University of New South Wales
School of Computer Science and Engineering
Sydney, Australia

Report for COMP9991,9993 Research Project A,C

DEEP LEARNING FOR IMPROVED BRONCHIAL TREE SEGMENTATION IN CT IMAGES

Oleg Kufirin
z5216713
oleg.kufirin@student.unsw.edu.au

Supervisor: Erik Meijering
University of New South Wales, Sydney, Australia

Assessor: Daniel Moses
University of New South Wales, Sydney, Australia

December 2020

Abstract

This project focuses on application of deep learning based approaches, particularly recurrent neural networks (RNN), in the problem of peripheral bronchial identification and tracking on computed tomography (CT) of human's chest. Potential of RNN for distal airway tracking has been explored and compared with conventional methods such as B-spline interpolation in three dimensional space. General framework and algorithm for airway localisation and tracing using combination of convolutional neural network (CNN) and RNN or B-spline interpolation are given.

The work is primarily based on the results obtained by professor Daniel Moses at University of New South Wales in his PhD dissertation "Computerised analysis of the bronchovascular anatomy of the lung on multidetector CT studies" (2018) and uses that work as the baseline method for research and improvements outlined in the current project.

Limitations of the proposed approach, such as inconsistency in sampling technique of cross-sectional images, imposed by the baseline method and data it generated, are explored and recommendations for improvements and future research are proposed.

Contents

List of Figures

List of Tables

1. Introduction

Many human diseases affect the bronchial tree (part of the airways toward the lungs), including chronic bronchitis, asthma, bronchiolitis and bronchiectasis. The latter, for example, results in the permanent abnormal dilatation of bronchi with thickening of their walls, and is seen in diseases such as cystic fibrosis. Chronic bronchitis, on the other hand, results in thickening of the bronchial walls without dilatation.

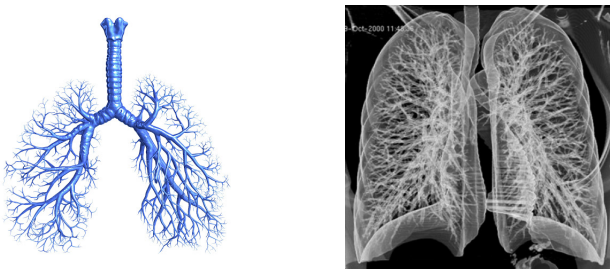

Figure 1: Left: graphic illustration of normal bronchial tree. Right: example CT scan of normal bronchial tree

Several of these diseases preferentially affect the small to medium airways. This can be visualized in patients by means of computed tomography (CT) imaging (see Fig. 1). Accurate segmentation and analysis of the bronchial tree, including the smaller airways, in CT images allows identification and quantitative assessment of the diseases and helps to determine optimal treatment.

A great challenge in segmenting the airways with medium to smaller diameters is the highly variable intensity levels within the lumen. In addition, the bronchi may sometimes be interrupted, caused by other pathologies (tumors or mucous plugging) or technical artifacts (movement of the lung by respiration or heart beat during scanning). Thus powerful segmentation methods are highly needed. [1]

In addition due to the COVID-19 pandemic in 2020 the number of generating lung CT scans in the world has exploded and hence the need for automatic processing and analysis of scans is of paramount importance.

The conventional extraction methods such as region growing can easily capture the main bronchi, but systematically fail at extracting the peripheral bronchi of smaller size. Other, more advanced methods, coupled traditional computer vision techniques with the supervised and semi-supervised machine learning approaches and obtained good separation results. [1] With the recent advent of deep learning (DL) the research community started gradually employing it for solving segmentation problem of tree-like anatomical structures such as neurons, vascular trees and an airway tree. The idea of using DL for segmentation is still relatively new and has a potential for exploration.

A common and straightforward approach is to employ convolutional neural networks (CNN) with various modifications to solve the task. Detailed overview of the conventional and modern approaches will be given in part 2 of this report. Despite being an active research area, very little work has been done to investigate potential of recurrent neural networks (RNN) in the context of given problem. Many proposed methods have an underlying assumption that the current prediction is only affected by the last tracking step (first-order Markovian process) ignoring effect of previous points along the path. Ability of RNN, such as LSTM, to remember states and use in temporal-like sequences, can potentially help to improve the tracking algorithm. The main purpose of this project is to explore the ability of RNN to improve segmentation. However, scope of this work is not limited to RNN only, but can explore other techniques along the way if necessary.

The ultimate goal of the work is an attempt do design and implement a novel method or improvement to the current methods of bronchial tree segmentation using DL with RNN component and compare with the current state of the art methods.

2. Related Work

In this chapter, we will give a short, but comprehensive review of the airway segmentation methods emerged in the last two decades. Roughly all the methods can be divided into two broad groups. First, is the methods based on conventional computer vision techniques commonly utilising hand-crafted features and machine learning classifiers. These approaches were extensively being explored and researched approximately by 2014. In 2012 the breakthrough paper "ImageNet Classification with Deep Convolutional Neural Networks" by Krizhevsky et al. [2] was published describing the power of CNN for image classification. Since that time, the main research direction in airway segmentation gradually started shifting to deep learning-based methodology. The following two sections will describe the key research ideas and primary outcomes in both periods.

2.1. Pre deep learning era (by 2014)

The dramatic rise of computing power from 2000 made application of automatic image processing possible at low cost. Sufficient processor characteristics and RAM are critical for processing medical images due to their high resolution, and they are often performed in 3D. In addition to machine learning algorithms, that previously had had a well-developed theoretical base, could be implemented and run within a reasonable time. All these aspects dictated the way the segmentation techniques were developed.

During that time, the traditional computer vision techniques were employed, such as region growing, thresholding, mathematical morphology and template matching. [3] In 2009, an extensive research competition called "Extract of Airways from CT" (EXACT'09) was launched under supervision from the research organisations in Denmark, the Netherlands and the USA. This work describes a framework for establishing a reference airway tree segmentation, which was used to quantitatively evaluate 15 different airway tree extraction algorithms in a standardised manner. The main techniques the teams used are listed above and were mainly centred around the region growing (RG). Some, like van Rikxoort et al. [4] coupled RG with rule-based algorithms or like Lo et al. with KNN machine learning and rule-based decision algorithms. [5]

A voxel-based fusion scheme for all fifteen methods was implemented by applying a sequential forward selection technique to assess whether the proposed algorithms complement each other. Fig. 2 (Left) shows average length detected (%) on the y-axis and false positive rate (%) on the x-axis for all fifteen methods and the fusion scheme is indicated with a star. It can be clearly seen that there exist a trade-off between accuracy (FPR) and tree length. The fusion scheme has superior performance among individual algorithms, and it proves that indeed all techniques provide complementary information and there is no clear winner in the competition. Fig. 2 (Right) shows that 13.7% of the branches were detected by one algorithm and only 18.3% were detected by all fifteen methods. Hence, there is a modest diversity in the outcomes of the algorithms.

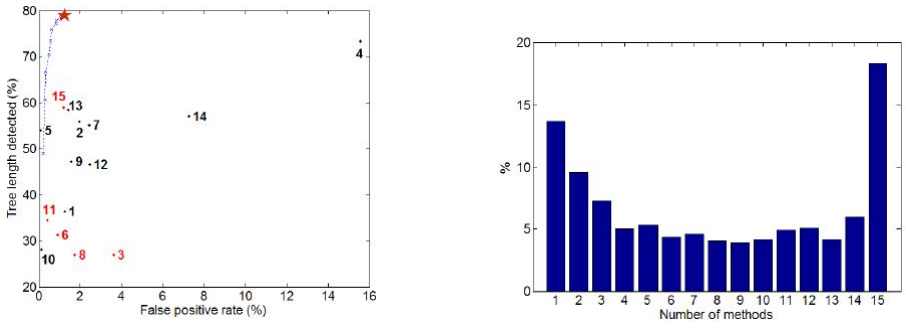

Figure 2: Left: Average tree length versus average false positive rate of all algorithms with the algorithms in the semi-automated category in red. Right: Bar plot shows the percentage of branches detected versus the number of algorithms that detected them, averaged across all test cases.

Since publishing the results in 2012, EXACT'09 has been a prevalent source of training data (publicly available online) and serves as a comparison baseline for novel approaches, including deep

learning. The reference data has not yet been revealed to the public since the challenge is still open for submissions at the time of writing this report (November 2020). Noticeably at present, there exist no single method that could achieve even close performance to the fusion scheme obtained in EXACT'09 despite trying more complex hypothesis and employing even greater computational power. It can be said that the fifteen algorithms nicely summarise the research ideas and directions of the decade. Next milestone was the advent of CNN in 2012 [2] and further application in the domain of interest of this report.

2.2. Deep learning era (2014 - now)

One of the first attempts to tackle the problem with deep learning was a work by Charbonnier at et al. [6] where they used 2D convolutional networks to classify initially extracted branches either as airway or leakage using cross-sectional image patches. In the core of the method, they used one of the algorithms from EXACT'09 and sampled 15 coarse extractions by varying the hyper-parameter. Leakage removal with CNN was performed for each sampled extraction, and then the union of all cleared branches was taken as the final result. The method was able to produce a deep tree (67% length detected) while having a relatively low false-positive rate (around 2%). Even though this method outperformed all the algorithms in EXACT'09, it can be noted that this is not a novel segmentation approach, but rather a way to lower the errors of the existing ones. Since it has to produce 15 coarse base segmentation the running time will also be increased by at least factor of 15.

In 2015 Ronneberber at al. [7] proposed novel convolutional network architecture called U-Net for the task of object segmentation in biomedical imaging. The proposed architecture has down-sampling and up-sampling paths within which series of convolution and max-pooling/up-convolution operations are performed. This has quickly become a gold standard approach for segmentation tasks and also found its application for the airway extraction problem. Jin et al. [8] extended the 2D approach in [7] to 3D (see Fig. 3) and used cropped cubical patches as inputs, sampled along the center-line extracted beforehand. Meng et al. [9] also used 3D U-Net based extraction, but based on dynamically generated 3D VOI and applying branch point detection using the gradient vector flow and tubular-likeness function as they go down the tree. Juarez et al. [10] came up with simple but robust end-to-end U-Net architecture using large patches of size 352x240 pixels. Zhao and Hamarneh [11] used a U-Net like architecture to predict branch direction of an anatomical tree. However, in relation to airways their work is limited to the third generation.

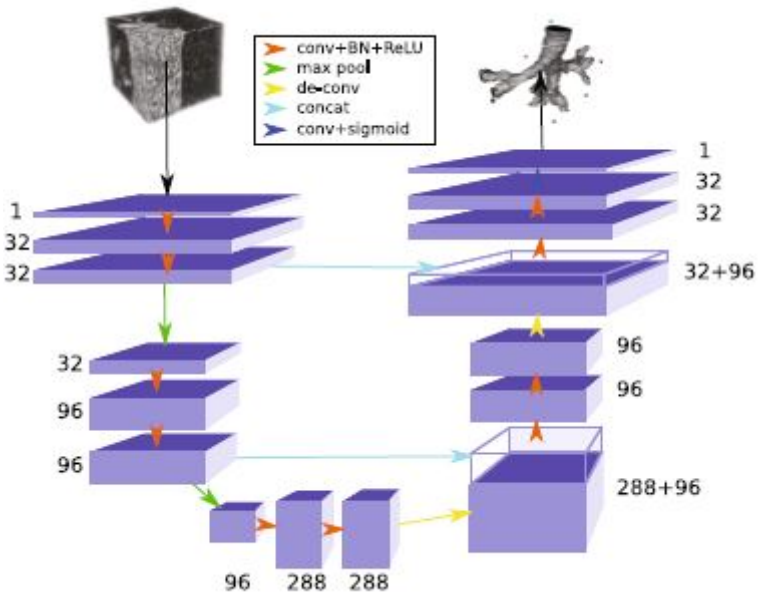

Figure 3: Example of a 3D U-Net architecture from Jin et al. [8]

In more recent works, i.e. Yun et al. [12] used standard contracting CNN, but each potential branch had three sampling points in 2D along three imaging planes. However, it remained unclear how airway-candidate patches were sampled before input into the network. The method was able to achieve tree length similar to top results in EXACT'09 - slightly above 60% but suffered high false-positive rate - above 4%. Zhao et al. [13] tackled the problem by combining 2D CNN and 3D U-net architectures for proximal and distal airways separately. Then they applied linear programming tracking to combine the results. Besides, they performed bronchi classification according to medical standards. Qin et al. [14] proposed a method based on 3D U-net and 26-voxel connectivity aware approach and outperformed the previous results of [8] and [10] by 4% in Dice coefficient.

Another promising direction of research is related to more exotic neural network architectures such as graph neural networks (GNN) and mean-field networks (MFN) [15]. The potential advantage of these methods is the ability to learn inter-dependencies in data for relational reasoning. Juarez expanded his previous work [10] and replaced CNN with GNN at the deepest level of the 3D U-net in [16]. The improvements in detection were minor, but the number of learned parameters reduced by a large margin. More research work in this direction is needed due to the novelty of such architectures and approaching segmentation of a tree-like structure in a structural way seems a reasonable and logical step.

Methods utilising recurrent neural networks (RNN) are still novel with regards to this problem. Zhao and Hamarneh [17] first explored the usage of LSTM architecture in the tracking of anatomical trees (either airway or vessels). In particular, they assessed the prediction of bifurcation points and direction by combining the LSTM module with the CNN module. The results showed improvement by at least 15% comparing with the standalone CNN evaluated on the EXACT'09 data-set. However, the method was evaluated only on the proximal parts of the airway such as the trachea, LMB and RMD, which are relatively easy to extract and get a ground truth.

More advances in RNN application have been made in the field of vessel tracking and segmentation. Kong et al. in their work [18] and [19] came up with a new architecture called 'Tree-structured Convolutional LSTM' and 'Tree-structured Convolutional GRU', where the dot product in LSTM/GRU was replaced with the convolutional operator to incorporate the image features. The RNN units are also connected with a child-parent relationship that makes them suitable for processing tree-structured data. However, the methods are not end-to-end and require prior centre-line extraction from 3D cardiac CTA, which is very challenging do to in 3D chest CT for airways.

In general, the research trend nowadays is going beyond context-less and memory-less approaches. More research and evaluation is needed to investigate the application of RNN to the problem of airway tracking and extraction due to its novelty.

3. Overview of the Baseline Method

This project is based on achievements in [1] from Prof. Moses's PhD thesis "Computerised Analysis of the Bronchovascular Anatomy of the Lung on Multidetector CT Studies". This chapter will give short summary of the results and methods used as it dictated the way the data was sampled for further analysis in the next sections. Significant portion of the thesis was devoted to detection, localisation and analysis of main pulmonary artery, but since it does not have direct relation to the current project, it will be excluded from the summary.

3.1. Central bronchial tree segmentation

The novel technique of using isosurface projection to obtain segmentation of the lumen of the central airway was developed. Details of the algorithm and theoretical base can be found in [1]. Overall, the method seems robust and showed consistent accurate segmentation results of the proximal bronchial tree. At present extraction of trachea and first three generations of bronchi is not a topic of research interest due to good results obtained before by various methods. Most commercially available software are able to do this task with acceptable accuracy and widely used in clinical practise.

3.2. Analysis of the peripheral pulmonary vasculature

The underlying idea of the method proposed in [1] is that bronchi run in parallel to arteries as opposed to veins. Hence, the vascular tree should be extracted at first for further analysis. Below the steps of the algorithm are given and overall diagram can be found in Fig. 4.

1. Lungs are segmented with thresholding and component analysis
2. Hessian filtering is applied (using a Frangi filter) to segment the vascular structure
3. The extracted structure is skeletonised with a 3D thinning algorithm
4. A graph is constructed in order to establish the relationship of all of the branches in the skeletonised data
5. The graph is transformed into a tree data structure to permit analysis of the branches and vessels within the vascular tree in a hierarchical fashion.The leaves (terminal branch points) of the vascular tree can be identified and analysis of the tree branches can occur backwards towards the root.
6. To enable vessel tracing, distance and diameter maps are created and combined into a composite map
7. Since the points in each path mostly lie on the centreline of the associated vessel, a least squares B-spline approximation is fitted to these centre points.
8. The derivative curve of the centreline spline is calculated to provide tangential information and at evenly spaced points a tangent vector is generated. A cross-sectional image is constructed through the original 3D data set at evenly spaced points perpendicular to the centreline.
9. The circular Hough transform is employed to identify bright circles in order to further analyse the vessels related to each path.
10. Hierarchical clustering is applied to the identified bright circles.By stitching the clusters together, and interpolating the missing circles, a good segmentation of the entire vessel is established.
11. An attempt is made to analyse vessels in order to separate vessels and veins by applying template matching and hierarchical clustering.

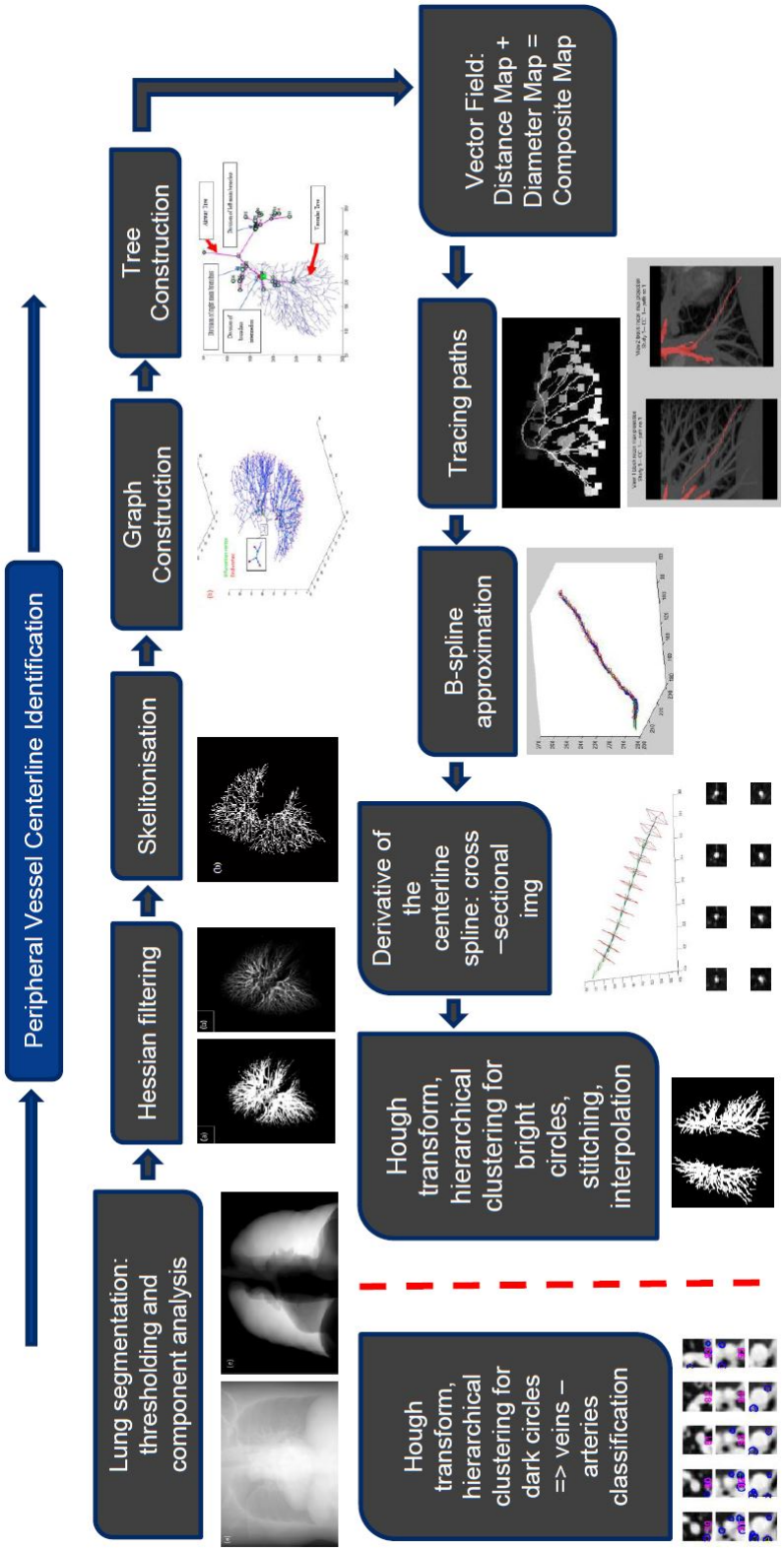

Figure 4: Pipeline of vessel identification and tracing from [1]

3.3. Peripheral bronchial tree identification

The steps below are graphically illustrated in Fig. 6.

1. Cross-sectional images are generated in the same fashion as in the previous chapter.
2. Cross-sectional images are interpolated from size 21 x 21 pixels to 161 x 161 pixels.
3. Circular Hough transform is applied to identify low intensity (dark) circles, which are potential cross-sections of bronchial lumens.
4. A novel ring filter is applied
5. Hierarchical clustering is used to further increase the specificity of the search for bronchial candidate centre points.

Overall, the method showed promising results in detecting distal airways achieving good high mean precision of 96.07% with standard deviation of 3.71%. Some examples of the final results are shown in Fig. 6.

3.4. Further work not included in the thesis

The following work has been done after the thesis was published and hence is omitted.

The developed method showed very high specificity and moderate sensitivity or, in other words, the method is conservative to some extent and good in not spotting false bronchi but meantime missing a few true positives. This enables to use of positively classified images with the bounding boxes in them as inputs to some deep learning network to try to learn bronchus representation. Pre-trained SSD network in Matlab was chosen to learn this task. Transfer learning was applied, so the weights in the deep levels remained unchanged, and only weights of the last dense layers were learned for this specific task. The learned model showed satisfactory results in spotting bronchi from 2D cross-sectional images and improved sensitivity of the previous method. New images from the test set were fed into the network, and output was assessed by a radiologist for the correctness of bronchi identification. Thus, a new method for generating new positively labelled images was developed and enabled the production of more labelled data for further improvements of the network or inputs to other methods requiring training.

Some examples of true positive and false negative images are shown in Fig. 5.

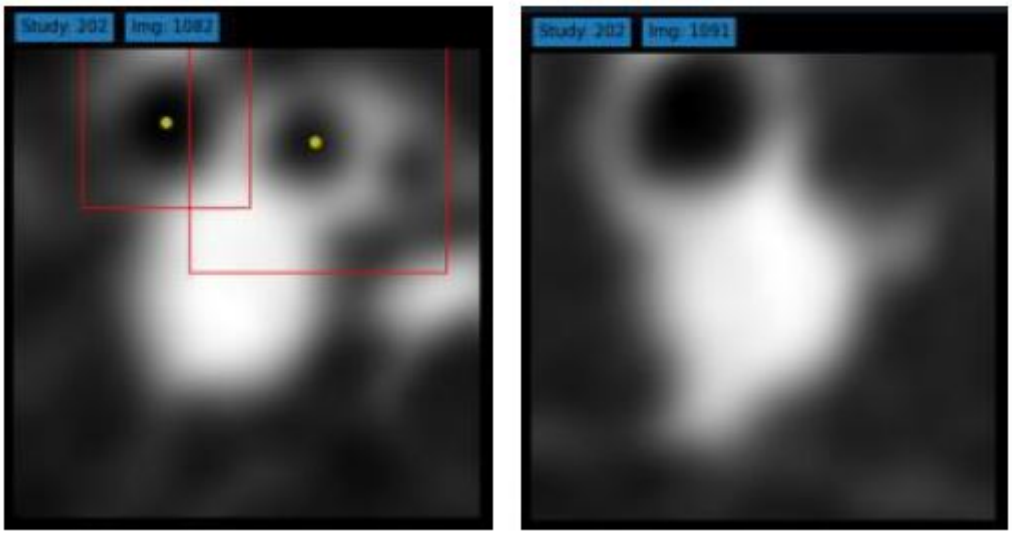

Figure 5: Examples of true positive (left) and false negative images (right) as a result of bronchi identification with SSD neural network

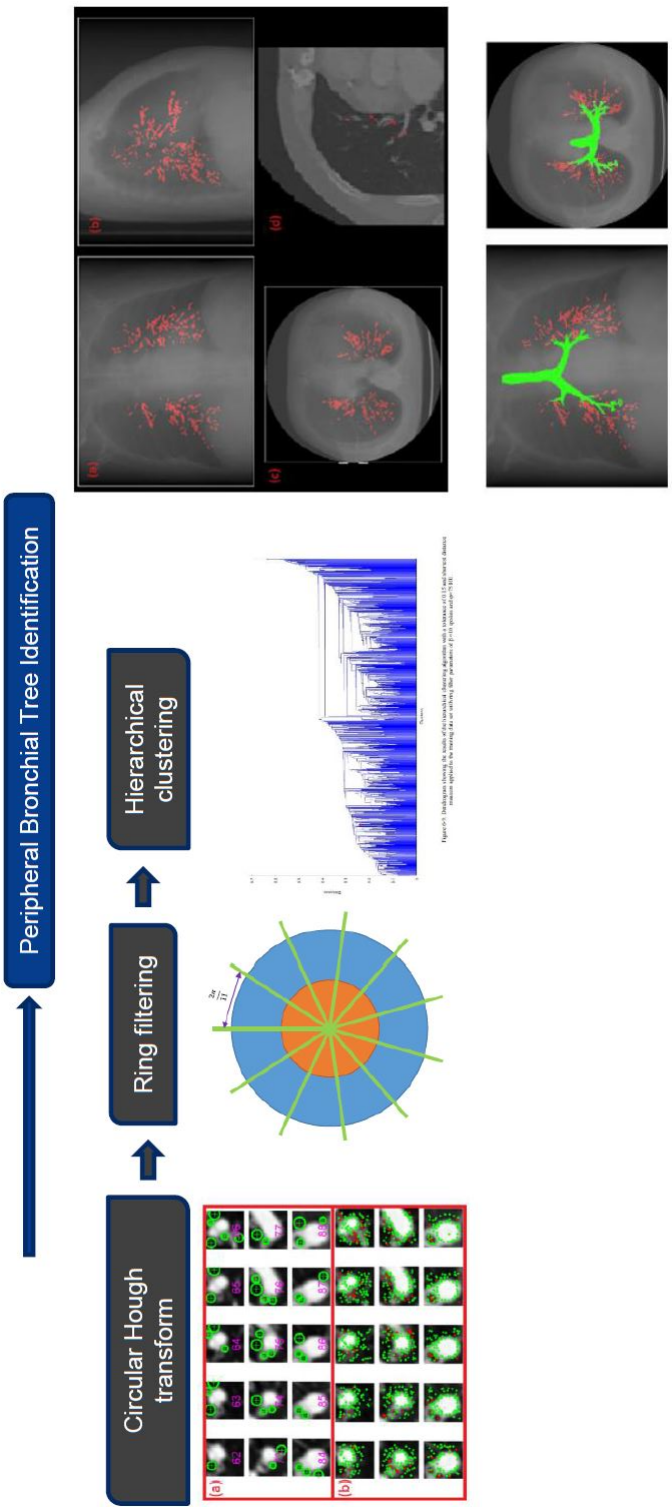

Figure 6: Pipeline of peripheral bronchial tree identification from [1]

4. Method

The overall proposed workflow is shown in Fig. 7. Sampled 2D cross-sectional images belonging to one edge are sequentially fed into the CNN introduced in the previous chapter. Standalone CNN, in the way it is used in the baseline method, treat each image independently without context, so it fails to detect continuous bronchi within an edge. We propose to accompany the current detection method with a tracking module that can potentially utilise recurrent neural networks (RNN) in order to help the CNN module to produce continuous detection in case it fails. Typically in the cross-sectional images of original size 21x21, one can observe from 1 to 4 bronchi depending on the spatial location of the samples. The challenge is to detect and track all of them within an edge.

The way the proposed method works is as follows. At each time stamp the detector, that consists of a single CNN, produces a list of detection in the form of bounding boxes in a local 2D coordinate system. The 2D local coordinates are translated into 3D global coordinates using spatial information stored for each cross-section. The association block either register new bronchi or link them to the existing traced paths based on the association algorithm. If for the existing path(s) the CNN failed to detect next location the current trajectory of bounding boxes are fed into the RNN block to estimate next point based on the history of previous positions in 3D space. The estimated location is sent back to the association block to complete the tracking step at the current time point. In such a way, continuous bronchi branches can be extracted from given vessel edges. Further, extracted bronchi branches can be connected to produce a reconstruction of full bronchial tree centre-line, but it is out of the scope of the current project.

More details on the association block will be given later in this chapter, but at this stage, it should be noted that association of bronchi detection should be done purely based on spatial information without including image features. The reason behind it is that bronchi look almost identical and indistinguishable to each other (see next chapter for image examples). This is opposed to many other tracking problems in computer vision where objects of interest have a different appearance, such as pedestrian tracking, where more sophisticated approaches can be used.

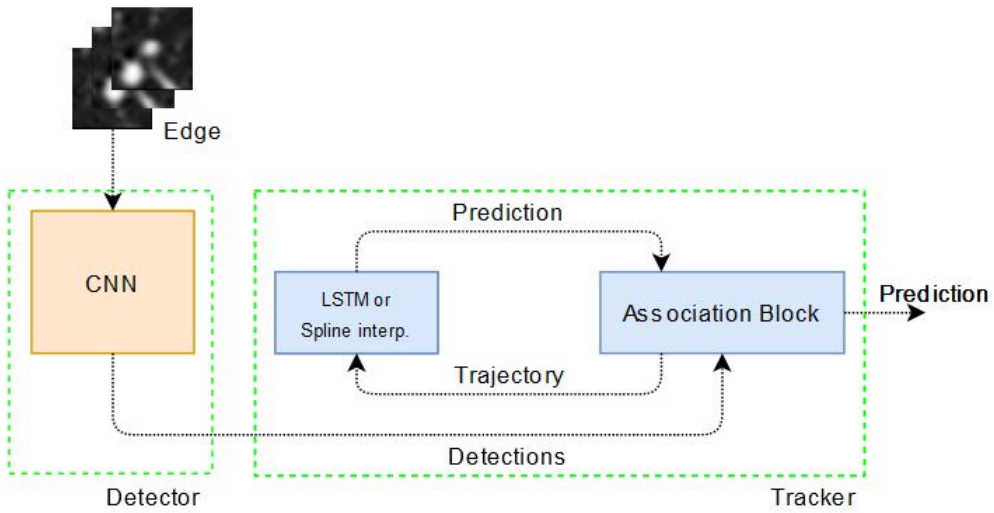

Figure 7: The proposed workflow

In this work, we will specifically explore the ability of RNN to predict the next point of trajectory in bronchi centre-line tracking. As an alternative hypothesis, we will compare RNN with 3D spline interpolation - a method that does not require any learning.

Since there is no labelled data set of continuous bronchi location withing an edge, it is necessary to create one. The next sub-section is dedicated to labelled data set creation, and the following two will discuss approaches related to RNN and 3D spline interpolation methodologies.

4.1. Annotation of the data-set

The current detection result cannot be used for training or processing sequential or temporal models due to a large number of missed detection and absence of connectivity information for the detected bronchi. Hence, the creation of a data-set of continuously annotated bronchi within a vessel edge is absolutely necessary. Moreover, as it can be seen in the next chapter, the current detection has many problems such as double detection and inconsistent bounding box size.

To create such data-set, first for each study, candidate edges have been extracted, that have detection in more than 50% of cross-sectional images. Then, for each edge, automatic detection from CNN has been loaded. To facilitate manual labelling, an interactive tool has been created. The user is presented with a cross-sectional image and loaded automatic detection. They are able to draw bounding box one at a time for each bronchi segment. The loaded automatic detection can guide the labelling process. The number of iterations for an edge is equal or more to the number of spotted bronchi by an observer. An example of the tool and labelling process can be seen in Fig. 8. For each bronchi segment, the following information is recorded: study number, connected component, edge number and a list consisted of the image number, x and y coordinates in 2D of the top left corner of a bounding box, length and width of a bounding box.

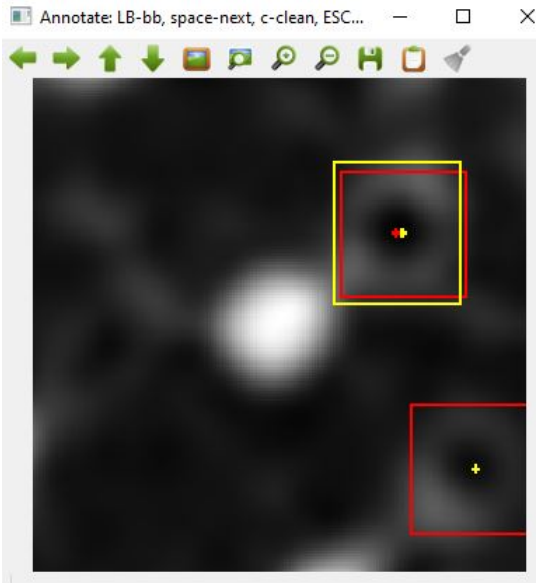

Figure 8: Screenshot of the annotation tool developed for labelling continuous bronchi segments. The red bounding boxes with yellow centre-points represent automatically detected bronchi by the CNN. The yellow bounding boxes with red centre-points represent manually annotated bronchi.

Due to the lack of labour resources and time constraints, the annotation has been done by the author and the number of annotated edges was 663. The length of the segments varies from 1 to approximately 30. The work took approximate three weeks to complete. Sample records from the output file are shown in Fig. 9.

annotations.txt

```
1  202;1;4;1937 93 25 51 53,1936 92 24 51 51,1935 91 25 47 46,1934 137 71 -48 -49,1933 136
2  202;1;18;1089 56 48 47 -59,1088 111 50 -53 -59,1087 117 52 -53 -62,1086 121 55 -56 -61,
3  202;1;18;1082 119 21 37 40,1081 130 24 34 37
```

Figure 9: Sample entries from the annotation file. Each row represents an annotated bronchi segment. From left to right: study number, connected component number, edge, number, list of image number, D x,y coordinates of the top left corner of the bounding box, length and with of the bounding box.

4.2. Recurrent neural networks for bronchi tracking

As described at the beginning of the chapter, the hypothesis is that recurrent neural networks will be able to learn a structural representation of a tree and in particular shape and direction of individual branches in order to correct mistakes made by the CNN. Typically RNN is used in a setting with existing time and context component, such as natural language processing, weather forecast prediction, stock price prediction etc. One can convert the problem of bronchi identification into such a setting by imagining that individual cross-sectional images are fed sequentially at each time point, and previous images give the context.

The most successful and popular RNN architecture at the time of writing this report is the long short term memory (LSTM) (see Fig. 10). The various modifications of LSTM and other RNN architectures, such as recurrent gated unit (GRU), might bring minor improvements in particular learning tasks, but to assess the general application of RNN for the current problem the classic standard architecture has been chosen.

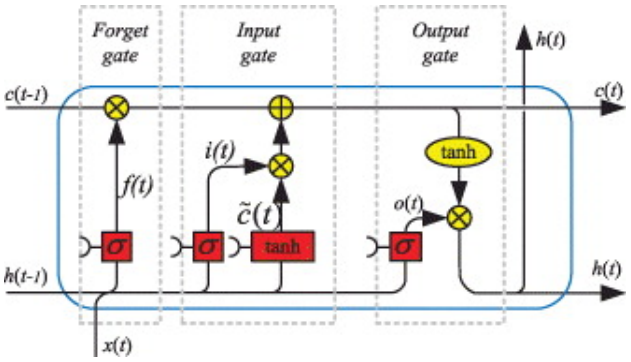

Figure 10: Standard LSTM architecture

Length of input sequences has to be determined in advance. We will parametrise our method by generating separate input sequences of lengths 3, 5, 7 and 9. First, original 2D annotations are shuffled and split into training, validation and test sets (65%, 15% and 20% respectively). Then, each set is processed separately, by converting 2D coordinates into 3D and generating sequences of lengths 3,5,7 or 9 shifted by one-time point. Original sequences of length two are ignored. Thus, from the original training set consisting of 347 bronchus branches of various length we are able to generate 2286 short sub-branches shifted by one-time point (and the respective amount for the validation and the test sets). The beginning of sub-sequences is padded with 0.

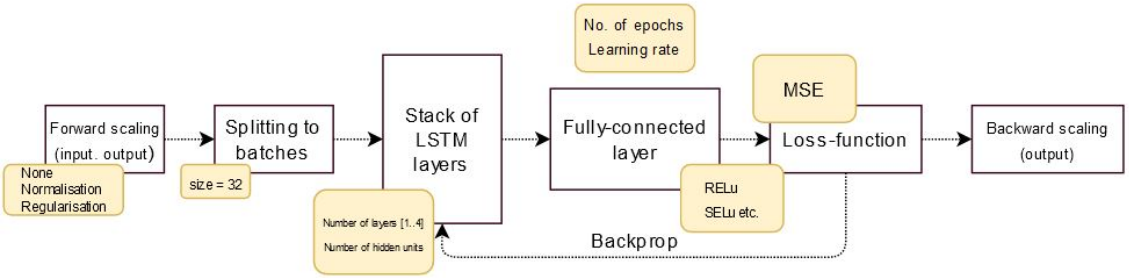

Figure 11: The learning framework and its parameters

The overall learning framework is shown in Fig. 11 along with its parameters we tried to optimise. First, the pre-processed data are read, and forward scaling of input and/or output is applied. The input and output data are three-dimensional. We tried three scaling options:

1. None. The raw global coordinates are used as inputs
2. Normalisation. Scale and translate each feature coordinate individually such that it is in the

range between 0 and 1. The scaler based on the training set is used for the validation and the test sets in order to avoid bias.

3. Regularisation. Standardise feature coordinates by removing the mean and scaling to unit variance. The scaler based on the training set is used for the validation and the test sets in order to avoid bias.

Then, scaled input are grouped into batches since it has been a good practise in the deep learning community to use batch learning as it helps to avoid getting stuck in a local minimum. We used recommended batches of size 32. After that, data is processed in the stack of LSTM layers. In common practise, it is feasible to have the number of layers between 1 and 4 since having too many layers lead to an explosion of learning parameters and significantly slows down learning. In addition, the number of hidden units is another parameter that can be explored.

A fully connected layer with an output of size 3 (since we have to predict one 3D coordinate) is followed by the stack of LSTM. Since we are doing a regression problem, one has to choose an unbounded activation function such as RELu or other variants. For the loss function, the most straightforward function is the mean squared error (MSE) which is the main choice of any regression type of problem. In the end, if we scaled to the output, we might want to scale them back to the original scale in 3D and then 2D if we want to visualise or compare the predictions.

The best choice of parameters and experimental results are outlined in the next chapter.

4.3. 3D spline approximation for bronchi tracking

An alternative approach to RNN for predicting the next centre-point would be finding an expression of the line of the best fit and then use analytical solution to find the point at the next timestamp. This can be done by using a B-spline approximation of N-D curve (3D in our case). Spline interpolation is a form of interpolation where the interpolant is a particular type of piecewise polynomial called a spline. A B-spline or basis spline is a spline function that has minimal support with respect to a given degree, smoothness, and domain partition. Any spline function of given degree can be expressed as a linear combination of B-splines of that degree. Cardinal B-splines have knots that are equidistant from each other. B-splines can be used for curve-fitting and numerical differentiation of experimental data. An example is shown in Fig. 12.

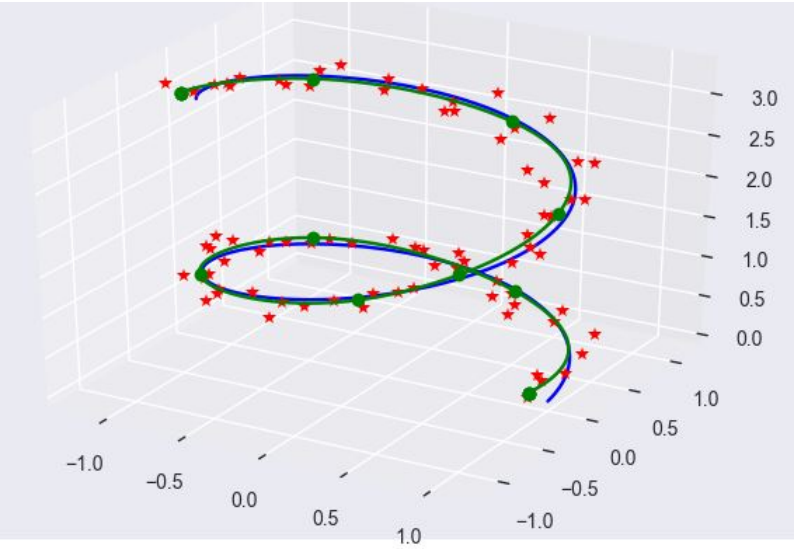

Figure 12: Example of B-spline approximation. The ground truth is shown as the blue line, sampling points as the red stars and the approximated curve as the green line with the dots representing knots.

The data pre-processing step remains the same as in the previous approach, so we generate data sets of sub-sequence lengths of 3,5,7 and 9. For each sub-sequence, we fit a 3D B-spline polynomial

of degree n and estimate its coefficients. Since we know the location of the cross-sectional image of the next time point where the point of interest lie (hence the normal vector), the point(s) of the intersection of the approximated curve and the plane can be found. This method becomes apparent after the data exploration step (discussed next) where it was noticed that in many situations 2D sampling is inconsistent, and planes intersect each other at different angles, so the plane at the next time stamp can, in reality, contain bronchi point from "the past" depending on the 3D geometry and curvature of the vessel and bronchi centre-lines. It was chosen to try to approximate curves with a line (polynomial of degree $n = 2$) since the analytical solution of line-plane intersection is simple and exists in closed form (see Fig. 13).

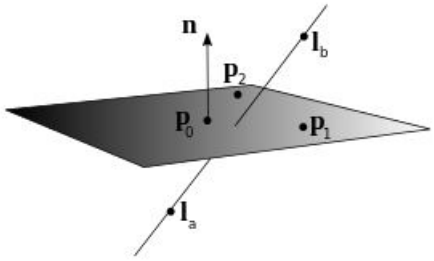

Figure 13: The intersection of line and plane.

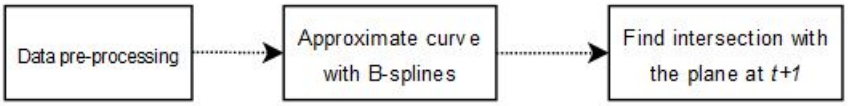

Figure 14: The B-spline approximation framework.

4.4. The association block and implementation

For matching newly detected bronchi with the trajectories stored in the previous step the Hungarian algorithm can be applied. Due to the scope of the project and time constraints only the part related to trajectory prediction is implemented as this is the main research focus. The other work is straightforward and comprises of software development activities rather than research.

4.5. Implementational aspects

All the work in this project has been done on a local computer with the following characteristics: Intel Core i7-6500U CPU @ 2.5GHz, 8 GB RAM, GPU NVIDIA GeForce 920MX 2GB.

The code was written in Python 3.7.7 in Spyder IDE v.4.1.4 with the help of the following libraries: scipy v.1.5.0, sklearn v.0.23.1, sortedcontainers v.2.2.2, numpy v.1.19.1, matplotlib v.3.2.2, cv2 v.4.3.0 and other. For the deep learning part of the project Pytorch v.1.6.0 framework was used utilising GPU resources.

5. Data Set

5.1. Ethics approval

The imaging studies have been collected at the Prince of Wales Hospital, Sydney and have been carefully anonymised to remove any personal information of the patients. The ethics approval has been received. Reference number HC200452. Protocol title is "Computerised Detection and Segmentation of Bronchovascular Structures of the Lung on Chest CT".

5.2. Data organisation

The data set comprises of 145 591 Matlab files in ".mat" format. Each file contains a table of meta information of the following parameters:

- Study Number
- Image Number within a study
- Connected component number
- Edge number within connected component
- Cross-sectional image number within edge

Another table contains image data and global 3D coordinates:

- Array of size 161 x 161 of image content values in HU
- Array of size 161 x 161 of global 3D x coordinates of image voxels
- Array of size 161 x 161 of global 3D y coordinates of image voxels
- Array of size 161 x 161 of global 3D z coordinates of image voxels

Coordinates of the bounding boxes of located bronchi are stored in two separate Matlab structures. Overall, each study can be viewed as a set of connected components which consists of edges which consist of cross-sectional images, as shown in Fig. 15. To correctly represent the hierarchical structure of the data object-oriented programming approach was chosen, and corresponding classes for connected components, edges, and cross-sectional images were created. A global class called BronchailTree was created that contains all the structural information read from the Matlab files and methods for reading, exploration, and other processing techniques described in other chapters were created. Due to the massive amount of image data and coordinates of voxels, this information is stored on a local drive, and only links to corresponding files are stored in memory. These data are processed on demand. Coordinates of detected bounding boxes in 2D are read from files and stored in corresponding instances of CrossSection class.

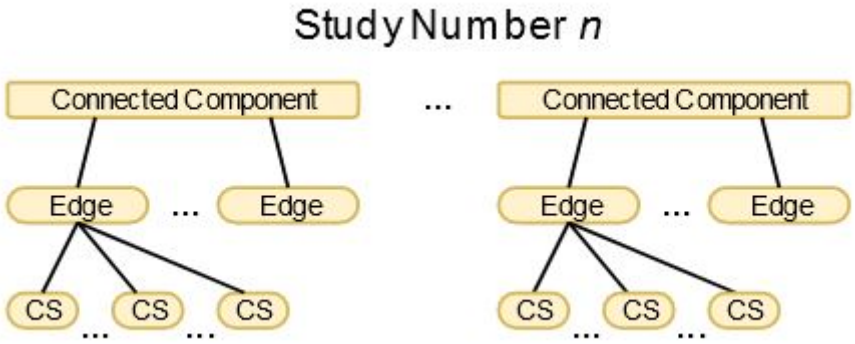

Figure 15: Structural representation of the data

5.3. Data exploration

Once the hierarchical data structure is created and data are read, one can perform quantitative analysis of data available. As shown in Table 1 there 11 studies available (number 200 to 210) with the first two studies having at least twice as many files comparing to the other nine. The number of identified vessel edges varies from 470 to 2055 per study, but should be noted that approximately half of them are veins and do not contain bronchi. In total there are 11,108 edges available. Among them, only 761 have a percentage of detected bronchi in cross-sectional images more than 50 across one single edge. This means there are quite a few missing bronchi that either the original method or the neural network failed to localise and improvement is needed.

Study No	No images	No edges	Edges>50% detection
200	26327	2055	62
201	21652	1749	57
202	7527	470	20
203	9874	738	61
204	14334	1134	134
205	11450	878	163
206	12221	889	41
207	12640	929	27
208	13267	989	72
209	7302	591	80
210	8981	686	44
Total	**145575**	**11108**	**761**

Table 1: Summary of data available for the project

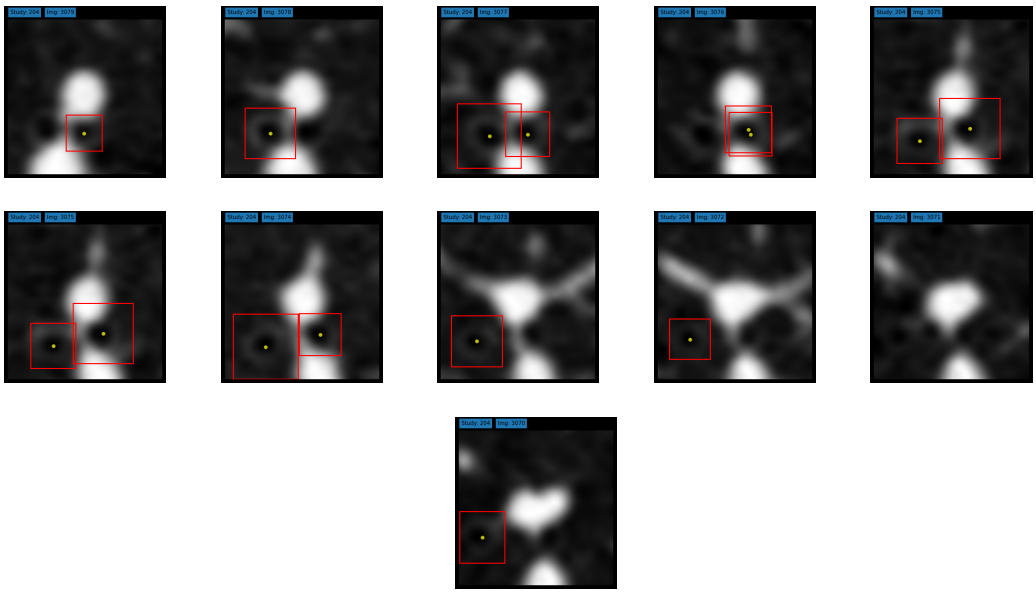

Figure 16: Example of cross-sectional images sampled along edge 150 from connected component 1, study 204. Red bounding boxes represent detection from the baseline method.

Length of individual branches was also assessed. It varies from 4 to over 50 cross-sectional images. The appearance of edges and detection was checked by looking at the sequence of images within an edge. Fig. 16 shows an example of an edge of length 11. There are two bronchi running in parallel to the vessel. One can spot a few problems with the current detection: images number 1, 2, 4, 8, 10 lack detection of one of the bronchi, image 9 lacks both, there is double detection of the same bronchus on image 4. Also, the bounding boxes are not consistent in size. The numbering of

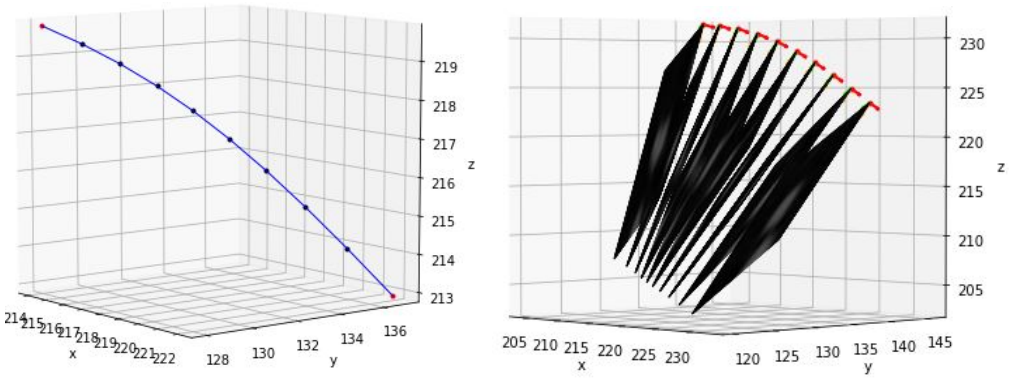

Figure 17: Examples of 2D cross-sectional patches visualised in 3D (Right) as well as their centre-lines (Left). The red arrows indicate normal vectors. Study 204, connected component 1, edge 150.

images starts from the top left corner and advances by column. 2D cross-sectional patches can be visualised in 3D space as well as their centre-points. Fig. 17 shows these images. The red arrows on the right picture position at the top left corner at original images and point towards the viewer. As it can be seen, the presented patches have consistent orientation and alignment in space and do not intersect each other.

Centre-lines of all edges in connected components and studies can be visualised in a similar fashion in 3D. The examples can be seen in Fig. 18. Note, the visualised points are centre-points of the images, hence centre-lines of the vessels.

After a thorough examination of edges of different shapes, lengths and location, it was found that no all cross-sectional images within an edge are positioned and aligned consistently.

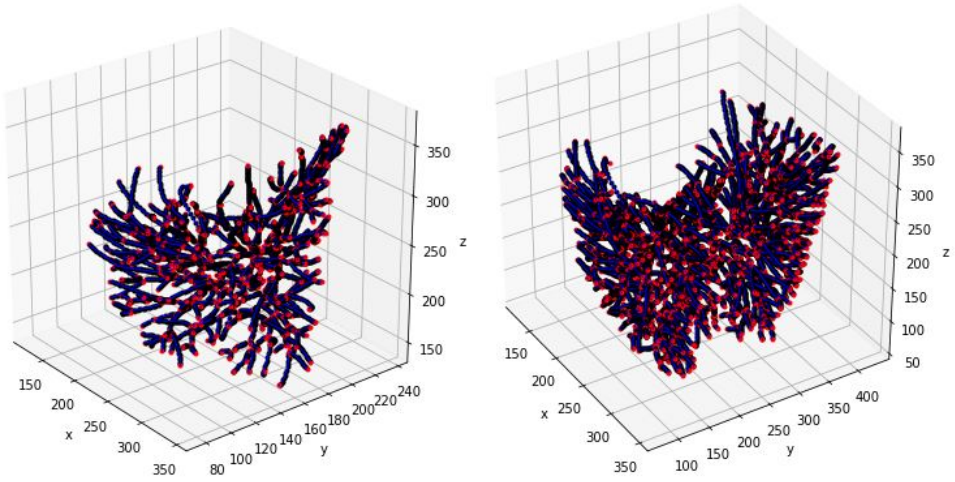

Figure 18: Centre-points of 2D patches in 3D space. Connected component 1, study 204 (Left). Study 205 (Right).

An example of inconsistent sampling is shown in Fig. 19. Note image flips from image number 1 to 2 and from 3 to 4. There is also no smooth transition from image 4 to 5 as a new vessel structure suddenly appears on image 5. Such artefacts make spacial analysis purely in 2D not possible.

3D visualisation of such sampling is depicted in Fig. 20. Attention should be paid to the red arrows representing normal vectors located at different angles of the image patches and many intersections of the planes due to visible non-linearity of the sampled vessel centre-line. Also, note a big difference between sampling points 4 and 5 comparing with the other pairs.

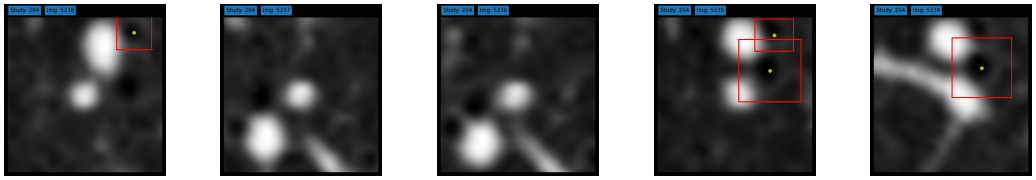

Figure 19: Example of inconsistent sampling of cross-sectional images sampled from edge 894 from connected component 1, study 204. Red bounding boxes represent detection from the baseline method.

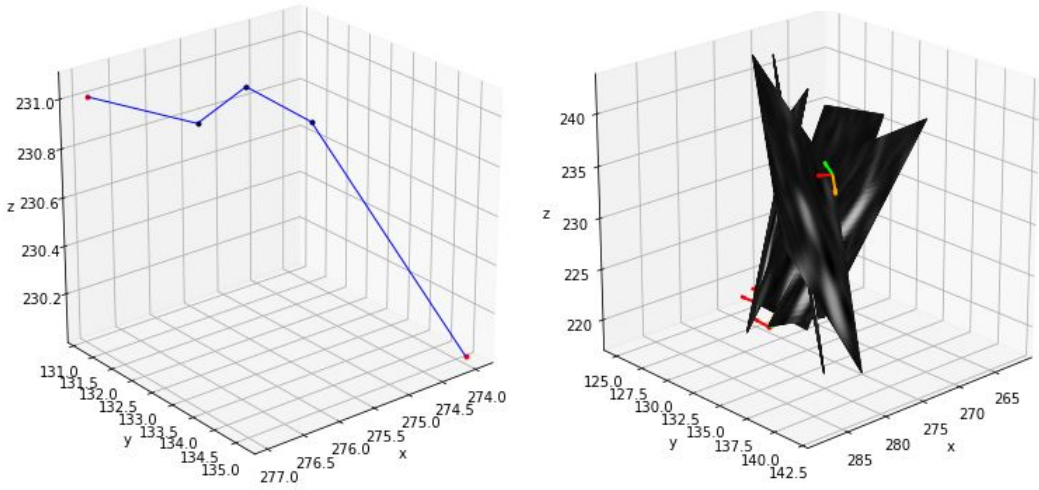

Figure 20: Examples of inconsistency in 2D cross-sectional patch sampling visualised in 3D (Right) as well as their centre-lines (Left). The red arrows indicate normal vectors. Study 204, connected component 1, edge 894.

The flips from Fig. 21 seem to have a random nature and also appear in edges whose image patches almost entirely parallel. This behaviour is due to the sampling process from the baseline method in [1]. Approximately 40% to 50% of the edges have errors as described above. The flips from Fig. 21 seem to have a random nature and also appear in edges whose image patches almost fully parallel. This behaviour is due to the sampling process from the baseline method in [1]. Approximately 40% to 50% of the edges have errors as described above.

As next step labelling of bronchi accompanying the vessel centre-lines has been done on edges. For more details, please see the previous chapter. Examples of labelled edges on previously shown consistent and inconsistent sampling can be found in Fig 22. As it can be seen on the left, bronchi centre points nicely to create a consistent trajectory, whereas on the right points 4 and 5 of the lower trajectory indeed are points between 1 and 2, so actual order of cross-sectional images should be 1,4,5,2,3 and similarly 1,2,4,3,5 for the upper trajectory. In addition, one can observe that trajectories of bronchi are not strictly parallel to vessels, and one 2D image can spot multiple bronchi with some of them belonging to the neighbouring vessels.

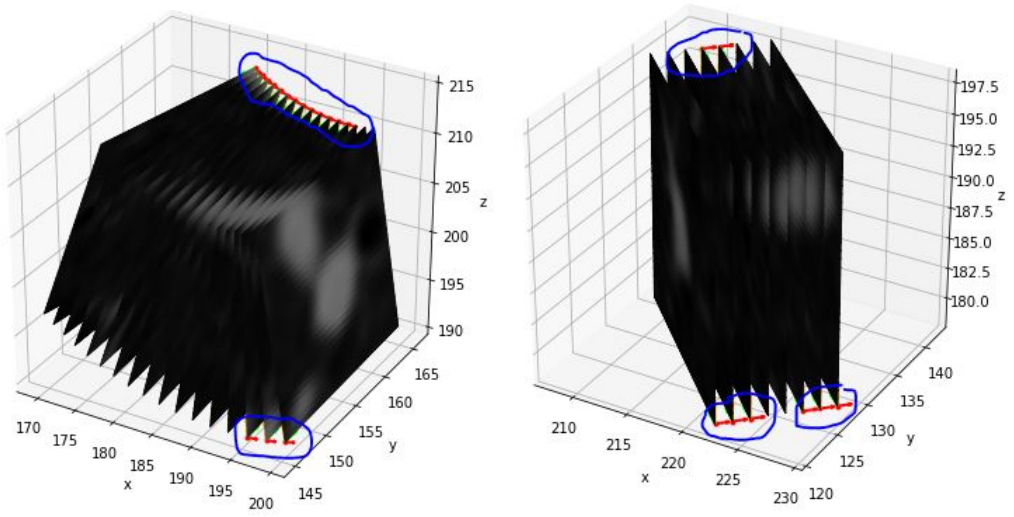

Figure 21: Examples of random flips of 2D patches where the planes are nearly parallel.

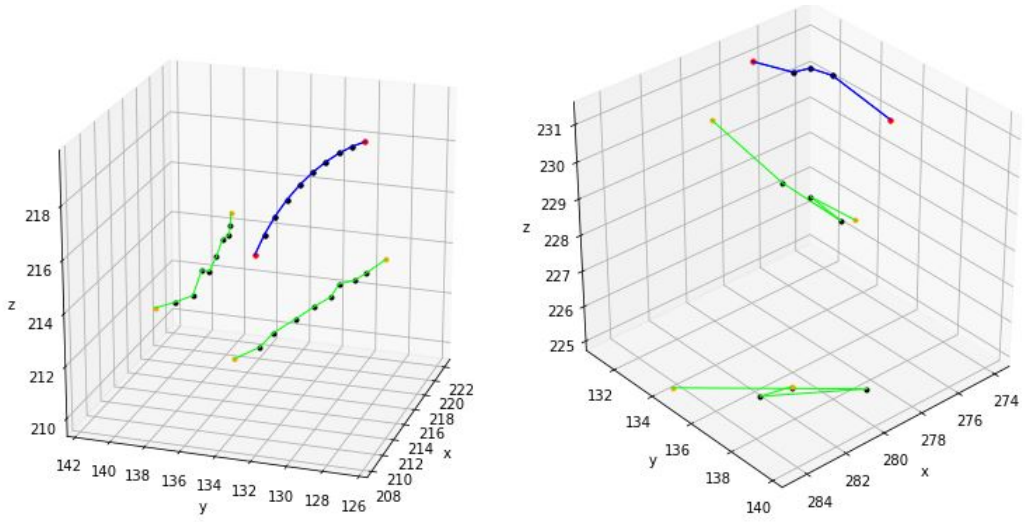

Figure 22: Examples of annotated bronchi centre-points on the edge of consistent sampling study 204, connected component 1, edge 150 (Left) and the edge of inconsistent sampling study 204, connected component 1, edge 894 (Right). Yellow dots represent end points of an edge.

6. Results

6.1. LSTM method

The raw input values are global 3D coordinates of voxels. The histogram of the x, y, z coordinate values is shown in Fig. 23. It can be seen that the values are of high magnitude that can cause instability of learning such as explosion of gradients as well as greater learning time. Nine various combinations of input and output scaling have been applied and the results are summarised in Tab. 2. It can be noted that applying Normalisation to both input and output as well as Standardisation fot input and Normalisation for output yielded the lowest MSE error and convergence happened much faster. For further research we will pick the latter.

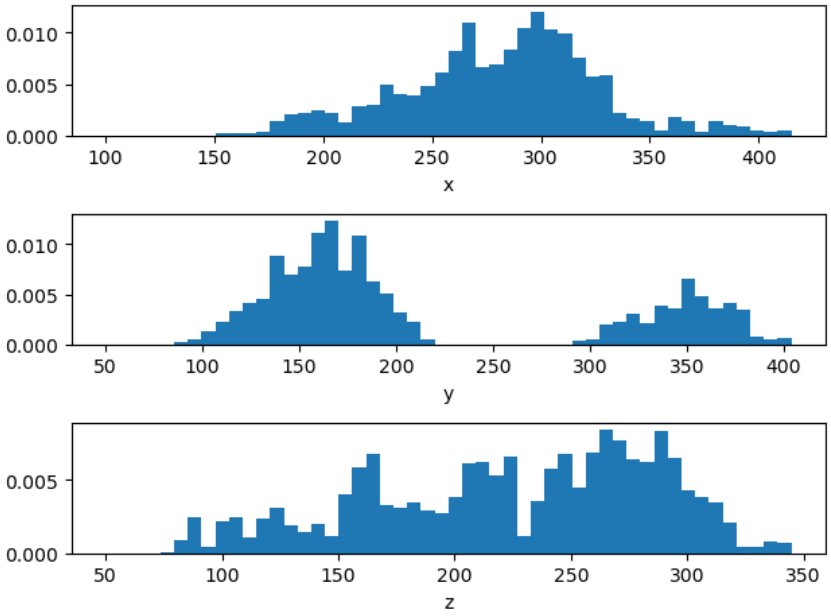

Figure 23: The histogram of input values for the x, y, z axis

Input Scaling	**Output Scaling**	**Train Loss**	**Val Loss**	**Train Loss (last 20)**	**Val Loss (last 20)**	**Epochs**
None	None	60.73	72.7	76.61	86.59	700
Norm	None	12,57	41,4	15,04	42,37	200
Std	None	3,71	14,76	4,82	17,54	200
None	Norm	47,11	87,03	59,64	81,07	200
None	Std	124,09	385,47	121,14	381,13	200
Norm	**Norm**	**2,66**	**2,69**	**3,46**	**4,32**	**200**
Std	Std	68,49	320,34	69,02	321,06	200
Norm	Std	70,08	322,61	71,33	324,65	200
Std	**Norm**	**2,27**	**3,11**	**2,92**	**3,64**	**200**

Table 2: Performance of the NN based on input and output scaling. Loss is calculated as MSE. Learning rate = 0.005. Number of LSTM layers = 3. Number of hidden layers = 128. Length of sequence = 5.

The optimal combination of parameters indicated in the method section has been found by running multiple iterations of the network. Below is the list of best parameters found.

- Number of LSTM layers = 1

- Number of hidden units in LSTM = 32
- Activation function CELU - Continuously Differentiable Exponential Linear Unit
- Sequence length = 9
- Fixed number of epochs = 200
- Fixed learning rate = 0.005

There have not been observed any significant deviation by varying these parameters. Searching for the best parameters has been performed on the validation set only. These parameters yielded the following final result on the validation and the test sets.

Train Loss	**Val Loss**	**Train Loss (last 20)**	**Val Loss (last 20)**	**Test Loss**
1.69	3.49	1.65	2.78	2.22

Table 3: The results of the final model fit

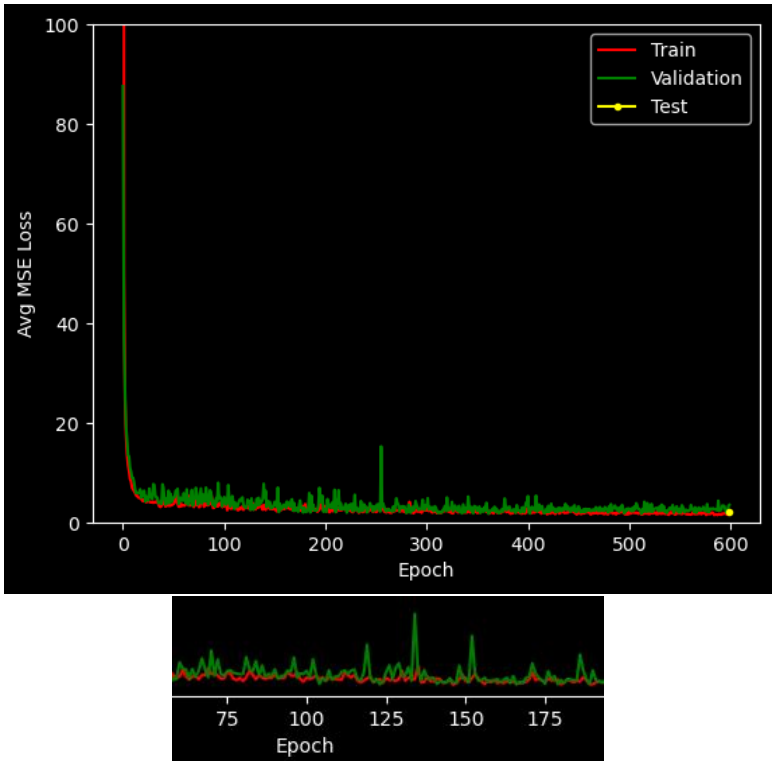

Figure 24: MSE loss of the final model fit

As it can be seen from Fig. 24 there is a slight over-fitting since the green line of the validation set is slightly above the red line representing the training set. The result of the final fit on the test set turned out to be marginally better than on the validation set.

6.2. 3D B-spline approximation method

The 3D B-spline approximation method does not require any learning, and hence all the annotated data can be used for testing. The data was split into 3617 sub-sequences of 3D coordinates and the method was applied on sub-sequences of lengths 3,5,7 and 9. The MSE losses are shown in Tab. 4. Better results on shorter sequence length can be explained by local linearity of bronchi centre-line, and vice versa poor results in longer sequences by non-linearity of long bronchi segments.

Sequence Length	**MSE Loss**
3	0.81
5	0.91
7	1.69
9	2.98

Table 4: MSE of the B-spline approximation method depending on the sequence length

The predicted coordinates in 3D were converted back into 2D local coordinates of the following cross-sectional plane for visualisation. Some examples can be seen in Fig. 25. Out of 3617 predicted points MSE of 39 points or 1% exceeded 2.2 which is the MSE of the deep learning method.

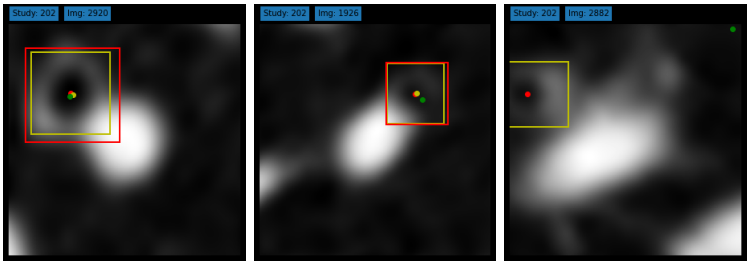

Figure 25: Examples of predicted bronchi centre-points based on the B-spline approximation method. The red bounding box and the yellow dots represent automatic detection from the CNN. The yellow bounding box and the red points represent annotations. The green points represent predictions made. Left: MSE = 0.0402. Middle: MSE = 0.2223. Right: MSE = 116.89

One can observe that points with high MSE loss belong to the segments significantly deviated from from the linear shape or the segments where inconsistent sampling was performed. Examples of such shapes are depicted in Fig. 26.

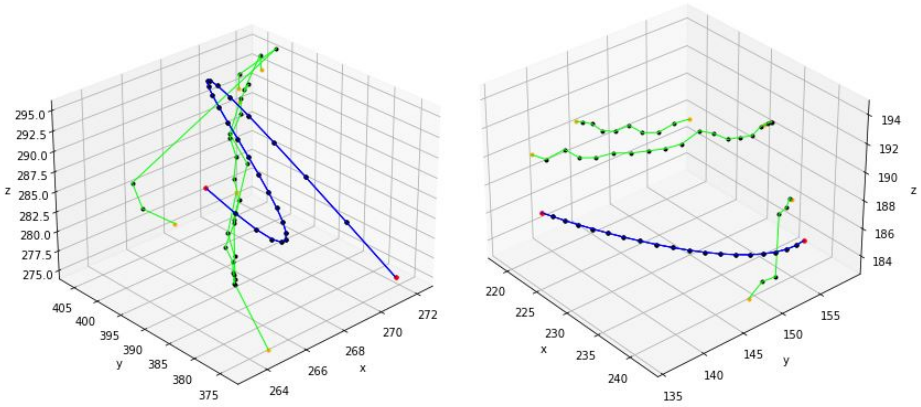

Figure 26: Examples of edges of nonlinear shape where B-spline approximation with linear functions yielded high MSE. The blue lines represent vessel centre-lines and the green lines represent annotated bronchi centre-lines.

7. Discussion

The result of the 3D B-spline approximation method exceeded the deep learning method by a large margin. In this section, we will discuss and compare the outcomes and give recommendation for improvements and future work.

7.1. Deep learning model

The deep learning method for predicting next centre-point of bronchi based on LSTM architecture requires a substantial amount of annotated data. In properly organised research studies, typically a team of trained radiologists is hired to annotate images. For instance, in [3] at least five trained specialists were involved. In this project, due to various organisational limitations, the author, who is not a radiologist, had to perform this task and due to time limitation, an only small amount of bronchial branches has been labelled. This approach may cause a lower quality of manual annotations and bias. Having a large amount of data for training is also imperative because of the sparsity of the input data. Ideally, many examples of bronchial branches from different spacial locations are needed.

Another issue with this method is the presence of inconsistent samples, as described in the data organisation chapter. Since the sampled cross-sectional patches may intersect each other at different angles, bronchi spotted at time-stamp t may actually lie behind the bronchi spotted at time $t+1$. Such a situation forces the network to learn trajectories of the wrong shape. An extreme situation is shown in Fig. 27 where the vessel trajectory is nearly linear, and the bronchi trajectory are almost U-shaped and first and last detection is nearly indicating the same spot.

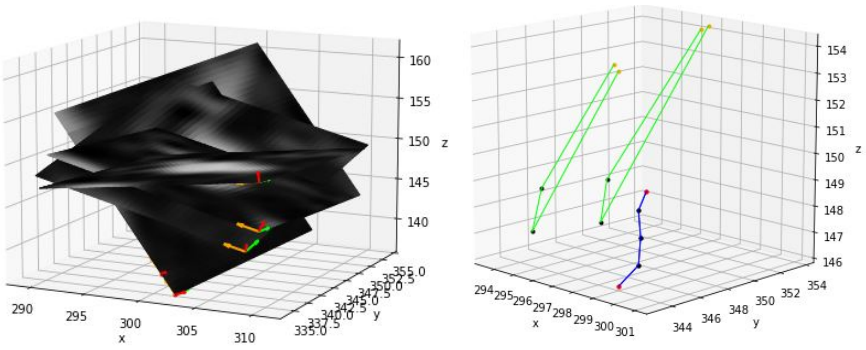

Figure 27: Examples of inconsistent sampling where the vessel centre-line is nearly linear and the bronchi centre-lines are U-shaped.

An issue, common for both methods, is the presence of residual bronchi captured in 2D cross-sectional patches, that do not run parallel to the tracked vessel centre-line. An example of such segment is shown in Fig. 26 (Right) at the bottom right corner. The additional algorithm has to be implemented to identify and keep only pairing bronchi to the tracked vessel.

However, there could be great potential benefits of learning geometry of bronchial branches with RNN if one can accurately sample enough consistent data for training. One of them is that RNN is able to learn very complex functions that can not be expressed analytically. For example, it could capture a different level of non-linearity at various bronchi generations. It can be hypothesised that smaller bronchi are more non-linear than larger ones. Also, it could automatically learn the direction of branches based on their spatial locations. If centre-lines are extended to bounding boxes, then it could learn narrowing parameter as well. Such model would allow not only more accurate estimate for tracking, but also use of the model for synthetic data generation.

To facilitate the creation of better RNN models, the sampling mechanism of cross-sectional images from the baseline method has to improved to create consistent samples. Alternatively, the inconsistent samples have to be removed manually or automatically, but that might be a grand challenge to do.

7.2. B-spline approximation model

The spline method showed robust and accurate results where the geometry of branches is approximately linear since we used linear splines for simplicity. To overcome the issue with inconsistent sampling, we find an intersection of the fitted spline with the plane of the next cross-sectional image.

The method can be improved by using quadratic or cubical splines; however, finding multiple intersections with a plane in 3D might not be a simple task to do.

The downside of the method is the pre-defined assumption of curves' shapes, i.e. linear, quadratic etc. without the ability to generalise.

7.3. The baseline method and general approach

Since this project of exploring RNN capabilities is based on the baseline method described in [1] and in the overview chapter, some critique and recommendation can be given since, as it has been observed, certain limitations were incurred from it.

First of all, the sampling process of 2D cross-sectional patches seems to be not robust and to some extent, uncontrolled as it was described in the data exploration chapter. The results do not suit further automatic process in a sequential fashion. The experiment with fitting the LSTM network showed that we force it to learn trajectories of the wrong shapes.

Secondly, there is great variability and inconsistency in sizes of bounding boxes produced by the CNN. It may confuse the neural network what it has to learn. Ideally, manual re-labelling is needed to have consistent data.

The process of generating new labelled data based on the outputs of the trained standalone CNN involves the only one trained person and observation of images without context. It would be beneficial to look at each image and its neighbours from the same edge as well have a better-designed protocol as proposed in [3].

Since the bronchial tree is a hierarchical structure, it would be a logical approach to process it in a hierarchical manner as opposed to a random draw of individual cross-sections or branches as it was done in [1] and in this project. This way, we could benefit from the inclusion of contextual and structural information. Algorithms based on convolutional LSTM or graph neural networks would be a good choice then.

Another area of improvement could be sampling 3D cubes as opposed to 2D cross-sections. This would be a natural way of solving the problem since the original data is three-dimensional and more information could be preserved. Alternatively, sampling 3D spheres in polar coordinates could be an interesting approach to explore.

Last but not least the CNN currently is not trained to identify bifurcation points. The challenge is that cross-sectional images of bronchi at bifurcation look different from a ring-shaped singular bronchus.

To sum up, it seems manual intervention and data correction is unavoidable if we want to tackle the problem in a hierarchical way. Researches must have confidence in the data they use for learning algorithms, but with the available data, it was not possible at this stage that resulted in poorer performance of the RNN method.

8. Conclusions

The task of bronchial tree identification and extraction from CT scans is a complex research problem that is still unsolved at present. The research community has been attempting to solve it for the last two decades by applying techniques from traditional computer vision to cutting edge deep-learning architectures in 3D. One of the main challenges is a lack of annotated data and standardised protocols that would allow comprehensive evaluation and comparison of the proposed methods.

This work is based on [1] and uses generated data set as input for the framework presented. In the core of the framework, there is a tracking module and a block, whose role is to predict next centre-point (or potentially bounding box) in case the CNN fails. To perform this task, two methods have been developed and compared: one is RNN-style network based on LSTM and the second one is B-spline approximation. The experimental results showed that the latter is preferred at this stage since it has a lower MSE error of 0.81 compared to 2.22 of the deep learning model.

Due to the hierarchical nature of an airway tree, it would be beneficial to perform identification and extraction in a sequential manner to include prior information in the decision process as well as substitute 2D to 3D. Methods based on convolutional LSTM and graph neural network seem to be promising in this regard, and this can be further research direction.

References

[1] D. Moses, 'Computerised analysis of the bronchovascular anatomy of the lung on multidetector ct studies,' Ph.D. dissertation, Computer Science & Engineering, Faculty of Engineering, UNSW, University of New South Wales, Sydney, NSW 2052, Australia, Nov. 2018.

[2] A. Krizhevsky, I. Sutskever and G. Hinton E., 'Imagenet classification with deep convolutional neural networks,' in *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol. 1, Curran Associates Inc., 57 Morehouse Lane, Red Hook, NY, United States, Oct. 2012, pp. 1097–1105.

[3] P. Lo, B. van Ginneken, J. Reinhardt M. and et al., 'Extraction of airways from ct (exact'09),' in *IEEE Transactions on Medical Imaging, issue 11*, vol. 31, IEEE, Jul. 2012, pp. 2093–2107.

[4] E. van Rikxoort, W. Baggerman and B. Ginneken, 'Automatic segmentation of the airway tree from thoracic ct scans using a multi-threshold approach,' *Proc. of Second International Workshop on Pulmonary Image Analysis*, pp. 341–349, Jan. 2009.

[5] P. Lo, J. Sporring and M. de Bruijne, 'Multiscale vessel-guided airway tree segmentation,' *Proc. 2nd Int. Workshop Pulmonary Image Anal*, pp. 323–332, Jan. 2009.

[6] J.-P. Charbonnier, E. van Rikxoort, A. Setio, C. Schaefer-Prokop, B. Ginneken and F. Ciompi, 'Improving airway segmentation in computed tomography using leak detection with convolutional networks,' *Medical Image Analysis*, vol. 36, pp. 52–60, Nov. 2016. DOI: 10.1016/j.media.2016.11.001.

[7] O. Ronneberger, P. Fischer and T. Brox, 'U-net: Convolutional networks for biomedical image segmentation,' *LNCS*, vol. 9351, pp. 234–241, Oct. 2015. DOI: 10.1007/978-3-319-24574-4_28.

[8] D. Jin, Z. Xu, A. Harrison, K. George and D. Mollura, '3d convolutional neural networks with graph refinement for airway segmentation using incomplete data labels,' *International Workshop on Machine Learning in Medical Imaging*, vol. 10541, pp. 141–149, Sep. 2017. DOI: 10.1007/978-3-319-67389-9_17.

[9] Q. Meng, H. Roth, T. Kitasaka, M. Oda, J. Ueno and K. Mori, 'Tracking and segmentation of the airways in chest ct using a fully convolutional network,' *International Workshop on Machine Learning in Medical Imaging*, vol. 10434, pp. 198–207, Sep. 2017. DOI: 10.1007/978-3-319-66185-8_23.

[10] A. Garcia-Uceda Juarez, H. Tiddens and M. de Bruijne, 'Automatic airway segmentation in chest ct using convolutional neural networks,' *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pp. 238–250, Aug. 2018. DOI: 10.1007/978-3-030-00946-5_24.

[11] M. Zhao and G. Hamarneh, 'Treenet: Multi-loss deep learning network to predict branch direction for extracting 3d anatomical trees: 4th international workshop, dlmia 2018, and 8th international workshop, ml-cds 2018, held in conjunction with miccai 2018, granada, spain, september 20, 2018, proceedings,' in. Sep. 2018, pp. 47–55, ISBN: 978-3-030-00888-8. DOI: 10.1007/978-3-030-00889-5_6.

[12] J. Yun, J. Park, D. Yu, J. Yi, M. Lee, H. Park, J.-G. Lee, J. B. Seo and N. Kim, 'Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net,' *Medical Image Analysis*, vol. 51, pp. 13–20, Oct. 2018. DOI: 10.1016/j.media.2018.10.006.

[13] T. Zhao, Z. Yin, J. Wang, D. Gao, Y. Chen and Y. Mao, 'Bronchus segmentation and classification by neural networks and linear programming,' *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pp. 230–239, Oct. 2019. DOI: 10.1007/978-3-030-32226-7_26.

[14] Y. Qin, M. Chen, H. Zheng, Y. Gu, M. Shen, J. Yang, X. Huang, Y. Zhu and G.-Z. Yang, 'Airwaynet: A voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks,' *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pp. 212–220, Oct. 2019. DOI: 10.1007/978-3-030-32226-7_24.

[15] R. Selvan, T. Kipf, M. Welling, A. Juarez, J. Pedersen, J. Petersen and M. de Bruijne, 'Graph refinement based airway extraction using mean-field networks and graph neural networks,' *Medical Image Analysis*, vol. 64, p. 101 751, Jun. 2020. DOI: 10.1016/j.media.2020.101751.

[16] A. Garcia-Uceda Juarez, R. Selvan, Z. Saghir and M. de Bruijne, 'A joint 3d unet-graph neural network-based method for airway segmentation from chest cts,' *International Workshop on Machine Learning in Medical Imaging*, pp. 583–591, Oct. 2019. DOI: 10.1007/978-3-030-32692-0_67.

[17] M. Zhao and G. Hamarneh, 'Tree-lstm: Using lstm to encode memory in anatomical tree prediction from 3d images,' in. Oct. 2019, pp. 637–645, ISBN: 978-3-030-32691-3. DOI: 10.1007/978-3-030-32692-0_73.

[18] B. Kong, X. Wang, J. Bai, Y. Lu, F. Gao, K. Cao, Q. Song, S. Zhang, S. Lyu and Y. Yin, 'Attention-driven tree-structured convolutional lstm for high dimensional data understanding,' *arXiv preprint arXiv:1902.10053*, Jan. 2019.

[19] B. Kong, X. Wang, J. Bai, Y. Lu, F. Gao, K. Cao, J. Xia, Q. Song and Y. Yin, 'Learning tree-structured representation for 3d coronary artery segmentation,' *Computerized Medical Imaging and Graphics*, vol. 80, p. 101 688, Dec. 2019. DOI: 10.1016/j.compmedimag.2019.101688.