SUPERVISED TISSUE CLASSIFICATION IN OPTICAL IMAGES: TOWARDS NEW APPLICATIONS OF SURGICAL DATA SCIENCE

Doctoral Dissertation of:

Sara Moccia

Supervisors:
Elena De Momì, PhD
Leonardo S. Mattos, PhD

Tutor:
Prof. Raffaele Dellacà

The Chair of the Doctoral Program:
Prof. Andrea Aliverti

2018 – XXX
«Marco Polo describes a bridge, stone by stone. “But which is the stone that supports the bridge?” Kublai Khan asks. “The bridge is not supported by one stone or another,” Marco answers, “but by the line of the arch that they form.” Kublai Khan remains silent, reflecting. Then he adds: “Why do you speak to me of the stones? It is only the arch that matters to me.” Polo answers: “Without stones there is no arch.”»

Italo Calvino, Invisible Cities
Abstract

INTRA-OPERATIVE tissue classification plays a fundamental role in different clinical fields. The automatic image-based tissue classification is a valuable solution to provide decision support and context awareness intra-operatively. The surgical data science (SDS) community is focusing more and more on machine learning (ML) to perform tissue classification in the operating room (OR). However, several technical challenges are still present, hampering the clinical translation of the developed methodology in the clinical practice. Indeed, robust and reliable tissue classification is not trivial due to noise in the image, varying illumination level, different camera pose with respect to the tissues, and intra- and inter-patient tissue variability. There are several aspects that can be tackled to potentially overcome these issues, including hardware design (to allow high-quality optical imaging), identification of images to be processed (to avoid the processing of uninformative images), and estimation of classification confidence (to improve system performance by excluding uncertain results).

On this background, the overall goal of this PhD thesis was to develop a framework for ML-based SDS algorithms for robust and reliable tissue classification in intra-operative optical images, as to offer decision support and provide context awareness during interventional-medicine processes. In particular, the contributions of this PhD work are:

1. A new method (M1) to automatic learning-based informative frame selection, which exploited a new set of features to retrieve informative frames to be processed by tissue classification algorithms

M1 exploited intensity, keypoint and image spatial-content features to classify, with multi-class support vector machines (SVM), endoscopic frames as informative, blurred, with saliva or specular reflections, or underexposed. When tested on a balanced set of 720 images from 18 different narrow-band laryngoscopic videos, a classification recall of 91% was achieved for informative frames, significantly overcoming the state of the art approaches (median recall = 57%, maximum recall = 81%).

2. A new method (M2) for reliable tissue classification in endoscopic images, which introduced a measure of confidence on classification to retrieve reliable classifi-
cation results and improve classification performance

M2 exploited textural information and supervised and semi-supervised approaches coupled with confidence estimation for pathological tissue classification. When tested on 33 narrow-band laryngoscopic videos, which refer to 33 different patients affected by early-stage laryngeal cancer, a median classification recall of 93% was achieved. The method was also tested on the hepatic district, achieving a recall of 82% in classifying 40 RGB images acquired in the OR of healthy and pathological liver donors’ grafts. The classification recall was increased to 98% (laryngeal tissue) and to 86% (hepatic tissue) by estimating the confidence of the SVM classification and excluding low-confidence results.

3. A new protocol (M3) for endoscopic-image analysis, which exploited multispectral imaging (MI) for data acquisition during endoscopic procedures with the goal of performing robust tissue classification and image tagging

M3 expanded M2 by exploiting for the first time in the literature intra-operative in vivo multispectral-imaging data acquisition and analysis. Multispectral data were acquired during laparoscopic procedures on 7 pigs. Six abdominal organs (abdominal wall, liver, spleen, gallbladder, diaphragm and intestine) were classified with multi-class SVM. The effect of using multispectral data was an increase in accuracy of 11% for the task of organ classification and an increase of 23% for the task of automatic image tagging with respect to standard RGB imaging. When exploiting the confidence measure proposed in M2, the result was a boost in classification accuracy of 38% (RGB) and 20% (MI).

4. Integration of the SDS ML-based methodologies within a handheld robotic tool to perform tissue avoidance in simulated robotic-assisted surgical procedures (M4)

With M4, ML-based tissue segmentation was integrated within a simulated robot-assisted surgical scenario. A forbidden-region virtual-fixture control was exploited, as to prevent unwanted robot-tissue interaction. In particular, phantom vascular structures were built and segmented using a deep–learning approach. When including the control, the error on the robotic tip position when it entered the forbidden zone (24% of the time) was small. Indeed, the median penetration error was 17 µm, which was 2 orders of magnitude smaller than median vessel diameter.

This PhD thesis work lies in the wider research field of SDS, which aims at providing the surgeons with decision support and context awareness during interventional-medicine procedures. The methodologies presented in this PhD thesis demonstrated the feasibility of using SDS ML-based algorithms for intra-operative tissue classification in several anatomical districts. The methodological progress made in this work highlights the potential of SDS ML-based algorithms in extracting useful information implicitly contained in intra-operative optical images, overcoming challenges typical of intra-operative tissue classification to support surgeons during interventional-medicine procedures.
A classificazione intra-operatoria di tessuti biologici gioca un ruolo fondamentale in diversi campi clinici. Metodi di classificazione automatica basata su immagini possono fornire supporto decisionale ed informazioni circa il contesto in cui il chirurgo sta operando. Con lo scopo di eseguire in maniera automatica la classificazione di tessuti da immagini estratte in sala operatoria (operating room, OR), la comunità che si occupa di scienza dei dati nell’ambito delle procedure di medicina interventistica (surgical data science, SDS) si sta concentrando sempre di più sull’apprendimento automatico (machine learning, ML). Tuttavia, le sfide tecniche che ostacolano la traslazione delle metodologie sviluppate nella pratica clinica sono molteplici. Le problematiche da affrontare al fine di ottenere una classificazione robusta ed affidabile sono: la presenza di rumore nelle immagini, il livello di illuminazione variabile, la diversa posa della camera rispetto ai tessuti ed infine la variabilità intrinseca dei tessuti (sia intra che inter-paziente). Questi problemi possono essere affrontati attraverso la progettazione hardware (per ottenere immagini di qualità sempre più alta), l’identificazione delle immagini da elaborare (per evitare l’elaborazione di immagini non informative rispetto al contesto analizzato) e la stima della confidenza della classificazione (per migliorare le prestazioni del sistema di identificazione dei tessuti attraverso l’esclusione di risultati di classificazione incerti).

Partendo da questi presupposti, l’obiettivo generale di questa tesi di dottorato è stato quello di sviluppare algoritmi di SDS basati su metodi di ML per classificare, in maniera robusta e affidabile, tessuti biologici a partire da immagini ottiche acquisite intraoperativamente. In particolare, i contributi di questo lavoro di dottorato sono:

1. **Sviluppo di un nuovo metodo (M1) per la selezione automatica di frame informativi da video di endoscopia.** M1 sfrutta l’apprendimento supervisionato e un nuovo set di feature per identificare frame informativi, dai quali verranno classificati i tessuti di interesse. M1 sfrutta caratteristiche di intensità e contenuto spaziale dell’immagine per classificare, con macchine a vettori di supporto (support vector machine, SVM) e secondo approcci a classi multiple, i frame endoscopici non informativi, sfocati, sottoesposti, e/o con riflessioni speculari o presenza di saliva. M1 è stato tes-
tato su un set bilanciato di 720 immagini estratte da 18 video laringoscopici in modalità narrow band, ottenendo una sensitività del 91% per i frame informativi e superando gli approcci dello stato dell’arte (sensitività mediana = 57%, massima = 81%) con evidenza statistica.

2. **Sviluppo di un nuovo metodo (M2) per la classificazione di tessuti a partire da immagini endoscopiche, che misura la confidenza di classificazione al fine di escludere risultati non affidabili incrementando l’accuratezza del risultato fornito ai chirurghi**

Al fine di identificare tessuti patologici, M2 sfrutta informazioni di texture, approcci supervisionati e semi-supervisionati ed un metodo per la stima della confidenza sulla classificazione. M2 è stato testato su 33 video di laringoscopia in modalità narrow band, registrati da 33 pazienti affetti da tumore laringeo in studio iniziale, ottenendo una sensitività mediana del 93%. Il metodo è stato testato anche su immagini del distretto epatico, ottenendo una sensitività dell’82% nella classificazione di 40 immagini RGB acquisite in sala operatoria raffiguranti fegati di donatori sani e con steatosi epatica. Stimando la confidenza della classificazione ottenuta con le SVM, ed escludendo i risultati con bassa affidabilità, la sensitività è stata incrementata fino a raggiungere il 98% (tessuto laringeo) e l’86% (tessuto epatico).

3. **Sviluppo di un nuovo protocollo (M3) per l’analisi di immagini endoscopiche, che sfrutta l’imaging multispettrale ( multispectral imaging, MI) per l’acquisizione di immagini durante procedure endoscopiche con l’obiettivo di eseguire classificazione dei tessuti ed tagging delle immagini in maniera robusta**

M3 è un’estensione di M2 in cui, per la prima volta in letteratura, si studia l’acquisizione e l’analisi di immagini ottenute in vivo sfruttando il MI. I dati multispettrali sono stati acquisiti durante procedure laparoscopiche su 7 suini. Le SVM con approccio a classi multiple sono state utilizzate per classificare sei organi addominali (parete addominale, fegato, milza, cistifellea, diaframma e intestino). I risultati ottenuti con l’utilizzo delle immagini multispettrali hanno mostrato un aumento di accuratezza del 11% nella classificazione dei tessuti e del 23% nel tagging automatico delle immagini, rispetto all’utilizzo di immagini standard di tipo RGB. Inoltre, con l’utilizzo della misura di confidenza proposta in M2, è stato riscontrato un aumento dell’accuratezza della classificazione del 38% (RGB) e 20% (MI).

4. **Integrazione delle metodologie basate su tecniche di ML nell’ambito della SDS all’interno di un sistema di chirurgia robotica al fine di evitare tessuti critici durante procedure simulate di chirurgia robotizzata (M4)**

Con M4, la segmentazione di tessuti basata su tecniche di ML è stata integrata in uno scenario simulato di chirurgia assistita da robot, utilizzando un controllo di tipo virtual fixture per impedire l’interazione indesiderata tra tessuto e robot. In particolare, sono state costruite strutture vascolari fantoccio, successivamente segnate in immagini di microscopia ottica utilizzando un approccio di apprendimento profondo (deep learning). Includendo il controllo, la distanza misurata fra la posizione della punta del robot e il punto più vicino della superficie vascolare (quando il robot entrava nella zona proibita (24% delle volte)) è stato trascurabile.
Infatti, l’errore di penetrazione mediano misurato è stato di 17 \( \mu m \), corrispondente a 2 ordini di grandezza in meno rispetto al diametro mediano dei vasi.

Questa tesi di dottorato si colloca nel piú ampio campo di ricerca della SDS, la quale mira a fornire ai chirurghi supporto decisionale e informazioni sul contesto chirurgico durante procedure di medicina interventistica. Nello specifico, le metodologie presentate in questa tesi di hanno dimostrato l’efficacia dell’uso di algoritmi di SDS basati sul ML per la classificazione intraoperatoria di tessuti in diversi distretti anatomici. I progressi metodologici compiuti in questo lavoro evidenziano il potenziale delle tecniche di SDS basate sul ML per l’estrazione di informazioni implicitamente contenute nelle immagini ottiche intraoperatorie, superando le sfide tipiche della classificazione tissutale in sala operatoria, con il fine di supportare i chirurghi durante le procedure di medicina interventistica.
Acknowledgments

There is nothing certain, but the uncertain. Nonetheless, I think I can say with certainty that this thesis exists thanks to the support of a group of people I would like to thank.

First of all, I thank Elena and Leo for being a constant example of perseverance and for having taught me that a good researcher is not the one who never falls, but the one who, after falling, finds the strength to get back up. Your passion for research has kept mine alive in the most difficult moments of this PhD.

I do really thank Professor Caldwell and Professor Ferrigno for allowing me to carry out my research activity in their prestigious laboratories.

A heartfelt thanks goes to Professor Lena Maier-Hein, to Sebastian and to the whole CAMI group of the DKFZ in Heidelberg. Thanks also to Professor Cameron Riviere and the group of the Robotic Institute of the CMU in Pittsburgh. You are an inexhaustible source of inspiration and knowledge for me, and working with you is a privilege.

I truly thank Dr. Laborai, together with Dr. Guastini, Dr. Peretti, Dr. Cesaretti, Dr. Kengott, Dr. Apitz, Dr. Mayer, Dr. Perin, Dr. Sekula and all the doctors with whom I have collaborated in these three years. Your wise advices guided my research.

An already nostalgic thank you goes to Andrea, Stefano, Matteo R., Alperen, and to Cheng, Francesca, Lorenza, Lara, Marco, Olmo, Joao, André, Giacinto, Jacopo, Lucia, Matteo S., Tommaso, Veronica and the whole group (present and past) of the Biomedical Robotics Lab at Istituto Italiano di Tecnologia for the breakfasts, confessions, advices and fraternal affection (I will always keep you in my heart).

I thank with the same intensity Hiren, Francesca, Alice, Jacopo, Alberto, Marco, Davide, Sara and the whole group of the NearLab at Politecnico di Milano. Our (not always scientific) discussion always canceled the physical distance between our desks.

I thank Giampaolo for the schnitzel, bretzels and the hot chocolates in which we have many times drowned doubts about our future, and Inna for our breaks under the snow of Heidelberg.

From the depth of my heart, I thank Anna, Arianna, Gabriele, Sara and Carlo. You have been my second family for eight years, petulant and morbid but above all warm and irreplaceable.

A big thank you goes to Fabio, Rossana, Annamaria e Alessandra, and to all the
friends of a lifetime: you are for me a fixed point when everything else around keeps changing.

I thank my flatmates Valeria, Zaira, Marco and Sep for having alleviated the most difficult evenings of this PhD.

I thank all the PhD students I met around the world: finding part of yourself in other people is amazing.

A special thanks goes to Francesca, Matteo, Marco, Simone, Gabriele, Michela, Ilaria, Akash, Anna, Michele, Riccardo and all the students I had the pleasure to supervise during these three years: “In learning you will teach, and in teaching you will learn”.

I deeply thank Fabiola, Gianmarco, the dear Claudio (proof that special people can leave a sign on this planet even if their time is limited), all the Tarantini’s / Bagnato’s, for welcoming me as one of the family.

Thanks to aunt Caterina and all the uncles/aunts, and to Alessio, Antonella and Masha: through your eyes I have always seen myself stronger and more intelligent than through mine.

I thank grandpa Pasquale and grandpa Franco, grandmother Giovanna and grandmother Sara. I know that (somewhere) you are proud of me.

The last thanks goes to my true love Daniele, and to mum, dad and Giovanna, for everything and for being my everything.

I love you!

Let’s start!
Un ricercatore ben sa che nulla di certo esiste al mondo. Tuttavia, aihmé, mi sento di dire con piena certezza che questa tesi non sarebbe esistita senza il supporto piú o meno esplicito di alcune persone. A queste persone voglio dire grazie.

Ringrazio Elena e Leo per essere stati costante esempio di perseveranza e per avermi insegnato che un buon ricercatore non é colui che non cade mai, ma colui che, dopo essere caduto, trova la forza di rialzarsi. La vostra passione per la ricerca ha mantenuto viva la mia nei momenti piú difficili di questo dottorato.

Ringrazio il professor Caldwell e il professor Ferrigno per avermi permesso di svolgere la mia attività di ricerca nei loro prestigiosi laboratori.

Un sentito grazie va alla professoressa Lena Maier-Hein, a Sebastian e a tutto il gruppo CAMI del DKFZ di Heidelberg. Un grazie anche al professor Cameron Riviere e al gruppo del Robotic Institute della CMU di Pittsburgh. Siete per me inesauribile fonte di ispirazione e conoscenza e lavorare con voi é un privilegio.


Un ringraziamento (giá velato di nostalgia) va ad Andrea, Stefano, Matteo R., Alperen, e a Cheng, Francesca, Lorenza, Lara, Marco, Olmo, Joao, André, Giacinto, Jacopo, Lucia, Matteo S., Tommaso, Veronica e a tutto il gruppo (presente e passato) del Biomedical Robotics Lab dell’Istituto Italiano di Tecnologia per le colazioni, le confessioni, i consigli e l’affetto fraterno (vi terro’ sempre nel mio cuore).

Ringrazio con la stessa intensité Hiren, Francesca, Alice, Jacopo, Alberto, Marco, Davide, Sara e tutto il gruppo del Nearlab del Politecnico di Milano. Le nostre chiacchierate (scientifiche e non) hanno sempre annullato la distanza fisica fra le nostre scrivanie.

Ringrazio Giampaolo per le schnitzel, i bretzel e le cioccolate calde in cui abbiamo mille volte affogato i dubbi sul nostro futuro, e Inna per le nostre pause sotto la neve ad Heidelberg.

Come non ringraziare Anna, Arianna, Gabriele, Sara e Carlo? Siete la mia seconda famiglia da otto anni, petulanti e morbosi ma soprattutto affettuosi e insostituibili.
Un grazie grande va a Fabio, Rossana, Annamaria, Alessandra, e a tutti gli amici di una vita: siete per me un punto fisso quando invece tutto intorno continua a cambiare.

Ringrazio i miei coinquilini Valeria, Zaira, Marco e Sep per aver alleviato con una cena sempre pronta e discorsi leggeri (ma non vuoti) le sere più difficili di questo dottorato.

Ringrazio tutti i dottorandi conosciuti in giro per il mondo: é bello riscoprire sé stessi in mille altri volti.

Un grazie speciale va a Francesca, Matteo, Marco, Simone, Michele, Gabriele, Michela, Ilaria, Akash, Anna, Riccardo e tutti gli studenti che ho avuto il piacere di seguire durante questi tre anni: “In learning you will teach, and in teaching you will learn”.

Ringrazio Fabiola, Gianmarco, il tanto caro Claudio (prova che le grandi persone non hanno bisogno di tanto tempo per lasciare un segno che non si cancella) e tutti i Tarantini/Bagnato per avermi accolto come una di famiglia.

Grazie a zia Caterina e a tutti gli zii, e ad Alessio, Antonella e Masha: attraverso i vostri occhi mi sono sempre vista più forte e intelligente che attraverso i miei.

Ringrazio nonno Pasquale e nonno Franco, nonna Giovanna e nonna Sara. So che da qualche parte siete fieri di me.

L’ultimo ringraziamento va al mio vero amore Daniele e a mamma, papa’ e a Giovanna per tutto e per essere il mio tutto.

Vi voglio bene!

Incominciamo!
Contents

1 Introduction 1
   1.1 Problem statement and motivation 1
   1.2 Introduction to surgical data science (SDS) 1
   1.3 Overview on the current status in SDS for tissue classification 5
   1.4 Aim of the thesis 7
      1.4.1 Structure of the thesis 7

2 Learning-based classification of informative frames with applications in laryngoscopy 11
   2.1 Introduction 12
   2.2 Related work 13
      2.2.1 Video clustering and keyframe extraction 14
      2.2.2 Uninformative frame removal 14
   2.3 Methods 16
      2.3.1 Feature extraction 16
      2.3.2 Classification 19
   2.4 Evaluation 21
      2.4.1 Experimental setup 22
   2.5 Results 24
   2.6 Discussion 29
   2.7 Conclusion 30

3 Texture-based tissue classification with confidence estimation for computer-assisted diagnosis 33
   3.1 Introduction 34
   3.2 Methods 35
      3.2.1 Feature extraction 35
      3.2.2 Classification 37
      3.2.3 Confidence estimation 37
   3.3 Application 1 - Laryngeal district 38
      3.3.1 Evaluation protocol 39
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.2 Results</td>
<td>44</td>
</tr>
<tr>
<td>3.3.3 Discussion</td>
<td>47</td>
</tr>
<tr>
<td>3.3.4 Conclusion</td>
<td>49</td>
</tr>
<tr>
<td>3.4 Application 2 - Hepatic district</td>
<td>49</td>
</tr>
<tr>
<td>3.4.1 Extension to semi-supervised classification</td>
<td>52</td>
</tr>
<tr>
<td>3.4.2 Evaluation</td>
<td>55</td>
</tr>
<tr>
<td>3.4.3 Results</td>
<td>58</td>
</tr>
<tr>
<td>3.4.4 Discussion</td>
<td>63</td>
</tr>
<tr>
<td>3.4.5 Conclusion</td>
<td>64</td>
</tr>
<tr>
<td>4 Uncertainty-aware organ classification in laparoscopy</td>
<td>65</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>66</td>
</tr>
<tr>
<td>4.2 Related work</td>
<td>66</td>
</tr>
<tr>
<td>4.3 Methods</td>
<td>69</td>
</tr>
<tr>
<td>4.3.1 Uncertainty-aware tissue classification</td>
<td>69</td>
</tr>
<tr>
<td>4.3.2 Automatic image tagging</td>
<td>73</td>
</tr>
<tr>
<td>4.4 In vivo validation</td>
<td>73</td>
</tr>
<tr>
<td>4.5 Results</td>
<td>75</td>
</tr>
<tr>
<td>4.6 Discussion</td>
<td>77</td>
</tr>
<tr>
<td>4.7 Conclusion</td>
<td>81</td>
</tr>
<tr>
<td>5 Toward improving safety in neurosurgery with an active handheld instrument</td>
<td>83</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>84</td>
</tr>
<tr>
<td>5.2 Related work</td>
<td>85</td>
</tr>
<tr>
<td>5.3 Methods</td>
<td>85</td>
</tr>
<tr>
<td>5.3.1 Micron architecture</td>
<td>86</td>
</tr>
<tr>
<td>5.3.2 Vessel segmentation algorithm</td>
<td>86</td>
</tr>
<tr>
<td>5.3.3 3D reconstruction algorithm</td>
<td>89</td>
</tr>
<tr>
<td>5.3.4 Forbidden-region virtual fixture</td>
<td>90</td>
</tr>
<tr>
<td>5.4 Evaluation protocol</td>
<td>91</td>
</tr>
<tr>
<td>5.4.1 Parameter settings and implementation</td>
<td>92</td>
</tr>
<tr>
<td>5.4.2 Evaluation dataset and metrics</td>
<td>94</td>
</tr>
<tr>
<td>5.5 Results</td>
<td>98</td>
</tr>
<tr>
<td>5.6 Discussion</td>
<td>98</td>
</tr>
<tr>
<td>5.7 Conclusion</td>
<td>99</td>
</tr>
<tr>
<td>6 Conclusion</td>
<td>101</td>
</tr>
<tr>
<td>6.1 Thesis Contributions</td>
<td>102</td>
</tr>
<tr>
<td>6.2 Impact</td>
<td>105</td>
</tr>
<tr>
<td>6.3 Future Perspective</td>
<td>107</td>
</tr>
<tr>
<td>Bibliography</td>
<td>127</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction

1.1 Problem statement and motivation

Intra-operative tissue classification plays a fundamental role in different clinical fields, such as laryngology [180], dermatology [76], oncology [31], ophthalmology [30], and neurosurgery [57], both for diagnosis, treatment planning and execution, and for treatment outcome evaluation and follow up. The importance of intra-operative tissue classification is supported by the gradual but constant introduction in clinical practice of medical technologies aimed at enhancing the visualization of tissues, such as multispectral [215], narrow-band [126], and spectroscopy imaging [55].

The automatic image-based classification of tissues is a valuable solution to provide surgeons with decision support or context awareness intra-operatively. Despite the fact that automatic or semi-automatic tissue classification approaches have been proposed by the surgical data science (SDS) community, several technical challenges are still present, hampering the clinical translation of the proposed methodology in the clinical practice. Indeed, performing robust and reliable classification is not trivial due to the high inter- and intra-patient variability, and to noise in the images, varying illumination level, and changing camera pose with respect to the tissues. In this context, the aim of this PhD thesis is to investigate machine-learning (ML) strategies to provide robust, reliable and accurate tissue classification for decision support and context awareness during interventional-medicine procedures.

1.2 Introduction to surgical data science (SDS)

SDS is an emerging interdisciplinary field that aims at extracting knowledge from interventional-medicine data. The goal of SDS is improving the quality of interventional healthcare through the organization, analysis, and modeling of the data [133].
Chapter 1. Introduction

![Diagram](image)

**Fig. 1.1:** Main blocks of interventional care processes. Images adapted from [222].

Interventional medicine deals with minimally-invasive treatment in several surgical fields (e.g., surgery, interventional radiology, gastroenterology, radiotherapy). In the last decades, these fields rapidly advanced, becoming to a greater and greater extent computer-based [222]. Imaging devices have been introduced in the operating room (OR), combining advanced sensors and increased computational power to provide more and more accurate information of patient’s anatomy and physiology [217]. Robotic systems and endoscopic cameras today allow performing minimally invasive procedures, improving post-operative patient’s prognosis and quality of life [68, 223]. Nowadays, as highlighted in [222] and shown in the schematic in Fig. 1.1, the surgeon’s decision process combines (i) qualitative analysis of patient-specific information from imaging systems and sensors and (ii) surgeon’s prior knowledge about medical rules and statistics. Such information is used to build an implicit patient’s model and define the surgical plan. The plan can be updated in the OR, exploiting intra-operative imaging and sensors. After-surgery, surgical outcomes are evaluated and statistically analyzed to improve treatment effectiveness and eventually change treatment protocol.

As a natural result of the massive introduction of imaging devices, sensors and robotic systems in the OR, almost unlimited amount of electronic patient records are available. Such data can be processed to further increase safety, effectiveness, and efficiency of surgical care. Moreover, as observed in [8, 9, 77], the Internet-of-things revolution elected the healthcare domain as one of the most promising field, with infinite opportunities arising from data sharing among hospitals, care-givers and patients. Indeed, data sharing can provide the surgeons with statistics from other patients shared among several care centers, and this information can integrate the patient-specific (local) data (Schematic in Fig. 1.2).

However, only a fraction of patient-related data is digitized and stored in a structured and standardized way, and data quality assessment is rarely performed [133]. This is probably the main reason why SDS only recently emerged as an active field of research. While shared databases are available for other research fields for advancing research (e.g. the PASCAL dataset for natural-image classification [59]), annotated datasets for the SDS community are still limited in number. This can be attributed to regulatory and sociological factors (e.g. data protection and privacy issues) [220]. A second factor deals with medical data annotation, which is typically an expensive process in terms of resources and time [40]. In the last years, several efforts have been made by the SDS community to support research-data sharing and develop crowd-algorithms for large-
1.2. Introduction to surgical data science (SDS)

Medical-data sharing influences the interventional care processes. With a focus on imaging data, data sharing is especially supported by international organizations, such as the MICCAI society, the IEEE Signal Processing Society and the IEEE Engineering in Medicine and Biology Society, which yearly organize Grand Challenges[1] and release annotated dataset for algorithm testing.

Apart from the problem of data sharing, a further SDS challenge is the integration of quantitative data analysis into the actual clinical practice [50, 138, 202]. In particular, the goal is developing adequate data-analysis technology to provide surgeons with quantitative support and effectively translate the technology into patient care workflows. As shown in the schematic in Fig. 1.3 SDS plays an important role in moving from (surgeon-specific) subjective to (computer-assisted) objective decision-making and from qualitative to quantitative assessment of surgical outcomes [133]. The integration of computer-aids facilitates the surgeon’s decision process [41], improved situation awareness [173], improved ergonomics [141] and reduced cognitive workload [32]. Thus, today the technological innovation in the OR aims at combining surgeons’ expertise with the computer ability of integrating and processing information to substantially improve surgical outcomes [65].

SDS methods and principles heavily builds on ML [101, 133]. The medical domain-specific knowledge can be encoded in a ML-based model through a learning process based on the description of cases solved in the past. The model can:

- Offer decision support [142], e.g. by assisting the physician when diagnosing new patients to improve the diagnostic speed, accuracy and/or reliability
- Provide context awareness [94], e.g. for autonomous assistance and collaborative robots to improve safety, quality, and efficiency of care

Here, as shown in Fig. 1.4, technical challenges include tissue tracking [216], 3D reconstruction [131], intra-operative registration [213], workflow modeling [142] and

---

1https://grand-challenge.org/all_challenges/
Chapter 1. Introduction

Fig. 1.3: Surgical data science integration in the interventional-medicine flow allows objective decision-making and quantitative evaluation of surgical outcomes.

Fig. 1.4: Some of the major opportunities that surgical data science offers to interventional medicine [133]. Blocks highlighted in green identify the main topics of this PhD thesis.
1.3. Overview on the current status in SDS for tissue classification

ML models for tissue classification typically (i) apply automated image analysis to extract a vector of quantitative features to characterize the relevant image content and (ii) apply a pattern classifier to determine the category to which the extracted feature vector belongs (e.g., malignant / healthy tissue) [103, 236].

Fig. 1.5 shows the most commonly adopted features in the field of tissue classification. Intensity-based features aim at encoding information related to the prevalent intensity components in the image. Intensity features are mainly based on intensity histogram, mean, variance and entropy [205] and are commonly combined with textural features, which encode tissue appearance, structure and arrangement [33]. Textural features include local binary pattern (LBP) [78], gray-level co-occurrence matrix (GLCM) [82], and histogram of oriented gradients (HOG) [63]. This class of features has been successfully used for several applications, such as tissue classification in gastric [118, 209] and colorectal images [67, 152]. Other popular features are obtained with filtering-based approaches. Filtering approaches build template-filters that correlates with tissue structure in the image. Common approaches are matched-filtering [85] and wavelets [232], which have been widely used for vessel detection and localization [62] and polyp classification [130]. Similarly, derivative-based approaches build derivative-filters to extract image spatial derivatives, such as gradient and Laplacian, e.g. to highlight tissue edges [29, 106]. Recently, learned features have been proposed and successfully used for tissue classification [58, 183]. Learned features refer to features that are automatically extracted from the image [166]. The most popular approach to automatic feature learning is using convolutional neural network (CNN), which showed remarkable performance in classifying skin cancer [58] and predict cardiovascular risk factors from retinal fundus photographs [183].

As for pattern classifiers, several solutions have been introduced in the last decades...
Chapter 1. Introduction

Fig. 1.6: Most popular approaches to supervised learning for tissue classification. The taxonomy is partially inspired by [102].

and an extensive review can be found in [102]. The most popular ones are shown in Fig. 1.6. First attempts were based on statistical approaches (i.e. naive Bayes (NB) [144], Bayesian networks [64] and Bayes classifiers [189]) and instance-based learning [47]. Applications for tissue classification include [16, 164, 206, 231]. Similarly, perceptron-based algorithms [194], have been widely used, e.g. for polyp detection in endoscopic images [93, 130, 208]. Logic-based algorithms [196, 235] and support vector machines (SVM) [42] are probably among the most widely used classifiers. These algorithms showed promising performance for tissue classification in several fields (e.g., [46, 71, 86, 88, 164, 200, 248]). More recently, deep learning for tissue classification drew the attention of the SDS community. Examples include skin-cancer classification [58], polyp detection [21], retinal image analysis [183], and vessel segmentation [62], where large and labeled datasets are publicly available for deep-learning model training.

As introduced in Sec. 1.1, despite a wide literature on tissue classification already exists, open technical challenges are still present [133]. Main challenges deal with the improvement of classification accuracy, robustness and reliability. Indeed, tissue classification can be challenging due to the high variability in patients’ anatomy (especially for pathological cases), imaging devices and acquisition protocols. Moreover, challenges associated with in vivo datasets include (i) wide range of illumination, (ii) variation of the camera pose with respect to the tissue, (iii) presence of specular reflections, (iv) presence of multiple organs in one image, (v) organ movement.

There are several aspects that can be tackled to potentially achieve the goal of robust and reliable tissue classification. The first aspect deals with hardware design. Indeed, the imaging field is constantly evolving thanks to new optical imaging technologies, such as narrow-band imaging (NBI) [198] and multispectral imaging (MI) [116]. These technologies potentially allow high-quality optical imaging (e.g., in terms of image noise and tissue-background contrast) and have already found interesting applications in the remote-sensing field [246]. However, the use of these technologies is still under-represented in the medical field.

A second aspect is related to the identification of images to be processed [179]. High noise level in the image, camera movements, tissue deformation and illumination
1.4. Aim of the thesis

Drop lower image quality and make the classification challenging also for the human eye. Similarly, classification algorithms are prone to error when processing uninformative frames. Solutions have been proposed in the literature, nonetheless the proposed methodologies still lack of robustness to patient’s and acquisition protocol variability. A third point deals with estimating the level of classification confidence with a view to improving system performance. This aspect has been widely highlighted in several research fields, such as face recognition [172] and spam-filtering [49]. In particular, it has been reported that allowing a system to produce “don’t know” results can potentially reduce the number of incorrectly classified cases [145]. However, the introduction of confidence estimation in the medical imaging field has been only marginally explored.

More generally, as SDS/ML strongly rely upon labeled data, the last aspect is related to the availability of labeled datasets. Indeed, the larger the training dataset, the bigger the chances the classification algorithm will be accurate in classifying unseen data. While the development of tissue-classification algorithms is strongly advancing in some specific fields (e.g. vascular district [62], and gastrointestinal tract [113]), there are other fields that are incredibly underrepresented in the literature. The most probable reason for this is indeed the lack of large and available labeled databases for algorithm training.

1.4 Aim of the thesis

The overall goal of this PhD thesis is to develop a framework for ML-based SDS algorithms for tissue classification in optical images, as to offer decision support and provide context awareness during interventional medicine processes.

In particular, starting from the analysis of the main open technical challenges in the field, the research hypotheses that guide the PhD work can be summarized as follows:

1. Hypothesis 1 (H1): ML supports the selection of informative frames to be processed by tissue classification algorithms
2. Hypothesis 2 (H2): Reliable tissue classification can be performed by exploiting ML and classification-confidence estimation
3. Hypothesis 3 (H3): Tissue classification with ML and confidence estimation can be further improved by using multispectral-imaging data
4. Hypothesis 4 (H4): ML-based tissue segmentation can be integrated in the flow of a simulated robot-assisted surgery, as to provide forbidden-region avoidance

The methodology developed to investigate the hypotheses will be described in this PhD dissertation. Sample applications will be proposed for experimentally testing the hypotheses.

1.4.1 Structure of the thesis

This dissertation is organized as follows:
Chapter 1. Introduction

Chapter 2 focuses on H1 and describes a new method (M1) to automatic learning-based informative frame selection, which exploits a new set of features to retrieve informative frames for tissue classification. The algorithm is tested on laryngeal endoscopic videos in NBI.

In Chapter 3, H2 is investigated and a new method (M2) for reliable pathological-tissue classification is proposed. The inclusion of classification confidence is proposed to discard low-confidence results and improve classification performance. Supervised and semi-supervised ML approaches are investigated. M2 is tested on (i) NBI laryngeal images and (ii) hepatic RGB images acquired in the OR.

Chapter 4 focuses on H3. A new protocol (M3) for image analysis is proposed, which expands M2 and exploits MI data acquisition for automatic tissue classification and image tagging. M3 is tested on laparoscopic abdominal images.

In Chapter 5, H4 is investigated. A new method (M4) is presented, which integrates a deep-learning strategy to tissue segmentation within a handheld robotic tool, with the goal of performing tissue avoidance in simulated robotic-assisted surgical procedures. M4 is experimentally tested on phantom neurovascular structures.

In Chapter 6, the conclusion, scientific and clinical implications, and future perspective of this PhD work are reported and discussed.

The graphical abstract of this PhD thesis is shown in Fig. 1.7.
1.4. Aim of the thesis

Fig. 1.7: Graphical abstract of the PhD thesis. Horizontal arrows indicate the anatomical region analyzed to experimentally test the hypotheses (H1, H2, H3, H4) that guide this PhD work. 

H1: Machine learning (ML) supports the selection of informative frames to be processed by tissue classification algorithms. 

H2: Reliable tissue classification can be performed by exploiting ML and classification-confidence estimation. 

H3: Tissue classification with ML and confidence estimation can be further improved by using multispectral-imaging data. 

H4: ML-based tissue segmentation can be integrated in the flow of a simulated robot-assisted surgery, as to provide forbidden-region avoidance.
CHAPTER 2

Learning-based classification of informative frames with applications in laryngoscopy

This Chapter describes a new method (M1) to classify informative endoscopic frames, to be processed by tissue classification algorithms. Indeed, frame selection strategies can benefit classification algorithms by lowering the overall amount of computational power required, and avoiding the processing of frames that do not show structures of interest. Several attempts to automatic frame selection can be found in the literature, but the proposed methodologies heavily rely on threshold-sensitive approaches and/or require intensive parameter tuning. By exploiting the generalization power of machine learning, the algorithm described in this Chapter is robust to different illumination level and noise in the images, and to variability in patients’ anatomy. The algorithm exploits intensity, keypoint and image spatial-content features. Multi-class support vector machines with the radial basis function and the one-versus-one scheme are used to classify frames as informative, blurred, with saliva or specular reflections, and underexposed. The algorithm is tested on 18 NBI laryngoscopic videos, which refer to 18 different subjects, acquired in the actual clinical practice.


The dataset of NBI laryngeal images used in this Chapter was released and it is now available online at https://zenodo.org/record/1162784#.WnCcQJOdUWo.
Chapter 2. Learning-based classification of informative frames with applications in laryngoscopy

Fig. 2.1: Visual examples of laryngeal video frames. (a) Informative frame (I); (b) Blurred frame (B); (c) Frame with saliva and specular reflections (S); (d) Underexposed frame (U).

2.1 Introduction

The relevance of visual analysis of tissues for screening purposes has led in the last decades to the development of new optical-biopsy techniques, such as NBI endoscopy [180]. It is recognized that, from the clinician’s side, reviewing an endoscopic video is a labour-intensive operation [201]. While focusing on particular structures during the video examination, clinicians may miss important clues indicating suspicious conditions (e.g., early-stage tumors). This process could be further compromised by the presence of uninformative video portions, which prolong the revision time of the endoscopic video. Developing a strategy to select informative frames has the potential to reduce the amount of data to review, lowering the surgeons’ workload.

The selection of informative frames can be beneficial also for computer-assisted diagnosis (CAD) algorithms for surgeon decision support, and for SDS algorithms that provide context awareness by identifying tissue and instruments in the camera field of view. Indeed, these algorithms usually rely on manual frame selection so that the frames to be processed show clearly structures of interest or, at least, have sufficient quality (e.g. in terms of low noise-level and appropriate illumination) to perform the diagnosis. Frame selection strategies can benefit the algorithms by (i) lowering the amount of computational power required, and (ii) avoiding the processing of frames that do not show structures of interest. Indeed, frames that do not show interesting structures can dilute any further post-processing (such as classification and segmentation) in CAD systems. This has the potential to significantly enhance the performance of the SDS algorithms.

In this Chapter, the problem of robust and automatic classification of informative
frames is addressed with a focus on laryngoscopy applications. Indeed, in the context of laryngoscopic video analysis, the identification of informative frames, such as the one in Fig. 2.1(a), is not trivial [20]. Apart from the well-known challenges associated to endoscopy, such as high camera-noise level in the images, major challenges typical of the laryngeal district include:

- Movement of swallowing muscles and vocal folds, as well as free and varying endoscope pose, which produces blurring in the images (Fig. 2.1(b))
- Presence of specular reflections (SR), due to the smooth and wet laryngeal surface, and saliva (Fig. 2.1(c))
- Varying illumination conditions, resulting in underexposed video frames (Fig. 2.1(d))

As introduced in Sec. 1.4, this Chapter investigates the hypothesis $H_1$ that ML supports the selection of informative frames, allowing to overcome issues related to the definition of threshold values to assess the quality of the image. Instead of focusing on the identification of just one class of uninformative frames, the proposed approach extends the classification process to four classes (as to deal with all the typically encountered types of uninformative frames in laryngeal endoscopic videos) namely:

- Underexposed frames ($U$)
- Frames with saliva or SR ($S$)
- Blurred frames ($B$)
- Informative frames ($I$)

In addition to identify informative frames, being able to identify these classes ($U$, $S$, $B$) of uninformative frames may help in:

1. Processing initially excluded frames to increase frame quality
2. Informing the clinician on the quality of images he/she is acquiring in real-time

In the first case, post-processing algorithms could be used to increase brightness/contrast for underexposed frame or try to extract residual useful information from frames with saliva or SR. In the second one, the clinician could perform corrective actions, e.g. increase the illumination level, move the endoscope slower to minimize motion blur, rinse the endoscope.

This Chapter is organized as follows: Sec. 2.2 surveys the methodologies proposed in the literature for frame selection; the proposed approach to the learning-based classification of informative frames is explained in Sec. 2.3. The materials used and the evaluation protocol are described in Sec. 2.4. Results are presented in Sec. 2.5 and discussed in Sec. 2.6. Major strengths, limitations and future work are summarized in Sec. 2.7 to conclude this Chapter.

### 2.2 Related work

Several attempts to automatic frame selection can be found in the literature. Many of the approaches exploit simple uniform frame sampling to reduce the amount of
Chapter 2. Learning-based classification of informative frames with applications in laryngoscopy

data to process (e.g., [15][234] for bladder endoscopy). Uniform sampling is fast in terms of run-time, but there is no guarantee that informative frames are extracted from all semantically-important video segments. At the same time, for long segments with identical content, a large number of redundant keyframes are selected. Moreover, also uninformative frames can be potentially elected as keyframes.

More advanced state of the art frame selection strategies applied to the endoscopic medical field can be roughly divided into two branches:

2.2.1 Video clustering and keyframe extraction

The goal of this class of algorithm is to cluster video frames with similar informative content, exploiting similarity measures between features extracted from the images.

In [199][230] keyframes are extracted using a keypoint-based approach. A keyframe is extracted if the distance between consecutive frames in the keypoint-space overcomes an user-defined threshold.

In [81][146][245] features based on color, texture and motion displacement are used to identify, with a threshold-sensitive approach, frames with redundant informative content. Instead of using simple thresholding, in [249] linear discriminant analysis (LDA) is applied in the feature space.

In [125], color and edge features are clustered with k-means. From each cluster, a representative frame is arbitrarily extracted as keyframe. In [87], clusters are obtained with uniform sampling, and non-negative matrix factorization is used to extract a keyframe from each cluster.

This class of algorithms potentially brings the advantage of summarizing the video content. Nonetheless, such algorithms do not make any assumptions about the presence of uninformative video portions, which can potentially represent a high percentage of the endoscopic video content without bringing any useful information for diagnosis.

2.2.2 Uninformative frame removal

This class of algorithms aims at evaluating if the content of a frame is of interest for a given application (e.g., its quality is sufficient to appreciate structures of interest).

In [6][19][91], after uniform sampling, uninformative frames are removed if the number of keypoints is lower than a threshold.

The work in [43][157] uses an intensity-based similarity score to assess the degree of images blur. Thresholding is applied to discard low-quality images. Shannon entropy is instead used in [107][174].

In [7][129], frames are clustered as informative and uninformative using features in the image frequency domain and k-means. GLCM-based features and Gaussian mixture model are used in [179].

A more advanced approach to uninformative frame removal, which is also the first attempt at using ML for this aim, has been proposed in [11][12]. The classification process exploits SVM trained on local color histogram features to discriminate between uninformative frames with residual food and potentially informative frames.
Learning-based classification of informative frames with applications in laryngoscopy

Fig 2.2: Workflow of the proposed approach to automatic learning-based classification of informative frames with applications in laryngoscopy.
Table 2.1: Tested feature vectors and corresponding number of features.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blind/referenceless image spatial quality evaluator (BRISQUE) [153]</td>
<td>1</td>
</tr>
<tr>
<td>Variance of the image Laplacian ($\Delta_{VAR}$)</td>
<td>3</td>
</tr>
<tr>
<td>Sobel-Tenengrad focus evaluation function score (TEN) [176]</td>
<td>3</td>
</tr>
<tr>
<td>Image entropy ($\text{entropy}$)</td>
<td>3</td>
</tr>
<tr>
<td>Local variance of the luminance channel intensity ($\text{VAR}$)</td>
<td>3</td>
</tr>
<tr>
<td>Image intensity variance ($G_{VAR}$)</td>
<td>3</td>
</tr>
<tr>
<td>Image histogram ($H$)</td>
<td>3</td>
</tr>
<tr>
<td>Number of detected keypoints ($N_P$)</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>20</strong></td>
</tr>
</tbody>
</table>

2.3 Methods

In this section, the proposed approach to learning-based classification of informative video frames is described. The feature extraction strategy is explained in Sec. 2.3.1, and the classification in Sec. 2.3.2. The workflow of the proposed approach is shown in Fig. 2.2.

2.3.1 Feature extraction

The aim of the classification features is to encode the main distinctive characteristics of the four frame classes introduced in Sec. 3.1. For each class, specific assumptions on the image content can be made. For instance, underexposed frames can be classified according to intensity-based features, since they contain high percentage of dark pixels (Fig. 2.1(d)). Informative frames have higher spatial-frequencies content than blurred frames, due to the presence of sharp edges, such as blood vessels, as can be seen by comparing Fig. 2.1(a) and Fig. 2.1(b). However, the presence of saliva or bubbles in the image creates similar components in the spatial-frequency domain, too. Frames with saliva or SR can be differentiated based on the color domain, as images with such content present high components in the green and blue color channels (Fig. 2.1(c)). In addition to such assumptions, features should be computationally cheap in order to minimize the effort with a view to real-time applications.

The set of features, which is summarized in Table 2.1, consisted of:

- **Blind/referenceless image spatial quality evaluator (BRISQUE):**

  The blind/referenceless image spatial quality evaluator (BRISQUE) \[153\] is a no-reference image quality assessment holistic metric that operates in the spatial domain. To obtain BRISQUE, the image normalized luminance coefficient, i.e. mean-subtracted contrast-normalized luminance pixel values, was first computed. Such coefficient was approximated by the asymmetric generalized Gaussian dis-
2.3. Methods

Distribution (AGGD):

\[
f(x, \alpha, \sigma_l^2, \sigma_r^2) = \begin{cases} 
\frac{\alpha}{(\beta_l+\beta_r)\gamma_B(1/\alpha)} \exp \left( -\frac{x}{\beta_l} \right)^\alpha & x < 0 \\
\frac{\alpha}{(\beta_l+\beta_r)\gamma_B(1/\alpha)} \exp \left( -\frac{x}{\beta_r} \right)^\alpha & x \geq 0
\end{cases}
\]  

(2.1)

where \( \alpha \) is a shape parameter, \( \sigma_l, \sigma_r \) are scale parameters and \( \gamma_B, \beta_l, \beta_r \) depend on \( \alpha, \sigma_l, \sigma_r \), as explained in [153]. \textit{BRISQUE} was computed by regression on the computed AGGD parameters, using a regressor trained on non-distorted natural images as in the original work [153].

- Variance of the image Laplacian (\( \Delta_{VAR} \)):

Since a high percentage of informative content (i.e., sharp edges, such as blood vessels) is encoded in high frequencies in the spatial frequency domain, a measure (\( \Delta_{VAR} \)) based on the Laplacian (\( L \)) of the image \( I \) was used as feature, as suggested in [176] for autofocusing in light microscopy videos. Given \( I \) of size \( M \times N \), \( \Delta_{VAR} \) is computed as:

\[
\Delta_{VAR} = \sum_{m}^{M} \sum_{n}^{N} (L(m,n) - \bar{L})^2
\]  

(2.2)

where \( \bar{L} \) is:

\[
\bar{L} = \frac{1}{MN} \sum_{m}^{M} \sum_{n}^{N} |L(m,n)|
\]  

(2.3)

and \( L \) is obtained by convolving \( I \) with the Laplacian kernel (\( K_L \)):

\[
K_L = \frac{1}{6} \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}
\]  

(2.4)

- Sobel-Tenengrad focus evaluation function (\( TEN \)):

The Sobel-Tenengrad focus evaluation function (\( TEN \)) [176] is another measure typically used in microscopy autofocusing [143], based on the image gradient magnitude value. Being \( G_x, G_y \) the image gradient along the \( x \) and \( y \) direction, respectively, \( TEN \) is defined as:

\[
TEN = \sum_{m}^{M} \sum_{n}^{N} [S(m,n)]^2, \quad \text{for } S(m,n) > T
\]  

(2.5)

where \( T \) is a threshold and:

\[
S(m,n) = \sqrt{[G_x(m,n)]^2 + [G_y(m,n)]^2}
\]  

(2.6)
Chapter 2. Learning-based classification of informative frames with applications in laryngoscopy

As to obtain $G_x$ and $G_y$, $I$ was convolved with the Sobel’s kernel ($K_S$) and its transpose, respectively, where:

$$K_S = \begin{bmatrix}
1 & 0 & -1 \\
2 & 0 & -2 \\
1 & 0 & -1
\end{bmatrix}$$

(2.7)

- **Image entropy (entropy):**
  Image entropy (entropy) is an effective measure of the amount of information in an image [191]. Here, it was used as feature, as suggested in [122] for quality assessment of natural images:

  $$\text{entropy} = -\sum_i h_i \log_2(h_i)$$

(2.8)

where $h_i$ refers to the $I$ intensity histogram counts of the $i \in [0, 255]$ bin.

- **Local variance of the luminance channel intensity (VAR):**
  The frame edge content can be roughly assessed also from variations in the local intensity variance (VAR) [176]. VAR is defined as:

  $$VAR = \frac{1}{MN} \sum_m \sum_n [lv(m, n) - \bar{lv}]^2$$

(2.9)

where:

$$lv(m, n) = \frac{1}{w_x w_y} \sum_i \sum_j [I(m + i, n + j) - \bar{I}_w]^2$$

(2.10)

$$\bar{lv} = \frac{1}{MN} \sum_m \sum_n lv(m, n)$$

(2.11)

and $\bar{I}_w$ is the mean intensity value on the window.

- **Image intensity variance (G_VAR):**
  In addition to the VAR local focus measure, a global intensity variance (G_VAR) focus measure was computed. G_VAR (Eq. 2.12) was used in addition to VAR to improve the feature robustness against noise.

  $$G_{VAR} = \frac{1}{MN} \sum_m \sum_n [I(m, n) - \bar{I}]^2$$

(2.12)

where $\bar{I}$ is the mean intensity of all pixel in $I$. 

18
2.3. Methods

- **Image histogram (H):**
  Inspired by [12] and to include intensity-related features, the first quartile, median and third quartile of the image histogram were added to the feature vector.

- **Number of keypoints (N_P):**
  The number of keypoints in a frame is a trivial measure of information, as suggested in [19]. Here, oriented fast and rotated brief (ORB) [195] was used to detect keypoints in the gray-scale version of I. ORB is a fast binary descriptor, rotation invariant and robust to noise.

\[ \Delta_{VAR}, \text{\textit{VAR}}, \text{\textit{TEN}}, \text{\textit{G}_{VAR}} \text{ and } H \] were computed for each I color channel in the image.

Prior to feature extraction, anisotropic diffusion filtering [147] was used to lower noise while preserving sharp edges in the images [157].

2.3.2 Classification

To perform tissue classification, SVM were used [28]. The SVM kernel-trick prevents parameter proliferation, lowering computational complexity and limiting over-fitting. Moreover, the SVM decisions are only determined by the support vectors, which makes SVM robust to noise in training data. Here, the SVM with Gaussian kernel (\(\psi\)) were used.

For a binary classification problem, given a training set \(T = \{y_t, x_t\}_{t \in T}\), where \(x_t\) is the \(t^{th}\) input feature vector and \(y_t\) is the \(t^{th}\) output label, the SVM decision function \((f)\), according to the “dual” SVM formulation, takes the form of:

\[
f(x) = \text{sign} \left[ \sum_{t \in T} a_t^* y_t \psi(x, x_t) + b \right]
\]

where:

\[
\psi(x, x_t) = \exp\left\{ -\gamma \|x - x_t\|^2 / \sigma^2 \right\}, \quad \gamma > 0
\]

\(b\) is a real constant and \(a_t^*\) is computed as follow:

\[
a_t^* = \max \left\{ -\frac{1}{2} \sum_{k,l=1}^{N} y_k y_l \psi(x_t, x_l) a_t a_l + \sum_{t \in T} a_t \right\}
\]

with:

\[
\sum_{t \in T} a_t y_t = 0, \quad 0 \leq a_t \leq C, \quad t \in T
\]

In this work, the SVM parameters \(\gamma\) and \(C\) were computed with grid search and cross-validation, as explained in Sec. 2.4. To implement multi-class SVM classification, the one-vs-one scheme was used, assigning ambiguous test points to the nearest decision boundary [151]. With the one-vs-one scheme, one binary SVM classifier was constructed for pairs of frame classes. For each binary learner, one class was considered positive, another was negative, and the rest was ignored. This design exhausted all combinations of class pair assignments. At prediction time, the class which received the most votes was selected.

Prior to classification, the feature matrices were standardized.
Table 2.2: Evaluation dataset. For each video (video ID), and for each class (I, B, S, U), the number of frames that contributed to build the dataset are reported. The dataset is split in 3 folds to perform robust estimation of the classification performance. The folds are balanced both at patient- and class-level. I: informative frame; B: blurred frame; S: frame with saliva or specular reflections; U: underexposed frame.

<table>
<thead>
<tr>
<th>video ID</th>
<th>I</th>
<th>B</th>
<th>S</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>0</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>40</td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>total</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Fold 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>28</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>8</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>21</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>total</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Fold 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>17</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>21</td>
<td>34</td>
<td>22</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>0</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>10</td>
<td>0</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>10</td>
<td>12</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>total</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>
2.4 Evaluation

In this study, 18 NBI endoscopic videos, referring to 18 different patients affected by squamous cell carcinoma (SCC), which is the most common cancer of the laryngeal tract \cite{139}, were retrospectively analyzed (average video length: 39s). Videos were acquired with a NBI endoscopic system (Olympus Visera Elite S190 video processor and an ENF-VH rhino-laryngo videoscope) with frame rate of 25 \textit{fps} and image size of 1920 × 1072 pixels.

A total of 720 video frames, 180 for each of the four classes (I, B, S, U) was extracted and labeled from the 18 videos, see Table 2.2. For each video, video frames were randomly extracted and presented to two human evaluators first. Then, the two evaluators were asked to label the frames. In case the two evaluators did not agree on the class, a third evaluator was asked to choose the ultimate class among the two proposed by the two evaluators. This process was repeated until all the 720 frames were extracted from the videos. For the manual labeling process, the following set of rules was defined: I frames should have an adequate exposure and clearly visible blood vessels; they may also present micro-blur and small portions of SR (up to 10% of the image area). B frames should show a homogeneous and widespread blur. S frames should present bright white/light-green bubbles or blobs, overlapping with at least half of the image area. Finally, U frames should present a high percentage of dark pixels, even though small image portions (up to 10% of the image area) with over- or normal exposure are allowed.

In addition, one of the videos was fully labeled (length = 17.64s; number of frames: I = 341, B = 7, S = 9, U = 84).

All the frames underwent the pre-processing step described in Sec. 2.3. The anisotropic diffusion filtering parameters were set as in \cite{147}.

From each frame, the features described in Sec. 2.3.1 were obtained using the following parameters:

- \textit{BRISQUE}: \textit{BRISQUE} code was downloaded from the \textit{Laboratory for Image & Video Engineering} website\footnote{http://live.ece.utexas.edu/research/quality/index.htm} and the parameters were set as in \cite{153}
- \textit{TEN}: the threshold \textit{T} of Eq. 2.5 was set to 0, as suggested in \cite{143} to include all pixels in the computation
- \textit{VAR}: to compute \textit{VAR}, the local window size was 5 × 5 pixels
- \textit{N-P}: the parameters of ORB were set as in the original paper \cite{195}

As for performing the classification presented in Sec. 2.3.2, the SVM hyper-parameters (\(\gamma, C\)) were retrieved via grid-search and 10 fold cross-validation on the training set. The grid-search space for \(\gamma\) and \(C\) was set to \([10^{-7}, 10^{-1}]\) and \([10^{-3}, 10^{3}]\), respectively, with seven values spaced evenly on \(\log_{10}\) scale in both cases.

The feature computation was implemented using OpenCV\footnote{http://docs.opencv.org /3.1.0/index.html}. The classification was implemented with scikit-learn\footnote{http://scikit-learn.org/stable/ index.html}.
Chapter 2. Learning-based classification of informative frames with applications in laryngoscopy

2.4.1 Experimental setup

To obtain a robust estimation of the classification performance of the frames reported in Table 2.2, 3-fold cross-validation was performed, separating data at patient level. Data were separated at patient level to ensure that frames from the same class were classified due to features that are peculiar to that class, and not due to features linked to the patient itself (e.g. vocal-fold anatomy). When the classification of the frames in fold 3 was performed, folds 1 and 2 were used to train the SVM. To retrieve the SVM parameters during the training phase, 10 fold cross-validation and grid-search were performed on the training set (i.e. using images from folds 1 and 2), as explained in Sec. 2.4. The same procedure was carried out for testing the classification of frames in fold 1 and 2, using fold 2 and 3, and fold 1 and 3 for hyper-parameter tuning, respectively.

A balanced dataset (both at patient level and frame class level) was built, as shown in Table 2.2. It can be noticed from Table 2.2 that, for some videos, selecting an equal number of frames for the four classes was not always possible, especially for the uninformative ones. The reason is that either the videos could contribute only with ambiguous frames (i.e. frames with mixed characteristics among the four classes) or a sufficient number of frames was not available for all the classes. When this was the case, the other videos in the fold contributed to balance the number of frames. The approach followed to balance the dataset is common for studies with limited amount of data. A similar approach was followed, for example, in [12].

In order to evaluate the classification performance, the class-specific recall ($Rec_{class,j} = \frac{TP_j}{TP_j + FN_j}$), the precision ($Prec_{class,j} = \frac{TP_j}{TP_j + FP_j}$), and the F1 score ($F1_{class,j} = \frac{2 \cdot Prec_{class,j} \cdot Rec_{class,j}}{Prec_{class,j} + Rec_{class,j}}$), were computed, where:

$TP_j$ the true positive of the $j^{th}$ class, $FN_j$ the false negative of the $j^{th}$ class, and $FP_j$ the false positive of the $j^{th}$ class.

The area (AUC) under the receiver operating characteristic (ROC) curve was also computed. Since the task was a multi-class classification problem and the dataset was balanced, the macro-average ROC curve was computed.

The computational time required to extract and classify the proposed features was computed, as well. Experiments were performed on a CPU Intel® Core™2 Duo @ 2.26GHz with 8GB of available RAM; Linux operative system, kernel 4.4.0-98-generic (x86_64) Ubuntu 16.04.3 LTS distribution.

The use of feature selection was also investigated. Principal component analysis (PCA) [238] was applied to the feature set to retrieve a relevant set of features. Then, the classification explained in Sec. 2.3.2 was performed. For the PCA implementation, principal components were retrieved as to explain the 99% of the variance encoded in the features.
2.4. Evaluation

Table 2.3: Classification performance of the proposed approach. Results are relative to support vector machines (SVM) and random forest (RF) classification on the proposed feature set. SVM results with principal component analysis (PCA) performed on the feature set are reported, too. Class-specific recall ($\text{Rec}_{\text{class}}$), precision ($\text{Prec}_{\text{class}}$), and F1 score ($\text{F1}_{\text{class}}$) are reported for the four different frame classes. I: informative frame; B: blurred frame; S: frame with saliva or specular reflections; U: underexposed frame. Median and inter-quartile range (IQR) of the metrics are reported, too.

<table>
<thead>
<tr>
<th>Proposed feature set and PCA + SVM</th>
<th>I</th>
<th>B</th>
<th>S</th>
<th>U</th>
<th>median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Prec}_{\text{class}}$</td>
<td>0.90</td>
<td>0.77</td>
<td>0.66</td>
<td>0.78</td>
<td>0.78</td>
<td>0.13</td>
</tr>
<tr>
<td>$\text{Rec}_{\text{class}}$</td>
<td>0.89</td>
<td>0.72</td>
<td>0.66</td>
<td>0.84</td>
<td>0.78</td>
<td>0.18</td>
</tr>
<tr>
<td>$\text{F1}_{\text{class}}$</td>
<td>0.90</td>
<td>0.74</td>
<td>0.66</td>
<td>0.81</td>
<td>0.78</td>
<td>0.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposed feature set + RF</th>
<th>I</th>
<th>B</th>
<th>S</th>
<th>U</th>
<th>median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Prec}_{\text{class}}$</td>
<td>0.89</td>
<td>0.76</td>
<td>0.61</td>
<td>0.72</td>
<td>0.74</td>
<td>0.16</td>
</tr>
<tr>
<td>$\text{Rec}_{\text{class}}$</td>
<td>0.78</td>
<td>0.73</td>
<td>0.59</td>
<td>0.86</td>
<td>0.76</td>
<td>0.16</td>
</tr>
<tr>
<td>$\text{F1}_{\text{class}}$</td>
<td>0.83</td>
<td>0.75</td>
<td>0.60</td>
<td>0.78</td>
<td>0.77</td>
<td>0.13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposed feature set + SVM</th>
<th>I</th>
<th>B</th>
<th>S</th>
<th>U</th>
<th>median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Prec}_{\text{class}}$</td>
<td>0.91</td>
<td>0.76</td>
<td>0.78</td>
<td>0.76</td>
<td>0.77</td>
<td>0.09</td>
</tr>
<tr>
<td>$\text{Rec}_{\text{class}}$</td>
<td>0.91</td>
<td>0.83</td>
<td>0.62</td>
<td>0.85</td>
<td>0.84</td>
<td>0.16</td>
</tr>
<tr>
<td>$\text{F1}_{\text{class}}$</td>
<td>0.91</td>
<td>0.79</td>
<td>0.69</td>
<td>0.80</td>
<td>0.80</td>
<td>0.12</td>
</tr>
</tbody>
</table>
Chapter 2. Learning-based classification of informative frames with applications in laryngoscopy

![Confusion matrices](image)

Fig. 2.3: Confusion matrices with (left) and without (right) applying principal component analysis to the proposed feature set. Frame classification was obtained with support vector machines. Matrices refer to the classification of the balanced dataset of 720 narrow-band imaging laryngoscopic video frames. The colorbar indicates the number of images.

For the sake of completeness, the performance of random forest (RF) [27] in classifying the proposed feature set was also investigated and compared with those obtained with SVM. The number of trees in the forest for RF was found with grid-search and cross-validation with a grid-search space set to [40,100] with six values spaced evenly.

The SVM performance were compared also with those obtained using the features commonly exploited in the state of the art. As explained in Sec. 2.2.2, commonly exploited features are (i) image keypoints, (ii) intensity-based similarity scores, (iii) color features and (iv) textural information. Therefore, the proposed method was compared with one research per feature category. Methods in [12,19,43,179] were considered, which use ORB keypoints, intensity-based similarity score, color histogram and GLCM, respectively. The parameters for the state of the art methods implemented for comparison were set as reported in their reference papers: ORB parameters for [19], thresholding values for the intensity-based similarity score for [43], histogram bin number for [12] and orientation and radius for GLCM computation for [179]. As stated in Sec. 2.1, such methods rely on thresholding instead of ML-based methods. However, the features from the state of the art were classified with SVM, for fair comparison.

The Wilcoxon signed-rank test (significance level \( \alpha = 0.05 \)) for paired samples was used to assess whether the classification achieved with the proposed feature vector (reported in Table 2.1) significantly differs from the ones achieved with the state of the art feature sets and with the proposed feature set using PCA. When significant differences were not found, the time required to extract the features was computed. The Wilcoxon signed-rank test was also used to assess if the performances of SVM and RF in classifying the proposed feature set were significantly different. In all cases, the Wilcoxon signed-rank test was performed by comparing the \( \text{Rec}_{\text{class}} \) vectors.

As for classifying the completely labeled video sequence, when training the SVM, all the frames from the three folds (excluding the ones relative to the specific analyzed video) were used, for a total of 689 training frames.

2.5 Results

With the proposed feature set and SVM classification, a median \( \text{Rec}_{\text{class}} = 84\% \) with inter-quartile range (IQR) = 9\% was obtained (Table 2.3 bottom). It is worth noting that misclassification occurred mainly when classifying uninformative frames, while
2.5. Results

Fig. 2.4: Macro-averaging receiver operating characteristic (ROC) curve analysis. ROC were obtained using support vector machines. No principal component analysis for feature reduction was performed. The mean (± standard deviation) curves obtained from the 3 cross-validation folds are reported by the orange solid lines (gray area). The mean (± standard deviation) area under the ROC curve is reported in the legend.

Fig. 2.5: Comparison of the proposed feature set performance with the state of the art feature performance. Classification was performed using support vector machines. No principal component analysis for feature reduction was performed. Boxplots of class-specific recall (Rec\textsubscript{class}) are reported. Stars indicate significant differences (Wilcoxon signed-rank test (significance level $\alpha = 0.05$) for paired samples). Dots indicate outliers.
Chapter 2. Learning-based classification of informative frames with applications in laryngoscopy

Table 2.4: Classification performance of the state of the art features using support vector machines. Class-specific recall ($\text{Rec}_{\text{class}}$), precision ($\text{Prec}_{\text{class}}$), and $F1$ score ($\text{F1}_{\text{class}}$) are reported for the four different frame classes. \textit{I}: informative frame; \textit{B}: blurred frame; \textit{S}: frame with saliva or specular reflections; \textit{U}: underexposed frame. Median and inter-quartile range (IQR) of the metrics are reported, too. The parameters for the state of the art methods implemented for comparison were set as reported in their reference papers: oriented fast and rotated brief (ORB) parameters for [19], thresholding values for the intensity-based similarity score for [43], histogram bin number for [12] and orientation and radius for GLCM computation for [179].

<table>
<thead>
<tr>
<th>Ref. [43]</th>
<th>I</th>
<th>B</th>
<th>S</th>
<th>U</th>
<th>median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Prec}_{\text{class}}$</td>
<td>0.31</td>
<td>0.61</td>
<td>0.30</td>
<td>0.33</td>
<td>0.32</td>
<td>0.09</td>
</tr>
<tr>
<td>$\text{Rec}_{\text{class}}$</td>
<td>0.49</td>
<td>0.41</td>
<td>0.33</td>
<td>0.21</td>
<td>0.37</td>
<td>0.13</td>
</tr>
<tr>
<td>$\text{F1}_{\text{class}}$</td>
<td>0.38</td>
<td>0.49</td>
<td>0.31</td>
<td>0.26</td>
<td>0.35</td>
<td>0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ref. [19]</th>
<th>I</th>
<th>B</th>
<th>S</th>
<th>U</th>
<th>median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Prec}_{\text{class}}$</td>
<td>0.62</td>
<td>0.52</td>
<td>0.16</td>
<td>0.44</td>
<td>0.48</td>
<td>0.18</td>
</tr>
<tr>
<td>$\text{Rec}_{\text{class}}$</td>
<td>0.61</td>
<td>0.90</td>
<td>0.02</td>
<td>0.51</td>
<td>0.56</td>
<td>0.30</td>
</tr>
<tr>
<td>$\text{F1}_{\text{class}}$</td>
<td>0.62</td>
<td>0.66</td>
<td>0.04</td>
<td>0.47</td>
<td>0.55</td>
<td>0.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ref. [12]</th>
<th>I</th>
<th>B</th>
<th>S</th>
<th>U</th>
<th>median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Prec}_{\text{class}}$</td>
<td>0.53</td>
<td>0.59</td>
<td>0.51</td>
<td>0.55</td>
<td>0.54</td>
<td>0.04</td>
</tr>
<tr>
<td>$\text{Rec}_{\text{class}}$</td>
<td>0.38</td>
<td>0.68</td>
<td>0.52</td>
<td>0.61</td>
<td>0.57</td>
<td>0.14</td>
</tr>
<tr>
<td>$\text{F1}_{\text{class}}$</td>
<td>0.45</td>
<td>0.63</td>
<td>0.52</td>
<td>0.58</td>
<td>0.55</td>
<td>0.09</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ref. [179]</th>
<th>I</th>
<th>B</th>
<th>S</th>
<th>U</th>
<th>median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Prec}_{\text{class}}$</td>
<td>0.95</td>
<td>0.60</td>
<td>0.66</td>
<td>0.90</td>
<td>0.78</td>
<td>0.29</td>
</tr>
<tr>
<td>$\text{Rec}_{\text{class}}$</td>
<td>0.95</td>
<td>0.73</td>
<td>0.52</td>
<td>0.90</td>
<td>0.81</td>
<td>0.30</td>
</tr>
<tr>
<td>$\text{F1}_{\text{class}}$</td>
<td>0.95</td>
<td>0.66</td>
<td>0.58</td>
<td>0.90</td>
<td>0.78</td>
<td>0.31</td>
</tr>
</tbody>
</table>
2.5. Results

AUC = 0.89

Sensitivity
1-Specificity

(a) ROC curve

(b) Confusion matrix

Fig. 2.6: Classification performance of the proposed method for a complete video sequence. (a) Macro-averaging receiver operating characteristic (ROC) curve analysis. The area under the ROC curve is reported in the legend. (b) Confusion matrix for a complete video sequence. The colorbar indicates the number of frames. The number of frame for each class is 341 (I, informative frame); 7 (B, blurred frame); 9 (S, frame with saliva or specular reflections); 84 (U, underexposed frame).

informative frames were classified with a recall of 91%. The relative confusion matrix is reported in Fig. 2.5 (right). From the ROC curve analysis (Fig. 2.4 left), a mean AUC of 91% was achieved. The computational time for the proposed feature set computation from one image was \( \sim 0.03 \) s. The classification process took \( \sim 10^{-5} \) s.

When applying PCA (Table 2.3 top, Fig. 2.3 left), no significant differences were found with respect to using non-reduced features (p-value > 0.05). Since SVM performances with and without PCA were comparable, we decided to exclude PCA from the analysis.

When using RF (Table 2.3 middle) to classify the proposed feature set, no significant difference (p-value > 0.05) were found with respect to SVM classification.

When applying the algorithm presented in [43] to the dataset, a mean AUC = 67% was obtained (Fig. 2.4). Worse performance with respect to the proposed approach was achieved also by the method in [19] (mean AUC = 81%) and in [12] (mean AUC = 82%), while [179] achieved a comparable value of AUC = 91%. The \( \text{Rec}_{\text{class}} \), \( \text{Prec}_{\text{class}} \), \( \text{F1}_{\text{class}} \) values for [43, 19, 12] and [179] are reported in Table 2.4. The complete statistics of \( \text{Rec}_{\text{class}} \) relative to the comparison of the proposed method with the state of the art is reported in Fig. 2.5. The proposed approach significantly outperformed [12, 19, 43] (p-value < 0.05). Comparable performances (p-value > 0.05) were instead achieved using GLCM as in [179]. The execution time to extract GLCM-based features from a single image using the scikit-image implementation [229] on the machine described in Sec. 2.4.1 was \( \sim 0.71 \) s.

Results relative to the automatic classification of the complete video sequence with the associated gold standard classification are reported in Fig. 2.6. The ROC curve is reported (AUC = 0.89), as well as the confusion matrix. The 83% of the I frames were correctly classified. To qualitative appreciate the classification results, a visual confusion matrix is shown Fig. 2.7.

Visual samples of the classification performance for four videos are shown in Fig. 2.8.
Chapter 2. Learning-based classification of informative frames with applications in laryngoscopy

**Fig. 2.7:** Visual confusion matrix for an entire laryngoscopic video sequence. Black boxes indicate the absence of misclassification between the true and predicted label. Numbers indicate the percentage of classified frames. **I:** informative frame; **B:** blurred frame; **S:** frame with saliva or specular reflections; **U:** underexposed frame.

<table>
<thead>
<tr>
<th>True Label</th>
<th>I</th>
<th>B</th>
<th>S</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td><strong>0.83</strong></td>
<td>0.0</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>B</td>
<td>0.0</td>
<td><strong>0.86</strong></td>
<td>0.14</td>
<td>0.0</td>
</tr>
<tr>
<td>S</td>
<td>0.0</td>
<td>0.33</td>
<td><strong>0.56</strong></td>
<td>0.11</td>
</tr>
<tr>
<td>U</td>
<td>0.13</td>
<td>0.02</td>
<td>0.31</td>
<td><strong>0.54</strong></td>
</tr>
</tbody>
</table>
2.6. Discussion

Fig. 2.8: Qualitative analysis of the classification outcomes. Each row shows examples of frames from a video sequence. I: informative frame; B: blurred frame; S: frame with saliva or specular reflections; U: underexposed frame.

2.6 Discussion

From the comparison with three state of the art methods, the proposed strategy proved to be a reliable and much better strategy for frame selection with respect to [12,19,43], with statistical evidence. Significant differences were not found when comparing the proposed performance with GLCM-based features [179]. However, it is worth noting that the GLCM computation time (∼0.71 s) for one image was 1 order of magnitude higher than the computation time required to compute the proposed feature set (∼0.03 s). This makes the proposed feature set more suitable for the task of informative frame selection with respect to GLCM-based features considering that time is a constraint with a view to real-time application. This is especially true if one considers that the computational time of CAD systems must be eventually added. Moreover, compared with the state of the art, the proposed learning-based method is simpler, as it eliminates the issue of setting thresholding values (which is required by [19,43,157,179]).

The SVM performance did not vary when applying feature selection with PCA, suggesting that PCA can be avoided to avoid increasing computational time. Significant differences between RF and SVM were not found, as expected considering results reported in the literature [25,44,52].

When testing the proposed approach on a complete labeled video sequence, the misclassifications occurred mainly for challenging frames, which were not trivial to classify also for humans (Fig. 2.7). I frames were never misclassified as B, while misclassification occurred with respect to U frames, especially when the (dark) area between vocal folds covered a large portion of the image, and to S frames, due to the presence in the frame of portions of leukoplakia, which is visually similar to SR. Leukoplakia is a pre-cancerous tissue alteration that implies thickening and whitening of the epithelial layer [90]. It is worth noting that frames with leukoplakia were misclassified as uninformative only when in presence of SR and saliva. An example of a frame that depicts a tissue with leukoplakia and that was correctly classified as informative is shown in Fig. 2.7 (top-left). Nonetheless, misclassification of informative frames is critical, as it could affect the judgment of diagnosis. Training on a larger set, which would include a wider range of laryngeal tissue conditions, should attenuate this issue. Moreover, with
Chapter 2. Learning-based classification of informative frames with applications in laryngoscopy

a larger dataset, more advanced tools may also be investigated, such as CNN, following the current trends in non-medical fields where large labeled dataset are available (e.g. [92][114]).

A limit of the proposed approach could be seen in the dimension of the evaluation dataset. The number of frames of the balanced dataset used to test the proposed approach was limited to ~700 images, to which ~500 frames from the fully labeled video were added. In fact, the manual labeling of frames for training and testing M1 was time consuming. Despite such number being much smaller than those available for the methods used for performance comparison, namely [19] (~3000 images) and [12] (~22000 images), it has the same order of magnitude of other methods in the literature, such as [43] (~300 images). Moreover, if from one side the tested dataset presents a lower intra-patient variability, it is worth noting that the dataset grants a more complete overview on the inter-patient variability, presenting a higher number of patients (18) compared to the 3 of [12], to the 2 of [19], and to the 6 of [157], which exploited the state of the art method in [43]. Therefore, to contribute to global research on laryngoscopic video analysis, the tested dataset was released and it is fully available online at https://zenodo.org/record/1162784#.WnCcQJodUWo.

As introduced in Sec. 2.1, uninformative frames initially excluded by the proposed algorithm could be further processed to increase frame quality, and informing the clinician on the quality of images he/she is acquiring in real-time. In the first case, post-processing algorithms could be used to increase brightness/contrast for underexposed frame (e.g. exploiting simple intensity-histogram equalization) or try to extract residual useful information from frames with saliva or SR (e.g. by automatically detecting SR areas as in [10] and perform texture analysis outside the detected SR areas). In case of slightly blurred frames, deblurring algorithms could be exploited, such [147].

The evaluation protocol was focused on laryngeal video endoscopy, however the proposed methodology can be easily and successfully integrated in the analysis of video of other anatomical districts. Sample applications, for which the presence of underexposed or blurred frames represent an issue, are hysteroscopy, gastrointestinal and coronary video analysis.

2.7 Conclusion

In this Chapter, the challenging topic of informative frame classification was addressed and H1 (Sec. 1.4) was experimentally validated with NBI laryngoscopic videos. The proposed method M1 was retrospectively applied to ~1200 frames from 18 videos of 18 different subjects recorded during the clinical practice. With the exploited experimental protocol, an overall median classification recall of 84% among four frame classes (i.e. blurred, underexposed, with saliva or SR, and informative frames) was achieved. Misclassification mainly occurred between classes of uninformative frames and informative video frames were classified with a recall of 91%. Such performances are significantly higher than those achieved applying other methods in the literature to the evaluation dataset. Moreover, the proposed approach is more robust, faster and simpler to implement since no parameter tuning is required.

It is recognized that future work is required to further ameliorate the algorithm performance. However, the results obtained here are expected to provide major contribu-
tion towards lowering the degree of manual intervention required by computer-assisted systems intended to analyze and summarize the endoscopic video content and increasing their performance.
CHAPTER 3

Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

This Chapter describes a new method (M2) for the classification of laryngeal and hepatic pathological tissues. With respect to the methodologies proposed in the literature in these two fields, the exploited approach tackles the challenges of (i) high intra- and inter-patient variability and (ii) non-standardized image acquisition protocol by exploiting texture analysis and supervised and semi-supervised machine learning. A measure of confidence on classification is included to increase classification reliability. The algorithm is tested on 33 NBI laryngeal videos and 40 RGB hepatic images acquired during the actual clinical practice.

Part of this work has been published as [156]: S. Moccia, E. De Momi, M. Guar naschelli, M. Savazzi, A. Laborai, L. Guastini, G. Peretti and L. S. Mattos, “Confident texture-based laryngeal tissue classification for early-stage diagnosis support,” Journal of Medical Imaging (DOI: [http://dx.doi.org/10.1117/1.JMI.4.3.034502](http://dx.doi.org/10.1117/1.JMI.4.3.034502)).

Part of this work has been submitted to *International Journal for Computer Assisted Radiology and Surgery* as: S. Moccia, L. S. Mattos, N. Poté, F. Dondero, F. Cauchy, A. Sepulveda, O. Soubrane, E. De Momi, A. Diaspro and M. Cesaretti, “Computer-assisted liver-graft steatosis assessment via learning-based texture analysis” and is currently under revision.

The dataset of NBI laryngeal images used in this Chapter was released and it is now
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

available online at https://zenodo.org/record/1003200#.WdeQcnBx0nQ

3.1 Introduction

As introduced in Chapter 1, the emerging and rich literature on SDS for tissue classification is focusing more and more on ML, and several solutions have been proposed in the literature. A first class of algorithms in the field deals with the extraction of handcrafted features, such as LBP and GLCM. The design of handcrafted features often involves finding the right trade-off between classification accuracy and computational efficiency [166]. Examples include [172], where the histogram \(H_{LBP}\) of LBP is exploited to classify ulcer and healthy regions in capsule-endoscopy images using multilayer perceptron. In [248], \(H_{LBP}\) is combined with intensity-based features to classify abdominal tissues in laparoscopic images by means of SVM. Similarly, in [118], intensity-based features and \(H_{LBP}\) are used to characterize lesions in gastric images. In [209], \(H_{LBP}\) is combined with GLCM-based features to classify gastroscopy images. AdaBoost is used to perform the classification. In [228], Gabor filter-based features are used to classify healthy and cancerous tissue in gastroscopy images by means of SVM. A recent work [152] exploits NBI data for colorectal image analysis. Colorectal tissues are classified as neoplastic or healthy by means of GLCM-based features and SVM.

The second class of algorithms deals with features that are automatically learned using a deep-learning approach. In this case, CNN can be exploited either to directly retrieve the classification output or as feature-extraction block. In both cases, features extracted by CNNs strongly depend on the training set. Thus, choosing a representative training dataset is critical and huge training sets are usually required to encode the complexity of the classification problem and avoid overfitting [166]. This is particularly true if one considers that the complexity of pathological tissues is high due to both inter-patient and intra-patient variability. For example, in [58] \(\sim130,000\) clinical images are used to train a CNN for skin cancer classification, while in [183] \(\sim300,000\) images are used for cardiovascular-risk assessment.

Inspired by the recent and promising studies in the literature, in this Chapter the hypothesis \(H_2\) that robust tissue classification can be performed by exploiting ML and classification-confidence estimation is investigated.

In particular, ML approaches based on tissue texture analysis applied to

1. Laryngeal tissue classification in NBI endoscopy

2. Hepatic tissue classification in RGB images acquired with smartphones in the OR

are investigated to test if they can provide reliable results, to be used as support for diagnosis. It is worth noting that estimating the classification confidence in the analyzed context would be particularly beneficial. Indeed, tissue biopsy would be required only for low-confidence regions in the image, lowering risks related to bleeding or infection [180].

To the best of author’s knowledge, these two applications have never been investigated in the literature of SDS so far, with the exception of some preliminary attempts that mostly do not exploit the generalization power of ML. As large and labeled dataset (such as the one used in [58] for skin cancer) are not publicly available for the analyzed
3.2 Methods

Following the classical ML workflow, the proposed approach consists of two main steps: (i) Feature extraction (Sec. 3.2.1) and (ii) Classification (Sec. 3.2.2). A third step is added to compute classification confidence, as explained in detail in Sec. 3.2.3.

3.2.1 Feature extraction

As images are captured under various illumination conditions and from different viewpoints, the features that encode the tissue texture information should be robust to the pose of the camera as well as to the lighting conditions. Furthermore, with a view of a real-time computer-aided application, they should be computationally cheap. In this work, the following descriptors were investigated to characterize pathological tissues:

**Texture-based global descriptors** Among classic texture-based global descriptors, LBP are widely considered as the state of the art for medical image texture analysis [167]. LBP are gray-scale invariant and provide low-complexity, well matching the requisite of CAD applications.

The first formulation of LBP ($LBP_{R,P}^{R,P}$) introduced in the literature requires to define, for a pixel $c = (c_x, c_y)$, a spatial circular neighborhood of radius $R$ with $P$ equally-spaced neighbor points ($\{p_n\}_{n \in (0, P-1)}$):

\[
LBP_{R,P}^{R,P}(c) = \sum_{n=0}^{P-1} s(g_{p_n} - g_c)2^n
\]

where $g_c$ and $g_{p_n}$ denote the gray values of the pixel $c$ and of its $n^{th}$ neighbor $p_n$, respectively, and $s$ is defined as:

\[
s(g_{p_n} - g_c) = \begin{cases} 
1, & g_{p_n} \geq g_c \\
0, & g_{p_n} < g_c
\end{cases}
\]

The most often adopted LBP formulation is the uniform rotation-invariant one ($LBP_{riu2}^{R,P}$) [170]. Rotation invariance is suitable for the purpose of this work since the camera pose during the tissue inspection is constantly changing. The $LBP_{riu2}^{R,P}$ is defined as:

\[
LBP_{riu2}^{R,P}(c) = \begin{cases} 
\sum_{p=0}^{P-1} s(g_{p_p} - g_c), & \text{if } U(LBP_{riu2}^{R,P}) \leq 2 \\
P + 1, & \text{otherwise}
\end{cases}
\]
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

Table 3.1: Tested feature vectors and corresponding number of features. \(Stat_1\): Intensity mean, variance, entropy; \(F_{GLCM}\): Gray-level co-occurrence matrix-based descriptors; \(H_{LBP}\): Normalized histogram of rotation-invariant uniform local binary patterns.

<table>
<thead>
<tr>
<th>Feature vector</th>
<th>(Stat_1)</th>
<th>(F_{GLCM})</th>
<th>(F_{GLCM} + Stat_1)</th>
<th>(H_{LBP})</th>
<th>(H_{LBP} + Stat_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of features</td>
<td>9</td>
<td>144</td>
<td>153</td>
<td>162</td>
<td>171</td>
</tr>
</tbody>
</table>

where \(g_c\) and \(g_{p_p}\) denote the gray values of the pixel \(c\) and of its \(p^{th}\) neighbor \(p_p\), respectively. \(s(g_{p_p} - g_c)\) is defined as:

\[
s(g_{p_p} - g_c) = \begin{cases} 
1, & g_{p_p} \geq g_c \\
0, & g_{p_p} < g_c 
\end{cases}
\]  

(3.4)

and \(U(LBP^{R,P})\) is defined as:

\[
U(LBP^{R,P}) = |s(g_{p_{p-1}} - g_c) - s(g_{p_0} - g_c)| + \sum_{p=1}^{P-1} |s(g_{p_p} - g_c) - s(g_{p_{p-1}} - g_c)|
\]  

(3.5)

In this work, from \(LBP^{R,P}_{riu}2\), the L2-normalized \(H_{LBP}\) was computed and used as feature vector.

For comparison, the GLCM was tested. GLCM calculates how often pair of pixels \((c, q)\) with specific values and in a specified spatial relationship occur in an image. The spatial relationship is defined by \(\theta\) and \(d\), which are the angle and distance between \(c\) and \(q\), respectively. The GLCM width \((W)\), equal to the GLCM height \((H)\), corresponds to the number of quantized image intensity gray-levels. For the \(w = h\) intensity gray-level, the GLCM computed with \(\theta\) and \(d\) is defined as:

\[
GLCM_{\theta,d}(h, w) = \begin{cases} 
1, & I(c) = h \text{ and } I(c_x + d \cdot \cos(\theta), c_y + d \cdot \sin(\theta)) = w \\
1, & I(c) = h \text{ and } I(c_x - d \cdot \cos(\theta), c_y - d \cdot \sin(\theta)) = w \\
0, & \text{otherwise} 
\end{cases}
\]  

(3.6)

From the normalized \(GLCM_{\theta,d}\), as suggested in [83], a feature set \((F_{GLCM})\) was extracted, which consisted of GLCM contrast, correlation, energy and homogeneity. The normalized \(GLCM_{\theta,d}\), which expresses the probability of gray-level occurrences, was obtained by dividing each \(GLCM_{\theta,d}\) entry by the sum of all entries.

First order statistics  Intensity mean, variance (Eq. 2.12) and entropy (Eq. 2.8) were computed and concatenated to form a single intensity-based feature set \((Stat_1)\).

In addition to these descriptors, two feature combinations were tested \((F_{GLCM} + Stat_1, H_{LBP} + Stat_1)\), as suggested in [171] for applications in colorectal endoscopy. All the tested feature vectors and their length are reported in Table 3.1.
3.2. Methods

Line of equality
Cumulative probability
Lorentz curve
Cumulative population

Fig. 3.1: The Gini coefficient (GC) is computed as twice the area (green area) between the line of equality and the Lorentz curve. The Lorentz curve represents the cumulative classification probability among the outcome classification states rank-ordered according to the decreasing values of their individual probabilities. A uniform discrete probability distribution has $GC = 0$, as the Lorentz curve overlays the line of equality, while for a state with probability 100% and the others at 0%, $GC = 1$.

In this Chapter, features were extracted from image patches, which were extracted as explained in Sec. 3.3.1 (laryngeal district) and Sec. 3.4.2 (hepatic district), respectively. Prior to feature extraction, anisotropic diffusion filtering \cite{147} was used to lower noise while preserving sharp edges in the images.

3.2.2 Classification

To perform tissue classification, SVM were used \cite{28}. SVM were chosen since, as introduced in Sec. 2.3.2, they allow overcoming the curse-of-dimensionality that arises analyzing the tested high-dimensional feature space \cite{44, 119}. Here, SVM with the Gaussian kernel were used.

In this work, $\gamma$ and $C$ were retrieved with grid search and 10-fold cross-validation, as explained in Sec. 3.3.1 (laryngeal district) and in Sec. 3.4.2 (hepatic district). To implement multi-class SVM classification, the one-vs-one (2.3.2) scheme was used.

Prior to classification, the feature matrices were normalized within each feature dimension. Specifically, the feature matrices were pre-processed by removing the mean (centering) and scaling to unit variance.

3.2.3 Confidence estimation

As a pre-requisite for the confidence estimation, the probability ($Pr_i(j)$) of the $i^{th}$ patch to belong to the $j^{th}$ class was computed, with $j \in [1, J]$ and $J$ the number of considered
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

tissue classes. For the probability computation, the Platt scaling method revised for multi-class classification problems is used [240]. The Platt scaling method consists of training the parameters of an additional sigmoid function to map SVM outputs to probabilities.

To estimate the reliability of the SVM classification of the $i^{th}$ patch, the dispersion of $Pr_i$ among the $J$ classes was evaluated using the Gini coefficient ($GC$) [137]:

$$GC = 1 - 2 \int_0^1 L(x)dx$$  (3.7)

where $L$ is the Lorentz curve, which is the cumulative probability among laryngeal classes rank-ordered by decreasing values of their individual probabilities. The $GC$ has value 0 if all the probabilities are equally distributed (maximum uncertainty) and 1 for maximum inequality (the classifier is 100% confident in assigning the label). The classification of a patch was considered to be confident if $GC$ was higher that a threshold ($\tau$):

$$\begin{cases} 
\text{Patch}(i) \text{ is confident,} & GC \geq \tau \\
\text{Patch}(i) \text{ is not confident,} & \text{otherwise}
\end{cases}$$

As can be seen from Fig. 3.1, in case of uniform discrete probability distribution (complete uncertainty), $L$ corresponds to the line of equality. Thus, the integral in Eq. 3.7 (red area in Fig. 3.1) has values 0.5 and $GC = 0$. On the contrary, for the case of a single state at 100% with the others at 0% (complete certainty), the integral value is 0 and $GC = 1$. The $GC$ computation can be also seen as twice the area (green area in Fig. 3.1) between the line of equality and the Lorentz curve.

### 3.3 Application 1 - Laryngeal district

As introduced in Sec. 2.4, SCC is the most common cancer of the laryngeal tract, arising from 95% to 98% of all cases of laryngeal cancer [139]. It is well known from medical literature that early-stage SCC diagnosis can lower mortality rate and preserve both laryngeal anatomy and vocal-fold function [227]. Histopathological examination of tissue samples extracted with biopsy is currently the gold-standard for diagnosis. However, tissue visual analysis for screening purpose is commonly performed with NBI endoscopy [180]. The identification of suspicious tissues during the endoscopic examination is, nonetheless, challenging due to the late onset of symptoms and to the small modifications of the mucosa, which can pass unnoticed to the human eye [182]. Main modifications occur to the mucosa vascular tree, with the presence of longitudinal hypertrophic vessels and dot-like vessels, known as intraepithelial papillary capillary loops (ICPL) [180]. Changes in the epithelium aspect not related to the vascular tree, such as thickening and whitening of the epithelial layer (i.e. leukoplakia, Sec. 2.6), are associated with increased risk of developing SCC, too [90]. Visual samples of laryngeal endoscopic video frames of patients affected by SCC are given in Fig. 3.2.

Considering the clinical challenges in diagnosis, some preliminary attempts of CAD have been presented [10, 226], despite only Barbalata et al. [10] specifically focus on early-stage diagnosis. The study proposes an algorithm for the classification of early-stage vocal fold cancer based on the segmentation and analysis of blood vessels. Vessel
segmentation is performed with matched filtering (MF) coupled with first order deriva-
tive of Gaussian. Vessel tortuosity, thickness and density are used as features to discrimi-
nate between malignant and benign tissue by means of LDA. Despite the good results
(overall classification accuracy $= 84\%$), the classification proposed in [10] is strongly
sensitive to a-priori set parameters, e.g., vessel width and orientation. Moreover, focusing
on vessels alone does not allow to take into account epithelial modifications that do
not affect the vascular tree, as in case of leukoplakia.

In this Section, the hypothesis $H_2$ is investigated in the specific context of early-
stage laryngeal cancer classification. In particular, $H_2$ is split in two subhypotheses:

- **Hypothesis 2.1 ($H_{2.1}$):** ML techniques can classify cancerous and pre-cancerous
  laryngeal tissues in NBI images by exploiting textural information

- **Hypothesis 2.2 ($H_{2.2}$):** By estimating the level of classification confidence and
discarding low-confidence samples, the number of incorrectly classified laryngeal
tissues can be lowered

The workflow of the approach is shown in Fig. 3.3

### 3.3.1 Evaluation protocol

In this study, four tissue classes, which are typically evaluated during early-stage di-
agnosis with NBI laryngoscopy, were considered: (i) tissue with IPCL-like vessels,
(ii) leukoplakia, (iii) tissue with hypertrophic vessels, and (iv) healthy tissue. 33 NBI
videos, which refer to 33 different patients affected by SCC, were retrospectively an-
alyzed. SCC was diagnosed with histopathological examination. Videos were ac-
quired with a NBI endoscopic system (Olympus Visera Elite S190 video processor

---

**Fig. 3.2:** Visual samples of narrow-band imaging laryngeal endoscopic frames of patients affected by squamous cell carcinoma.
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis.

Fig. 3.3: Workflow of the proposed approach to laryngeal tissue classification in narrow-band imaging endoscopic video frames.
3.3. Application 1 - Laryngeal district

Table 3.2: Evaluation dataset. For each of the 33 patients’ video, 10 images are used for a total of 330 images. From each image, 4 tissue patches are extracted for a total of 1320 patches relative to the 4 considered tissue classes: healthy tissue, tissue with hypertrophic vessels, leukoplakia, tissue with intraepithelial papillary capillary loop-like vessels. For a robust evaluation, the dataset is split at patient level to perform 3-fold cross-validation. In each fold, 11 patients are included, for a total of 110 images per fold. Each fold contains 440 patches equally distributed among the laryngeal tissue classes.

<table>
<thead>
<tr>
<th></th>
<th>Fold 1</th>
<th>Fold 2</th>
<th>Fold 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>patient ID</td>
<td>1-11</td>
<td>12-22</td>
<td>23-33</td>
<td>33</td>
</tr>
<tr>
<td>n. of images</td>
<td>110 (10 per patient)</td>
<td>110 (10 per patient)</td>
<td>110 (10 per patient)</td>
<td>330</td>
</tr>
<tr>
<td>n. of patches</td>
<td>440 patches (4 per image)</td>
<td>440 patches (4 per images)</td>
<td>440 patches (4 per image)</td>
<td>1320</td>
</tr>
</tbody>
</table>

Fig. 3.4: Four patches, relative to the four analyzed laryngeal tissue classes, are manually cropped from the image. Blue: tissue with intraepithelial papillary capillary loop-like vessels; Yellow: tissue with leukoplakia; Green: healthy tissue; Red: tissue with hypertrophic vessels.

and an ENF-VH rhino-laryngo videoscope) with frame rate of 25 fps and image size of 1920 × 1072 pixels.

A total number of 330 in-focus images (10 per video) was manually selected from the videos, in such a way that the distance between the endoscope and the tissue could be considered constant and approximatively equal to 1 mm for all the images. This distance is suggested in clinics for correct evaluation of tissues during NBI endoscopy examination [124].

For each of the 330 images, 4 patches were manually cropped with a size of 100 × 100 pixels, for a total of 1320 patches, equally distributed among the four classes (Table 3.2). Each patch was cropped from a portion of tissue relative to only one of the four considered classes (tissue with IPCL-like vessels, leukoplakia, hypertrophic vessels and healthy tissue), thus avoiding tissue overlap in one patch. The selection was performed under the supervision of an expert clinician (otolaryngologist specialized in head and neck oncology). A visual example of 4 patches cropped from a NBI frame is shown in Fig. 3.4. Only one patch per tissue class was selected because, in most of the images, selecting more than a single patch was not possible for the IPCL-like class. This is due to the small extension of this vascular alteration in early-stage cancer.
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

For the feature extraction described in Sec. 3.2.1, the \( LBP_{R;P} \) were computed with the following \((R; P)\) combinations: \((1; 8)\), \((2; 16)\), \((3; 24)\), and the corresponding \( H_{LBP} \) were concatenated. Such choice allows a multi-scale, and therefore a more accurate description of the texture, as suggested in [248]. Twelve \( GLCM_{\theta,d} \) were computed using all the possible combinations of \((\theta, d)\), with \( \theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\} \) and \( d \in \{1, 2, 3\} \), and the corresponding \( F_{GLCM} \) sets were concatenated. The chosen interval of \( \theta \) allows to approximate rotation invariance, as suggested in [83]. The values of \( d \) were chosen to be consistent with the scale used to compute \( LBP_{R;P} \), \( GLCM_{\theta,d} \) and \( Stat_1 \) were computed for each channel in the NBI image.

As for performing the classification presented in Sec. 3.2.2, the SVM hyper-parameters \((\gamma, C)\) were retrieved via grid-search and cross-validation on the training set. The grid-search space for \( \gamma \) and \( C \) was set to \([10^{-7}, 10^{-4}]\) and \([10^{-3}, 10^{3}]\), respectively, with six values spaced evenly on \( \log_{10} \) scale in both cases.

For the sake of completeness, the performance of other classifiers, such as \( k\) nearest neighbors (kNN) [96], NB [111], and RF [27], were also investigated. The number of neighbors for kNN was retrieved with a grid-search space set to \([2, 10]\) with nine values spaced evenly, and the number of trees in the forest for RF with a grid-search space set to \([40, 100]\) with six values spaced evenly.

The computation of \( H_{LBP} \), \( F_{GLCM} \) and \( Stat_1 \) was implemented using OpenCv [1]. The classification was implemented with scikit-learn [2].

Investigation of H2.1 In order to assess the hypothesis that ML techniques can characterize laryngeal tissues in NBI images by exploiting textural information, the classification performance of the texture descriptors without confidence estimation (Base case) were first evaluated.

To obtain a robust estimation of the classification performance, 3-fold cross-validation was performed, separating data at patient level to prevent data leakage. The 1320-patch dataset was split to obtain well-balanced folds both at patient-level and tissue-level, as shown in Table 3.2. Each time, two folds were used for training and the remaining one for testing purpose only. This evaluation does not lead to biased results since the tested dataset is balanced over the three folds.

\[
\text{Rec}_{\text{class}} = \{R_{\text{class}}\}_{j \in [1, J=4]} \quad \text{(Eqs. 2.17)}, \quad \text{Prec}_{\text{class}} = \{P_{\text{class}}\}_{j \in [1, J=4]} \quad \text{(Eqs. 2.18)} \quad \text{and} \quad F_{\text{1class}} = \{F_{\text{1class}}\}_{j \in [1, J=4]} \quad \text{(Eqs. 2.19)}
\]

were computed to evaluate the classification performance. For a comprehensive analysis, the AUC was computed. Since the task is a multi-class classification problem and the tested dataset is balanced, the macro-average ROC curve was computed. The gold-standard classification was obtained by labeling the patches under the supervision of an expert clinician.

The Wilcoxon signed-rank test (significance level \( \alpha = 0.05 \)) for paired sample was used to assess whether the classification achieved with the best performing (highest \( \text{Rec}_{\text{class}} \) median value) feature vector significantly differs from the ones achieved with the other feature sets in Table 3.1. Similarly, the Wilcoxon signed-rank test with \( \alpha = 0.05 \) was used to evaluate whether the classification achieved with SVM differed from the ones achieved with the other tested classifiers (kNN, NB, RF).

For the sake of completeness, the performance of the best-performing feature set was
3.3. Application 1 - Laryngeal district

Table 3.3: Median (first quartile - third quartile) class-specific recall (Rec\textsubscript{class}), precision (Prec\textsubscript{class}), and F1 score (F1\textsubscript{class}) obtained testing different feature vectors for the Base case (i.e., without the inclusion of confidence on classification estimation). Classification is obtained with support vector machines. Stat\textsubscript{1}: Intensity mean, variance, entropy; F\textsubscript{GLCM}: Gray-level co-occurrence matrix-based descriptors; H\textsubscript{LBP}: Normalized histogram of rotation-invariant uniform local binary patterns.

<table>
<thead>
<tr>
<th>Stat\textsubscript{1}</th>
<th>F\textsubscript{GLCM}</th>
<th>F\textsubscript{GLCM} + Stat\textsubscript{1}</th>
<th>H\textsubscript{LBP}</th>
<th>H\textsubscript{LBP} + Stat\textsubscript{1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rec\textsubscript{class}</td>
<td>72 (54-82)</td>
<td>75 (72-81)</td>
<td>78 (71-86)</td>
<td>90 (87-92)</td>
</tr>
<tr>
<td>Prec\textsubscript{class}</td>
<td>67 (57-80)</td>
<td>75 (71-80)</td>
<td>78 (72-84)</td>
<td>90 (88-92)</td>
</tr>
<tr>
<td>F1\textsubscript{class}</td>
<td>70 (56-81)</td>
<td>74 (71-80)</td>
<td>79 (72-85)</td>
<td>90 (89-91)</td>
</tr>
</tbody>
</table>

Fig. 3.5: Comparison of different features without including the classification confidence estimation. Classification is obtained with support vector machines. (a) Boxplots of Rec\textsubscript{class} for different features. Stat\textsubscript{1}: Intensity mean, variance, entropy; F\textsubscript{GLCM}: Gray-level co-occurrence matrix-based descriptors; H\textsubscript{LBP}\textsubscript{Stat\textsubscript{1}}: Histogram of rotation-invariant uniform local binary patterns. The stars indicate significant differences (Wilcoxon test, α = 0.05). (b) Normalized confusion matrix for H\textsubscript{LBP} + Stat\textsubscript{1}. The colorbar indicates the number of patches. The total number of patches (n. of patches) is reported.
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

Fig. 3.6: Macro-average receiver operating characteristic (ROC) curves. The mean (± standard deviation) curves obtained from the 3 cross-validation folds are reported in bold (transparent area). The mean (± standard deviation) area under the ROC curve is reported in the legend. (a) ROC curves for the tested features. Classification is obtained using support vector machines. Stat1: Intensity mean, variance, entropy; $F_{GLCM}$: Gray-level co-occurrence matrix-based descriptors; $H_{LBP_{rise}}$: Histogram of rotation-invariant uniform local binary patterns. (b) ROC curves for the tested classifiers. Classification is obtained using the histogram of local binary pattern and first order statistics. kNN: k-nearest neighbors, NB: naive Bayes, RF: random forest, SVM: support vector machines.

compared with those of the most recent - and so far the only one - method [10] published on the topic of laryngeal tissue classification in NBI endoscopy, applying the latter to the tested dataset. As introduced in Sec. 3.1, the method requires to set the vessel segmentation parameters, which were here set as in [10]. The feature classification was performed with SVM, instead of LDA, for fair comparison. The comparison was repeated excluding the leukoplakia class, to avoid privileging the proposed method. Indeed, the method in [10] focuses on the analysis of vessels, which however are not visible in case of leukoplakia due to the thickening of the epithelial layer.

Investigation of H2.2 To investigate the hypothesis that, by estimating the level of classification confidence and discarding low-confidence samples, the number of incorrectly classified cases can be lowered, it was evaluated how $Rec_{class}$, $Prec_{class}$, $F1_{class}$ obtained with the best performing feature vector changed considering different thresholds ($\tau \in [0.6 : 0.1 : 1]$) on the $GC$ value. Since, once the low-confidence patches were excluded, the balance between classes did not hold, the ROC curves for each of the four laryngeal classes was computed (and not the macro-average ones as for H2.1).

3.3.2 Results

For the Base case, the best performance (median $Rec_{class} = 93\%$, inter-quartile range (IQR) = 6%) was obtained with $H_{LBP} + Stat1$ and SVM classification, as shown in Table 3.3. The same was observed also when considering $Prec_{class}$ (median = 94%, IQR = 4%) and $F1_{class}$ (median = 92%, IQR = 4%). The classification statistics relative to all the analyzed features are reported in Fig. 3.5(a). Significant differences (p-value < 0.05) were found when comparing $H_{LBP} + Stat1$ with $Stat1$, $F_{GLCM}$, and $F_{GLCM} + Stat1$. The normalized confusion matrix for $H_{LBP} + Stat1$ is shown in Fig. 3.5(b).
3.3. Application 1 - Laryngeal district

Table 3.4: Comparison of different classifiers. Median (first quartile - third quartile) class-specific recall ($\text{Rec}_{\text{class}}$), precision ($\text{Prec}_{\text{class}}$), and F1 score ($\text{F1}_{\text{class}}$) are reported for the four different tissue classes. Classification is obtained using the histogram of local binary patterns and first order statistics. kNN: k-nearest neighbors, NB: naive Bayes, RF: random forest, SVM: support vector machines.

<table>
<thead>
<tr>
<th></th>
<th>kNN</th>
<th>NB</th>
<th>RF</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Rec}_{\text{class}}$</td>
<td>90 (84-93)</td>
<td>78 (74-82)</td>
<td>89 (84-91)</td>
<td>93 (90-96)</td>
</tr>
<tr>
<td>$\text{Prec}_{\text{class}}$</td>
<td>89 (86-91)</td>
<td>81 (73-84)</td>
<td>87 (86-89)</td>
<td>94 (91-95)</td>
</tr>
<tr>
<td>$\text{F1}_{\text{class}}$</td>
<td>89 (86-91)</td>
<td>79 (74-83)</td>
<td>89 (86-90)</td>
<td>92 (91-95)</td>
</tr>
</tbody>
</table>

Fig. 3.7: Performance of the state of art. (a) Visual samples of the vessel segmentation obtained applying Barbalata et al. [10] algorithm to patches with hypertrophic vessels (first row), healthy tissue (second row) and intraepithelial papillary capillary loop-like vessels (third row). From left to right, original patch, vessel mask and vessel mask superimposed on the original patch. (b) Normalized confusion matrix obtained applying Barbalata et al. [10] algorithm to the tested laryngeal dataset. Colorbar indicates the number of patches. The total number of patches (n. of patches) is reported.

In Fig. 3.6(a), the macro-average ROC curves are reported for all tested features and SVM classification. The mean AUC across the three folds was 0.99 for $H_{LBP} + \text{Stat}_1$.

As shown in Table 3.4, SVM has shown comparable performance with respect to kNN and RF in terms of $\text{Rec}_{\text{class}}$, $\text{Prec}_{\text{class}}$, $\text{F1}_{\text{class}}$, while SVM outperformed (p-value < 0.05) NB. The same can be noticed from the ROC curve analysis in Fig. 3.6(b).

When applying the algorithm proposed by Barbalata et al. [10] to the tested dataset, a median $\text{Rec}_{\text{class}}$ value of 42% was obtained, with IQR of 48%. Significant differences (p-value $\ll 0.05$) were found when comparing the algorithm results with those obtained exploiting $H_{LBP} + \text{Stat}_1$. Visual examples of the vessel segmentation obtained with the method proposed in [10] are reported in Fig. 3.7(a) for patches with hypertrophic vessels, healthy tissue and IPCL-like vessels. The confusion matrix for the classification obtained with the method in [10] is reported in Fig. 3.7(b). Barbalata et al. algorithm correctly labeled leukoplakias and abnormal IPCL only in the 7% and 26% of all cases, respectively. Almost half of leukoplakias and abnormal IPCL were misclassified as healthy tissues. When excluding the leukoplakia class, the $\text{Rec}_{\text{class}}$ was: 62% (healthy tissue), 70% (tissue with hypertrophic vessels), 28% (tissue with IPCL-like vessels).

As shown in Table 3.5, when varying $\tau$ in $[0.6 : 0.1 : 1]$, the median $\text{Rec}_{\text{class}}$ for $H_{LBP} + \text{Stat}_1$ monotonically increased from 93% (Base case) to 98% ($\tau = 0.9$). The corresponding statistics are shown in Fig. 3.8(a). The same trend was observed for
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

Table 3.5: Median (first quartile - third quartile) class-specific recall ($\text{Rec}_{\text{class}}$), precision ($\text{Prec}_{\text{class}}$), and F1 score ($\text{F1}_{\text{class}}$) are reported at different level of confidence ($\tau$) on support vector machines classification.

<table>
<thead>
<tr>
<th></th>
<th>$\tau = 0$</th>
<th>$\tau = 0.60$</th>
<th>$\tau = 0.70$</th>
<th>$\tau = 0.80$</th>
<th>$\tau = 0.90$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Rec}_{\text{class}}$</td>
<td>93 (90-96)</td>
<td>95 (91-97)</td>
<td>96 (92-99)</td>
<td>98 (93-99)</td>
<td>98 (95-100)</td>
</tr>
<tr>
<td>$\text{Prec}_{\text{class}}$</td>
<td>94 (91-95)</td>
<td>95 (92-96)</td>
<td>97 (93-98)</td>
<td>97 (95-98)</td>
<td>99 (96-100)</td>
</tr>
<tr>
<td>$\text{F1}_{\text{class}}$</td>
<td>92 (91-95)</td>
<td>94 (93-96)</td>
<td>95 (94-97)</td>
<td>96 (95-97)</td>
<td>98 (97-99)</td>
</tr>
</tbody>
</table>

Fig. 3.8: Effect of varying the threshold ($\tau$) on the classification confidence level. Classification is obtained using local binary pattern and first order statistics with support vector machines. (a) Boxplot of the class-specific accuracy-rate ($\text{Rec}_{\text{class}}$) for different $\tau$. The percentage of confident patches for each $\tau$ is reported above each boxplot. $\tau = 0$ refers to classification without confidence estimation. (b) Normalized confusion matrix for $\tau = 0.9$. The number of patches for each class is reported in parenthesis.

Fig. 3.9: Receiver operating characteristic (ROC) curves at different level $\tau$ of confidence on classification. Each curve refers to one of the laryngeal tissue classes. He: healthy tissue; Hbv: tissue with hypertrophic vessels; Le: leukoplakia; IPCL: tissue with intraepithelial papillary capillary loop-like vessels. The area under the ROC curve, for each curve, is reported in the legend. Classification is obtained using local binary pattern and first order statistics with support vector machines. (a) ROC curves for $\tau = 0.6$. (b) ROC curves for $\tau = 0.9$. 

46
3.3. Application 1 - Laryngeal district

Prec_{\text{class}} and F_{1\text{class}}. The ROC curves for $\tau = 0.6$ and $\tau = 0.9$ are shown in Fig. 3.9. The AUC is reported for each of the analyzed classes (AUC = 0.99 $\pm$ 0.01 ($\tau = 0.9$)). The Rec_{\text{class}} increment came at the cost of a reduction of the percentage of confident patches to 80% ($\tau = 0.9$) of all the patches in the testing set, which corresponds to $\sim$ 1056 patches. However, as shown in Fig. 3.7(b) with the exclusion of low-confidence patches, even in the worst case (classification of tissue with IPCL-like vessels), the accuracy still reached 93%.

Fig. 3.10 shows visual samples of patches in the tested dataset (Fig. 3.10(a)), as well as samples of patch classification results at the Base case (Fig. 3.10(b)) and after the introduction of the confidence measure ($\tau = 0.9$) (Fig. 3.10(c)).

3.3.3 Discussion

In this Chapter, a new paradigm for the computer-aided classification of laryngeal tissues in NBI laryngoscopy was presented and fully evaluated. Different textural features were tested to investigate the best feature set to characterize malignant and healthy laryngeal tissues: texture-based global descriptors ($F_{\text{GLCM}}$ and $H_{\text{LBP}}$) and first order statistics ($\text{Stat}_1$). A confidence measure on the SVM-based classification was used to estimate the reliability of the classification results.

When comparing non-combined features ($F_{\text{GLCM}}$, $\text{Stat}_1$, $H_{\text{LBP}}$), the highest classification performance was obtained with $H_{\text{LBP}}$. In general, $F_{\text{GLCM}}$ performed worse with respect to $H_{\text{LBP}}$. This is probably due to the GLCM lack of robustness to illumination condition changes, which are typically encountered during endoscopic examination.

SVM showed comparable performance with respect to RF and NB, while significant differences (p-value < 0.05) were found with respect to NB. This is probably due to NB not being able to handle high-dimensional feature spaces such as the ones here tested. This is in accord with previous findings in the literature, e.g., [25, 44, 52].

When comparing the proposed method with the state of the art, the classification based on $H_{\text{LBP}}$ significantly outperformed (p-value $\ll$ 0.05) the one proposed by Barbata et al. [10], also when excluding the leukoplakia class. Since the method in [10] relies on accurate vessel segmentation (to extract vascular shape-based features), a possible reason of such result could be related to the challenging nature of the validation dataset, which however well summarizes the diagnostic scenario. Indeed, vessel segmentation was not trivial (Fig. 3.7(a)) due to (i) the noisy nature of NBI data, (ii) the low contrast of vessels in patches with healthy tissue and leukoplakia and (iii) the irregular shape of IPCL-like vessels. With texture-based features, higher classification performance was achieved with respect to shape-based features since texture-based feature computation does not require vessel segmentation. Moreover, the texture-based features here used are invariant to illumination changes and endoscope pose, which makes them suitable for the analyzed scenario.

The classification performance obtained with $H_{\text{LBP}}$ was further increased by estimating the confidence of the SVM classification, with few misclassification of confident patches that mainly occurred with high-challenging vascular patterns, whose classification is not trivial also for the human eye (Fig. 3.10(c)). Such results support the hypothesis that the proposed approach is suitable for classifying laryngeal tissues with high reliability, since it automatically estimates its own confidence level and provides
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

Fig. 3.10: Visual samples of classification results. Classification is obtained using local binary pattern and first order statistics with support vector machines. Examples of patches for the four tissue classes in the tested dataset. Visual confusion matrices for the Base case (b) and with the inclusion of confidence estimation and after including the confidence estimation with $T = 0.9$ (c). Black squares indicate the absence of misclassification between the true and predicted label.
high classification accuracy for confident patches.

A limitation of the proposed study could be seen in its patch-based nature. Note, however, that the choice of focusing on patches manually extracted under the supervision of an expert clinician was driven by the necessity of having a controlled and representative dataset to fairly evaluate different features. As future work, instead of manually selecting squared patches, more automatic strategies, such as superpixel segmentation [117], could be integrated. The features could be directly extracted from superpixels, as to classify each superpixel as belonging to one of the analyzed laryngeal classes. Moreover, it would be interesting to exploit also CNN as feature extractor for comparison once a larger dataset will be collected.

The expectation is that research on the classification of laryngeal tissues will be empowered by the proposed work, becoming a topic of interest for the scientific community, which until now has mainly focused on other anatomical sites, such as the gastrointestinal tract. Moreover, this study will motivate a more structured and widespread data collection in clinics and the sharing of such data through public databases. Despite the dimension of the analyzed dataset (330 images) is comparable with that of similar researches (e.g., Barbalata et al. [10] with 120 images, Turkmen et al. [226] with 70 images), larger amounts of data would bring the possibility of further exploring ML classification algorithms, e.g., to classify a larger number of laryngeal malignant tissues. In this perspective, the testing dataset was released and it is now available online at https://zenodo.org/record/1003200#.WdeQcnBx0nQ.

### 3.3.4 Conclusion

In conclusion, the most significant contribution of this work is showing that LBP-based features and SVM can differentiate laryngeal tissues accurately. This is highly beneficial for practical uses. Comparing with other state-of-the-art methods in the area, the proposed method is simpler and the result is more accurate. It is acknowledged that further research is required to further ameliorate the algorithm as to offer all possible support for diagnosis, but the results presented here are surely a promising step towards a helpful endoscope-integrated processing system to support the diagnosis of early-stage SCC.

### 3.4 Application 2 - Hepatic district

Liver transplantation (LT) is the treatment of choice for patients with end–stage liver disease for which no alternative therapies are available [36]. Due to increasing demand and shortage in organ supply, expanded donor selection criteria are applied to increase the number of grafts for LT. Since extended criteria donors generates augmented morbidity and mortality in recipient population, liver-graft quality assessment is crucial.

Hepatic steatosis (HS) is one of the most important donor characteristic that can influence graft function and so LT outcome, mostly because of severe ischemia reperfusion injury [100]. Defined as the intracellular accumulation of triglycerides resulting in the formation of lipid vesicles in the hepatocytes, HS is commonly assessed by histopathological examination of liver tissue samples extracted with biopsy. Through visually analyzing the quantities of large sized lipid droplets in the sample, an HS score is assigned to the sample in a semiquantitative fashion. Livers classified as with 5%–30%
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

Fig. 3.11: Sample RGB liver images acquired in the operating room. Images are captured with different lighting conditions and different tissue-camera pose. Images refer to a transplanted liver graft (left) and a discarded one (right).

Fatty infiltration are associated with decreased patient and graft survival, but are still considered suitable for transplantation due to the limited donor availability. Severe HS (≥60%) is instead associated with primary graft dysfunction or non–function and is not compatible with transplantation.

Despite histopathological analysis of biopsied liver tissue being currently the gold reference standard for diagnosis and grading of HS in liver grafts, it is invasive, time-consuming and expensive. Due to the short time availability between liver procurement and transplantation, the surgeon usually performs HS assessment through clinical evaluation (medical history, blood tests) and qualitative visual graft assessment. In this context, visual liver texture analysis is recognized as crucial in grading HS: livers that cannot be transplanted due to high HS (Fig. 3.11 right) are usually characterized by inhomogeneous texture and are more yellowish than the transplantable ones (Fig. 3.11 left). It is nonetheless recognized that the precise estimation of HS remains challenging even in experienced hands.

On this background, the development of a robust, quantitative, practical, cost-effective and rapid method to support the surgeon in deciding whether to accept or discard liver grafts is mandatory.

In the last years, preliminary efforts to the automated or semi-automated HS assessment have been proposed. In contrast-enhanced CT imaging, it has been reported that the attenuation value of a healthy liver organ is about 50–57 Hounsfield unit (HU), and it is about 10 HU higher than the attenuation value of a healthy spleen. By evaluating the liver-to-spleen HU attenuation ratio, a sensitivity of 79% was reported in recognizing significant steatosis. However, the attenuation value of liver does not depend only on fat contents. Different factors, such as fibrosis, edema, accumulation of glycogen or increased deposition of iron affect the CT-based HS assessment.

Fibroscanning has been proposed for HS assessment. The idea is to use liver stiffness as features for HS assessment, nonetheless a strategy to classify transplantable livers has not been proposed. By applying thresholds on the liver stiffness values, the authors reported an AUC of 75%.

Liver bioelectrical impedance analysis was used to assess HS through resistance and reactance analysis. Despite a significative correlation between liver impedance...
Fig 3.12: Proposed workflow for graft hepatic steatosis assessment. From 40 RGB liver images of 40 different donors, a dataset of patches with size $100 \times 100$ is extracted. From each patch, a set of textural features is computed. The dataset is divided in training and testing patches. The features from the training patches are used to train a supervised classifier model. The trained model is used to assess HS from the testing patches. Classification confidence is estimated and low-confidence patches are excluded.
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

and HS level being demonstrated, also in this case a classification strategy was not proposed.

More recently, animals’ experimental studies showed that Raman spectroscopy coupled with principal component analysis (PCA) can provide accurate and rapid assessment of the hepatic fat content for bulk liver \[84\]. Unfortunately, there is not human study that confirmed the research findings and the authors only reported a correlation \(R^2 = 0.96\) between the principal component score and the gold standard HS score.

MRI-based techniques for HS quantification have been proposed and a substantial correlation between the fat fraction determined histopathologically and radiologically has been demonstrated \[239\]. A semi-automatic HS grading approach that exploits MR spectroscopy has been proposed in \[187\]. The method demonstrated that a correlation exists between a measure based on the average pixel values of liver and spleen and the gold standard HS level (Spearman correlation = 0.90).

It is worth noting that all the proposed methodologies require additional imaging instrumentation, which may be not always available in the remote graft-procurement hospitals. Moreover, at most the methods concluded that there is a correlation between liver physical characteristics (e.g., liver HU, stiffness, impedance) and HS grade, without providing a solution for liver-graft quality assessment.

Despite the relevance of visual liver-texture analysis being crucial for clinical HS assessment \[244\], to the best of authors’ knowledge no efforts have been done to develop a computer-assisted diagnostic tool that exploits automatic texture analysis to assess graft steatosis. Moreover, liver-texture analysis has the advantage of being performed on standard RGB optical imaging, without requiring additional instrumentations. It is worth noting that modern cellphone cameras provide decent quality images for liver assessment and are ubiquitous. Therefore, they could be the solution for automatic HS assessment not only in remote hospitals, but also in low-income countries where other imaging equipment may not be available. Indeed, the use of RGB cameras for tissue classification is becoming quite popular in different fields, such as skin cancer diagnosis \[58\].

In this Section, the hypothesis \(H2\) is investigated in the context of pathological liver-issue classification. In particular, \(H2\) is divided in:

- **Hypothesis 2.3 \((H2.3)\)**: Liver texture analysis from RGB images acquired with smartphones in the OR coupled with ML can provide reliable results, to be used as support for LT decision

- **Hypothesis 2.4 \((H2.4)\)**: By estimating the level of classification confidence and discarding low-confidence samples, the number of incorrectly classified hepatic tissues can be lowered

Moreover, a semi-supervised approach to \(H2.3\), which is an extension of the method described in Sec. \[3.2\] is here investigated, too, as described in Sec. \[3.4.1\]

### 3.4.1 Extension to semi-supervised classification

In Sec. \[3.2.2\] the classification is performed on a patch-basis. However, it is worth noting that the ground-truth biopsy-based HS level is provided for the whole image, and
Fig. 3.13: Workflow of the proposed approach to liver-donor's steatosis assessment from RGB images acquired in the operating room. Image patches are extracted from the RGB images. From each patch, textural, intensity and blood-test features are extracted. The features extracted from training patches are used to train a semi-supervised support vector machine (SVM)-single instance learning (SIL) model (green boxes). The approach is semi-supervised as the ground-truth (histopathological-based) label is assigned to the whole image and not to the single patch. The proposed semi-supervised approach is compared with a naive patch-based supervised SVM classification followed by clustering (orange boxes). Clustering is required to obtain image classification from patch-classification results.
not for the single patch. Thus, in the method explained in Sec. 3.2.2 the assumption is made that all patches extracted from pathological images are associated with pathological labels. In fact, this is not always true as HS is not homogeneous in the hepatic tissue.

The problem of diagnosing HS from unlabeled RGB image patches extracted from labeled images is a semi-supervised problem. The problem can be addressed with semi-supervised techniques, which have recently drawn the attention of the computer-assisted diagnosis community [133].

Thus, this Chapter also investigates if multiple instance learning (MIL), a semi-supervised machine-learning technique, can support HS diagnosis from (unlabeled) patches extracted from (labeled) images acquired with RGB cameras in the OR.

Among MIL algorithms, the use of single instance learning (SIL) [188] is investigated. SIL is an exclusively bag-level algorithm according to the categorization proposed in [188]. Exclusively bag-level algorithms have the strong advantage of allowing the fusion of patch-level features (such as textural features) with image-level features (such as blood-sample features) [188]. Indeed, in addition to the features described in Sec. 3.2.1, here features from blood tests (Blo) were used, too, as suggested by the clinical literature [244]. In particular, alanine aminotransferase, aspartate aminotransferase, bilirubin, liver HU, difference between the liver and the spleen HU, and gamma globuline were considered. Further, patient’s age, weight and height were also considered.

For being consistent with the method introduced in Sec. 3.2.2, in which SVM are used, here SVM-SIL classification [190] was exploited.

For a semi-supervised binary classification problem, like the one here addressed, let, as in Sec. 2.3.2, $T$ be the set of training images. $T_p \subseteq T$ is the set of positive images (rejected grafts), and $T_n \subseteq T$ the set of negative images (accepted grafts). Let $\tilde{T}_p = \{ t \mid t \in T \in T_p \}$ and $\tilde{T}_n = \{ t \mid t \in T \in T_n \}$ be the patches from positive and negative images, respectively. Let $L = L_p + L_n = |\tilde{T}_p| + |\tilde{T}_n|$ be the total number of patches. For any patch $t \in T$ from an image $T \in T$, let $x_t$ be the feature vector representation of $t$. Thus, $x_T = \sum_{t \in T} x_t$ is the feature vector representation of image $T$.

The SVM-SIL optimization, here written in the “primal” SVM formulation for better readability, aims at minimizing:

$$J(w, b, \xi) = \frac{1}{2} \|w\|^2 + \frac{C}{L} \sum_{X \in X} \sum_{x \in X} \xi_x$$  \hspace{1cm} (3.8)

subject to:

$$w x_t + b \leq -1 + \xi_t, \quad \forall t \in \tilde{T}_n$$ \hspace{1cm} (3.9)

$$w x_t + b \geq +1 - \xi_t, \quad \forall t \in \tilde{T}_p$$ \hspace{1cm} (3.10)

$$\xi_t \geq 0$$ \hspace{1cm} (3.11)

where $\xi_t$ is the relaxing term introduced for the soft-margin SVM formulation, $b$ is a real value, $w$ the SVM weight vector and $C$ is a regularization parameter, trade off between maximizing the SVM margins and minimizing the training error, as introduced in Sec. 2.3.2. Also in this case, $C$ was retrieved with grid-search and cross-validation, as explained in Sec. 3.4.2.
3.4. Application 2 - Hepatic district

**Fig. 3.14:** Dataset sample images. The images refer to transplanted livers. Images were acquired at different distance and orientation with respect to the liver. Images present different illumination levels. Specular reflections are present due to the smooth and wet liver surface.

**Fig. 3.15:** Liver and liver mask obtained through manual segmentation

For comparing the supervised classification explained in Sec. 3.2.2, the output of which refers to patches, with SVM-SIL, the supervised classification in Sec. 3.2.2 was performed and followed by a clustering procedure (SVM + Clustering). Clustering was required as post-processing to assign a label to the whole image starting from patch-classification results. After SVM-based classification, each image was labeled as positive if at least one patch was classified as positive.

The workflow of the proposed approach to HS assessment through semi-supervised learning is showed in Fig. 3.13 along with the SVM + Clustering comparison method.

### 3.4.2 Evaluation

In this study, 40 RGB images, which refer to 40 different potential liver donors, were analyzed. HS was assessed with histopatological analysis performed after liver biopsy.

Biopsy was performed during procurement, taking surgical triangular hepatic samples up to 2 cm. One pathologist analyzed the histological sections. Steatosis was visually assessed based on the percentage of hepatocytes with intracellular large lipid droplets by using a semi-continuous scale [0%:5%:100%].
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

Fig. 3.16: Dataset sample patches. The green and red boxes refer to patches extracted from transplanted and non-transplanted livers. Each row in a box refers to patches extracted from the same liver.

Table 3.6: Area under the receiving operating characteristic (ROC) curve obtained with support vector machines (SVM) and different feature vectors for patch classification. $Stat_1$: intensity mean, variance and entropy; $F_{GLCM}$: Gray-level co-occurrence matrix-based descriptors; $H_{LBP_{f_{iu}}}$: Normalized histogram of rotation-invariant uniform local binary patterns.

<table>
<thead>
<tr>
<th>Feature Vectors</th>
<th>Area under the ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Stat_1$</td>
<td>0.65</td>
</tr>
<tr>
<td>$F_{GLCM}$</td>
<td>0.71</td>
</tr>
<tr>
<td>$H_{LBP_{f_{iu}}} + Stat_1$</td>
<td>0.76</td>
</tr>
<tr>
<td>$H_{LBP}$</td>
<td>0.77</td>
</tr>
</tbody>
</table>

From the dataset, 20 livers referred to discarded grafts, as with a HS $\geq 60\%$. The remaining 20 livers had a HS $\leq 20\%$ and were transplanted. Images were acquired with a smartphone RGB camera. Image size was $1920 \times 1072$ pixels. Challenges associated with the dataset included:

- Wide range of illumination
- Varying camera pose
- Presence of SR
- Different organ position

Visual samples of liver images are shown in Fig. 3.14.

From each image, liver manual segmentation was performed to separate the hepatic tissue from the background (Fig. 3.15). The whole image was then divided in non-overlapping patches of size $100 \times 100$ pixels starting from the top-left image corner. The rightest part of the image, for which it was not possible to select full patches, was discarded. This did not represent a problem since the liver was always displayed at the center of the image. A patch was considered valid for the analysis if it overlapped with at least 90% of the liver mask.

To have the same number of patches from each patient, the minimum number of image patches that could be extracted among all images was first computed, which was
3.4. Application 2 - Hepatic district

Fig. 3.17: Receiving operating characteristic (ROC) curve for patch classification with the normalized histogram of rotation-invariant uniform local binary patterns and support vector machines.

15. Then, 15 patches were randomly extracted from all the other images. As result, the testing patch dataset was composed of 300 patches extracted from transplanted liver and 300 from non-transplanted ones. Sample patches for transplanted and non-transplanted livers are shown in Fig. 3.16.

For the feature extraction described in Sec. 3.2.1, the $LBP_{R;P}^{riu_2}$ were computed with the following $(R;P)$ combinations: $(1; 8)$, $(2; 16)$, $(3; 24)$, and the corresponding $H_{LBP}$ were concatenated. Such choice allows a multi-scale, and therefore a more accurate description of the texture. The $LBP_{R;P}^{riu_2}$ were computed for each RGB image channel.

Nine $GLCM_{\theta,d}$ were computed for each RGB channel using all the possible combinations of $(\theta, d)$, with $\theta \in \{0^\circ, 45^\circ, 90^\circ\}$ and $d \in \{1, 2, 3\}$, and the corresponding $F_{GLCM}$ sets were concatenated. The chosen interval of $\theta$ allows to approximate rotation invariance, as suggested in [83]. The values of $d$ were chosen to be consistent with the scale used to compute $LBP_{R;P}^{riu_2}$.

For performing the classification presented in Sec. 3.2.2, the SVM hyper-parameters $(\gamma, C)$ were retrieved via grid-search and cross-validation on the training set. The grid-search space for $\gamma$ and $C$ was set to $[10^{-10}, 10^{-1}]$ and $[10^0, 10^{10}]$, respectively, with 10 values spaced evenly on $\log_{10}$ scale in both cases.

For the sake of completeness, the performance of RF [27] was also investigated. The number of trees for RF training was retrieved with with a grid-search space set to $[40, 100]$ with six values spaced evenly.

As for SVM-SIL (Sec. 3.4.1), the same protocol here described to train supervised SVM-based method of Sec. 3.2.2 was used.
Chapter 3. Texture-based tissue classification with confidence estimation for
computer-assisted diagnosis

The feature extraction and classification were implemented with scikit-image\(^3\) and
scikit-learn\(^4\).

**Investigation of H2.3** The classification performance was evaluated using leave-one-
patient-out cross-validation. Each time, the patches extracted from one patient were
used for testing the performance of the SVM trained with the images of all the other
patients. The separation at patient level was necessary to prevent data leakage. \(\text{Rec}_{\text{class}}\) (Eq. 2.17), \(\text{Prec}_{\text{class}}\) (Eq. 2.18) and \(\text{F1}_{\text{class}}\) (Eq. 2.19) were computed to evaluate the
classification performance, here considering \(J = 2\) (binary classification problem).
For the semi-supervised image classification, the overall image classification accuracy
\(\text{Acc}\) was also computed. \(\text{Acc}\) counts the number of correct classification results di-
vided by the total number of tested data.

The Wilcoxon signed-rank test (significance level \(\alpha = 0.05\)) for paired sample was
used to assess whether the classification achieved with the best-performing (highest
recall in classifying discarded liver patches) feature vector significantly differed from
the ones achieved with the other feature sets in Table 3.1.

**Investigation of H2.4** To investigate the hypothesis that, by estimating the level of classi-
fication confidence and discarding low-confidence samples, the number of incorrectly
classified cases can be lowered, it was evaluated how \(\text{Rec}_{\text{class}}\), obtained with the best-
performing feature vector changed considering different thresholds \(\tau \in [0.1 : 0.1 : 0.4]\) on the \(GC\) value. With respect to Sec. 3.3.1, here lower \(\tau\) values were investi-
gated, as the problem is a two-class problem \[^{137}\].

**3.4.3 Results**

Table 3.6 shows the AUC for the supervised SVM patch classification (Sec. 3.2.2)
obtained with the feature vectors in Table 3.1. The higher AUC (0.77) was obtained
with \(H_{\text{LBP}}\). The relative ROC curve is shown in Fig. 3.17. The SVM classification
performance obtained with SVM and \(\text{Stat}_1, F_{\text{GLCM}}, H_{\text{LBP}}\) and \(H_{\text{LBP}} + \text{Stat}_1\) are shown in Table 3.7. The best performance was obtained with \(H_{\text{LBP}}\), with mean
\(\text{Rec}_{\text{class}} = 0.73\), \(\text{Prec}_{\text{class}} = 0.74\) and \(\text{F1}_{\text{class}} = 0.73\). It is worth noting that a recall
of 0.82 was obtained for non-transplanted patches. Using only \(\text{Stat}_1\) features led to the worst
classification performance, with a recall of 0.58 for the non-transplanted patches.

Significant differences were found when comparing the best performing feature
\((H_{\text{LBP}})\) with \(\text{Stat}_1\) and \(F_{\text{GLCM}}\). The confusion matrices for feature comparison are
reported in Fig. 3.18. When classifying \(H_{\text{LBP}}\) with RF, mean \(\text{Prec}_{\text{classes}} = 0.67\), \(\text{Rec}_{\text{classes}} = 0.67\), and \(\text{F1}_{\text{classes}} = 0.67\) were obtained. A recall of 0.72 was ob-
tained for non-transplanted patches (Table 3.8). Significant differences with respect to
SVM performance were not found. The visual confusion matrix for the classification
performed with SVM and \(H_{\text{LBP}}\) is shown in Fig. 3.19.

When introducing the confidence measure, the median recall for was improved from
0.82 (rejected-liver patches) and 0.64 (accepted-liver patches) for the \(\text{Base}\) case to 0.86
and 0.72 for \(\tau = 0.4\), respectively. The same trend was observed for \(\text{Prec}_{\text{class}}\) and
\(\text{F1}_{\text{class}}\).

\(^3\)http://scikit-image.org/  
\(^4\)http://scikit-learn.org
Table 3.7: Patch-classification performance measures obtained with support vector machines (SVM) and different feature vectors. \( \text{Stat}_1 \): Intensity mean, variance and entropy; \( F_{\text{GLCM}} \): Gray-level co-occurrence matrix-based descriptors; \( H_{\text{LBP}} \): Normalized histogram of rotation-invariant uniform local binary patterns. Metrics are reported for transplanted and non-transplanted liver patches. Mean metrics are reported, too. \( \text{Prec}_{\text{classes}} \) = Precision; \( \text{Rec}_{\text{classes}} \) = Recall, \( \text{F1}_{\text{classes}} \) = \( F_1 \) score.

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Transplanted</th>
<th>Non-transplanted</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Stat}_1 )</td>
<td>0.66</td>
<td>0.74</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.58</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>0.72</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>( F_{\text{GLCM}} )</td>
<td>0.69</td>
<td>0.65</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.73</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>( H_{\text{LBP}} + \text{Stat}_1 )</td>
<td>0.76</td>
<td>0.69</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.80</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>( H_{\text{LBP}} )</td>
<td>0.78</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.82</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.73</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.8: Patch-classification performance measures obtained using the histogram of rotation-invariant local binary patterns as feature with different classifiers. \( \text{Rec}_{\text{class}} \) = Precision; \( \text{Prec}_{\text{class}} \) = Recall, \( \text{F1}_{\text{class}} \) = \( F_1 \) score; SVM = Support vector machines, RF = Random forest. Metrics are reported for transplanted and non-transplanted liver patches. Mean metrics are reported, too.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Transplanted</th>
<th>Non-transplanted</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>0.69</td>
<td>0.65</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>0.72</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>0.65</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>0.78</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.82</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.73</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

Fig. 3.18: Confusion matrix (CM) for classification of patches from transplanted (T) and non-transplanted (NT) liver. (a) CM for gray-level co-occurrence based-features ($F_{GLCM}$). (b) CM for intensity-based features ($Stat_1$). (c) CM for intensity-based and local binary pattern features ($H_{LBP} + Stat_1$). (d) CM local binary pattern-based features ($H_{LBP}$).

Table 3.9: Image-classification performance measures for support vector machines and clustering (SVM + Clustering), and SVM-single instance learning (SVM-SIL). Classification accuracy (Acc), recall ($Rec_{classes}$), precision ($Prec_{classes}$) and F1-score ($F1_{classes}$) are reported. Metrics are reported for transplanted and non-transplanted liver images.

<table>
<thead>
<tr>
<th></th>
<th>$Rec_{classes}$</th>
<th>$Prec_{classes}$</th>
<th>$F1_{classes}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM + Clustering</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transplanted</td>
<td>0.65</td>
<td>0.81</td>
<td>0.72</td>
</tr>
<tr>
<td>Non-transplanted</td>
<td>0.85</td>
<td>0.71</td>
<td>0.77</td>
</tr>
<tr>
<td>SVM-SIL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transplanted</td>
<td>0.80</td>
<td><strong>0.94</strong></td>
<td>0.86</td>
</tr>
<tr>
<td>Non-transplanted</td>
<td><strong>0.95</strong></td>
<td>0.83</td>
<td><strong>0.88</strong></td>
</tr>
</tbody>
</table>
3.4. Application 2 - Hepatic district

Fig. 3.19: Visual confusion matrices for SVM-based patch classification obtained with the histogram of rotation-invariant uniform local binary patterns and support vector machines. NT: Non-transplanted patches; T = Transplanted patches.

Fig. 3.20: Confusion matrices (CM) for image classification of transplanted (T) and non-transplanted (NT) liver grafts. CM are obtained with the best performing feature set ($H_{LBP}$ + Stat$_1$ + Blo). The classification is performed with (left) support vector machines (SVM) + Clustering and (right) SVM-single instance learning (SVM-SIL). The colorbars indicate the number of correctly classified images.
Chapter 3. Texture-based tissue classification with confidence estimation for computer-assisted diagnosis

Fig. 3.21: Samples of classification outcomes for transplanted (first row) and non-transplanted (second row) liver grafts. Classification refers to support vector machines (SVM) - single instance learning (SIL) with the best performing feature set \( \text{H}_{10}^\text{LBP} + \text{Stat}_1 + \text{Blo}_1 + \text{Blo}_5 \). The green and red boxes refer to correct and wrong classification outcomes, respectively. SVM-SIL wrongly classified a rejected liver only once.
For the semi-supervised image-wise classification (Sec. 3.4.1), among all the tested feature sets, \( H_{LBP} + Stat_1 + Blo \) showed the best classification performance for both SVM-SIL and SVM + Clustering. The SVM-SIL and SVM + Clustering performances for \( H_{LBP} + Stat_1 + Blo \) are reported in Table 3.9. SVM-SIL outperformed SVM + Clustering, with a recall of 0.95 (vs 0.85) for non-transplanted livers and an overall Acc of 0.88 (vs 0.75). The Acc values achieved for rejected liver images with SVM-SIL for the other tested feature sets were 0.63 (Stat1), 0.55 (\( H_{LBP} \)), 0.63 (\( H_{LBP} + Stat_1 \)), 0.78 (\( H_{LBP} + Blo \)). The confusion matrices for the comparison of SVM-SIL and SVM + Clustering for the best performing feature set are reported in Fig. 3.20.

Visual samples of liver classification outcomes with \( H_{LBP} + Stat_1 + Blo \) for SVM-SIL are showed in Fig. 3.21. It is worth noting that SVM-SIL failed in classifying images of non-transplanted livers only in one case.

### 3.4.4 Discussion

In this Chapter, an innovative approach to the computer-aided assessment of HS in RGB images acquired with smartphones in the OR was investigated. With respect to the approaches in the literature, the proposed methods only requires RGB images. Moreover, it provides the surgeons with a classification outcome on whether to accept or discard a liver graft.

For the experimental analysis, the highest classification performance was obtained with \( H_{LBP} \). \( F_{GLCM} \) performed worse with respect to \( H_{LBP} \) and this is probably due to the GLCM lack of robustness to illumination condition changes. In fact, when acquiring liver images, no assumption on keeping the illumination constant was done, resulting in different levels of illumination in the images. Similarly, also \( Stat_1 \) features were not able to face such variability in the illumination.

When using \( H_{LBP} \), RF performed in the same way as SVM, as significant differences were not found when comparing RF and SVM performance. This is something expected, if one compares the obtained results with the literature (e.g. [52]).

A direct comparison with the state of the art results was not possible, due to the lack of benchmark datasets. Moreover, as reported in Sec. 3.4, the methods in the literature only correlated hepatic physical characteristics with the HS level, not providing a method for graft quality assessment.

By visually inspecting the wrongly classified patches (Fig. 3.19), it emerged that misclassification occurred for patches that are challenging to classify also for the human eye. In fact, images were acquired without a controlled acquisition protocol, making the classification not trivial. With a larger training dataset, the classification results can be further improved.

Also in this case, introducing the confidence measure and discarding low-confidence patches helped in improving classification recall for rejected-liver patches.

The classification approach was extended to semi-supervised classification, as the ground-truth biopsy HS classification was provided for the whole image, and not for the single patch. With the semi-supervised method, the highest classification performance was obtained with textural features combined with blood-test features, for both SVM + Clustering and SVM-SIL classification. Indeed, blood-test feature inclusion helped increasing the classification accuracy with respect to using only textural features. However, \( Blo \) alone was not sufficient to achieve accurate HS diagnosis. Indeed, during
a preliminary analysis, an overall $Acc$ of 0.75 was achieved with simple SVM-based classification when considering $Blo$ alone. This demonstrated that textural-information inclusion is a valid support for HS diagnosis.

As for the simple patch-based classification proposed in Sec. 3.2, the semi-supervised approach proposed in Sec. 3.4.1 showed promising results in assessing HS without the need of additional imaging-instrumentation, as indeed required by other methods presented in the literature [74]. Nonetheless, with respect to the approach in Sec. 3.2, the semi-supervised approach did not make any assumption of patch ground-truth labels, resulting in a more reliable and robust classification outcome. Indeed, SVM-SIL misclassified non-transplanted liver images only in one case.

A limitation of this work could be the reduced number of images and this, indeed, is a typical problem of the computer-assisted diagnosis community [133]. Hence, as future work, we aim at enlarging the training dataset. Having a wider dataset will also allow investigating more advanced machine learning methods. Nonetheless, it has been reported that SVM-SIL often achieves competitive results when compared with other more sophisticated semi-supervised methods [188].

3.4.5 Conclusion

In conclusion, the most significant contribution of this work is showing that LBP-based features and semi-supervised SVM can be used as support for HS assessment. This is highly beneficial for practical uses as the method can be potentially developed to run in real-time, being compatible with the short time available between the time of liver procurement and the LT. Moreover, the only required imaging source is a standard RGB camera, which can be easily used in the OR without requiring additional imaging sources such as CT, Raman spectrometer or US.

It is acknowledged that further research is required to further ameliorate the algorithm as to offer all possible support for diagnosis and achieve classification performance comparable with those obtained with biopsy. However, the results presented here are surely a promising step towards a helpful processing system to support the decision process for HS assessment in liver procurement setting.
CHAPTER 4

Uncertainty-aware organ classification in laparoscopy

This Chapter expands the methodology M2 presented in Chapter 3 by investigating the use of multispectral imaging data for tissue classification. Multispectral imaging is an optical technique that enables us to capture both spatial and spectral information on structures. Thus, MI can potentially reveal tissue-specific optical characteristics better than standard RGB imaging systems. In particular, this Chapter describes for the first time in the literature, a framework (M3) for multispectral imaging data acquisition and processing during in vivo laparoscopy, with the goal of performing abdominal-tissue classification. Six abdominal organs are considered, namely abdominal wall, liver, spleen, gallbladder, diaphragm and intestine. Classification outcomes obtained with RGB and multispectral data are compared. From confident regions in the processed images, an approach to automatic image tagging with the organs in the image is presented, too.

A journal paper on this work has been published as [159]: S. Moccia, S. Wirkert, H. Kenngott, A. S. Vemuri, M. Apitz, B. Mayer, E. De Momi, L. S. Mattos and L. Maier-Hein, “Uncertainty-Aware Organ Classification for Surgical Data Science Applications in Laparoscopy,” IEEE Transactions on Biomedical Engineering) (DOI: 10.1109/TBME.2018.2813015).
Chapter 4. Uncertainty-aware organ classification in laparoscopy

4.1 Introduction

As introduced in Chapter [1], SDS has recently emerged as a new scientific field which aims to improve the quality of interventional healthcare [133]. In endoscopy, some of the major opportunities that SDS offers to improve surgical outcomes are surgical decision support [142] and context awareness [94]. Here, technical challenges include the detection and localization of anatomical structures. To date, however, clinical translation of the developed methodology continues to be hampered by the poor robustness of the existing methods. In fact, a grand international initiative on SDS [133] concluded that the robustness and reliability of SDS methods are of crucial importance.

In this context, the aim of this Chapter is investigating the hypothesis $H_3$ that the tissue classification with ML and confidence estimation presented in Chapter [3] can be further improved by using multispectral imaging data. $H_3$ will be experimentally tested in the specific context of organ classification and image tagging in laparoscopic video images. Thus, guided by the subhypotheses that ($H_{3.1}$) automatic confidence estimation can significantly increase the accuracy and robustness of automatic laparoscopic image labeling methods, and that ($H_{3.2}$) MI data are more suitable for in vivo anatomical structure labeling than RGB data, the contributions of this Chapter are summarized as follows:

1. Uncertainty-aware organ classification (Sec. 4.3.1): Development of a new method for superpixel ($Spx$)-based anatomical structure classification, which features an intrinsic confidence measure for self-performance estimation and which can be generalized to MI data;

2. Automatic image tagging (Sec. 4.3.2): Development of an approach to automatic image tagging, which relies on the classification method and corresponding confidence estimation to label endoscopic RGB/multispectral images with the organs present in that image;

3. In vivo validation (Sec. 4.4): A comprehensive in vivo study is conducted using seven pigs to experimentally investigate hypotheses $H_{3.1}$ and $H_{3.2}$.

It is worth noting that image tagging refers to the action of identifying organs present in an image. Instead, organ classification refers to the classification of the organ present in an $Spx$.

To the best of the author’s knowledge, this is the first at using MI data for in vivo abdominal-tissue classification.

4.2 Related work

First attempts at image-guided classification of tissues in RGB endoscopic images primarily used parameter-sensitive morphological operations and intensity-based thresholding techniques, which are not compatible with the high levels of inter-patient multi-organ variability (e.g. [110, 149]). The method for multiple-organ segmentation in laparoscopy reported in [169] relied on non-rigid registration and deformation of pre-operative tissue models on laparoscopic images using color cues. This deformation was achieved using statistical deformable models, which may not always represent the patient-specific tissue deformation, thus resulting in a lack of robustness in terms of
4.2. Related work

Automatic image tagging (Sec. 4.3.2)

Data acquisition and pre-processing

Feature extraction

Superpixel-based classification

Confidence estimation

Uncertainty-aware tissue classification (Sec. 4.3.1)

Superpixel $n$:
- Abdominal wall
- Liver
- Spleen
- Gallbladder
- Diaphragm
- Intestine

High-confidence gallbladder

High-confidence liver

Low-confidence diaphragm

Tagging result:
- Liver
- Gallbladder

Fig. 4.1: Workflow of proposed approaches for uncertainty-aware organ classification and automatic image tagging.
Chapter 4. Uncertainty-aware organ classification in laparoscopy

inter-patient variability. Recently, machine learning based classification algorithms for tissue classification have been proposed to attenuate this issue. The method described in [38] exploited a machine learning approach to segment the uterus. Gabor filtering and intensity-based features were exploited to segment the uterus from background tissues with SVM and morphology operators. However, this approach is limited to single organ segmentation and the performance is influenced by the position of the uterus. Similarly, the method presented in [185] was specifically designed for segmentation of fallopian tubes, as it exploits tube-specific geometrical features, such as orientation and width, and cannot be transferred to other anatomical targets.

As introduced in Chapter [1], in parallel to the development of new computer-assisted strategies to tissue classification, the biomedical imaging field is also evolving thanks to new technologies such as MI [116]. MI is an optical technique that enables us to capture both spatial and spectral information on structures. MI provides images that generally have dozens of channels, each corresponding to the reflection of light within a certain wavelength band. Multispectral bands are usually optimized to encode the informative content which is relevant for a specific application. Thus, MI can potentially reveal tissue-specific optical characteristics better than standard RGB imaging systems [116].

One of the first in vivo applications of MI was proposed by Afromowitz et al. [2], who developed a MI system to evaluate the depth of burns on the skin, showing that MI provides more accurate results than standard RGB imaging for such application. For abdominal-tissue classification, Akbari et al. [3] and Triana et al. [225] exploited pixel-based reflectance features in open surgery and ex vivo tissue classification. The work that is most similar to the present study was recently presented by Zhang et al. [248]. It pointed out the advantages of combining both reflectance and textural features. However, the validation study for this focused on patch-based classification and was limited to ex vivo experiments in a controlled environment, including only 9 discrete endoscope poses to view the tissues, with only single organs in the image and without tissue motion and deformation. Furthermore, the challenges of confidence estimation were not addressed.

As for automatic laparoscopic image tagging, there is no previous work in the literature that has specifically addressed this challenging topic. However, it has been pointed out that there is a pressing need to develop methods for tagging images with semantic descriptors, e.g. for decision support or context awareness [24, 79]. For example, context-aware augmented reality (AR) in surgery is becoming a topic of interest. By knowing the surgical phase, it is possible to adapt the AR to the surgeon’s needs. Contributions in the field include [94, 95]. The AR systems in [94, 95] provide context awareness by identifying surgical phases based on (i) surgical activity, (ii) instruments and (iii) anatomical structures in the image. This is something that is commonly assumed as standard [168]. However, a strategy for retrieving the anatomical structures present in the image was not proposed.

A possible reason for such a lack in the literature can be seen in the challenging nature of tagging images recorded during in vivo laparoscopy. Tissues may look very different across images and may be only partially visible. The high level of endoscopic image noise, the wide range of illumination and the variation of the endoscope pose with respect to the recorded tissues further increase the complexity of the problem. As a result, standard RGB systems may be not powerful enough to achieve the task, even
when exploiting advanced machine learning approaches to process the images. With H3.1 and H3.2, this Chapter aims at investigating if the use of MI and the introduction of a measure of classification confidence may face such complexity.

4.3 Methods

Figure 4.1 shows an overview of the workflow of the proposed methods for uncertainty-aware organ classification (Sec. 4.3.1) and automatic image tagging (Sec. 4.3.2).

4.3.1 Uncertainty-aware tissue classification

The steps comprising the proposed approach to organ classification are presented in the following paragraphs.

Pre-processing

To remove the influence of the dark current and to obtain the spectral reflectance image \( (S_r(\lambda_i)) \) for each MI channel \( i \in [1, N_C] \), where \( N_C \) is the number of MI bands, the raw image \( (I(\lambda_i)) \) was pre-processed by subtracting the reference dark image \( (D(\lambda_i)) \) of the corresponding channel from the multispectral image. \( \lambda_i \) refers to the band central wavelength of the \( i^{th} \) channel. This result was then divided by the difference between the reference white image \( (W(\lambda_i)) \) of the corresponding channel and \( D(\lambda_i) \), as suggested in \[136\]:

\[
S_r(\lambda_i) = \frac{I(\lambda_i) - D(\lambda_i)}{W(\lambda_i) - D(\lambda_i)}
\]

(4.1)

Note that \( W(\lambda_i) \) and \( D(\lambda_i) \) had to be acquired only once for a given camera setup and wavelength. These images were obtained by placing a white reference board in the field of view and by closing the camera shutter, respectively. Each reflectance image was additionally processed with anisotropic diffusion filtering to remove noise while preserving the sharp edges \[105\]. The specular reflections were segmented by converting the RGB image into hue, saturation, value (HSV) color space and thresholding the V value. They were then masked from all channels \[157\].

Feature extraction

In the method proposed in this study, features were extracted from \( S_{px} \). \( S_{px} \) were selected because, compared to regular patches, they are built to adhere to image boundaries better \[117\]. This characteristic is particularly useful considering the classification of multiple organs within one single image. To obtain the \( S_{px} \) segmentation, linear spectral clustering (LSC) \[117\] was applied to the RGB image and then used the obtained \( S_{px} \) segmentation for all multispectral channels.

Inspired by the recently published ex vivo study by Zhang et al. \[248\], both textural and spectral reflectance features were extracted from each multispectral channel. Indeed, as stated in Sec. 4.1, the authors demonstrated that incorporating textural information improved the classification performance with respect to single pixel-based features in their controlled experimental setup. As laparoscopic images are captured from various viewpoints under various illumination conditions, the textural features
Chapter 4. Uncertainty-aware organ classification in laparoscopy

A feature vector is extracted from each $n_{th}$ superpixel ($S_{px_n}$), where $N$ is the number of superpixels in the image. The feature vector for $S_{px_n}$ is obtained by concatenating the histogram ($H_{LBP}$) of uniform rotation-invariant local binary pattern ($LBP_{R;P}^{riu}$) and the average spectrum ($AS$), for each $i \in N$. $N$ image channel, where $N_C$ is the number of channels in the image.
4.3. Methods

should be robust to the pose of the endoscope as well as to the lighting conditions. Furthermore, their computational cost should be negligible to enable real-time computation with a view to future clinical applications.

The $H_{LBP}$ (Sec. 3.2.1), which fully meets these requirements, was here used to describe the tissue texture of an $Spx$. $H_{LBP}$ was normalized to the unit length to account for the different pixel numbers in an $Spx$.

Spectral reflectance information was encoded in the average spectrum ($AS$), which is the average spectral reflectance value in an $Spx$. The $AS$ for the $i$th channel and the $n$th $Spx$ ($Spx_n$), with $n \in (1, N)$ and $N$ the total number of $Spx$, is defined as:

$$AS_{Spx_n}(\lambda_i) = \frac{1}{M} \sum_{p \in Spx_n} Sr_p(\lambda_i) \quad (4.2)$$

where $M$ is the number of pixels in $Spx_n$ and $Sr_p(\lambda_i)$ is the reflectance value of the $p$th pixel of $Spx_n$ in the $i$th channel.

The L2-norm was applied to the $AS$ in order to accommodate lighting differences. $AS$ was exploited instead of the simple spectral reflectance at one pixel to improve the feature robustness against noise, although this is detrimental to spatial resolution.

The steps for obtaining the feature vector are shown in Fig. 4.2.

Superpixel-based classification

To classify the $Spx$-based features, SVM with the radial basis function (Sec. 2.3.2) were used. In this work, SVM hyperparameters $\gamma$ and $C$ were computed with grid search, as explained in Sec. 4.4.

Since the classification task is a multiclass classification problem, SVM with the one-against-one scheme. In particular, six organ classes were involved in the SVM training process, as described in Sec. 4.4. Prior to classification, the feature matrix was standardized within each feature dimension.

Similarly to the strategy presented in Sec. 3.2.3 as a prerequisite for the confidence estimation, the probability $Pr(Spx_n = j)$ for the $n$th $Spx$ to belong to the $j$th organ ($j \in [1, J]$) was retrieved, where $J$ is the number of considered organs. In particular, $Pr(Spx_n = j)$ was obtained, according to the pairwise comparison method proposed in [240] (which is an extension of [181] for the binary classification case), by solving:

$$Pr(Spx_n = j) = \sum_{i=1,i\neq j}^{J} \frac{Pr(Spx_n = j) + Pr(Spx_n = i)}{J - 1} r_{ji}, \forall j \quad (4.3)$$

subject to:

$$\sum_{j=1}^{J} Pr(Spx_n = j) = 1, \quad Pr(Spx_n = j) \geq 0, \quad \forall j \quad (4.4)$$

where $r_{ji}$ is the estimates of $Pr(Spx_n = j|Spx_n \in \{i, j\})$ with $r_{ji} + r_{ij} = 1, \forall j \neq i$. The estimator $r_{ji}$ was obtained according to [181], mapping the SVM output to probabilities by training the parameters of a sigmoid function.
Chapter 4. Uncertainty-aware organ classification in laparoscopy

Table 4.1: Camera light-filter central wavelengths and full width at half maximum (FWHM) for each \( i (= 1–8) \) band in multispectral imaging (MI) and RGB.

<table>
<thead>
<tr>
<th></th>
<th>MI</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>470 nm</td>
<td>480 nm</td>
<td>511 nm</td>
<td>560 nm</td>
<td>580 nm</td>
<td>600 nm</td>
<td>660 nm</td>
<td>700 nm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RGB</td>
<td>470 nm</td>
<td>-</td>
<td>-</td>
<td>560 nm</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>700 nm</td>
</tr>
<tr>
<td></td>
<td>FWHM</td>
<td>20 nm</td>
<td>25 nm</td>
<td>20 nm</td>
<td>20 nm</td>
<td>20 nm</td>
<td>20 nm</td>
<td>20 nm</td>
<td>20 nm</td>
</tr>
</tbody>
</table>

Fig. 4.3: Challenges of the evaluation dataset. Four samples of images showing the gallbladder (first row) and spleen (second row) are reported. Images were recorded with varying endoscope pose and illumination level. Specular reflections are present in the images due to the smooth and wet organ surfaces. Multiple organs can be present in a single image. All images refer to the same multispectral channel.

Confidence estimation

To estimate the SVM classification performance, two intrinsic measures of confidence were evaluated: (i) \( GC \) (Eq. 3.7), which has been introduced in Sec. 3.2.3, and (ii) a measure based on the normalized Shannon entropy \( E \), called posterior probability certainty index (PPCI) \([137]\).

For the \( n^{th} \) \( Spx \), \( PPCI(Spx_n) \) is defined as:

\[
PPCI(Spx_n) = 1 - E(Spx_n)
\]  

(4.5)

where \( E \) is:

\[
E(Spx_n) = -\sum_{j=1}^{J} Pr(Spx_n = j)\log(Pr(Spx_n = j))\frac{1}{\log(J)}
\]  

(4.6)

and:

\[
\log(Pr(Spx_n = j)) = \begin{cases} 
\log(Pr(Spx_n = j)), & \text{if } Pr(Spx_n = j) > 0 \\
0, & \text{if } Pr(Spx_n = j) = 0
\end{cases}
\]  

(4.7)
4.4 In vivo validation

4.3.2 Automatic image tagging

Automatic image tagging uses the SVM $Sp_x$-based classification and the corresponding confidence estimation. Specifically, test images were tagged considering $Sp_x$ labels with high confidence values only. Similarly to what was introduced in Sec. 3.2.3, the value of $GC(Sp_x_n)$ was thresholded to obtain binary confidence information. An $Sp_x$ was considered to have an acceptable confidence level if $GC(Sp_x_n) > \tau$, for the threshold $\tau$. The same procedure was performed using $PPCI$ instead of $GC$.

4.4 In vivo validation

Seven pigs were used to examine the H3.1 and H3.2 introduced in Sec. 4.1. Raw multispectral images ($I$) were acquired using a custom-built MI laparoscope. In this study, the multispectral laparoscope was comprised of a Richard Wolf (Knittlingen, Germany) laparoscope and a 5–MP Pixelteq Spectrocam (Largo, FL, USA) multispectral camera. The $\lambda_i$ for each $i^{th}$ band index and the corresponding full widths at half maximum (FWHM) are reported in Table 4.1. The filters were chosen according to the band selection strategy for endoscopic spectral imaging presented in [237]. The method makes use of the Sheffield index [207], which is an information theory based band selection method originally proposed by the remote sensing community. The 700, 560 and 470 nm channels were chosen to simulate RGB images as the camera did not provide RGB images directly. The image size was $1228 \times 1029 \times 8$ for MI and $1228 \times 1029 \times 3$ for RGB.

The physical size of the multispectral camera was 136 x 124 x 105 mm, with a weight of 908 g. The acquisition time of one multispectral image stack took 400 ms.

From the seven pigs, three pigs were used for training (29 images) and four for testing (28 images). The number of images used to test the SVM performance on RGB and MI data was the same, as RGB data were directly obtained from MI data by selecting 3 of the 8 MI channels. The total number of $Sp_x$ in the training and testing dataset, for both MI and RGB data, was 1382 and 1559, respectively.

Six porcine organ tissues typically encountered during hepatic laparoscopic surgery were considered: the liver, gallbladder, spleen, diaphragm, intestine, and abdominal wall. These tissues were recorded during in vivo laparoscopy. Challenges associated with the in vivo dataset include:

- Wide range of illumination
- Variation of the endoscope pose
- Presence of specular reflections
- Presence of multiple organs in one image
- Organ movement

Visual samples of the dataset challenges are shown in Fig. 4.3.

The multispectral images were pre-processed as described in Sec. 4.3.1. The $Sp_x$ segmentation with LSC was achieved using an average $Sp_x$ size of $150^2$ pixels and an $Sp_x$ compactness factor of 0.1. Accordingly, 55 $Sp_x$ on average were obtained for each image. The $LBP_{\text{H,}r_iu_2}^{P}$ were computed considering the following $(r, p)$ combinations: (1,
Chapter 4. Uncertainty-aware organ classification in laparoscopy

Table 4.2: Median superpixel-based accuracy rate ($\text{Acc}_{\text{Spx}}$) and inter-quartile range (in brackets) for RGB and multispectral imaging (MI) using different features for the Base case (i.e., without confidence inclusion). $H_{LBP}$: Histogram of local binary patterns; AS: Average spectrum.

<table>
<thead>
<tr>
<th></th>
<th>$H_{LBP}$</th>
<th>AS</th>
<th>$H_{LBP}$ + AS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RGB</td>
<td>MI</td>
<td>RGB</td>
</tr>
<tr>
<td>$\text{Acc}_{\text{Spx}}$</td>
<td>63% (17%)</td>
<td>77% (13%)</td>
<td>76% (39%)</td>
</tr>
</tbody>
</table>

Fig. 4.4: Effect of confidence threshold ($\tau$) on the superpixel-based organ classification accuracy rate ($\text{Acc}_{\text{Spx}}$) for RGB and multispectral imaging (MI). Base refers to classification without confidence estimation. The stars indicate significant differences. Dots indicate outliers. The confidence is computed with: (a) the Gini coefficient ($GC$), (b) the posterior probability certainty index ($PPCI$).

8), (2, 16), and (3, 24). The feature vector for an $Spx$ was obtained by concatenating the $H_{LBP}$ with the $AS$ value for all 8 multispectral channels (for MI) and for $\lambda_i = 700$, 560 and 470 nm (for RGB). The feature vector size for an $Spx$ was:

$$\left(l_{H_{LBP}} + l_{AS}\right) \times N_C$$

where $l_{H_{LBP}}$ is the length of $H_{LBP}$, equal to 54, $l_{AS}$ is the length of $AS$, equal to 1, and $N_C$ is the number of channels, 3 for RGB and 8 for multispectral data.

The SVM kernel parameters were retrieved during the training phase via grid-search and 10-fold cross-validation on the training set. The grid-search spaces for $\gamma$ and $C$ were set to $[10^{-8}, 10^1]$ and $[10^1, 10^{10}]$, respectively, with 10 values spaced evenly on the $\log_{10}$ scale in both cases. The determined values for the hyperparameters were subsequently used in the testing phase.

The feature extraction was implemented using OpenCV\(^1\). The classification was implemented using scikit-learn\(^2\).

Investigation of H3.1

To investigate whether the inclusion of a confidence measure increases $Spx$-based organ classification accuracy ($\text{Acc}_{\text{Spx}}$), the $\text{Acc}_{\text{Spx}}$ dependence on $\tau \in [0.5 : 0.1 : 1)$, applied to both $GC$ and $PPCI$, was evaluated. $\text{Acc}_{\text{Spx}}$ is defined as the ratio of correctly classified confident $Spx$ to all confident samples in the testing set. To evaluate

---

\(^1\)http://opencv.org/
\(^2\)http://scikit-learn.org/
4.5. Results

Fig. 4.5: Confusion matrix for confidence threshold $\tau = 0.9$ on the Gini coefficient and multispectral imaging. The values are in percentages and the colorbar indicates the number of superpixels.

whether differences existed between $Acc_{\text{sp}}$ obtained applying $GC$ and $PPCI$ on the SVM output probabilities, the Wilcoxon signed-rank test for paired samples ($\alpha = 0.05$) was used.

The SVM performance with the inclusion of confidence when leaving one organ out of the training set was also investigated. Specifically, six SVMs were trained, leaving each time one organ out. For each of the six cases, the percentage ($\%LC_{\text{sp}}$) of low-confidence $\text{sp}$ (considering $\tau = 0.9$) was computed. This was done both for the organ that was excluded ($Ex$) from the training set and for the included organs ($In$). For image tagging, the tagging accuracy ($Acc_{\text{tag}}$) for different $\tau$ was computed, where $Acc_{\text{tag}}$ is the ratio of correctly classified organs in the image to all organs in the testing image.

Investigation of H3.2

To investigate whether MI data are more suitable for anatomic structure classification than conventional RGB video data, the same analysis was performed for RGB and compared the results with those from the MI. To complete the evaluation, the performance of $H_{\text{LBP}}$ alone and $AS$ alone was also evaluated for $\tau = 0$, which corresponds to the $Base$ case, i.e., SVM classification without a confidence computation. Since the analyzed populations were not normal, the Wilcoxon signed-rank test for paired samples was used to assess whether differences existed between the mean ranks of the RGB and MI results (significance level $= 0.05$).

4.5 Results

The descriptive statistics of $Acc_{\text{sp}}$ for the analyzed features are reported in Table 4.2. For the $Base$ case, the highest $Acc_{\text{sp}}$ (median = 90%, IQR = 6%) was obtained with $H_{\text{LBP}} + AS$ and MI. The other results all differ significantly (p-value < 0.05) from those obtained with $H_{\text{LBP}} + AS$ and MI.

When $\tau$ applied to $GC$ (Fig. 4.4(a)) and $PPCI$ (Fig. 4.4(b)) was varied in [0.5 :
Chapter 4. Uncertainty-aware organ classification in laparoscopy

Fig. 4.6: Image tagging accuracy ($\text{Acc}_{Tag}$) for RGB and multispectral imaging (MI) for Base case and following introduction of confidence measure ($\tau = 0.9$ on the Gini coefficient). The stars indicate significant differences. Dots indicate outliers.

0.1 : 1), the median $\text{Acc}_{Spx}$ for the MI data increased monotonously to 99% ($\tau = 0.9$), when using both GC and PPC1. The same trend was observed for the RGB data, with an overall improvement of the median from 81% to 93% (using GC) and 91% (using PPC1). For both the Base case and after introduction of the confidence measures, the MI outperformed the RGB (p-value < 0.05). No significant differences were found when comparing the classification performance obtained with GC and PPC1. Therefore, as GC computation is more sensitive to high values and faster to compute than PPC1, GC was used.

Figure 4.5 shows the confusion matrix for MI and $\tau = 0.9$ on GC. Note that, in the case yielding the least accurate result, which corresponds to spleen classification, the accuracy rate still achieved 96%, whereas for RGB the lowest accuracy rate was 69%.

The $\% \text{LC}_{Spx}$ boxplots relative to the leave-one-organ out experiment are shown in Fig. 4.7. The $\% \text{LC}_{Spx}$ is significantly higher for organs that were not seen in the training phase (MI: 42% (Ex) vs. 23% (In); RGB: 36% (Ex) vs. 40% (In)).

When applied to endoscopic image tagging, the mean $\text{Acc}_{Tag}$ values in the experiments were increased from 65% (RGB) and 80% (MI) to 90% (RGB) and 96% (MI) with the incorporation of the confidence measure (using GC). The descriptive statistics are reported in Fig. 4.6. In this instance, the MI also outperformed the RGB both in the Base case and with the confidence measure (p-value < 0.05). Figure 4.8 shows the influence of low-confidence $Spx$ exclusion on the image tagging: after low-confidence $Spx$ exclusion, all $Spx$ in the image were classified correctly.

Sample results for the SVM classification and the corresponding confidence map (using GC) are shown in Fig. 4.9. For low-confidence $Spx$, the probable cause of uncertainty is also reported. The main sources of uncertainty are specular reflections, camera sensor noise at the image corner, and the partial organ effect, i.e., when two or more organs correspond to one $Spx$. 

76
4.6. Discussion

The emerging field of surgical data science [133] aims at observing the entire patient workflow in order to provide the right assistance at the right time. One important prerequisite for context-aware assistance during surgical treatment is to correctly classify the phase within an intervention. While a great amount of effort has been put into automatic instrument detection (e.g. [5, 24, 221]), the problem of automatic organ classification has received extremely little extension. This can be attributed to the fact that the task is extremely challenging. In fact, the related problem of organ boundary detection was regarded so challenging by participants of the MICCAI 2017 endoscopic vision challenge (https://endovis.grand-challenge.org/) that only a single team decided to submit results for the sub-challenge deadline with kidney boundary detection. In this work, this problem was tackled by two previously unexplored approaches:

- **Accuracy:** The image acquisition process was slightly changed using a multispectral camera as opposed to a standard RGB camera in order to increase the quality of the input data (for the classifier). The effect of this measure was an increase in accuracy of 11% for the task of organ classification and an increase of 23% for the task of automatic image tagging.

- **Robustness:** A superpixel-based measures of confidence was derived to increase the reliability of image tagging. The result was a boost in accuracy of 38% (RGB) and 20% (MI) absolute.

With the tested validation dataset, MI significantly outperforms standard RGB imaging in classifying abdominal tissues. Indeed, as the absorption and scattering of light in tissue is highly dependent on (i) the molecules present in the tissues, and (ii) the wavelength of the light, the multispectral image stack was able to encode the tissue-specific optical information, enabling higher accuracy in distinguish different abdominal struc-
Chapter 4. Uncertainty-aware organ classification in laparoscopy

Fig. 4.8: Image tagging examples for Base case (left) and following introduction of confidence measure with $\tau = 0.9$ on the Gini coefficient (right). The low-confidence superpixels (in gray) are excluded from the image tagging. The crosses indicate erroneously classified superpixels.
4.6. Discussion

(a) Test image

(b) Superpixel segmentation

(c) Classification for confident superpixels

(d) Confidence map with associated colorbar

Fig. 4.9: (a) Test image, (b) test image with superpixel segmentation, (b) corresponding classification for superpixels with acceptable confidence level and (c) confidence map obtained with confidence threshold \( \tau = 0.9 \) on the Gini coefficient. The symbols give examples of the probable causes of uncertainty.
Chapter 4. Uncertainty-aware organ classification in laparoscopy

tures in comparison to standard RGB.

With the introduction of the confidence measure, the classification accuracy can be improved, for both RGB and MI. This happened when exploiting both GC and PPCI. Since no significant differences were found between GC and PPCI, GC was used as it is more sensitive at higher values than PPCI and its computation is faster. In fact, a major advantage of the proposed method is its high classification accuracy, which attained 93% (RGB) and 99% (MI) in the regions with high confidence levels, with a significant improvement compared to the Base case. Few misclassifications of high-confidence Spx occurred, and where they did then this was mainly with tissues that are also challenging to distinguish between for the human eye, e.g. liver and spleen (Fig. 4.5). It is worth noting that GC and PPCI were two examples of confidence estimation measures to investigate H3.1. We decided against using simple thresholding on the maximum (Max) value of $Pr(Spx_n = j)$ computed among the $J$ organ classes as GC and PPCI are generally known for being more sensitive at higher values [137]. This assumption was confirmed in additional experiments, where image tagging performed with confident Spx according to GC/PPCI was substantially more robust than tagging based on confident Spx according to Max.

The results obtained with the introduction of the confidence measure are comparable with those obtained by Zhang et al. [248] for ex vivo organ classification in a controlled experimental setup. Zhang et al. reported a median classification accuracy of 98% for MI, whereas the classification accuracy for the Base case only achieved 90% due to the challenging nature of the in vivo dataset. An accuracy level comparable to the one of [248] was, however, restored for the tested dataset once the low-confidence Spx were excluded.

When excluding one organ from the training set, $\%LC_{Spx}$ relative to the excluded organ was significantly higher than the number of low-confidence superpixels obtained for the remaining organs. This indicates that the confidence inclusion helped in handling situations where unknown structures appeared in the field of view of the camera. These results are in keeping with those found in the literature for case reasoning [35, 98]. Indeed, the importance of the estimation of the level of confidence of the classification with a view to improving system performance has been widely highlighted in several research fields, such as face recognition [49], spam-filtering [172], and cancer recognition [247]. Similarly, the results achieved with M3 are in keeping with those obtained in Chapter 3 with M2. However, the use of confidence metrics had not been exploited in the context of laparoscopic image analysis, up until now.

Although several Spx misclassifications occurred at the Base case, which had a negative effect on tagging performance, the low-confidence Spx exclusion significantly increased tagging accuracy. Indeed, regions affected by camera sensor noise, specular reflections, and spectral channel shift due to organ movement were easily discarded based on their confidence value. The same process was implemented when the Spx segmentation failed to separate two organs. Also in this case, MI showed that it performs better than standard RGB.

While this is the first attempt at addressing the challenges of in vivo image labeling, including the large variability of illumination, variation of the endoscope pose, the presence of specular reflections, organ movement, and the appearance of multiple organs in one image, one disadvantage of the validation setup is that the testing database was
4.7. Conclusion

not recorded during real surgery. Hence, some of the challenges typically encountered when managing real surgery images were absent (e.g., blood, smoke, and occlusion). Moreover, as the MI camera does not provide RGB data directly, synthetic RGB images were obtained by merging three MI channels. It should be noted, however, that the RGB image encodes more specific information, as the bands used to obtain these data are considerably narrower than those of standard RGB systems (FWHM = 20 nm). A second limitation of the proposed work could be seen in the relatively small number of training images (29). However, analyzing researches on the topic of tissue classification in laparoscopy, such number is comparable with the one of Chhatkuli et al. [38], which exploited 45 uterine images, and Zhang et al. [248], which recorded 9 poses of just 12 scenes (3 pigs \(\times\) 4 ex-vivo organs). Further, it is worth noting that the training was performed at \(Sp_{px}\)-level, meaning that the training set sample size was about \(55 \times 29\), where 55 is the average number of \(Sp_{px}\) in an image.

Considering that the proposed study was not aimed at evaluating the system performance for clinical translation purpose, the clinical requirements of the proposed method performance were not analyzed. Despite recognizing the relevance of such analysis, it should be performed in relation to the specific application. For example, with reference to [95], it would be interesting to analyze and evaluate the requirements of a context-aware AR system supported by the proposed methodology. However, when discussing with the clinical partners, it emerged that the end-to-end accuracy should be close to 100% (i.e. for recognizing the surgical state). However, it has to be further investigated how errors in image tagging affect the error of the final task.

With the exploited MI laparoscope prototype, the image stack acquisition time (400 ms) was faster than most systems commonly presented in literature, like e.g. (e.g. [39] with \(\sim 3\) s), which makes it more advantageous for clinical applications. A further solution to be investigated is the use of loopy belief propagation [89, 165] as post-processing strategy to include spatial information with respect to how confident classification labels appear in the image. This would be particularly useful for images where the tagging failed due to few confident misclassified \(Sp_{px}\) surrounded by correctly classified confident \(Sp_{px}\). Future work will also deal with the real-time implementation of the classification algorithm, which was not the aim of this work. Recent advancements in tissue classification research suggest that the use of CNN could be also investigated for comparison [210]. Indeed, uncertainty in deep learning is an active and relatively new field of research, and standard deep learning tools for classification do not capture model uncertainty [69]. Excluding popular dropout strategies (e.g. [97, 214]), among the most recently proposed solutions, variational Bayes by Backpropagation [23, 175] is drawing the attention of the deep learning community.

4.7 Conclusion

In this Chapter, the challenging topic of robust classification of anatomical structures in \textit{in vivo} laparoscopic images was investigated.

With the first \textit{in vivo} laparoscopic MI dataset, the two hypotheses were confirmed: (H3.1) the inclusion of a confidence measure increases the \(Sp_{px}\)-based organ classification accuracy substantially and (H3.2) MI data are more suitable for anatomic structure classification than conventional video data. To this end, the first approach to anatomic
structure labeling was proposed. The approach features an intrinsic confidence measure and can be used for high accuracy image tagging, with an accuracy of 90% for RGB and 96% for MI.

Despite the proposed experimental protocol being focused on in vivo laparoscopic data, there is the expectation that the potentiality of MI can be easily transferred to other anatomical districts, e.g. for tissue diagnosis. Indeed, tissue optical properties usually change with pathology onset. Using MI would help gathering the optical property changes and thus improving CAD-system performance, bringing the research closer to clinical translation.
CHAPTER 5

Toward improving safety in neurosurgery with an active handheld instrument

During robot-assisted procedures, automatic tissue detection and localization can be exploited for safe tool-trajectory definition or forbidden-region avoidance. Indeed, tissue damaging due to tool misplacement is a well-recognized issue during surgical procedures. In this Chapter, a solution (M4) is proposed to integrate learning-based tissue segmentation in a simulated robot-assisted neurosurgical scenario. In particular, a deep-learning approach to phantom brain-vessel segmentation is investigated to perform vessel avoidance. With respect to the methodology presented in Chapter 3 and Chapter 4, where tissue detection is performed on a patch- or Spx-base, here deep learning allows image-wise tissue segmentation. A handheld robotic tool is exploited and virtual-fixture control is implemented to avoid the robot tip to enter vessel regions.


A journal paper on this work has been submitted to Annals of Biomedical Engineering as: S. Moccia, S. Foti, A. Routraym F. Prudente, A. Perin, R. Sekula, L.S. Mattos, J. Balzer, W. Fellows Mayle, E. De Momi and C. Riviere, “Toward Improving Safety in Neurosurgery with an Active Handheld Instrument” and is currently under revision.
Chapter 5. Toward improving safety in neurosurgery with an active handheld instrument

5.1 Introduction

The use of robotic systems for interventional healthcare has rapidly increased during the last decade [4]. Nowadays, research in robotic systems aims at improving the execution of interventional procedures through providing tremor compensation and virtual-fixture forbidden-region avoidance [222]. Indeed, by selecting regions to be preserved (such as healthy tissues, vessels, nerves) the robot can be prevented to enter such regions, reducing risks related to tissue damaging due to tool misplacement.

Several attempts at introducing virtual-fixture in healthcare processes have been done, but the majority of the proposed methodologies still require the manual definition of regions to be avoided [178], which is a tedious, time-consuming process that would prolong the duration of the surgical procedures. SDS can help solving this issue by performing automatic tissue segmentation. In this context, the goal of this Chapter is to investigate the hypothesis H4 that ML-based tissue segmentation can be integrated in the flow of a simulated robot-assisted surgery, as to provide forbidden-region avoidance.

In particular, this Chapter is focused on neurosurgical procedures. Indeed, during neurosurgery procedures, surgeons perform accurate and minute operations with limited visibility [161]. Such procedures are currently performed under microscope magnification. Bleeding requiring transfusion is recognized as one of the most common complications in cranial surgeries (5.4%) [150]. Detaching the tumor from tiny or large arterial and venous vessels is a daunting challenge; damaging (and then blocking) the physiological blood flow in such structures usually causes brain infarction and a consequent grim neurological performance and prognosis, being the worst neurosurgical scenario to be faced [109, 203]. In particular, petroclivial meningioma resection is known to be among the most technically challenging neurosurgical procedures, due to the meningioma proximity to major blood vessels that serve key and vital nervous structures [51].

As preserving large vessels is of primary importance for lowering postoperative morbidity, an assistive robotic device could be used to filter surgeons’ unintended movements of the human hand and prevent entering forbidden regions such as vascular structures [162]. Examples of assistive devices developed for forbidden-region robotic control include neuroArm [218, 219], a magnetic resonance-compatible robot for image-guided, ambidextrous microneurosurgery. Cooperative control for precision targeting in neurosurgery was explored in [17, 18]. In [212, 223] and in [72], cooperatively controlled robots were proposed with applications in retinal surgery.

Recently, handheld robotic systems have been proposed [75]. Compared to teleoperated and cooperative systems, the biggest advantages of handheld systems are (i) intuitive operation, (ii) safety, and (iii) economy [14]. Moreover, handheld tools have the strong advantage of offering the same intuitive feel as conventional unaided tools. This results in improving the likelihood of acceptance by the surgeon [75]. Among handheld tools, the Navio PFS [123] is a sculpting tool designed for knee arthroplasty. The system combines image-free intraoperative registration, planning, and navigation for bone positioning. KYMERAX has been designed for laparoscopic applications [80]. The system is made of a console, two handles, and interchangeable surgical instruments that can be attached to the handles. Micron [127] has emerged as a powerful
5.2 Related work

actively stabilized handheld tool for applications in retinal vessel microsurgery. Micron is equipped with an optical tracking system and a microscopy stereocamera vision system.

Despite Micron control for tremor compensation being widely investigated (e.g. [127]), forbidden-region virtual fixture control has been investigated only for retinal microsurgery applications [26]. In [26], vessel segmentation and tracking are performed using simultaneous localization and mapping [53]. However, more advanced solutions to the problem of vessel segmentation have been proposed in the last years, with a growing interest in deep-learning strategies [62]. Deep learning allows fast and accurate segmentation also in the presence of challenging vascular architectures, such as bifurcations, and of high noise level and intensity inhomogeneities, typical of microscopy images recorded during surgery [62].

Thus, in this Chapter the use of deep learning to perform fast and accurate vessel segmentation from microscopy images will be investigated as to implement robust forbidden-region virtual-fixture Micron control.

The Chapter is organized as follows: Sec. 5.2 surveys vessel segmentation strategies, with a focus on deep learning. Sec. 5.3 explains the proposed approach M4 to vessel collision avoidance. Sec. 5.4 deals with the experimental protocol used to test the proposed methodology. Results are presented in Sec. 5.5 and discussed in Sec. 5.6. Finally, strength, limitations and future work of the proposed approach are reported in Sec. 5.7. 

5.2 Related work

Despite the most popular vessel segmentation approaches in the past being focused on deformable models (e.g. [34,37]) or enhanced approaches (e.g. [61,197]), in the last years researchers are focusing more and more on machine learning. A comprehensive review of blood vessel segmentation algorithms can be found in [155].

As introduced in Sec. 1, among supervised segmentation models, CNNs have emerged as a powerful tool in many visual recognition tasks [73,104,193]. A CNN is a feed-forward artificial neural network inspired by the organization of the human visual cortex. The typical use of a CNN for vessel segmentation is based on pixel-wise classification: each pixel in the image is assigned to the class vessel or background. The CNN is either used to extract image features then classified with other supervised model approaches [70,233,241], or trained to directly obtain the pixel classification by encoding one or more fully-connected layers [120,184,211].

More complex architectures have been recently proposed to directly deal with the segmentation task [66,115,135]. A quite innovative solution is exploited in [48,148,154], where fully convolutional networks (FCNNs) are used for vessel segmentation. With respect to classification CNNs, FCNNs allow faster training time, lower computational cost during the testing phase and higher segmentation performance.

5.3 Methods

In this section, the proposed approach to safety improvement in neurosurgery through vessel collision avoidance is described. A forbidden-region virtual fixture strategy was designed to control Micron and prevent its tip from colliding with vascular structures.
Chapter 5. Toward improving safety in neurosurgery with an active handheld instrument

Fig. 5.1: (a) Micron handle and moving platform on which infrared LEDs (highlighted in red) are mounted for Micron tip tracking. (b) Micron tip tracking system (ASAP), stereo-cameras for microscopy image acquisition and 3D reconstruction and Micron are shown.

To retrieve vessel segmentation, the deep-learning approach presented in Sec. 5.3.2 was used. The vessel 3D reconstruction and registration to the Micron reference system are explained in Sec. 5.3.3. Finally, the control strategy is reported in Sec. 5.3.4.

5.3.1 Micron architecture

As introduced in Sec. 5.1, Micron is an actively stabilized handheld surgical robot. It is a 6-degree-of-freedom (DOF) system with a Gough-Stewart platform tip manipulator [243]. The Micron actuators are piezoelectric linear motors with 400 μm range of motion, 1N force capability, and bandwidth over 100 Hz (Squiggle SQL-RY-1.8, NewScale Technologies, Victor, N.Y., USA). Micron has two sets of infrared LEDs, of which three are mounted on the moving platform and three are fixed to the handle. The moving and fixed platforms are showed in Fig. 5.1(a). The LEDs are optically tracked by a custom-built tracking system called Apparatus to Sense Accuracy of Position (ASAP) [128]. The ASAP provides Micron position and orientation with a resolution of 4 μm at 2000 samples/s over a 27 cm³ workspace [242]. The Micron microscopy stereocamera vision system is equipped with two Flea2 1024 × 768 cameras (Point Grey Research, Inc., Richmond, B.C.). The complete Micron setup, comprising Micron, ASAP and the stereomicroscope used for image acquisition, is shown in Fig. 5.1(b).

5.3.2 Vessel segmentation algorithm

As introduced in Sec. 5.2, FCNNs have been successfully used for vessel segmentation [48, 148, 154]. In particular, the U-shaped FCNN (Fig. 5.2) architecture proposed in [193] was investigated. The architecture in [193] outperformed other FCNNs in the literature for the task of neuron segmentation in microscopy images. Despite the FCNN being proposed for neuron segmentation, the hypothesis is that the FCNN can be successfully exploited also for the vessel segmentation task, as both axons and vessels have similar tubular architectures.
5.3. Methods

Fig. 5.2: Fully convolutional neural network architecture exploited for vessel segmentation. Conv: Convolution; Pad: Padding; ReLU: Rectified linear unit; MaxPool: Max pooling; Up-Conv: Up-convolution; Copy: Copy layer. The number of feature maps are shown on top of the boxes.

Table 5.1: Nomenclature relative to the 3D reconstruction.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_i$</td>
<td>$i^{th}$ circumference that approximates the 3D vessel surface section</td>
</tr>
<tr>
<td>$C_i$</td>
<td>3D center position of $C_i$</td>
</tr>
<tr>
<td>$\vec{N}_i$</td>
<td>3D orientation versor of $C_i$</td>
</tr>
<tr>
<td>$R_i$</td>
<td>radius of $C_i$ in space</td>
</tr>
<tr>
<td>$L_c_i$</td>
<td>2D center position of $C_i$ in the left image plane</td>
</tr>
<tr>
<td>$L_r_i$</td>
<td>radius of $C_i$ in the left image plane</td>
</tr>
<tr>
<td>$L_e_i$</td>
<td>2D rightmost vessel edge position on $C_i$ associated to $L_c_i$ the left image plane</td>
</tr>
<tr>
<td>$R_c_i$</td>
<td>2D center position of $C_i$ in the right image plane</td>
</tr>
<tr>
<td>$R_r_i$</td>
<td>radius of $C_i$ in the right image plane</td>
</tr>
<tr>
<td>$R_e_i$</td>
<td>2D rightmost vessel edge position on $C_i$ associated to $R_c_i$ in the right image plane</td>
</tr>
<tr>
<td>$E_i$</td>
<td>3D point on $C_i$ obtained triangulating $L_e_i$ and $R_e_i$</td>
</tr>
</tbody>
</table>
Chapter 5. Toward improving safety in neurosurgery with an active handheld instrument

**Fig. 5.3:** The vessel surface can be discretized as a set of points laying on I circumferences. The $i^{th}$ ($i \in [0, I - 1]$) circumference ($C_i$) was characterized by its center 3D position ($C_i$), orientation ($\vec{N}_i$) and radius ($R_i$).

**Fig. 5.4:** 3D reconstruction algorithm workflow. First, the homography transformation ($H$) between the left and right image planes is computed. After homography computation, the assumption was made that the 3D shape of vessels can be approximated by a bent tubular structure with variable radius. Thus, the vessel surface in 3D can be discretized as a set of points lying on I different juxtaposed circumferences ($C_i$, $i \in [0, I - 1]$). From the (left) segmented image, the 2D position of the center ($Lc_i$) and the 2D corresponding rightmost vessel edge point ($Le_i$) of $C_i$ are computed. From $Le_i$ and $Lc_i$, the homography is used to estimate the 2D position of the center ($Rc_i$) and edge point ($Re_i$) of $C_i$ ($i \in [0, I - 1]$) in the right image plane. Direct linear transformation is used to triangulate and obtain $C_i$ and $R_i$. From $C_i$ and $R_i$, the 3D vessel radius $R_i$ and direction $\vec{N}_i$ are computed, $\forall i$. Once $C_i$, $\vec{N}_i$ and $R_i$ are known, $C_i$ is automatically identified. Uniform sampling is used to sample 36 equally spaced points on $C_i$. The set of 36 points on $C_i$, $\forall i$, represents the 3D vessel surface point cloud.
5.3. Methods

Fig. 5.5: The forbidden-region virtual-fixture control algorithm requires computation of the distance between the tool tip resting position ($p_{\text{rest}}$) and the 3D vessel point ($p_{\text{closest}}$) closest to $p_{\text{rest}}$. $p_{\text{rest}}$ is the position Micron takes when no control is triggered. To understand if the control command has to be triggered, first, (left image) the closest point $C_i$ on the 3D vessel centerline ($C_i, i \in [0, I - 1]$, with $I$ equal to the number of centerline points) is found. Then, (central image) $p_{\text{closest}}$ is found among the points on the circumferences $[C_i - 2, C_i + 2]$. Finally, (right image) if $d$ is lower than a predefined threshold ($d_{\text{safety}}$), the Micron control is activated and the new Micron tip position ($p_{\text{control}}$) is defined, which lays on the line passing through $p_{\text{closest}}$ and the current $p_{\text{rest}}$ and is distant $d_{\text{safety}}$ from $p_{\text{closest}}$.

The FCNN consisted of a contractive (descending) path and an expansive (ascending) one. The contractive path consisted of the repeated application of two $3 \times 3$ convolutions ($\text{Conv}$), each followed by a rectified linear unit ($\text{ReLU}$) and a $2 \times 2$ max-pooling ($\text{MaxPool}$) operation with stride 2 for downsampling. At each downsampling step, the number of feature channels was doubled. Every step in the expansive path consisted of an upsampling of the feature map with a $2 \times 2$ up-convolution ($\text{UpConv}$) that doubled the size of the feature map and halved the number of feature channels, followed by two $3 \times 3$ convolutions and a $\text{ReLU}$ each. The copy layers ($\text{Copy}$), peculiar of the exploited architecture, were introduced in the expansive path to retrieve the information lost in the contractive path due to the $\text{MaxPool}$ operations. In the original work [193], the introduction of $\text{Copy}$ proved to be particularly useful for improving segmentation accuracy.

The proposed CNN topology had only four out of the five original layers described in [193], to shorten the time consumption. At the final layer, a $1 \times 1 \text{Conv}$ was used to map each feature vector to the two classes (0: background, 1: vessel).

For training purposes, ADAM [99] was used. ADAM is an algorithm for first-order gradient-based optimization of stochastic objective functions and is based on the adaptive estimates of lower-order moments. The method is computationally efficient, has little memory requirements and is invariant to gradient diagonal rescaling.

Image segmentation was performed on the left microscopy image.

5.3.3 3D reconstruction algorithm

To compute the 3D vessel position in the Micron reference frame, vessel images were acquired from the surgical microscope (Fig. 5.1(b)). Following [242], the camera calibration was performed by matching the points corresponding to the Micron tip projections on the image planes of the two cameras and the real Micron tip coordinates acquired with the ASAP positioning system.

First, the homographic transformation ($H$) between the left (L) and right (R) images was computed using speeded up robust features (SURF) [13] and the fast library for
approximate nearest neighbors (FLANN) \([163]\).

Once \(H\) was computed, the assumption was that the 3D shape of vessels can be approximated by a bent tubular structure with variable radius, as is commonly done in the literature \([62]\). Thus, the vessel surface in 3D can be discretized as a set of points lying on \(I\) circumferences with different radius. The \(i^{th}\) \((i \in [0, I - 1])\) circumference \((C_i)\) was characterized by its center 3D position \((C_i)\), orientation \((N_i)\) and radius \((R_i)\), as shown in Fig. 5.3.

To compute \(C_i\), \(\tilde{N}_i\) and \(R_i\), the vessel segmentation mask was thinned with the Zhang Suen algorithm \([108]\) to obtain the vessel centerline. Each \(i^{th}\) pixel of the centerline \((Lc_i)\) represented the 2D projection of \(C_i\) in the left image plane. The 2D radius \((Lr_i)\) of the projection of \(C_i\) in the left image plane was determined using the Euclidean distance transform \([60]\). The versor of each radius was computed as the normal to the tangent of the centerline computed in \(Lc_i\). Once the versor and \(Lr_i\) were obtained, the 2D coordinates of the rightmost point \((Lc_i)\) on the vessel edge associated to \(Lc_i\) was retrieved. The \(H\) transformation was applied to \(Lc_i\) and \(Lc_i\) \(R_i\) to obtain the corresponding approximated centerline \(\tilde{C}_i\) and edge \(\tilde{E}_i\) points \(\forall i\) on the right image.

Using the direct linear transform \([54]\), \(Lc_i\) and \(R_i\) \(\forall i\) were triangulated, obtaining \(C_i\), \(\forall i\). The same procedure was carried out for \(Lc_i\) and \(R_i\), obtaining the 3D edge point \(E_i\), \(\forall i\). Then, \(R_i\) was estimated as the 3D distance between each \(\tilde{C}_i\) and \(\tilde{E}_i\). To obtain \(\tilde{N}_i\), the tangent to the 3D vessel centerline in each \(\tilde{C}_i\) was computed, using the finite differences approach to approximate the tangent. Once \(\tilde{N}_i\), \(\tilde{C}_i\) and \(R_i\) were calculated \(\forall i\), the \(C_i, i \in [0, I - 1]\) was obtained with uniform circumference sampling with 36 points \((\tilde{p}_j), j \in [0, 35]\).

The workflow of the 3D reconstruction algorithm is shown in Fig. 5.4. The nomenclature list that describes the 3D reconstruction algorithm is reported in Table 5.1.

### 5.3.4 Forbidden-region virtual fixture

As shown in the schematic in Fig. 5.5, the Micron control command for the forbidden-region virtual fixture was triggered when the distance \((d)\) between the tool tip resting position \((p_{rest})\) (i.e the position of the tip, when control was not triggered) and the 3D vessel position \((p_{closest})\) closest to the tool tip resting position was lower than a predefined threshold \((d_{safety})\):

\[
    d \leq d_{safety}
\]

(5.1)

To speed up the estimation of \(p_{closest}\), we first searched among the \(C_i, i \in [0, I - 1]\),

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum number of iterations</td>
<td>18000</td>
</tr>
<tr>
<td>Batch size</td>
<td>10</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.001</td>
</tr>
<tr>
<td>Exponential decay rate for the first moment estimates</td>
<td>0.9</td>
</tr>
<tr>
<td>Exponential decay rate for the second moment estimates</td>
<td>0.999</td>
</tr>
</tbody>
</table>
5.4. Evaluation protocol

Fig. 5.6: Phantom vascular samples used to test the proposed approach to vessel collision avoidance using Micron. Phantoms are reported in the first two rows. The last row shows images recorded during real petroclival meningioma surgery. Vessels are highlighted with respect to the background.

An experimental set-up was developed in order to simulate Micron–vessel interaction in a surgical environment. 230 phantoms were created, which contained modeling clay vessels lying on a background obtained from microscopy images recorded during real petroclival meningioma surgery. Vessels were different in shape and size, varying from straight cylinders to structures with variable radius and multiple branches. The diameter of these structures was created between 2 mm and 4 mm (median diameter = 2.58 mm). Phantom design was inspired by vascular structures recorded during real petroclival meningioma surgery images, as shown in Fig. 5.6.

Each phantom was recorded with the microscope, obtaining 230 images of size 600 × 800 × 3 pixels. Challenges in the dataset included inhomogeneous background, instrumentation in the camera field of view, varying vessel orientation, width and dis-
Fig. 5.7: Evaluation of the 3D reconstruction algorithm. The 3D cloud was reprojected on the left image plane, obtaining a set of reprojected points \( L^{\text{proj}} \). \( L^{\text{proj}} \) was made of several reconstructed circumferences \( L^j_{C_j} \) with \( j \in [0 : I - 1] \) and \( I \) the number of reconstructed circumferences) with center \( L^j_{C_j} \). For each \( L^j_{C_j} \), the 2 extremes were retained, obtaining a new set of points \( L^{\text{extremes}_j} \). The Euclidean distances between \( L^{\text{extremes}_j} \) and the points of \( L^j_{C_j} \) closest to the vessel mask were computed. Two distances were considered, \( L^j_{d_{in}} \) and \( L^j_{d_{out}} \), depending on whether the extreme lied inside or outside the vessel mask, respectively.

To train the FCNN, the dataset was split in 220 training and 10 testing images. Data augmentation was performed. In particular, to each image in the original dataset, 9 transformations were applied: 45°, 120° and 180° rotations, vertical and horizontal mirroring, barrel distortion, sinusoidal distortion, shearing and a combination of sinusoidal distortion, rotation and mirroring, as suggested in [104]. Data augmentation was performed only for \( M^4 \), and not for \( M^1, M^2, M^3 \), because \( M^1, M^2, M^3 \) exploit classical texture descriptors, such as \( LBP_{R;P}^{r;u} \), which are invariant to rotation. Therefore, applying data augmentation would not increase the informative content available to the classification models.

After data augmentation, the augmented dataset was composed of 2200 training and 100 testing images. Among the training images, 1800 were used for FCNN training and 400 for FCNN parameter tuning.

5.4.1 Parameter settings and implementation

The training images and the corresponding segmentation masks were used to train the FCNN using Caffe [1]. The training parameters are reported in Table 5.2. The maximum number of iterations and the batch size were determined with a trial-and-error procedure. The learning rate and the exponential decay rates were set as in [99].

The FCNN was trained on an Amazon Web Server (AWS) bitfusion Amazon Machine Image (AMI) controlled via SSH. The instance used was a p2.xlarge, which has 12 EC2 Compute Units (4 virtual cores), 61GB of memory, plus a GPU NVIDIA K80 (GK210).

The FCNN-based segmentation, 3D reconstruction and virtual fixture control were
Fig. 5.8: Visual samples of vessel segmentation outcomes. Original RGB phantom vessel images are shown in the first row. The second and third row refer to the ground-truth manual segmentation and the segmentation outcome, respectively.
implemented in C++ using OpenCV\footnote{https://opencv.org/}. An Intel Xeon CPU E5-1607 (3.10GHz with 4 Core), 16 Gb of memory and a GPU NVIDIA Quadro K420 was used.

5.4.2 Evaluation dataset and metrics

The segmentation algorithm was tested on the whole set of 100 testing images. The segmentation algorithm was evaluated in terms of $Acc$ (Eq. 3.4.2), sensitivity ($Se = \text{Rec}_\text{class}$ of Eq. 2.17 for class “vessel”), specificity ($Sp$) and Dice similarity coefficient ($DSC = F1_\text{score}$ of Eq. 2.19 for class “vessel”), computed with respect to gold-standard manual segmentation performed by an expert, as commonly done in the literature \cite{62}. $Sp$ is defined as the number of correctly-classified background pixels divided by the total number of background pixels.

To evaluate the performance of the 3D reconstruction, the 10 testing phantoms were recorded using the stereo microscope. For each stereoImage, once the vessel segmentation and 3D reconstruction were obtained as explained in Sec. 5.3.2 and Sec. 5.3.3 respectively, the number of points ($N_{\text{points}}$) of the 3D point cloud ($p_{\text{cloud}} = \{p_{\text{cloud},i}\}$ with $i \in [0; N_{\text{points}}]$) was computed. The time required to reconstruct $p_{\text{cloud}}$ was measured, too. Further, the $p_{\text{cloud}}$ was reprojected on the left image plane, obtaining a set of reprojected points ($L_{\text{proj}} = \{L_{\text{proj},i}\}$ with $i \in [0; N_{\text{points}}]$). $L_{\text{proj}}$ was made of several reconstructed circumferences ($L^C_j$ with $j \in [0 : I - 1]$) with center $L^c_j$, as shown in Fig. 5.7. For each $L^C_j$, only the two extremes ($L^c_{\text{extremes},j}$) were retained. The Euclidean distances between $L^c_{\text{extremes},j}$ and the points of $L^C_j$ closest to the vessel mask ($\forall j$) were computed. Once the distances were computed for each $L^C_j$, the set of distances was split in two subsets, depending on whether the extreme lay inside ($L^d_{\text{in}}$) or outside ($L^d_{\text{out}}$) the vessel mask, to understand whether the true vessel edges were underestimated or overestimated.

For the forbidden-region virtual-fixture algorithm performance assessment, Micron interacted at random with 25 different vascular phantoms for $\sim 80$ s per trial. The $d_{\text{safety}}$ was set equal to 2 mm, which was comparable with the median vessel diameter. First, for each trajectory, the number of times the tool tip entered the forbidden zone was evaluated. Specifically, the percentage penetration error ($err\%$) was computed, which is defined as:

$$ err\% = \frac{N_{\text{in},\text{forbidden}}}{N_{\text{control}}} \times 100 $$

where $N_{\text{in},\text{forbidden}}$ is the number of $p_{\text{in},\text{forbidden}}$, $N_{\text{control}}$ is the number of $p_{\text{control}}$, and $p_{\text{in},\text{forbidden}}$ is defined as:

$$ p_{\text{in},\text{forbidden}} = \{p_{\text{tip}} \mid \|p_{\text{tip}} - p_{\text{closest}}\| < d_{\text{safety}}\} $$

$err\%$ was expressed as a percentage with respect to the total number of points in which the Micron control was triggered as to exclude Micron position far away from the safety zone (i.e. for which no control is required).

Then, for each trajectory, the absolute error distance ($err$) was computed:

$$ err = d_{\text{safety}} - \|p_{\text{in},\text{forbidden}} - p_{\text{closest}}\| $$

The $err$ allowed understanding of how much the Micron tip entered the forbidden region.
5.4. Evaluation protocol

Fig. 5.9: Point cloud reprojected on left and right image plane plotted over the left (left column) and right (right column) ground-truth vessel segmentation images.
Chapter 5. Toward improving safety in neurosurgery with an active handheld instrument

Table 5.3: Vessel segmentation performance and computational cost for the 100 testing images. Median, first and third quartiles are reported for each metric. DSC: Dice similarity coefficient; Se: Sensitivity; Sp: Specificity; Acc: Accuracy; Time: Computational time [s].

<table>
<thead>
<tr>
<th>Metric</th>
<th>DSC</th>
<th>Se</th>
<th>Sp</th>
<th>Acc</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>First quartile</td>
<td>0.94</td>
<td>0.95</td>
<td>0.99</td>
<td>0.98</td>
<td>0.77</td>
</tr>
<tr>
<td>Median</td>
<td>0.96</td>
<td>0.91</td>
<td>0.99</td>
<td>0.99</td>
<td>0.77</td>
</tr>
<tr>
<td>Third quartile</td>
<td>0.97</td>
<td>0.96</td>
<td>1.00</td>
<td>0.99</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Table 5.4: Number of points \(N_{points}\) and computational cost (Time [s]) for each of the 10 reconstructed 3D vessel clouds considered for evaluating the 3D reconstruction algorithm. Median values are reported, too. The values are reported for the proposed algorithm and for the 3D reconstruction performed with speeded up robust features (SURF) \[13\] and the fast library for approximate nearest neighbors (FLANN) \[163\]. Significant differences were found when comparing \(N_{points}\) computed with the proposed approach and with SURF and FLANN (Wilcoxon signed-rank test (significance level = 0.05)).

<table>
<thead>
<tr>
<th>(N_{points}) (SURF + FLANN)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time [s] (SURF + FLANN)</td>
<td>0.10</td>
<td>0.10</td>
<td>0.11</td>
<td>0.13</td>
<td>0.10</td>
<td>0.11</td>
<td>0.13</td>
<td>0.11</td>
<td>0.11</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>(N_{points}) (Proposed)</td>
<td>10906</td>
<td>15190</td>
<td>12058</td>
<td>19654</td>
<td>17710</td>
<td>14362</td>
<td>16630</td>
<td>17746</td>
<td>21778</td>
<td>17170</td>
<td></td>
</tr>
<tr>
<td>Time [s] (Proposed)</td>
<td>0.53</td>
<td>0.47</td>
<td>0.47</td>
<td>0.66</td>
<td>0.45</td>
<td>0.49</td>
<td>0.44</td>
<td>0.54</td>
<td>0.64</td>
<td>0.38</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Fig. 5.10: Distance \(d\) between Micron tip and 3D vessel point closest to the Micron tip. \(d\) with and without the virtual-fixture control is reported in green and orange, respectively.
5.4. Evaluation protocol

Fig. 5.11: Sample trajectory for virtual fixture evaluation. The reconstructed vessel point cloud is represented in blue; the trajectories with and without the virtual fixtures control are depicted in green and orange, respectively.
Chapter 5. Toward improving safety in neurosurgery with an active handheld instrument

5.5 Results

Visual samples of vessel-segmentation outcome are shown in Fig. 5.8. Table 5.3 reports median, first and third quartile of $DSC$, $Se$, $Sp$ and $Acc$ computed for the vessel segmentation performed on the 100 testing images as well as the computational time required for the segmentation. Segmentation performances were: 0.96 ($DSC$), 0.91 ($Se$), 0.99 ($Sp$) and 0.99 ($Acc$). The computational time required for segmenting one image was 0.77 s on the deployed GPU.

The number of reconstructed points for each cloud and the computational cost required to obtained the clouds are reported in Table 5.4. A median $N_{points}$ of 17170 was obtained, with a median computational cost of 0.48 s. When compared to SURF- and FLANN-based 3D reconstruction, the computational cost was higher than the one required by SURF and FLANN (0.11 s). However, the $N_{points}$ was significantly higher (Wilcoxon signed-rank test, significance level = 0.05) than the $N_{points}$ obtained using SURF and FLANN, equal to 12. The median values of $L_{d_{in}}$ and $L_{d_{out}}$, computed among all the 10 point clouds, were 139.6 $\mu$m and 70.4 $\mu$m, respectively. The 71% of $L_{extremes}$ lied inside and the 28% of $L_{extremes}$ lied outside the vessel mask.

A median value of 114.16 $\mu$m, computed considering both $L_{d_{in}}$ and $L_{d_{out}}$ for all the 10 point clouds, was obtained. Visual samples of the point clouds reprojected on the left and right ground-truth image are shown in Fig. 5.9.

A median $err\%$ among the 25 tested Micron trajectories of 24% (IQR = 10%) was obtained. Median $err$ was 16.9 $\mu$m (IQR = 5.86 $\mu$m). The $d$ for a sample trajectory is reported in Fig. 5.10. Corresponding $err_{rel}$ and $err$ were 21% and 6.7 $\mu$m. Sample Micron trajectories with and without the virtual-fixture are shown in Fig. 5.11.

5.6 Discussion

The segmentation performance (median $DSC = 0.96$) proved that the neural network was able to encode the variability in the simulated surgical environment. The major source of error, as shown in Fig. 5.8, was the presence of specular reflections (i.e. bright and saturated areas) within the vessels. It is worth noting that complex vascular structures (i.e. bifurcation), which are usually challenging to segment [62], were correctly segmented. The median time required to segment each image (0.77 s on the deployed GPU) was still not negligible for real-time control purposes. This problem may be overcome using a more powerful GPU. Indeed, during the experiments, the segmentation time was reduced of $\sim 1/7$ with the Amazon web server (Sec. 5.4.1) that was used to perform FCNN training.

The 3D reconstruction algorithm produced dense point clouds, with a median $N_{points}$ of 17170 points, which was crucial to enable robust Micron control. This can be also seen from the visual samples of Fig. 5.9 where the reprojected points covered the vessel area homogeneously. The analysis of $L_{d_{in}}$ and $L_{d_{out}}$ showed that the median value of all the distances (114.16 $\mu$m) was two orders of magnitude lower than the median vessel diameter (2.58 mm). However, this came at the cost of a median reconstruction time of 0.48 s, which was still not negligible when trying to reach real-time control performance. A possible solution to lower the computational time could be assuming simpler vascular architectures (e.g. constant vessel radius) and having a parametric expression of the vessel.
5.7 Conclusion

When including tremor compensation and forbidden-region virtual-fixture control, the error on the tip position when it entered the forbidden zone was small (the median error was 16.9 μm). However, error outliers were present due to limits in the range of motion of Micron. When trying to reach points outside Micron’s range of motion, the forbidden-region virtual-fixture control algorithm brought the Micron motors to saturation. When this was the case, the tip entered the forbidden zone to a greater depth.

Improving the control strategy and the virtual fixture algorithm could be the first natural evolution of the project. For example, once retrieved, the velocity of Micron could be used to influence the possible control choice to adopt. Faster movements of the handle are more likely to be unwanted and the last tip position before the abrupt movement might be maintained. Moreover, using the center of the range of motion might help to prevent the saturation of the motors. For example, instead of rejecting the tip along the direction between the closest point of the cloud and the resting position of the tip, other directions away from both cloud and saturation zone might be used. Finally, ergonomics tests with surgeons actually performing tumor resection while avoiding vessels in a moving phantom will be performed.

5.7 Conclusion

With this Chapter, H4 was validated experimentally in the context of simulated neurosurgery. The work presented provided a new approach to vessel avoidance for safe robotic assisted neurosurgery, which exploits a handheld tool to reliably constrain surgeon movements outside predefined forbidden zones.

The feasibility of the method M4 was proved in a simulated scenario with phantom vessels, obtaining a median segmentation accuracy of 99% and avoiding the tool tip penetrating the forbidden region 76% of the time the control algorithm was active (median penetration depth 2 orders of magnitude lower than vessel median diameter).

To conclude, it would be interesting to adapt and test the developed methodology to other surgical fields. Indeed, several surgical procedures would benefit from the definition of virtual fixture [56], and the developed methodology would surely help in improving the surgery outcomes.
CHAPTER 6

Conclusion

With the goal of improving the quality of interventional healthcare, the overall aim of this PhD thesis was to develop ML-based SDS algorithms for tissue classification in optical images.

Despite the fact that interventional medicine is increasingly and rapidly advancing, becoming to a greater and greater extent computer-based \[222\], the core of interventional medicine processes still heavily (and almost only) relies on surgeon’s expertise and experience. The surgeon’s decision process is still a subjective process and the assessment of surgical outcomes is still almost entirely qualitative \[133\]. In fact, the integration of quantitative data processing into the actual clinical flow is still hampered by the poor reliability and robustness of the algorithms proposed in the literature.

Thus, through the organization, analysis, and modeling of healthcare data, the methodologies presented in this dissertation were developed with a view to (i) supporting surgeons’ decision process with reliable classification data and (ii) providing robust context awareness.

In Chapter 2, the problem of the selection of informative frames, to be processed by tissue classification algorithms, was addressed. \(M_1\), a new approach to ML-based informative frame selection, was proposed and \(H_1\) was experimentally validated in the context of NBI laryngoscopic videos. Chapter 3 tackled the problem of robust pathological tissue classification by exploiting ML and classification-confidence estimation. \(H_2\) was investigated and confirmed for pathological laryngeal and hepatic tissues with \(M_2\). With Chapter 4, tissue classification performance was further increased, with \(M_3\), by exploiting multispectral imaging data, confirming \(H_3\) for \textit{in vivo} laparoscopic images. Finally, Chapter 5 dealt with the integration of SDS technology in a simulated robot-assisted neurosurgical scenario (\(M_4\)), confirming \(H_4\).
Chapter 6. Conclusion

6.1 Thesis Contributions

The PhD research contributed to:

- Development of a new strategy (M1) to automatic learning-based informative frame selection, which exploits a new set of features to retrieve informative frames to be processed by tissue classification algorithms.

It is widely recognized that, from the clinician’s side, reviewing an endoscopic video is a labour-intensive operation. While focusing on particular structures during the video examination, clinicians may miss important clues indicating suspicious conditions (e.g., early-stage tumors). This process could be further compromised by the presence of uninformative video portions, which prolong the revision time of the endoscopic video. Developing a strategy to select informative frames has the potential to reduce the amount of data to review, lowering the surgeons’ workload. The selection of informative frames can be beneficial also for SDS algorithms that offer decision support, for classifying pathological tissues, and context awareness, e.g. for classifying tissue and surgical tool in the camera field of view. In particular, frame selection strategies can benefit SDS algorithms by (i) lowering the amount of computational power required, and (ii) avoiding the processing of frames that do not show structures of interest. Chapter 2 described a new method to classify endoscopic frames based on intensity, keypoint and image spatial-content features. Support vector machines with the radial basis function and the one-versus-one scheme were used to classify frames as informative, blurred, with saliva or specular reflections, or underexposed. The algorithm was experimentally tested on NBI laryngoscopic videos. Apart from the well-known challenges associated to endoscopy, such as high camera-noise level in the images, major challenges typical of the laryngeal district include movement of swallowing muscles and vocal folds, as well as free and varying endoscope pose, which produces blurring in the images, presence of specular reflections, due to the smooth and wet laryngeal surface, and saliva, varying illumination conditions, resulting in underexposed video frames. The proposed method was retrospectively applied to 1200 frames from 18 videos of 18 different subjects recorded during the clinical practice. The method demonstrated to be robust to inter-patient variability, achieving an overall median classification recall of 84% among four frame classes (i.e. blurred, underexposed, with saliva or SR, and informative frames). Misclassifications mainly occurred between classes of uninformative frames and informative video frames were classified with a recall of 91%. The performances of the proposed approach were significantly higher than those achieved applying other methods in the literature to the evaluation dataset (median recall = 57%, maximum recall = 81%). Moreover, the proposed approach is more robust, faster and simpler to implement since no parameter tuning is required. Despite the evaluation protocol being focused on laryngeal videos, there is the potentiality that the proposed methodology can be successfully integrated as pre-processing step in other anatomical regions.

- Development of a new paradigm (M2) for reliable pathological tissue classifica-
6.1. Thesis Contributions

tion in endoscopic images, which introduces a measure of confidence on classification to retrieve reliable classification results.

Detection and localization of anatomical structures during interventional health-care processes in among the main challenges of SDS. Despite a wide literature on tissue classification already exists, open technical challenges are still present. Main challenges deal with the improvement of classification accuracy, robustness and reliability. Indeed, the classification is not trivial, especially for pathological tissues, due to the high inter- and intra-patient variability. Moreover, varying illumination level, noise in the image, and different camera position with respect to tissues pose further challenges for classification algorithms. Chapter 3 tackled the problem of robust tissue classification by exploiting supervised and semi-supervised SVM-based classification coupled with classification confidence estimation. The confidence measure was based on the dispersion of classification probabilities among the tissue classes. The methodology was tested on laryngoscopy images in NBI and hepatic images acquired in the OR with smartphones, to diagnose early-stage laryngeal cancer and steatosis in hepatic tissue, respectively. These two applications had never been investigated in the literature of SDS so far, with the exception of some preliminary attempts that mostly do not exploit the generalization power of ML. It is worth noting that estimating the classification confidence for these pathological tissues is particularly beneficial. Indeed, tissue biopsy would be required only for low-confidence regions in the image, lowering biopsy risks related to bleeding or infection. When tested on 33 videos, which refer to 33 different patients affected by early-stage laryngeal cancer, a median classification recall of 93% achieved. For the hepatic district, a recall of 82% was achieved when classifying 40 images of healthy and pathological livers. Both for the laryngeal and the hepatic district, the classification recall was increased to 98% (laryngeal tissue) and to 86% (hepatic tissue) by estimating the confidence of the SVM classification, with few misclassification that mainly occurred with high-challenging images, whose classification is not trivial also for the human eye. For the hepatic district, a semi-supervised approach was investigated, too, which showed a recall of 95% in recognizing images with rejected liver grafts. With the work presented in Chapter 3, pathological tissue classification in the laryngeal and hepatic districts was addressed for the first time in the literature. The introduction of the classification confidence helped in increasing the reliability of classification outcome in both cases. The most significant contribution of Chapter 3 is showing that LBP-based features and SVM can differentiate laryngeal and hepatic tissues accurately. This is highly beneficial for practical uses. Comparing with other state-of-the-art method in the area, the proposed method is simpler and the result is more accurate.

- Development of a new protocol (M3) for image analysis, which expands M2 and exploits multispectral imaging for automatic tissue classification and image tagging.

In parallel to the development of new computer-assisted strategies to tissue classi-
Chapter 6. Conclusion

Classification, the biomedical imaging field is also evolving thanks to new technologies such as MI. MI is an optical technique that enables us to capture both spatial and spectral information on structures. MI provides images that generally have dozens of channels, each corresponding to the reflection of light within a certain wavelength band. Multispectral bands are usually optimized to encode the informative content which is relevant for a specific application. Thus, MI can potentially reveal tissue-specific optical characteristics better than standard RGB imaging systems. Chapter 4 explored, for the first time in the literature, the MI potentiality in the context of tissue classification during in vivo laparoscopy. The laparoscopic-image acquisition process was changed using a multispectral camera as opposed to a standard RGB camera in order to increase the quality of the input data (for the classifier). In particular, multispectral data were acquired during laparoscopic procedures on 7 pigs. Six abdominal organs (abdominal wall, liver, spleen, gallbladder, diaphragm and intestine) were classified. To provide context awareness, a solution for image tagging with the organs in the images was proposed, too. The effect of using multispectral data was an increase in accuracy of 11% for the task of organ classification and an increase of 23% for the task of automatic image tagging. When exploiting a superpixel-based measures of confidence (similarly to what introduced with M2), the result was a boost in accuracy of 38% (RGB) and 20% (MI) absolute. M3 significantly enhances the state of art in automatic labeling of endoscopic videos by being the first study to use MI data for in vivo laparoscopic tissue classification. Moreover, despite the proposed experimental protocol being focused on in vivo laparoscopic data, there is the expectation that the potentiality of MI can be easily transferred to other anatomical districts, e.g. for tissue diagnosis. Indeed, tissue optical properties usually change with pathology onset. Using MI would help gathering the optical property changes and thus improving CAD-system performance, bringing the research closer to clinical translation.

- Development of a new method (M4), which integrates a deep-learning strategy to tissue segmentation within a handheld robotic tool, with the goal of performing tissue avoidance in simulated robotic-assisted surgical procedures.

The use of robotic systems for interventional healthcare has rapidly increased during the last decade. Nowadays, research in robotic systems aims at improving the execution of interventional procedures through providing tremor compensation and virtual-fixture forbidden-region avoidance. Indeed, by selecting regions to be preserved (such as healthy tissues, vessels, nerves) the robot can be prevented to enter such regions, reducing risks related tissue damaging due to tool misplacement. Several attempts at introducing virtual-fixture in healthcare processes have been done, but the majority of the proposed methodologies still require the manual definition of regions to be avoided, which is a tedious, time-consuming process that would prolong the duration of the surgical procedures. Chapter 5 investigated methodologies to solve this issue, focusing on performing automatic tissue segmentation for vessel preservation in simulated robot-assisted neurosurgical environments using vascular phantoms. In particular, Chapter 5 investigated the use of a handheld robotic tool (Micron) to automate vessel avoidance in microsurgery. A deep-learning based vessel segmentation in microscopy images was implemented.
and integrated with a feature-based passive 3D reconstruction algorithm to obtain accurate and robust vessel position. A virtual-fixture–based strategy to control the handheld robotic tool was defined to implement vessel avoidance. Clay vascular phantoms, lying on a background obtained from microscopy images recorded during meningioma surgery, were used for testing the segmentation and control algorithms. The phantom segmentation performance (median \(DSC = 0.96\)) proved that the exploited deep learning approach was able to encode the variability in the simulated surgical environment. When including tremor compensation and forbidden-region virtual-fixture control, the error on the robotic tip position when it entered the forbidden zone (24% of the time) was small (the median penetration error was \(16.9 \, \mu m\), which was 2 orders of magnitude smaller than median vessel diameter). Results suggested the proposed solution can assist surgeons in performing safe vessel avoidance during neurosurgical procedures. Moreover, as several surgical procedures would benefit from the automatic definition of virtual fixture, the developed methodology could be adapted for improving procedure safety and after-treatment patient’s recovery.

- **Sharing of SDS data.**

As a further contribution of this PhD thesis, some of the datasets built to test the research hypotheses were released, with the goal of contributing to data sharing among the SDS community. The datasets are available at:

- The *Laryngeal dataset*:  
  [https://zenodo.org/record/1003200#.WdeQcnBx0nQ](https://zenodo.org/record/1003200#.WdeQcnBx0nQ)

- The *NBI-InfFrames* dataset:  
  [https://zenodo.org/record/1162784#.WnCcQJOdUWo](https://zenodo.org/record/1162784#.WnCcQJOdUWo)

Table 6.1 lists all the *in vivo* datasets collected for testing the methodologies presented in this PhD thesis.

### 6.2 Impact

The impact of the proposed work lies in the improvement of interventional-medicine procedures through providing context awareness and decision support.

This topic is of major importance in SDS research, as pointed out during a grand international initiative [133]. Indeed, researchers are constantly working to improve the performance of intraoperative imaging devices and robotic systems as to attenuate the actual shortcomings in interventional medicine. In this perspective, computer-assisted procedures are expected to rapidly increase in number in the near future [223]. However, as highlighted in [4, 224], adverse events are still being experienced during the procedures. For example, during the study period analyzed in [4], 144 deaths (1.4% of the 10,624 reports), 1,391 patient injuries (13.1%), and 8,061 device malfunctions (75.9%) were reported during robot-assisted minimally-invasive procedures. Adverse events mainly dealt with system errors, unintended operation of instruments and video/imaging problems. Malfunctions impacted patients in terms of injuries or death, or procedure interruptions or conversion to more complex surgery.
Table 6.1: Datasets collected for testing the methodologies presented in this PhD Thesis.

<table>
<thead>
<tr>
<th>Name</th>
<th>Imaging modality</th>
<th>Anatomical region</th>
<th>Number of images</th>
<th>Imaging modality</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Laryngeal Dataset</td>
<td>Laryngoscopy</td>
<td>Larynx</td>
<td>33 (1320 patches)</td>
<td>Laryngoscopy</td>
<td>The Laryngeal Dataset</td>
</tr>
<tr>
<td>MI_EndoDB</td>
<td>RGB-camera imaging</td>
<td>Liver</td>
<td>40</td>
<td>RGB-camera imaging</td>
<td>MI_EndoDB</td>
</tr>
<tr>
<td>NBI-IIFrames</td>
<td>Laryngoscopy</td>
<td>Larynx</td>
<td>720</td>
<td>Laryngoscopy</td>
<td>NBI-IIFrames</td>
</tr>
<tr>
<td>NBI-IIFrames</td>
<td>Laryngoscopy</td>
<td>Larynx</td>
<td>720</td>
<td>Laryngoscopy</td>
<td>NBI-IIFrames</td>
</tr>
</tbody>
</table>

https://zenodo.org/record/1003200#.WdeQcnBx0nQ

https://zenodo.org/record/1162784#.WnCcQJOdUWo

https://zenodo.org/record/1003200#.WdeQcnBxonq
6.3. Future Perspective

In this context, the methodologies proposed in this PhD thesis will help in solving some of the open issues in interventional-healthcare processes by providing decision support and context awareness. This will impact positively on the quality of the delivered treatments and, therefore, on the after-treatment quality of patients' life.

The performance level achieved by M1, M2, M3, and M4 is encouraging for methodology translation into the actual clinical practice. In fact, when discussing with the clinical partners it appeared clear that the end-to-end accuracy level of a computer-assisted support system should be as close as possible to 100%, or at least to the performance level of clinicians. Nonetheless, such level is not trivial to estimate, as it varies with the clinician's experience, and according to the image quality provided by the exploited imaging systems. The required end-to-end accuracy level could be achieved by developing online platforms that offer support to clinicians in performing tissue classification while collecting medical data, possibly asking clinicians to correct the algorithm performance with manual annotations.

6.3 Future Perspective

SDS goal is enabling “the understanding of surgical procedures, their variability, crucial parameters, hidden structures, dependencies, optimal pathways, the importance of each parameter and keys to success and failure of methodologies and the basic principles driving the surgical education, training and practice” [133].

This goal cannot be achieved without performing robust and reliable image analysis, to achieve intra-operative registration, 3D reconstruction, tissue tracking, workflow modeling and, of course, tissue detection and localization. However, several open technical challenges are still present, hampering the translation of the SDS methodology into the clinical practice. The PhD work described in this dissertation dealt with a portion of this wider and interconnected field of research. In particular, the algorithms described in this dissertation were developed with a view to reduce the gap between research and clinics by focusing on robust and reliable tissue classification in optical images.

Considering how intensive and rapid the progress in SDS is, I truly expect that researches in intra-operative tissue classification will keep on significantly advancing in the future. In my view, new-generation imaging-hardware design will allow the acquisition of images at higher and higher resolution. For example, as discussed in this dissertation, multispectral systems showed the potentiality of encoding tissue-specific information better than standard RGB systems, and I expect this technology will be soon integrated in interventional-healthcare procedures.

Despite the SDS community increasingly exploiting advanced machine learning for several applications, such as image registration, generation and enhancement, content-based image retrieval, and also for tissue classification itself [121], the intra-operative application scenarios in research are still limited to too few anatomical districts. In my view, as soon as large and labeled datasets will be publicly available, the use of advanced learning approaches to solve the problem of tissue classification will be explored also for the underrepresented fields. Moreover, advanced machine-learning approaches have already been integrated (and are currently in use) in several processes outside the medical research fields, where large annotated datasets are already available to encode the complexity of the classification problem without overfitting. On this background,
Chapter 6. Conclusion

I believe that exploiting the generalization power of advanced machine-learning strategies will help the medical computer-vision community in reducing the gap between the SDS research and its use in actual clinical practice.
Part of the methodologies presented and tested in this PhD dissertation were presented in the following peer-reviewed publications.

**Journal publications**


Chapter 6. Conclusion

...under revision.


Conference proceedings


Workshop abstracts


6.3. Future Perspective

## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Significance level</td>
</tr>
<tr>
<td>Acc</td>
<td>Accuracy</td>
</tr>
<tr>
<td>AR</td>
<td>Augmented reality</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the receiver operating characteristic</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer-assisted diagnosis</td>
</tr>
<tr>
<td>CNN</td>
<td>Convolutional neural network</td>
</tr>
<tr>
<td>CT</td>
<td>Computer tomography</td>
</tr>
<tr>
<td>DSC</td>
<td>Dice similarity coefficient</td>
</tr>
<tr>
<td>entropy</td>
<td>Image entropy</td>
</tr>
<tr>
<td>$F_{1_{\text{class}}}$</td>
<td>Class-specific F1 score</td>
</tr>
<tr>
<td>$FN_j$</td>
<td>False negative of class $j$</td>
</tr>
<tr>
<td>$FP_j$</td>
<td>False positive of class $j$</td>
</tr>
<tr>
<td>GC</td>
<td>Gini coefficient</td>
</tr>
<tr>
<td>GLCM</td>
<td>Gray-level co-occurrence matrix</td>
</tr>
<tr>
<td>$H_{LBP}$</td>
<td>Normalized histogram of rotation-invariant uniform local binary patterns</td>
</tr>
<tr>
<td>HOG</td>
<td>Histogram of oriented gradients</td>
</tr>
<tr>
<td>HU</td>
<td>Hounsfield unit</td>
</tr>
<tr>
<td>IQR</td>
<td>Inter-quartile range</td>
</tr>
</tbody>
</table>
Chapter 6. Conclusion

$J$ Number of classes
kNN k-nearest neighbors
LBP Local binary pattern
$LBP^{{R, P}_{riu2}}$ Rotation-invariant uniform local binary patterns
LDA Linear discriminant analysis
LSC Linear spectral clustering
MI Multispectral imaging
ML Machine learning
MIL Multiple instance learning
MRI Magnetic resonance imaging
NB naive Bayes
NBI Narrow-band imaging
OR Operating room
$PPCI$ Posterior probability certainty index
$\text{Prec}_{\text{class}}$ Class-specific precision
$\text{Rec}_{\text{class}}$ Class-specific recall
RF Random forest
ROC Receiver operating characteristic
SCC Squamous cell carcinomas
SDS Surgical data science
$Se$ Sensitivity
SIL Single instance learning
$Sp$ Specificity
$Sp_{px}$ Superpixel
SR Specular reflections
SVM Support vector machines
T Training set
$TN_j$ True negative of class $j$
$TP_j$ True negative of class $j$
List of Figures

1.1 Main blocks of interventional care processes. Images adapted from [222].
1.2 Medical-data sharing influences the interventional care processes.
1.3 Surgical data science integration in the interventional-medicine flow allows objective decision-making and quantitative evaluation of surgical outcomes.
1.4 Some of the major opportunities that surgical data science offers to interventional medicine [133]. Blocks highlighted in green identify the main topics of this PhD thesis.
1.5 Most popular image-derived features for tissue classification.
1.6 Most popular approaches to supervised learning for tissue classification. The taxonomy is partially inspired by [102].
1.7 Graphical abstract of the PhD thesis. Horizontal arrows indicate the anatomical region analyzed to experimentally test the hypotheses (H1, H2, H3, H4) that guide this PhD work. H1: Machine learning (ML) supports the selection of informative frames to be processed by tissue classification algorithms. H2: Reliable tissue classification can be performed by exploiting ML and classification-confidence estimation. H3: Tissue classification with ML and confidence estimation can be further improved by using multispectral-imaging data. H4: ML-based tissue segmentation can be integrated in the flow of a simulated robot-assisted surgery, as to provide forbidden-region avoidance.
2.1 Visual examples of laryngeal video frames. (a) Informative frame (I); (b) Blurred frame (B); (c) Frame with saliva and specular reflections (S); (d) Underexposed frame (U).
2.2 Workflow of the proposed approach to automatic learning-based classification of informative frames with applications in laryngoscopy.
2.3 Confusion matrices with (left) and without (right) applying principal component analysis to the proposed feature set. Frame classification was obtained with support vector machines. Matrices refer to the classification of the balanced dataset of 720 narrow-band imaging laryngoscopic video frames. The colorbar indicates the number of images.

2.4 Macro-averaging receiver operating characteristic (ROC) curve analysis. ROC were obtained using support vector machines. No principal component analysis for feature reduction was performed. The mean (± standard deviation) curves obtained from the 3 cross-validation folds are reported by the orange solid lines (gray area). The mean (± standard deviation) area under the ROC curve is reported in the legend.

2.5 Comparison of the proposed feature set performance with the state of the art feature performance. Classification was performed using support vector machines. No principal component analysis for feature reduction was performed. Boxplots of class-specific recall (Rec\textsubscript{class}) are reported. Stars indicate significant differences (Wilcoxon signed-rank test (significance level \( \alpha = 0.05 \)) for paired samples). Dots indicate outliers.

2.6 Classification performance of the proposed method for a complete video sequence. (a) Macro-averaging receiver operating characteristic (ROC) curve analysis. The area under the ROC curve is reported in the legend. (b) Confusion matrix for a complete video sequence. The colorbar indicates the number of frames. The number of frame for each class is 341 (I, informative frame); 7 (B, blurred frame); 9 (S, frame with saliva or specular reflections); 84 (U, underexposed frame).

2.7 Visual confusion matrix for an entire laryngoscopic video sequence. Black boxes indicate the absence of misclassification between the true and predicted label. Numbers indicate the percentage of classified frames. I: informative frame; B: blurred frame; S: frame with saliva or specular reflections; U: underexposed frame.

2.8 Qualitative analysis of the classification outcomes. Each row shows examples of frames from a video sequence. I: informative frame; B: blurred frame; S: frame with saliva or specular reflections; U: underexposed frame.

3.1 The Gini coefficient (GC) is computed as twice the area (green area) between the line of equality and the Lorentz curve. The Lorentz curve represents the cumulative classification probability among the outcome classification states rank-ordered according to the decreasing values of their individual probabilities. A uniform discrete probability distribution has GC = 0, as the Lorentz curve overlays the line of equality, while for a state with probability 100% and the others at 0%, GC = 1.

3.2 Visual samples of narrow-band imaging laryngeal endoscopic frames of patients affected by squamous cell carcinoma.

3.3 Workflow of the proposed approach to laryngeal tissue classification in narrow-band imaging endoscopic video frames.
3.4 Four patches, relative to the four analyzed laryngeal tissue classes, are manually cropped from the image. Blue: tissue with intraepithelial papillary capillary loop-like vessels; Yellow: tissue with Leukoplakia; Green: healthy tissue; Red: tissue with hypertrophic vessels.

3.5 Comparison of different features without including the classification confidence estimation. Classification is obtained with support vector machines. (a) Boxplots of class-specific recall ($\text{Rec}_{\text{class}}$) for different features. $\text{Stat}_1$: Intensity mean, variance, entropy; $F_{\text{GLCM}}$: Gray-level co-occurrence matrix-based descriptors; $H_{\text{LBPR}_{\text{uni}}}^2$: Histogram of rotation-invariant uniform local binary patterns. The stars indicate significant differences (Wilcoxon test, $\alpha = 0.05$). (b) Normalized confusion matrix for $H_{\text{LBPR}}^2 + \text{Stat}_1$. The colorbar indicates the number of patches. The total number of patches (n. of patches) is reported.

3.6 Macro-average receiver operating characteristic (ROC) curves. The mean ($\pm$ standard deviation) curves obtained from the 3 cross-validation folds are reported in bold (transparent area). The mean ($\pm$ standard deviation) area under the ROC curve is reported in the legend. (a) ROC curves for the tested features. Classification is obtained using support vector machines. $\text{Stat}_1$: Intensity mean, variance, entropy; $F_{\text{GLCM}}$: Gray-level co-occurrence matrix-based descriptors; $H_{\text{LBPR}_{\text{uni}}}^2$: Histogram of rotation-invariant uniform local binary patterns. (b) ROC curves for the tested classifiers. Classification is obtained using the histogram of local binary pattern and first order statistics. kNN: k-nearest neighbors, NB: naive Bayes, RF: random forest, SVM: support vector machines.

3.7 Performance of the state of art. (a) Visual samples of the vessel segmentation obtained applying Barbalata et al. [10] algorithm to patches with hypertrophic vessels (first row), healthy tissue (second row) and intraepithelial papillary capillary loop-like vessels (third row). From left to right, original patch, vessel mask and vessel mask superimposed on the original patch. (b) Normalized confusion matrix obtained applying Barbalata et al. [10] algorithm to the tested laryngeal dataset. Colorbar indicates the number of patches. The total number of patches (n. of patches) is reported.

3.8 Short caption

3.9 Receiver operating characteristic (ROC) curves at different level $\tau$ of confidence on classification. Each curve refers to one of the laryngeal tissue classes. He: healthy tissue; Hbv: tissue with hypertrophic vessels; Le: leukoplakia; IPCL: tissue with intraepithelial papillary capillary loop-like vessels. The area under the ROC curve, for each curve, is reported in the legend. Classification is obtained using local binary pattern and first order statistics with support vector machines. (a) ROC curves for $\tau = 0.6$. (b) ROC curves for for $\tau = 0.9$.

3.10 Short caption
3.11 Sample RGB liver images acquired in the operating room. Images are captured with different lighting conditions and different tissue-camera pose. Images refer to a transplanted liver graft (left) and a discarded one (right). ............................................. 50

3.12 Proposed workflow for graft hepatic steatosis assessment. From 40 RGB liver images of 40 different donors, a dataset of patches with size $100 \times 100$ is extracted. From each patch, a set of textural features is computed. The dataset is divided in training and testing patches. The features from the training patches are used to train a supervised classifier model. The trained model is used to assess HS from the testing patches. Classification confidence is estimated and low-confidence patches are excluded. .................................................. 51

3.13 Workflow of the proposed approach to liver-donor’s steatosis assessment from RGB images acquired in the operating room. Image patches are extracted from the RGB images. From each patch, textural, intensity and blood-test features are extracted. The features extracted from training patches are used to train a semi-supervised support vector machine (SVM)-single instance learning (SIL) model (green boxes). The approach is semi-supervised as the ground-truth (histopathological-based) label is assigned to the whole image and not to the single patch. The proposed semi-supervised approach is compared with a naive patch-based supervised SVM classification followed by clustering (orange boxes). Clustering is required to obtain image classification from patch-classification results. ............................................. 53

3.14 Dataset sample images. The images refer to transplanted livers. Images were acquired at different distance and orientation with respect to the liver. Images present different illumination levels. Specular reflections are present due to the smooth and wet liver surface. ................................. 55

3.15 Liver and liver mask obtained through manual segmentation ................. 55

3.16 Dataset sample patches. The green and red boxes refer to patches extracted from transplanted and non-transplanted livers. Each row in a box refers to patches extracted from the same liver .................................. 56

3.17 Receiving operating characteristic (ROC) curve for patch classification with the normalized histogram of rotation-invariant uniform local binary patterns and support vector machines. ............................................. 57

3.18 Confusion matrix (CM) for classification of patches from transplanted (T) and non-transplanted (NT) liver. (a) CM for gray-level co-occurrence based-features ($F_{GLCM}$). (b) CM for intensity-based features ($Stat_1$). (c) CM for intensity-based and local binary patter features ($H_{LBP} + Stat_1$). (d) CM local binary patter-based features ($H_{LBP}$). .................. 60

3.19 Visual confusion matrices for SVM-based patch classification obtained with the histogram of rotation-invariant uniform local binary patterns and support vector machines. NT: Non-transplanted patches; T = Transplanted patches. ............................................. 61
3.20 Confusion matrices (CM) for image classification of transplanted (T) and non-transplanted (NT) liver grafts. CM are obtained with the best performing feature set ($H_{LBP} + Stat_1 + Blo$). The classification is performed with (left) support vector machines (SVM) + Clustering and (right) SVM-single instance learning (SVM-SIL). The colorbars indicate the number of correctly classified images.

3.21 Samples of classification outcomes for transplanted (first row) and non-transplanted (second row) liver grafts. Classification refers to support vector machines (SVM) - single instance learning (SIL) with the best performing feature set ($H_{LBP} + Stat_1 + Blo$). The green and red boxes refer to correct and wrong classification outcomes, respectively. SVM-SIL wrongly classified a rejected liver only once.

4.1 Workflow of proposed approaches for uncertainty-aware organ classification and automatic image tagging.

4.2 A feature vector is extracted from each $n \in N$ superpixel ($Spx_n$), where $N$ is the number of superpixels in the image. The feature vector for $Spx_n$ is obtained by concatenating the histogram ($H_{LBP}$) of uniform rotation-invariant local binary pattern ($LBP_{rin}^{riu}$) and the average spectrum ($AS$), for each $i \in N_C$ image channel, where $N_C$ is the number of channels in the image.

4.3 Challenges of the evaluation dataset. Four samples of images showing the gallbladder (first row) and spleen (second row) are reported. Images were recorded with varying endoscope pose and illumination level. Specular reflections are present in the images due to the smooth and wet organ surfaces. Multiple organs can be present in a single image. All images refer to the same multispectral channel.

4.4 Effect of confidence threshold ($\tau$) on the superpixel-based organ classification accuracy rate ($Acc_{Spx}$) for RGB and multispectral imaging (MI). Base refers to classification without confidence estimation. The stars indicate significant differences. Dots indicate outliers. The confidence is computed with: (a) the Gini coefficient ($GC$), (b) the posterior probability certainty index ($PPCI$).

4.5 Confusion matrix for confidence threshold $\tau = 0.9$ on the Gini coefficient and multispectral imaging. The values are in percentages and the colorbar indicates the number of superpixels.

4.6 Image tagging accuracy ($Acc_{Tag}$) for RGB and multispectral imaging (MI) for Base case and following introduction of confidence measure ($\tau = 0.9$ on the Gini coefficient). The stars indicate significant differences. Dots indicate outliers.

4.7 Effect of previously unseen target structures on the uncertainty estimation. Percentage ($\%LC_{Spx}$) of low-confidence $Spx$ ($\tau = 0.9$) for organs that were seen ($In$) and not seen ($Ex$) in the training process. Dots indicate outliers.
List of Figures

4.8 Image tagging examples for Base case (left) and following introduction of confidence measure with $\tau = 0.9$ on the Gini coefficient (right). The low-confidence superpixels (in gray) are excluded from the image tagging. The crosses indicate erroneously classified superpixels. 78

4.9 (a) Test image, (b) test image with superpixel segmentation, (b) corresponding classification for superpixels with acceptable confidence level and (c) confidence map obtained with confidence threshold $\tau = 0.9$ on the Gini coefficient. The symbols give examples of the probable causes of uncertainty. 79

5.1 (a) Micron handle and moving platform on which infrared LEDs (highlighted in red) are mounted for Micron tip tracking. (b) Micron tip tracking system (ASAP), stereo-cameras for microscopy image acquisition and 3D reconstruction and Micron are shown. 86

5.2 Fully convolutional neural network architecture exploited for vessel segmentation. Conv: Convolution; Pad: Padding; ReLU: Rectified linear unit; MaxPool: Max pooling; UpConv: Up-convolution; Copy: Copy layer. The number of feature maps are shown on top of the boxes. 87

5.3 The vessel surface can be discretized as a set of points laying on $I$ circumferences. The $i^{th} (i \in [0, I - 1])$ circumference ($C_i$) was characterized by its center 3D position ($C_i$), orientation ($\hat{N}_i$) and radius ($R_i$). 88

5.4 3D reconstruction algorithm workflow. First, the homography transformation ($H$) between the left and right image planes is computed. After homography computation, the assumption was made that the 3D shape of vessels can be approximated by a bent tubular structure with variable radius. Thus, the vessel surface in 3D can be discretized as a set of points lying on $I$ different juxtaposed circumferences ($C_i, i \in [0, I - 1]$). From the (left) segmented image, the 2D position of the center ($^L c_i$) and the 2D corresponding rightmost vessel edge point ($^L e_i$) of $C_i$ are computed $\forall i$. From $^L c_i$ and $^L e_i$, the homography is used to estimate the 2D position of the center ($^R c_i$) and edge point ($^R e_i$) of $C_i (i \in [0, I - 1])$ in the right image plane. Direct linear transformation is used to triangulate and obtain $C_i$ and $E_i$. From $C_i$ and $E_i$, the 3D vessel radius $R_i$ and direction $N_i$ are computed, $\forall i$. Once $C_i$, $N_i$ and $R_i$ are known, $C_i$ is automatically identified. Uniform sampling is used to sample 36 equally spaced points on $C_i$. The set of 36 points on $C_i, \forall i$, represents the 3D vessel surface point cloud. 88
5.5 The forbidden-region virtual-fixture control algorithm requires computation of the distance $d$ between the tool tip resting position ($p_{rest}$) and the 3D vessel point ($p_{closest}$) closest to $p_{rest}$. $p_{rest}$ is the position Micron takes when no control is triggered. To understand if the control command has to be triggered, first, (left image) the closest point $C_i$ on the 3D vessel centerline ($C_i, i \in [0, I-1]$, with $I$ equal to the number of centerline points) is found. Then, (central image) $p_{closest}$ is found among the points on the circumferences $[C_{i-2}, C_{i+2}]$. Finally, (right image) if $d$ is lower than a predefined threshold ($d_{safety}$), the Micron control is activated and the new Micron tip position ($p_{control}$) is defined, which lays on the line passing through $p_{closest}$ and the current $p_{rest}$ and is distant $d_{safety}$ from $p_{closest}$.

5.6 Phantom vascular samples used to test the proposed approach to vessel collision avoidance using Micron. Phantoms are reported in the first two rows. The last row shows images recorded during real petroclival meningioma surgery. Vessels are highlighted with respect to the background.

5.7 Evaluation of the 3D reconstruction algorithm. The 3D cloud was reprojected on the left image plane, obtaining a set of reprojected points ($L^{proj})$. $L^{proj}$ was made of several reconstructed circumferences ($L^C_j$ with $j \in [0 : I-1]$ and $I$ the number of reconstructed circumferences) with center $L^c_j$. For each $LC_j$, the 2 extremes were retained, obtaining a new set of points ($L^{extremes}_j$). The Euclidean distances between $L^{extremes}$ and the points of $LC_j$ closest to the vessel mask ($\gamma_j$) were computed. Two distances were considered ($L^d_{in}$ and $L^d_{out}$), depending on whether the extreme lied inside or outside the vessel mask, respectively.

5.8 Visual samples of vessel segmentation outcomes. Original RGB phantom vessel images are shown in the first row. The second and third row refer to the ground-truth manual segmentation and the segmentation outcome, respectively.

5.9 Point cloud reprojected on left and right image plane plotted over the left (left column) and right (right column) ground-truth vessel segmentation images.

5.10 Distance ($d$) between Micron tip and 3D vessel point closest to the Micron tip. $d$ with and without the virtual-fixture control is reported in green and orange, respectively.

5.11 Sample trajectory for virtual fixture evaluation. The reconstructed vessel point cloud is represented in blue; the trajectories with and without the virtual fixtures control are depicted in green and orange, respectively.
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Tested feature vectors and corresponding number of features</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>Evaluation dataset. For each video (video ID), and for each class (I, B, S, U), the number of frames that contributed to build the dataset are reported. The dataset is split in 3 folds to perform robust estimation of the classification performance. The folds are balanced both at patient- and class-level. I: informative frame; B: blurred frame; S: frame with saliva or specular reflections; U: underexposed frame.</td>
<td>20</td>
</tr>
<tr>
<td>2.3</td>
<td>Classification performance of the proposed approach. Results are relative to support vector machines (SVM) and random forest (RF) classification on the proposed feature set. SVM results with principal component analysis (PCA) performed on the feature set are reported, too. Class-specific recall ($\text{Rec}<em>{\text{class}}$), precision ($\text{Prec}</em>{\text{class}}$), and F1 score ($\text{F1}_{\text{class}}$) are reported for the four different frame classes. I: informative frame; B: blurred frame; S: frame with saliva or specular reflections; U: underexposed frame. Median and inter-quartile range (IQR) of the metrics are reported, too.</td>
<td>23</td>
</tr>
<tr>
<td>2.4</td>
<td>Classification performance of the state of the art features using support vector machines. Class-specific recall ($\text{Rec}<em>{\text{class}}$), precision ($\text{Prec}</em>{\text{class}}$), and F1 score ($\text{F1}_{\text{class}}$) are reported for the four different frame classes. I: informative frame; B: blurred frame; S: frame with saliva or specular reflections; U: underexposed frame. Median and inter-quartile range (IQR) of the metrics are reported, too. The parameters for the state of the art methods implemented for comparison were set as reported in their reference papers: oriented fast and rotated brief (ORB) parameters for [19], thresholding values for the intensity-based similarity score for [43], histogram bin number for [12] and orientation and radius for GLCM computation for [179].</td>
<td>26</td>
</tr>
</tbody>
</table>
List of Tables

3.1 Tested feature vectors and corresponding number of features. \( Stat_1 \):
Intensity mean, variance, entropy; \( F_{GLCM} \): Gray-level co-occurrence
matrix-based descriptors; \( H_{LBP} \): Normalized histogram of rotation-invariant
uniform local binary patterns. ........................................... 36

3.2 Evaluation dataset. For each of the 33 patients’ video, 10 images are
used for a total of 330 images. From each image, 4 tissue patches are
extracted for a total of 1320 patches relative to the 4 considered tissue
classes: healthy tissue, tissue with hypertrophic vessels, leukoplakia,
tissue with intraepithelial papillary capillary loop-like vessels. For a
robust evaluation, the dataset is split at patient level to perform 3-fold
cross-validation. In each fold, 11 patients are included, for a total of
110 images per fold. Each fold contains 440 patches equally distributed
among the laryngeal tissue classes. ........................................... 41

3.3 Median (first quartile - third quartile) class-specific recall (\( Rec_{class} \)),
precision (\( Prec_{class} \)), and F1 score (\( F1_{class} \)) obtained testing different
feature vectors for the \( Base \) case (i.e., without the inclusion of confi-
dence on classification estimation). Classification is obtained with sup-
port vector machines. \( Stat_1 \): Intensity mean, variance, entropy; \( F_{GLCM} \):
Gray-level co-occurrence matrix-based descriptors; \( H_{LBP} \): Normalized
histogram of rotation-invariant uniform local binary patterns. .......... 43

3.4 Comparison of different classifiers. Median (first quartile - third quar-
tile) class-specific recall (\( Rec_{class} \)), precision (\( Prec_{class} \)), and F1 score
(\( F1_{class} \)) are reported for the four different tissue classes. Classification
is obtained using the histogram of local binary patterns and first order
statistics. kNN: k-nearest neighbors, NB: naive Bayes, RF: random for-
est, SVM: support vector machines. ........................................... 45

3.5 Median (first quartile - third quartile) class-specific recall (\( Rec_{class} \)),
precision (\( Prec_{class} \)), and F1 score (\( F1_{class} \)) are reported at different
level of confidence (\( \tau \)) on support vector machines classification. .......... 46

3.6 Area under the receiving operating characteristic (ROC) curve obtained
with support vector machines (SVM) and different feature vectors for
patch classification. \( Stat_1 \): Intensity mean, variance and entropy; \( F_{GLCM} \):
Gray-level co-occurrence matrix-based descriptors; \( H_{LBP} \): Normalized
histogram of rotation-invariant uniform local binary patterns. ........... 56

3.7 Patch-classification performance measures obtained with support vector
machines (SVM) and different feature vectors. \( Stat_1 \): Intensity mean,
variance and entropy; \( F_{GLCM} \): Gray-level co-occurrence matrix-based
descriptors; \( H_{LBP} \): Normalized histogram of rotation-invariant uniform
local binary patterns. Metrics are reported for transplanted and non-
transplanted liver patches. Mean metrics are reported, too. \( Prec_{classes} \)
= Precision; \( Rec_{classes} \) = Recall, \( F1_{classes} \) = F1 score. .................. 59
List of Tables

3.8 Patch-classification performance measures obtained using the histogram of rotation-invariant local binary patterns as feature with different classifiers. \( \text{Rec}_{\text{class}} = \) Precision; \( \text{Prec}_{\text{class}} = \) Recall; \( F_1_{\text{class}} = \) F1 score; SVM = Support vector machines, RF = Random forest. Metrics are reported for transplanted and non-transplanted liver patches. Mean metrics are reported, too. ............................................. 59

3.9 Image-classification performance measures for support vector machines and clustering (SVM + Clustering), and SVM-single instance learning (SVM-SIL). Classification accuracy (\( \text{Acc} \)), recall (\( \text{Rec}_{\text{classes}} \)), precision (\( \text{Prec}_{\text{classes}} \)) and F1-score (\( F_1_{\text{classes}} \)) are reported. Metrics are reported for transplanted and non-transplanted liver images. ............................................. 60

4.1 Camera light-filter central wavelengths and full width at half maximum (FWHM) for each \( i (= 1–8) \) band in multispectral imaging (MI) and RGB. 72

4.2 Median superpixel-based accuracy rate (\( \text{Acc}_{\text{Spx}} \)) and inter-quartile range (in brackets) for RGB and multispectral imaging (MI) using different features for the Base case (i.e., without confidence inclusion). \( H_{LBP} \): Histogram of local binary patterns; \( AS \): Average spectrum. 74

5.1 Nomenclature relative to the 3D reconstruction. ............................................. 87

5.2 Training parameters for vessel segmentation. ............................................. 90

5.3 Vessel segmentation performance and computational cost for the 100 testing images. Median, first and third quartiles are reported for each metric. \( DSC \): Dice similarity coefficient; \( Se \): Sensitivity; \( Sp \): Specificity; \( Acc \): Accuracy; Time: Computational time [s]. 96

5.4 Number of points (\( N_{\text{points}} \)) and computational cost (Time [s]) for each of the 10 reconstructed 3D vessel clouds considered for evaluating the 3D reconstruction algorithm. Median values are reported, too. The values are reported for the proposed algorithm and for the 3D reconstruction performed with speeded up robust features (SURF) [13] and the fast library for approximate nearest neighbors (FLANN) [163]. Significant differences were found when comparing \( N_{\text{points}} \) computed with the proposed approach and with SURF and FLANN (Wilcoxon signed-rank test (significance level = 0.05)). ............................................. 96

6.1 Datasets collected for testing the methodologies presented in this PhD thesis. ............................................. 106
Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


