initialization: $h_p^0 = \frac{1}{N}\sum_i^N h_i^0$ on N samples, G is randomly initialized;
for t =*1:max iteration* **do**
  $G^{(t)} \leftarrow argmin_G$ Eq.2;
  $\bar{G} \leftarrow G^{(t)}$;
  $h_p^{(t)} \leftarrow argmin_{h_p}$ Eq.3;
  $h_p^0 \leftarrow h_p^{(t)}$;
end
Result: G, h_p *// The GCN function and the prototype*

4 Experiments

Dataset: We collected datasets from 3 resources: The Lung Image Database Consortium (LIDC) [8], The Lung Tissue Research Consortium (LTRC) [9], and The National Lung Screening Trial (NLST) [10], leading to 62 normal samples and 23 anomaly samples in total. The available airway data in our community is quite limited and it is even costly to label node-level anomaly. Meanwhile, we observe three most common anomaly patterns: switch, shift and cut. Thus, we propose three augmentation methods to generate more anomaly samples from the collected dataset: (1) **Switch**: randomly switch two bronchus within the same lobe; (2) **Shift**: a bronchus is randomly shifted to one of its descendants; and (3) **Cut**: cut an airway branch and re-label the airway by the bronchus classification algorithm [7]. After the augmentation, we have 2396 training samples with 40 of them being normal samples, and 193 testing samples with 22 samples being normal samples. In total, a dataset of 2589 samples is established.

Evaluation Metric: We use two evaluation metrics in the experiment: the sensitivity that is the proportions of the correctly predicted anomaly data, and the specificity that is the proportions of the correctly predicted normal data.

Ablation Study: We evaluate the effectiveness of GCN by comparing it with CNN and evaluate the effectiveness of prototype by comparing networks with or without it. The ablation study is summarized in Table 1. Since our dataset has more anomaly samples than normal samples, the CNN w/o prototype method gets a high sensitivity but a low specificity, which means the method tends to predict the anomaly. With the prototype, specificity is increased, since the prototype represents and emphasizes the normal samples during training. The GCN w/o prototype method has a better average score than the CNN w/o prototype method. Compared to the CNN model, GCN can exchange information between neighboring nodes. In the anomaly case, only a small portion of the nodes are anomaly nodes. The GCN can make the prediction of each node more robust. Overall, our GCN with the prototype method has the best performance.

Table 1. The ablation study results.

method	sensitivity	specificity	average
CNN w/o prototype	**0.964**	0.681	0.822
CNN w prototype	0.935	0.772	0.853
GCN w/o prototype	0.923	0.818	0.871
GCN w prototype	0.912	**0.954**	**0.933**

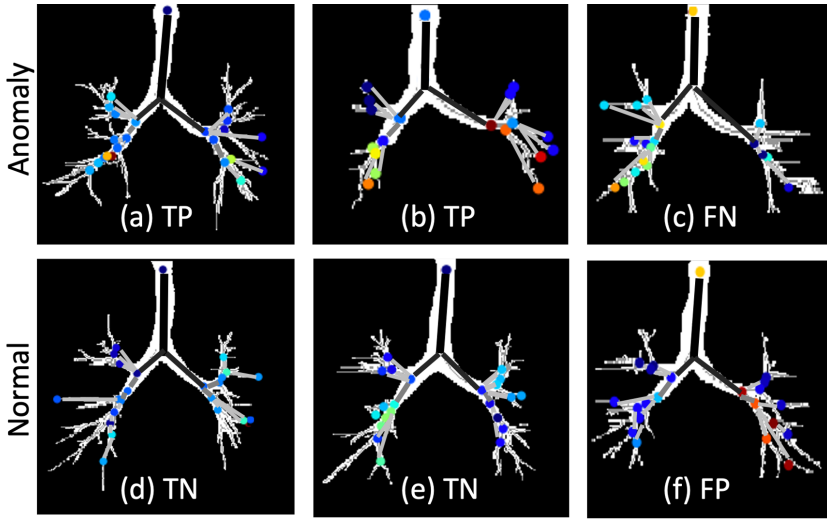

Fig. 5. Anomaly score of each node overlaid on the airway via heatmap color scheme.

Qualitative Analysis: Some qualitative results are provided in Fig. 5. The segmented airway tree is represented by white pixels. The anomaly score is represented by color dot at each node (blue: low anomaly score; red: high anomaly score). In Fig. 5(a), one bronchus is correctly predicted as the anomaly and in Fig. 5(b), multiple bronchi are correctly predicted as the anomaly. A false negative example is given in Fig. 5(c). The anomaly in this example is very minor, where some bronchi are too thin to detect. Three normal samples are given in Fig. 5(d-f). Most of the bronchi are correctly predicted as normal, except some bronchi in the left inferior lobe of Fig.5(f). In this case, multiple segmental bronchi are very close to each other, causing the confusing clutter.

5 Conclusion

We propose a prototype-based anomaly detection algorithm to predict the anomaly score for each node of an airway through a graph neural network, which can aid the lung disease diagnosis. We initialize the airway anomaly detection problem, and the newly created dataset along with our source codes will be released to the community.

References

1. Zhao, Tianyi, Dashan Gao, Jiao Wang, and Zhaozheng Yin. "Lung segmentation in CT images using a fully convolutional neural network with multi-instance and conditional adversary loss." ISBI, pp. 505-509, 2018.
2. Ding, Jia, Aoxue Li, Zhiqiang Hu, and Liwei Wang. "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks." MICCAI, pp. 559-567, 2017.
3. Jin, Dakai, Ziyue Xu, Adam P. Harrison, Kevin George, and Daniel J. Mollura. "3D Convolutional neural networks with graph refinement for airway segmentation using incomplete data labels."MLMI, 2017.
4. Juarez, Antonio Garcia-Uceda, Harm AWM Tiddens, and Marleen de Bruijne. "Automatic airway segmentation in chest CT using convolutional neural networks. " Image Analysis for Moving Organ, Breast, and Thoracic Images. 238-250, 2018.
5. Meng, Qier, Holger R. Roth, Takayuki Kitasaka, Masahiro Oda, Junji Ueno, and Kensaku Mori. "Tracking and segmentation of the airways in chest CT using a fully convolutional network." MICCAI, 2017.
6. Yun, Jihye, Jinkon Park, Donghoon Yu, Jaeyoun Yi, Minho Lee, Hee Jun Park, June-Goo Lee, Joon Beom Seo, and Namkug Kim. "Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net." Medical Image Analysis (MedIA), 51:13-20, 2019.
7. Zhao, Tianyi, Zhaozheng Yin, Jiao Wang, Dashan Gao, Yunqiang Chen, and Yunxiang Mao. "Bronchus segmentation and classification by neural networks and linear programming." MICCAI, 2019.
8. Armato III, Samuel G., Geoffrey McLennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, Binsheng, Zhao *et al.* "The lung image database consortium (lidc) and image database resource initiative (idri):A completed reference database of lung nodules on ct scans." Medical Physics, 38:915–931.
9. Bartholmai, B., R. Karwoski, V. Zavaletta, R. Robb, and D. R. I. Holmes. "The Lung Tissue Research Consortium: An extensive open database containing histological, clinical, and radiological data to study chronic lung disease." MICCAI Open Science Workshop, 2006.
10. Clark, Kenneth, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore *et al.* "The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository." Journal of Digital Imaging, 26(6):1045-1057, 2013.
11. Vyas, Apoorv and Jammalamadaka, Nataraj and Zhu, Xia and Das, Dipankar and Kaul, Bharat and Willke, Theodore L. "Out-of-distribution detection using an ensemble of self supervised leave-out classifiers." ECCV, pp. 550-564, 2018.
12. Dan Hendrycks and Kevin Gimpel. "A baseline for detecting misclassified and out-of-distribution examples in neural networks." ICLR, 2017.
13. Eskin, Eleazar. "Anomaly detection over noisy data using learned probability distributions." ICML, 2000.
14. Bruna, J., Zaremba, W., Szlam, A., and LeCun, Y. "Spectral networks and locally connected networks on graphs." arXiv preprint arXiv:1312.6203, 2013.
15. Bruna, J., Zaremba, W., Szlam, A., LeCun, Y. "Semi-supervised classification with graph convolutional networks." arXiv preprint arXiv:1609.02907, 2016.
16. Veličković, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. "Graph attention networks." arXiv preprint arXiv:1710.10903, 2017.

17. Wang, Nanyang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. “Pixel2Mesh: generating 3D mesh models from single RGB images.” , ECCV, 2018.
18. Hamilton, William L., Rex Ying, and Jure Leskovec. “Inductive representation learning on large graphs.” arXiv preprint arXiv:1706.02216, 2017.
19. Ying, Rex, Jiaxuan You, Christopher Morris, Xiang Ren, William L. Hamilton, and Jure Leskovec. “Hierarchical graph representation learning with differentiable pooling.” arXiv preprint arXiv:1806.08804, 2018.
20. Huang, Jingjia, Zhangheng Li, Nannan Li, Shan Liu, and Ge Li. “AttPool: Towards hierarchical feature representation in graph convolutional networks via attention mechanism.” ICCV, pp. 6480-6489. 2019.