Universitat Autònoma de Barcelona

# An Endoscopic Navigation System for Lung Cancer Biopsy

**Antonio Esteban Lansaque**

Universitat Autònoma
de Barcelona

# An Endoscopic Navigation System for Lung Cancer Biopsy

A dissertation submitted by **Antonio Esteban Lansaque** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor of Philosophy**.

Bellaterra, July 24, 2019

| | |
|---|---|
| Co-Director | **Dr. Debora Gil**<br>Dept. Ciències de la computació & Centre de Visió per Computador<br>Universitat Autònoma de Barcelona,UAB |
| Co-Director | **Dr. Carles Sanchéz**<br>Dept. Ciències de la computació & Centre de Visió per Computador<br>Universitat Autònoma de Barcelona,UAB |
| Thesis committee | **Dr. Mario Ceresa**<br>Department of Information and Communications Technologies<br>Universitat Pompeu Fabra, Upf<br><br>**Dr. Katerine Diaz**<br>Dept. Ciències de la computació & Centre de Visió per Computador<br>Universitat Autònoma de Barcelona,UAB<br><br>**Dr. Thomas Langø**<br>Health Research<br>SINTEF |

This document was typeset by the author using LaTeX $2_\varepsilon$.

The research described in this book was carried out at the Centre de Visió per Computador, Universitat Autònoma de Barcelona. 

# Acknowledgements

My stage as a PhD has felt like a roller coaster. We all know the PhD gets you to your best and worst days. Even so to write this makes me feel like I am in one of the best days. There are tons of people that I should thank so even if I forget to mention you, please, don't think it is because I don't appreciate it but because of the sake of the length of this thesis.

I want to start with the academic part of this acknowledgments. Thanks Debora, for all the understanding and patience. You gave me freedom to prioritize other aspects in life when I needed it the most. I also want to thank you for all the help you provide. I can tell not every supervisor does the same. Besides, you also slightly pushed me to continue when I felt I could not and here I am. Even though you joined later, Carles, I am also very thankful to you. You also provide help in tasks that not everyone would help with. You two and Agnes have guided me and given me ideas which allowed me to complete this thesis.

I spent three years in the CVC so I cannot name everyone thankful for in this paragraph. Suport staff: Ana Maria, Claire, Raquel, Montse, Alexandra, Marc, Silvia,... And my dear slaves (PhD students): Marc, Laura, Xim, David,... A special thanks to Marc for dealing with me at home everyday.

I have been living three years in Barcelona, more than enough time to meet a lot of people. But actually, I didn't need them. I had Amaia, Ricard and Dani who make me enjoy weekends and weekdays since we met in the master in computer vision. Ricard helped me for almost three years here in the CVC and without Amaia these whole journey would have been pointless.

Of course, my scholarship ended with no time to write this thesis. Hence, I moved abroad and tried to combine my new job and this thesis. A very challenging task that without the support (and pushing) of Elisa and Sara I would have quit for sure. But every person I met abroad help me to keep going. Besides Sara and Elisa, I also want to thank to Brais, Alfonso, Cayetano and dann. Downs were better with you all.

I am very volatile sometimes so I decided to come back home, after more than 4 years to keep working on my PhD was a tough tasks. The last phase of my Phd... Thank you to all my friend in my home town. You also helped me to go on. A huge thanks for the support of all my family, but my dad, mum and sister in particular who made me continue in these last months. I also made it because of them, I am sure they will be happier than I am.

To conclude, for all those I mentioned and for the ones I could not, THANK YOU!

# Abstract

Lung cancer is one of the most diagnosed cancers among men and women. Actually, lung cancer accounts for 13% of the total cases with a 5-year global survival rate in patients. Although Early detection increases survival rate from 38% to 67%, accurate diagnosis remains a challenge. Pathological confirmation requires extracting a sample of the lesion tissue for its biopsy. The preferred procedure for tissue biopsy is called bronchoscopy. A bronchoscopy is an endoscopic technique for the internal exploration of airways which facilitates the performance of minimal invasive interventions with low risk for the patient.

Recent advances in bronchoscopic devices have increased their use for minimal invasive diagnostic and intervention procedures, like lung cancer biopsy sampling. Despite the improvement in bronchoscopic device quality, there is a lack of intelligent computational systems for supporting in-vivo clinical decision during examinations. Existing technologies fail to accurately reach the lesion due to several aspects at intervention off-line planning and poor intra-operative guidance at exploration time. Existing guiding systems radiate patients and clinical staff, might be expensive and achieve a suboptimlal 70% of yield boost.

Diagnostic yield could be improved reducing radiation and costs by developing intevention support systems able to guide the bronchoscope to the lesion. The goal of this thesis is to develop an image-based navigation system for intra-operative guidance of bronchoscopists to a target lesion across a path previously planned on a CT-scan.

We propose a 3D navigation system which uses the anatomy of video bronchoscopy frames to locate the bronchoscope within the airways. Once the bronchoscope is located, our navigation system is able to indicate the bifurcation which needs to be followed to reach the lesion. In order to facilitate an off-line validation as realistic as possible, we also present two methods for augmenting simulated virtual bronchoscopies with the appearance of intra-operative videos. Experiments performed on augmented and intra-operative videos, proves that our algorithm can be speeded up for an on-line implementation in the operating room.

**Key words:** *lung cancer, bronchoscopy, navigation system,...*

# Resumen

El cáncer de pulmón es uno de los canceres más diagnosticados entre hombres y mujeres. De hecho, el 13 % del total de casos con una tasa de supervivencia de 5 años son de cáncer de pulmón. Aunque la detección precoz incrementa la tasa de supervivencia del 38 % al 67 %, un diagnostico acertado aun es un reto. Para una confirmación patológica se requiere de la extracción de una muestra de tejido para su biopsia. El procedimiento preferido para la biopsias de tejido pulmonar se llama broncoscopia. Una broncoscopia es una técnica endoscópica para la exploración interna de las vías respiratorias que facilita la realización de intervenciones de bajo riesgo para el paciente.

Avances recientes en broncoscopios han incrementado su uso en diagnósticos minimamente invasivos e intervenciones como en la extracción de muestras de tejido. A pesar de estos avances, hy una falta de sistemas inteligentes que ayuden decisiones clínicas in-vivo durante los procedimientos. Las tecnologías actuales fallan en alcanzar ciertas lesiones debido a algunos aspectos en la planificación off-line y a un mal guiado durante los procedimientos. Algunos sistemas de guiado ya existentes radian a pacientes y el equipo clínico, pueden ser muy caros y solo consiguen un incremento de la eficiencia de hasta el 70 %.

La eficiencia del diagnostico podría mejorarse reduciendo la radiación y costes de la intervención desarrollando sistemas de soporte capaces de guiar el broncoscopio hasta una determinada lesión. El objetivo de esta tesis es el de desarrollar un sistema de navegación para broncoscopias que permita alcanzar una lesión siguiendo un camino previamente calculado en un CT.

Proponemos un sistema de navegación que usa información anatómica de imágenes de broncoscopias para localizar el broncoscopio dentro de las vías respiratorias. Una vez que el broncoscopio es localizado, nuestro sistema de navegación es capaz de indicar la bifurcación que se necesita atravesar para alcanzar una determinada lesión. Para facilitar una validación off-line tan realista como sea posible, también presentamos dos métodos para aumentar imágenes de broncosopias virtuales con apariencia real. Los experimentos realizados tanto en imágenes aumentadas como en broncoscopias reales demuestran que nuestro algoritmo puede acelerarse para ser usado dentro de un quirófano.

**Palabras clave:** *cáncer de pulmón, broncoscopia, sistema de navegación,...*

# Resum

El càncer de pulmó és un dels càncers més diagnosticats entre homes i dones. En realitat, el càncer de pulmó representa el 13% del total de casos amb una supervivència global de 5 anys. Tot i que la detecció precoç augmenta la supervivència del 38% al 67%, el diagnòstic precís continua sent un repte. La confirmació patològica requereix extreure una mostra del teixit de la lesió per a la seva biòpsia. El procediment preferit per a la biòpsia del teixit extret s'anomena broncoscòpia. Una broncoscòpia és una tècnica endoscòpica per a l'exploració interna de les vies respiratòries que facilita el funcionament d'intervencions mínimament invasives amb un baix risc per al pacient.

Els avenços recents en aparells broncoscòpics han incrementat el seu ús per procediments diagnòstics i d'intervenció, com ara, el mostreig de biòpsia per càncer de pulmó. Malgrat la millora de la qualitat del dispositiu broncoscòpic, hi ha una manca de sistemes computacionals intel·ligents per donar suport a la decisió clínica en el moment de la intervenció. Les tecnologies existents no aconsegueixen assolir la lesió amb precisió a causa de diversos aspectes en la planificació de la intervenció i un mal guiatge en el moment d'exploració. Els sistemes de guia existents irradien els pacients i el personal clínic, podrien ser cars i aconsegueixen un 70

Es podria millorar el rendiment del diagnòstic reduint la radiació i els costos mitjançant el desenvolupament de sistemes de suport a la intevenció capaços de guiar el broncoscopi a la lesió. L'objectiu d'aquesta tesi és desenvolupar un sistema de navegació basat en imatges per a l'orientació intraoperatòria dels broncoscopistes a una lesió objectiu a través d'un recorregut prèviament planificat en un escànner CT.

Proposem un sistema de navegació en 3D que utilitza l'anatomia dels frames en els vídeos de broncoscòpia per localitzar el broncoscopi dins de les vies respiratòries. Un cop situat el broncoscopi, el nostre sistema de navegació pot indicar la bifurcació que cal seguir per arribar a la lesió. Per tal de facilitar una validació el més realista possible, també es presenten dos mètodes per augmentar les broncoscopies virtuals simulades amb l'aparença de vídeos intraoperatoris. Els experiments realitzats en vídeos augmentats i intraoperatoris demostren que el nostre algoritme es pot accelerar per a una implementació en temps real a la sala d'operacions.

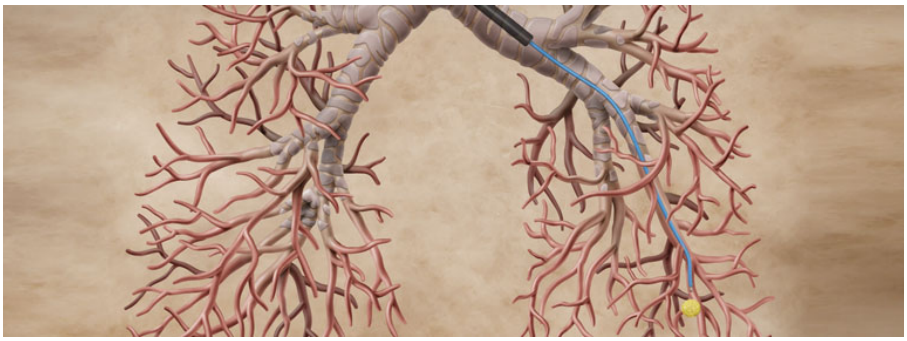
**Paraules clau:** *càncer de pulmó, broncoscòpia, sistema de navegació,...*

# Contents

# List of Figures

# List of Tables

# 1 Introduction



Lung cancer is one of the most diagnosed cancers among men and women. Actually, lung cancer accounts for 13% of the total cases with a 5-year global survival rate in patients. However if lung cancer is detected in early stages of the disease the survival rate over 5-years increases from 38% to 67% [23]. This implies that early-stage detection is critical to increse the survival rate. Despite its importance, early-stage detection is still a challenge in many countries [49] by either financial or technical problems. Currently, cancer diagnosis can only be achieved by analysis of tissue sampled on the lesion. Accurate sampling is a challenge in the case of lung cancer due to the difficulty to reach the lesions (called nodules) for their biopsy.

## 1.1 Procedures for Lung Cancer Biopsy

Computed tomography (CT) screening programs may significantly reduce the risk of lung cancer death. Screening programs are based on the detection of small pulmonary lesions with low dose chest computed tomography (CT) and its pathological confirmation. Pathological confirmation requires extracting a sample of the lesion tissue for its biopsy. Such sample can only be obtain by either a transthoracic needle aspiration (TTNA) [18, 26] or an endoscopic examination [3].

TTNA is a diagnostic procedure that is done to determine the cause of an abnormality on the lung. After injecting a local anesthetic, a long needle is inserted into the chest wall, between the ribs, to take a sample of lung tissue for a biopsy. Ultrasound or computed tomography (CT) are used to guide the needle directly to the area of the lung where the sample should be taken, as illustrated in figure 1.1 which shows the insertion of the needle into the lungs. Once the biopsy needle has punctured the target nodule, a sample tissue is taken to further analysis. Although

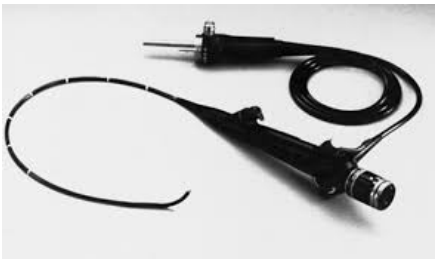
Figure 1.1 – Transthoracic needle aspiration procedure. A long needle is inserted into the chest wall, between the ribs and through the pleura towards the target nodule. Once the needle punctures the target nodule a sample is taken for further analysis.

TTNA is a minimally invasive procedure, the insertion of the needle across the patients' ribs and pleura is a delicate procedure that can generate pneumothorax in 20% of the cases.

An alternative (also minimally invasive) to TTNA is to perform a bronchoscopic explorations. A bronchoscopy is an endoscopic technique for the internal exploration of airways which facilitates the performance of minimal invasive interventions with a low risk for the patient. Bronchoscopic procedures are performed routinely, with about 261 millions of interventions around the world per year. Bronchoscopes are slender, tubular devices with a small electric light and a camera mounted on the tip of them. Bronchoscopes are inserted into the airways through the nose or mouth and allow the examination of airways for assessment of abnormalities like bleeding, inflammation or tumors. The tube of the bronchoscope has a hole called working channel that allows the insertion of miniaturized surgical instruments (like clamps or aspiration needels) for minimally invasive interventions such as tissue sampling for biopsy. Figure 1.2 (a) shows a bronchoscopy procedure. It can be seen how the bronchoscope is introduced into the lungs through the mouth in order to reach a specific lesion. The whole procedure is rendered in monitors which are used by the doctors to explore the airways and guide the bronchoscope

(a) Bronchoscopy procedure

(b) Rigid bronchoscope



(c) Flexible bronchoscope

Figure 1.2 – Illustration of a bronchoscopy procedure and tools.

to the target lesion.

There are several kinds of bronchoscopes ranging from rigid metal tubes with attached lighting devices to flexible optical fiber instruments with real time video equipment. Rigid bronchoscopes [6] like the one shown in Figure 1.2 (b) are rigid, straight, hollow metal tubes available in several sizes. Its purpose is to provide access to the airways. The external diameter of a rigid bronchoscope varies from 2 to 14 mm, wall thickness ranges from 2 to 3 mm, and length varies from a very short tube (for pediatric cases) to a long or extra-long tube (for adults). Rigid bronchoscopes can be used for either diagnostic or therapeutic reasons. Modern use is almost

exclusively for therapeutic indications such as retrieving inhaled foreign objects.

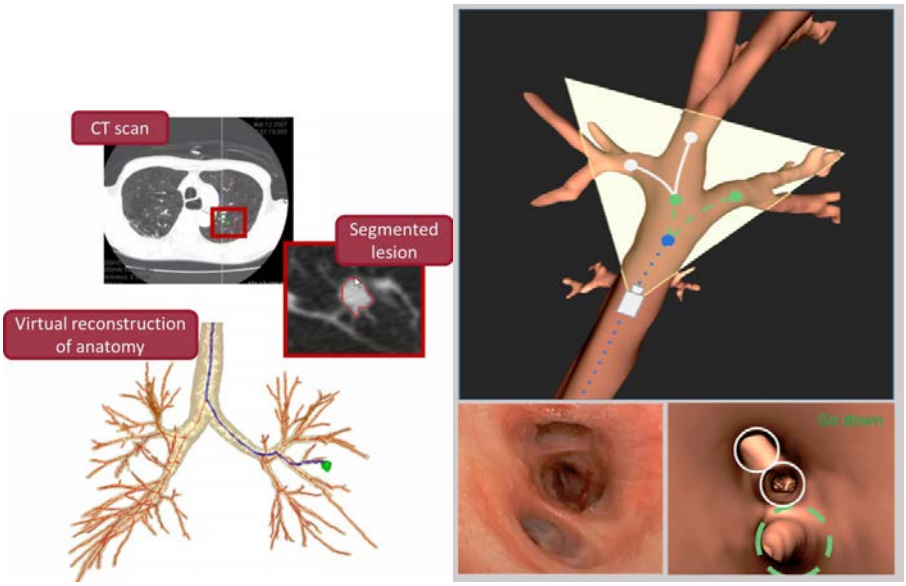
Flexible bronchoscopes [9] like the one shown in Figure 1.2 (c) have a fiberoptic system that transmits an image from the tip of the bronchoscope to the opposite end. Since they are long and thin, they allow patients to breathe during the intervention. In addition, they permit access to deeper bronchi. Therefore,flexible bronchoscopes are the preferred choice for both diagnostic and therapeutic procedures such as tissue sampling for biopsy.

A main difficulty for biopsy sampling is the navigation through the airways to reach the lesion. Conventional bronchoscopic diagnostic procedures are visually guided using radiating fluoroscopy and render a suboptimal 34% of positive results for lesions < 2 cm [10]. New endoscopy (like electromagnetic navigation) techniques are expensive, require either manual intervention or special gadgets, only increase diagnostic yield to 70%, and still radiate the patient. The 30% undiagnosed pulmonary lesions need CT follow-up or futile surgery procedures such as thoracoscopies, which induces patient anxiety, radiation exposure, invasive surgery along with associated pain, disability and rarely death.

Diagnostic yield could be improved reducing radiation and costs by developing intervention support systems able to guide the bronchoscopist to the lesion. Despite the improvement in endoscopic device quality, there is a lack of intelligent computational systems for supporting in-vivo clinical decision during examinations. Existing technologies fail to accurately reach the lesion due to several aspects at intervention off-line planning and intra-operative difficulties at exploration time.

At planning, CT imaging is used in bronchoscopy to create a planning of the intervention in a methodology known as Virtual Bronchoscopy (VB) [21]. Virtual bronchoscopic navigation systems [49] are used to reconstruct computed tomography (CT) data into three dimensional representations of the tracheobronchial tree. The reconstruction of the bronchial tree is used to compute the shortest path across brochial levels to each target nodule. Current Virtual Bronchoscopy systems [2] ignore bronchoscope mechanical constraints and bronchi breathing deformations and, thus, cannot provide information on whether the target nodule is reachable or not in practice.

Figure 1.3 illustrates the main functionalities of a Virtual Bronchoscopic system. Figure 1.3 shows the 3D reconstruction of the shortest path to the lession (left) and the virtual navigation (right) simulating the bronchoscopy exploration. The path to the target nodule is also indicated with a solid line inside the reconstructed volume in the lower part of Figure 1.3 (a)). The reconstruction of the 3D anatomy as well as the location of the lesion is obtained from the CT scan (shown in the upper images of Figure 1.3 (a)). The lesion is shown in the enlarged region enclosed by the red square. To help brochoscopist during exploration time, VBs also displays a simulation of the planned navigation indicating at each bronchial level the branch




(a) Virtual bronchoscopy planning.

(b) Virtual bronchoscopy guidance.

Figure 1.3 – Virtual bronchoscopy system for biopsy intervention guidance. At the planning stage, the volume and the centerline of the airways is extracted from a CT which is used to plan the actual intervention (a). At the intervention stage, the direction to follow at each bifurcation is shown in order to guide the bronchoscopist to reach the target lesion (b).

that needs to be followed (Figure 1.3 (b)).

Even with a pre-planned route extracted from a CT, bronchoscopists need to reproduce such a route by visual identification of bronchial levels and branch orientation in the intra-operative bronchoscopy video. Even for expert bronchoscopists it is difficult to reach a lesion due to the lung's anatomical structure. There are up to 21 bronchial levels with a branching geometry which has large intra-subject variability from the 4th level on. Besides, the projection of bronchi have a circular symmetrical geometry which makes difficult their proper identification in case that there is a rotation between VB navigation and the actual exploration. A rotational misalignment is common in order to reach some of the lungs lobes such as the superior lobes. Therfore, the further the lesion is, the hardest is for the bronchoscopists to reach it. Despite the improvement in endoscopic device quality, there is a lack of

intelligent computational systems for supporting in-vivo clinical decisions during bronchoscopic examinations. To assess the navigated path, bronchoscopists use a technique called fluoroscopy (Figure 1.4 (b)) to obtain real-time X-ray images of the lungs. Aside from errors arising from visual interpretation, fluoroscopy implies repeated radiation for, both, clinical staff and patients [55]. Protocols relying on fluoroscopy have a diagnostic yield around 40%, last 20 min per intervention and require 5-10 min of repetitive patient and medical staff radiation.

## 1.2 Goal of the thesis

The goal of this PhD work is to develop an image-based navigation system for intra-operative guidance of bronchoscopists to a target lesion across a path previously planned on a CT-scan.

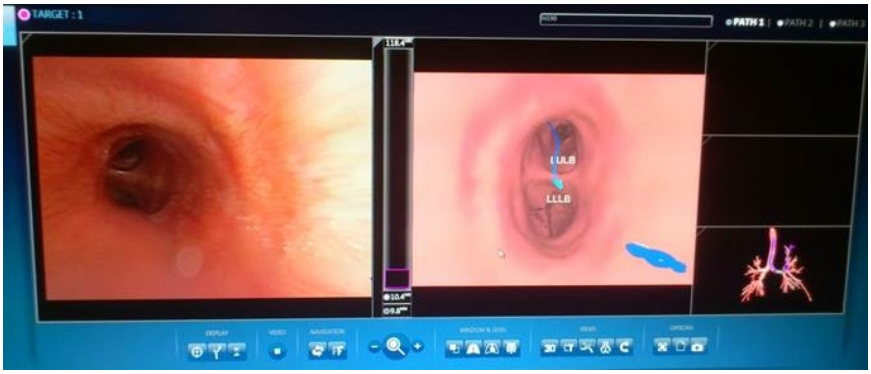
## 1.3 Existing guiding systems for bronchoscopy biopsy

During the past years, several technologies have been developed for off-line planning the best path to a lesion and on-line guiding the bronchoscopist through the planned path. Existing systems can be splitted into purely image-based navigation systems and systems using specific tools that provide additional information helpful in the guidance process such as electromagnetic navigation systems.

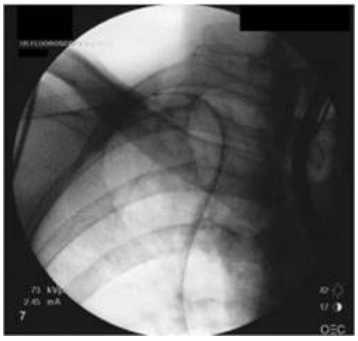
Electromagnetic Navigation Bronchoscopy [19] (ENB) is a medical procedure designed to localize and guide both bronchoscope and bronchoscopic tools through the bronchial tree by means of electromagnetic waves. ENB consist of two stages: planning and navigation. Figure 1.4 (c) shows an electromagnetic navigation procedure At planning stage, ENB uses CT virtual bronchoscopy to mark target locations and plan paths to reach target nodules within the lungs. At navigation stage, patients are placed on the magnetic navigation board and bronchoscopists navigate the sensor probe and extends the working channel to the desired target location. Sensor probes return 5-DoF positions which in combination with the CT virtual bronchoscopy allow bronchoscopists to navigate through the lungs to reach the target lesion. Once at the target position, bronchoscopists locks the extended working channel in place and the steerable sensor probe is removed. The extended working channel provides access to the target lesion for standard bronchoscopic tools or catheters. The main disadvantage of electromagnetic navigation systems is that these specific gadgets increase the costs of interventions limiting it use to resourceful entities. In addition, they might not be accurate enough due to interfer-

ences between the electromagnetic waves and human tissues and lack of rotational information.

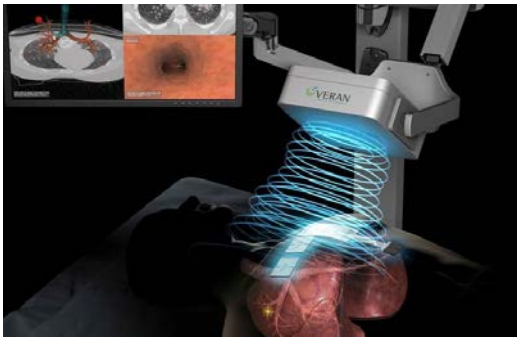
Image-based navigation systems try to put into correspondence video-bronchoscopy images and CT virtual bronchoscopy images (Figure 1.4 (a)) using multimodal registration techniques most of the times. Such registration method optimize a virtual bronchoscope pose (rotation and position) in a 3D CT volume. For each virtual bronchoscope pose, 2D projections are generated using render techniques and these images are compensated with intra-operative images to obtain the pose of the virtual camera that is closest to the scope tip pose.




(a) Lung Point© navigation system.



(b) Fluoroscopy procedure.




(c) Electromagnetic navigation

Figure 1.4 – Existing techniques supporting guidance during a bronchosopic procedure. commercial image-based guidance system, (a), standard radiating fluoroscopy, (b) and electromagnetic navigation system, (c).

Image registration techniques are prone to fail, specially in case of large motion, to a suboptimal solution given by a local minima. To reduce the impact of local minima and surgical artefacts, most methods compute an initial affine transformation. In [33], the initial transformation is computed using SIFT feature matches and epipolar constraints. A main inconvenience is that SIFT matching require detection of accurate feature points assuming textured images with singular points like corner or junctions. However, medical scans of anatomical structures are usually textureless volumes with smooth level sets. In particular, in the case of endoscopic videos, SIFT features might correspond to specular shines introduced by the camera lighting. Thus, they move along with the scope and should not be used to match images because they illuminate points that correspond to different anatomical sites. Other systems replace SIFT feature matches by optical flow [39] or Kalman filtering [41] in order to estimate a rough first bronchoscope position by applying dynamical models. In any case, the initial affine transformation it is refined applying an intensity-based image registration algorithm. The main disadvantages of such methods is that they are very time consuming and can lead to a mismatch when images are obscured by blood or mucus and bronchi are deformed by patient's coughing.

The synchronization of simulated navigation with the actual intra-operative video at interventional time is such a chanllenge that there is not any comercial software able to provide a solution clinically valid yet. Currently, the only product available on the market is the Lung Point pltaform [12]. Lung Point (See Figure 1.4 (a)) is an image-based navigation commercial system which helps to localize a target lesion at intervention time. Given a pre-computed CT scan, the system displays two images during intervention: the live intervention video (left image in Figure 1.4 (a)) and CT-VB showing the pathway to follow to reach the target nodule (right image in Figure 1.4 (a)). The system aligns the virtual images with the images from the bronchoscopy video by manual synchronization. Once the two images are synchronized, Lung Point calculates the position of the bronchoscope in the airway and overlays the pathway to the target on the bronchoscopic video image. The main disadvantages of Lung point are that it is extremely expensive and the image-based synchronization is not powerful enough to automatically navigate through the whole path. It requires a second doctor who corrects the position of the virtual bronchoscope which increases both intervention cost and waiting lists.

Recent approaches are starting to use techniques, like Simultaneous Localization and Mapping (SLAM), that are well-known in other aplication areas such as robotics or advanced driver-assistance systems (ADAS). SLAM constructs and updates a 3D map of an unknown environment while at the same time keeps track of the camera location. For the construction of the 3D maps it is necessary to find matches between consecutive frames. Existing SLAM methods used in ADAS and

robotics use SIFT-based matchings which implies that images need to be textured so that similar feature points can be found in every frame. In this context, ORBSLAM [40] achieves great performance with respect to other state-of-the-art monocular SLAM approaches because it has adapted feature detectors to the specific surgical conditions of endoscopic interventions. In particular, it uses ORB which is a fast robust local feature detector that fuses FAST keypoint detector and BRIEF descriptor with specific modifications to enhance performance under surgical conditions. Although the method works in some surgical procedures, like laparoscopic interventions, it is not guaranteed an optimal performance when applied to bronchoscopy procedures. This follows from the observation that aside textureless bronchoscopic images, the bronchoscope movements are basically backward and forward along center lines. Such camera motion, makes monocular SLAM a bad-posed problem since small errors in the camera pose estimation, introduce large errors in the 3D estimation.

In recent years, several alternatives to classical intensity-based registration methods have begun to be explored. These new image-based methods include Convolutional Neural Networks (CNNs) to compute 3D point clouds from depth maps learned from images and synchronization based on visual features that can be identified in, both, CT scans an videos.

A feasible alternative to cope with mismatches due to images obscured by blood-/mucus and bronchi deformations is the use of anatomical landmarks. The main advantages of using anatomical landmarks are that they are more robust to illumination changes than multi-modal registration methods and there is no need of using huge annotated datasets in order to train complex methods. A previous work [53] has already proved the capability of anatomical landmarks applied to bronchoscopic navigation. In that work, center lines are used to detect the route followed by the bronchoscope and indicating the path that needs to be followed to reach a target lesion. A weakness of the method is that center lines are extracted frame by frame from the centers of the airways luminal areas. Adding temporal information could improve drastically the outcome of such algorithm. Despite of promising results shown in [53], the method was only proved in virtual environments due to a lack of robustness under surgical conditions.

In a very recent work [58], CNNs are trained to generate depth maps from bronchoscopy images. Depth maps allows to generate 3D point clouds that are registered to the CT volume using Iterative Closest Point (ICP) or similar methods. By using image depth maps instead of 2D features point matches, the method is more stable against intensity changes and surgical artifacts.

In spite of the promising results[5, 16, 57] of the use of CNNs in the medical imaging domain, in particular in navigation systems [58], a main inconvenience is the limited annotated data available. The quality and quantity of annotations is

a bottle neck for training CNNs and other state of the art classification methods. Although fine-tuning of existing CNNs trained with natural images can achieve better results that handcrafted features [27], the required level of fine-tuning is highly application dependent with some applications requiring a simplified network architecture be trained from scratch [38]. It follows that some sort of data augmentation should be applied to complement the available intra-operative data.

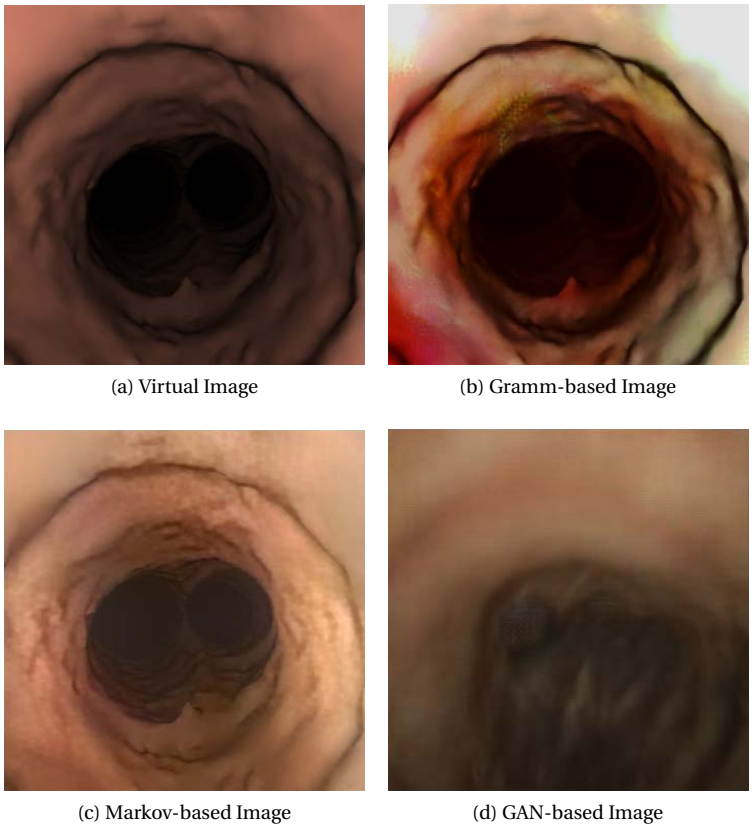
### 1.3.1 Data augmentation for virtual bronchoscopy

Augmenting small datasets is important and challenging. Many approaches apply simple transformations (like rotation, mirroring and changing contrasts among others) to the available samples to augment their number at the training stage. This makes data analysis invariant to such transformations, but the lack of independence in the augmented set could limit the amount of new information that is actually added. Alternatives to this kind of data augmentation could be the use of either images having similar content or virtual data obtained by computational simulations.

In the context of health applications, the selection of images having similar content can be a challenging problem, especially in case of pathological cases or surgical conditions. Although this could be solved using data augmentation techniques by means of virtual environments, data augmentation using virtual images is a delicate issue in medical applications. This follows from the observation that the appearance and content of simulated images are substantially different from images gathered in interventions so algorithms can not be trained and validated under true clinical conditions. We consider that to satisfactorily use simulated data for training, virtual simulations should be augmented to preserve the content and the appearance of real intervention videos.

A possible way to generate new valid independent samples for training stages and algorithm verification is to use style transfer techniques. Style transfer [13, 31] consist in the recomposition of images and texture using the style of a different set of (style) images. Such techniques have been widely used in applications which modify an input images into stylized images similar such as well-known painter's drawings or into any different domain [24]. This same idea can be used to transform virtual images into intra-operative images in a smart way so that stylized images do not loose any content.

State-of-art techniques for style transfer based on CNNs follow an auto-encoder architecture to transfer into input images the appearance of style images. Existing techniques can be split into methods transferring style between image pairs and methods transferring style between image domains. The first group [22, 29, 43] requires the identification of the style image that best resembles the input image.



(a) Virtual Image (b) Gramm-based Image

(c) Markov-based Image (d) GAN-based Image

Figure 1.5 – Virtual image stylized with different state-of-art methods.

This is not possible in most of the cases. Such pairs are not available in most of the medical imaging problem. For that reason, the second group looks for the transformation that maps the space of input images into the space of stylized images and, thus, do not require image pairs. In order to do so, a system of two different neural networks (a generative network and a discriminative one) to obtain stylized images achieving a compromise between preservation of input image content and style images texture and appearance is used. The first network is an auto-encoder that generates stylized images from the input images. The output of this auto-encoder is the input of a discriminative network that classifies between style and

input images to assess how much stylized images appearance matches the actual style. This general scheme has several variants concerning, especially, the kind of loss function used to train the network.

In [17, 25] the generative network minimizes a loss function that includes two terms. One term (feature reconstruction loss) penalizes that stylized images deviate in content from input images, while the second one (style reconstruction loss) measures the similarity between stylized images and style appearance. The feature reconstruction loss is given by the $L_2$ difference between feature maps of input and stylized images. The style reconstruction loss is given by the Frobenius norm of the difference between the Gram matrices of the stylized and style images feature maps. As explained in both works, since Gram matrices encode probabilistic distribution correlations, minimizing the style reconstruction loss preserves stylistic features from the style image but does not necessarily preserve its spatial structure (as illustrated in Figure 1.5 (b)).

This invalidates the use of Gramm metrics for augmentation of simulated data using interventional images, since the spatial structure of their appearance is strongly related to anatomical content. This phenomena is illustrated in Figure 1.5 for a bronchoscopic image simulated from a CT-scan using virtual bronchoscopy. This simulation (shown in Figure 1.5 (b)) has been augmented with the appearance of the intra-operative image of Figure 1.5 (a)) using [17]. We observe that in the stylized image shown in Figure 1.5 (c)) dark areas have become bright areas and the other way around. Thus, any algorithm for anatomy detection or modelling (such as image-based navigation systems [15, 54]) would fail.

Other approaches like [30] are based on deep Markovian models and transform images locally instead of globally. To do so, feature maps are split into patches which are the input of the classifier that discriminates between real and virtual appearances. Like [17, 25] the loss function also includes a content regularization term to preserve the spatial structure of images. However, the fact that style is locally transferred also leads to a loss of anatomical content in style images.

New approaches such as [61] use Generative Adversarial Networks (GANs) in order to transform images from one domain $A$ (like virtual simulations) into a domain $B$ (like interventional videos). The novelty of [61] is that a cyclic term is added in order to make the domain transfer bijective ($A \rightarrow B \rightarrow A$ and $B \rightarrow A \rightarrow B$). Although, the method also adds a regularization term to preserve the spatial structure of the stylized virtual images, content information is still lost as shown in Figure 1.5 (d). Figure 1.5 (d) shows the augmentation of Figure 1.5 (a) using GANs. The image has no illumination artifacts and looks quite real but the main anatomy of Figure 1.5 (a) has been significantly distorted. This alteration of anatomical content invalidates GAN for clinical use. Figure 1.5 (c) shows the image in Figure 1.5 (a) augmented using [25]. We observe that, although, the main anatomy is prserved

like in Figure 1.5 (b), the augmented image presents some artificial shines and shadows.

## 1.4 Contributions of this thesis

This thesis contributes to image-based navigation systems and data augmentation in medical imaging in the following aspects:

- **Contribution 1: A navigation system based on bronchi anatomical landmarks**.

  The anatomy of airways can be represented using a tree data structure with nodes bronchi branching points. At the planning stage, such structure is obtained from CT scans using the branching structure of the skeleton of the segmented airways. This provides a static representation of the global anatomy of each patient. In bronchoscopy videos, airways anatomical structure is encoded in a collection of luminal regions with centers corresponding to airways centerline. Luminal regions of different bronchial levels projected onto the same reference results in a hierarchy of regions that matches the global anatomy of the patient. This hierarchy is represented in a tree data structure constructed dynamically by computing and tracking anatomical landmarks from in-vivo intra-operative bronchoscopic procedures. The dynamical tree is matched to the anatomical structure of the lung extracted from a CT in order to locate the bronchoscope inside the lungs. Once the bronchoscope is located, our navigation system is able to indicate the bifurcation which needs to be followed to reach the lesion.

  Figure 1.6 illustrates the whole pipeline of the proposed navigation systems. In the top section of Figure 1.6, a set of anatomical landmarks are detected in a videobronchoscopy frame. The set of landmarks is processed and the anatomical hierarchy is encoded in a tree structure. The bottom left section of Figure 1.6 shows how the anatomical hierarchy is tracked across the whole videobronchoscopy in order to build dynamically an on-line interventional tree. Finally, the bottom right section of Figure 1.6 describes how the online exploration tree is put in correspondence with tree representing the whole airways anatomy and the planned path.

  In this work we contribute to the extraction of the exploration intra-operative path from the analysis of videobronchoscopy frames in the following aspects:

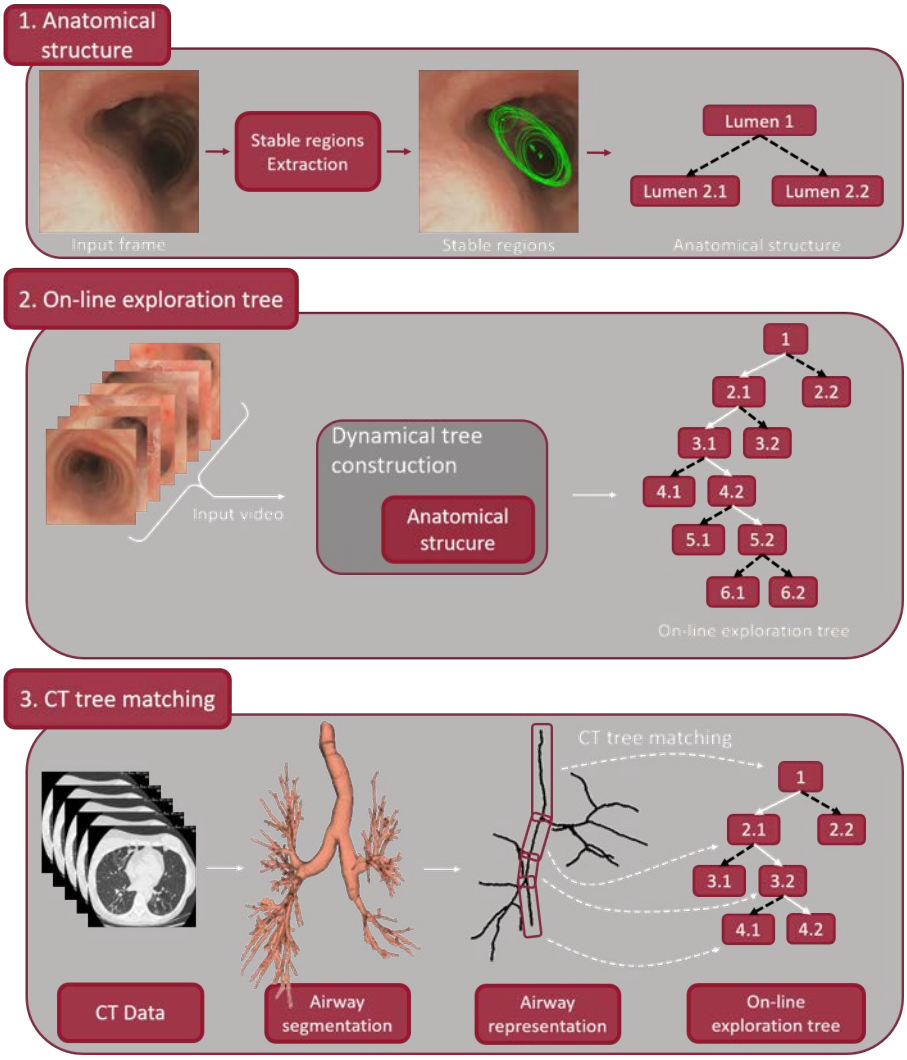



Figure 1.6 – Pipeline of the proposed navigation system.

– **Contribution 1.1: Extraction of bronchial anatomical structure in single videobronchoscopy frames.** We have improved the method presented in [51] by developing a video processing algorithm that dynami-

cally extracts stable anatomical landmarks from bronchoscopic images. Such extracted anatomical landmarks encode the local structure of the bronchial tree in the image. For each frame, a hierarchy of the projected luminal regions is extracted and codified in a tree structure

  - **Contribution 1.2: On-line exploration tree construction.** The bronchial path followed during a bronchoscopic exploration is encoded using a binary tree dynamically constructed from the hierarchy of anatomical landmarks. The hierarchy is extracted for each of the frames of the intervention and it is tracked along the whole sequence in order to endow temporal consistency. Temporal consistency allows us to build an on-line exploration tree which represents the path followed during the intervention. We call the tree on-line exploration tree and it represents the bronchial bifurcations observed during the intervention.

- **Contribution 2: Augmentation of virtual images with intra-operative appearance.**

  We present a two-stage algorithm for the augmentation of virtual endoscopic images using intra-operative videos based on convolutional neural networks. First, we use cycleGAN in a multi-objective optimization scheme to obtain pairs of virtual and intra-operative style images that share some content information. The content and appearance of these image pairs are blended using a siamese u-net architecture that modulates skip connections by a measure of neuron activation content.

  A general overview of our proposed algorithm for augmenting of virtual bronchoscopy images with intra-operative appearance can be seen in Figure 1.7. In the top section of Figure 1.7, the GAN architecture used to generate augmented virtual images is presented. The network is trained using adversarial training. During training, a two-terms loss is optimized. The bottom left section of Figure 1.7 plots the Pareto front. Pareto front is a technique used in multi-objective optimization problems to automatically select the epoch with the best trade-off between the loss terms of the adversarial training. GAN architecture generates images with intra-operative appearance but it changes the content of the image. To cope with this effect, an extra step, presented in the bottom right part of Figure 1.7, to blend the content and appearance of these image pairs using a siamese u-net architecture that modulates skip connections by a measure of neuron activation content.

  The contributions of our work are the following:

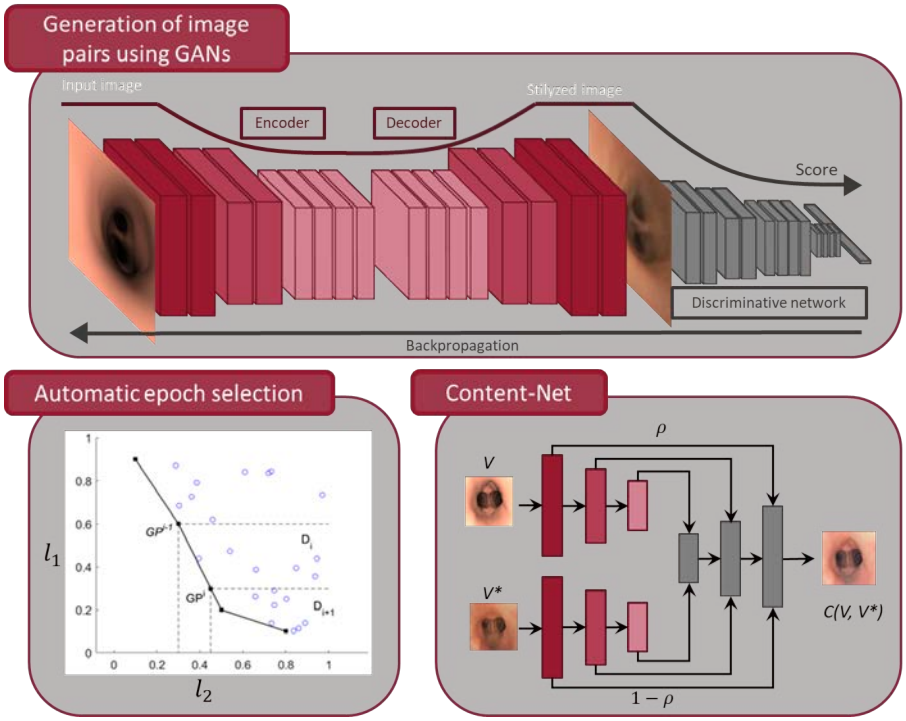



Figure 1.7 – Pipeline of the proposed method for augmenting virtual bronchoscopy images with intra-operative appearance.

– **Contribution 2.1: A multi-objective approach for Generative Adversarial Networks**. A main challenge with GANs [46] is the selection of the epochs most suitable for a given problem. Due to the oscillating behavior of adversarial training, most of the cases this is done manually, which leads to subjective errors depending on the observer. In this work, we propose a multi-objective optimization approach based on the Pareto front [37] to select the epoch achieving the best compromise between content preservation and style transfer.

– **Contribution 2.2: A siamese u-net network (Content-net) for blending content and appearance of image pairs**. We introduce an auto-encoder with siamese encoders and skip connections to the decoder. Skip connections are modulated by a measure of the information accord-

ing to the type of information filters respond to. Modulation is used to fine tune the amount of virtual anatomical content and intra-operative appearance of final simulations. This way we can produce images with several degrees of interventional artifacts from the same pair of images obtained from multi-objective GANs.

The remains of this thesis are structured as follows: Chapter 2 details the algorithms to generate virtual bronchoscopy images with intra-operative appearance. Chapter 3 is dedicated to explain in detail the 3D navigation system, including the extraction of bronchial anatomical structure from videobronchoscopy is, as well as, the matching between a pre-planned path tree and an on-line exploration tree. Chapter 4 reports the experiments and validation performed in the algorithms developed in previous chapters (Chapter 2 and Chapter 3). Finally, Chapter 5 closes the thesis by exposing the main conclusions that can be extracted along with sketching the opened future lines of research.

# 2 Augmentation of virtual bronchoscopy images with intra-operative appearance

In order to obtain realistic data useful for data augmentation and validation of machine learning and image processing methods, simulations should resemble intra-operative recordings. We consider that style transfer could be used to endow virtual endoscopic images with the content and texture of intra-operative videos using modern techniques for artistic style transfer.

The main difference between artistic style transfer and realistic simulations of endoscopic procedures is that in the latter, stylized images should preserve the structure of simulated data. This comes from the fact that style structure encodes the anatomical content of the image and, thus, it should be preserved. Therefore, the spatial structure of style target images should be similar to match the content of input images if such data has to be used for data augmentation or ground truth generation.

In this chapter, we present a method for the generation of endoscopic images using intra-operative video data. Our strategy is a two stage method based on CNNs that maps virtual images to the intra-operative domain preserving their anatomical content. In a first stage, pairs of virtual and intraoperative images sharing anatomical content using GANs are generated. We propose a multi-objective optimization based on Pareto front strategy for selecting GAN epochs ensuring the best compromise between content preservation and style transfer of virtual images and intra-operative images. In a second step, the content and appearance of such pairs

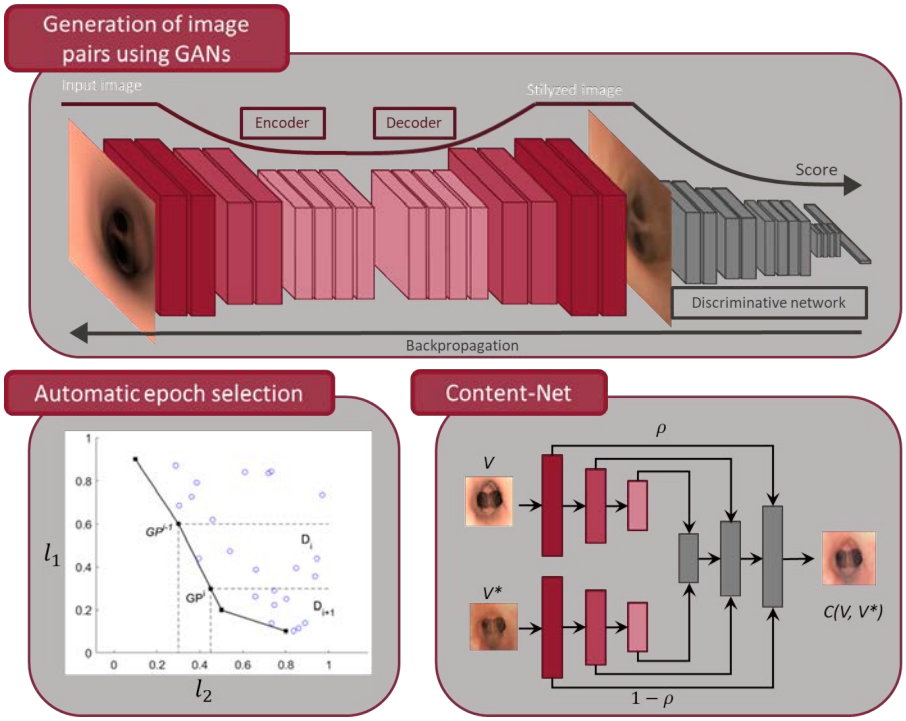



Figure 2.1 – Pipeline of our proposed method for augmenting of virtual bronchoscopy images with intra-operative appearance.

are blended using a siamese u-net architecture trained to modulate the amount of content and texture that is taken from each pair. The pipeline of our proposed algorithm is illustrated in Figure 2.1.

This chapter is organized as follows. Section 2.1 explains the multi-objective approach for selection of cycleGAN epochs achieving the best trade-off between the different terms of the cycleGAN loss function. Section 2.2 is dedicated to improve the compromise between content preservation and style transfer given by cycleGAN. Section 2.2 defines Content-net architecture, as well as, the functions for blending content and appearance of the image pairs obtained from our multi-objective cycleGAN.

## 2.1 Multi-Objective Generative Adversarial Networks

Given two domains $Virtual$ and $Real$, the main goal is to learn two (bijective) mapping functions from one domain onto the other one:

$$G_r : Virtual \rightarrow Real \quad G_v : Real \rightarrow Virtual \tag{2.1}$$

with the map composition $G_r(G_v)$ and $G_v(G_r)$ being the identity on each domain. Following [61], maps are given by auto-encoders trained to optimize:

$$\boldsymbol{\ell}(G_r, G_v, D_r, D_v) = \boldsymbol{\ell}_{\mathbf{GAN}}(G_v, G_r, D_r, D_v, V, R) + \lambda \boldsymbol{\ell}_{\mathbf{cyc}}(G_v, G_r) \tag{2.2}$$

for $R = \{r, r \in Real\}$, $V = \{v, v \in Virtual\}$ two sets of, respectively, real and virtual images. The term $\boldsymbol{\ell}_{\mathbf{GAN}}$ measures how good are $G_v, G_r$ transferring images from one domain to the other one and is expressed as:

$$\boldsymbol{\ell}_{\mathbf{GAN}}(G_v, G_r, D_r, D_v, V, R) = \boldsymbol{\ell}_{\mathbf{GAN}}(G_r, D_r, R, V) + \boldsymbol{\ell}_{\mathbf{GAN}}(G_v, D_v, R, V)$$

$$\boldsymbol{\ell}_{\mathbf{GAN}}(G_r, D_r, R, V) = D_r(r)^2 + (D_r(G_r(v)) - 1)^2$$
$$\boldsymbol{\ell}_{\mathbf{GAN}}(G_v, D_v, R, V) = D_v(v)^2 + (D_v(G_v(r)) - 0)^2$$

for $D_v$ and $D_r$ two adversarial discriminative networks trained to distinguish between real, $r$, and transformed virtual, $G_r(v)$, images and viceversa. The term $\boldsymbol{\ell}_{\mathbf{cyc}}$ is a "cycle consistency loss" introduced to force mappings to be bijective:

$$\boldsymbol{\ell}_{\mathbf{cyc}}(G_r, G_v) = ||G_v(G_r(v)) - v||_1 + ||G_r(G_v(r)) - r||_1$$

for $||\cdot||_1$ the $L^1$ norm.

The minimization problem is given by adversarial training as:

$$G_r^*, G_v^* = \underset{G_r, G_v}{\operatorname{argmin}} \left( \underset{D_r, D_v}{\operatorname{argmax}} \boldsymbol{\ell}(G_r, G_v, D_r, D_v) \right) \tag{2.3}$$

This way, $G_r^*$ and $G_v^*$ are optimized so that $G_r$, $G_v$ minimize while the adversarial $D_r$, $D_v$ maximize it (Equation 2.2). Thus, adversarial training can be considered as an optimization problem that searches for the discriminators ($D_r$ and $D_v$) that maximize the term $\boldsymbol{\ell}_{\mathbf{GAN}}$ and the mapping functions ($G_r$ and $G_v$) that minimize the full objective including the maximal term $\boldsymbol{\ell}_{\mathbf{GAN}}$.

These conditions (minimize while maximizing at the same time) might be in conflict considered into a single optimization process. Such a conflict is prone to introduce an oscillating behavior of, both, the solution and the objective function

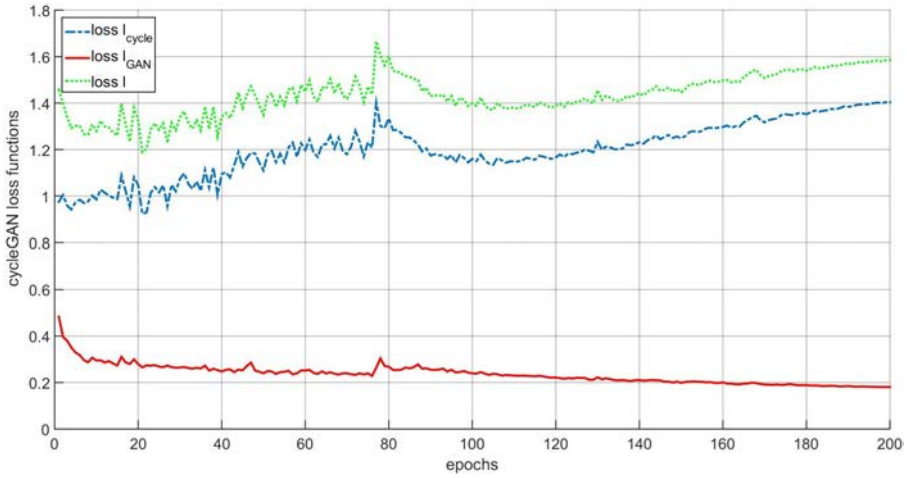



Figure 2.2 – Profile of $l_{cycle}$ and $l_{GAN}$ loss functions during GAN training stage. Such GAN was trained for transforming images from VB simulation images ($Virtual$) and intra-operative images ($Real$).

in a gradient descent iterative scheme. The oscillatory behavior hinders the convergence of cycleGAN training stage and invalidates the use of most stop criteria [62, 63] for single optimization problems for the selection of GANs epoch, since they assume smooth convergence to the solution.

Figure 2.2 illustrates the oscillatory behavior in cycleGAN losses curing training. We plot the two losses terms ($l_{GAN}$ and $l_{cycle}$ that are combined in Equation 2.2), as well as, the global loss $l$ combining the two terms. We observe that the partial losses $l_{GAN}$ and $l_{cycle}$ are not comparable in terms of either magnitude or behavior (increasing, decreasing) across training epochs. Therefore, the magnitude sum $l$ presents an oscillatory behavior which prevents the training stage to converge in the sense of numerical methods. This oscillation implies that cycleGAN generates random mapping functions every single epoch since the loos function does not decrease towards a global solution.

The unstable behavior across epochs is illustrated in Figure 2.3. Figure 2.3 shows three different virtual images augmented using four consecutive epochs. It can be appreciated that augmented images anatomical structure and image appearance are completely different for consecutive epochs. In particular, they might have luminal areas not present in the original virtual image (like in the first raw of Epoch 105), as well as, absent lumens of virtual images (like in the last raw of Epochs 105

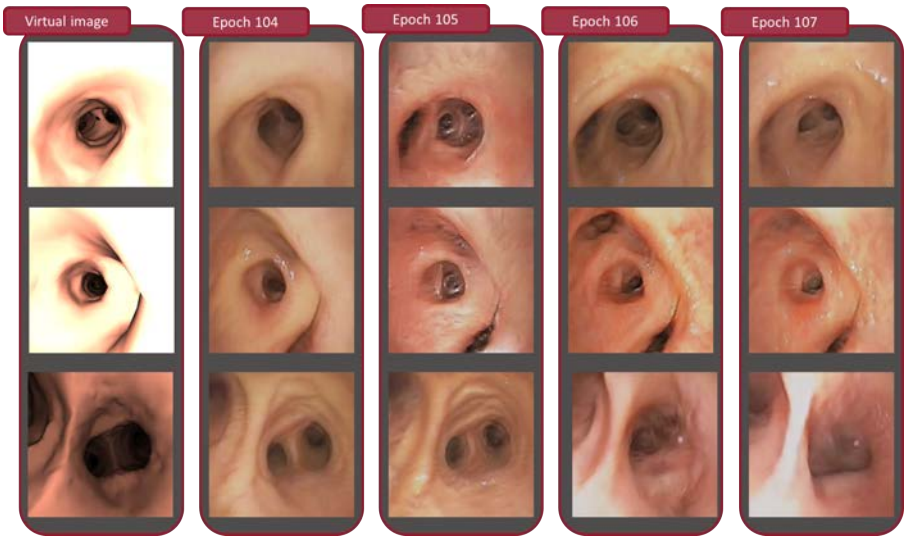



Figure 2.3 – Collection of augmented images for consecutive epochs.

and 107).

A main inconvenient of the lack of convergence in an iterative numeric schema is that the selection of the optimal iteration (epoch) has to be manually adjusted for each problem since no stabilization stopping criteria can be satisfactory applied. This is a main inconvenient for the systematic application of cycleGAN (GANs in general) since such optimal epoch is highly dependent on the training set. In order to provide a numeric criteria for the automatic selection of cycleGAN optimal epochs, we propose to consider adversarial training of Equation 2.2 as a multi-objective optimization [37] problem. Therefore, We propose to consider separately the optimization of each of the terms in the objective function $\boldsymbol{\ell}$ which allows to pose adversarial training as the following multi-objective optimization problem:

$$\begin{aligned} G_r^*, G_v^* = & \operatorname{argmin}_{G_r,G_v}(\boldsymbol{\ell}_{\mathbf{cyc}}, \boldsymbol{\ell}_{\mathbf{GAN}}) = \\ & \operatorname{argmin}_{G_r,G_v}(\boldsymbol{\ell}_{\mathbf{cyc}}(G_r, G_v), \operatorname{argmax}_{D_r,D_v} \boldsymbol{\ell}_{\mathbf{GAN}}(G_v, G_r, D_r, D_v)) \end{aligned} \tag{2.4}$$

Since a multi-objective optimization problem involves the optimization of multiple objective functions, there does not exist, in general, a solution simultaneously minimizing all of them. The expected situation is to have a set of solutions that outperform in any of the objectives without degrading the other ones. These dominating solutions is called Pareto front and is defined as the set of solutions, $x_P$, such

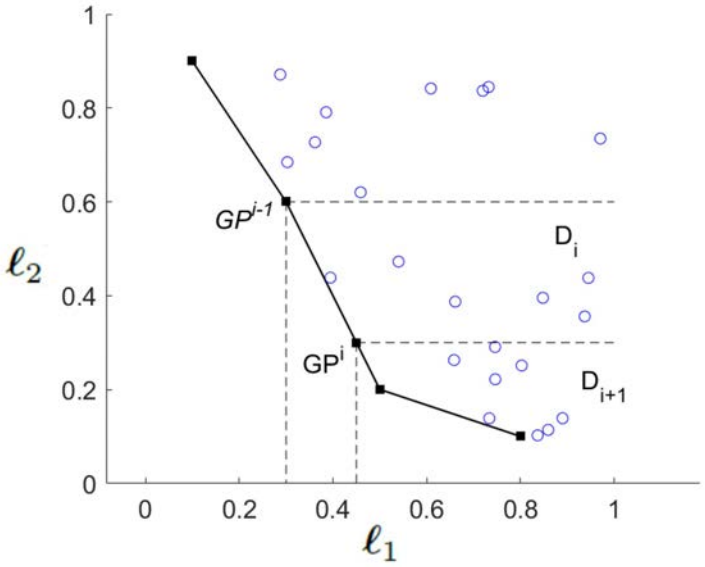



Figure 2.4 – Pareto front of cycle-GAN 2-objective optimization.

that:

$$\forall x \text{ such that } \boldsymbol{\ell}_i(x) \leqslant \boldsymbol{\ell}_i(x_P)\ \forall i \in \{1,2,...,k\}$$
$$\exists j \in \{1,2,...,k\}\ \boldsymbol{\ell}_j(x_P) < \boldsymbol{\ell}_j(x)$$

being $\boldsymbol{\ell}_1,\ldots,\boldsymbol{\ell}_k$ the set of functions to be optimized. In our case $\ell_1 = \ell_{cyc}$ and $\ell_2 = \ell_{GAN}$.

The condition of the Pareto front can be used to select cycleGAN epoch as follows. Let $\boldsymbol{G}^k := (G_r^k, G_v^k)$ be the transformation maps at the $k$-th epoch and $\boldsymbol{GP} = (\boldsymbol{GP}^i)_{i=1}^{NP}$ be the set of epochs belonging to the Pareto front. Such Pareto maps can be iteratively computed from the values of the objective functions as:

$$\boldsymbol{GP}^i := \underset{\boldsymbol{G}^i \in \boldsymbol{D}_i}{\operatorname{argmin}}(\boldsymbol{\ell}_{\mathbf{cyc}}(\boldsymbol{G}^i)) \tag{2.5}$$

for $\boldsymbol{D}_1$ the set of maps for all epochs and $\boldsymbol{D}_i$, $i > 2$, the set of maps dominating $\boldsymbol{GP}^{i-1}$. In our case, $\boldsymbol{G}^i \in \boldsymbol{D}_i$ if it satisfies the following conditions:

1. $\boldsymbol{\ell}_{\mathbf{cyc}}(\boldsymbol{G}^i) > \boldsymbol{\ell}_{\mathbf{cyc}}(\boldsymbol{GP}^{i-1})$
2. $\boldsymbol{\ell}_{\mathbf{GAN}}(\boldsymbol{G}^i) < \boldsymbol{\ell}_{\mathbf{GAN}}(\boldsymbol{GP}^{i-1})$

Figure 2.4 shows the Pareto front associated to the two-objective problem given by (Equation 2.4) with the dashed lines enclosing the region that includes the set of epochs that dominate a given $\boldsymbol{GP}^{i-1}$. The pairs of function values $(\boldsymbol{\ell}_{\mathbf{cyc}}(\boldsymbol{GP}^i), \boldsymbol{\ell}_{\mathbf{GAN}}(\boldsymbol{GP}^i))$, $\boldsymbol{GP}^i \in \boldsymbol{GP}$, given by the two objectives evaluated at the Pareto front are shown in solid squares joined with a solid line. The set of Pareto epochs achieve the best trade-off between the two objective functions and, thus, are equivalent from the point of view of the GAN. The epoch from the Pareto front best suited for augmentation of virtual endoscopic images is selected as the one that minimizes the $L^2$-difference:

In order to select the epoch from the Pareto front best suited for a given problem, it is necessary to define a cost objective function measuring the adequacy of a GAN to the application specific requirements. In our case, since the anatomical content of virtual images should be preserved, our cost function for the final selection of the Pareto map is given by the $L^2$-norm of their difference:

$$\boldsymbol{\ell}_{\mathbf{Cont}} = \underset{v \in V}{mean}\left(||v - G_r(v)||_2\right) + \underset{r \in R}{mean}\left(||r - G_v(r)||_2\right)$$

for $||\cdot||_2$ the $L^2$-norm and $\underset{v \in V}{mean}$, $\underset{r \in R}{mean}$ the average values for the training set of virtual, $V$, and real images, $R$. The epoch selected, $\boldsymbol{G}_P^*$ is the one in the Pareto front achieving the minimum value of $\boldsymbol{\ell}_{\mathbf{cont}}$:

$$\boldsymbol{G}_P^* = (G_r^*, G_v^*) = \underset{\boldsymbol{G}_P^i}{\operatorname{argmin}}\left(\boldsymbol{\ell}_{\mathbf{Cont}}(\boldsymbol{G}_P^i)\right) \tag{2.6}$$

We will note the images transformed by these maps by $v^* := G_r^*(v)$ and $r^* := G_v^*(r)$.

## 2.2 Content-Net

The proposed content-net (labelled $C$) is an auto-encoder with siamese encoders (one for each image domain) that have skip connections to the decoder in order to blend content and style of image pairs. Siamese encoders follow a VGG-19 architecture. These are build using the first three convolutional blocks with two max pooling as shown in the sketch of Figure 2.5. The first block consist of 2 convolutional layers with 64 filters. Each convolutional layer is followed by a ReLU layer. The second block follows the same architecture as the first block but each convolutional layer consist of 128 filters. Finally, the third block consist of 4 convolutional layers with 256 filters. As in the first two blocks, each of the convolutional layers is followed by a ReLU layer. In addition to the convolutional and ReLU layers, there is a max pooling layer at the end of the first and the second block. Each max pooling between the

first and the second and the second and the third block is computed with a window size and a stride of two pixels. We decided to use only three convolutional networks since all the texture that we want to transfer is removed after two max pooling. Content-net architecture is sketched in Fig.2.6.

In order to selectively blend the content of one image with the appearance of the other one, skip-connections are weighted by a function, namely $\rho$, that quantifies the amount of content that filters respond to. Feature maps of the first siamese encoder are weighted with $\rho$, while its sister encoder are weighted with $1-\rho$. This way each siamese encoder contributes with either image content (first siamese encoder) or appearance (second siamese encoder).

The contrastive loss function that content-net minimizes is given by a content loss, $\ell_{\mathbf{Cont}}$, defined as:

$$\ell_{\mathbf{Cont}} = ||v - C(v, v^*)||_2 + ||v^* - C(v^*, v)||_2 \tag{2.7}$$

for $v$ and $v^*$ an image pair produced by our multi-objective cycleGAN and $C(v, v^*)$, denoting the output of content net with $v$, $v^*$ being inputs of, respectively, the first and second siamese encoders.

The function $\rho$ weighting skip-connections is learned from a training set of intra-operative images by comparing the input image to the activation of each neuron in an encoder trained to yield the identity map. The similarity measure chosen to compare input images to its neuron activations is their mutual information [45]. Mutual information compares the correlation between random variables and is

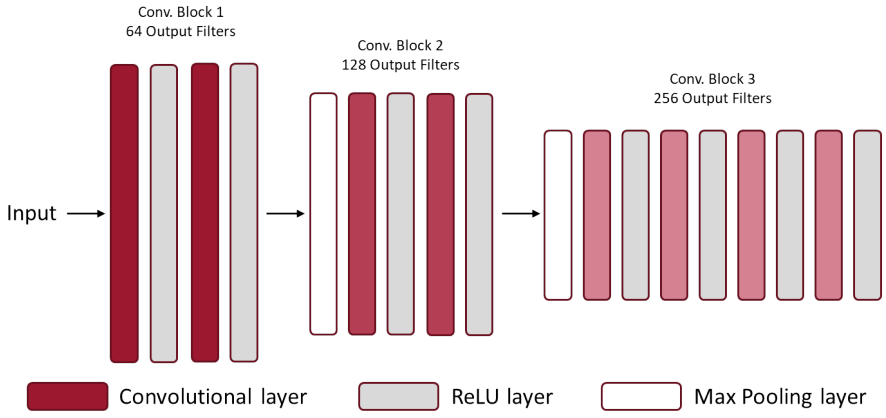



Figure 2.5 – VGG-19 architecture followed by siamese encoders.

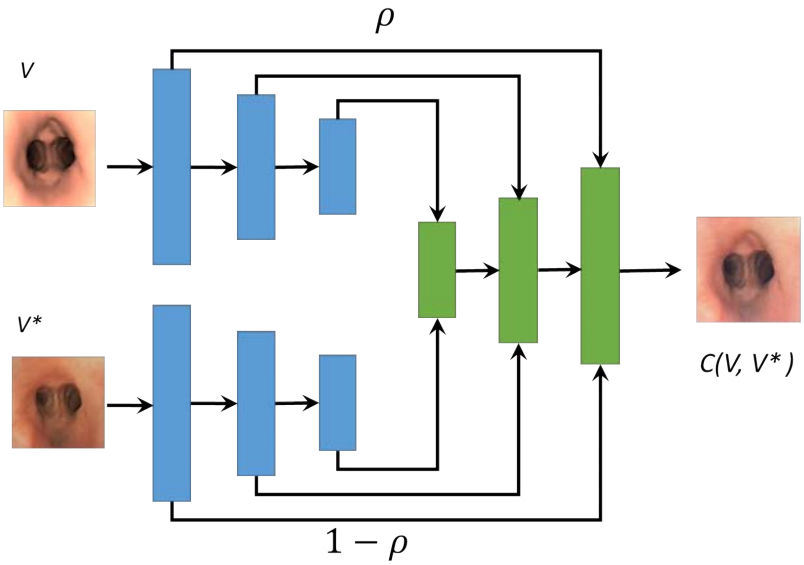



Figure 2.6 – Content-net Architecture used for blending image pairs.

used in multimodal registration to compare images with equal content but different appearance. If $W_{ij}$ and $b_{ij}$ indicates the weights and bias of the j-th convolutional layer of the i-th block, the input $X$ and output $Y$ of the convolutional layer is given by:

$$Y_{i,j} = W_{i,j} X_{i,j-1} + b_{i,j} \tag{2.8}$$

Therefore, if $X$ denotes the random variable of the input and $Y$ the random variable of the output of a convolutional layer, then their mutual information, $I(X,Y)$ is given by:

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) \tag{2.9}$$

where $log$ denotes the natural logarithm, $p(x,y)$ is the joint probability function of $X$ and $Y$, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of $X$ and $Y$, respectively. Mutual information measures the independence between the two random variables $X$, $Y$. In particular, a value close to 0 indicates that input image and activations do not share information. Meanwhile, in case of sharing

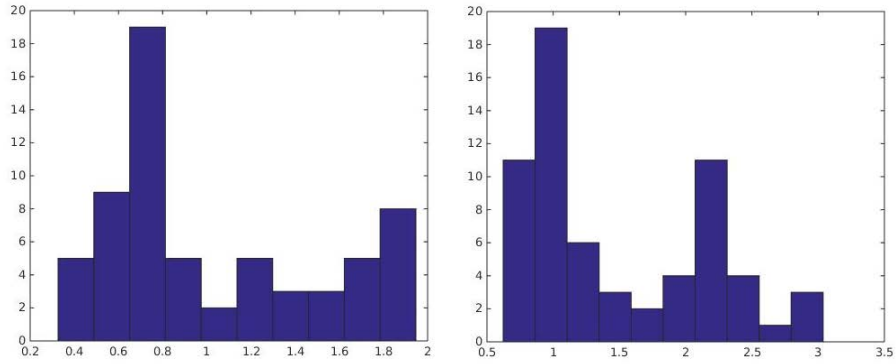



Figure 2.7 – Examples of mutual information distributions for $conv_{11}$ and $conv_{12}$ VGG layers.

content, $X$ and $Y$ would be dependent and $I$ would be equal to the input image entropy.

Mutual infomration $I(X,Y)$ is a random variable which is bimodal over the activations of each layer. This is illustrated in Figure 2.7 which shows the mutual information distributions for $conv_{11}$ and $conv_{12}$ VGG layers. Being $I(X,Y)$ bimodal, a thresholding like Otsu [42] allows to split the neurons of the j-th layer into two groups. These groups represents the amount of content which is preserved between the input and output of a neuron, either high or low. The function $\rho$ is given by the probability that a neuron is classified into the high content set and it is approximated as the percentage of times each activation is considered above the Otsu threshold $th$ for a set of training images:

$$\rho_{i,j} = \frac{|I(X_{i,j-1}, Y_{i,j}) > th_{i,j}|}{|V|} \tag{2.10}$$

for $|\cdot|$ denoting the cardinality of a set and $V$ the training set.

Figure 2.8 shows a collage of neuron activations for a given intra-operative image and layer ($conv2\_1$). Neuron activations have been sorted in ascendant order according to the contribution of appearence from left to right and from top to bottom. The top left corner represents the neuron which contributes the most to preserve the content of the input. Contrary, the bottom right neuron represents the neuron which contributes the least to preserve the content of the input.

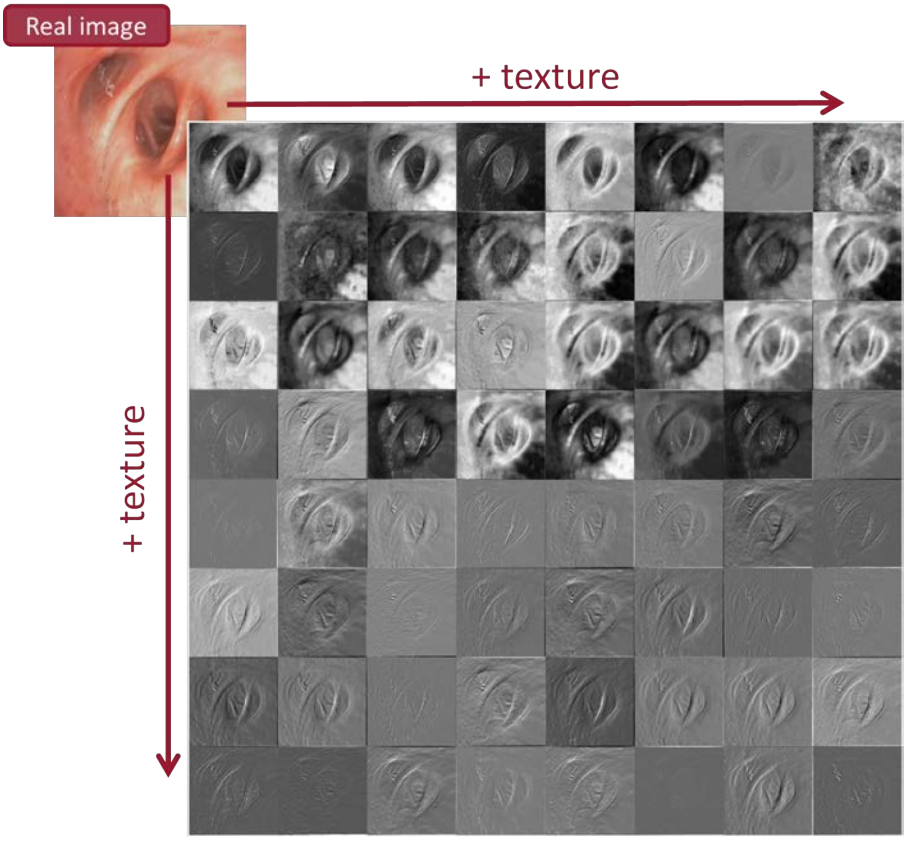
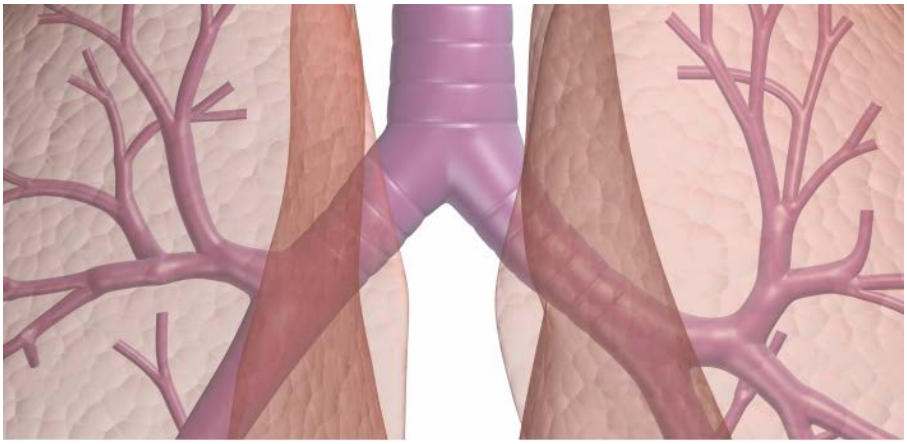



Figure 2.8 – Collection of activations for the $conv1\_2$ VGG layer. Activations are sorted in ascendant order according to the contribution of appearance from left to right and from top to bottom.

# 3 Navigation system based on bronchi anatomical landmarks



The proposed navigation system is based on the identification in both CT scans and video-bronchoscopy the identification of anatomical lansmarks describing airways geometry. To select the best anatomical landmark which allows us to characterize the airway anatomy in in-vivo bronchoscopic images is vital to understand the airway anatomy of the respiratory system. Airway anatomy can be subdivided into two different groups:

- The extrathoracic (superior) airway, which includes the supraglottic, glottic, and infraglottic regions
- The intrathoracic (inferior) airway, which includes the trachea, the mainstem bronchi, and multiple bronchial generations (which have as their main function the conduction of air to the alveolar surface)

Since doctors just need guiding systems in distal bronchial levels from now on we will focus on the intrathoracic inferior airway. The lungs are part of the intrathoracic inferior airway which starts at the trachea. The trachea is a cartilaginous and fibromuscular tube that extends from the inferior aspect of the cricoid cartilage (sixth cervical vertebra level) to the main carina (fifth thoracic vertebra level). The airways divide by dichotomous branching, with approximately 21 generations of branches from the trachea to the alveoli. The first bronchi to branch from the trachea are the right mainstem bronchi and the left mainstem bronchi. From there,

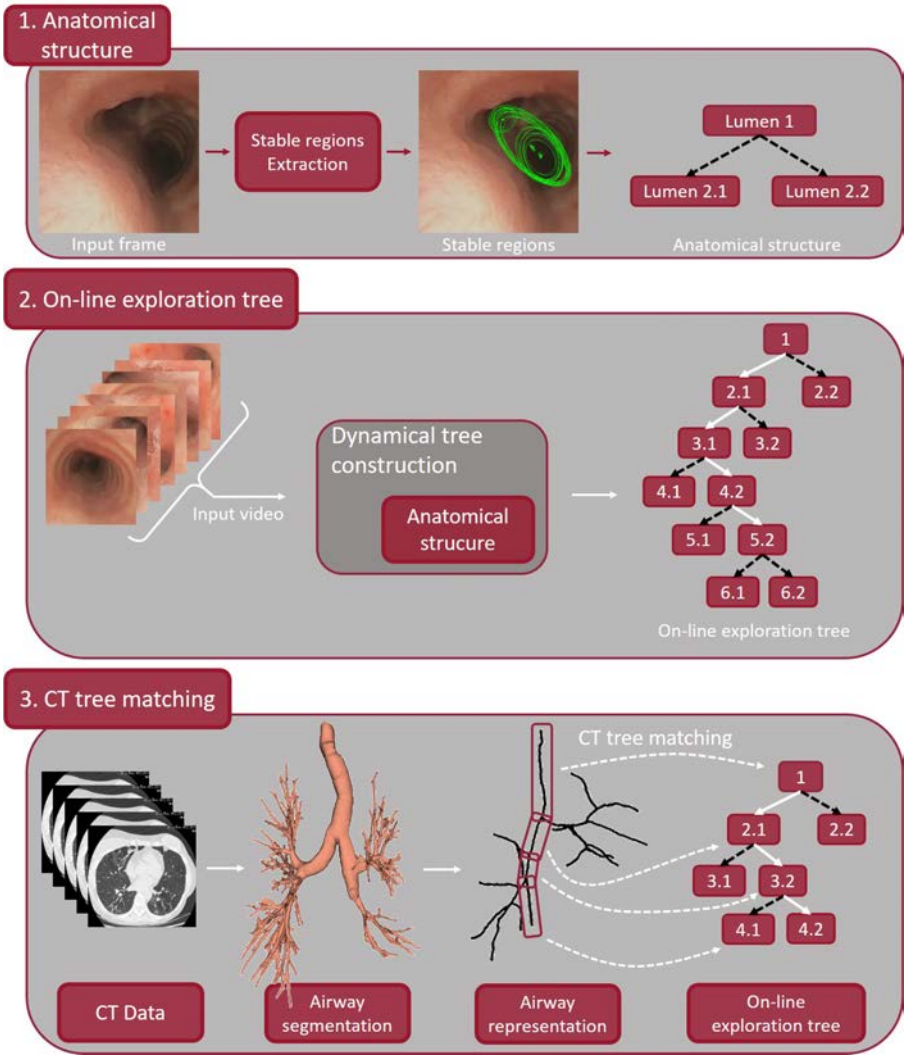



Figure 3.1 – Pipeline of the proposed navigation system.

they branch into narrower and narrower bronchi when advancing to deeper levels of the bronchial tree.

We claim that identification of bronchial tree key-points in, both, CT scans and

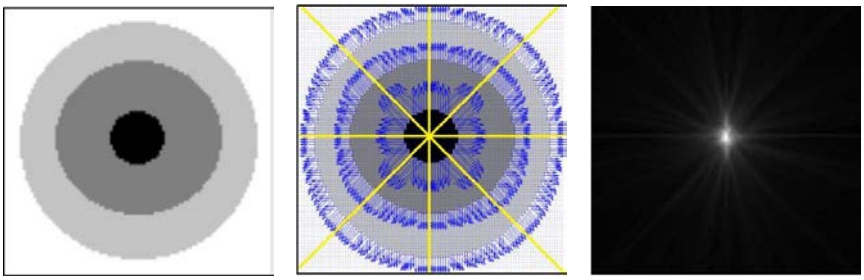
videobronchoscopy provides accurate matching between off-line planed path and the current endoscopic navigation. For navigation purposes, the best anatomical landmarks are lumen regions since they allow to detect bifurcations which helps to detect when a bifurcations has been crossed. We present in this chapter our strategy to detect stable luminal regions which allow accurate registration between CT-scans and videobronchoscopy procedures. Firstly, the branching airways anatomy is extracted for each frame. Secondly, the extracted anatomy is tracked across the sequence to obtain a dynamic representation of the interventional path across bronchial levels. The on-line interventional path is encoded as a binary tree which contains all the bifurcations seen during the whole procedure. Finally, the on-line interventional tree is dynamically matched using edit distance to the tree representing the whole airways anatomy and the planned path. This matching allows to keep a record of whether the planned path is being followed during the intervention and to indicate the next lumen to follow in order to reach the target lesion.

Figure 3.1 illustrates the whole pipeline of the proposed navigation systems. In the top section of Figure 3.1, a set of anatomical landmarks are detected in a videobronchoscopy frame. The set of landmarks is processed and the anatomical hierarchy is encoded in a tree structure. The bottom left section of Figure 3.1 shows how the anatomical hierarchy is tracked across the whole videobronchoscopy in order to build dynamically an on-line interventional tree. Finally, the bottom right section of Figure 3.1 describes how the online exploration tree is put in correspondence with tree representing the whole airways anatomy and the planned path.

This chapter is organized as follows. Section 3.1 explains the anatomical hierarchy extraction from videobroncoscopy frames. Section 3.3 describes the use of such anatomical hierarchy in order to create an on-line interventional tree representing the path and branching points seen during the intervention. Section 3.3 also develops the matching between the on-line interventional tree and the pre-planned path.

## 3.1 Extraction of anatomical structure in single videobronchoscopy frames

The first step to encode the anatomical structure of bronchoscopic images is to find image region candidates to enclose airways lumens. Extraction of lumen regions is based on likelihood maps [51] which indicate the probability of a point to be a lumen centre. In [51], such maps are computed by combining two operators



(a) Virtual Image (b) ∇ Virtual Image (c) Virtual raw DGA

Figure 3.2 – Graphical explanation of DGA algorithm: Original synthetic image (a); Corresponding gradient vectors superimposed to the image (b) and resulting DGA accumulation map (c).

that characterize lumen regions: Directed gradient accumulation (DGA) and Dark region identification (DRI).

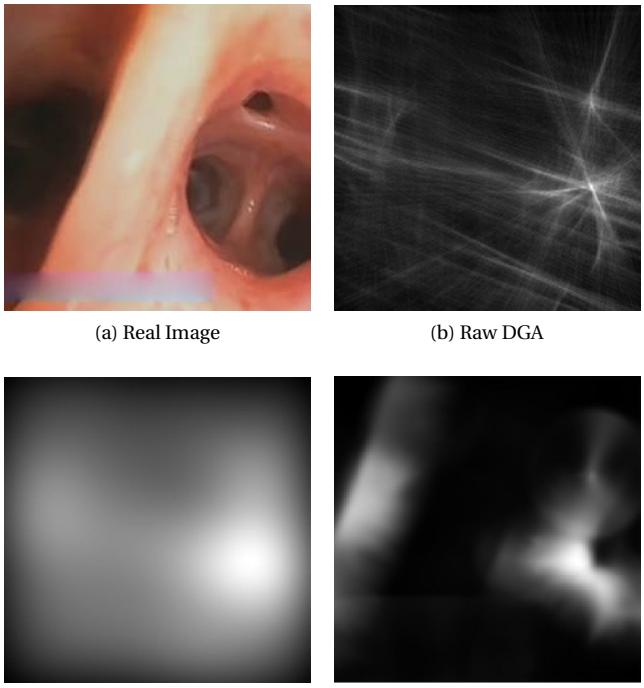
DGA and DRI describe luminal geometry and appearance. They define a 2-dimensional feature space that characterize several elements of the endoluminal scene. Pixels with a high value of DRI (dark regions) and DGA (high accumulation of gradients) are pixels belonging to lumen regions. High DGA and low DRI are pixels which belong to specular highlights and other bright protruding. Finally, folds and tracheal rings have low DGA and DRI. Therefore, a classification of the (DRI,DGA) feature space into these three classes characterize the luminal region. In particular, the local maxima of the likelihood maps corresponds to the lumen center. In spite of the good results that the (DGA,DRI) space exhibits for the detection of lumen centers and region for the main bronchi [52], as deeper levels are reached the main hypothesis supporting DGA and DRI is not fulfilled.

DGA maps are based on the idea that gradients direction meet in the center of the center of the lumen. In [51], DGA was calculated as the accumulation of every gradient directed line that cross each pixel:

$$DGA(x,y) := \sum_{\exists\ (x,y)=(x_0,y_0)+\lambda\nabla I(x_0,y_0)} \|\nabla I(x_0,y_0)\|_2 \tag{3.1}$$

where $(x_0,y_0)+\lambda\nabla I(x_0,y_0)$ corresponds to the parametric formulation of a line through $(x_0,y_0)$ oriented across its image gradient $\nabla I(x_0,y_0)$, $\lambda$ is the free parameter of the gradient line equation and $\|\cdot\|_2$ denotes the $L_2$ quadratic norm.

The synthetic image in Figure 3.2 (a) illustrates how DGA works. In this example



(a) Real Image (b) Raw DGA

(c) Smoothed DGA (d) Improved DGA

Figure 3.3 – Example of DGA on an intra-operative image: Original intra-operative image (a); Resulting DGA accumulation map (b); Resulting smoothed DGA accumulation map and resulting improved DGA accumulation map (d).

all gradient vectors are directed from the center of the image (darkest part) to the brightest external part and, thus, DGA maximum response corresponds to the center of the image (lumen). Figure 3.2 (c) shows the resulting DGA accumulation map extracted from 3.2 (a).

In case of perfect circular shapes like trachea, carina and main bronchi, DGA accumulation has a clear maximum in the center of the circle. However, this does not hold fro deeper bronchial levels because in case of lateral or distal bronchi the shape becomes more elliptical or some parts are missing. Since a gradient only affects to pixels within the same direction of the gradient, DGA likelihood maps results in a mess of straight lines with a lot of local maxima. The authors in [51] proposed to smooth DGA using a Gaussian filter before applying it to detect luminal

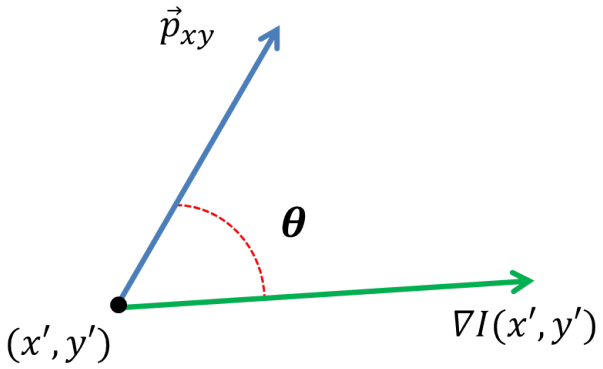



Figure 3.4 – Scheme of DGA calculation for a pair of points $(x, y)$ and $(x', y')$. It is shown how the pixel $(x', y')$ is affected by a point $(x, y)$ modulated by the angle $\theta$ between the gradient and the vector given by $\vec{p_{xy}}$ and $\nabla I(x', y')$

structures in intra-operative videos. The intra-operative image in Figure 3.3 (a) illustrates how DGA is computed and post-processed in case of distal navigation. While Figure 3.2 (b) shows the raw DGA likelihood map, Figure 3.2 (c) shows the Gaussian smoothing of raw DGA accumulation map extracted from Figure 3.2 (a).

We propose an improved DGA likelihood map which solves the directionality of the original DGA. Instead of accumulating pixels within the gradient direction, all pixels are affected by each gradient modulated by the angle $\theta(x, y)$ between the gradient of the image and the vector given by the current pixel position, $(x, y)$, where DGA is computed and the position, $(x', y')$, of the pixel where the gradient to be accumulated is computed. If we note $\vec{p_{xy}} = (x, y) - (x', y')$ such vector and $\nabla I(x', y')$ the gradient at $(x', y')$, then the improved DGA at $(x, y)$ is given by:

$$DGA(x, y) := \sum_{\forall (x', y') \in I(x', y')} \left( \frac{< \nabla I(x', y') \cdot \vec{p_{xy}} >}{\|\nabla I(x', y')\|_2 \|\vec{p_{xy}}\|_2} \right)^2 \|\nabla I(x', y')\|_2 =$$
$$= \sum_{\forall (x', y') \in Image} (\|\cos\theta(x, y)\|)^2 \|\nabla I(x', y')\|_2 \tag{3.2}$$

for $< \cdot, \cdot >$ denoting the scalar product and $\|\cdot\|_2$ the quadratic norm.

On the other hand, in [51], a single DRI map was computed convolving images, $I = I(x, y)$, with a Gaussian isotropic kernel $g_\sigma$ with the scale $\sigma$ related to the size of

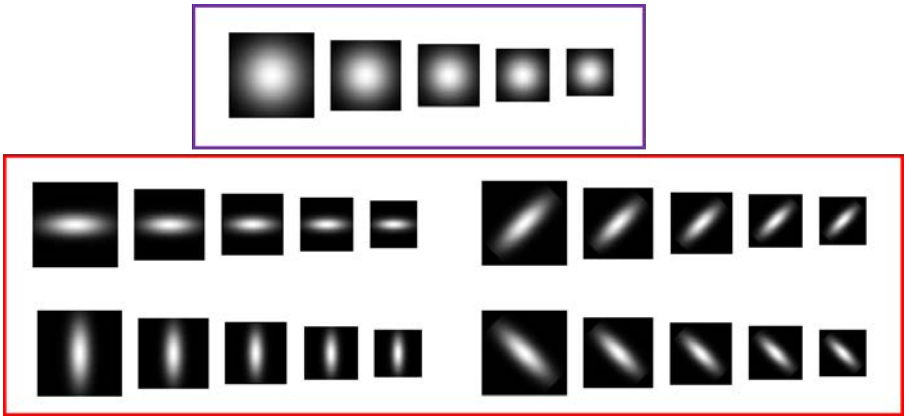


Figure 3.5 – Bank of anisotropic oriented Gaussian filters for the modeling of elliptic luminal regions.

the lumen:

$$DRI(x,y) := g_\sigma * I = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2}{2\sigma^2}+\frac{y^2}{2\sigma^2}\right)} * I(x,y) \tag{3.3}$$

for $*$ the convolution operator. The response to DRI enhances dark values and, thus, luminal areas.

The formulation of DRI uses a single isotropic Gaussian kernel under the assumption that in central navigation the appearence of the projected airways lumens are dark circular areas. The use of one single Gaussian kernel limits the extraction of lumen regions to circular regions of the same size which is not fully sensible in interventional videos. To model non-circular lateral bronchi and small distal levels, we compute several likelihood maps using a bank of anisotropic Gaussian filters as the ones shown in Fig 3.5 with different orientations, $\theta$, and scales $\sigma = (\sigma_x, \sigma_y)$. The improved DRI is given by

$$DRI_{\Theta\theta}(x,y) := \frac{g_{\Theta\theta}}{\|g_{\Theta\theta}\|_2} * I = \frac{1}{\|g_{\Theta\theta}\|_2} \cdot \frac{1}{2\pi\sigma_x\sigma_y} \cdot e^{-\left(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right)} * I(x,y) \tag{3.4}$$

with $g_{\sigma\theta}$ denoting an anisotropic filter of scale $\Theta = (\sigma_x\sigma_y)$ orientated along the angle $\theta$.

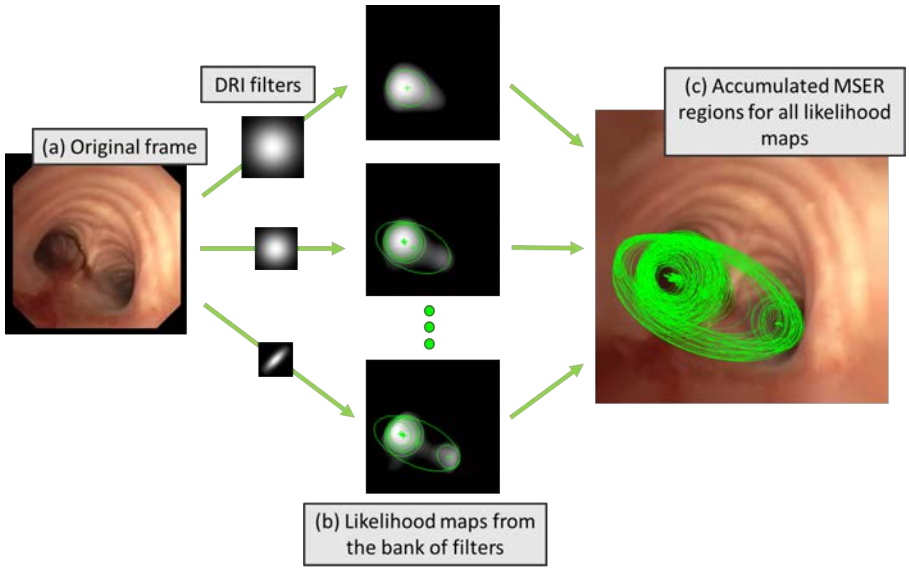



Figure 3.6 – Scheme of the extraction of the anatomical hierarchy from a set of MSER regions.

The different $\sigma_x$ are given by the size of the image divided by 16, 18, 20, 22 and 24, while $\sigma_y$ takes the values of $\sigma_x$ and $\sigma_x/3$. In addition, each of the anisotropic filter is rotated by $\sigma = \{0°, 45°, 90°$ and $135°\}$ . Gaussian filters have been normalized by their L2 norm ($\|\cdot\|_2$) to obtain more uniform responses comparable across different scales and degrees of anisotropy [32]. The convolution of videobronchoscopy frames with each one of the filters provides a description of either circular or non-circular luminal areas that we note by $DRI_{\Theta\theta}$

For each of the anisotropic oriented filters, we compute a different likelihood map given by the classification in the space $(DRI_{\Theta\theta}, DGA)$. Figure 3.6 shows the likelihood maps computed by convolving the left-hand side frame, labeled (a), with the bank of Gaussian filters shown in the left hand side of the image frame. The corresponding likelihood maps obtained for each of the filters are shown in the central central column of Figure 3.6 labeled (b).

We observe that by using more and smaller filters likelihood maps are more sensitive to changes of illumination, highlights and even non-luminal structures. This results in liklihood maps prone to have non-prominent local maxima with unstable pattern across video frames. Hence, in order to improve the detection of

lumen centers, it is mandatory to ignore small local maxima and endow detections with temporal consistency. To suppress outlying small local maxima, likelihood maximal regions are computed using maximally stable extremal regions (MSER) [35]. Maximally stable extremal regions is a method for detection of blobs in images like the local maxima that characterize lumen regions in our likelihood maps. The concept of MSER can be explained by thresholding. All the pixels below a given threshold are 'black' and all those above or equal are 'white'. Given an image $I$, if a sequence of thresholded result images $I_{th}$ is generated where each image $I_{th}$ corresponds to an increasing threshold $th$, first a white image would be seen, then 'black' spots corresponding to local intensity minima will appear then grow larger. These 'black' spots will eventually merge, until the whole image is black. The set of all connected components in the sequence is the set of all extremal regions.

The collection of MSER regions for all likelihood maps, as shown in the most right image of Figure 3.6 labeled (c), is a set of elliptical regions following a hierarchy of inclusions that correspond to the projection of airways from different bronchial levels. We codify the structure of such hierarchy as a binary tree that represents the anatomical structure of the airways projected in the image. Since each MSER region should be represented as a node of the tree, we iteratively construct the tree by keeping a list of root and children regions. First, MSER regions are sorted regarding their area in ascending order and the first region of the sorted list is added to the root node list and marked as current root. Then, we iteratively consider the next region in the sorted list, add it to the root list and update the children list according to whether the region contains any of the current roots. All roots contained in it are added in the tree structure as child of the node we are examining and are removed from the root list.

**Algorithm 1** Tree structure pruning

```
1: procedure PRUNING(tree)          ▷ tree is sorted by area in descending order
2:     for i=length(tree):-1:1 do
3:         parent ← tree(i).Parent
4:         if length(tree(parent).Children) ≤ 1 then ▷ the parent has only one child
5:             tree ← remove(tree,i)        ▷ Node i does not belong to a bifurcation
6:         end if
7:     end for
8: end procedure
```

Ideally, we would like that each of the bronchial branches that represents a lumen region would correspond to a tree node. This is not the case due to the multiple MSER regions coming from different likelihood maps that lie on a bronchial

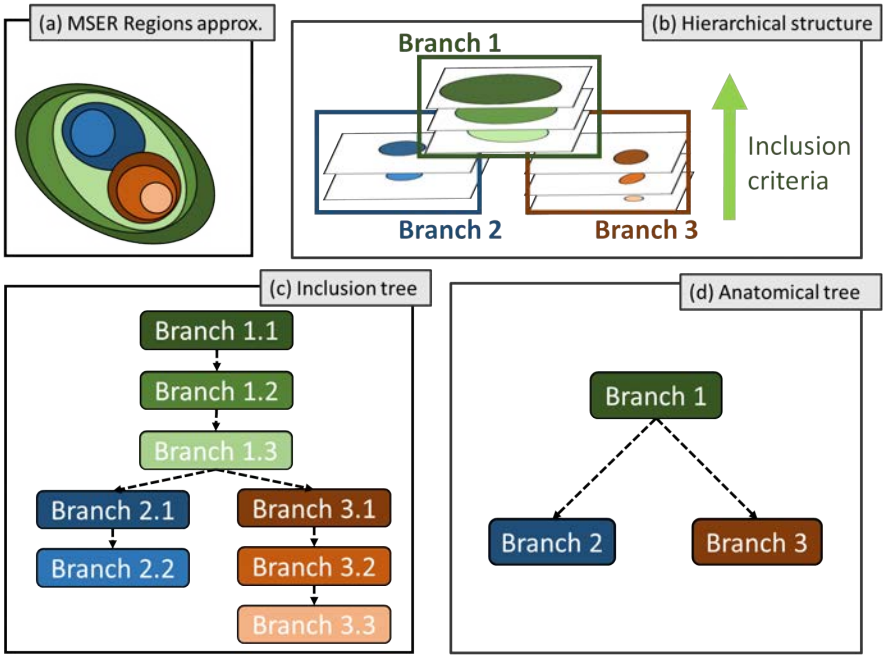



Figure 3.7 – Tree structure from MSER regions.

lumen. This introduces intermediate nodes in the represented MSER hierarchy that do not actually correspond to a new bronchial level (see Figure 3.7 (c)). In order to remove them and keep a tree representing the main bronchial anatomy, the inclusion tree is pruned according to each node adjacency order. Those nodes having a single child are iteratively removed from the tree, starting by the leafs. The pseudo-code can be found in Algorithm 1.

Figure 3.7 sketches the main steps in the extraction of the anatomical hierarchy from a set of MSER region into a binary tree structures. To better illustrate the tree creation we show a synthetic image (Figure 3.7 (a)) that simplifies the image in Figure 3.6 (c). The ellipses sorted by area and inclusion shown in Figure 3.7 (b) produce the tree in Figure 3.7 (c). In this tree, nodes are labeled according to its branch and its order inside it. Branches of the inclusion tree consist of a set of nodes which are pruned to obtained one representative ellipse per branch (ellipse with the highest area). Thus, the final anatomical tree is illustrated in Figure 3.7 (d) which represents the anatomical structure of airways of the image in Figure 3.7 (a).

## 3.2 Extraction of the on-line exploration tree

The anatomical structure extracted from single frames is key because it can be used for creating an on-line exploration tree. The on-line exploration tree encodes all the bifurcations (called bronchial path) seen during the intervention and the position inside the airways at any interventional time. The on-line exploration tree is dynamically constructed from the hierarchy of MSER ellipses extracted from each video bronchoscopy frame as follows.

The nodes of the on-line exploration tree represent the lumnal regions explored during the intervention while edges encode the hierarchy of branching levels. For each node we store luminal anatomical information given by the parameter of the MSER ellipse representing the luminal region such as the center position, $(x, y)$, mayor and minor axis, $(a, b)$, and its orientation, $\theta$. The anatomical information also stores the image quadrant, $Q$, the ellipse is in. Image quadrants are stored in order to provide guidance instructions indicating the bronchi to follow at each bronchial level. Finally, we also store a flag ($active$) indicating whether the node corresponds to a lumen that is currently being observed in the video frame or it corresponds to a bifurcation seen in previous frames. A given node is activated when it has been matched for $N_{Fr}$ frames and it is deactivated when it has been lost for $N_{Fr}$ frames. This, we call this extra information, temporal information. Temporal information allows as to infer the position of the bronchoscope inside the airways in terms of the current bronchial level the scope is placed at. The sub-tree of nodes whose variable $active$ is set to true is called active nodes sub-tree. The active nodes sub-tree represents the anatomy currently observed in the interventional video and its root gives the level inside the whole exploration tree.

Figure 3.8 represents the on-line exploration tree at different non-consecutive frames. The right upper part represents the MSER anatomical hierarchy at a certain frame i. In frame i, there is only one active lumen represented by Node 1 $(x_1, y_1, a_1, b_1, \theta_1, Q_1, active)$. The left part of the image represent the on-line exploration tree at frame i + 50. In frame i + 50, the active nodes sub-tree consist of three nodes: Node 2 $(x_2, y_2, a_2, b_2, \theta_2, Q_2, active)$, Node 4 $(x_4, y_4, a_4, b_4, \theta_4, Q_4, active)$ and Node 5 $(x_5, y_5, a_5, b_5, \theta_5, Q_5, active)$. Since Node 1 has not been detected for $N_{Fr}$ frames, it is inactive and it does not belong to the active nodes sub-tree. Thus, the on-line exploration tree is updated and constructed by continuously matching MSER ellipse hierarchy to the active nodes sub-tree. An important remark is that the active nodes sub-tree are surrounded by a blue area in which the nodes marked with a double line indicates the bronchial level the bronchoscope is in.

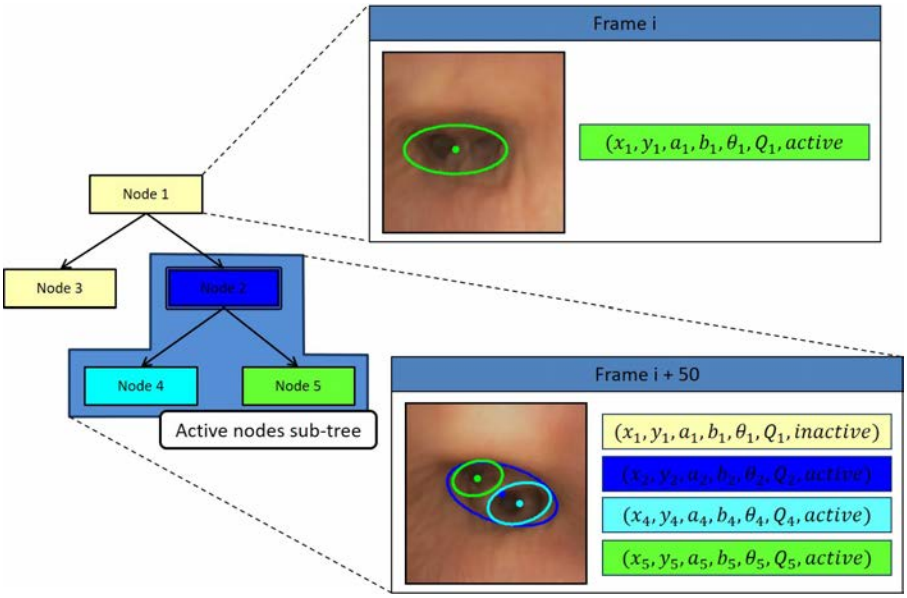


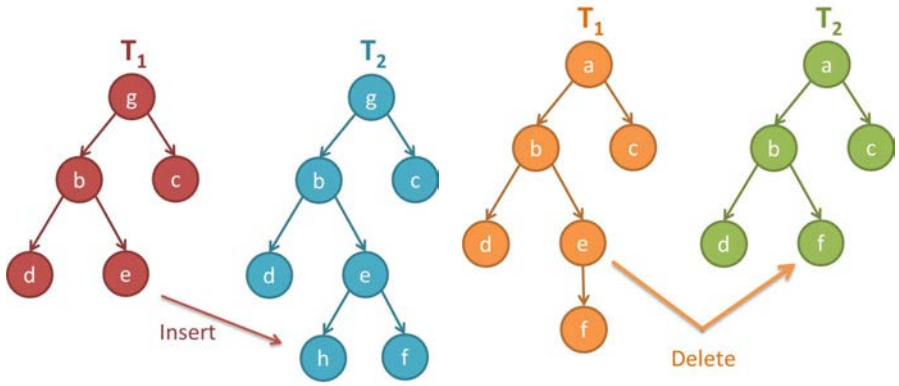

Figure 3.8 – On-line exploration tree schema at different frames. The left part of the image represent the on-line exploration tree at frame i + 50. The right upper part represents the MSER anatomical hierarchy at frame i. The right lower part represents the MSER anatomical hierarchy at frame i.

There are three aspects of the exploration tree that must be updated during intervention time:

1. **Anatomical Information.** The ellipse parameters (center $(x_0, y_0)$, axis $(a, b)$ and orientation $\theta$) and its position (quadrant $Q$) of the active nodes is updated according to the current image MSER hierarchy.

2. **Temporal Information.** The activation and deactivation of on-line tree nodes is updated by tracking the MSER hierarchy across the interventional video.

3. **Creation of New Nodes.** New nodes representing deeper bronchial levels not observed in previous frames are created according to changes in MSER hierarchy.




(a) Insertion of node $h$ and $f$ (b) Deletion of node $e$

Figure 3.9 – Available edit operations to transform $T_1$ in $T_2$. Insert operation to transform $T_1$ in $T_2$ (a); Delete operation to transform $T_1$ in $T_2$ (b)

### 3.2.1 Matching of MSER hierarchy and active nodes sub-tree

In order to update the on-line exploration tree, the active nodes sub-tree is matched to the MSER hierarchy across the interventional video. To do so, we propose a modification of the tree edit distance for graph transformation and comparison [4, 44].

The tree edit distance computes the optimal set of operations that transforms a tree $T_1$ into another tree $T_2$. Three edit operations to transform one tree into another can be applied:

- **Insert** a node between an existing node and a subsequence of consecutive children of this node. In our problem, the insertion operation is applied when new deeper lumens appear in the intraoperative video. Figure 3.9 (a) sketches the transformation of $T_1$ into $T_2$ obtained by insertion of nodes $h$ and $f$ as children of node $e$.
- **Delete** a node and connect its children to its parent maintaining the order. In our case this corresponds to a region which could not be detected in one frame because the bronchoscope is in a different bronchial level, a sudden change in the illumination, a collision to the bronchial walls,... Figure 3.9 (b) sketches the transformation of $T_1$ into $T_2$ obtained by deletion of node $e$.
- **Replace** the label of a node. This operation does not apply in our case since

we cannot modified the detected regions.

In order to define a criterion for the selection of the optimal transformation, each edit operation is assigned a cost which is used to computed the total cost of the transformation and define a distance between the trees. Therefore, ff we note $N_i$ the set of $m_i$ nodes of $T_i$,

$$N_i = \{n_i^j \in T_i\}_{j=1}^{m_i} \tag{3.5}$$

and $sim(n_1^{j_1}, n_2^{j_2})$ denotes the similarity between nodes $n_1^{j_1} \in N_1$ and $n_2^{j_2} \in N_2$, we have that:

- **Node deletion.** A node $n_1^{j_1}$ is removed if $sim(n_1^{j_1}, n_2^{j_2}) < th$ for all $n_i^{j_2} \in N_2$ with $th$ a threshold on minimum similarity.
- **Node insertion.** Those nodes in $T_2$ that have not been matched to any node in $T_1$ are inserted according to $T_2$ tree adjacency.

In our case, the similarity between nodes is given by the overlap between the ellipses they represent. This overlap is computed using the following equation of the interior ellipse region that represents a node $n$:

$$\frac{(x\cos(-\theta) - y\sin(-\theta)) - x_0}{a} + \frac{(y\cos(-\theta) - x\sin(-\theta)) - y_0}{b} \leq 1 \tag{3.6}$$

for $n = (x_0, y_0, a, b, \theta)$ the parameters of the ellipse.

We note the left hand side of the above inequality by $En(x, y)$. Thus, the overlap between two nodes $n_1^{j1} = (x_1, y_1, a_1, b_1, \theta_1)$ and $n_2^{j2} = (x_2, y_2, a_2, b_2, \theta_2)$ is given by the following score:

$$sim(n_1^{j1}, n_2^{j2}) := En_1(x, y) \cap En_2(x, y) = \frac{2 \cdot |(En_1 < 1) \cap (En_2 < 1)|}{|En_1 < 1| + |En_2 < 1|} \tag{3.7}$$

where $|\cdot|$ indicates the area of an ellipse region.

We note that since Equation 3.6 defines an image mask of the interior of the ellipse, Equation 3.7 counts the number of pixels within both ellipses over their total area and, thus, its percentage of overlap. This is known as Volumetric similarity ($VS$) or Volumetric Overlap Error ($VOE$) [56] and it is a quantity commonly used to validate shape similarity in the context of medical imaging segmentation [11, 60]. We observe that $overlap = 1$ if there is full overlap between ellipses and overlap = 0 if there is no overlap between them.

Let $del(\cdot)$ represents the cost of deleting a node from the tree and $add(\cdot)$ represents the cost of inserting a new node to the tree. Given $M$ a matching between

nodes of $T_1$ and $T_2$ using the similarity metric and $O_2$ the sets of nodes in $N_1$ and $N_2$ not appearing in $M$, the tree edit distance between two trees $T_1$ and $T_2$ can be stated as

$$d(T_1, T_2) = \underset{M}{\operatorname{argmin}} \left\{ \sum_{n_1 \in O_1} del(n_1) + \sum_{n_1 \in O_2} add(n_2) \right\} \tag{3.8}$$

Both cost functions are set to 1 if a node is removed or added.

Insert and delete operations of the tree edit distance gives us information about anatomical changes between frames. Let´s assume the optimal tree edit distance sequence between two consecutive frames contains the deletion of one node. The information that we can extract is that one lumen which was present in frame i,

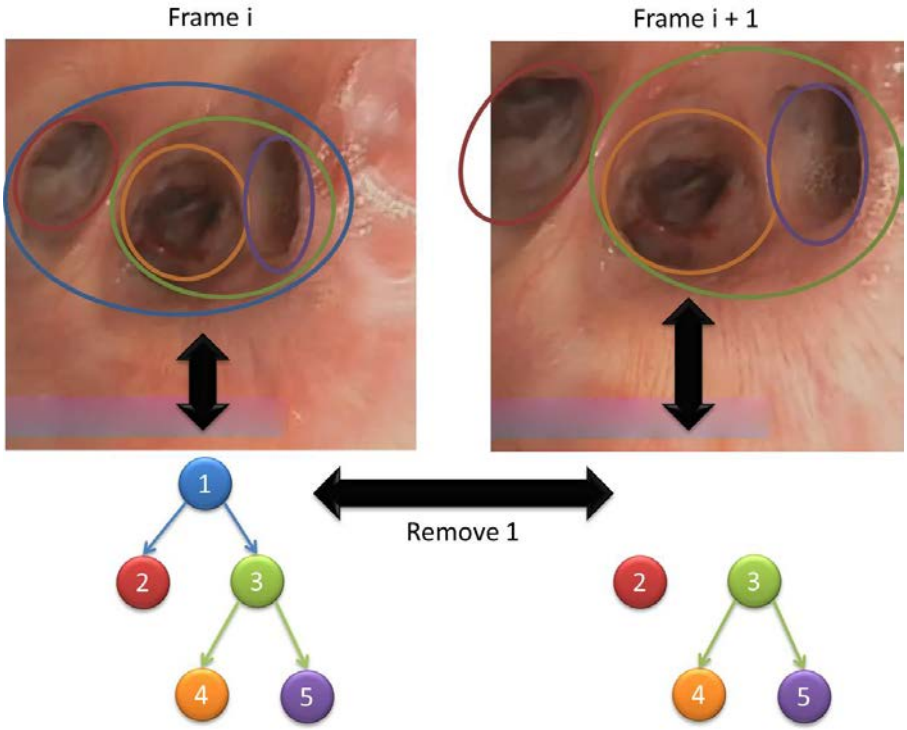



Figure 3.10 – Example of delete edit operation between anatomical trees of two consecutive frames.

it has not been detected in frame i+1. Figure 3.10 illustrates this case. Figure 3.10 simulates that we are approaching to a bifurcation so the main bronchi region is lost. Therefore, in Figure 3.10, we can see that to transform the tree at Frame i into the tree at Frame i+1, node number 1 needs to be removed.

The opposite way, the optimal tree edit distance sequence between two consecutive frames could contain the insertion of one node. This means that a new lumen has appeared in the new frame which represents a region of a new bronchial

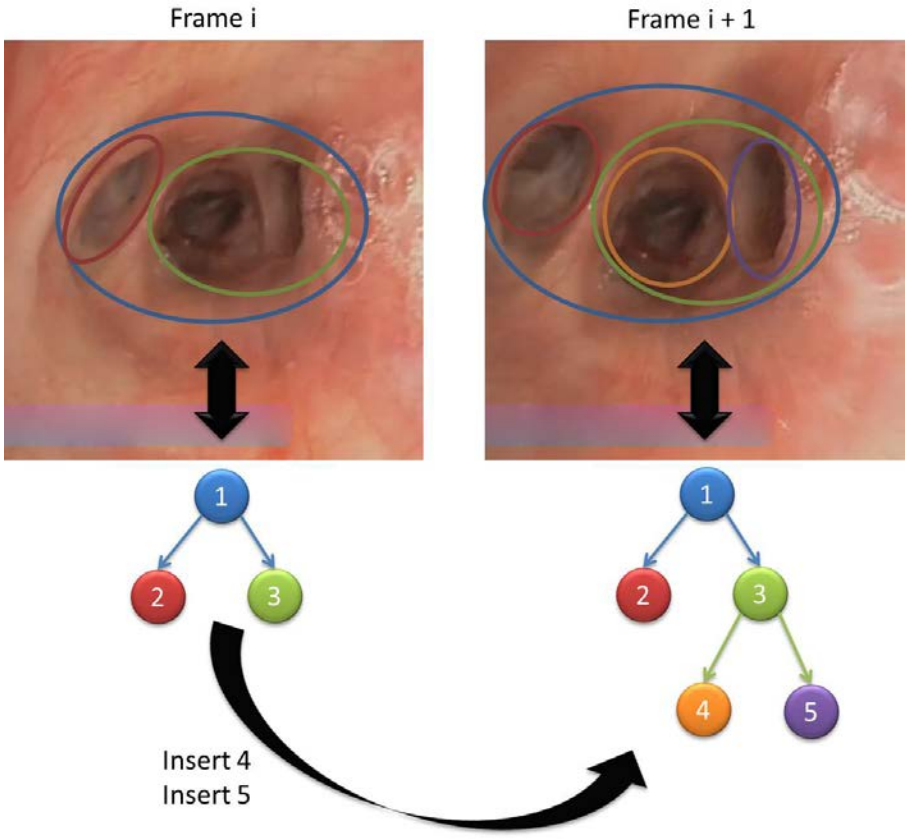



Figure 3.11 – Example of insert edit operation between anatomical trees of two consecutive frames.

level lumen. Figure 3.11 illustrates this example. In Figure 3.11, we can see that to transform the tree at Frame i into the tree at Frame i+1, nodes number 4 and 5 need to be added (node 4 and 5 represents new detected luminal tubes in Frame i+1).

### 3.2.2 On-line exploration tree updating

The anatomical edit distance defined in Section 3.2.1 is used to update the on-line tree from the MSER hierarchy extracted across the interventional videos, as follows:

- **Anatomical information.** The anatomical information of active nodes sub-tree is updated using the center and position of MSER hierarchy ellipses. Active nodes anatomical information is set to the values of the MSER hierarchy ellipses they have been matched to according to the anatomical edit distance.

- **Positional information.** The image quadrant the ellipse is in the MSER hierarchy tree is also updated.

- **Temporal information.** The temporal information is updated according to the number of frames the node has been matched to MSER hierarchy trees following the anatomical edit distance. A node in the on-line tree matched to a MSER hierarchy ellipse is activated if the node has been matched to a node in the MSER hierarchy for a given number of frame. The same way, a node is deactivated if it is not matched for $N_{Fr}$ frames.

Figure 3.12 illustrates the update of the on-line exploration tree between consecutive frames. The top left and right hand-side of Figure 3.12 shows the active nodes sub-tree and the MSER hierarchy at frame i and i+1 respectively. The anatomical information of the associated regions of such trees is shown in the middle section of Figure 3.12. In such section, dotted lines denote the matches between the active nodes sub-tree and the MSER hierarchy tree obtained according to the similarity metric, $sim$. Then, the bottom section of Figure 3.12 describes how nodes in the on-line exploration tree are updated at frame i + 1 using the anatomical info of the MSER hierarchy tree following the matches found.

A very important step of the algorithm is the creation of new nodes representing deeper bronchial levels not observed in previous frames. When the bronchoscope approaches a bifurcation, two new children nodes appear in the MSER hierarchy. These children nodes are also inserted in the active nodes sub-tree as children of the node matched to its parent in the MSER hierarchy. The anatomical infromation of the new created nodes is given by the MSER ellipse parameters and the $active$ lable is set to false. The $active$ label will set to true only when the new nodes have been matched for $N_{Fr}$ frames.

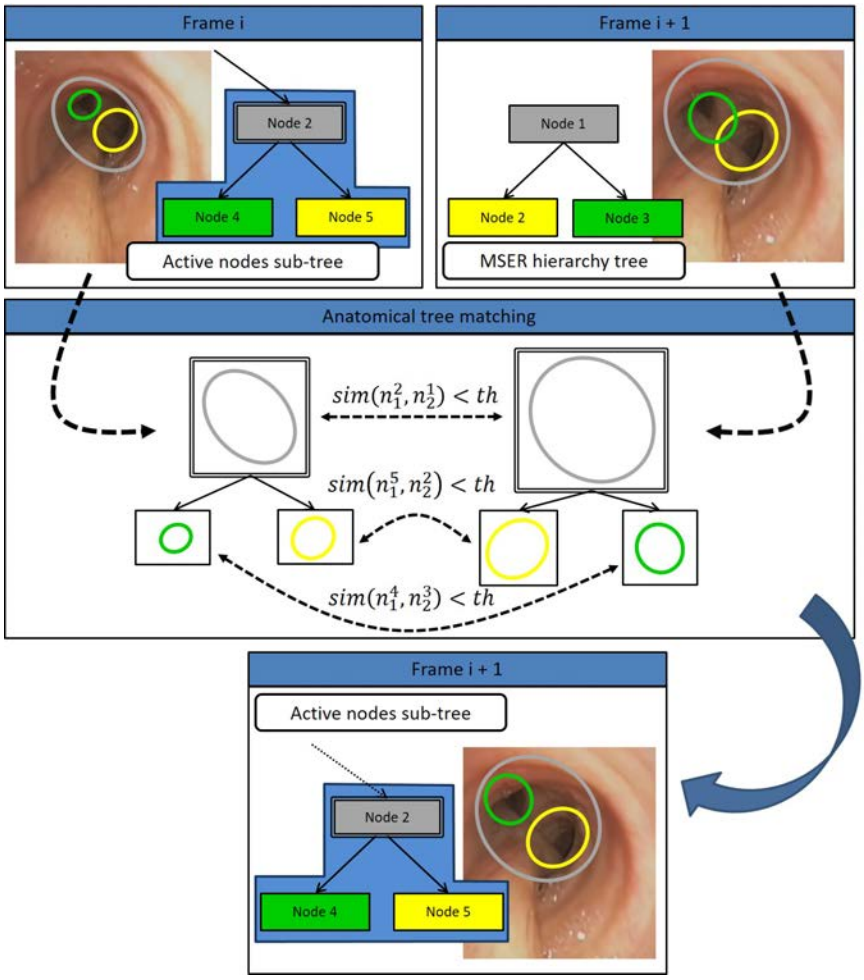



Figure 3.12 – Example of the on-line exploration tree update in consecutive frames. The left section of the figure contains the active nodes and the MSER hierarchy tree extracted at frame i. The right section shows how the nodes of the active nodes sub-tree are updated following the matching and the anatomical info of the MSER hierarchy tree.

Figure 3.13 illustrate the creation and activation of nodes in the on-line exploration tree of a sequence of frames approaching a bronchial bifurcation. The

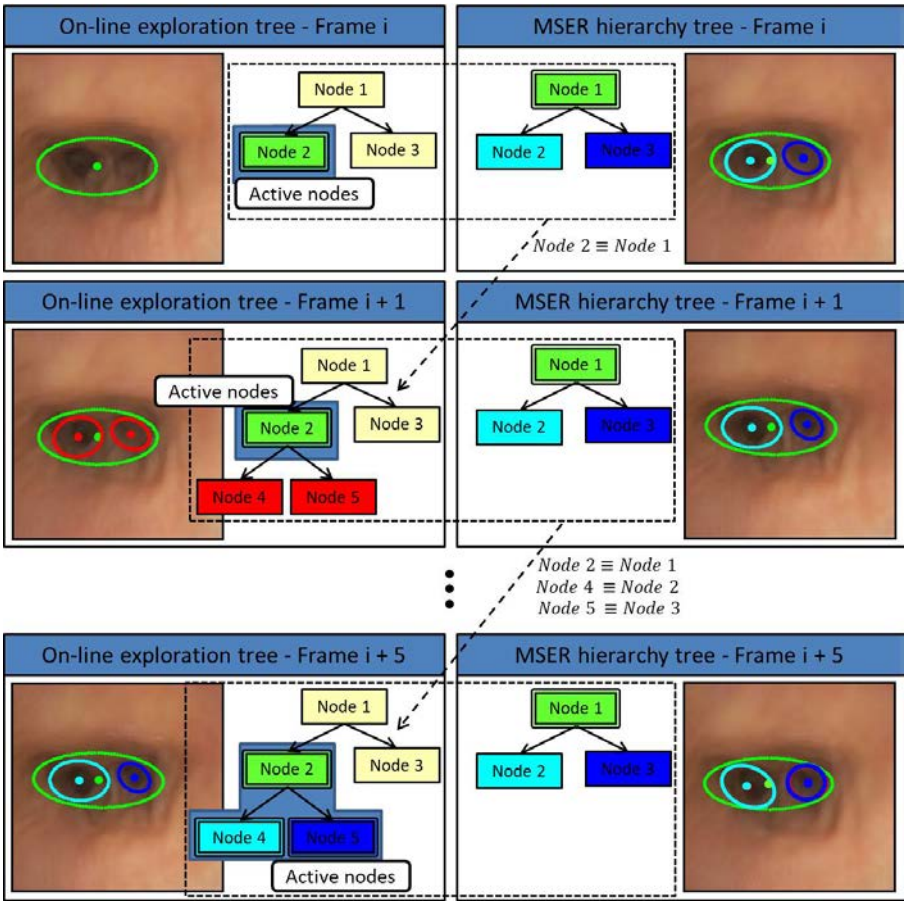



Figure 3.13 – Creation and activation of new nodes in the on-line exploration tree.

sequence starts with one single active node representing we are in the intermediate area of the bronchial tube. In the MSER hierarchy tree at Frame i two new nodes appear within the already detected region in the on-line exploration tree of Frame i. Hence, the two new nodes are added in the online-exploration tree in Frame i+1 as childs of the matched node. After the new nodes have been matched during $N_{Fr}$ frames, they are marked in the on-line exploration tree as active nodes.

The temporal information stored in every node in the tree allows us to detect the anatomy currently observed in the interventional video. The set of active nodes

of the interventional videos tells us the position of the bronchoscope within the bronchi. In particular, the scope of the current bronchial level is the level inside the whole on-line tree of the active nodes sub-tree root. The sequence of active nodes sub-tree roots also allows to detect the time the scope enters into a deeper bronchial level. When the scope approaches a new level, the active sub-tree root representing the incoming bronchi is deactivated and its two children become two roots. The moment the scope enters into one of the children bronchi, one of the roots is deactivated. Therefore, our criteria, for detecting that the scope has traversed a new level, is an increase in the active nodes sub-tree roots, followed by a decrease.

This is illustrated in Figure 3.14. The figure starts in an i-nth frame in which a single lumen is detected in the second bronchial level. This lumen is represented in the on-line exploration tree at Frame i as an active node. The MSER hierarchy at frame i contains a lumen with two smaller inside of it. After tracking them for $N_{Fr}$ frames, those lumens are also added to the active nodes sub-tree in Frame i + 5. MSER hierarchy tree at Frame i+10 detects two lumens without a common root. Therefore, we can assume we are getting closer to a bifurcation. After $N_{Fr}$ frames, this configuration is transferred to the on-line exploration tree at Frame i+15. Since the bronchoscope keeps moving forward, at some point one of the root is lost. This is illustrated in the MSER hierarchy tree at Frame i+20. In the next frame, the on-line exploration tree is updated according to that configuration. The change of configuration from two roots in the active nodes sub-tree to a single root sub-tree means the bronchoscope has crossed the bifurcation.

## 3.3 Matching between CT-planning and on-line exploration tree

At the planning stage, CT imaging is used to create a planning of the intervention video based on the structure of the airways branching anatomy. Figure 3.15 summarizes our strategy for the codification of airways anatomy from CT-scans. The upper part of Figure 3.15 illustrates the extraction of airways volume from CT scans. First, CT scans are segmented to provide a binary volume of airways. Then this volume is skeletonized to obtain airways centerline which represents the essential branching anatomy. Airways skeleton branching structure is encoded in a binary tree as shown in the bottom of Figure 3.15 that contain information about the disposition of airways in central navigation explorations.

Airways volumes geometrically follow a tubular structure. The centerlines of such structures correspond to the skeleton of segmented volumes, and they allow

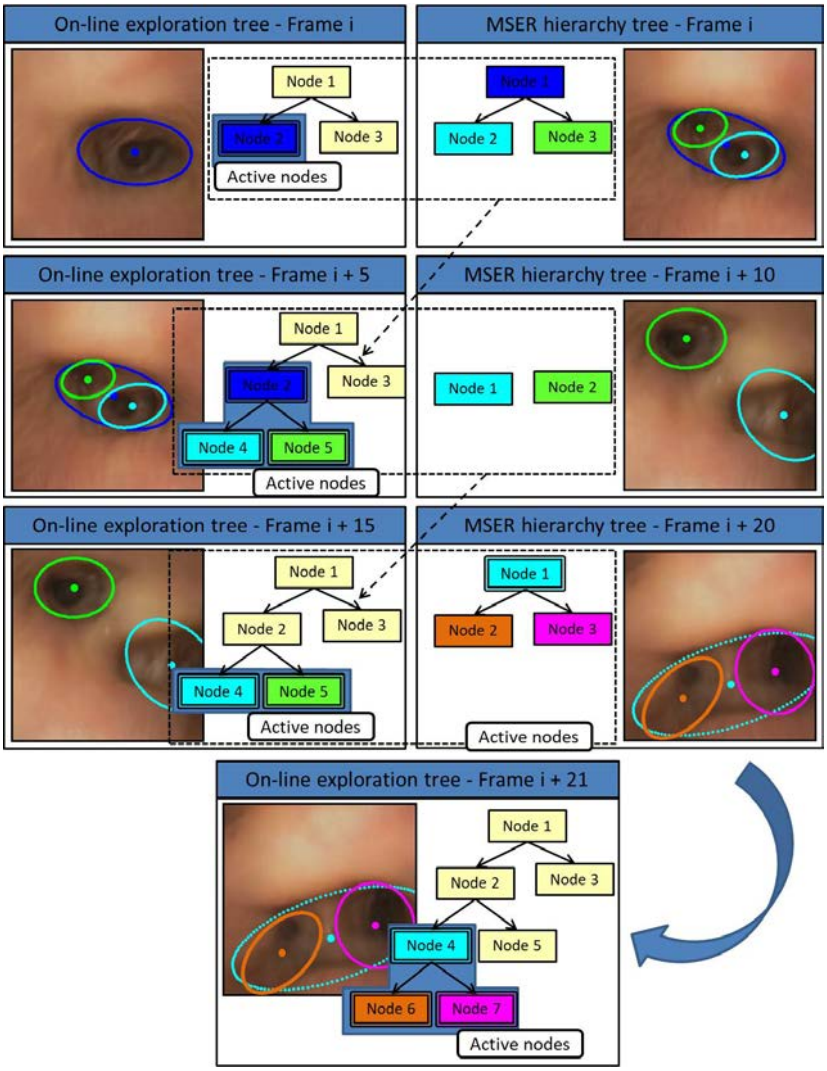



Figure 3.14 – Videobronchoscopy sequence where new nodes are added to the on-line intervention tree, the bronchoscope cross a bifurcation. This figures also proves we can extrapolate the location of the bronchoscope from the on-line exploration tree.

the construction of tree-based structures. These tree-based structures encode branching geometry using nodes and edges. The nodes of the graph correspond to the skeleton branching points and its edges represent branch connectivity. For each node in the CT-tree we store its 3D position $(x, y, z)$ in the CT scan and the image quadrant they would be in, if they were projected during an exploration in central navigation.

In order to compute bronchi quadrant, we use the same approach as [48] in order to provide an edge labeling consistent with intra-operative visual information. In [48], the path is created across the segmented volume and projects the segmented 3D geometry at each traversed bronchial level to obtain a collection of virtual images of the intra-operative path. The position of the projected bronchi in such virtual images provides the edge labeling. In order to label the nodes, the trajectory from a node before a bifurcation $v_k$ towards the nodes of the next bifurcation ($v_{k-1}$ and $v_{k+1}$) is projected onto the camera frame.

The camera viewpoint is givenby a point of the airways skeleton oriented according to the Frenet-Serret frame [20, 59] of the skeleton curve. The Frenet-Serret

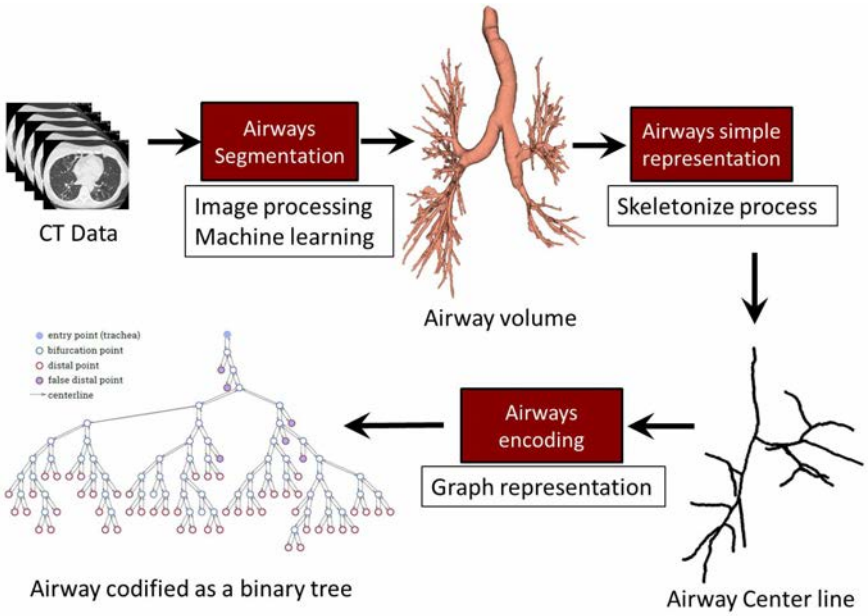



Figure 3.15 – Process of codification airways as a graph from CT data.

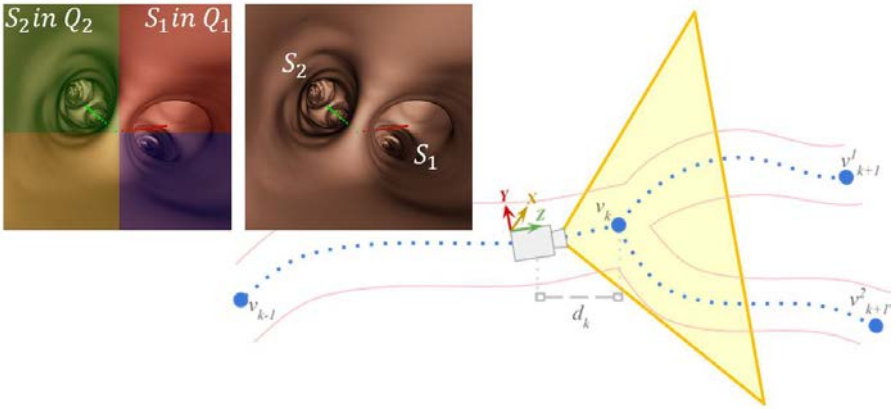



Figure 3.16 – Node Labelling Procedure: the camera captures an image at distance $d_k$ from $v_k$ , projecting lines to children at level $k+1$. Image is split into quadrants to label line segments S1, S2.

frame is the natural reference on the curve associated to its interior geometry. Then, the projected virtual image is split into four quadrants centered at the projected position of $v_k$. Since quadrants represent the spatial distribution of airway lumens in bronchoscopic frames during traversal of bronchial levels, each projected segment will be labeled according to the quadrant it belongs to. Then for each point of the projected trajectory, the position of the quadrant they belong to is computed. For this, the mode and average values are considered. The mode indicates the predominant quadrant where each segment belongs. When two or three segments lie on the same quadrant, they are counterclockwise ordered according to their average.

The labeling process is described in the visual scheme shown in Figure 3.16. Figure 3.16 shows an outline of airways and a navigation path with its branching nodes labeled. The figure also displays a camera positioned at distance $d_k$ from the node $v_k$ , and the segments S1, S2 colored in red and green respectively. The rectangular images show the simplified scene projected over the complete airway anatomy. The most left image is split into four colored quadrants: $Q_1$=red, $Q_2$=green, $Q_3$=yellow, $Q_4$=blue, to illustrate that $S_1$ lies in $Q_1$ and $S_2$ in $Q_2$ .

The planned path is given by sequence of nodes of the "Dijkstra" distance [8] between the root and the node closest to the biopsy point. Planning and on-line exploration paths are matched using the sequence of quadrants of the nodes representing each path. The navigation path is given by the sequence of active root

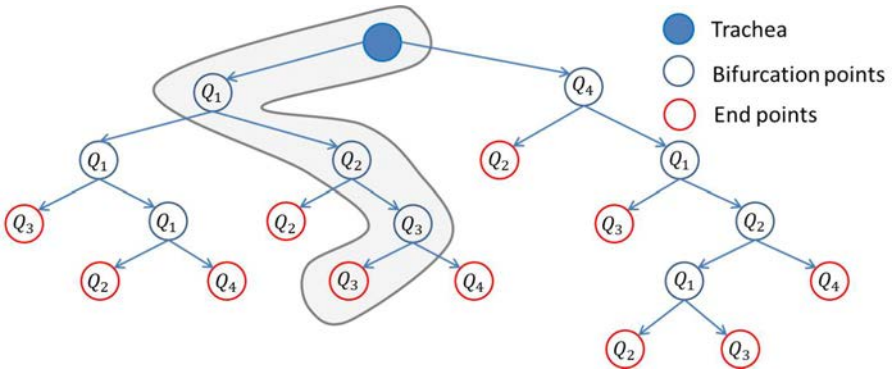



Figure 3.17 – Labeling of airways tree nodes in such a way that the branch to follow is identified into intra-operative bronchoscopy according to [48].

nodes of the on-line exploration tree. We recall that according to our hierarchical definition of the anatomical structure, if the local anatomical tree structure is a single rooted tree, we can extrapolate that we are in the top/middle section of the bronchi. In addition, when our local anatomical tree structure is a set of two trees each of them with one root, we interpret that the bronchoscope is in the bottom part of a bronchi next to a bifurcation. Each of the roots represents the lumens of the bifurcation.

Figure 3.18 illustrates schematically how our proposed navigation system works. In this recreation we will assume we want to follow the path defined in Figure 3.17. An example of labeled airways graph is shown in Figure 3.17. A navigation path across airways can be given by the sequence of the navigated bronchial levels labeled in such a way that the branch to follow is identified in the intra-operative bronchoscopy videos. In Figure 3.17, a navigation path has been highlighted. In order to reach the endpoint of such a path, at each bifurcation we only need to follow the direction of lumens detected at quadrants $Q_1$, $Q_2$, $Q_3$, and $Q_3$ respectively. The navigtion system starts in carina and the bronchoscope moves forward until it reaches the first bifurcation. Regarding the first bifurcation in the navigation path, the first direction to follow is $Q_1$ so lumen detected in $Q_1$ is highlighted to indicate the intervention should continue that way. The procedure follows until the second bifurcation in which the direction is $Q_2$. $Q_3$ is the direction to be followd in the third bifurcation. Finally the direction for the last bifurcation according to our navigation path is $Q_3$. Therefore after crossing the last bifurcation we are in the target bronchi.

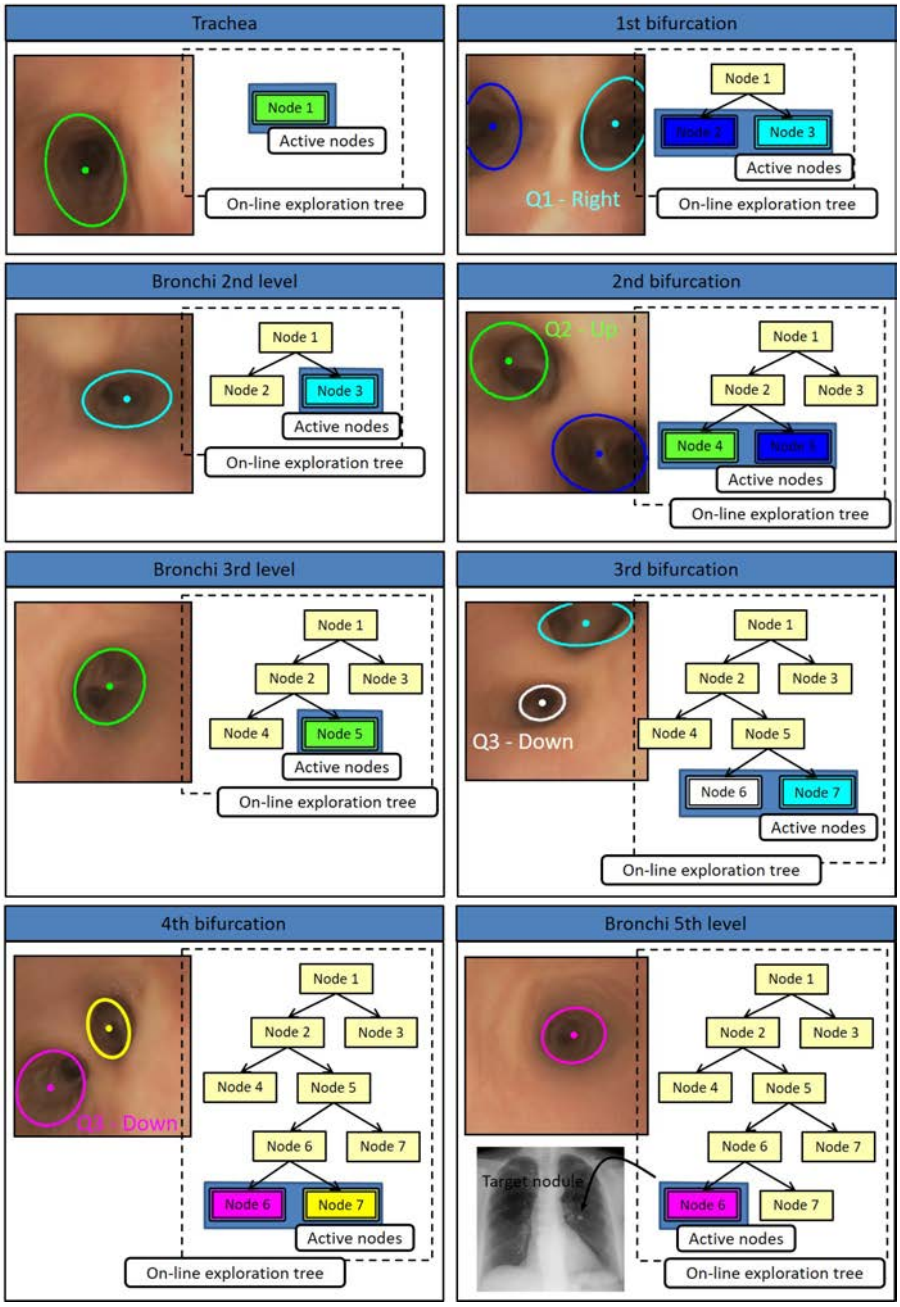
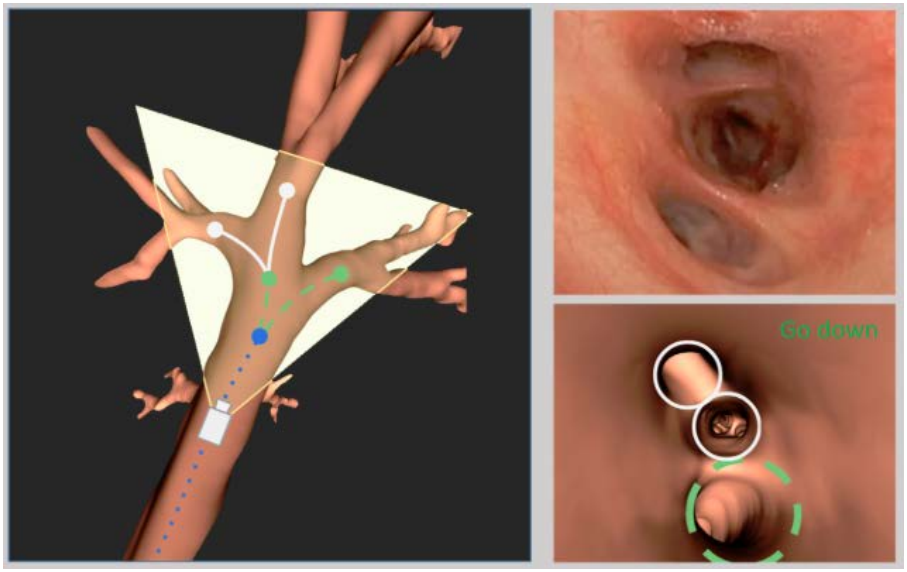





Figure 3.18 – Recreation of an intervention in our navigation system. When the bronchoscope approach a bifurcation the system provides the direction according to the pre-planned route.

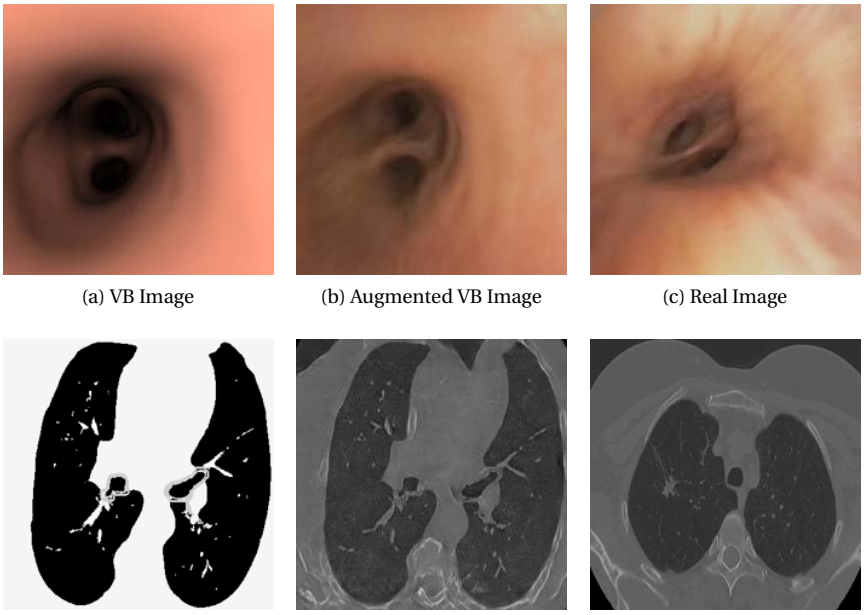
# 4 Results




In this chapter the experimental setups and the results for each of the algorithms developed within this thesis are presented. Experiments have been grouped following the same structure of the main chapters of this thesis. First, Section 4.1 explains the validation of data augmentation for two different kind of images, synthetic virtual broncoscopy images and CT-scan images. In addition, the validation of the navigation system based on bronchi anatomical landmarks is also described in Section 4.2.

## 4.1 Validation of data augmentation

This section addresses data augmentation in two different applications. Data augmentation for virtual bronchoscopy with intra-operative appearance and data augmentation for CT-scans from patient anatomy segmentations. Each of the application has been tested for two kind of images, synthetic images and images generated using simple graphical primitives. On the one hand, Section 4.1.1 shows the results for synthetic virtual bronchoscopy images and segmentation of CT-scans. On the other hand, in order to demonstrate the generalization of the previous models, they were tested on images generated using simple graphical primitives. Such



(a) VB Image (b) Augmented VB Image (c) Real Image

(d) CT-scan segmentation (e) Augmented CT-scan segmentation (f) Real CT-scan

Figure 4.1 – Virtual, augmented and real images for both VB and CT-scans domains.

generalization tests are illustrated in Section 4.1.2.

### 4.1.1 Augmentation from synthetic data

This section shows the results of the augmentation of synthetic virtual bronchoscopy images with intra-operative appearance and the augmentation of CT-scans from patient anatomy segmentations. Figure 4.1 illustrates and example of the image domains and the augmented image for both types of synthetic images. The first row of Figure 4.1 shows an example of the augmentation of synthetic virtual bronchoscopy images with intra-operative appearance. Figure 4.1 (a), (b) and (c) illustrates a Virtual Bronchsocopy images, an augmented Virtual Bronchoscopy images and an image with intra-operative appearance respectively. Similarly, the second row of Figure 4.1 shows an example of the augmentation of CT-scans from patient anatomy segmentations. Figure 4.1 (d), (e) and (f) shows a CT-scan segmentation, an augmented CT-scan segmentation and a real CT-scan respectively.

**Virtual bronchoscopy with intra-operative appearance**

Our method has been used to simulate realistic bronchoscopic interventions. Virtual bronchoscopies defining the $Virtual$ domain were generated using [47] from CT scans acquired with an Aquilion ONE (Toshiba Medical Systems, Otawara, Japan). Scans were selected from 10 patients in the CPAP study [7] conducted at Hospital Bellvitge (Barcelona, Spain). For each scan, we simulated 4 paths, one for each main lung lobule (left and right upper lobes; left and right lower lobes). Intra-operative videos defining the $Real$ domain were acquired also at Hospital Bellvitge (Barcelona, Spain) during biopsy interventions using an Olympus Exera III HD Ultrathin videobronchoscope (6 videos) and CAO diagnostic procedures using a Olympus Exera III HD Therapeutic videobronchoscope (4 videos).

Methods were trained on 4 CT anatomies (16 virtual bronchoscopies) and video recordings from 3 ultrathin explorations and 2 diagnostic procedures. The remaining data were left for validation and testing. The multi-objective cycleGAN was trained from scratch using the whole training set. After 200 epochs, our multi-objective approach selected epoch 50 as the one achieving the best compromise between intr-operative appearance and preservation of virtual anatomical content. Content-net was fine-tuned on the set of intra-operative recordings from an auto-encoder trained to yield the identity map on the $Real$ domain. The weighting function $\rho$ was also learned from the latter auto-encoder.

We performed two diferent experiments which evaluates the quality of the enhanced virtual images in terms of intra-operative appearance and preservation of each patient's anatomy acquired by CT scans. For comparison purposes, simulations were also enhanced using the 200th epoch network, labelled GAN200, and the network achieving the least value of the cost defined in (Equation 2.2), labelled GANLeast. This last network corresponded to the 21st epoch of cycleGAN training.

- **EXP 1: Intra-operative appearance in augmented virtual bronchoscopy images**

  For the first experiment, we trained a network, $D_r$, to discriminate between real and virtual images. This network was evaluated on the test sets of real images and virtual images enhanced using the proposed method, GAN200, and GANLeast.

  For each enhancing method, $D_r$ values were compared to the ones obtained by the test real images to assess their similarity in appearance. Main analysis was performed using a Student T-test for unpaired data. We computed p-values and 95% confidence interval (CI) for the difference. A p-value $< 0.05$ was considered statistically significant.

  Table 4.1 summarizes statistics for the comparison of appearances. For each

enhancing method, we report p-values, 95% CIs of the difference with real image and descriptive statistics (mean and standard deviation (SD)) of $D_r$ values. According to T-tests, none of the methods has an appearance significantly different from intra-operative videos (p-value> 0.05) with all discrimination values $D_r$ in similar ranges comparable to the values achieved by the test set of real images (mean=0.90 and SD=0.11).

| Method | Descriptive | | T-test | |
|---|---|---|---|---|
| | mean | SD | p-val | 95% CI |
| ContentNet | 0.88 | 0.11 | 0.305 | (-0.023,0.073) |
| GAN200 | 0.91 | 0.08 | 0.576 | (-0.057,0.032) |
| GANLeast | 0.91 | 0.09 | 0.438 | (-0.065,0.028) |

Table 4.1 – $D_r$ Statistics for Assessment of Intra-operative Appearance.

Fig 4.2 shows representative images of virtual bronchoscopies enhanced using our method (Figure 4.2 (a)), the 200th epoch network (Figure 4.2(b)) and the least cost one (Figure 4.2(c)). For each case, we show two consecutive

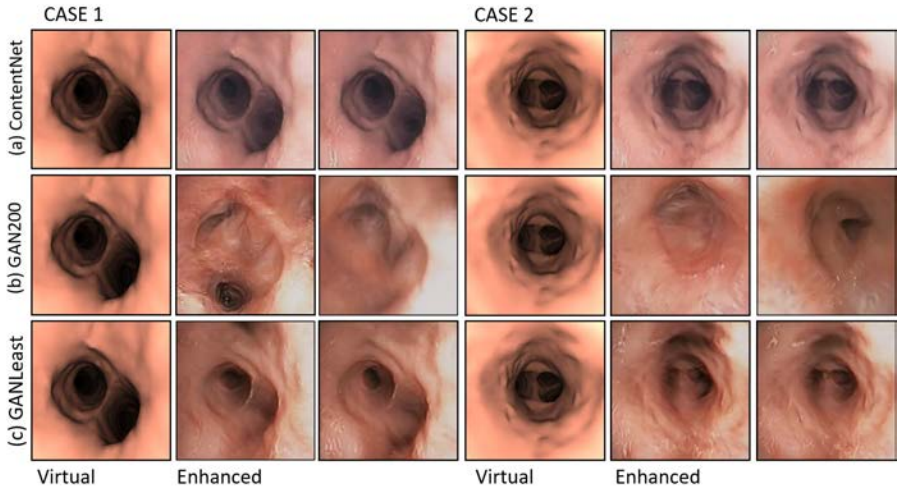



Figure 4.2 – Virtual bronchoscopy enhanced using ContentNet, (a), 200th epoch network, (b) and the least cost one, (c).

frames of the enhanced virtual sequence which should be very similar in appearance and content. GANLeast images have sudden dark artifacts, while GAN200 yields highly unstable images that do not always match the original anatomy. ContentNet provides a stable appearance in images which are the most consistent with the original anatomical content of virtual images.

- **EXP 2: Preservation of patient's anatomy in augmented virtual bronchoscopy images.**

  For the second experiment, we applied the lumen center detector from Chapter 3 to ContentNet, GAN200, GANLeast and the non-enhanced original virtual images to verify that original lumen position and structure is preserved. The center detector was applied using two different sets of parameters, one learned on interventional videos and the other one learned on simulated bronchoscopies. Interventional parameters were used on enhanced images, while simulation parameters were applied to original virtual images. Detections were plot on original virtual images and shown to 2 independent observers for the identification of false detections and missed centres. Inspired in crowd sourcing strategies [34], to statistically compare our tracker, ground truth was produced by intersecting the experts' annotations as illustrated in Figure 4.3 and used to compute precision and recall.

  Scores obtained for ContentNet, GAN200, GANLeast were compared to the ones obtained for virtual non-enhanced images using a Student T-test for

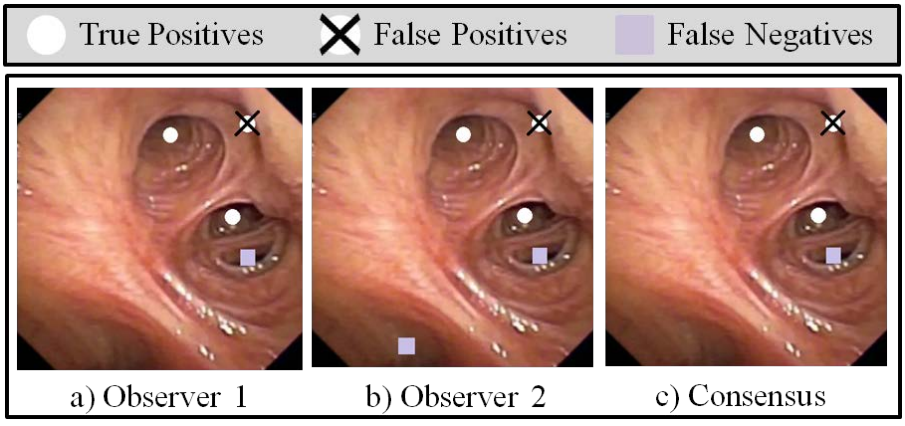



Figure 4.3 – Manual Annotation of Tracked Centres.

paired data. As in the first experiment, we computed p-values and 95% confidence intervals and a p-value < 0.05 was considered statistically significant.

Table 4.2 and 4.3 summarize statistics for Prec and Rec reported as in Table 4.1. ContentNet outperforms the two cycleGAN, both, in terms of precision and recall.

The recall of GAN200 and GANLeast is under 0.75 and is significantly lower (p-value < 0.01) than the recall obtained for the non-enhanced virtual images. This drop in the recovery of the original anatomical structure indicates that enhanced images systematically distort the anatomical content. ContentNet recall is almost equal to the one obtained in virtual images with a difference less than 0.04, which shows that it consistently preserves the anatomical content along the sequence.

Concerning precision, ContentNet averages are a bit higher (p-value=0.025 and CI=(-0.1150, -0.0082)) than the average precision of original virtual images. This might be attributed to artifacts in the segmentation of distal airways that are prone to introduce shadows in virtual sequences resembling the appearance of luminal distal areas. Such shadowing artifacts do not appear in intra-operative videos and, thus, they are significantly reduced in ContentNet enhanced images. In fact, ContentNet precision ranges ($0.94 \pm 0.09$) are closer to the full precision of intra-operative videos [54] than ranges for

| Method | Descriptive | | T-test | |
|---|---|---|---|---|
| | mean | SD | p-val | 95% CI |
| ContentNet | 0.94 | 0.09 | 0.025 | (-0.1150, -0.0082) |
| GAN200 | 0.89 | 0.15 | 0.695 | (-0.0719, 0.0487) |
| GANLeast | 0.84 | 0.19 | 0.265 | (-0.0332, 0.1148) |

Table 4.2 – Precision Statistics for Assessment of Anatomical Content.

| Method | Descriptive | | T-test | |
|---|---|---|---|---|
| | mean | SD | p-val | 95% CI |
| ContentNet | 0.90 | 0.14 | 0.939 | ( -0.0384, 0.0413) |
| GAN200 | 0.70 | 0.24 | <0.01 | (0.1400, 0.3559) |
| GANLeast | 0.73 | 0.24 | <0.01 | (0.1081, 0.3058) |

Table 4.3 – Recall Statistics for Assessment of Anatomical Content.

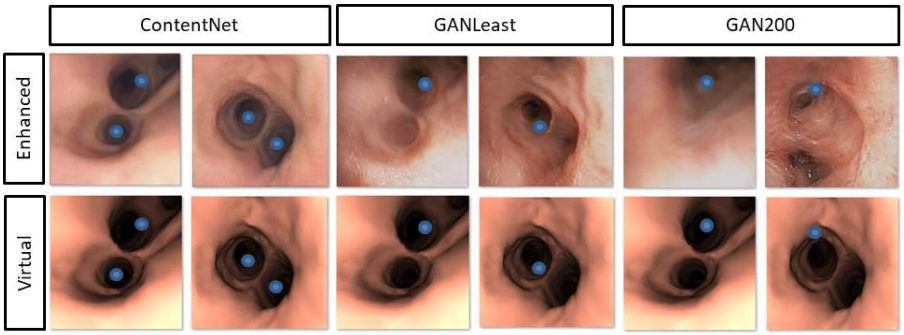



Figure 4.4 – Detected lumens on virtual bronchoscopy enhanced using our method (ContentNet), 200th epoch network (GAN200) and the least cost one (GANLeast).

virtual images ( $0.8843 \pm 0.1364$). Finally, precision average ranges for GAN200 and GANLeast compares to virtual images precision which indicates that they present some kind of artifact (like the shadows of the images shown in Fig 4.2) that intra-operative videos do not have.

Fig 4.4 shows two representative examples of lumens detected on images enhanced with each method. Lumen centers are shown in blue points on, both, original virtual images and the enhanced ones. In the top case, GAN200 and GANLeast enhanced images miss the bottom branch. In the bottom case, in spite of a slight deviation in the location of its center, GANLeast succeeds in keeping the upper branch, while GAN200 enhancement has such a distorted anatomy that none of the original virtual branches can be identified. The proposed ContentNet preserves the virtual anatomy in both cases and, thus, all branches are properly detected.

**CT-scans from patient anatomy segmentations**

For this experiment we trained a cycleGAN on a set of 11 chest CT volumes (defining the $Real$ domain) and their segmentation (defining the $Virtual$ domain) of the main lung structures. CT scans were acquired in inspiration with an Aquilion ONE (Toshiba Medical Systems, Otawara, Japan) using slice thickness and interval of 0.5 and 0.4 mm respectively. Segmentations included the body, lungs, pulmonary vessels and airways and were computed using an own-designed software. Each volume (original CT scan and its segmentation) was uniformly sampled in 250 short axis planes covering the whole volume of lungs for training cycleGAN from scratch.

A total of 10 CT-scans segmentation have been inspected in order to prove the

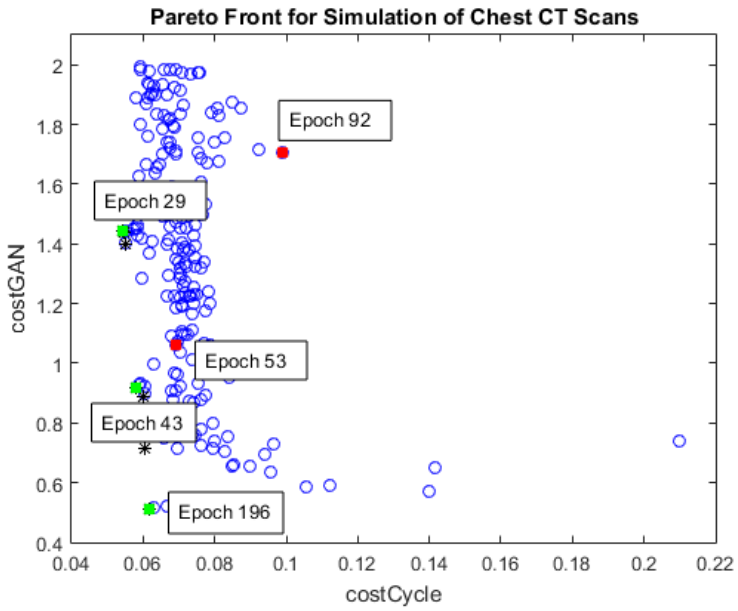



Figure 4.5 – Pareto front for CT-scans.

improvement of the automatic epoch selection using the Pareto front epochs. In this case, we have selected 2 epochs not belonging to the Pareto front, one (epoch 92) far from it, another one (epoch 53) close to it. The epochs selected on the Pareto front are the boundary ones (196, 29, which is the one achieving the minimum value of (Equation 2.2)) and the epoch (43) closest to epoch 53.

Figure 4.5 plots the two cost functions for all epochs with the Pareto epochs in black crosses. The epochs selected for these experiment are shown in squares, green ones for epochs on the Pareto front, red ones, otherwise.

Figure 4.6 shows a mosaic of representative segmentations and their augmentation using these 5 epochs. The epochs selected for these experiment are shown in squares, green ones for epochs on the Pareto front, red ones, otherwise. Images generated with epochs not belonging to the Pareto front present different artifacts, like a stripe pattern (epoch 92), lack of texture in lungs (epoch 53) and bright structures in the body (epoch 53). Meanwhile, images generated with Pareto epochs keep a stable appearance similar to the one of a chest CT scan.

As shown in Figure 4.6, images from Pareto epochs present a spatial continuity which is not present in non-Pareto epochs. It is important to mention that the GAN

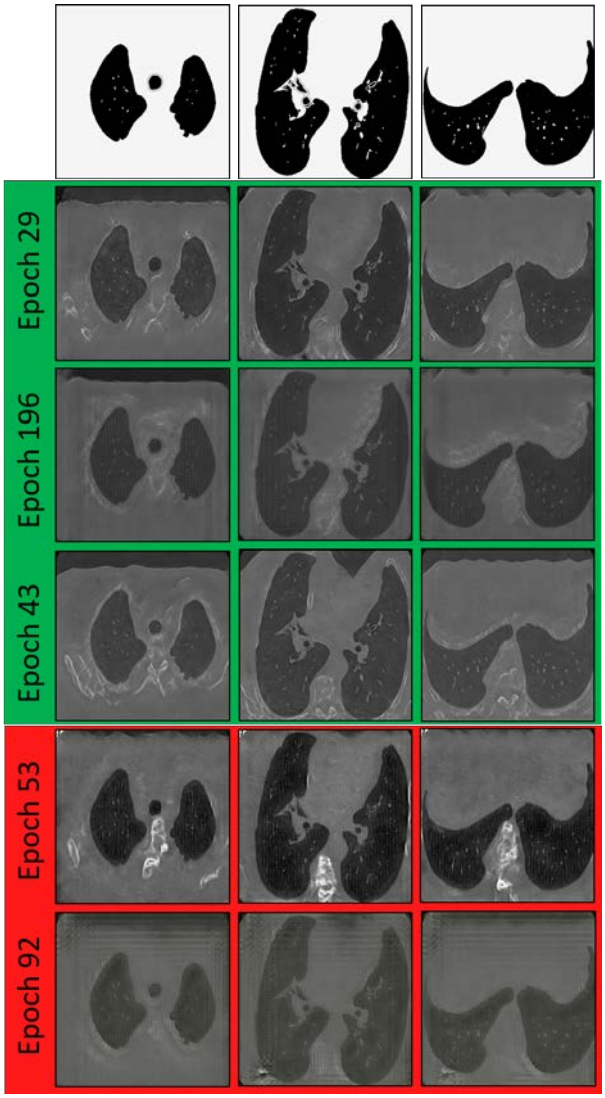



Figure 4.6 – Generated CT virtual scans for 3 representative CT segmentations (basal, mid and distal): images from Pareto front epochs in green frame and images from non-Pareto epochs in red one.

which transforms CT segmentations into augmented images with CT appearance and viceversa has been trained with transversal images of CT-scans. In order to prove the stability of the Pareto front epochs, some coronal and sagittal slices are shown.

Figure 4.7 shows a mosaic of representative augmented coronal CT-scans slices for Pareto and non-Pareto epochs. Pareto epochs are shown in squares, green ones

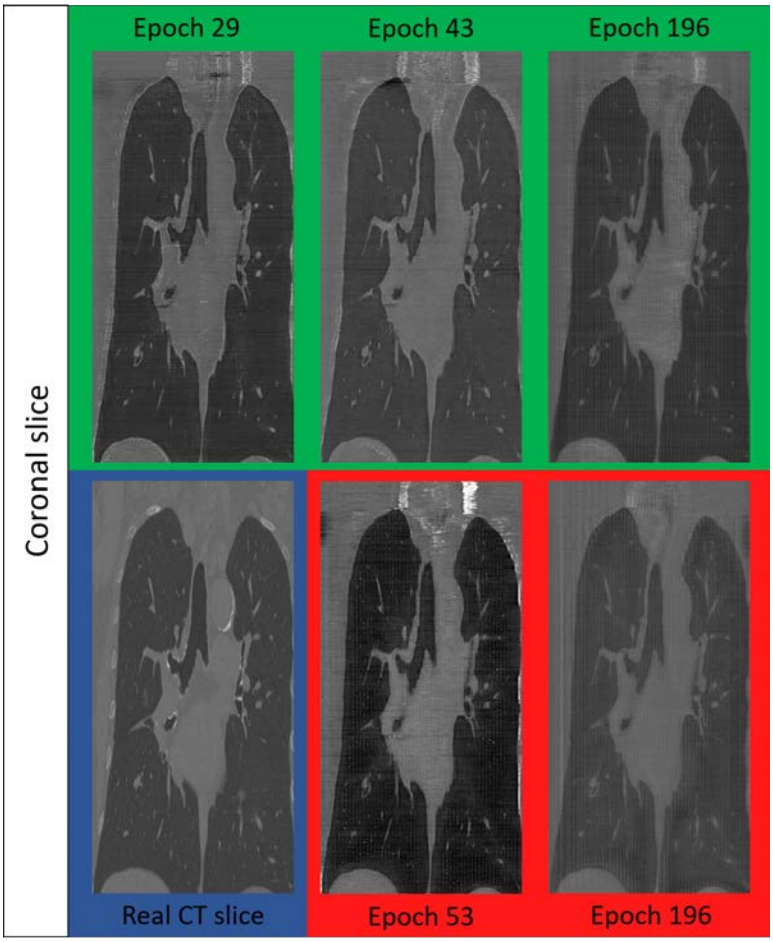



Figure 4.7 – Augmented coronal CT-scans slices for Pareto and non-Pareto epochs.

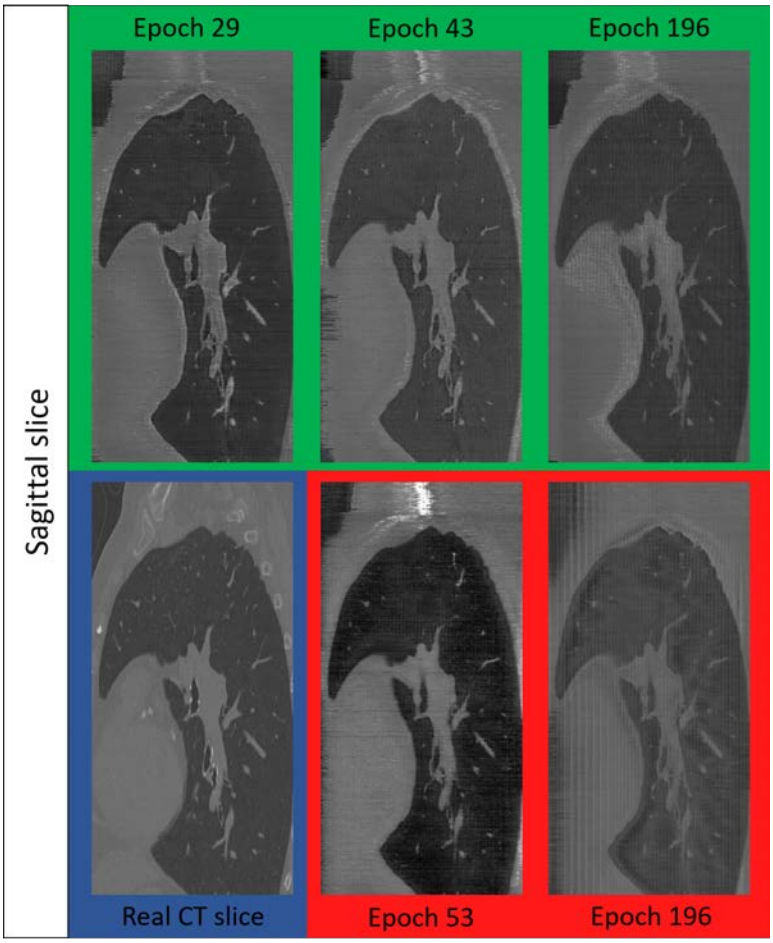



Figure 4.8 – Augmented sagittal CT-scans slices for Pareto and non-Pareto epochs.

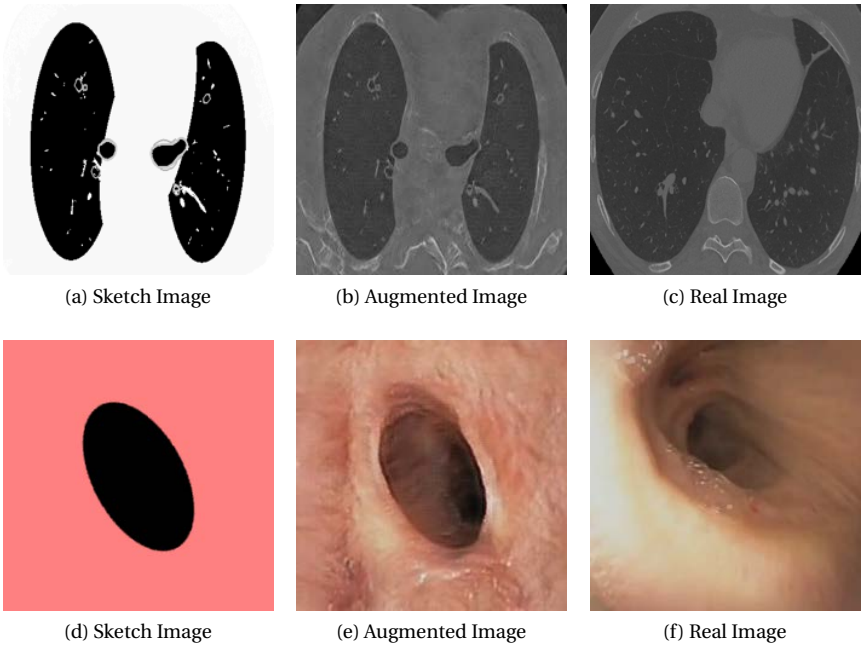
for epochs on the Pareto front, red ones, otherwise. The real CT slice correspondent to the augmented one is shown in a blue square. We can see that in pareto epochs (epoch 29, epoch 43 and epoch 196) there is continuity in the slices even though it was trained for transversal slices. We can also see that for non-pareto epochs (epoch 53 and epoch 92), there is not continuity due to some strip and square patterns in the image.

The same experiment for sagittal slices were performed. Figure4.8 shows a mosaic of representative augmented sagittal CT-scans slices for Pareto and non-Pareto epochs. Image frame colors are like in Figure 4.7. As in Figure 4.7, we can see that in pareto epochs (epoch 29, epoch 43 and epoch 196) there is continuity in the slices while in non-pareto epochs (epoch 53 and epoch 92) there is not.

### 4.1.2 Augmentation from graphical primitives

This section shows the results of the augmentation of virtual bronchoscopy images with intra-operative appearance and the augmentation of CT-scans both from images generated with graphical primitives. We call such images sketch images. The goal of these experiments is to prove the generalization of the GANs trained in Section 4.1.1.

Figure 4.9 illustrates and example of the image domains and the augmented



(a) Sketch Image (b) Augmented Image (c) Real Image

(d) Sketch Image (e) Augmented Image (f) Real Image

Figure 4.9 – Virtual, augmented and real images for both VB and CT-scans from sketch domains.

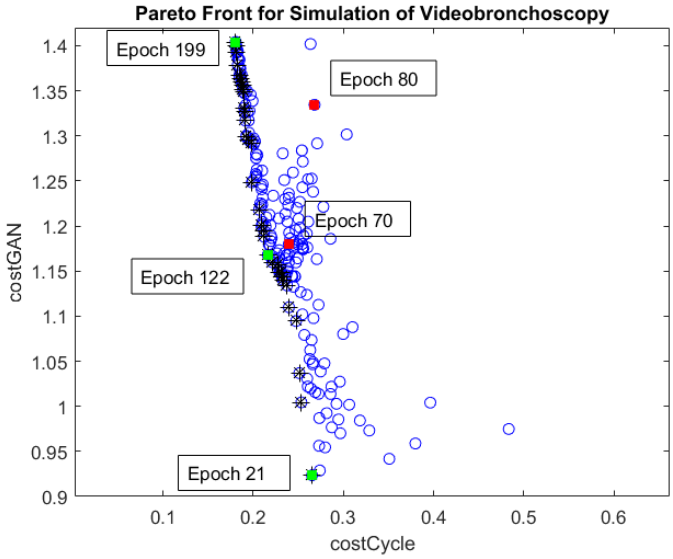



Figure 4.10 – Pareto front for Intra-operative virtual bronchoscopies.

image for both types of generated images. The first row of Figure 4.9 shows an example of the augmentation of a generated virtual bronchoscopy sketch images with intra-operative appearance. Figure 4.9 (a), (b) and (c) illustrates a Virtual Bronchsocopy sketch images, an augmented Virtual Bronchoscopy images and an image with intra-operative appearance respectively. Similarly, the second row of Figure 4.9 shows an example of the augmentation of CT-scans from sketches. Figure 4.9 (d), (e) and (f) shows a CT-scan sketch segmentation, an augmented CT-scan segmentation and a real CT-scan respectively.

**Video bronchoscopy data augmentation from graphical primitives**

In order to validate our proposal, we generated 10 sequences using sketches of airways in bronchoscopic frames. Sketches of airways were produced using elliptical shapes of different sizes and eccentricity and considering several configurations in number and disposition in the image. For each configuration, a sequence of affine transformations was applied to simulate bronchoscopic navigation.

Figure 4.10 plots the two cost functions for all epochs with the 35 Pareto epochs in black crosses. The epochs selected for these experiment are shown in squares, green ones for epochs on the Pareto front, red ones, otherwise. We have selected 2

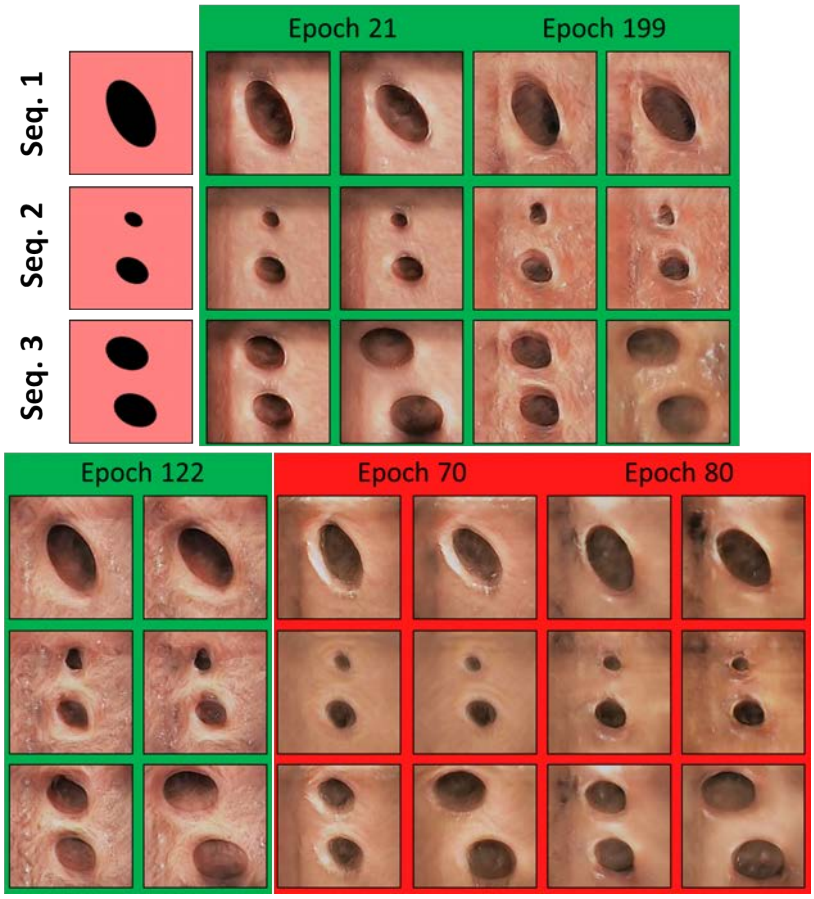



Figure 4.11 – Generated intra-operative virtual bronchoscopies for 3 representative sketch sequences: images from Pareto front epochs in green frame and images from non-Pareto epochs in red one.

epochs not belonging to the Pareto front, one (epoch 80) far from it, another one (epoch 70) close to it. The epochs selected on the Pareto front are the boundary ones (199, 21, which is the one achieving the minimum value of (Equation 2.2)) and the epoch (122) closest to epoch 70.

Figure 4.11 shows a mosaic of representative sketches and their augmentation using the 5 epochs. For each sketch, we show consecutive frames to check the sta-

bility of the augmented images. Images obtained from Pareto epochs have a green frame, red otherwise. As it can be seen, the appearance of images generated with Pareto front epochs are more stable and keep better texture. Also brightness, color and the shape of holes are more realistic. Epoch 21 artifacts on image boundary and lack of texture can be attributed to an early stage of cycleGAN training. It is worth noticing that images generated with epochs not belonging to the Pareto front (like epoch 80) present an unstable anatomy with new holes that were not in the sketches.

#### CT scans augmentation from graphical primitives

A total number of 10 anatomical sketches were generated using 3D graphical primitives for each lung structure. Body was modelled with a cylinder, lungs with ellipsoids and, both, vessels and airways were modelled as tubular structures with several branching levels. The thickness of the tubes was set depending on the level to account for different lumen sizes. To endow sketches with more realistic geometry, the cardiac and abdominal cavities were also modelled using ellipsoids. The ranges for primitives sizes were statistically learned from segmented CTs. For each sketched anatomy 250 short axis planes were sampled for testing.

In this case, the Pareto front had 7 epochs. As in the previous experiment, we selected the Pareto boundary epochs (196 and the least cost 29), two epochs not belonging to the Pareto front (epoch 92 far from it and epoch 53 close to it) and the Pareto epoch (43) closest to epoch 53. Figure 4.12 shows a mosaic of representative sketches and their augmentation using these 5 epochs. Image frame colors are like in fig.4.11. As before, images generated with epochs not belonging to the Pareto front present different artifacts, like a stripe pattern (epoch 92), lack of texture in lungs (epoch 53) and bright structures in the body (epoch 53). Meanwhile, images generated with Pareto epochs keep a stable appearance similar to the one of a chest CT scan.

## 4.2 Validation of navigation system based on bronchi anatomical landmarks

This section addresses the validation of our navigation system based on bronchi anatomical landmarks. Section 4.2.1 is dedicated to analyze the extraction of bronchial anatomical hierarchy in single videobronchoscopy frames. The next section, Section 4.2.2, shows the accuracy of the on-line exploration tree extraction. Finally, the last section, Section 4.2.3, proves the capabilities of our navigation systems for guiding purposes in augmented virtual bronchoscopy videos.

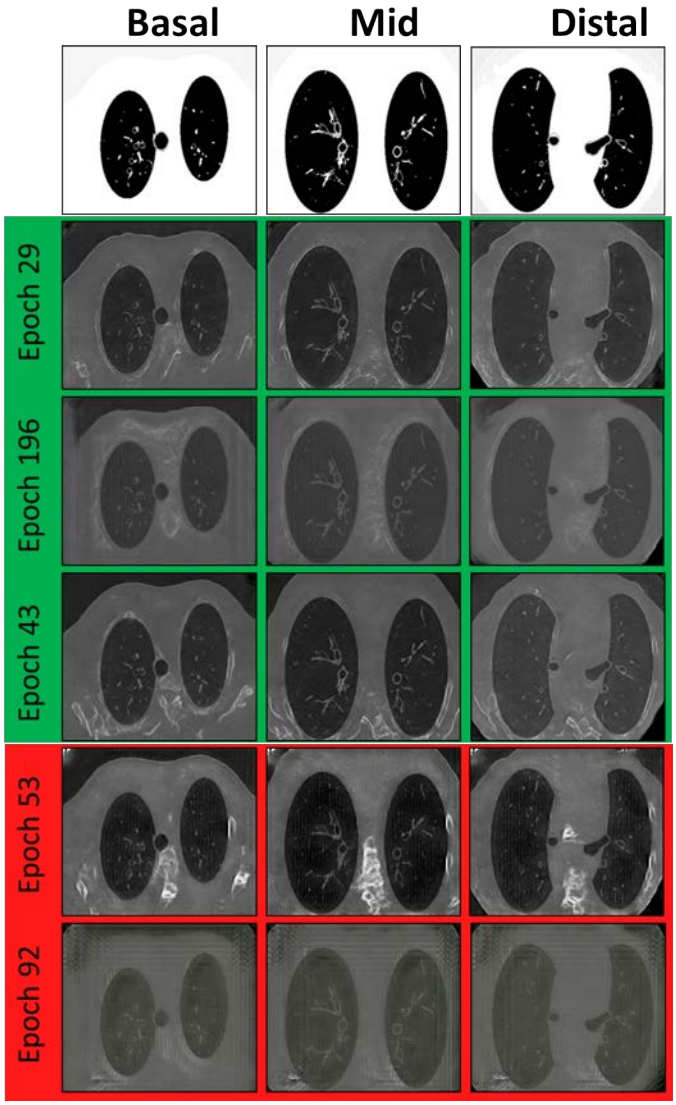



Figure 4.12 – Generated CT virtual scans for 3 representative sketches: images from Pareto front epochs in green frame and images from non-Pareto epochs in red one.

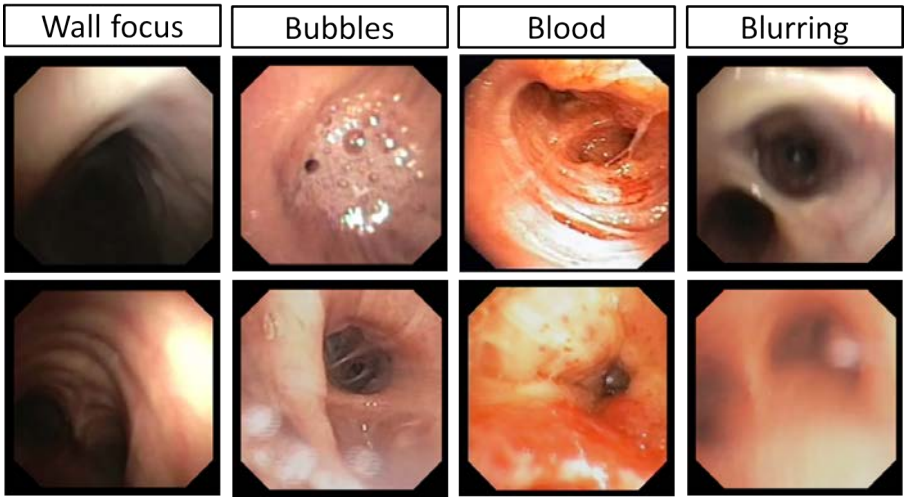



Figure 4.13 – Challenges in anatomical landmark extraction: wall focus, bubbles, blood, blurring.

### 4.2.1 Extraction of bronchial anatomical structure in single video-bronchoscopy frames

Our method has been applied to 8 sequences extracted from 3 ultrathin bronchoscopy videos performed for the study of peripheral pulmonary nodules at Hospital de Bellvitge. Videos were acquired using an Olympus Exera III HD Ultrathin videobronchoscope. We have split the 8 sequences into proximal (up to 6th division) and distal (above 6th) sets to compare also the impact of the distal level. The maximum bronchial level achieved in our ultrathin explorations was within 10th and 12th, which is in the range of the maximum expected level reachable by ultrathin navigation. Sequences contain bronchoscope collision with the bronchial wall, bubbles due to the anaesthesia and patient coughing, blood and blurred images. Such phenomena are illustrated in Figure 4.13.

For each sequence, we sampled 10 consecutive frames every 50 frames. Such frames were annotated by 2 clinical experts to set false detections and missed centres. Like in Subsection 4.1.1, to statistically compare our tracker, ground truth was produced by intersecting the experts' annotations as illustrated in fig.4.3. Ground truth sets were used to compute precision (Prec) and recall (Rec) for each set of consecutive frames. These scores are taken for all such sets in distal and proximal fragments for statistical analysis. We have used a Wilcoxon test data to assess

| | Proximal | Distal | Total |
|---|---|---|---|
| Prec | [1.0, 1.0] | [1.0, 1.0] | [1.0, 1.0] |
| Rec | [0.84, 0.99] | [0.91, 0.98] | [0.90, 0.97] |

Table 4.4 – Average precision and recall confidence intervals for region tracking.

significant differences and confidence intervals, CI, to report average expected ranges.

Table 4.4 reports CIs for each set of consecutive frames score at proximal, distal and all together (both proximal and distal) levels. According to these results, it

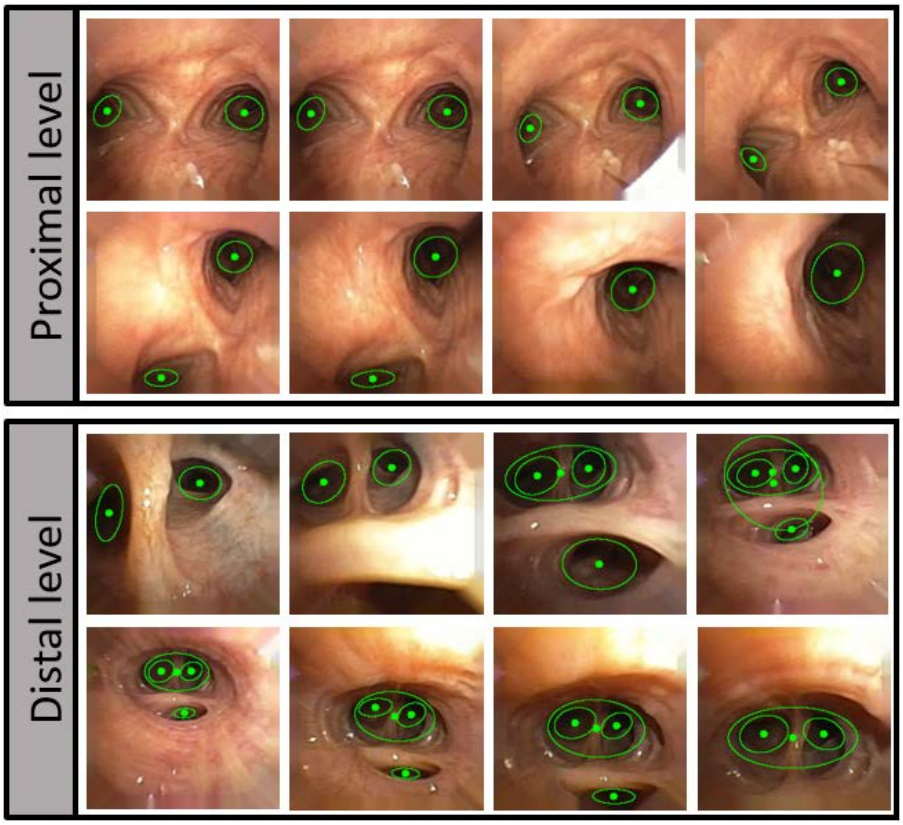



Figure 4.14 – Frames of tracked regions at proximal and distal levels.

is worth noticing that the proposed method always has a 100% of precision and a recall over 84%. We can see that Recall at proximal levels is a bit smaller than recall at distal levels. This is due to more frames with collisions at proximal levels that distort the likelihood model (see Discussion Section). Even so, proximal and distal levels present non-significant differences between them ($p-val > 0.8$ for a Wilcoxon test).

Figure 4.14 shows regions tracked in consecutive frames selected at distal and proximal levels. It is worth noticing the capability of our strategy to capture most distal and lateral bronchi without introducing false positives.

### 4.2.2 Extraction of the on-line exploration tree in intra-operative videos

The extraction of the on-line exploration tree has been tested on 8 interventional videos. Videos were collected from four different patients. For each patient 2 videos with different paths were recorded.

In order to validate the extraction of the on-line exploration tree, each of the interventional videos were inspected and the ground truth tree with all the bronchi seen across the video was created. Both trees were compared to compute the precision and recall of the on-line exploration tree.

| Experiments | FP | FN | Num. Nodes |
|---|---|---|---|
| Patient 1 - Video 1 | 2 | 0 | 27 |
| Patient 1 - Video 2 | 6 | 2 | 21 |
| Patient 2 - Video 1 | 0 | 0 | 17 |
| Patient 2 - Video 2 | 0 | 2 | 11 |
| Patient 3 - Video 1 | 0 | 2 | 11 |
| Patient 3 - Video 2 | 0 | 2 | 19 |
| Patient 4 - Video 1 | 8 | 2 | 15 |
| Patient 4 - Video 2 | 4 | 0 | 11 |

Table 4.5 – TN, FN and total number of nodes for the on-line exploration tree extraction.

Table 4.5 reports the total number of nodes in the ground truth, as well as, the number of false positive (FP) and false negative (FN) nodes in the on-line exploration tree extracted for each patient and exploration. Table 4.6 reports CIs for the average precision and recall. CIs indicate that our algorithm detects most of the bronchi in the explorations. In terms of precision, our algorithm has more variability since the number of FP is highly dependent on either patient or physicians.

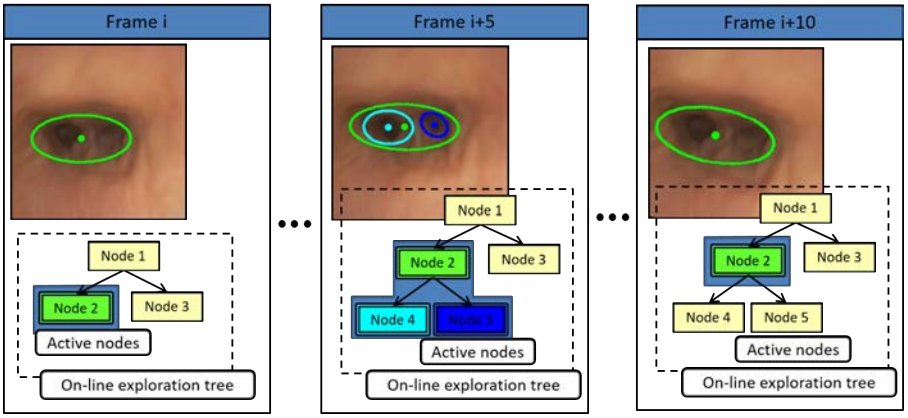



Figure 4.15 – Robustness of the on-line exploration tree creation.

For example, some patients produce bubbles during the exploration. In case of the doctors, they don't perform the explorations in the same way. Inexperience physicians perform explortions very smoothly due to their lack of experience while very experienced physicians perform the exploration abruptly because they are used to perform such procedures. These abrupt changes ruin our algorithm since sudden changes in position leads to failures in the matching step of the algorithm.

| Precision | Recall |
|---|---|
| [66.80%, 98.43%] | [84.71%, 100.0%] |

Table 4.6 – Precision and recall statistics for the on-line exploration tree extraction.

It is worth to notice that even when the local anatomical structure changes at some frames, the current scope of the bronchoscope is not affected. Since the scope of the bronchoscope is determined by the roots of the available nodes sub-tree, as long as these nodes are not affected, the navigation system remains unaffected.

Figure 4.15 shows the robustness of the on-line exploration tree creation. The sequence start with a frame in which only one lumen has been detected in the left-hand side image. In such frame, the root node of the active nodes sub-tree indicates that the bronchoscope is in the second bronchial level. After a few frames in the middle image of Figure 4.15, two new lumens are added as children of the root of the active nodes sub-tree. in the right-hand image of Figure 4.15 we can see that as the bronchoscope moves forwards changes in illumination, bubbles or blood cause that the children of the root node are not detected anymore. However,

since the root of the active nodes sub-tree remains the same, the on-line exploration tree indicates that the bronchoscope is still in the second bronchial level.

### 4.2.3 Matching between CT-planning and on-line exploration tree in augmented sequences

We have compared the matching between the CT-planning and the on-line exploration path in augmented virtual bronchoscopies with intra-operative appearance.

Virtual bronchoscopies were generated using an own developed software from CT scans acquired for four different patients with an Aquilion ONE (Toshiba Medical Systems, Otawara, Japan) using slice thickness and interval of 0.5 and 0.4 mm respectively. For each Virtual bronchoscopy of a patient, four virtual explorations were generated, covering the four main lobes: left and right upper lobes, noted LUL, RUL, and left and right lower lobes, noted LLL, RLL. These paths were performed using central navigation without rotation around the scope. In addition, explorations were considered with paths reaching between the sixth and twelfth bronchial level. Each virtual explorations was augmented according to the algorithms explained in Chapter 2 and the ground truth for comparison purposes was extracted following the same approach as [48].

We have compared the quality of the retrieved paths extracted from the on-line exploration tree using Section 3.3 (on-line paths) against the paths using method in [54]. For each exploration, the path followed by the truth camera positions was compared to the path extracted from both algorithms. The metrics used for the comparison of both methods are True Positives Nodes (TPN) and True Path Representations (TPR). For a given exploration, a node is considered to be a TPN if its label coincides with the GT node label. The number of consecutive TPN achieved from the 1st node divided by the path node length defines TPR. We have used a T-test to detect significant differences across methods and CIs at significance $\alpha$ = 0.05 to report average TPN and TPR ranges.

| | On-line path | Paths from [54] | p-val |
|---|---|---|---|
| TPN | [78.06%, 94.43%] | [71.07%, 80.95%] | 0.0235 |
| TPR | [47.31%, 85.28%] | [54.25%, 74.82%] | 0.8086 |

Table 4.7 – Comparison of TPN and TPR at proximal and distal levels.

Table 4.7 reports CIs for average TPN and TPR percentages for both methods and p-values of the T-test for the difference. The detection rate of the correct bifurcations along the whole path (TPN) is significantly different with a CI for the difference equal to [1.44, 19.03]. This improvement is due to the fact that the proposed

Figure 4.16 – Ternary bifurcation in an augmented bronchoscopy image. The image was generated from a virtual bronchoscopy extracted from a CT-scan of a real patient.

algorithm not only detects luminal regions but it also encodes the hierarchy relationships between them. Hierarchy relationships allow to increase the robustness when luminal regions are not detected in some frames. Concerning the percentage of correct paths reached from the trachea (TPR), there are not significant differences with a CI for the difference equal to $[-16.39, 20.92]$. Although, our algorithm outperforms the approach from [54] in terms of TPN, there are not significant differences due to the TPR metric itself. Given a path with all the bifurcations but the first one correct, the TPR value is 0 since there is none correct value starting from the trachea. Both methods use binary trees to encode the path followed during the intervention but this does not hold for some ternary bifurcations. Such ternary bifurcations appear in the proximal levels leading to low TPR values for both methods.

Figure 4.16 shows a ternary bifurcation in an augmented image. This images was generated from CT-scan data of a real patient. Although in general, bronchial trees can be represented by a binary tree, there are a few exceptions where some bifurcation are ternary. Figure 4.16 illustrates clearly a ternary bifurcation but ternary bifurcations are very challenging since there is a lot of intra-observer variability when ground truth is generated.

# 5 Conclusions and Future work


## 5.1 Conclusions

A main limitation of flexible bronchoscopy is the difficulty to determine the correct pathway to peripheral lesions. Despite all the available tools, the diagnostic yield is still under 70% and drops for small lesions [36].

Standard guiding protocols such as fluoroscopy have a diagnostic yield around 40%, last 20 min per intervention and require 5-10 min of repetitive patient and medical staff radiation [50]. Existing alternatives like image based systems (LungPoint, NAVI) or electromagnetic navigation (inReachTM,SPinDrive®) are far from meeting clinician expectations. Image systems are based on multimodal registration of CT virtual projections to bronchoscopy frames and require manual intra-operative adjustments of the guidance system. Electromagnetic systems require specific gadgets, that alter the operating protocol and increase intervention cost. Aside, they lack of any correction in orientation changes. In order to cope with such difficulties, we proposed an image-based navigation system for intra-operative guidance of bronchoscopists to a target lesion across a previously planned path on a CT-scan.

This thesis contributes in two aspects. The first contribution is related to the augmentation of synthetic images with intra-operative appearance that can be used for testing or training complex algorithms that need tons of images. For the augmentation of virtual endoscopic images, we propose a two-stage algorithm using intra-operative videos based on Generative Adversarial Networks. In addition, we use techniques of multi-objective optimization schemes to provide a numeric criteria for the automatic selection of cycleGAN optimal epochs.

Furthermore, the second and most important contribution of this thesis is the development of a navigation system based on anatomical landmarks. By detecting and tracking anatomical landmarks across the intervention we are able to construct an on-line exploration tree. A sub-tree of the on-line exploration, active nodes sub-tree, represents the anatomy currently observed in the interventional video and its root gives the level inside the whole exploration tree. This allows us to identify all the bifurcation and properly indicate the pre-planned path that needs to be followed at intervention.

The main conclusions for the contributions of the augmentation of synthetic images with intra-operative appearance are:

- **Numeric criteria for the automatic selection of GAN optimal epochs.** Artifacts of cycleGANs images might be partially attributed to the adversarial training. On one hand, the combination of two loss functions with different (opposite, indeed) goals (minimization and maximization) introduces an oscillating behavior across training epochs and, thus, consecutive epochs might produce very different results. On the other hand, it is not guaranteed that both losses will have equal influence during training since the magnitude of one of the two might be predominant in the back-propagation of their gradients.

  The proposed multi-objective approach allows the join optimization of both losses ensuring equal influence on the cycleGAN, regardless of their magnitude or gradient. This way, our multi-objective cycleGAN produces stylized images that share enough anatomical structure with virtual images as to be the input for a network blending both image pairs. The weighted skip connections of ContentNet provide selective blending of the structure and texture of these image pairs. This allows enhancing the patient specific anatomical content acquired by CT scans, while keeping an intra-operative appearance. In this context, it is worth noticing that ContentNet precision and recall ranges achieved in the detection of airways structure (centers) is very close to the ranges obtained in intra-operative videos [54].

- **Augmentation of synthetic images.** We have presented a method for style transfer that preserves the structure of original input images and, thus, it is suitable for endowing virtual endoscopy with intra-operative appearance.

  Content-net has been compared to state-of-art style transfer methods based on GANs. In particular, we chose cycleGAN after 200 epochs and the cycleGAN achieving the minimum cost. Two experiments were conducted. The first one assessed whether the appearance of enhanced images compares to intra-operative videos for their use in classification problems. The 2nd experiment

assessed whether the anatomical content of the virtual images extracted from patient's CT is preserved after their enhancement for their use in image processing problems.

Results obtained for the first experiment show that, like cycleGAN, Content-net has an appearance close enough to intra-operative videos as to be classified real by a discriminative network. This validates our method for data augmentation in classification problems. The 2nd experiment shows that images enhanced using both cycleGANs have a significant distortion in anatomical content (see Figure 4.4) and have larger temporal artifacts (see Figure 4.2) in comparison to Content-net. In fact, according to T-tests Content-net anatomical structure is not no significantly different from original virtual images extracted from patient's CT anatomy. This validates our method for data augmentation in image processing problems.

In summary, two interesting conclusions can be inferred from our experiments. First, the use of multi-objective optimization strategies can be an effective alternative to back-propagation for the optimization of adversarial networks and other networks relying on multiple loss functions. In this context, the Pareto front condition could also be adapted for the selection of the most appropriate task in sequential multi-task learning. Second, the structure of neuron activations can be measured by the amount of information shared with input images. This measure of their content provides a description easy to interpret in terms of classical computer vision. We envision that this could be useful to define more specific and interpretable representation spaces based on CNNs.

The main conclusions for the contributions of the novel strategy for a bronchoscopic navigation system based on anatomical landmarks are presented next:

- **Extraction of bronchial anatomical hierarchy in single video-bronchoscopy.** Luminal regions of different bronchial levels projected in the image results in a hierarchy of regions. We have introduced a method that extracts such anatomical hierarchy from videobronchoscopy frames.

  Results in ultrathin bronchoscopy videos indicate high equal performance of our anatomical hierarchy extractor at proximal and distal levels. Particularly, there are not any false detections (Prec=1) and the rate of missed lumen regions is under 16% (Rec > 0.84). Although, non-significant according to a Wilcoxon test, we can appreciate a slight deviation between proximal and distal recall. The reason for such bias is that our model does not satisfy the illumination conditions in carina when collisions happen. This could be solved by making the likelihood maps less restrictive at proximal levels, but does not invalidate our system for bronchoscopic navigation. Clinicians need

guiding systems for distal levels in which we obtain a recall grater than 90%, at proximal levels, they navigate without any tool just by visually assessing the path.

We conclude that results are promising enough (see the full exploration at to encourage the use of anatomical landmarks in a biopsy guidance system. In Fig. 4.14 we can see 8 sample images from two videos at distal and proximal level. Images are ordered according to its occurrence in time from left to right and from up to down. As we can see, at proximal levels the anatomical structure of bronchi is easy but at distal levels it becomes more complex.

- **Extraction of the on-line exploration tree in intra-operative videos.** The robustness of the anatomical landmark extraction allows to accurately detect the luminal hierarchy of video broncshocopy frames. Such luminal hierarchy describes key points of the whole anatomical structure of the lungs. The identification of bronchial tree key-points in CT scans and videobronchoscopy frames provide accurate matching between planed path and endoscopic navigation.

  By continuously matching the local anatomical hierarchy of single bronchoscopic frames across the video, we are able to construct an on-line exploration tree. The online-exploration tree encodes the key-points in videobronchoscopy frames seeing during the whole intervention.

- **Matching between CT-planning and on-line exploration tree in augmented sequences** The online-exploration tree encodes the anatomical structure seen at any time across an interventional video. This means, it also encodes the path that the bronchoscope followed during the intervention. Hence, the on-line exploration path can be matched to a previously computed CT-planning path in order to provide guidance.

  We have compared the matching between the CT-planning and the on-line exploration path in augmented virtual bronchoscopies with intra-operative appearance. The Results in augmented virtual bronchoscopies shows high performance estimating the paths followed during interventions. Particularly, the algorithm correctly detects the bifurcation that needs to be followed in 86% of the total number of bifurcations. This promising results validate our assumption of a navigation system based on anatomical landmarks without using neither radiation nor additional gadgets.

Several interesting conclusions can be extracted from the experiments. Our algorithm detects robustly the anatomical structure in single frames. The robustness of the anatomical extraction allows to create an on-line exploration tree which

encodes the position of the bronchoscope and the path followed at any time of the intervention. By putting into correspondences the on-line exploration path and the pre-planned CT path we can provide guidance to reach a target lesion. In summary, experiments shows that anatomical landmarks are stable enough to provide accurate guidance during bronchoscopic procedures.

## 5.2 Future Work

As described in previous chapters of this thesis, Medical imaging applications are challenging for machine learning and computer vision methods, in general, for two main reasons: it is difficult to generate reliable ground truth and databases are usually too small in size for training state of the art methods. We have presented an algorithm for the augmentation of virtual endoscopic images using intra-operative videos based on convolutional neural networks. Although, it shows promising results there is still some room for improvement:

- **Customized virtual bronchoscopies with intra-operative.**

  In the scope of this thesis we have only use one appearance to transfer to synthetic images. This limits the use of the algorithm for real purposes. A classifier could be trained to differentiate between different patient appearance. The classifier could be an input to the GAN so that it could change the appearance of the output image. This would allow to give virtual bronchosocpies with intra-operative appearance at intervention time. In addition, since we can collect depth maps, optical flow, camera position,... we could generate a virtual bronchoscopy database with different intra-operative appearance that could be use for navigation testing purposes or for training complex algorithms which need huge amount of images.

- **Augmentation of synthetic images with intra-operative appearance using 3D information.**

  GANs have proved its efficiency in the generation of new images. In the scope of this thesis, new images were generated with intra-operative appearance. However, content of such images was slightly lost so a new step was added in order to preserve more content in the generated images. Since we use images from virtual bronchoscopies, there is a lot of additional information such as depth maps, optical flow that could be used. We could use depth maps in order to preserve content instead of using $L_2 norms$ which reduce the texture transferred by the net. Depth maps might be a smart way to preserve content since content appearance is a huge contribution in the depth estimation.

The main contribution of this thesis is the development of a navigation system based on anatomical landmarks. But it is unrealistic to perform bronchoscopy procedures smoothly. This is translated into wall collisions or sudden movements backward and forward. Such events might cause our system to loose precision. The methods proposed in this PhD thesis pave the way for several studies:

- **Multi-modal navigation system based on anatomical landmarks.**

  We have proved that bronchoscopic navigation based on anatomical landmarks is a good approach to detect and indicate paths to reach a lesion. It uses the detected bifurcations to indicate the path to follow. Therefore, anatomical landmarks between bifurcation would not been necessary in theory. A multi-modal navigation system could drastically improve the physician experience in the operating room. The system could add speech recognition to allow

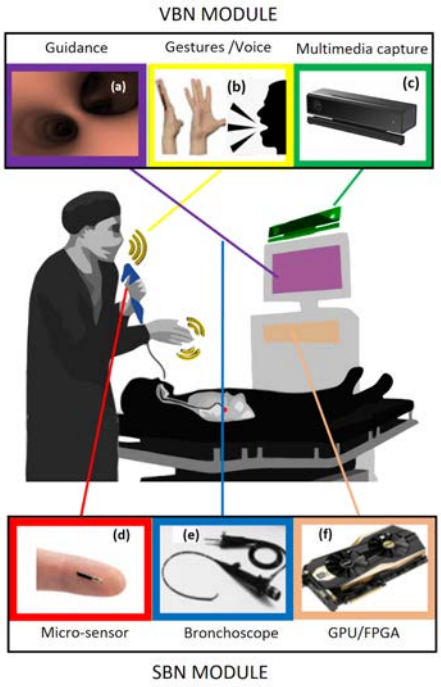



Figure 5.1 – Schema of the future multi-modal navigation system. Guidance screen, (a), gesture/voice recognition, (b), multimedia capture, (c), micro-sensor, (d), bronchoscope, (e), gpu/fpga, (f).

the physician to interact with the system without stopping the intervention. The physician could inform the system that the bronchoscope is at a bifurcation. Then the system would extract the anatomical hierarchy and the path to follow would be highlighted. The physician could even correct rotation deviations between the interventional video and the VB to ease the process.

Figure 5.1 shows a proposal for a multi-modal navigation system. Figure 5.1 (a) shows the guidance support system to the physician who can interact with it by means of either gesture/voice recognition (Figure 5.1 (b)) captured by multimedia systems, Figure 5.1 (c). In addition, the bronchoscope (Figure 5.1 (e)) will contain a micro-sensor (Figure 5.1 (d)) which allows to detect bronchoscope rotations. The system will be implemented on gpu/fpga to fulfill real-time constraints.

- **Improve performance of the algorithm.**

  Although our method is by far not the most time consuming method, the time for processing a single image makes the algorithm not usable for a real prototype that can be tested in a real operating room. But even so, we could use the hierarchy structure of bronchoscopic images using our algorithm and use it as a ground truth to train a neural network which would make the whole pipeline usable for real time applications.

- **First prototype of a multi-modal navigation system based on anatomical landmarks.**

  After a superficial feasibility study of a multi-modal navigation system based on anatomical landmarks, a first prototype could be build and tested on a real operating room. This would imply the use of GPU for fast computation and the definition of protocols which make the prototype acceptable for broad use in hospitals.

## 5.3 Scientific Articles

This dissertation has led to the following communications:

### 5.3.1 Journals

- Enhancing virtual bronchoscopy with intra-operative data using a multi-objective gan, rennes, france, june 18–21, 2019. *International Journal of Computer Assisted Radiology and Surgery*, 14(1):S7, Jun 2019 (Accepted)

- Marta Diez-Ferrer, Debora Gil, Cristian Tebe, Carles Sanchez, Noelia Cubero, Rosa López-Lisbona, Jordi Dorca, Antoni Rosell, PAP enhanced CT study group, et al. Positive airway pressure to enhance computed tomography imaging for airway segmentation for virtual bronchoscopic navigation. *Respiration*, 96(6):525–534, 2018 (Accepted)

- Antonio Esteban Lansaque, Carles Sánchez Ramos, Agnes Borràs, and Debora Gil Resina. Augmentation of virtual endoscopic images with intra-operative data using content-nets. *bioRxiv*, page 658807, 2019 (Submitted)

### 5.3.2 International Conferences and Workshops

- Carles Sánchez, Antonio Esteban-Lansaque, Agnès Borràs, Marta Diez-Ferrer, Antoni Rosell, and Debora Gil. Towards a videobronchoscopy localization system from airway centre tracking. In *VISIGRAPP (4: VISAPP)*, pages 352–359, 2017

- Antonio Esteban-Lansaque, Carles Sánchez, Agnés Borràs, Marta Diez-Ferrer, Antoni Rosell, and Debora Gil. Stable anatomical structure tracking for videobronchoscopy navigation. In *Workshop on Clinical Image-Based Procedures*, pages 18–26. Springer, 2016

- A Esteban-Lansaque, C Sanchez, A Borras, M Diez-Ferrer, A Rosell, and D Gil. Stable airway center tracking for bronchoscopic navigation. *28th Conference of the international Society for Medical Innovation and Technology*, 2016

## 5.4 Scientific Dissemination

### 5.4.1 In the media

The work of this thesis has lead to scientific dissemination in different media. The different interviews are listed as follows:

- *El GPS Del Pulmón: Imagen Médica Y Salud Conectada, CVC Outreach, January, 2019*

- *"En nuestro sistema GPS, usamos IA para localizar y contabilizar los bronquios que se ven en cada imagen», BigData Magazine, October, 2018*

- *Tratamientos a medida y cirugía guiada con GPS: el 'big data' llega a la medicina, Retina, El Pais, September, 2018*

- *El GPS que ayudará a los médicos a recorrer el cuerpo humano, La Vanguardia, November, 2017*
- *How video game tech, AR, and 3D models help these surgeons do their job, ZDNet, April, 2017*
- *Un «videojoc» per millorar les broncoscòpies, Butlletí RECERCAT, February, 2017*
- *Videogame Technology And Computer Vision For The Bronchoscope Of Tomorrow, CVC Outreach, January, 2017*
- *Broncoscopia en 3D, Catalunya Vanguardista, January, 2017*
- *Idean un GPS bronquial para la detección precoz del cáncer de pulmón, BioTech, January, 2017*

### 5.4.2 Demos

- *Planning and intervention Demo for lung cancer biopsy - http://datasets.cvc.uab.es/ust/agnesba/MWC2019/BronchoX_Demo_MWC/Exe/index.html*

# Bibliography


[1] Enhancing virtual bronchoscopy with intra-operative data using a multi-objective gan, rennes, france, june 18–21, 2019. *International Journal of Computer Assisted Radiology and Surgery*, 14(1):S7, Jun 2019.

[2] Fumihiro Asano, Ralf Eberhardt, and Felix JF Herth. Virtual bronchoscopic navigation for peripheral pulmonary lesions. *Respiration*, 88(5):430–440, 2014.

[3] British Medical Association, Tony Smith, et al. *Endoscopy. The British Medical Association Complete Family Health Encyclopedia Complete Family Health Encyclopedia*. Dorling Kindersley, 1990.

[4] Philip Bille. A survey on tree edit distance and related problems. *Theoretical computer science*, 337(1-3):217–239, 2005.

[5] Jean-Paul Charbonnier, Eva M Van Rikxoort, Arnaud AA Setio, Cornelia M Schaefer-Prokop, Bram van Ginneken, and Francesco Ciompi. Improving airway segmentation in computed tomography using leak detection with convolutional networks. *Medical image analysis*, 36:52–60, 2017.

[6] Rick Daniels. *Delmar's guide to laboratory and diagnostic tests*. Cengage Learning, 2014.

[7] Marta Diez-Ferrer, Debora Gil, Cristian Tebe, Carles Sanchez, Noelia Cubero, Rosa López-Lisbona, Jordi Dorca, Antoni Rosell, PAP enhanced CT study group, et al. Positive airway pressure to enhance computed tomography imaging for airway segmentation for virtual bronchoscopic navigation. *Respiration*, 96(6):525–534, 2018.

[8] Edsger W Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.

[9] J Dionisio. Diagnostic flexible bronchoscopy and accessory techniques. *Revista Portuguesa de Pneumologia (English Edition)*, 18(2):99–106, 2012.

[10] Edwin F Donnelly. Technical parameters and interpretive issues in screening computed tomography scans for lung cancer. *Journal of thoracic imaging*, 27(4):224–229, 2012.

[11] Qi Dou, Lequan Yu, Hao Chen, Yueming Jin, Xin Yang, Jing Qin, and Pheng-Ann Heng. 3d deeply supervised network for automated segmentation of volumetric medical images. *Medical image analysis*, 41:40–54, 2017.

[12] Ralf Eberhardt, Nicolas Kahn, Daniela Gompelmann, Maren Schumann, Claus Peter Heussel, and Felix JF Herth. Lungpoint—a new approach to peripheral lesions. *Journal of Thoracic Oncology*, 5(10):1559–1563, 2010.

[13] Alexei A Efros and Thomas K Leung. Texture synthesis by non-parametric sampling. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1033–1038. IEEE, 1999.

[14] A Esteban-Lansaque, C Sanchez, A Borras, M Diez-Ferrer, A Rosell, and D Gil. Stable airway center tracking for bronchoscopic navigation. *28th Conference of the international Society for Medical Innovation and Technology*, 2016.

[15] Antonio Esteban-Lansaque, Carles Sánchez, Agnés Borràs, Marta Diez-Ferrer, Antoni Rosell, and Debora Gil. Stable anatomical structure tracking for video-bronchoscopy navigation. In *Workshop on Clinical Image-Based Procedures*, pages 18–26. Springer, 2016.

[16] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[17] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2414–2423, 2016.

[18] Patricia R Geraghty, Stephen T Kee, Gillian McFarlane, Mahmood K Razavi, Daniel Y Sze, and Michael D Dake. Ct-guided transthoracic needle aspiration biopsy of pulmonary nodules: needle size and pneumothorax rate. *Radiology*, 229(2):475–481, 2003.

[19] Thomas R Gildea, Peter J Mazzone, Demet Karnak, Moulay Meziane, and Atul C Mehta. Electromagnetic navigation diagnostic bronchoscopy: a prospective study. *American journal of respiratory and critical care medicine*, 174(9):982–989, 2006.

[20] Gabriel A Hare, Keith Mertens, Matias Perez, ROY Neeloy, and David Holz. Non-linear motion capture using frenet-serret frames, January 2 2018. US Patent 9,857,876.

[21] James P Helferty, Chao Zhang, Geoffrey McLennan, and William E Higgins. Videoendoscopic distortion correction and its application to virtual guidance of endoscopy. *IEEE transactions on medical imaging*, 20(7):605–617, 2001.

[22] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[23] Ahmedin Jemal, Freddie Bray, Melissa M Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: a cancer journal for clinicians*, 61(2):69–90, 2011.

[24] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 2019.

[25] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.

[26] Nagi Fouad Khouri, FP Stitik, YS Erozan, PK Gupta, WS Kim, WW Scott Jr, Ulrike Maria Hamper, RB Mann, JC Eggleston, and R Robinson Baker. Transthoracic needle aspiration biopsy of benign and malignant lung lesions. *American journal of roentgenology*, 144(2):281–288, 1985.

[27] Agnieszka Korchut, Sebastian Szklener, Carla Abdelnour, Natalia Tantinya, Joan Hernández-Farigola, Joan Carles Ribes, Urszula Skrobas, Katarzyna Grabowska-Aleksandrowicz, Dorota Szczęśniak-Stańczyk, and Konrad Rejdak. challenges for service robots—requirements of elderly adults with cognitive impairments. *Frontiers in Neurology*, 8, 2017.

[28] Antonio Esteban Lansaque, Carles Sánchez Ramos, Agnes Borràs, and Debora Gil Resina. Augmentation of virtual endoscopic images with intra-operative data using content-nets. *bioRxiv*, page 658807, 2019.

[29] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[30] Chuan Li and Michael Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *European Conference on Computer Vision*, pages 702–716. Springer, 2016.

[31] Lin Liang, Ce Liu, Ying-Qing Xu, Baining Guo, and Heung-Yeung Shum. Real-time texture synthesis by patch-based sampling. *ACM Transactions on Graphics (ToG)*, 20(3):127–150, 2001.

[32] Tony Lindeberg. Scale-space: A framework for handling image structures at multiple scales. 1996.

[33] Xióngbião Luó, Marco Feuerstein, Daisuke Deguchi, Takayuki Kitasaka, Hirotsugu Takabatake, and Kensaku Mori. Development and comparison of new hybrid motion tracking for bronchoscopic navigation. *Medical image analysis*, 16(3):577–596, 2012.

[34] Lena Maier-Hein, Daniel Kondermann, Tobias Roß, Sven Mersmann, Eric Heim, Sebastian Bodenstedt, Hannes Götz Kenngott, Alexandro Sanchez, Martin Wagner, Anas Preukschas, et al. Crowdtruth validation: a new paradigm for validating algorithms that rely on image correspondences. *International journal of computer assisted radiology and surgery*, 10(8):1201–1212, 2015.

[35] Jiri Matas, Ondrej Chum, Martin Urban, and Tomás Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10):761–767, 2004.

[36] Jessica S Wang Memoli, Paul J Nietert, and Gerard A Silvestri. Meta-analysis of guided bronchoscopy for the evaluation of the pulmonary nodule. *Chest*, 142(2):385–393, 2012.

[37] Kaisa Miettinen. *Nonlinear multiobjective optimization*, volume 12. Springer Science & Business Media, 2012.

[38] Saad Minhas, Aura Hernández-Sabaté, Shoaib Ehsan, Katerine Díaz-Chito, Ales Leonardis, Antonio M López, and Klaus D McDonald-Maier. Lee: A photorealistic virtual environment for assessing driver-vehicle interactions in self-driving mode. In *Computer Vision–ECCV 2016 Workshops*, pages 894–900. Springer, 2016.

[39] Kensaku Mori, Daisuke Deguchi, Jun Sugiyama, Yasuhito Suenaga, Jun-ichiro Toriwaki, CR Maurer, Hirotsugu Takabatake, and Hiroshi Natori. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. *Medical Image Analysis*, 6(3):321–336, 2002.

[40] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015.

[41] Jiro Nagao, Kensaku Mori, Tsutomu Enjouji, Daisuke Deguchi, Takayuki Kitasaka, Yasuhito Suenaga, Jun-ichi Hasegawa, Jun-ichiro Toriwaki, Hirotsugu Takabatake, and Hiroshi Natori. Fast and accurate bronchoscope tracking using image registration and motion prediction. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004*, pages 551–558, 2004.

[42] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[43] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

[44] Mateusz Pawlik and Nikolaus Augsten. Tree edit distance: Robust and memory-efficient. *Information Systems*, 56:157–173, 2016.

[45] Josien PW Pluim, JB Antoine Maintz, and Max A Viergever. Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging*, 22(8):986–1004, 2003.

[46] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[47] Esmitt Ramírez, Carles Sánchez, Agnés Borràs, Marta Diez-Ferrer, Antoni Rosell, and Debora Gil. Bronchox: bronchoscopy exploration software for biopsy intervention planning. *Healthcare technology letters*, 5(5):177–182, 2018.

[48] Esmitt Ramírez, Carles Sánchez, Agnés Borràs, Marta Diez-Ferrer, Antoni Rosell, and Debora Gil. Image-based bronchial anatomy codification for biopsy guiding in video bronchoscopy. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, pages 214–222. Springer, 2018.

[49] Pall J Reynisson, Håkon O Leira, Toril N Hernes, Erlend F Hofstad, Marta Scali, Hanne Sorger, Tore Amundsen, Frank Lindseth, and Thomas Langø. Navigated bronchoscopy: a technical review. *Journal of bronchology & interventional pulmonology*, 21(3):242–264, 2014.

[50] M Patricia Rivera and Atul C Mehta. Initial diagnosis of lung cancer: Accp evidence-based clinical practice guidelines. *Chest*, 132(3):131S–148S, 2007.

[51] Carles Sánchez, Jorge Bernal, Debora Gil, and F Javier Sánchez. On-line lumen centre detection in gastrointestinal and respiratory endoscopy. In *Clinical image-based procedures. Translational Research in Medical Imaging*, pages 31–38. Springer, 2013.

[52] Carles Sánchez, Jorge Bernal, F Javier Sánchez, Marta Diez, Antoni Rosell, and Debora Gil. Toward online quantification of tracheal stenosis from videobronchoscopy. *International journal of computer assisted radiology and surgery*, 10(6):935–945, 2015.

[53] Carles Sánchez, Marta Diez-Ferrer, Jorge Bernal, F Javier Sánchez, Antoni Rosell, and Debora Gil. Navigation path retrieval from videobronchoscopy using bronchial branches. In *Workshop on Clinical Image-Based Procedures*, pages 62–70. Springer, 2015.

[54] Carles Sánchez, Antonio Esteban-Lansaque, Agnès Borràs, Marta Diez-Ferrer, Antoni Rosell, and Debora Gil. Towards a videobronchoscopy localization system from airway centre tracking. In *VISIGRAPP (4: VISAPP)*, pages 352–359, 2017.

[55] Ray W Shepherd. Bronchoscopic pursuit of the peripheral pulmonary lesion: navigational bronchoscopy, radial endobronchial ultrasound, and ultrathin bronchoscopy. *Current opinion in pulmonary medicine*, 22(3):257–264, 2016.

[56] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1):29, 2015.

[57] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312, 2016.

[58] Marco Visentini-Scarzanella, Takamasa Sugiura, Toshimitsu Kaneko, and Shinichiro Koto. Deep monocular 3d reconstruction for assisted navigation in bronchoscopy. *International journal of computer assisted radiology and surgery*, 2017.

[59] Michael G Wagner and Bahram Ravani. Curves with rational frenet-serret motion. *Computer Aided Geometric Design*, 15(1):79–101, 1997.

[60] Weiwei Wu, Zhuhuang Zhou, Shuicai Wu, and Yanhua Zhang. Automatic liver segmentation on volumetric ct images using supervoxel-based graph cuts. *Computational and mathematical methods in medicine*, 2016, 2016.

[61] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.

[62] Karin Zielinski and Rainer Laur. Stopping criteria for a constrained single-objective particle swarm optimization algorithm. *Informatica*, 31(1), 2007.

[63] Karin Zielinski, Dagmar Peters, and Rainer Laur. Stopping criteria for single-objective optimization. In *Proceedings of the Third International Conference on Computational Intelligence, Robotics and Autonomous Systems*, 2005.