Automatic Tissue Characterization from Optical Coherence Tomography Images for Smart Laser Osteotomy

Inaugural dissertation

to

be awarded the degree of Dr. sc. med.

presented at
the Faculty of Medicine
of the University of Basel

by

Yakub Aqib Bayhaqi
from Jakarta, Indonesia

Basel, 2023

Approved by the Faculty of Medicine
On application of

Prof. Dr. Azhar Zam, University of Basel, First examiner
Prof. Dr. Philippe C. Cattin, University of Basel, Second examiner
Prof. Dr. Alexander A. Navarini, University of Basel, Further advisor
Dr. Ferda Canbaz, University of Basel, Further advisor
Prof. Dr. Yifan Jian, Oregon Health and Science University, External expert

Basel, 20$^{\text{th}}$ March 2023

Dean Faculty of Medicine

Prof. Dr. Primo Leo Schär

To my parents, Haulian Siregar and Egi Legiati,
my loving wife, Purwinda Marliasih,
and all of my teachers and friends ...

Acknowledgments

CONTENTS

Acronyms and abbreviations

AC   Attenuation Coefficient

ANN Artificial Neural Network

CNN Convolutional Neural Network

CO$_2$ Carbon Dioxide

CPU Central Processing Unit

CT   Computed Tomography

DC   Direct Current

DenseNet Densely Connected Convolutional Networks

DNN Deep Neural Network

Er:YAG Erbium-doped Yttrium Aluminium Garnet

Er-Cr:YSGG Erbium-Chromium-dopped Yttrium-Scandium-Gallium Garnet

FD-OCT Fourier Domain Optical Coherence Tomography

FWHM Full Width Half Maximum

GPU Graphical Processing Unit

LCI Low Coherence Interferometry

LIBS Laser-Induced Breakdown Spectroscopy

NA   Numerical Aperture

OCM Optical Confocal Microscopy

OCT Optical Coherence Tomography

ReLu Rectifier Linear Unit

ResNet Residual Neural Network

RL   Random Lasing

SD-OCT Spectral Domain optical Coherence Tomography

SGD Stochastic Gradient Descent

SNR Signal to Noise Ratio

SS-OCT Swept Source Optical Coherence Tomography

SVM Support Vector Machine

TD-OCT Time Domain Optical Coherence Tomography

Ti:Sapphire Titanium-doped Sapphire

MIRACLE Minimally Invasive Robot-Assisted Computer-guided LaserosteotomE

Nd:YAG Neodymium Yttrium Aluminium Garnet

Nd:YLF Neodymium-doped Yttrium Lithium Fluoride

VCSEL Vertical-Cavity Surface-Emitting Laser

VGG Visual Geometry Group

Summary

Fascinating experiments have proved that in the very near future, laser will completely replace mechanical tools in bone surgery or osteotomy. Laser osteotomy overcomes mechanical tools' shortcomings, with less damage to surrounding tissue, lower risk of viral and bacterial infections, and faster wound healing. Furthermore, the current development of artificial intelligence has pushed the direction of research toward smart laser osteotomy. This thesis project aimed to advance smart laser osteotomy by introducing an image-based automatic tissue characterization or feedback system. The Optical Coherence Tomography (OCT) imaging system was selected because it could provide a high-resolution subsurface image slice over the laser ablation site.

Five experiments were conducted and published to show the feasibility of the feedback system. In the first part of this thesis project, a deep-learning-based OCT image denoising method was demonstrated and yielded a faster processing time than classical denoising methods, while maintaining image quality comparable to a frame-averaged image. Next part, it was necessary to find the best deep-learning model for tissue type identification in the absence of laser ablation. The results showed that the DenseNet model is sufficient for detecting tissue types based on the OCT image patch. The model could differentiate five different tissue types (bone, bone marrow, fat, muscle, and skin tissues) with an accuracy of 94.85%. The last part of this thesis project presents the result of applying the deep-learning-based OCT-guided laser osteotomy in real-time. The first trial experiment took place at the time of the writing of this thesis. The feedback system was evaluated based on its ability to stop bone cutting when bone marrow was detected. The results show that the deep-learning-based setup successfully stopped the ablation laser when bone marrow was detected. The average maximum depth of bone marrow perforation was only $216\,\mu$m.

This thesis project provides the basic framework for OCT-based smart laser osteotomy. It also shows that deep learning is a robust approach to achieving real-time application of OCT-guided laser osteotomy. Nevertheless, future research directions, such as a combination of depth control and tissue classification setup, and optimization of the ablation strategy, would make the use of OCT in laser osteotomy even more feasible. Finally, as a sub-project of the multidisciplinary MIRACLE project at the University of Basel, this thesis project supports the drive towards replacing mechanical tools with safer and less invasive laser osteotomy.

ZUSAMMENFASSUNG

In faszinierenden Experimenten wurde nachgewiesen, dass in der Knochenchirurgie oder Osteotomie mechanische Werkzeuge in naher Zukunft vollständig durch Lasertechnologien ersetzt werden können. Die Laserosteotomie übertrifft die mechanischen Werkzeuge durch eine geringere Schädigung des umliegenden Gewebes, verringert das Risiko viraler und bakterieller Infektionen und führt zu einer schnelleren Wundheilung. Darüber hinaus haben die aktuellen Entwicklungen im Rahmen der künstlichen Intelligenz die Forschung im Bereich der Laserosteotomie in Richtung der intelligenten Laserosteotomie gelenkt. In dieser Dissertation wird sich der intelligenten Laserosteotomie durch die Einführung einer bildbasierten automatischen Gewebecharakterisierung genähert. Für das Überwachungs- bzw. Feedback-System wurde die optische Kohärenztomographie (Optical Coherence Tomography, OCT) gewählt, da sie einen hochauflösenden Bildausschnitt der Laserabtragungsstelle liefern kann.

Es wurden fünf Experimente durchgeführt und veröffentlicht, um die Realisierbarkeit des Feedback-Systems zu zeigen. Im ersten Teil dieser Arbeit wird eine auf Deep Learning basierende OCT-Bildentrauschung vorgestellt. Diese bietet nachweislich eine schnellere Verarbeitungszeit als die klassischen Entrauschungsmethoden und dank der Bildmittelung eine vergleichbare Bildqualität. Anschließend wird das beste Deep-Learning-Modell zur Erkennung von Gewebetypen ohne Laserablation ermittelt. Das Ergebnis zeigt, dass das DenseNet-Modell ausreicht, um die Gewebetypen anhand des OCT-Bild-Patches zu erkennen. Das Modell konnte fünf verschiedene Gewebetypen (Knochen-, Knochenmark-, Fett-, Muskel- und Hautgewebe) mit einer Genauigkeit von 94,85 % unterscheiden. Der letzte Teil dieser Arbeit zeigt das Endergebnis der Echtzeitanwendung der OCT-gesteuerten Laserosteotomie auf Basis von Deep Learning. Das erste Experiment wurde zum Zeitpunkt der Erstellung dieser Arbeit durchgeführt. Das Feedback-System wurde derart justiert, dass mit dem Schneiden in den Knochen gestoppt wird, sobald Knochenmark erkannt wurde. Das Ergebnis zeigt, dass Deep Learning den Ablationslaser erfolgreich stoppte, wenn Knochenmark erkannt wurde. Die durchschnittliche maximale Tiefe der Perforation des Knochenmarks betrug nur $216\,\mu$m.

Das Hauptergebnis dieser Dissertation liefert den grundlegenden Rahmen für OCT-basierte intelligente Laserosteotomie. Es zeigt auch, dass Deep Learning ein geeigneter Ansatz ist und in der Echtzeitanwendung der OCT-gesteuerten Laserosteotomie eingesetzt werden kann. Dennoch ist davon auszugehen, dass künftige Forschungsprojekte, zum

Beispiel eine Tiefenkontrolle für den Gewebeklassifikator oder die Optimierung der Ablationsstrategie, die Realisierbarkeit des OCT-Einsatzes bei der Laserosteotomie verbessern werden. Als Teilprojekt des multidisziplinären MIRACLE-Projekts an der Universität Basel unterstützen die Ergebnisse dieser Arbeit schließlich die Realisierbarkeit des Ersatzes mechanischer Werkzeuge durch eine sicherere und weniger invasive Laserosteotomie.

CHAPTER 1

Introduction

1.1 Motivation

Bone surgery or osteotomy has been practiced to treat bone and joint abnormalities for centuries. This standard procedure has been used for clinical treatments such as knee and hip-joint replacement [1, 2], spine correction [3, 4], and maxillofacial surgery [5, 6]. Despite the procedure's effectiveness, however, the conventional osteotomy procedures usually involve mechanical tools (saw, drill, chisel, and hammer). In the process, mechanical contact and movement apply excessive pressure that, along with friction and vibration, removes more tissue than necessary [7–9]. Additionally, preventing bacterial contamination is a serious concern, given the direct contact between tissues and metals [10–12].

Over the last decades, the tool technology has developed rapidly to solve these problems. One cutting-edge technology is the use of laser for osteotomy or laser osteotomy for short. Laser osteotomy offers an efficient way to address the problems of mechanical tool, with contactless intervention and a higher precision level. The focused cut of laser enables surgeons to go beyond straight cuts and perform more complex cuts like circular, diamond, and dovetail shapes [13]. Compared to conventional saws or drills, laser osteotomy cuts with low mechanical vibration, thus leading to less material loss and less potential microorganism contamination, both of which improve the bone's regenerative ability [14, 15].

While laser osteotomy holds many advantages over mechanical tools, there is an inherent risk of collateral damage to surrounding critical tissue (such as nerves). Sufficient information about the depth of the cut or the tissue type being ablated is needed so that the surgeon can cut the bone precisely while avoiding damage to critical tissue at the bottom of the cut. One solution is to implement a feedback system that detects both the depth of the cut and the tissue type so that surgeons could perform laser osteotomy precisely, safely, and without damaging critical tissues.

The tissue type being ablated can be detected by analyzing the secondary light and/or the acoustic emission generated during ablation or by using external diagnostic (imaging) modalities. Analyzing the secondary light emissions via Random Lasing (RL) [16, 17]

and Laser-Induced Breakdown Spectroscopy (LIBS) [18–21] can differentiate tissue types with a high degree of accuracy. Furthermore, analyzing the acoustic signals emitted by photoablation (optoacoustic) can detect tissue types with an accuracy comparable to that of RL and LIBS [22–25]. However, each of these methods has its own limitations. For example, RL depends on a laser dye that might be harmful to biological tissue, while LIBS is more accurate for detecting tissue surfaces. Additionally, LIBS and optoacoustic techniques require a high-energy pulsed laser to produce a measurable signal and usually results in unavoidable damage to critical tissues. Another method for tissue detection involves integrating the ablation laser with external optical spectroscopy. The non-contact nature of laser allows for integration with external optical sensors such as Raman [26, 27], autofluorescence [28, 29], and diffuse reflectance spectroscopies [30–33]. Like LIBS, Raman spectroscopy accurately differentiates between the tissue's molecular bonds. However, it requires a few seconds of integration time to get an observable signal and is therefore unsuitable for use in a real-time feedback system. Moreover, autofluorescence and diffuse reflectance spectroscopy, could be used to detect tissue types with higher detection rates than Raman spectroscopy. However, these techniques are limited by the fact that they give feedback about tissue type only for a single point and lack cutting-depth information.

Alternatively, the laser ablation process can be monitored with an integrated, ablation-free imaging system. The basic idea underlying this thesis project was to use an Optical Coherence Tomography (OCT) imaging system to monitor the tissue anatomy at the subsurface level and control the laser automatically. This imaging technology is analogous to ultrasound imaging, which performs cross-sectional tomography using light interference phenomena. The echo time delay of the back-reflected or backscattered light from the tissue's internal microstructures is measured using interferometry of partially coherent light. Although the light's penetration depth in OCT is limited in comparison with ultrasound, the contactless nature of OCT makes it easily integrated with laser osteotomy. Unlike ultrasound, OCT does not require impedance matching or contact with the target tissue.

OCT has become a standard diagnostic tool in ophthalmology for diagnosing retinal diseases [34]. It has also become an alternative to ultrasound for high-resolution intra-venous imaging [35]. When integrating OCT into a laser osteotomy system, the image provides information about both the depth of ablation (for precise depth measurement) and the tissue texture (for tissue type detection). Moreover, recent advances in computation technology have led to the adaptation of artificial intelligence (deep learning) for medical image analysis, opening a new era for smart feedback systems. Specifically, the Convolutional Neural Network (CNN) marks a significant breakthrough for automatic image analysis and recognition. In the field of OCT, CNN has been shown to be an accurate method for retinal image segmentation [36–39] and plaque detection in intravenous OCT images [40, 41]. The deep learning model classifies images faster by skipping the classical feature extraction process, thereby enabling its use in real-time tissue detection for laser osteotomy.

More importantly, this thesis was initiated under a flagship project called MIRACLE, short for "Minimally Invasive Robot-Assisted Computer-guided LaserosteotomE", at the

University of Basel. The result of this thesis demonstrate the feasibility of a computer-guided laser osteotomy which is emphasized as "smart" laser osteotomy by using deep learning method.

1.2 Contribution and Research Scope

Research into the clinical applications of OCT as a feedback system for monitoring the depth of laser ablation has been reported, such as tumor resection [42, 43], cochlear implant surgery [44, 45], and bone osteotomy [46–48]. Recently, it has been shown that OCT helps to shape the endosteal layer's curvature for cochlear implantation with a mean absolute accuracy of around $20\,\mu$m [44]. This achievement is a strong evidence of OCT's suitability for precise laser osteotomy. However, this finding has been applied as a real-time depth monitoring system. Its application for real-time tissue detection, specifically for laser osteotomy, remains a challenge worth further research. Changes in the tissue's optical properties due to high-power laser effects are also another crucial challenge to achieving optimal differentiation during the laser-cutting process.

This thesis project focused on using OCT for tissue detection, to guide laser osteotomy by fulfilling the basic requirement of real-time feedback. Tissue detection can be used to stop the laser ablation process when it reaches critical tissue. The investigation also included the application of artificial intelligence (deep learning) to classify the tissue and emphasized the smart laser osteotomy approach. This thesis project contributes to finding the best deep learning models for image quality improvement and tissue detection, and to demonstrating the real-time feedback system.

It would also be interesting to explore the benefit of deriving both depth control and tissue detection from one OCT image. However, this research aims to give the basic building blocks for tissue-specific laser osteotomy. It provides a "proof of principle" and reveals the limitations of a real-time and automatic tissue-detection concept. In the future, combining both depth control and tissue detection based on the OCT image would lead to greater precision and safety of laser osteotomy.

1.3 Outline

The main concepts and strategies underlying this thesis project are outlined in seven chapters, starting with this introduction chapter (Chapter 1). Chapter 2 discusses the basic principle of OCT, as well as methods for OCT signal processing, image denoising, and some classical approaches to tissue detection. The next chapter, Chapter 3, briefly discusses approaches for automatic tissue characterization using machine learning and deep learning. The same chapter also briefly describes the integration of the OCT acquisition system with the deep learning inference for real-time feedback.

The outcomes of this thesis project have been published in several conference proceedings and journals. Two publications, presented in Chapter 4, describes the deep learning-based approach to fast image denoising to improve tissue detection accuracy. In

Chapter 5, two publications describes the best deep learning models to differentiate tissue types. The last publication, in Chapter 6, evaluates the performance and accuracy of the closed-loop real-time feedback system. Finally, the last chapter of the thesis, Chapter 7, discusses the outlook for future research, specifically in combining depth control and tissue detection to have both precise and safe laser osteotomy, and ends with some closing remarks.

Optical Coherence Tomography

Optical Coherence Tomography (OCT) has become the preferred modality for non-invasive and non-contact biomedical imaging. It is analogous to ultrasound imaging, except that it uses light instead of sound. It performs a subsurface two-dimensional cross section and three-dimensional volumetric image scan of biological tissue. The latest light source technology enables OCT to provide real-time imaging with an axial resolution of less than $3\,\mu$m in air [49,50]. The use of infrared and visible light is safer for most biological samples, compared to the ionizing radiation used in micro-computed tomography (micro-CT) for microscopic-scale tissue imaging.

This chapter briefly reviews the basic principles of OCT and the signal processing steps. The discussion starts with the basic concept of signal formation in the Time Domain Optical Coherence Tomography (TD-OCT). It continues with the latest OCT technology, Fourier Domain Optical Coherence Tomography (FD-OCT), which was used in this thesis project. The benefit of using FD-OCT over TD-OCT is also discussed. This chapter also discusses the necessary signal processing steps and the requirement for a high-performance computational resource to reconstruct the OCT image in real-time. The origin of speckles in OCT and several methods to enhance image quality are also explained in the last section of the chapter. Reducing image speckles is important as it could improve the tissue classifier.

2.1   Basic Principle of OCT

OCT imaging is a tomographic technique that uses interference phenomena to measure the magnitude and echo time delay of light traveling through the internal micro-structure of biological tissue. The basic concept of OCT is based on the classical optical measurement technique, known as Low Coherence Interferometry (LCI), which uses the Michelson interferometer configuration (Figure 2.1a) [51]. The interference signals are formed by splitting light into two paths (called reference and sample arms) using a beam-splitter and reflecting them back using a mirror at each end to recombine them to interfere with each other constructively or destructively. The interference signal, referred as interferogram,
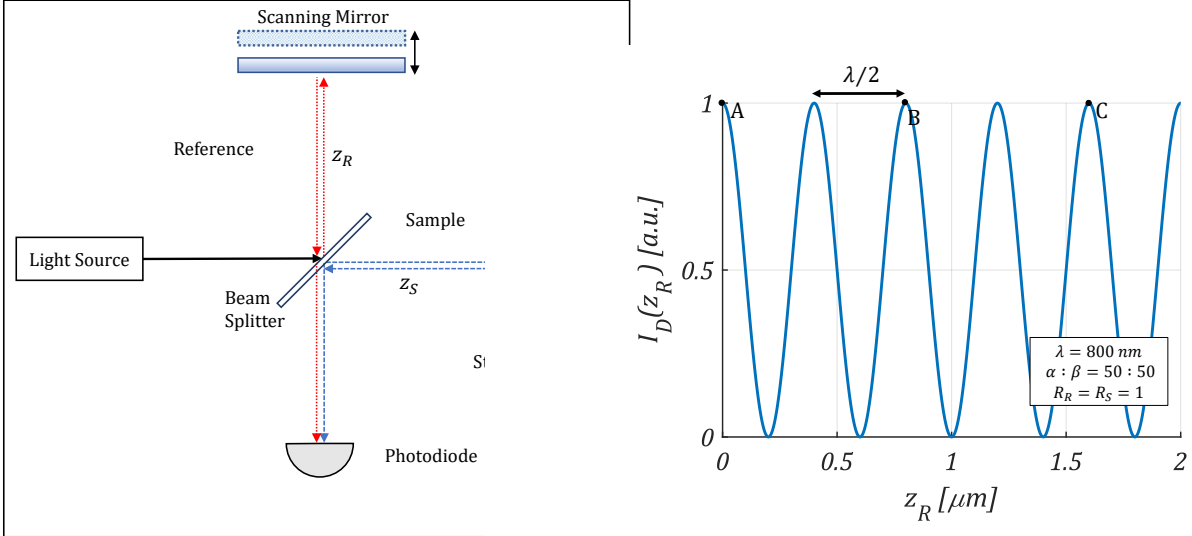
(a) Michelson interferometer configuration.

(b) Interferogram with monochromatic light.

Figure 2.1: A Michelson interferometer configuration (a). The light is split into two paths, the reference arm path (red dashed line) and the sample arm path (blue dashed line), by a beam splitter. Mirrors are placed at each arm to reflect the light back. The same beam splitter recombines the light reflected from the two paths. A photodiode detects the intensity of the recombined light, while the mirror at the reference arm is moving. An example of the detected interference signal from a monochromatic light source is shown in (b). However, this interference pattern lacks location information. A reflector in the sample arm will appear multiple times (e.g., at points A, B, or C), assuming that each peak corresponds to a path length match between the sample and reference arm. Images adapted from [35].

is detected by a photodetector while scanning (moving) the mirror at the reference arm. This technique is called Time Domain OCT (TD-OCT). The time terminology comes from the definition of the reference mirror location as a function of time.

The interferogram from a monochromatic light source with a wavenumber of k and intensity of I_0 follows the equation [35]

$$I_D\left(z_R\right) = \alpha\beta I_0 \left[R_R + R_S + 2\sqrt{R_R R_S}\cos(2k(z_R - z_S))\right]. \tag{2.1}$$

$I_D(z_R)$ defines the detected intensity as a function of the location of the reference mirror. Fraction of $\alpha : \beta$ indicates the splitting ratio between incident light power sent to the reference arm and the sample arm, respectively. R_R and R_S indicate the power reflectivity of the mirror at the reference and sample arms, respectively, and are given by the magnitude squared of the electric field reflectivity ($R = |r|^2$). The intensity of the detected signal will either increase or decrease depending on the path length difference between the two arms ($z_R - z_S$). However, this sinusoidal signal lacks information about the mirror location in the sample arm. The mirror reflection appears multiple times for every path length difference, with a multiplication of $\pm\pi/k$, for example, at points A, B, or C in Figure 2.1b.

Most light sources are never truly monochromatic. Even laser can only be quasi-

monochromatic. Here, the coherence property of the light source plays an important role in determining the axial location of the mirror in the sample arm or a reflection site in the biological sample. The frequency (wavenumber) variation shortens the coherency of the light source, thus, limiting the interference signal detection. Figure 2.2 shows an example where the light coherency limits the interference signal by enveloping the interference fringe into a package or sync signal. The packaged signal provides a better estimation of location from a reflective boundary, as full interference only arises when the two path lengths are matched within the coherence length of the light source. Other reflections originating outside the coherence length will go undetected (coherence gating).
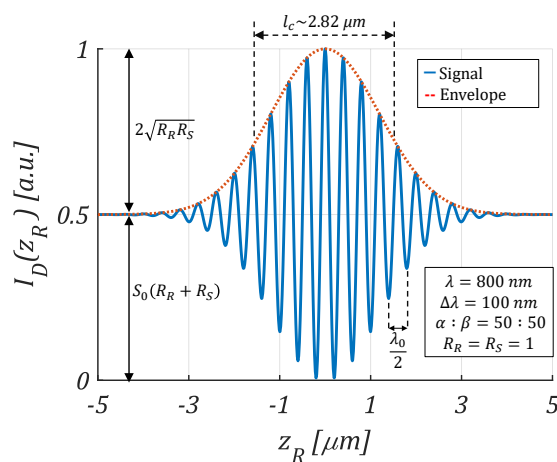
Figure 2.2: Example of detected interference signal with a Gaussian polychromatic light source ($\lambda_0 = 800\,nm$ and $\Delta\lambda = 100\,nm$). A reflector in the sample arm will be detected only in the coherence length (l_c) and most likely at the peak of the envelope. Images adapted from [35].

The interferogram, when using a light source with a Gaussian spectrum, can be rewritten from Equation 2.1 and becomes [51]

$$I_D\left(z_R\right) = \alpha\beta\,S_0\left[R_R + R_S + 2\sqrt{R_R R_S}\cos(2k_0(z_R - z_S))\,e^{-((z_R - z_S)\Delta k)^2}\right]. \qquad (2.2)$$

Here, k_0 represents the central wavenumber of the light source spectrum, and Δk represents its spectral bandwidth. The total power spectral emitted by the light source is defined as $S_0 = \int_0^\infty S(k)$. The exponential component of the equation defines the envelopes of the signal. It is associated with the complex degree of coherence of the light source, which is dependent on the spectral bandwidth of the light source.

In clinical practice, the mirror in the sample arm would be replaced with a biological sample. The sample under investigation contains profiles of reflectivity along the depth (see Figure 2.3). The internal sample reflectivity profile is reconstructed by scanning the reference mirror, along with the path length difference. This scan is analogous to A-Scan
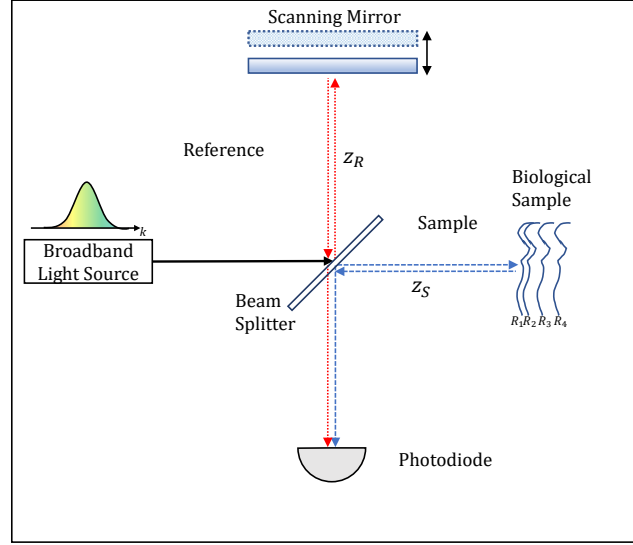
Figure 2.3: TD-OCT uses a broadband light source to have a low coherence length. Practically, it uses the Michelson interferometer setup and replaces the mirror in the sample arm with a biological sample. The internal sample reflectivity profile is reconstructed by scanning the reference mirror. The signal formation is illustrated in Figure 2.4. Images adapted from [51].

in ultrasound imaging. The scan motion moves the envelope, whose amplitude corresponds to the reflectivity at each location. The interferogram detected by the photodiode with multiple reflectors is then defined as [51]

$$
\begin{aligned}
I_D\left(z_R\right) = \alpha\beta\ S_0\Big[&R_R + \sum_{n=1}^{N} R_{S_n} && \text{(DC)}\\
&+ 2\sum_{n=1}^{N}\sqrt{R_R R_{S_n}}\cos(2k_0(z_R - z_{S_n}))\ e^{-((z_R - z_{S_n})\Delta k)^2} && \text{(Cross-correlation)}\\
&+ \sum_{n\neq m=1}^{N}\sqrt{R_{S_n} R_{S_m}}\cos(2k_0(z_{S_n} - z_{S_m}))\ e^{-((z_{S_n} - z_{S_m})\Delta k)^2}\Big]. && \text{(Auto-correlation)}
\end{aligned}
$$
$$(2.3)$$

The multiple reflectors in this equation distinguish three signal components [51]: (1) *Direct Current* (DC)[1] signal is the most significant component, proportional to the reflectivity of the reference and sample arms. However, it is typically proportional to the mirror power reflectivity at the reference arm, as its magnitude is usually larger than the reflectivity of the biological tissue at the sample arm ($R_R \gg R_S$). This could be considered as the base signal, with the absence of interference. (2) *Cross-correlation* signal is the primary information carrier. This component carries the interference relation between individual reflectors at the sample arm and the mirror at the reference arm. (3) *Auto-correlation* signal is the signal resulting from self-interference between reflectors in the biological

---

[1]The direct current terminology comes from the fact that the detector measures the signal as a direct electrical current.
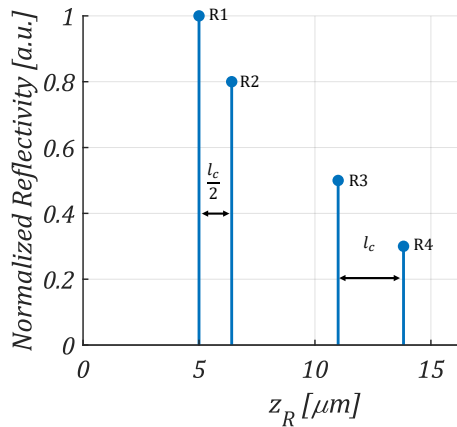
(a) Discrete reflectors in a sample.
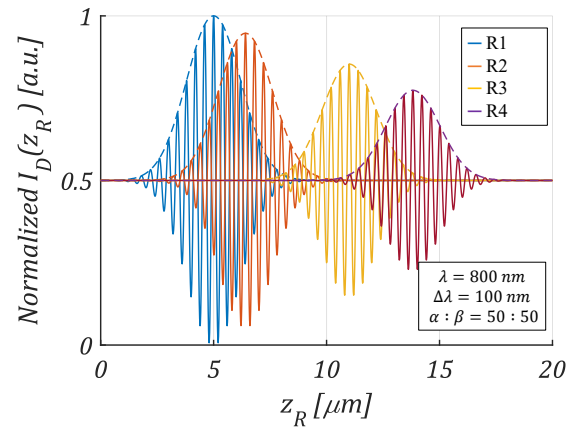
(b) Individual interference signals.
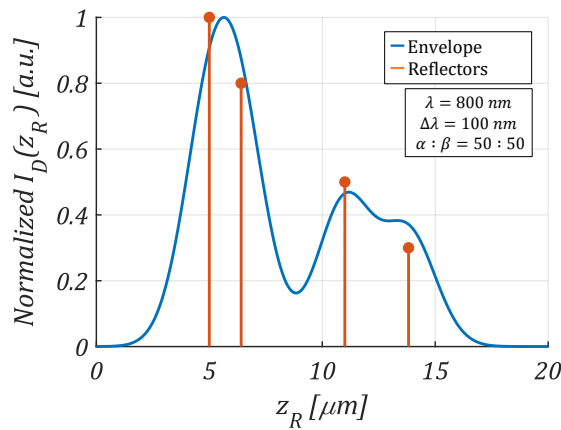
(c) The envelope of the combined interference
signal.

*Figure 2.4: Illustration of the sample reflectivity profile reconstruction in TD-OCT. Figure (c) is
the envelopes of the combined interference signals (b) coming from the individual reflectors (a).
Two reflective boundaries can be fully distinguished if located at a distance $\geq l_c$ (e.g., R3 and
R4, with distance l_c). Still, it is difficult to separate two reflectors with a distance $< l_c$ (e.g., R1
and R2, with distance $l_c/2$). Images adapted from [51].*

tissue under investigation. This signal is typically neglected, using the same argument for
the DC signal that $R_R \gg R_S$.

The axial resolution of a TD-OCT is characterized by Full Width Half Maximum
(FWHM) of the signal envelope. It is defined as the round-trip coherence length of a
Gaussian-shaped light source (l_c) given by [51]

$$l_c = \Delta z = \frac{2\sqrt{ln(2)}}{\Delta k} = \frac{2\,ln(2)}{\pi}\frac{\lambda_0^2}{\Delta\lambda}, \tag{2.4}$$

where λ_0 and $\Delta\lambda$ are the center and bandwidth wavelength of the light source. The signal

formed is illustrated in Figure 2.4. Two reflective boundaries can be fully distinguished if they are located at a distance more than or equal to the coherent length ($\geq l_c$); for example, the reflectors at points R3 and R4. In contrast, two reflective boundaries with a distance less than coherent length ($< l_c$) may be hard to separate, as is the case with the reflectors at points R1 and R2.

OCT extends the benefit of LCI by performing multiple A-Scans in the lateral direction to capture a two-dimensional image, called a B-Scan image, analogous to ultrasound B-Mode. Lateral resolution in OCT is determined by the optical system that focuses the light and is akin to Optical Confocal Microscopy (OCM). The lateral resolution is defined by [51]

$$\Delta x = \frac{4\lambda}{\pi} \frac{f}{d} \, , \tag{2.5}$$

where d and f are the size of the incident beam on the objective lens and the focal length of the objective lens, respectively. The lateral resolution is proportional to wavelength (λ) and inversely proportional to the Numerical Aperture (NA), defined as f/d. The lateral resolution also determines the depth of field (b), whereby increasing the lateral resolution (smaller Δx) decreases the depth of field. The relation is illustrated in Figure 2.5 and defined by [51]

$$b = \frac{\pi \Delta x^2}{\lambda} \, . \tag{2.6}$$

The selection of the depth of field or NA imposes the benefit of OCT compared to OCM. The OCT axial resolution is technically independent of the focusing lens's depth of field
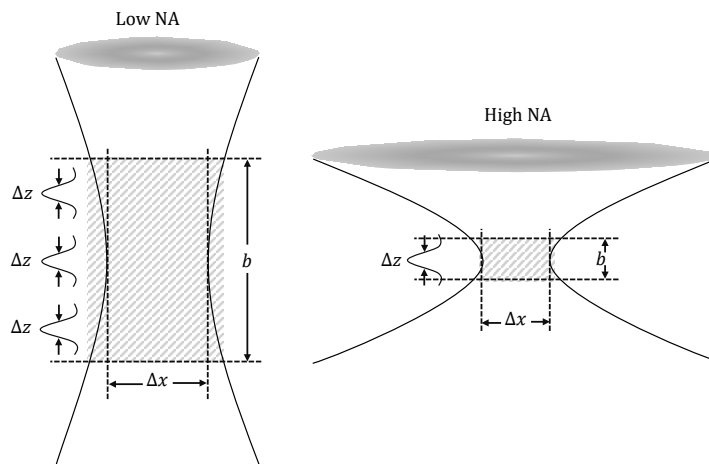
Figure 2.5: Lateral resolution in OCT is determined by the focused spot size of the OCT beam. OCT imaging is usually performed with a low NA, with a depth of field (b) much longer than the coherence length, to scan the reflectivity function over the sample cross-section. Although focusing with a high NA results in high lateral resolution, it reduces the depth of field and limits the cross-sectional range of the OCT. Images adapted from [51].

or NA. Rather, the axial resolution is given by the envelope's width or governed by the light source's coherence length (l_c). Conversely, the OCM axial resolution is limited by the confocal gating, which is defined by the depth of focus (b). Thus, a high NA is preferable in OCM. Most OCT imaging systems, however, work in a low NA condition to attain a considerable depth of field. The depth of field needs to be larger than the coherence length, $b > \Delta z$, to benefit from the coherence length resolution, although the low NA consequently limits the lateral resolution. This condition is useful in a limited NA application, such as catheter imaging

OCT has the added benefit of sending light via an optical fiber, which gives more flexibility when placing the sample arm, usually embedded into a handheld probe. The use of optical fiber allows easier integration with a variety of medical instruments, such as a handheld probe for skin imaging or a catheter for intravenous imaging. TD-OCT is limited by the movement speed and range of the scanning mirror in the reference arm. The next section discusses Fourier domain OCT, which is faster than TD-OCT.

2.2 Fourier Domain OCT

Fourier domain OCT (FD-OCT) uses a faster signal collection technique and offers a solution to the scanning mirror limitation of TD-OCT. In FD-OCT, depth information is obtained by evaluating the spectrum of the returning light without moving the mirror in the reference arm. Instead, it detects sample reflectivity at different locations by distinguishing the frequency components (Fourier domain) in the output signal. A spectrometry technique is used, which can be performed in two ways: either changing the detector to a spectrometer, called Spectral Domain optical Coherence Tomography (SD-OCT), or using a rapidly tunable laser source, called Swept Source Optical Coherence Tomography (SS-OCT). Schematics of the SD-OCT and SS-OCT are illustrated in Figure 2.6. The interferogram formation principle is similar for both techniques. In this thesis project, we specifically used SS-OCT to benefit from certain advantages it has over SD-OCT; these advantages will be discussed in the next subsection. However, to understand the FD-OCT working principle, one must first understand how the signal formation is reconstructed from the spectral domain to the spatial domain in SD-OCT.

The interferogram in SD-OCT can be determined by referring to Equation 2.1, where the sinusoidal signal is formed by either the variation in path length difference, $z_R - z_S$, or wavenumber k. In TD-OCT, the single photodiode implicitly measures the integrated spectral power emitted by the light source. Thus, the sinusoidal signal only depends on the moving reference mirror $I(z_R)$. In SD-OCT, however, both the biological sample and reference mirror remain static. A spectrometer is used instead of a photodiode. Each pixel element in the spectrometer measures the back-reflected intensity for each wavenumber, $I(k)$, and can be rewritten based on Equation 2.3 by considering the individual wavenumber,

(a) Spectral do

(b) Swept source OCT.
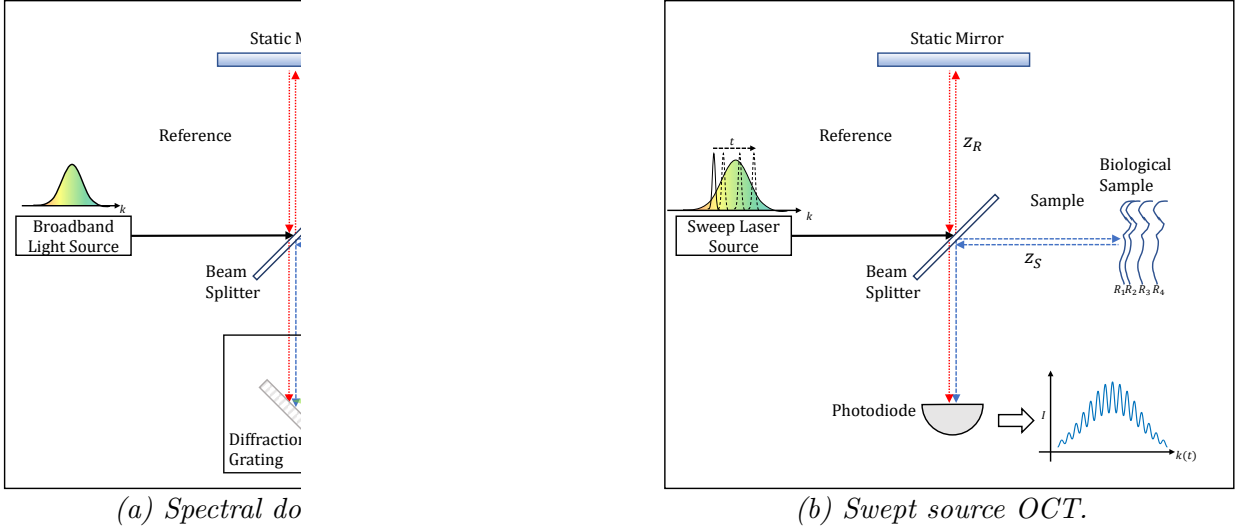
Figure 2.6: Schematic for SD-OCT (a) and SS-OCT (b). As with TD-OCT, a broadband light source is used, but the mirror in the reference arm is kept static. For SD-OCT (a), a spectrometer replaces the photodiode to analyze the output light from the interferometer. The interference signal is embedded in the spectrogram signal, called the interferogram. For SS-OCT (b), a tunable light source is used to produce a variation of wavenumber while a photodiode is used to analyze the interference of each wavenumber. Images adapted from [52].

k, as a monochromatic light source, as follows [51]:

$$
\begin{aligned}
I_D\left(k\right) =&\alpha\beta \; S(k)\left[R_R + \sum_{n=1}^{N} R_{S_n}\right.\\
&+ 2\sum_{n=1}^{N} \sqrt{R_R R_{S_n}} \cos(2k(z_R - z_{S_n}))\\
&\left.+ \sum_{n\neq m=1}^{N} \sqrt{R_{S_n} R_{S_m}} \cos(2k(z_{S_n} - z_{S_m}))\right].
\end{aligned}
\tag{2.7}
$$

$S(k)$ refers to the intensity of wavenumber k. Note that the reference mirror is kept static so that z_R is constant. Here, each pixel element in the spectrometer implicitly measures the sum interference of individual wavenumber, k, from all reflectors along the biological tissue (both cross-correlation and autocorrelation).

    The reflectivity profile of the sample is estimated from the *inverse Fourier transform* $\mathcal{F}^{-1}\{I(k)\}$. We can transform Equation 2.7 from the k-space domain into the spatial domain by using the property of Fourier pair $cos(kz_0) \overset{\mathcal{F}}{\longleftrightarrow} \delta[z \pm z_0]$, the convolution property of the Fourier transform $X(k)Y(k) \overset{\mathcal{F}}{\longleftrightarrow} x(z) \otimes y(z)$, and the sifting property of the convolution of the delta function $y(z) \otimes \delta[z \pm z_0] = y(z_0)$. Therefore, the inverse Fourier

transformation of Equation 2.7 is written as [51]

$$
\begin{aligned}
i_D\left(z\right) = & \alpha\beta\left[\gamma(z)\,R_R + \gamma(z)\sum_{n=1}^{N} R_{S_n}\right] \\
& + 2\alpha\beta\sum_{n=1}^{N}\sqrt{R_R R_{S_n}}\,\gamma[2(z_{S_n}\pm z_R)] \\
& + \alpha\beta\sum_{n\neq m=1}^{N}\sqrt{R_{S_n}R_{S_m}}\,\gamma[2(z_{S_m}\pm z_{S_n})]\,,
\end{aligned}
\tag{2.8}
$$

where $\gamma(z)$ is the *complex degree* function of the source, which is the inverse Fourier transform of the source spectrum $S(k)$. As with Equation 2.3, the transformed signal also consists of the DC, cross-correlation, and auto-correlation signals. The signal reconstruction is illustrated in Figure 2.7. The displacement of each reflector is $(2(z_{S_n}\pm z_R))$ because the interferometer measures the round-trip distance to each reflector. Here, the sample reflectivity profile is interpreted as the cross-correlation term $\sqrt{R_S(z_S)} = \sum_{n=1}^{N}\sqrt{R_{S_n}}\,\gamma[2(z_{S_n}\pm z_R)]$. Although the auto-correlation factor still surrounds it, it is negligible based on the same reasoning applied to Equation 2.3. The FD-OCT axial resolution is technically similar to that of TD-OCT, which is characterized by the coherence length of the light source l_c (Equation 2.4). An example of the reconstructed signal with multiple reflectors is shown in Figure 2.8.

(a) Spectral interferogram.

(b) Spatial domain interferogram.
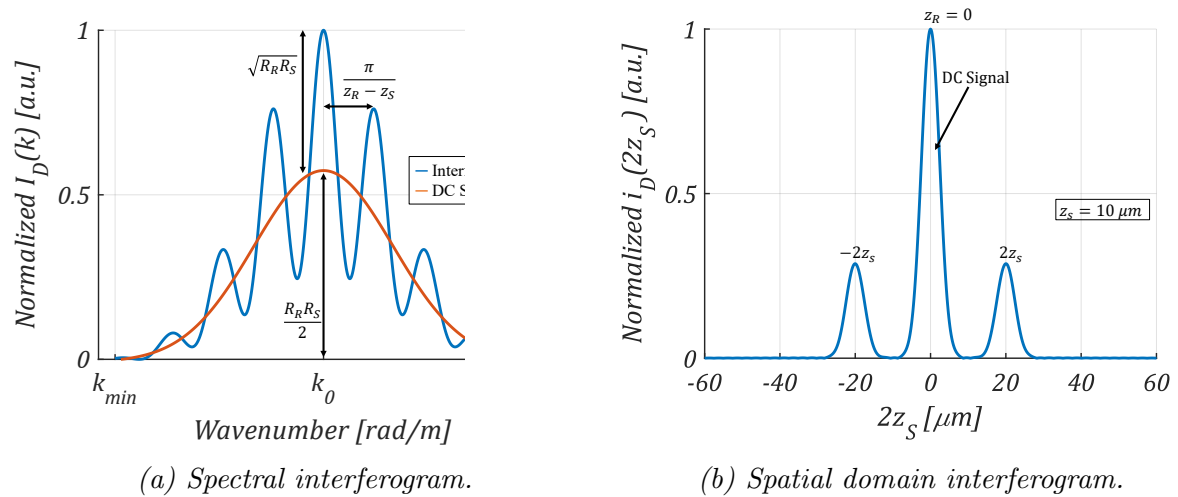
Figure 2.7: Illustration of the detected interferogram (a) and the inverse Fourier-transformed signal (b). The interference fringes are encoded on the spectrogram with the light source spectrum defining the DC signal. The transformed signal is doubled due to round-trip scaling. The transformed DC signal appears as a large artifact signal centered at zero path length difference. Images adapted from [51].
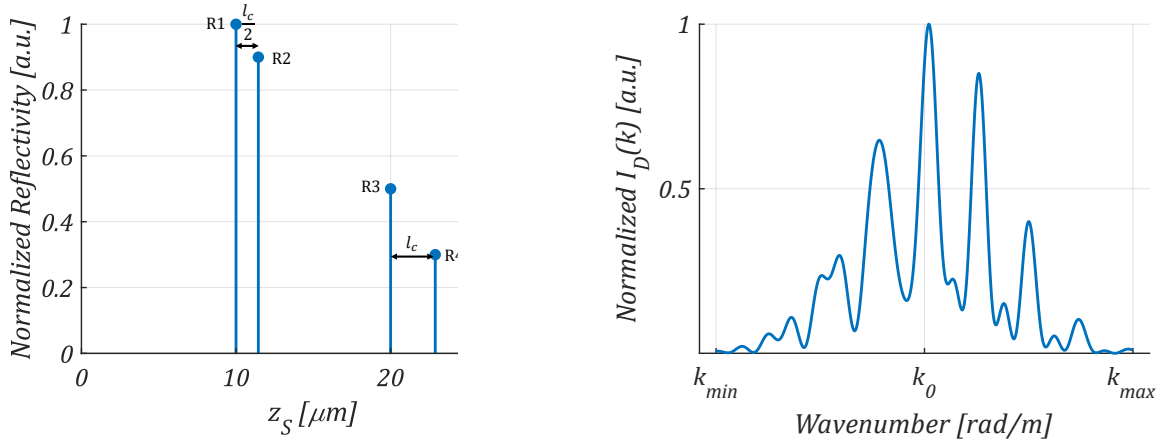
(a) Discrete reflectors in a sample.

(b) Interferogram with multiple reflectors.
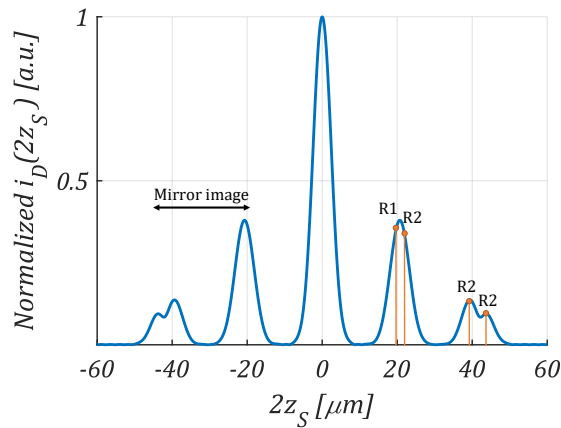
(c) Inverse Fourier transform of the interfero-
gram.

Figure 2.8: Illustration of the sample reflectivity profile reconstruction in FD-OCT. Figure (c) is
the inverse Fourier transform of the interferogram signals (b) coming from the reflectors (a).
As with TD-OCT, two reflective boundaries can be fully distinguished if located at a distance
$\geq l_c$ (e.g., R3 and R4, with distance l_c). It is difficult, however, to separate two reflectors with a
distance $< l_c$ (e.g., R1 and R2, with distance $l_c/2$). Images adapted from [51].

2.3 Signal to Noise Ratio and Sensitivity in FD-OCT

With the inverse Fourier transformation principle, the A-Scan rate of SD-OCT depends on the spectrometer's response time, which is usually faster than the scanning speed of the reference mirror in TD-OCT. SD-OCT uses multiple pixel elements in the spectrometers to acquire signal simultaneously[2] [53–55], which accounts for its faster scanning time. Consequently, SD-OCT has a higher Signal to Noise Ratio (SNR) than TD-OCT. Furthermore, since both OCT methods use the same principle of interference of coherence light, they yield similar axial and lateral resolutions.

The well-known expression for the SNR of a TD-OCT set up is given by [51]

$$SNR_{TD-OCT} = \frac{\langle I_D \rangle^2}{\sigma^2} = \frac{\rho S_0 R_S}{2eB_D},\tag{2.9}$$

where $\langle I_D \rangle^2$ is the mean-square peak signal power; it occurs at $z_R = z_S$ in Equation 2.2 for a single reflector, R_S, assuming that the power of the source, S_0, is split equally between the reference and sample arms ($\alpha : \beta = 50 : 50$). Both ρ and B_D are the responsivity and bandwidth of the detector, respectively. This equation was derived upon the assumption of *shot noise* limited detection, with shot noise variance σ^2.

Conversely, in SD-OCT, the backscattered light from all pixels in the A-Scan line (that contributes to the signal) is acquired simultaneously, so the noise in each spectral channel is uncorrelated. Thus, the noise variances add up incoherently in the inverse discrete Fourier summation. It has been shown both theoretically and experimentally that the SNR for both SD- and SS-OCT is defined as [51]

$$SNR_{SD-OCT} = SNR_{SS-OCT} = SNR_{TD-OCT}\frac{M}{2},\tag{2.10}$$

where M is the number of active pixel elements in the spectrometer. The M-factor improvement arises because all the pixel elements capture the interference signal in parallel [2], while in TD-OCT, the photodiode sequentially samples the signal. The division of factor M by 2 accounts for the fact that the inverse Fourier transform generates redundant data for positive and negative sample displacement.

This thesis project specifically focused on using SS-OCT to monitor the laser osteotomy process. Therefore, it is worth briefly discussing its advantages over TD-OCT and SD-OCT. SS-OCT uses a principle similar to SD-OCT, except that the wavenumber is swept linearly over time, $k = k_0 + mt$, where k_0 is the initial wavenumber and $m = \Delta k/\Delta t$ is the rate of change of the wavenumber over time (see Figure 2.6b). This mechanism is almost analogous to TD-OCT. However, during implementation, sweeping is relatively faster and more convenient than moving the reference mirror. Most of the current swept laser sources can sweep an A-Scan with a repetition rate of more than $100\,\text{kHz}$, which corresponds to $10\,\mu s$ per A-Scan. This instantaneous tuning speed gives SS-OCT benefits similar to

---

[2] Electronically, the digital acquisition in the Charge-Coupled Device (CCD) array happens sequentially along the pixel line array. However, the acquisition rate is relatively fast, usually in the magnitude of MHz, and is therefore considered simultaneous acquisition.

those of SD-OCT with a higher SNR than TD-OCT, as defined in Equation 2.10, where M in SS-OCT is defined as the total number of wavenumber samples per A-Scan.

FD-OCT suffers from sensitivity fall-off along the depth of the A-Scan line. Based on Equation 2.7, it is expected that the SD-OCT spectrometer could resolve the discrete individual wavenumber of the light source. However the spectrometer has a finite number of pixel elements and pixel size, which ultimately defines the spectral resolution (denoted as δk). The spectral resolution at which the signal can be detected introduces the fall-off in sensitivity. This limitation affects the spectral interferogram by convolution to the rectangular function of the spectral resolution, $rect(\delta k)$, and results in a Sinc function over depth, $sinc(\delta k\, z)$, after the inverse Fourier transformation as follows [51]:

$$I_D(k) \otimes rect(\delta k) \xleftrightarrow{\mathcal{F}} i_D(z) \cdot sinc(\delta k\, z)\,. \qquad (2.11)$$

Sensitivity fall-off causes the peak signals of a reflector to decrease over depth. For SS-OCT, δk depends on the instantaneous line width of the light source and the detection bandwidth of the photodiode and digitizer. Moreover, we could predict that the detected signal would fall close to zero at a distance point, denoted as z_{max}. Following the *Nyquist criterion*, z_{max} is defined as

$$z_{max} = \pm \frac{\pi}{2\delta k}\,. \qquad (2.12)$$

Although this fall-off effect applies to both SD- and SS-OCT, current OCT technology makes it more convenient to achieve an instantaneous spectral width with a narrower resolution using a sweeping light source (SS-OCT) than to increase the number of pixel elements in a SD-OCT spectrometer, which will technically require smaller pixel elements. Current commercial swept source lasers in particular, such as Vertical-Cavity Surface-Emitting Laser (VCSEL) [56] and akinetic swept laser [57], have a coherence length corresponding to an OCT depth range of centimeters.

2.4   Signal Preprocessing

Image reconstruction in an FD-OCT set up is relatively simple, requiring an inverse Fourier transform of the interferogram from the k-space domain to the spatial domain. However, it is necessary to preprocess the interferogram to further improve the image quality in FD-OCT. This section aims to explain the three major steps of preprocessing the FD-OCT interferogram before performing the inverse Fourier transform.

2.4.1   DC and background removal

The first step is to remove the DC and background signal. Recalling Equation 2.7, the interference signal is composed of the DC, cross-correlation, auto-correlation, and additional

dark background signals [58]

$$I_D \propto I_{DC} + I_{cross-correlation} + I_{auto-correlation} + B_{dark}. \qquad (2.13)$$

For image reconstruction, we are most interested in the cross-correlation signal, which contains the sample reflectivity profile. Therefore, we need to subtract the DC, auto-correlation, and dark background signals from the interferogram. The auto-correlation signal can be neglected, given that the reflectivity of the reference arm is usually greater than the reflectivity of the biological tissue in the sample arm ($R_R \gg R_S$).

Here, we introduce the term dark background (B_{dark}). This signal can be measured by capturing the signal while closing both the sample and reference arms. The dark background signal represents the external noise that might arise due to external factors, such as electrical noise and environmental light sources. The DC signal consists of signals coming from both the reference and sample arms. The signal from the reference arm, I_{ref}, can be measured by closing the sample arm completely. Likewise, the sample background signal, I_{samp}, can

(a) Interfer...   ...ubtracted.

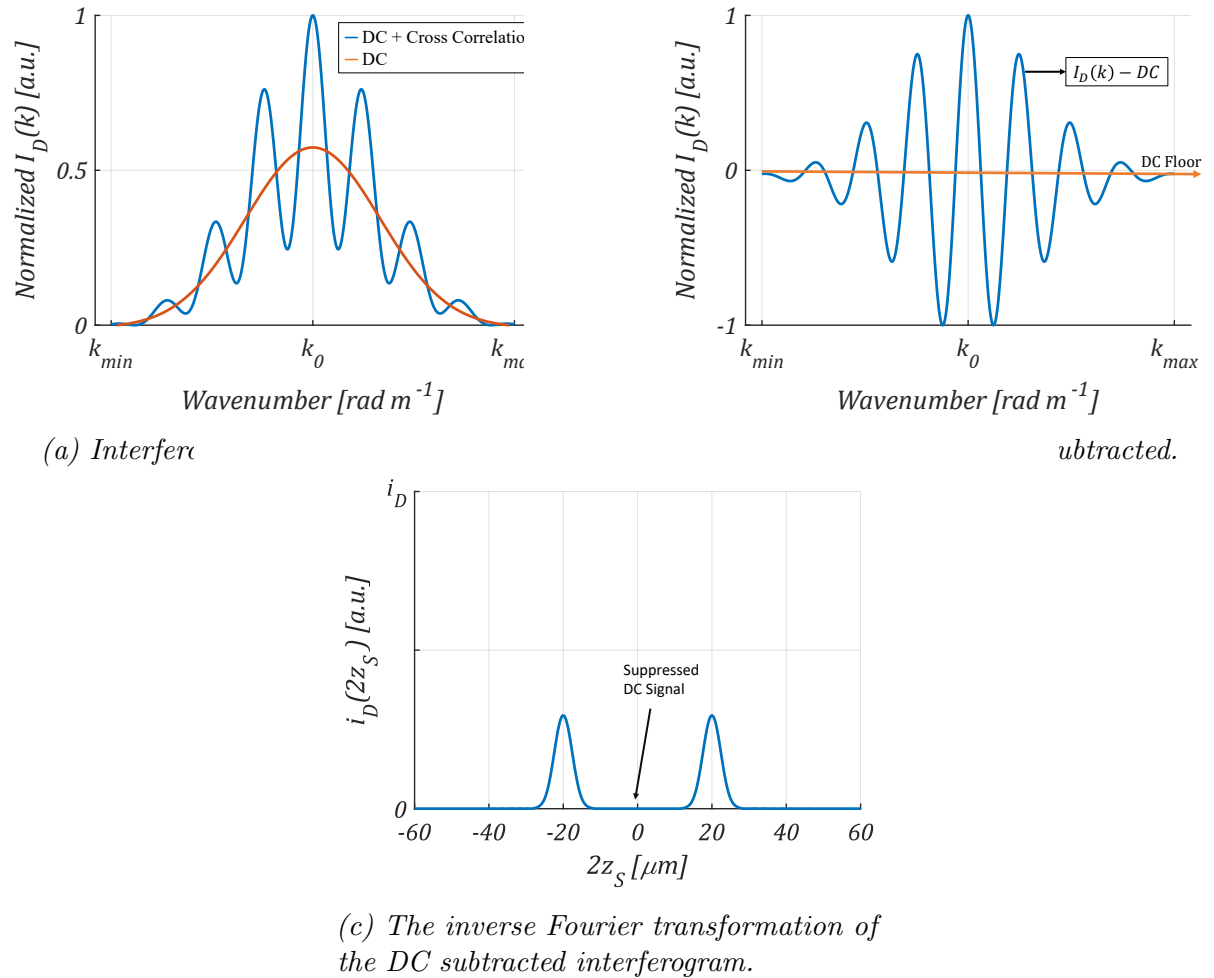(c) The inverse Fourier transformation of
the DC subtracted interferogram.

Figure 2.9: Illustration of the DC subtraction. DC signal subtraction (b) from the original interferogram (a) leads to suppression of the DC peak on the inverse Fourier transformation (c).

be measured by closing the reference arm. However, the dark background signal must be subtracted as it was measured with each one of the signals, $I_{DC} = I_{ref} + I_{samp} - 2B_{dark}$. This method preserves the cross-correlation signal with a high degree of accuracy; however, it is sensitive to changes in the DC signal (e.g., due to changes in the optical setup or the light source). Therefore, it is necessary to measure the total background signal before every measurement with an actual biological sample.

Alternatively, it is more convenient to measure the background signal by taking the average interferogram over all A-Scans in a B-scan image [58,59]. This averaging technique corresponds to measuring the sum of DC and dark background signals, which results in a static signal, $\langle I_D \rangle_t \simeq I_{DC} + B_{dark}$. The fluctuating interference (cross-correlation) signal will have a relatively small impact on the static signal. This method is preferred in most applications because it is more adaptive to changes in the DC signal.

2.4.2 k-space linearization

The second major step in processing the OCT signal is to remap the spectral wavelength (λ) into wavenumber (k). Fourier transformation relates the physical distance, z, to the wavenumber, k, of the reflected light wave. However, the spectra obtained with the spectrometer in SD-OCT have a unit of wavelength. The wavelength is inversely proportional to the wavenumber ($k = 2\pi/\lambda$), which will induce phase non-linearity over the signal and leads to the broadening of the point spread function and deteriorates the reconstructed image resolution [60]. This is similar to the dispersion problem, which will be discussed in the next section. Therefore, before applying the Fourier transform, the data must be resampled so that it is evenly (linearly) spaced in the k dimension (k-space). The remapping process usually involves interpolation from the wavelength function as the wavenumber function. Typically, the interpolated spectrum is fitted to the third-order polynomial, and the nonlinear part is used to correct the wavelength. Practically, SS-OCT provides a finer and more convenient way of tuning the light source with a linearly spaced wavenumber (calibrated for linearity), making the remapping process unnecessary. Therefore, in this thesis project, the k-space linearization step is skipped.

2.4.3   Windowing

The last significant step in preprocessing the OCT signal is windowing the signal. A major artifact in OCT A-Scans comes from sidelobes in the reconstructed signal, which could reduce image sharpness and contrast [61]. These sidelobes arise from the imperfection of the Gaussian-shaped spectrum for most broadband light sources. Specifically, in SS-OCT, most of the light source's spectrum is flat or close to a rectangular function. Therefore, the interferogram must be reshaped (windowing) to reduce the sidelobes. The windowing process involves reshaping the interferogram envelope to match the Gaussian-shaped spectrum. There are several different windows that can be applied. Several studies have explored various window filters, all with nearly similar effects. For most cases, a Hanning filter [62] is sufficient. Figure 2.10 illustrates the sidelobe effect resulting from the inverse Fourier transform of an interferogram with a near linear DC envelope.
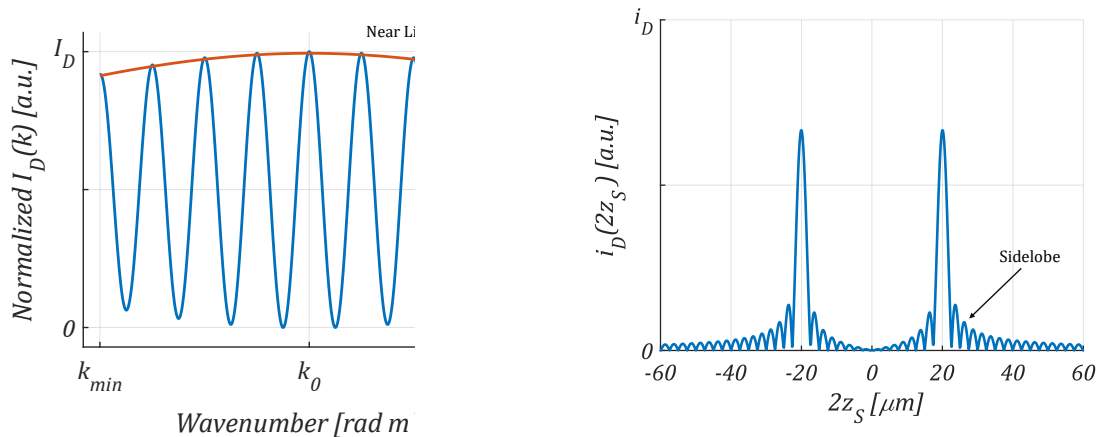
(a) Interferogram with near flat envelope.

(b) The inverse Fourier transform.

Figure 2.10: Illustration of an inverse Fourier transformation of an interferogram with a near linear light source spectrum. The reconstructed reflector location is embedded with sidelobe artifacts due to the transformation of an imperfect Gaussian-shaped interferogram. Practically, a windowing filter could digitally reshape the signal to get an ideal signal, as shown in Figure2.9b.

2.5   Speeding Up the Signal Processing

FD-OCT technology has been widely used in modern OCT systems. Higher sensitivity and speed account for much of FD-OCT's success over TD-OCT. The previous sections highlight the key aspects of physics that have enabled an increase in sensitivity and, consequently, increased imaging speed without loss of image quality. Increasing the speed allowed OCT to shift from 1D point sampling measurements to comprehensive 4D imaging [63, 64]. Scanning speeds increased from 2 A-Scans/second (TD-OCT) to a few million A-Scans/second, using SS-OCT with high-speed swept-source laser technology [65].

To fully take advantage of FD-OCT's speed, a high-performance computational resource is also required for real-time signal processing and OCT image visualization. Real-time visualization and image analysis would improve pathology diagnosis and extend the scanning ability from 3D volume images to 4D dynamic volumes, such as in OCT-Angiography. Fast image reconstruction is also necessary to monitor the laser osteotomy process, with its high repetition rate. Significant computational resources are required for complex FD-OCT interferometric data processing, especially for the inverse Fourier transformation process, which demands an extremely high data throughput. To address this problem, one must consider additional aspects for improving OCT performance beyond the optical components that have been discussed in the previous sections. While OCT image resolution depends on the coherence length of the light source and the aperture of the focusing optics, its visualization performance depends on the clock speed of the computer processor, which is currently limited. In other words, OCT visualization performance is limited by both the image acquisition rate and the digital processing speed.

Most modern OCT systems have a high acquisition rate of up to a few million A-

Scans/second. Nevertheless, the clock speed of a single processor to process the signal is limited. Therefore, the computational speed to process A-Scan signals one by one (sequentially) in a B-Scan is also limited. A multi-threading technique, or parallel processing, plays an important role in real-time OCT image visualization. This technique takes advantage of multiprocessors that can work together simultaneously. The method is strongly compatible with FD-OCT's high data throughput, which usually slow with a sequential process. Parallelization in OCT mainly employs a Graphical Processing Unit (GPU), which contains more coprocessors and has a higher memory bandwidth than most Central Processing Unit (CPU) clusters.

For comparison, an NVIDIA GPU GTX1050Ti has 768 virtual processors in one board, with an individual processor clock speed of 1.4 GHz. Meanwhile, an Intel i7-7700 Quadcore CPU has eight virtual processors with an individual processor clock speed of 3.6 GHz. A single fast inverse Fourier transformation (A-Scan) in Matlab 2020b, with 2048 swept points per A-Scan, running on this CPU took ∼6.8 µsec (without CPU parallelization). The same process took ∼48 µsec in the GPU, 7x slower than the CPU. However, when transforming 128 A-Scans, the GPU took ∼435 µsec, almost 4x faster than the CPU, which took ∼1,143 µsec[3].

Fast image reconstruction was necessary for this thesis project in order to monitor the laser ablation process, which uses a high repetition rate. Furthermore, parallel computing in a GPU has become the standard for deep learning inference, where deep learning methods are used in both speckle reduction and tissue differentiation. In this thesis project, almost all of the computational processes to reconstruct the OCT image were executed in a GPU. Additionally, it is more convenient to both reconstruct images and detect tissue using GPU processing to avoid redundant communication between the GPU and CPU.

2.6   Speckle in OCT and Reduction Methods

This section will discuss the origin of speckles and explains a few methods for reducing them. Speckle in OCT images appears as a random granularity that has no obvious relationship to the texture of the biological tissue. One could easily assume that it is just random noise outside the OCT system. However, this characteristic appearance of OCT images arises from the fact that we are dealing with interference of coherent light. The coherence property of light is both a strength and a weakness of OCT [66]. The signal modulation resulting from light coherency enables us to detect structural interfaces with very high resolution. However, it is also the cause of speckle patterns in OCT images, which ultimately reduces the resolution and sensitivity. Note that it is important to distinguish between speckle noise and the noise described in Section 2.3.

The origin of the speckles can be traced from the interference Equation 2.2. The detected signal is modulated as an interference pattern. Coherency limits the interference

---

[3]Note that this example is based on a personal benchmark with a specific computer. Results may differ with different computer specifications and different number of Fourier transformations.

pattern to a package of signals (coherent envelopes). In other words, the detection system packed the continuous structure into a discrete function of the path length difference (sample volume). As a result, an ideal object would need to have one scatterer for each discrete path length difference to get a perfectly modulated signal. However, there will be microscopic structures between the sample surface and path length locations (l_c), which lead to reflections.

In principle, the detected signals should only come from multiple backscattered lights within the sample volume, as illustrated in Figure 2.11a. However, the microscopic structures consequently cause a detection delay due to multiple forward-scattering effects during propagation through the sample volume[4]. The forward-scattered lights contribute to random delays and distort the path length detection in the axial direction. Conversely, multiple backscattering occurs within the sample volume and contributes least to signal delays. The forward-scattered lights give rise to secondary interference between the multiple backscattered lights within the sample volume and distort the path length detection in the lateral direction. As a result of these two effects, the returning wavefront is distorted, from an ideal spherical wavefront to a wavefront with localized regions of constructive and destructive interference, and detected as random delays in the signal. Lights scattered from any point outside the focal volume (due to deflection from the forward scatter) also generate speckles. As long as these altered path lengths fall within the coherence length, the interference signal will still be observed. Since these scattering events are distributed randomly, the interference pattern is modulated randomly to form a speckle pattern.

From this explanation, speckle can be classified into two categories. The first category is the signal-carrying speckles, originating from multiple backscattered lights within the sample volume (focal zone). This speckle mostly appears with large granularity, illustrated in Figure 2.11b. The second category is the signal-degrading speckles, originating from forward-scattering lights during propagation and the scattered light outside the focal zone. This speckle induces small granularity in the image. The signal-degrading speckle can be reduced by increasing the NA of the focusing optics. However, the optimal NA should be set such that the depth of field is larger than the coherence length, as explained in Section 2.1, to benefit from the coherence length resolution.

At this point, we have discussed the origin of speckles as an interference phenomenon. Another explanation that has been put forward to account for speckles is the missing frequency problem. This problem appears as a limitation of the coherency tomographic method. OCT acts as a bandpass filter. OCT scanners can detect only those objects whose spatial frequency spectra overlap the band of spatial frequencies, from $2/\lambda_0 - 1/l_c$ to $2/\lambda_0 + 1/l_c$ [66]. Only biological structures within this spatial frequency band will be detected uniformly by OCT and appear as image granularity due to the modulation of the signals. Structures with smaller sizes are absent in OCT images. At the same time, the higher frequency limit appears as the axial resolution of the OCT system. One solution

---

[4]In principle, some of the lights would also experience multiple microscopic (premature) backscatters during propagation through the sample. However, these backscatters are outside of the coherence length and will not deteriorate the image.
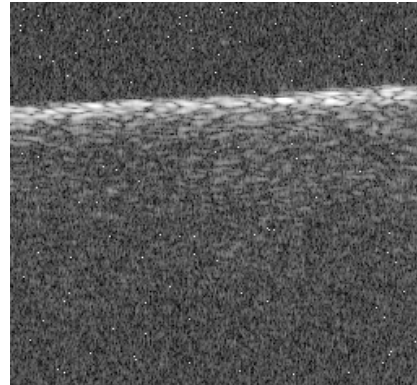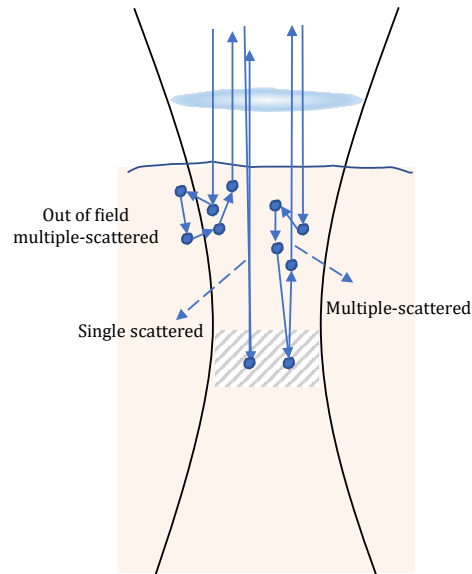
(b) An OCT speckled image example.

Figure 2.11: Illustration of light scattering in biological tissue (a). Speckle in OCT images originates from multiple backscattering on the sample volume. Meanwhile, random speckle noise appears when the signal originates from multiple forward-scattering and out-of-field scattering. Image (b) shows an OCT bone tissue image containing speckles that have no obvious relation to the bone structure. Image (a) was adapted from [66].

to this problem is to use a broader bandwidth light source, which also improves the axial resolution.

Ideally, we would like to remove speckles from the OCT image completely. However, completely removing speckles from the OCT image also means removing both the signal-carrying and signal-degrading speckles, which consequently remove important structural information about the biological tissue. This information is important for texture-based tissue differentiation, which will be explained in the next chapter. Here, OCT speckle reduction is addressed as a process to minimize signal-degrading speckle while preserving the signal-carrying speckle. In practice, there are four basic techniques that can be used for speckle reduction. However, none of them are theoretically ideal and each of them has its respective shortcomings. The practicality of implementing the technique must also be considered. Nevertheless, we will discuss the techniques briefly without expanding on how they could be implemented in the OCT system.

- **Polarization Diversity**

  The first technique is polarization diversity, which can be implemented by rotating the linear polarization state during the exposure time. The spectral changes will partially average out the speckled signal. However, this approach only improves the SNR by a factor of $\sqrt{2}$ at most [67], and is therefore quite limited.

- **Spatial Compounding/Frame-Averaging**

  The second technique is spatial compounding or frame-averaging, which involves averaging a series of images from the same or slightly displaced sample volume. Displacement must be minimal so that the signal-carrying speckles in the image remain correlated between frames. Otherwise, the resulting image would show the effects of smoothing and loss of sharpness. The effectiveness of this approach depends on the number of averaged images. This technique yields an improvement of the SNR by a factor of $\sqrt{N}$ [68]. The problem with this approach is the need for multiple image acquisitions, which usually require a relatively longer acquisition time. Nevertheless, this approach usually serves as the gold standard because it can be implemented without additional changes or modifications to the OCT components.

- **Frequency Compounding**

  The third technique takes advantage of the reduced correlation between speckles within different frequency bands. There are two ways to create a variety of frequency bands (frequency compounding). First, the light source can be split with a bandwidth of $\Delta\lambda$ and centered at λ_0 into N equally non-overlapping frequency bands. The reconstruction process is similar to frame averaging, where each frame is acquired with a different individual frequency band. Frequency compounding also improves the SNR by a factor of $\sqrt{N}$ [69], but partitioning the spectral bandwidth reduces the axial resolution of the OCT system.

- **Digital Image Processing**

  The last technique for speckle reduction is often implemented because it can be applied simultaneously after the Fourier transformation of the interferogram. Several algorithms have been studied to perform this technique. The most popular, simple, and relatively fast methods are image filtering, including median, Gaussian, Wiener, and wavelet filters [70–73]. However, these image filters typically result in over-smoothing the image and reduce its sharpness.

In this project, reducing speckles in the OCT image is important. The tissue textural information strongly discriminates among different tissue types, especially for distinguishing between uniform and non-uniform tissue. However, a fast speckle reduction technique is required to monitor the laser ablation process at a high repetition rate. The frame averaging technique is easily adapted to different OCT systems and produces a relatively higher quality image, but it requires a longer image acquisition time. A post-processing filter works relatively faster than frame averaging. This thesis project tried to use a deep learning algorithm as a black box that mimics the frame-averaging method [74, 75]. The current deep learning technology makes it possible to reduce signal-degrading speckle with a faster filtering process while maintaining the signal-carrying speckle. A more detailed explanation of the algorithm is explained in Chapter 4.

CHAPTER 3

Tissue Characterization and Feedback System with OCT

The main objective of this thesis project involves tissue characterization. Tissue characterization aims to quantify tissue properties or features that can be used to classify tissue type. Tissue characterization for medical images, especially OCT images, is usually performed by extracting the features associated with different tissue types and validating the result with standard histology images. Generally, human experts or medical doctors perform these steps. Nowadays, however, the vast majority of research in the medical imaging field, including OCT, seeks to automate these processes. Even more, improvements in deep learning led to a paradigm shift, from manually engineered feature extraction to using a Deep Neural Network (DNN) as a black box capable of extracting discriminating features automatically.

This chapter begins by discussing classical tissue characterization, including feature extraction from the OCT image, which relies on the tissue's optical properties and geometrical texture. It is followed by the application of machine-learning algorithms to differentiate tissue types based on the extracted features. Such automated tissue characterization systems benefit medical doctors by providing fast and repeatable image analyses. The work in this thesis moves away from the traditional approaches, using deep learning methods, a cutting-edge branch of machine learning capable of automatically learning complex and robust features from raw data without the need for feature engineering. The second section of this chapter discusses the basic working principle underlying the deep learning model. The last section of this chapter briefly discusses the ablation laser and its integration with the deep-learning-based OCT tissue characterization setup for real-time osteotomy monitoring.

3.1 Classical Tissue Characterization

Tissue characterization is the classification of tissue based on its typical appearance in an OCT image. Traditionally, medical doctors with specialized expertise are trained to analyze

OCT images based on tissue anatomy. However, the analysis is often subject to both intra- and inter-observer variability. Automating these processes with a computer would help surgeons achieve consistent results. The typical steps of building the tissue profile or characterization are shown in Figure 3.1. Here, the subjective high-level knowledge of tissue characteristics (features) is often difficult to quantify. Therefore, interpretable a

t

p

l

Figure 3.1: Common steps of building an automatic tissue characterization system in OCT. It starts with collecting images from an OCT device and manually labeling the tissue type of the corresponding image. It is followed by feature extraction to find unique characteristic features of a tissue embedded in the image. The extracted features are then split into training and validation datasets. A machine-learning model will be trained and validated before it is applied in clinical conditions. Additional testing datasets are usually used as a benchmark by which to evaluate the performance of the machine-learning model.

3.1.1   Attenuation Coefficient

The use of light to carry the signal poses a drawback in terms of limiting attenuation in biological tissue. In addition to the confocal properties of the focusing optics and spectral resolution in FD-OCT, light attenuation also governs the maximum depth of an A-Scan. This optical property, however, could be used as one parameter by which to characterize biological tissue. The optical Attenuation Coefficient (AC) is a tissue property that defines how incident light is attenuated when passing through a medium. It is an important parameter that enables quantitative analysis of tissue property signals and acts as a strong tissue discriminant for several OCT applications [76]. Additionally, the optical setup in most OCT research is designed such that the imaging depth should be greater than the light attenuation through tissue.

    The AC defines how incident light is attenuated when passing through a medium and is dependent on the underlying medium's optical properties. Light loss in tissue can be caused by absorption, scattering, or a combination of the two. The intensity detected over depth, $I_D(z)$, in OCT can be modeled as a single scattering event that incorporates the

axial point spread function and follows the Lambert-Beer law equation [77]

$$I_D(z) = I_0 \, h(z) \, s(z) \, e^{-\mu_t z}, \tag{3.1}$$

where z and I_0 indicate the penetration depth and source intensity, respectively. The attenuation coefficient, μ_t, is the result of scattering and absorption. Here, $h(z)$ is the confocal function defined by the focusing optics. Specifically, in FD-OCT, $s(z)$ accounts for the sensitivity fall-off function, while in TD-OCT, it is considered a constant.

Multiple methods have been proposed to extract the AC from OCT images. The simplest method is the curve fitting method [78]. This method estimates μ_t by simplifying Equation 3.1 to a logarithmic function,

$$\log[I_D(z)] = \log[I_0] + \log[h(z)] + \log[s(z)] - \mu_t z. \tag{3.2}$$

This equation presents a linear relation between the logarithmic intensity $\log[I_D(z)]$ and a function of depth $\mu_t z$, with μ_t indicating the slope. Hence, we could estimate μ_t using the linear least-square fitting method. Note that we need to consider $\log[h(z)] + \log[s(z)]$ as a constant value and $\log[I_0]$ as an offset intensity. One drawback of this method is that it needs a large amount of data to accurately fit the curve. Additionally, this method is unable to extract values over small tissue regions or multilayered tissue.

Another approach to extracting the AC is the depth-resolve method proposed by Vermeer et al. to overcome the resolution limit of the curve fitting method [79]. This method enables a pixelwise AC extraction and provides a better estimation of the AC in multilayered tissue. It introduces the depth-dependent attenuation coefficient, $\mu_t(z)$. Considering the confocal $h(z)$ and sensitivity $s(z)$ functions as a constant function, we could rewrite Equation 3.1 into

$$I_D(z) = I_0 \, e^{-2 \int_0^z \mu(u) du}. \tag{3.3}$$

Here, the $-\int_0^z \mu(u) du$ denotes the sum of the attenuation coefficient of the biological tissue from 0 to depth z. The factor of 2 accounts for the round-trip attenuation of light in tissue. Solving the definite integral of this equation from depth z to infinity yields an AC estimation of

$$\mu(z) \approx \frac{I(z)}{2 \int_z^\infty I(u) du} \approx \frac{I(z)}{2 \int_z^D I(u) du}, \tag{3.4}$$

where D is the maximum imaging range of FD-OCT. The pixelwise AC values are then

estimated by calculating the discrete integral and are expressed as

$$\mu\left[i\right] = \frac{1}{2\Delta} \log \left(1 + \frac{I[i]}{\sum\limits_{j=i+1}^{N} I[j]} \right) \approx \frac{I[i]}{2\Delta \sum\limits_{j=i+1}^{N} I[j]}, \tag{3.5}$$

where i denotes the pixel index location on an A-Scan and is bounded to the last pixel index N, $(i \leq N-1)$. Thus, the number of reconstructed AC pixels is reduced by one pixel for each A-Scan. Notation Δ is the pixel size. The right side of the equation is the first-order linearization, with $\log(1+x) \approx x$, assuming x is small.

This pixelwise method provides a more convenient way of mapping the AC for heterogeneous tissue. One limitation of this method is that it fails to extract the AC map when the light does not completely attenuate at the last pixel location. Liu et al. optimized the method by adding a constant AC value from the last data point N [80]. They rewrite Equation 3.5 as

$$\mu\left[i\right] \approx \frac{I[i]}{2\Delta \sum\limits_{j=i+1}^{N} I[j] + \frac{I[N]}{\mu[N]}}. \tag{3.6}$$

The AC at the last point $\mu[N]$ is approximated using the curve fitting method.

Up to this point, we have discussed the commonly used method for extracting the AC from OCT images based on a single scattering event model. Another method attempts to model the light interaction as a multi-scattering event. Almasian et al. developed a comprehensive AC extraction method with a multi-scattering model [81]. However, due to its complexity, it would be difficult to implement this method for real-time extraction and use in a laser osteotomy monitoring system.

3.1.2 Textural Features

Another simple and straightforward way to characterize tissue objects in most medical images is by analyzing the texture appearance. There are several quantitative parameters that can be extracted from the textural appearance of tissue in an OCT image. The most commonly used and interpretable parameters are as follows:

- **First-order statistics** are related to the distribution of the pixel intensity and ignore inter-pixel correlations. These parameters include mean value, median value, standard deviation, skewness, and kurtosis of the tissue texture [82].

- **Gray level co-occurrence matrix** is the second-order statistical parameter for texture analysis. It examines the spatial relationship among pixels and defines how often different combinations of pixel brightness values (gray levels) occur in an image. Based on this matrix, the features that can be extracted include contrast, correlation, dissimilarity, energy, entropy, homogeneity, and maximum probability [83].

- **Neighborhood gray-tone difference matrix** corresponds to the visual properties of the texture. It quantifies the difference between a gray value and the average gray value of its neighbors' pixels within a kernel matrix. Five texture features can be extracted from this matrix, including busyness, contrast, complexity, coarseness, and texture length [84].

- **Fractal dimension** is a characterization index of a fractal pattern, established by quantifying its complexity as a ratio of the change in detail to the change in scale. Typically, the differential box-counting algorithm is used to quantify the pattern change [85].

3.1.3 Classical Machine Learning Methods for Tissue Classification

Traditionally, automating tissue characterization involves a machine-learning algorithm to classify tissue types based on the features described in the previous sections. The main purpose of machine learning is to implement an algorithm that can classify tissue types by learning from the feature data, without explicit programmed[1]. In practice, the data is divided into two main datasets, the training set and the testing set. The training set serves as the learning data to optimize the machine-learning algorithm, which is then referred to as the classifier model. The testing set is used to evaluate the performance of the optimized model.

There are two main approaches to training the classifier model. The first approach is supervised learning, whereby each of the samples in the dataset has a true label or ground truth, and the model is trained to give a label as similar as possible to the true label. The second approach is unsupervised learning, where the data does not have any associated labels or prior knowledge of the ground truth. Although unsupervised learning seems powerful, employing autonomous learning to discover the inherent structure of unlabeled data, the approach is rarely used for classification problems. Ethically, it is difficult to justify the validity of unlabeled data for medical applications. In this thesis project, the discussion focuses on supervised learning, due to its simplicity. Among the available machine learning algorithms commonly used in research of tissue characterization, two supervised learning algorithms have been selected for illustration: the random forest and Support Vector Machine (SVM). Both of these algorithms have been studied for automated tissue characterization in intravascular plaque detection [86, 87], malignant detection [88, 89], and glaucoma detection [90].

- **Random forest** is an ensemble classifier that combines numerous decision trees. It combines multiple binary tree predictors, where each tree consists of a randomly sampled subset of features. Choosing the predictors in a random subspace of the training data reduces the correlation between the individual decision trees and

---

[1]Although some of the algorithms were developed for regression problems, the discussion focuses on their application to the classification problem.

provides a variety of trees. This strategy is better at reducing the risk of overfitting than using a normal decision tree. For a classification task, a majority vote from multiple predictors' outputs, where the most frequent categorical label, will yield the predicted class [91].

- **Support Vector Machine** SVM is a supervised learning algorithm that analyses data by constructing a set of hyperplanes used for separating different classes of data in the feature space. The linear SVM classifier can be viewed as an extension of the perceptron, an earlier classical machine-learning algorithm. Although perceptron guaranteed finding a separating hyperplane on separable data, it mostly failed to separate noisy data (embedded with outliers) that may be inseparable with a linear hyperplane. In contrast, the SVM algorithm tries to find the maximum margin of a separating hyperplane (support vectors) and adds a penalty term for violating the classification constraints; thus, it performs better in noisy data. Furthermore, SVM can be implemented using a nonlinear kernel function, such as polynomial, sigmoid, and radial basis functions, to provide optimal separation for more complex data [92].

3.2 Deep Learning

Deep learning is a subset of machine learning. It is regarded as an improvement over conventional Artificial Neural Network (ANN) by building networks with multiple (more than two) hidden layers, which are then called Deep Neural Network (DNN). As the backbone of deep learning models, ANN is inspired by neurobiological aspects, although its main goal is not to artificially rebuild biological brain systems. Technically, mimicking the connectivity of neurons in the brain helps ANN to automatically discover hierarchical feature representations, such that higher-level features can be derived from lower-level features [93]. Its success in computer vision has inspired its use in several medical imaging analysis studies, such as image denoising, classification, and segmentation.

This thesis project focused on applying a DNN for tissue classification, based on OCT images. Compared to classical machine-learning methods, DNN can be considered as a black box, capable of identifying feature representations automatically, without any feature engineering or feature extraction techniques applied to the dataset [93]. Skipping the feature extraction process is a key benefit for real-time feedback. This superiority of DNN over other classical machine-learning algorithms can be regarded from the approximation theorem, which describes ANN as a universal function approximation [94]. This theorem states that an ANN with a single hidden layer is capable of approximating any continuous functions with arbitrary complexity if it has a continuous input-output relation and contains a sufficiently large number of neurons. Furthermore, DNN employs more than one hidden layer, which theoretically offers more benefits than a single hidden layer ANN, especially when the data has discontinuities. Several deep learning studies have shown that approximation capacity grows faster with multilayer neural network models than with single layer ones [95, 96].

3.2.1 Deep Neural Network General Representation

The artificial (deep) neural network model mimics the behavior of a biological neural network by building a generalization of mathematical models of human neuronal biology. The very basic building blocks of DNN are the adaptive computational units or neurons. Figure 3.2 illustrates the mathematical model of a neuron with the vector x_i representing the input of the neuron that receives signals from the external sensory system or other neurons. The weights vector, $w = [w_1, w_2..., w_i]$, is the synaptic weights that modify the received signal. These weights can be interpreted as gains that attenuate or amplify the input values. The parameter b_k is the bias of a neuron. Here, the learning process of an ANN refers to the process of modifying the weights of connections between the neurons. The output of the neuron is the sum of the multiplication between the inputs vector and the weights vector and is activated by an activation $\psi(.)$, written as

$$y = \psi(\sum w_i x_i).\tag{3.7}$$

The activation function determines whether the neuron is activated or deactivated. Several options are available, such as the linear, sigmoid, tanh, and Rectifier Linear Unit (ReLu) activation functions. Nowadays, the ReLu activation function is one of the most popular functions due to its simplicity. Although it has a linear relationship, it provides nonlinear transformation. With enough neurons, it can be used to approximate arbitrary nonlinear functions, such as the exclusive disjunction (XOR) and sinusoidal functions, which cannot be approximated by simple linear activation functions [97]. Furthermore, unlike other smooth activation functions, ReLu is insensitive to the vanishing gradient problem [98], a condition where the weight is either extremely small or extremely big.

The beginning of this section states that complex functions can be approximated with an ANN with a large finite number of neurons, which practically implies the stacks of neurons that make up a neural layer. The DNN extends the ANN with multiple (two
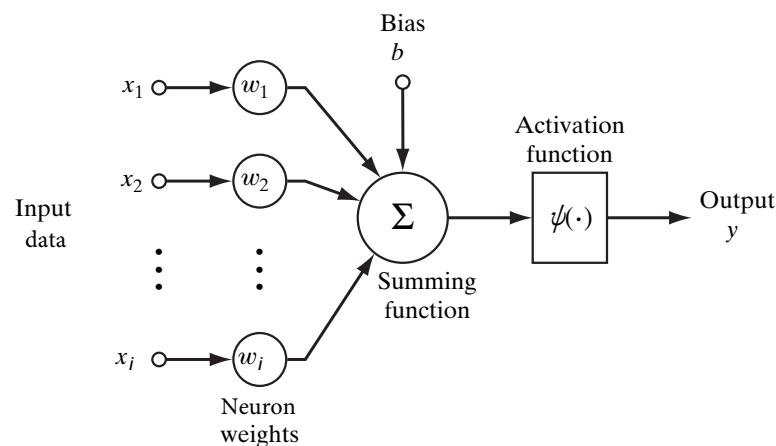
Figure 3.2: General model of a neuron in deep learning. Image is redrawn from [94].

or more) neural layers, and the practice of training this type of network is called deep learning. The general representation of a DNN can be seen in Figure 3.3. It is important to understand the layers of a DNN model, which can be distinguished into one of three types of layers:

(1) The *input layer* consists of an input vector $(x_1, x_2, ..., x_i)$ that supplies information coming from the external sources of the network. In computer vision, it is the image that needs to be classified. It is usually followed by at least one neural layer called the hidden layer.

(2) The *hidden layer* contains the internal neurons of the network that have no connection with the outside world. For DNN, the number of this layer should be more than one. The input of this layer comes from the input layer or output from the previous neural layer. The hidden layers are key components for capturing data's complex nonlinear behaviors. Complex features extracted from the training data are stored by the weight vector $w_j = [w_{1,j}, w_{2,j}..., w_{i,j}]$.

(3) The last layer is the *output layer*, which consists of a set of neurons that transfers the most discriminating features learned from the training data to the outside world. The output can be continuous, binary, or ordinal, depending on the specified activation function. For classification problems, the output layer is typically equipped with a ridge function, such as a sigmoid function, to translate the output information into a probability distribution of classes.
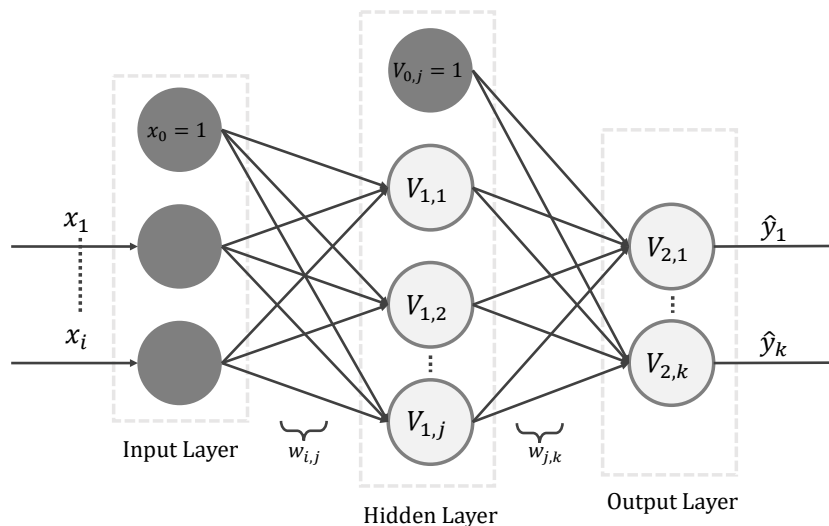
Figure 3.3: General representation of a multilayer deep neural network, where w_0 is the bias of a layer and $V = \psi(\sum(...))$ [see Equation 3.8]. The dark gray node represents a constant value, and the light gray node represents a trained variable. Image is redrawn from [94].

The mathematical form of the model represented in Figure 3.3 with output layer $\hat{y}$,

inputs layer x, and a hidden layer V, is given by the following equations

$$
\begin{aligned}
V_{1,j} &= \psi\left(\sum w_{i,j}x_i\right) \qquad \text{[hidden layer]}, \\
\hat{y} = V_{1,k} \circ V_{1,j} &= \psi\left(\sum w_{j,k}V_{1,j}\right) \qquad \text{[output layer]},
\end{aligned}
\tag{3.8}
$$

where the connections between layers are defined as a composition function. The learning process is obtained by updating the weight w.

3.2.2 Loss Function and Optimization

A machine learning model undergoes a training process, defined as the process that tries to minimize a loss function. In general, a loss function is a function that quantifies how close the predicted values produced by a machine learning model are to the true values. While training a DNN model, the output layer is connected to a loss function. Practically, a convex function is preferable, such as the mean squared error (MSE) or the logistic error function. For classification problems, logistic loss functions are preferred, since we are mostly interested in the probabilities of success rather than in just the hard classifications.

This logistic loss function originated as the negative log-likelihood of the product of Bernoulli distributions. It is also known as cross-entropy loss. For binary classification (0 or 1), it is defined as [99]

$$
L(w) = -\sum_{i=1}^{n}[y_i \times \log(\hat{y}_i) + (1 - y_i) \times \log(1 - \hat{y}_i)],
\tag{3.9}
$$

where $\hat{y}$ is the predicted output and y is the true output. Note that n is the size of the training dataset. When the number of output classes is more than two, the loss function is known as categorical cross-entropy [99] and is equal to

$$
L(w) = -\sum_{i=1}^{n}\sum_{j=1}^{m}[y_{i,j} \times \log(\hat{y}_{i,j})],
\tag{3.10}
$$

where m defines the number of targets (class outputs) that the network has to predict.

The optimization process in deep learning is denoted as updating the neural weights such that the loss function is minimized. Optimization algorithms used for training DNN models differ from traditional machine-learning optimization algorithms in several ways. Most of the time, it is not possible to use an analytical solution to estimate the weights. Therefore, optimization in deep learning is mostly obtained through iterative optimization algorithms. One iterative optimization algorithm that is commonly used in deep learning is gradient descent.

The straightforward gradient descent optimizer family is the batch gradient descent which uses the entire training dataset and tries to minimize the average total loss at one

training set. This method is mathematically written as [100]

$$w^{k+1} = w^k - \eta \, \nabla L(w^k), \tag{3.11}$$

where $\nabla L(w^k)$ is the derivative of the loss function evaluated at w^k, and η is the learning rate parameter.

This approach is prone to overfitting. Deep learning models with high capacity can simply memorize the training dataset. In principle, overfitting could be prevented by training the DNN with a dataset bigger than the model's capacity. Unfortunately, it can be expensive and is not always possible in practice. It is also computationally expensive because it requires evaluating the model for every example in the entire dataset. Alternatively, the gradient descent can be computed by randomly sampling a number of small sample subsets from the dataset and taking the average loss from those subsets. This optimization method is known as the minibatch Stochastic Gradient Descent (SGD) method [100].

The learning rate is a very important parameter for the SGD algorithm. It is difficult to set the optimal value. On the one hand, if the learning rate is too small, then the algorithm will take longer to converge. On the other hand, if the learning rate is too high, the algorithm will jump across the optimum loss value. Therefore, most optimization algorithms have been developed to individually adapt the learning rates for different parameters. For example, the AdaDelta is an adaptive learning rate method that adapts learning rates based on a moving window of gradient updates [101]. This method later improved by the Adam[2] optimizer, which adapts the learning rates from the estimation of the first and second moments of the gradients [102].

3.2.3   Backpropagation

A deep neural network learns its weights and biases using optimizers such as SGD, as explained in the previous subsection. In the traditional optimization with feed-forward process, one must calculate the gradient of the loss function for each individual weight, which requires a lot of computational resources and effort, especially in a deep neural network with deep layers. Therefore, it is important to have a fast algorithm for computing such gradients.

A well-known algorithm for obtaining fast gradient computation is the backpropagation algorithm. Its strength was demonstrated in 1986, in a study by Rumelhart et al. [103]. The team found that backpropagation learning works faster than feed-forward learning. Today, the backpropagation algorithm is considered a fundamental building block of DNN. The idea of this algorithm is that the deep neural network is considered as a compositional function. Therefore, we could intuitively use the chain rule of derivative ($\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x}$) to propagate the partial derivative from the lost function, output layer, hidden layer, and the input layer, respectively. This derivative process is called backpropagation. An illustration of this gradient computation is shown in Figure 3.4.

---

[2]Abbreviation for adaptive moment estimation.

Forward Mode
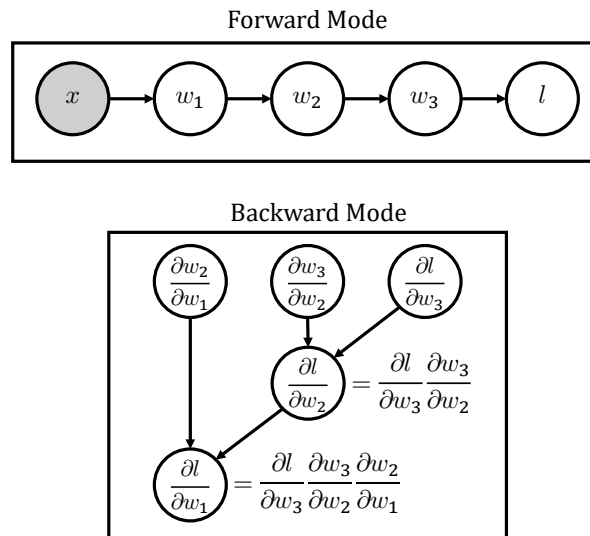
Backward Mode

Figure 3.4: Comparison between the feed-forward mode (top) and the backpropagation mode (bottom) to compute the gradient. In the forward mode, the weight gradient must be computed for every individual node, while keeping the other weights constant. In the backward mode, the weight gradient can be reused from the higher node to the lower node at one pass. Thus, faster computation can be achieved.

The backpropagation algorithm is faster than forward propagation. One forward pass computes the gradient of all weights and outputs with respect to one parameter (weight). Thus, when computing the full gradient, the computation amounts to the number of all weights in the network. In contrast, one backward differentiation pass computes the derivative of one output with respect to all parameters. When computing the full gradient thusly, the computation only amounts to the number of neurons in the output layer. Since the quantity of outputs is usually much smaller than the number of trainable parameters in a DNN model, the backward mode is faster and has become the standard algorithm for fast gradient computation.

3.2.4 Convolutional Neural Network

Just as the neural structure of the brain inspired the development of ANN, studies on how our brain processes visual information from the eye[3] led to the development of the Convolutional Neural Network (CNN). CNN is a specialized form of deep neural network for analyzing spatial structure. It has been used primarily to solve problems of computer vision, such as self-driving cars, robotics, and medical diagnoses. CNN has been designed specifically to recognize two-dimensional shapes with a high degree of invariance to translation, scaling, skewing, and other forms of distortion.

Technically, CNN depends on two main operational layers, the convolution layer and

---

[3]The inspiration came from a cat's visual cortex [104], which has small regions of cells that are sensitive to specific regions in the visual field.

the pooling layer. The convolutional operation in the convolutional layer is defined as the integral operation of two functions such that the first function acts as a filter (denoted as the kernel) to the second function. The convolution output is usually interpreted as a modified (filtered) version of the second function. For discrete signals, the integral operation becomes a summation which is, in general, defined as

$$y(i,j) = I(i,j) * K(m,n) = \sum_m \sum_n I(i-m, j-n) \, K(m,n), \tag{3.12}$$

with I representing the input image to be convolved with a kernel matrix ,K. Here, indices i and j are within the image matrix dimension, while those of m and n deal with the kernel matrix dimension. The operation is illustrated in Figure 3.5. In the form of a DNN, the connecting weights are stored in the kernel; thus, a convolutional layer has more weight capacity compared to a fully connected layer with the same height [105].

Convolution operation extracts feature maps or activation maps that operate as filters for spatial pattern detection, for example, vertical and horizontal edge detectors. After applying the convolution operation, the pooling layer replaces the convolution output with a statistical value over the local region (neighborhood). For example, the max pooling layer summarizes the output as the maximum within a rectangular neighborhood. Convolutional layers followed by pooling layers are applied to reduce the spatial size (width and height) of the data represented and enables feature extraction with higher-level representation [105]. Typically, the last layer of the CNN is connected to the conventional fully connected layer that encodes the extracted feature map to the output (vector) layer, as illustrated in Figure 3.6.

The CNN is one of the most successful and frequently used deep neural network model us

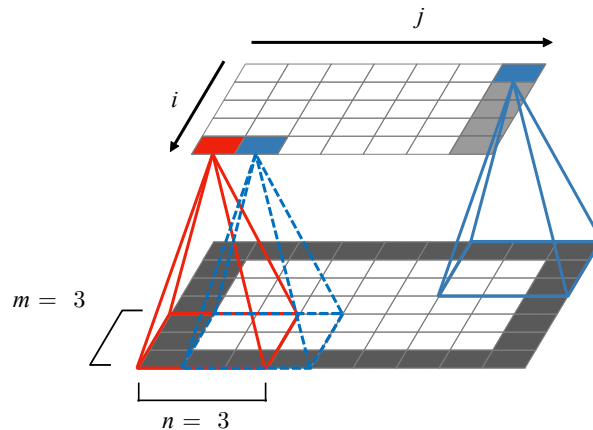*Figure 3.5: Illustration of the convolution operation. The 3x3 kernel matrix (dashed blue) slides over the input image (bottom) starting from the red kernel (red pyramid) location to the blue kernel (blue pyramid) location and reconstructs the output image (top). Note that the dark gray region at the input image is the out-of-border region which is usually filled with zero value (zero padding).*
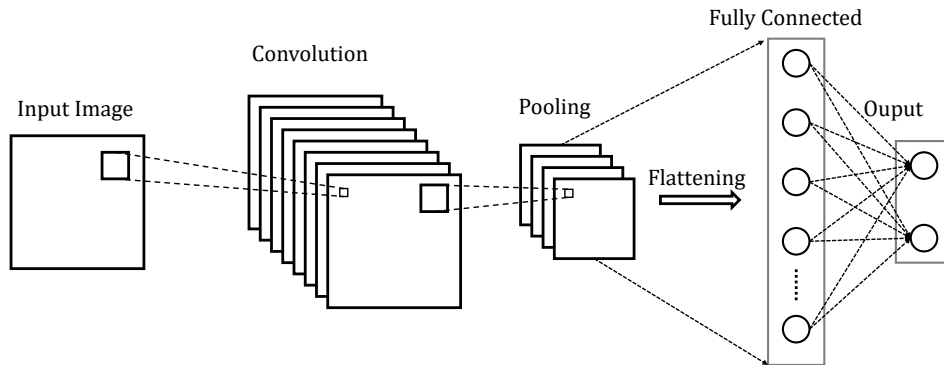
Figure 3.6: *Typical architecture of CNN. The convolutional layer is paired with a pooling layer to extract higher feature representation. After pooling, the extracted feature is connected to a fully connected layer that encodes the map to the output (vector) layer. The flattening step transforms the 2-dimensional feature map into a 1-dimensional weight vector. The image is redrawn from [106]*

.

was LeNet-5, used by banks to recognize handwritten digits [106]. It was followed by AlexNet, which won the ImageNet contests in 2012 [107]. Over time, many advanced CNN architectures have since been developed. One example of a modern CNN model is Residual Neural Network (ResNet) [108], which improves CNN performance by adding a skipping connection to transfer the identity function to a deeper layer. This concept was later extended to a more advanced network, Densely Connected Convolutional Networks (DenseNet), which applied a skipping connection layer to almost all layers in the CNN [109].

3.3 Real-Time Deep Learning Based Feedback System

The basic idea behind the feedback system proposed in this project is to use OCT images to monitor tissue type at the subsurface level and to control the laser automatically. This section discusses the general steps and components needed to realize the automatic feedback system. The steps are divided into three main subprocesses, illustrated in a schematic diagram in Figure 3.7. The complete implementation of these processes is discussed in Chapter 6.

The first subprocess is acquiring and enhancing (denoising) the image. Monitoring laser ablation in real-time requires an OCT system with fast acquisition time. OCT technology at the time of writing this thesis suggests the use of SS-OCT for a higher acquisition rate and deeper imaging range compared to other OCT options (see Section 2.3). The acquisition step is followed by an image denoising step to improve the SNR of the acquired image, which will further improve the accuracy of the feedback system. The second subprocess is tracking the ablation crater. A region of interest (image patch) from the OCT image is used as input for a DNN to classify tissue type at the ablation
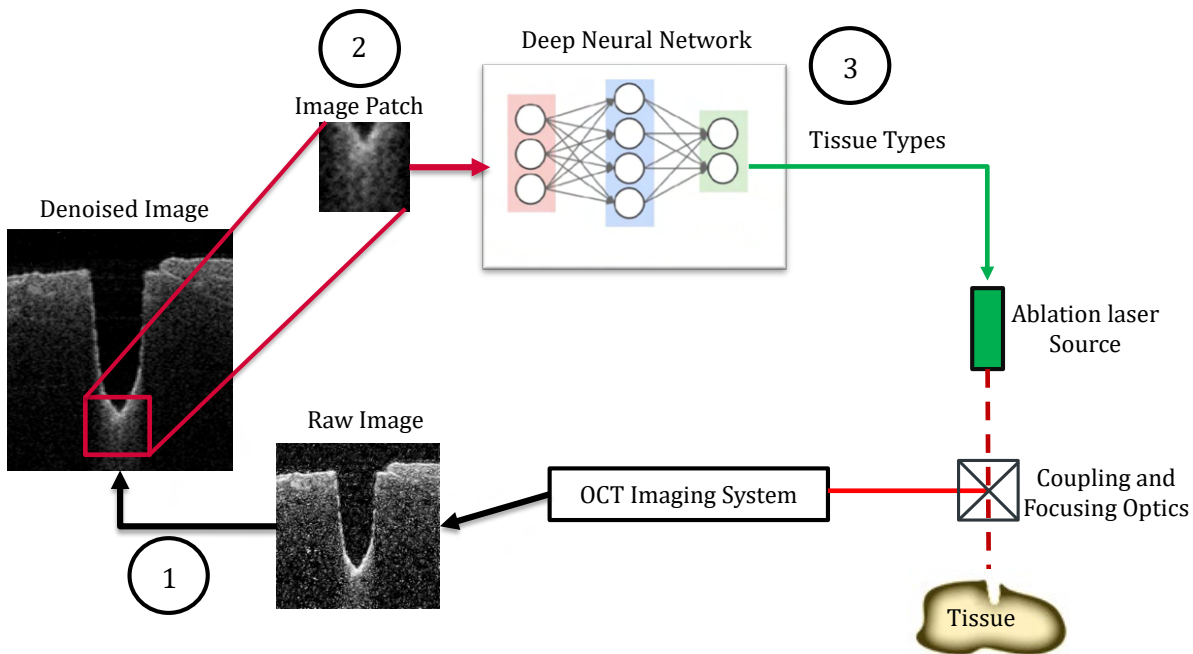
Figure 3.7: Schematic diagram of the proposed smart laser surgery system. Step 1: OCT images are acquired and enhanced during laser ablation. Step 2: A region of interest (image patch) from the OCT image is used as input for the deep neural network to classify tissue type at the ablation position. Step 3: The deep neural network output provides feedback for the ablation laser to either stop or continue ablating.

location. The third subprocess is classifying tissue type. A DNN model classifier predicts the current tissue type being ablated and provides feedback to the laser to either stop or continue ablating.

3.3.1 Er:YAG Ablation Laser

Lasers have been used extensively in medicine, especially high-power lasers for osteotomy. High-power lasers such as CO_2, Er:YAG, Er-Cr:YSGG, Nd:YAG, Nd:YLF, and Ti:Sapphire lasers are among those that have been clinically applied in dentistry and rapidly studied for application in laser osteotomy. Of the lasers mentioned, the Er:YAG laser was selected as the ideal laser for laser osteotomy, based on its optimal absorption by bone and water. Figure 3.8 shows the absorption coefficient of the Er:YAG laser and other medical lasers for hydroxyapatite materials (the main component of bone), water (H_2O), melanin, and hemoglobin. The plot shows that the Er:YAG laser has a high peak of absorption in both water and hydroxyapatite. It is important to note that the high absorption of Er:YAG laser has an indirect effect on ablation efficiency.

Based on the plot in Figure 3.8, CO_2 laser has a higher absorption coefficient than Er:YAG laser for hydroxyapatite material, which is higher than its absorption coefficient in water. In contrast, the absorption coefficient of water is higher than that of bone for
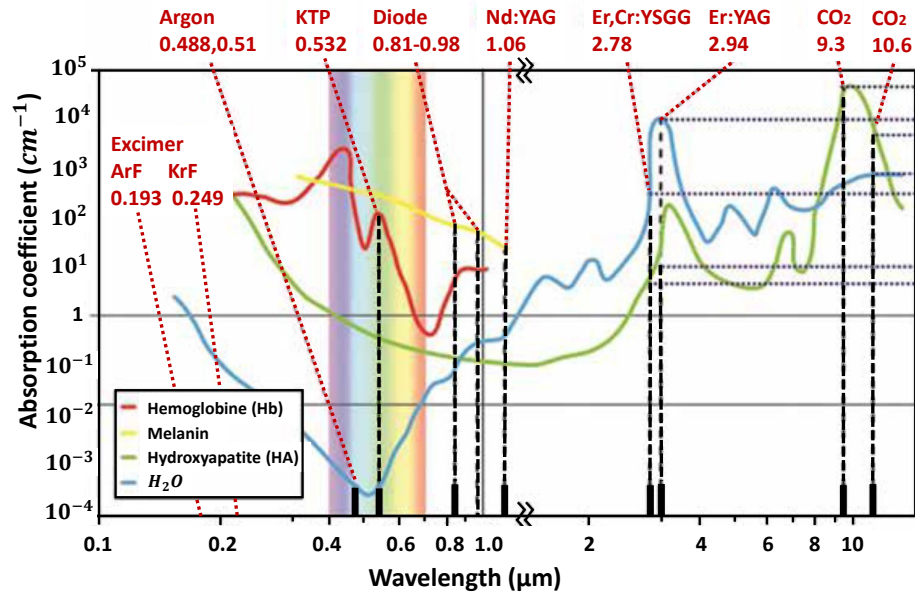
Figure 3.8: Absorption coefficients of several medical lasers and other existing lasers for hemoglobin, melanin, hydroxyapatite, and water (H_2O). Figure adapted and reprinted with permission from [110].

Er:YAG laser. The differences between the respective water and hydroxyapatite absorption coefficients affect ablation efficiency during laser osteotomy, since the ablation mechanism is mostly based on the tissue-water phase explosion. The water existent in tissue is an explosive material when high energy laser is deposited. The tremendous changes in temperature cause high pressure bubbles to build, which then explode, resulting in bone material ejection. Under water-cooled ablation conditions, Er:YAG laser light is mostly absorbed by the water, while the CO_2 laser light is mostly absorbed by the hydroxyapatite. Therefore, the probability of a tissue-water phase explosion is higher when using Er:YAG rather than CO_2 laser.

Studies of CO_2 laser use for bone cutting show heavy carbonization at the margin of the cut and, consequently, took a longer healing time compared to osteotomy using conventional drills or saws [14,15]. Additionally, carbonization and excessive debris obstruct the view of the ablation site. Conversely, the Er:YAG laser has relatively higher bone ablation rates per pulse with little or no carbonization on the adjacent ablation margins, making it an optimal system for bone cutting [111]. A previous study found that a bone ablation rate of up to 0.1 mm/pulse, could be easily achieved with Er:YAG laser [112]. The thermal damage to the tissue is low, so wound healing time is comparable to that of conventional water-cooled osteotomies performed with mechanical saws or drills [13–15]. Therefore, Er:YAG laser was chosen for the investigation into developing a reliable laser osteotomy set up with a feedback system to detect the tissue to be cut.

3.3.2 Image Denoising

Image denoising is one of the tasks discussed in Chapter 2 Section 2.6 and is essential to reducing the speckle in OCT images. The gold standard method has been spatial compounding, which works by sampling numerous images at a single sampling point and calculating the mean value for each pixel (frame-averaging). This method improves the SNR of the image by a factor of $\sqrt{N}$. However, it requires a longer acquisition time. In this thesis project, a denoising autoencoder neural network, one of the deep learning models, was used to enhance OCT image quality. The performance of the model is explained in Chapter 4, where the deep learning model mimics the spatial compounding method and improves the classifier algorithm's accuracy.

3.3.3 Ablation Tracker

Ablation crater tracking is a critical task in this project because the tissue classifier system needs to focus on the area below the ablation crater. In the application, it was assumed that the laser ablation spot was fixed at the center of the OCT image. Edge detection of the OCT image is required to define the tissue surface and to locate the patch image position over depth. In Chapter 5 and Chapter 6, the Canny edge method [113] is shown to be sufficient for detecting the surface of bone tissue in the OCT image.

3.3.4 Challenges of Real-Time Tissue Characterization During Laser Ablation

Tissue classification or characterization is an essential part of the proposed feedback system. Implementation of this thesis project specifically focused on the use of deep learning methods. However, like other machine-learning algorithms, the deep learning method is highly dependent on the datasets used for training. The quality and quantity of the training data determine the accuracy and performance of a machine-learning model. Since photothermal ablation is the primary ablation mechanism for microsecond Er:YAG lasers, increases in temperature during laser ablation induce changes in the optical properties of the tissue and indirectly lead to textural deformation. Therefore, it is important to train the deep learning model with training datasets collected while ablating the tissue. An implementation of this step is discussed in Chapter 6.

Chapter 4

OCT Image Denoising with Deep Learning

Denoising is one of the major steps in most medical image analyses. Especially in OCT, it refers to the suppression of speckle noise and preservation of carrier speckles. In this thesis, improvement in image quality was expected to improve the accuracy of an image-based tissue classifier. This chapter presents two publications that proposed the use of deep learning for denoising OCT images.

4.1 Image Denoising with A Deep Learning Model for Bone Images

The first publication demonstrated our experiment with using a deep learning model to resemble the frame-averaging method. Specifically, we were implementing a U-Net model, inspired by Ronneberger et al., as an autoencoder to denoise the biomedical images [114]. As a preliminary experiment, denoising was done for OCT images of bone tissue only. The results show that the images denoised with U-Net have higher SNR, CNR, and sharpness than those denoised with classical filters, BM3D and DD-CDWT [115, 116]. Additionally, the processing time was faster with U-Net than with the classical denoising filters.

Publication : Y. A. Bayhaqi, G. Rauter, A. A. Navarini, P. C. Cattin, and A. Zam, "Fast optical coherence tomography image enhancement using deep learning for smart laser surgery: preliminary study in bone tissue", *Proc. SPIE*, Vol. 11207, in *Fourth International Conference on Applications of Optics and Photonics*, 2019.

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Fast optical coherence tomography image enhancement using deep learning for smart laser surgery: preliminary study in bone tissue

Bayhaqi, Yakub A., Router, Georg, Navarini, Alexander, Cattin, Philippe C., Zam, Azhar

Fast Optical Coherence Tomography Image Enhancement using Deep Learning for Smart Laser Surgery: Preliminary Study in Bone Tissue

Yakub A. Bayhaqi[1a], Georg Rauter[b], Alexander Navarini[c], Philippe C. Cattin[d], and Azhar Zam[2a]

[a]Biomedical Laser and Optics Group (BLOG), Dept. of Biomedical Engineering, University of Basel, Gewerbestrasse 14, 4123 Allschwil, Switzerland
[b]Bio-Inspired RObots for MEDicine-Lab (BIROMED-Lab), Dept. of Biomedical Engineering, University of Basel, Gewerbestrasse 14, 4123 Allschwil, Switzerland
[c]Dermatology, Dept. of Biomedical Engineering, University of Basel, Hebelstrasse 20, 4031 Basel, Switzerland
[d]Center for medical Image Analysis and Navigation (CIAN), Dept. of Biomedical Engineering, University of Basel, Gewerbestrasse 14, 4123 Allschwil, Switzerland

ABSTRACT

One of the most common image denoising technique used in Optical Coherence Tomography (OCT) is the frame averaging method. Inherent to this method is that the more images are used, the better the resulting image. This approach comes, however, at the price of increased acquisition time and introduced sensitivity to motion artifacts. To overcome these limitations, we proposed an artificial neural network architecture able to imitate an averaging method using only a single image frame. The reconstructed image has an improvement in the signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) parameters compared to the original image. Additionally, we also observed an improvement in the sharpness of the denoised images. This result shows the possibility to use this method as a pre-processing step for real-time tissue classification in smart laser surgery especially in bone surgery.

Keywords: Neural Network, Optical Coherence Tomography, Denoising, Frame Averaging

1. INTRODUCTION

Optical coherence tomography (OCT) is rapidly becoming the preferred method for in vivo investigation of thin tissues or subsurface imaging. This modality is a clinically viable alternative for real-time high resolution in situ investigation of tissue structures [1]. In laser surgery, where critical tissues such as nerves and blood vessels need to be avoided from being cut, OCT is useful to monitor the tissue anatomy at the subsurface level before each ablation pulse is applied. The basic idea of the smart laser surgery system proposed in this study is to use OCT images to monitor tissue anatomy at the subsurface level during laser surgery. A trained artificial neural network able to differentiate tissues based on the OCT images could be used to determine if the ablating laser should stop or continue ablating [2]. However, in the process, OCT image enhancement is necessary to minimize tissue classification errors. Figure 1 shows the schematic diagram of the laser surgery system whereas we focus in this study on Step 1.

Speckle is one of the causes of the grainy appearance of OCT images and is dependent on both the wavelength of the imaging beam and structural characteristics of the tissue [3]. Similar to ultrasound, as the signal carrier, speckle in OCT also contains information about the textural characteristics and clinical features of the tissues. However, random interference introduces interference artifacts called *speckle noise* that degrades the image quality and can mask the texture features of the tissues carried by the signal carrying speckle [3]. The sensitivity of the interferometer to the phase of a light makes OCT susceptible to *speckle noise*. Furthermore, another source of noise in OCT is the missing frequency problem. Due to the nature of the coherent detection method, OCT scanners can only detect those objects whose spatial frequency spectra (v) overlap with the band of spatial frequencies between $[2v - 1/l_c, 2v + 1/l_c]$, where l_c is the coherence length

---

[1]yakub.bayhaqi@unibas.ch; phone +41 (0)61 207 54 61; dbe.unibas.ch
[2]azhar.zam@unibas.ch; phone +41 (0)61 207 54 60; dbe.unibas.ch

of the source. As a result, the interior of structures with smoothly varying refractive-index profiles are absent in OCT images [3]. Separation of these signal carrying speckle from the signal degrading speckle (*speckle noise*) and the missing frequency problem are ongoing challenge in OCT. Multiple approaches have been suggested to address these challenges in the literature. The approaches can be roughly classified into frame averaging methods, and digital suppression methods.
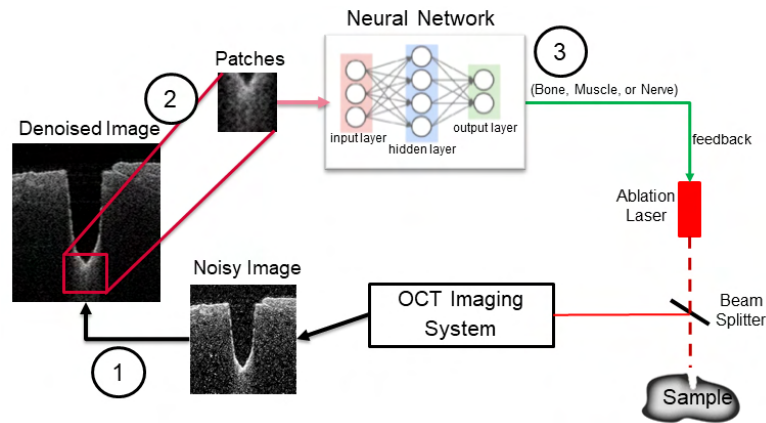
Figure 1. Schematic diagram of the proposed smart laser surgery system. Step 1: OCT images are acquired and enhanced during laser ablation. Step 2: A region of interest (image patch) from the OCT image is used as input for the neural network to classify the tissue type at the ablation position. Step 3: The Neural network output provides feedback for the ablation laser to either stop or continue ablating.

The frame averaging method operates based on registrations of the images and elimination of the randomness by averaging between repeated acquisitions [4]. However, this method is susceptible to motion artifacts if the registration of movement is not well solved. Additionally, the performance of the method depends on the acquisition speed and number of repetitions hence may increase the acquisition time. Another method to reduce noise in OCT is using digital suppression such as wavelet transform filtering and spatial domain filtering approaches which can be applied directly to a single frame image [5, 6]. However, these methods are often computationally complex and often remove small structural features from the image which leads to lower image quality as compared to the frame averaging method.

In this work, we present an implementation of a deep learning based artificial neural network (ANN) model to reduce noise in OCT images. In image pattern recognition and computer vision, ANNs are already known for their ability to classify, retrieve, detect and segment images. ANNs are also used to correct or denoise images which is useful in most medical imaging applications as it could lead to a better diagnostic assessment. In ultrasound imaging, ANNs were demonstrated to solve the problem of recovering ultrasound signals from undersampled measurements by utilizing stacked denoising autoencoders [7]. Furthermore, ANNs are also studied to enhance low dose CT Scan images which leads to a solution for reducing the risk of X-ray radiation [8]. The success of such methods arises from their ability to exploit the spatial correlation at multiple resolutions using a hierarchical network structure.

In OCT imaging, recently, an implementation of a noise suppression algorithm using deep learning was proposed by Halupka *et al.* in 2018 by utilizing a convolutional neural network [9]. In 2019, Laves *et al.* also proposed another review on this problem which employed a deep convolutional autoencoder with a prior trained ResNet image classifier as regularizer [10]. It was shown that regularized autoencoders are capable of denoising retinal OCT images without blurring details of diseases. For smart laser surgery, where short and predictable computation time is required, ANNs would be the way to go. In this study, we proposed to use the U-Net model as an extension to deep convolutional autoencoders for our denoiser function which has been recently reviewed for its ability to enhance medical images [11].

2. MATERIAL AND METHOD

2.1 Data acquisition and OCT system

The OCT system used was a custom developed Axsun swept source OCT System which has a center wavelength (λ_0) of 1060 nm and a sweep rate of 100 kHz. The image size acquired from this OCT system has a size of 300 pixels (7.5 mm) in

width and 1024 pixels (1.6 mm) in depth. The output image format from the OCT system is 32-bit grayscale TIFF images. This system was used to obtain a subsurface image from a porcine femur bone sample. OCT B-Scan images were taken in 120 different locations. In each acquisition location, 300 frames were acquired and averaged to reduce the noise of the OCT (frame averaging image), while ten images were randomly selected out of these frames (noisy images). Overall, a total of 1200 images were acquired and splitted into 480 training, 120 validation, and 600 testing datasets. Figure 2 shows the custom OCT system, the bone image sample and a representative image sample taken with this system.

Figure 2. The custom developed Axsun swept source OCT system (left) used in this work, a porcine femur bone, sample (middle) and the raw OCT image example of the bone (right).
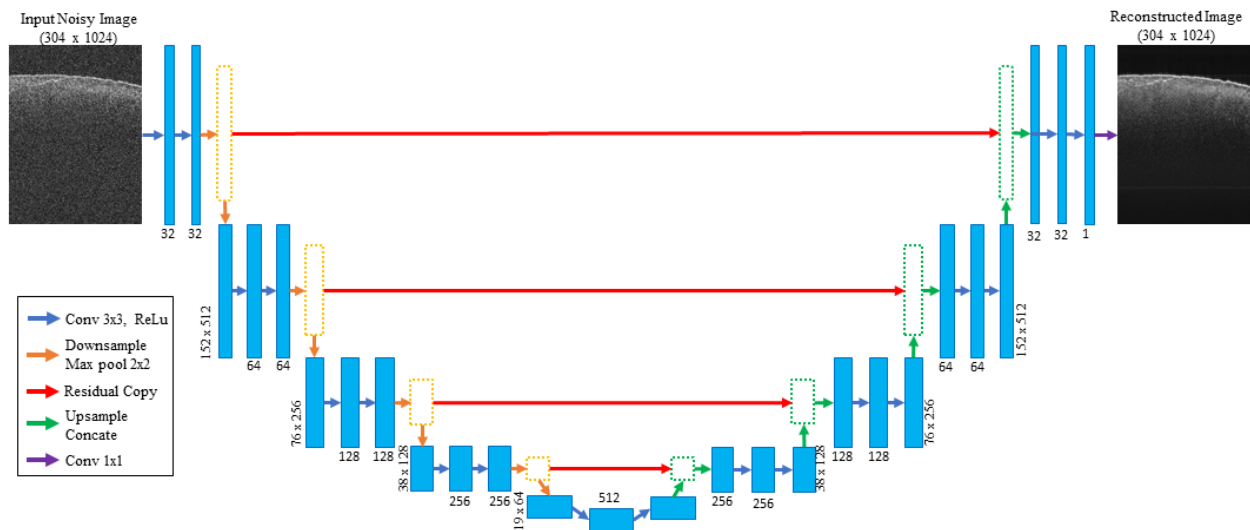
Figure 3. The U-Net model proposed in this work. It consisted of the encoding path (left side) and decoding path (right side).

2.2 U-Net Architecture

The objective of this study is implementing a U-Net convolutional neural network model as an autoencoder to denoise OCT images. The main structure of the model proposed was an implementation of the neural network inspired by Ronneberger *et al.* in 2015 [11]. However, changes to the U-Net input size were done mainly to fit our OCT image size. The encoding path (left side) and decoding path (right side) were adapted accordingly. The encoding path was followed by a typical architecture of a convolutional neural network. It consists of a threefold repetition of two convolutional layers with small receptive 3x3 pixels filters. Figure 3 illustrates the network architecture. Each repetition followed a 2x2 max pooling layer with a stride of 2 for downsampling, and we doubled the number of features after each downsampling step. As the opposite, the decoding path consisted the same repetition of upsampling layers of the feature map followed by two convolutional layers and half the number of feature channels, a concatenation layer with the correspondingly cropped

feature map from the encoding path, and two convolutional layers with small receptive 3x3 pixel filters. We equipped all hidden layers with the rectification nonlinearity (ReLU) activation function. The input and output size of the network were adapted such that the max-pooling operations are applied with an even x and y size. Therefore, the OCT images were resized to 304x1024 pixels using replicate padding method which extend the image by repeating border elements of array. Furthermore, to prevent overfitting, we put a dropout layer (ratio of 0.1) after each convolutional layer. For the final layer, we used a 1x1 convolutional layer to map the pixel value as similar to the reference pixel.

2.3 Training details

We implemented our approach in Keras-GPU environment, using the TensorFlow backend for neural network constructing and training. The noisy images were the input for the U-Net, while the averaged images were used as a label image. We ran the training on a workstation equipped with an Intel Xeon E5620 and two NVIDIA GEFORCE GTX 1080Ti GPUs, which enabled us to perform parallel computation to speed up the computation time. We trained the U-Net using 4000 epochs. The training was done in mini-batches, with a batch size of 4 samples to fit our GPUs memory capacity. For the loss function, we defined the mean squared error function as already been proposed by Halupka *et al.* in 2018 [9]. The loss function implicitly tells the similarity between the reconstructed image and the label image. We also defined Adam (learning rate = 1.0×10^{-4}) as the training optimizer. It took approximately 96 hours (4 days) to train the network.

2.4 Evaluation and comparison

We compared the U-Net autoencoder with two analytical denoising methods, block-matching 3D (BM3D) and double-density dual-tree complex wavelet transform (DD-CDWT) [5, 6]. Both denoising methods were implemented using MATLAB 2017b. We compared all methods in term of image quality and computation time. We used standard image quality metrics, such as signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), sharpness, and mean squared error, to evaluate their performance. The sharpness parameter was proposed by Sattar *et al.* [12] as a measure of edge preservation between the enhanced image and the correspond frame averaging image which defined as

$$\beta = \frac{\Gamma(\Delta s - \overline{\Delta s}, \widehat{\Delta s} - \overline{\widehat{\Delta s}})}{\sqrt{\Gamma(\Delta s - \overline{\Delta s}, \Delta s - \overline{\Delta s}) . \Gamma\left(\widehat{\Delta s} - \overline{\widehat{\Delta s}}, \widehat{\Delta s} - \overline{\widehat{\Delta s}}\right)}}$$

where $\Delta s\,(i,j)$ is the filtered version of $s(i,j)$, obtained with a 3x3 pixels approximation of the Laplacian operator. Furthermore, the mean squared error metric (MSE) was used to measure the similarity between the enhanced image and the correspond frame averaging image. To fairly evaluate between the different denoising methods, we tested all of the methods on a personal computer with 2.8 GHz Intel Core i7 processor, 16 GB 1867 MHz DDR3 memory and a GPU NVIDIA GTX 1050Ti.

3. RESULT AND DISCUSSION

Representative images of each enhancement technique within the test set are shown in Figure 4. For display purposes, the images were resized while the aspect ratio was maintained during processing. Furthermore, the image quality comparison is shown in Table 1. We observed that the BM3D and DD-CDWT were out-performed by the U-Net autoencoder to denoise OCT images. The average SNR and CNR were 12.92 dB and 4.72 dB respectively for the raw images; 16.64 dB and 6.34 dB for the 300-fold averaged images; 15.5 dB and 6.4 dB for the BM3D; 15.11 dB and 6.34 dB for the DD-CDWT; and 15.87 dB and 7.31 dB for the U-Net filtered images. Additionally, we also observed an improvement in the sharpness of 0.54 for the frame averaging images to 0.82 for the U-Net reconstructed images. The frame averaging method tends to blend the image points with the surrounding area. This smears the edges and induces a loss of general detail [4]. Meanwhile, we applied dropout regularization in the U-Net to prevent overfitting and support the generalization capabilities for denoising the images instead [13]. Therefore, the U-Net was able to remove the speckle noise while maintaining the sharpness of single frame image. Other than that, we also measured the MSE of the reconstructed images. The BM3D and DD-CDWT method had higher MSE as compared to U-Net. This is because the U-Net was trained to modify the pixel as close to the reference (frame-averaging image), while the standard methods are based on an analytical approach.

The averaged computation time required for all the denoising techniques is shown in Table 2. The performance time was a time average of all the 600 test set images. The result shows that the U-Net was faster since it was run in parallel on the GPU while the other existing methods were run on the CPU. This result shows the possibility of using U-Net to denoise OCT images in real-time and increase tissue type classification accuracy for our smart laser surgery system. However, the number of OCT images used in this study was limited, and further studies are needed to train and test the neural networks using thousands of OCT images and more images from different tissue types to be denoised.

| (a) | (b) | (c) | (d) |
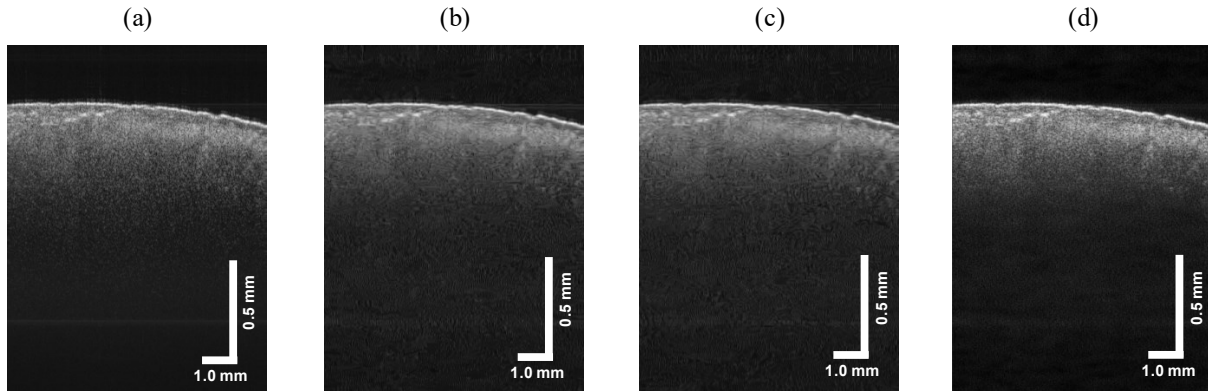
Figure 4. Image comparison between (a) Frame Average method, (b) BM3D, (c) DD-CDWT, and (d) U-Net.

Table 1. Image quality comparison between the standard techniques and the proposed U-Net network.

Denoising Method	Average SNR (dB)	Average CNR (dB)	Average Sharpness	Average MSE (pixel value)
Raw	12.92	4.72	-	-
Frame Averaged-300	16.64	6.33	0.54	-
BM3D	15.50	6.4	0.72	1.23
DD-CDWT	15.11	6.34	0.68	1.36
U-NET	15.87	7.32	0.82	1.04

Table 2. Denoising performance comparison between all method.

Denoising Method	Time (seconds)
Frame Averaged-300	2.730
BM3D	2.350
DD-CDWT	0.860
U-NET	0.068

4. CONCLUSION AND FUTURE WORK

The study demonstrates the ability of the U-Net to mimic the frame averaging method in denoising OCT images. The output images from the U-Net autoencoder have a higher image quality than those from the standard analytical methods. The reconstructed image showed an improvement in the signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) parameters compared to the original image. The MSE result shows that the U-Net reconstructed images are similar to the

image denoised using the frame averaging method. Furthermore, the performance comparison shows that the proposed method can be used for real-time image denoising which is suitable for our smart laser surgery system. However, the authors believe that a further study is needed to investigate the performance of the U-Net autoencoder in denoising OCT images of other tissue types such as muscle, nerve, and bone marrow.

ACKNOWLEDGMENT

REFERENCES

[1] Brezinski, M.E., Tearney, G.J., Bouma, B.E., Izatt, J.A., Hee, M.R., Swanson, E.A., Southern, J.F. and Fujimoto, J.G., "Optical coherence tomography for optical biopsy. Properties and demonstration of vascular pathology," Circulation 93, 1206–1213 (1996).

[2] Bayhaqi, Y. A., Navarini, A., Rauter, G., Cattin, P. C., and Zam, A., "Neural network in tissue characterization of Optical Coherence Tomography (OCT) image for smart laser surgery: preliminary study," Proc. SPIE 11044, 1104402 (2019).

[3] Schmitt, J. M., Xiang, S. H., and Yung, K. M., "Speckle in optical coherence tomography," Journal of Biomedical Optics 4(1), (1999).

[4] Jørgensen, T. M., Thomadsen, J., Christensen, U., Soliman, W., and Sander, B., "Enhancing the signal-to-noise ratio in ophthalmic optical coherence tomography by image registration–method and clinical examples," J. Biomed. Opt. 12, 041208 (2012).

[5] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K., "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," IEEE Transactions on Image Processing, 16 (8), 2080-2095 (2007). (http://www.cs.tut.fi/~foi/GCF-BM3D)

[6] Selesnick, I. W., "The double-density dual-tree DWT," IEEE Transactions on Signal Processing, 52(5), 1304-1314 (2004). (http://eeweb.poly.edu/iselesni/WaveletSoftware/dt2D)

[7] Perdios, D., Besson, A., Arditi, M., and Thiran, J.P., "A deep learning approach to ultrasound image recovery," IEEE International Ultrasonics Symposium (IUS), 1–4, (2017).

[8] Gondara, L., "Medical image denoising using convolutional denoising autoencoders," IEEE 16th International Conference on Data Mining Workshops (ICDMW), 241–246, (2016).

[9] Halupka, K. J., Antony, B. J., Lee, M. H., Lucy, K. A., Rai, R. S., Ishikawa, H., Wollstein, G., Schuman, J. S., and Garnavi, R., "Retinal optical coherence tomography image enhancement via deep learning," Biomed. Opt. Express 9, 6205-6221 (2018).

[10] Laves, M-H., Sontje, I., Kahrs, L. A., Ortmaier, T., "Semantic denoising autoencoders for retinal optical coherence tomography," Accepted for publication at the SPIE/OSA European Conferences on Biomedical Optics (ECBO) (2019).

[11] Ronneberger O., Fischer P., Brox T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," Medical Image Computing and Computer-Assisted Intervention – MICCAI, 9351, 234-241 (2015).

[12] Sattar, F., Floreby, L., Salomonsson, G. and Lovstrom, B., "Image enhancement based on a nonlinear multiscale method," IEEE Transactions on Image Processing, 6(6), 888-895 (1997).

[13] Liang, J., Liu, R., "Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network," IEEE 8th International Congress on Image and Signal Processing (CISP), 697-701 (2015)

4.2 Image Denoising with Deep Learning for Five Tissues

After demonstrating the denoising performance for bone images, the experiment is extended to by denoising OCT images of marrow, fat, muscle, and skin. Two deep learning models, U-Net [114] and ResNet [108] models, were compared. The training of the models with supervised learning and semi-supervised learning were also compared, represented by the Generative Adversarial Network (GAN) [117]). The quantitative and qualitative measurements show that the deep learning models with supervised learning perform similarly to the frame-averaging method and better than the classical denoising filters. The results also prove that image denoising improves the accuracy of the tissue classifier.

Publication : Y. A. Bayhaqi, A. Hamidi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Deep-Learning-Based Fast Optical Coherence Tomography (OCT) Image Denoising for Smart Laser Osteotomy", *IEEE Transactions on Medical Imaging*, Vol. 41, No. 10, pp. 2615-2628, 2022.

Deep-Learning-Based Fast Optical Coherence Tomography (OCT) Image Denoising for Smart Laser Osteotomy

Yakub A. Bayhaqi, Arsham Hamidi, Ferda Canbaz, Alexander A. Navarini,
Philippe C. Cattin, *Member, IEEE*, and Azhar Zam, *Member, IEEE*

Abstract—**Laser osteotomy promises precise cutting and minor bone tissue damage. We proposed Optical Coherence Tomography (OCT) to monitor the ablation process toward our smart laser osteotomy approach. The OCT image is helpful to identify tissue type and provide feedback for the ablation laser to avoid critical tissues such as bone marrow and nerve. Furthermore, in the implementation, the tissue classifier's accuracy is dependent on the quality of the OCT image. Therefore, image denoising plays an important role in having an accurate feedback system. A common OCT image denoising technique is the frame-averaging method. Inherent to this method is the need for multiple images, i.e., the more images used, the better the resulting image quality. However, this approach comes at the price of increased acquisition time and sensitivity to motion artifacts. To overcome these limitations, we applied a deep-learning denoising method capable of imitating the frame-averaging method. The resulting image had a similar image quality to the frame-averaging and was better than the classical digital filtering methods. We also evaluated if this method affects the tissue classifier model's accuracy that will provide feedback to the ablation laser. We found that image denoising significantly increased the accuracy of the tissue classifier. Furthermore, we observed that the classifier trained using the deep learning denoised images achieved similar accuracy to the classifier trained using frame-averaged images. The results suggest the possibility of using the deep learning method as a pre-processing step for real-time tissue classification in smart laser osteotomy.**

Index Terms—**Deep learning, image denoising, image processing, optical coherence tomography.**

I. INTRODUCTION

LASER osteotomy offers many advantages over mechanical tools, such as a reduced risk of bacterial contamination (due to its contactless nature), less tissue loss, and high precision cutting [1]–[4]. In addition, the small focused cut of a laser osteotome enables surgeons to go beyond straight cuts and perform more complex cuts like circular, diamond, and dove-tail shapes [5]. Furthermore, the non-contact laser osteotome provides a possibility to introduce a feedback system to prevent cutting unwanted tissues or damaging critical tissues, such as bone marrow and nerve. Several such feedback systems have been developed, such as optical spectroscopy [6]–[8] and acoustic feedback induced by the laser ablation process [9]–[12]. However, these methods rely on signals emitted from the laser ablation process and permit some damage to the critical tissues. An alternative and ablation-free approach for monitoring the laser ablation process can be implemented by coupling the ablation laser with an Optical Coherence Tomography (OCT) imaging system.

OCT is an emerging technology that performs non-invasive cross-sectional tomography using light propagation properties in media and interference phenomena. This imaging technology is analogous to ultrasound imaging, except that it uses light instead of sound. The signal reconstruction is performed by measuring the magnitude and echo time delay of back-reflected or back-scattered light from internal micro-structures in the tissue. Thus, OCT is a viable alternative for real-time, high-resolution, and in-situ investigations of thin tissue structures [13], [14].

Critical tissues such as bone marrow and nerve must be avoided in laser osteotomy. Therefore, OCT could help to monitor tissue anatomy at the subsurface level during the laser ablation process. Fig. 1 presents a schematic diagram of our proposed system. This process consists of three main subprocesses. The first subprocess is the acquisition and denoising of the images. Next, the second subprocess is tracking the ablation crater. The tracked region of interest (image patch) from the OCT image will be used as an input
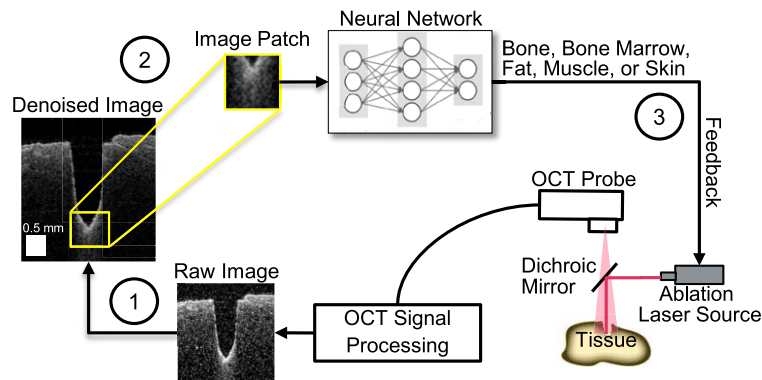
Fig. 1. Schematic of the proposed smart laser osteotomy system. Step 1: OCT images of tissue are acquired (light red) and denoised during laser ablation (dark red line). Step 2: A region of interest or image patch (yellow box) from the OCT image is used as input for the artificial neural network to classify the tissue types at the ablation position. Step 3: An artificial neural network provides feedback for the ablation laser to either stop or continue ablating based on the patch image.

for a classifier to classify tissue type at the ablation position. The third and final subprocess is the classification process. A deep-learning artificial neural network (ANN) is proposed to predict the current tissue type during the ablation and provides feedback to the ablating laser to stop or continue ablating.

In the proposed schematic, the tissue classifier's accuracy is dependent on the quality of the OCT image that is used for training. Although the ANN could identify tissue type from the raw image directly, we believe that the ANN could identify tissue type from the denoised image with better accuracy. Furthermore, besides using an ANN as a tissue classifier, our approach also proposed to use an ANN for the image denoising process. This paper demonstrated the benefit of using ANN as an image denoiser compared to the classical denoising methods. We also investigated the effect of denoising the OCT image to improve the tissue classifier's accuracy.

II. SPECKLE AND IMAGE DENOISING IN OCT

Like most narrow-band detection systems such as radar and ultrasound, speckle is the fundamental OCT image source. Speckle is the cause of the reconstructed image's grainy appearance. It depends on the size and temporal coherence of the light source and the tissue's structural characteristics. The phenomenon was found as the result of random interference between mutually coherent reflected waves from multiple back- and forward-scattering [15]. Consequently, speckle plays a dual role as a source of noise (speckle noise) and as a carrier of information about the tissue microstructure (signal-carrying speckle). Speckle is considered as noise when destructive interference happens and reduces the correspondence between the local density of scatterers and the intensity variations. If all the reflected waves from the tissue could be forced to interfere constructively, the noise would vanish, and the image contrast would be significantly improved. The OCT image denoising in this paper aimed toward this ideal of speckle-noise reduction.

Separating the signal carrier speckle from speckle noise (image despeckle) is an ongoing challenge in OCT. Several approaches have been suggested to address this problem. A common technique for reducing speckle noise in OCT is frame-averaging, where absolute magnitudes of repeated signals from the same location are averaged to form a new noise-reduced signal [16]. However, this method is susceptible to motion artifacts if compensation of movement is not resolved. Additionally, the resulting image quality depends on the number of repetitions. The more images used, the better the resulting image quality, but this inherently leads to increased acquisition time.

Other than that, several classical digital filters were also suggested to reduce speckle noise in OCT. Sparse and wavelet transform filtering approaches, for example, can be applied directly to a single frame image [17]–[20]. However, these methods are often computationally complex and remove small structural features from the image, resulting in lower image quality than the frame-averaging method.

In this work, besides using an ANN to classify the tissue type, we also used an ANN to reduce speckle noise in the OCT image. ANN is already known for its ability to classify, retrieve, detect, and segment images in image pattern recognition and computer vision. ANN has also been used to correct or denoise images, making them useful for most medical imaging applications and potentially leading to better diagnostic assessments. In ultrasound imaging, ANN has been shown to solve recovering ultrasound signals from under-sampled measurements by utilizing stacked autoencoders [21]. ANN has also been used to enhance low-dose Computed Tomography images, which may offer a solution for reducing X-ray radiation [22]. These endeavors' success arises from exploiting the spatial correlation at multiple resolutions, using a hierarchical network structure.

OCT image denoising using ANNs has been proposed and implemented mostly based on the convolutional neural network (CNN) models [23]–[25]. The resulting images had similar image quality to those denoised with the frame-averaging method. The CNN could denoise retinal OCT images from a single image without blurring the retinal tissue structure's details, which reduces acquisition time. However, these references are only from the field of ophthalmology with their retinal OCT images. Here, the CNN model has to learn to denoise an image while retaining (memorize) the retinal tissue structure rather than just a general noise reduction. The CNN model might fail to denoise images in different OCT domains (e.g., intravascular, dental, or dermatology OCT images).
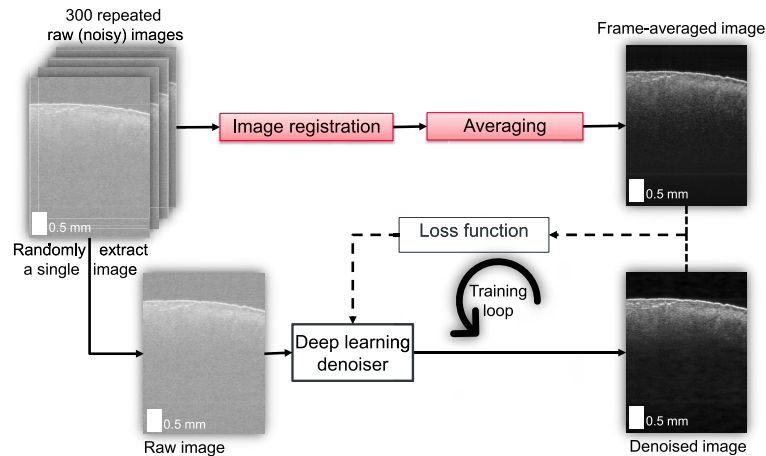
Fig. 2. Illustration of the training process using the deep learning method to denoise raw OCT images. The frame-averaged (reference) image is reconstructed by registering and averaging 300 repeated raw images of the same position on the sample. The deep learning denoiser is trained to modify the raw images to make them as similar as possible to the frame-averaged image, based on the loss function.

In our application, we trained and tested the performance of the CNN method using several tissues with various textures so that the CNN model learns more generalization of the noise. We extended the ability of the CNN to increase OCT image quality for normal tissues, which are encountered during laser osteotomy. Additionally, this paper gives an overview of the CNN model's performance to improve OCT image quality compared to frame-averaging and some classical digital filter methods. We also outlined the effectiveness of the image denoiser to increase the accuracy of a subsequent CNN tissue classifier.

III. OCT Image Denoising Methods

This work aimed to train a CNN that takes raw (noisy) images as input and generates images with the same quality level as the corresponding frame-averaged images. To achieve this, the CNN was trained to minimize the defined loss function, e.g., the mean squared error (MSE) between the corresponding raw and frame-averaged images. Furthermore, since noise in OCT also applies in the temporal domain, we also trained the CNN to generalize the temporal noise by using training images that were extracted in different (random) temporal locations over the repeated frames in the same spatial acquisition location. These training steps are shown in Fig. 2. At the end of the experiment, we investigated the benefit of denoising the image to improve the tissue classifier's accuracy.

A. Frame-Averaging Method

The frame-averaging method is still one of the most effective ways to reduce speckle noise in OCT imaging. In this paper, we used the frame-averaged image as the reference image or label. The reference image is generated by registering and averaging repeated scans of the same location. Averaging of N images improves the SNR by a factor $\sqrt{N}$ [26]. Hence, the higher the number of images, N, the lower the noise level. Nevertheless, since the frame-averaging method is susceptible to motion artifact, image registration is necessary before averaging the frames.

B. Deep-Learning Models

We compared two CNN models for denoising OCT images: a UNet autoencoder model [27] and a residual network (ResNet) model [28]. Additionally, we also compared two different loss functions for both CNN models. First, we trained both UNet and ResNet models with the MSE loss function only. Then, we investigated the model's performance by combining MSE, perceptual, and Wasserstein-adversarial loss. This combination was previously suggested in [22], [23], since the MSE loss function alone may skip some embedded details in the reference image.

1) UNet-Based Autoencoder: Starting with the UNet autoencoder, we adopted the structure of the CNN reported by Ronneberger *et al.* in 2015 [27]. We changed the size of the UNet input to the size of the acquired OCT images. The encoding path (left side) and decoding path (right side) were adapted accordingly, as shown in Fig. 3. Each side consists of five folded convolutional blocks. There are two convolutional layers (kernel size of 3×3) for each block; the number of filters gradually increases from 32, to 64, 128, 256, and 512, respectively, for the encoding path and vice versa for the decoding path. A 2×2 max-pooling layer (stride of two) downsamples the features after each convolutional block in the encoding path, except for the last (deepest) block. On the other hand, a 2×2 upsampling2D layer was applied after each convolutional block in the decoding path. Each convolutional block in the encoding path forwards a residual feature (copy) to the corresponding convolutional block in the decoding path. We equipped all the convolutional layers with a rectified linear unit (ReLU) as the activation function. We added a dropout layer (ratio of 0.1) after each convolutional block to prevent overfitting [29]. The final layer was a 1×1 convolutional layer to reconstruct the decoded image with a similar size as the input and activated with the sigmoid function.

2) Residual Network: The architecture of the residual network (ResNet) was suggested by He *et al.* [28]. The model consisted of a pre-residual layer, ten residual connecting blocks with identical structures, and a post-residual layer (shown in Fig. 4). The pre-and post-residual layers are 2D
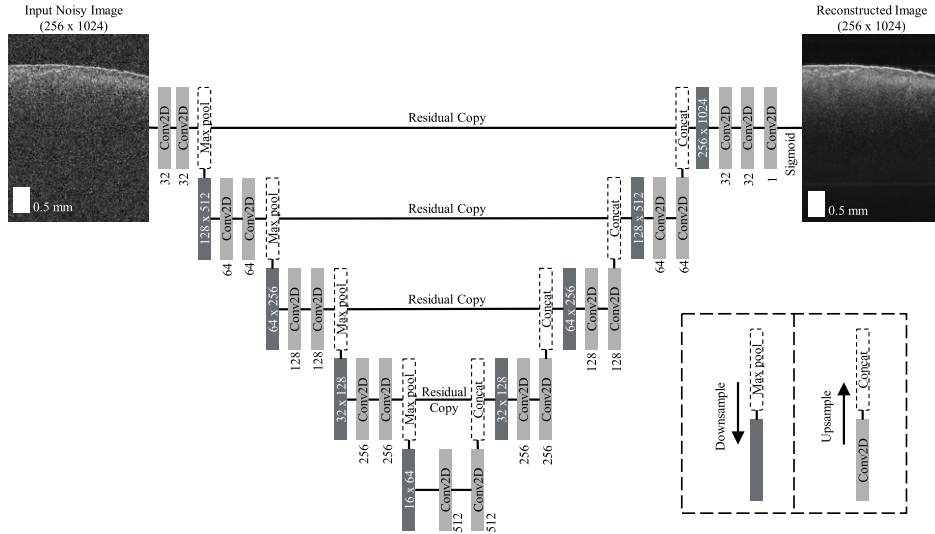
Fig. 3. The architecture of the UNet model. The main structure is separated between the downsampling encoder (left side) and the upsampling decoder (right side). Each side consists of four folded convolutional blocks. Each convolutional block in the upsampling concatenates a residual feature (copy) forwarded from the corresponding downsampling block. There are two convolutional layers for each block, the size of which depends on the depth of the factor of two, starting with 32 filters.
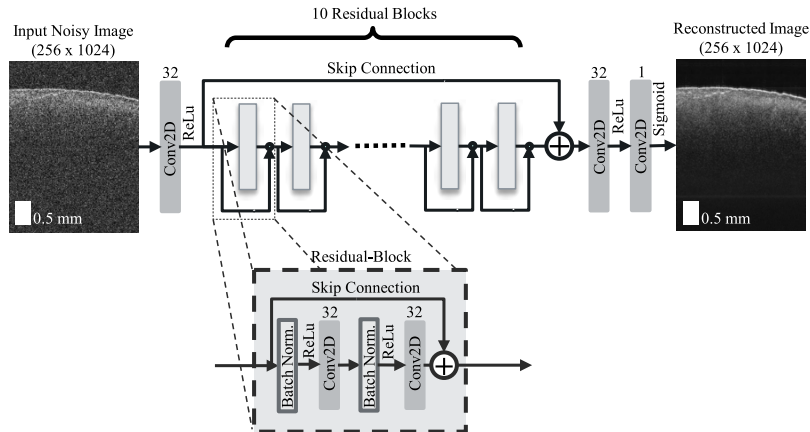
Fig. 4. The architecture of the ResNet model. It consists of 10 residual blocks placed between pre- and post- convolutional layers. Parallelly, a residual copy (skip) connects the pre- and post-layers. The residual block consists of two stacked convolutional layers. All the layers have 32 filters with a kernel size of 3 x 3 pixels.

convolutional layers. Each layer has 32 filters with 3 × 3 kernel size and is activated with the ReLU function. The residual blocks consisted of two stacked 2D convolutional layers of the same size as the pre-residual layer. Batch normalization is applied before each convolutional layer. A skip connection between blocks was introduced by He *et al.*. [28], which added a signal between the pre-and post-processing of a block (shown in Fig. 4). This identity mapping improves information flow through the network during feed-forward and back-propagation. Another skip connection was added between the signal before the residual blocks after processing through addition and followed by the post-residual layer. The final layer was the reconstruction layer, a 1 × 1 convolutional layer activated with a sigmoid function. The input size of the ResNet was set according to the size of the image in our datasets.

3) Mean Squared Error Loss: The most intuitive way of measuring the similarity between two images is by using the MSE. MSE measures the quadratic mean of the overall pixel difference between the corresponding reference image and the predicted image. This measurement is defined as:

$$MSE = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} (\hat{p}_{i,j} - p_{i,j})^2}{mn} \quad (1)$$

The MSE is the squared mean deviation of the pixel value (p) in the frame-averaged image and the pixel value ($\hat{p}$) in the denoised image at the i, j-th position with the same width m and height n. The value shows the general similarity per pixel between these images. The aim of training the UNet and ResNet is to minimize MSE to zero. However, studies show that the MSE loss may result in over-smoothed image [22].
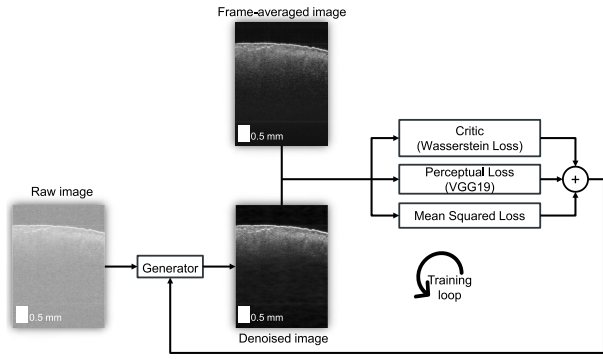
Fig. 5. Illustration of the adversarial and perceptual learning process. The training involves measuring Wasserstein, perceptual, and mean squared error loss of the denoised image produced by a generator. The critic is trained in a binary manner that tries to distinguish frame-averaged and denoised image [29]. Meanwhile, the perceptual loss is the measure of the relative perspective difference between the frame-averaged and denoised images, which is extracted from the high-level feature representation of a pre-trained VGG19 image-Net [30].

4) Adversarial and Perceptual Learning: First, we trained the models using the MSE loss function alone. We defined these models as UNet-MSE and ResNet-MSE, respectively. In this specific situation, the models were trained to generate pixel values similar to those of the frame-averaged image, based on the MSE measurements. Furthermore, Yang *et al.* [22] proposed to train the CNN models by introducing the Wasserstein loss (critics) [29] and perceptual loss [30]. The solution raised to tackle MSE loss problems which are associated with over-smoothed edges and loss of details. A MSE-based CNN overlooks subtle tissue texture in the image, which is critical for human perception. These additional losses have been demonstrated to improve the CNN for denoising images with better image quality and statistical properties than the MSE-based CNN. Furthermore, Halupka *et al.* [23] used the same method to denoise OCT retinal images. Our application of these losses in the CNN learning process is illustrated in Fig. 5. Therefore, we trained additional UNet and ResNet models with the Wasserstein and perceptual loss. We defined them as UNet-WGAN and ResNet-WGAN, respectively.

A generative adversarial network (GAN) consists of a discriminator (D) and a generator (G) network. During training, the discriminator learns to distinguish between the frame-averaged image and the image denoised with the generator. Simultaneously, the generator will try to generate a high-quality denoised image from a raw image that would be indistinguishable by the discriminator. The discriminator network architecture is illustrated in Fig. 6. On the other side, the generator network is the investigated UNet or ResNet.

We applied the improved version of the original GAN, which used the Wasserstein distance [29] as the discriminator loss function to criticize or score the performance of the generator. A Wasserstein GAN (WGAN) would have more stability during the training process compared to the original GAN. Gulrajani *et al.* also suggested that using the gradient penalty term to enforce the Lipschitz constraint would even
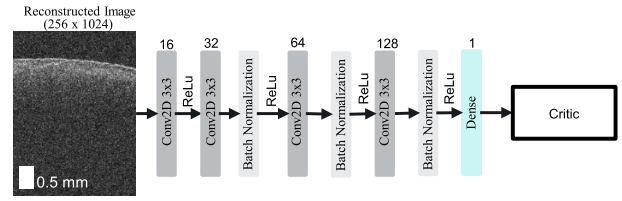
Fig. 6. The discriminator network for measuring the Wasserstein distance (critic) of the reconstructed image.

improve more training stability [31]. Both the discriminator and generator are trained with the min-max objective defined as:

$$\min_{G} \max_{D} L_{WGAN}(D, G) = \underset{\tilde{I}_{raw}}{\mathbb{E}} [D(\tilde{I}_{raw})] - \underset{I_{ref}}{\mathbb{E}} [D(I_{ref})] \\ + \lambda \underset{\hat{I}_{raw}}{\mathbb{E}} [(\|\nabla_{\hat{I}_{raw}} D(\hat{I}_{raw})\|_2 - 1)^2] \quad (2)$$

The Wasserstein distance is calculated between the denoised image ($\tilde{I}_{raw}$) and the frame-averaged image (I_{ref}). The denoised image is the reconstructed image by the generator from the raw image ($\tilde{I}_{raw} = G(I_{raw})$). The final term is the gradient penalty which enforces the Lipschitz constraint to have gradient norm ($\|\nabla_{\hat{I}_{raw}} D(\hat{I}_{raw})\|_2$) at most 1, where,

$$\hat{I}_{raw} = pI_{ref} + (1 - p)\tilde{I}_{raw}. \quad (3)$$

p is a uniform random number between 0 and 1 ($p \sim U[0, 1]$). The gradient penalty is weighted with a coefficient λ.

The perceptual loss is calculated from the high-level features extracted from a pre-trained VGG19 network [30]. The perceptual loss measures the similarity (MSE) of feature representations between the frame-averaged and the denoised images. This loss offers a more robust training approach because the feature extracted from the VGG19 network represents an external or alternate perspective, such as the content or style of the image. The perceptual loss function obliges the generator to denoise raw images with similar feature representations rather than requiring the pixels to match exactly the pixel of the frame-averaged image. The feature representation loss is the mean squared error (Euclidean distance) between the extracted features:

$$L_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\varphi_{i,j}(I_{ref})_{x,y} - \varphi_{i,j}(\hat{I}_{raw})_{x,y})^2 \quad (4)$$

where, $\varphi_{i,j}$ indicates the ReLU activated feature map obtained by the j-th convolutional layer before the j-th pooling layer within the VGG19 network. $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the respective feature maps of the corresponding layer. In the implementation, we used the fifth convolutional and pooling layers to measure the perceptual loss. Furthermore, since the pre-trained VGG19 network worked with 3-channels image, we converted the 1-channel OCT image to 3-channels image by repeating the first channel to the second and third channel. The adversarial network model minimized the combination of MSE, perceptual, and Wasserstein loss, and
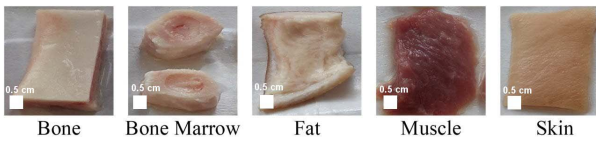
Fig. 7.   Examples of the tissue types used in the experiments.

established a ratio between them [22], [23]. Thus, the min-max objective of the WGAN training is given by:

$$Loss = \min_{G} \max_{D} \omega_1 L_{WGAN}(D, G) + \omega_2 L_{VGG(G)} + L_{MSE}$$
$$(5)$$

Here, ω_1 and ω_2 are the weights balancing the different contributions of the loss function. This loss will force the generator to match exactly the corresponding pixel value and force the output to be perceptually similar to the reference image.

IV. Experiment

A. Image Acquisition and Processing

A custom-made swept-source OCT system was used for this experiment. The system was equipped with an Axsun swept laser source with a center wavelength (λ_0) of 1060 nm and sweep rate of 100 kHz. The sensitivity (SNR_{max}) of our OCT system was 96.46 dB. The OCT B-scan image sizes are 1024 pixels (3.6 mm) deep and 256 pixels (4.1. mm) wide. The output image format is 16-bit grayscale with TIFF-formatted image.

The OCT image datasets for the denoising experiment were acquired from three different pigs, which were used for training, validation, and testing datasets, respectively. The tissue sample types being investigated include the femur bone, bone marrow, fat, muscle, and skin tissues (Fig. 7). We randomly selected the scan locations over the surface area of each tissue. Thus, the images obtained vary in shape and surface location. The training image datasets were acquired from the first pig at 140 scan locations for each tissue type, yielding a total of 700 scans. Meanwhile, the validation image datasets were acquired from the second pig at only 28 scan locations for each tissue type. Therefore, the validation datasets contain only 140 scans. Furthermore, the testing image datasets were acquired from the last pig with similar number of scans as the training image datasets (700 scans).

We acquired 300 repeated B-scan frames for each scan location, the number of B-scans was selected to provide the highest signal strength (as a default setting by the OCT scanner software). These images at each scan location were then registered to remove motion artifacts. Rigid image registration was used because motion artifacts in our OCT images mainly originate from the object's translational motion. We used the fast normalized cross-correlation similarity measure to detect shifts between two images [32], [33]. This method was demonstrated for its application for fast-image-template matching. This registration method was implemented using the *normxcorr2* function in MATLAB® to calculate the correlation coefficient matrix of two images. We used the first frame

for each scan as the static (fixed) image and the other frame as the moving image. The predicted translation is given by the location of the maximum correlation coefficient [34]. The motion-corrected frames were then averaged (frame-averaged) and labeled as the "ground truth" image.

Moreover, we randomly extracted 10 raw images from the 300 repeated images for each scan location and paired them with the same ground-truth frame-averaged images. We called them the raw images and used them for training, as explained in Section III and Fig. 2.

In summary, the training image datasets consisted of 7000 raw images with 700 corresponding ground truth images (7000 image pairs). The validation image datasets consisted of 1400 raw images with 140 corresponding ground-truth frame-averaged images (1400 image pairs). Furthermore, the testing image datasets consisted of only a single image randomly (instead of 10) extracted from each of the 300 repeated frame images. Therefore, only 700 image pairs were used for the testing.

B. Performance Comparison Methods

We compared the image quality and measured the similarities between the images that were denoised using the defined CNN models (UNet-MSE, Resnet-MSE, UNet-WGAN, and Resnet-WGAN) and the reference (frame-averaged) images. The performance of the CNN models were also compared with three classical digital filters—the median filter, block-matching 3D (BM3D) [17], and double-density complex wavelet transform (DD-CDWT) [19]. The image quality evaluation of the denoised images were done quantitatively and qualitatively. Processing time comparisons were also investigated to show the possibility of using the CNN for real-time image denoising.

Additionally, we also investigate the role of training dataset size to the performance of the CNN models. We trained each of CNN model with three variations of dataset size (2000, 5000, and 7000 image pairs). This investigation is intended to demonstrate the ability of the CNN models to generalize noise in the OCT images. Here, the WGAN based loss is expected to improve the noise generalization better than the MSE only loss. In summary, 12 CNN models were defined and compared along with the BM3D and DD-CDWT denoising methods. Definition of the evaluated denoising methods is explained in Table I.

C. CNN Model Training Details

We train all the CNN models in Keras-GPU environment with TensorFlow backend [35]. The training took place on an NVIDIA DGX A100 workstation equipped with NVIDIA A100 GPUs, which enabled us to perform parallel computations to speed up the training process. We trained all of our models using 1000 epochs. The training was done in mini-batches, with a batch size of 8. We selected the adaptive learning rate optimization algorithm (Adam) as the training optimizer [36], with the step size $\alpha = 1 \times 10^{-5}$ and decay parameters $\beta_1 = 0.9$ and $\beta_2 = 0.9$. The loss-weighting parameters, ω_1 and ω_2, for the WGAN-based models were similarly set to 1×10^{-3}. The gradient penalty (λ) was set

TABLE I
DEFINITION OF THE IMAGE AND DENOISING METHODS
THAT WERE INVESTIGATED

Method	Definition
Raw Image	The raw image taken from single scan
Frame-Averaged	The frame-averaged image from 300 scans on the same location
Median Filter	The median filtered image of the raw image
BM3D	Block-matching 3D filter
DD-CDWT	Double-density complex wavelet transform
UNet-MSE 1	UNet model with MSE loss function and trained with 2000 images.
UNet-MSE 2	UNet model with MSE loss function and trained with 5000 images.
UNet-MSE 3	UNet model with MSE loss function and trained with 7000 images.
UNet-WGAN 1	UNet model with WGAN loss function and trained with 2000 images.
UNet-WGAN 2	UNet model with WGAN loss function and trained with 5000 images.
UNet-WGAN 3	UNet model with WGAN loss function and trained with 7000 images.
ResNet-MSE 1	ResNet model with MSE loss function and trained with 2000 images.
ResNet-MSE 2	ResNet model with MSE loss function and trained with 5000 images.
ResNet-MSE 3	ResNet model with MSE loss function and trained with 7000 images.
ResNet-WGAN 1	ResNet model with WGAN loss function and trained with 2000 images.
ResNet-WGAN 2	ResNet model with WGAN loss function and trained with 5000 images.
ResNet-WGAN 3	ResNet model with WGAN loss function and trained with 7000 images.

to 10, as suggested by the original paper [31]. Furthermore, online data augmentation was performed by small random geometrical (translation) shifts and flipping each image horizontally. We used the reflection mode to fill the points outside the boundaries of the image after translation. The data augmentation generators for all classifiers were set to have similar random seeds for fair training.

D. Quantitative Image Quality and Similarity Evaluation

1) Image Quality Metrics: We used the signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) to measure the image quality. These metrics show the noise magnitude of an image. The SNR is defined as the logarithmic ratio of the mean pixel values to the standard deviation of pixel values over the image foreground as follows:

$$SNR = 20 \log \frac{\mu_{fg}}{\sigma_{fg}} \tag{6}$$

The mean μ_{fg} and the standard deviation σ_{fg} were measured over a defined foreground area of the image, consisting of the tissue structure. We applied the Canny edge detection algorithm to define the foreground and background areas [37], [38]. Defining the area enabled us to measure the contrast to noise ratio between tissue textural features and general noise, defined as:

$$CNR = 10 \log \frac{\mu_{fg} - \mu_{bg}}{\sqrt{\sigma_{fg}^2 + \sigma_{bg}^2}} \tag{7}$$

The tissue textural feature is defined as the deviation of the mean value of the foreground μ_{fg} and background μ_{bg}. The general noise is defined as the square root of the total foreground μ_{fg} and background noise μ_{bg}.

2) Similarity Metrics: In addition to the image quality, we also measured the relative similarity of the images to evaluate the denoising methods in terms of noise suppression performance. The similarity is the relative measurement between the frame-averaged and the denoised image. In this work, we defined similarity based on three metrics. The first metric is the peak signal-to-noise ratio (PSNR). Unlike the previous SNR, PSNR is defined as the logarithmic ratio, which is a relative measurement with respect to the reference image:

$$PSNR = 10 \log \frac{MAX_I^2}{MSE} \tag{8}$$

where MAX_I is the peak intensity or maximum pixel value that exists in the frame-averaged image and MSE is the mean squared error between the frame-averaged and the denoised image.

The second metric that we used to measure the similarity was the structural similarity index (SSIM). SSIM is the measure of the perceived visual difference between two images, which was difficult to estimate with PSNR alone. The metric describes similarity based on three main properties: luminance, contrast, and structure [39]. The simplified version of the SSIM is:

$$SSIM(I, \hat{I}) = \frac{(2\mu_I \mu_{\hat{I}} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)} \tag{9}$$

where μ_I, $\mu_{\hat{I}}$, σ_I, $\sigma_{\hat{I}}$, and $\sigma_{I\hat{I}}$ are the local means, standard deviations, and cross-covariance for images I and $\hat{I}$. The constants C_1 and C_2 are the regularization coefficients, used to avoid instability in image regions where the local mean or standard deviation is close to zero. In this work, we set the C_1 and C_2 parameters to $(0.01 \times L)^2$ and $(0.03 \times L)^2$, respectively, where L is the maximum possible pixel intensity range (65535) of our particular OCT image.

The final metric was the edge preservation index (EPI), proposed by Sattar *et al.* [40] to measure edge preservation between the denoised image and the corresponding frame-averaged image. The EPI is defined as follows:

$$EPI = \frac{\Gamma(\Delta s - \overline{\Delta s}, \widehat{\Delta s} - \overline{\widehat{\Delta s}})}{\sqrt{\Gamma(\Delta s - \overline{\Delta s}, \Delta s - \overline{\Delta s}) . \Gamma(\widehat{\Delta s} - \overline{\widehat{\Delta s}}, \widehat{\Delta s} - \overline{\widehat{\Delta s}})}} \tag{10}$$

where Δs and $\widehat{\Delta s}$ are the Laplacian filtered version of the frame-averaged and denoised images, respectively. The gamma function $\Gamma(x, y)$ is the pixel-wise summation function and defined as follows:

$$\Gamma(x, y) = \sum_{i=1}^{w} \sum_{j=1}^{h} x(i, j) . y(i, j) \tag{11}$$

with w and h are the image width and height, respectively.

In addition to measure the image quality and similarity between the frame-averaged and denoised images, we also

measured the processing performance of each method. This was done by averaging the processing time of each image over the testing dataset.

E. Qualitative Image Evaluation

The evaluation of the denoising performance was also conducted with qualitative experiments. Experts mean opinions were collected to quantify the image quality subjectively. The experts were selected who mostly work with medical image processing. This selection was chosen because it is difficult to find an expert who is working specifically on OCT images for normal tissues. Therefore, we defined three non-diagnostic evaluation points to rank the denoised images.

1) Sharpness: This first point evaluates if the tissue surface (border) is clearly visible or blurred. The highest score of 6 indicates that the image has clearly visible (sharp). On the other hand, the lowest score of 1 indicates difficulties in distinguishing tissue surfaces due to blurred images.

2) Contrast Details: This point evaluates the ability to discriminate structures below the tissue surface. The highest score of 6 indicates a clearly visible structural pattern. The lowest score of 1 indicates no structural pattern was noticeable.

3) Noise Level: This last point evaluates the noise level, such as shot noise, salt and paper noise, and Gaussian noise, on both the background and foreground area of the image. The highest score of 6 indicates a noise-free image. The lowest score of 1 indicates that the noise level is too high and hides the tissue structure.

Each expert was asked to evaluate a set of denoised images for each tissue type. A set of denoised images consisted of 17 images as described in section III.B (Table I).

F. Accuracy Comparison of Tissue Classifiers

Tissue classification is the primary aim of our proposed smart laser osteotomy scheme. Apart from showing the performance of the CNN for image denoising, we also explored the effect of image denoising on the accuracy of the tissue classifier. The CNN denoisers are expected to surpass the frame-averaging method with faster processing time and similar image quality, without significant change in its effectiveness of increasing the tissue classifier's accuracy.

We acquired additional image datasets to evaluate changes in the tissue classifier's accuracy due to image denoising. Tissue images of a pig were used to train and validate the classifier. The number of tissues and scan locations were comparable to the datasets used to train the CNN image denoisers (five tissue types and 140 scan locations per tissue). Both raw and frame-averaged image pairs were extracted for each scan location. In total, 700 image pairs (scan locations) were used to train and validate the classifier (with a fraction of 0.7 and 0.3, respectively).

For each scan location, we also denoised the raw images using each denoising method investigated in this study, separately. However, unlike the denoising performance experiments, we only tested the deep learning models which were trained with 7000 training datasets. In summary, each scan
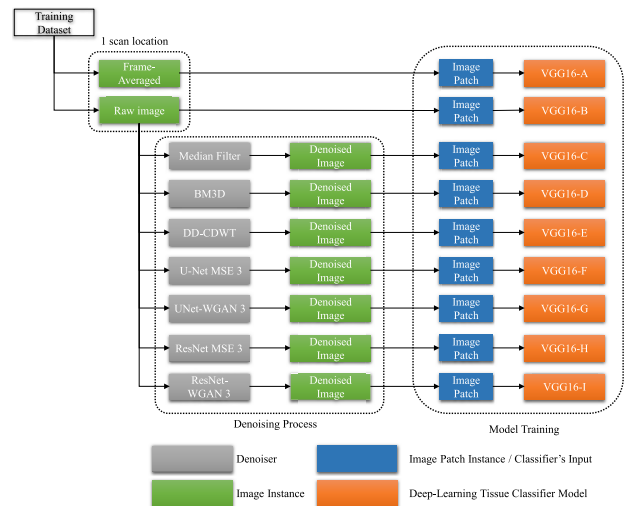
Fig. 8. Training Process of the tissue classifier. A pair of raw and frame-averaged images were acquired for each scan location. Then, the raw image was denoised using the denoisers that were investigated in this paper (median-filtered, BM3D,DD-CDWT, UNet-MSE 3, UNet-WGAN 3, ResNet-MSE 3, and ResNet-WGAN 3). A total of nine image instances (green) were acquired for each scan location. A tissue classifier (orange) was trained to differentiate tissue type based on an image patch (blue) input that is extracted from each image instance (green). The input of the classifier was an image patch (blue) with sized of 128 × 128 pixels grayscale image.

location or image pair consists of nine images (raw, frame-averaged, median-filtered, BM3D, DD-CDWT, UNet-MSE 3, UNet-WGAN 3, ResNet-MSE 3, and ResNet-WGAN 3). Additionally, another tissue image from a different pig was used to test the classifier. Similarly, with the classifiers' training datasets, additional 700 image pairs (five tissue types and 140 scan locations per tissue) were acquired and used for testing.

Furthermore, different from the deep-learning models for image denoising, which uses the full (1024 x 256 pixels) OCT image as the input, the input for the classifier was an image patch. The image patch was a 128 x 128 pixels grayscale image, selected at the center of the tissue surface border on the full OCT image (Fig. 1, step 2). This location represents the location of the laser ablation which is designed to be coaxially aligned with the center of the OCT image. We defined the ablation spot as always in the lateral center of the image. Vertical Canny edge detection method was used to trace the tissue surface in axial direction because of its simplicity and low sensitivity to noise [37], [38]. To have a similar patch location across the classifiers' dataset, we extracted an image patch at the same location for all nine images in each scan location.

Therefore, we trained all of the nine classifiers separately and compared their accuracies in classifying tissue type. The process of denoising and model training is illustrated in Fig. 8. We trained the first classifier to classify tissue type using the frame-averaged image patches. Then, we trained the second classifier using the raw image patches to classify tissue type. We used the second classifier as a reference classifier to compare the performance of the other classifiers. Next, the

third classifier was trained to classify tissue type using the median filtered image patches. The fourth and fifth classifiers were trained separately to classify tissue type based on image patches that denoised using the BM3D and DD-CDWT, respectively. Furthermore, we also trained classifiers using the image patches that were denoised with the CNN models. The sixth and seventh classifiers were trained separately to classify tissue based on image patches that were denoised with the UNet-MSE 3 and UNet-WGAN 3, respectively. Finally, we trained the eighth and ninth classifiers separately based on image patches that were denoised with the ResNet-MSE 3 and ResNet-WGAN 3, respectively.

We used the VGG16-Net [41] model as the base model for all classifiers which are often used for image-based object recognition. We customized the size of the input layers to fit the image patch size. The output layer was a soft-max activated layer, consisting of five neurons for classifying the tissue type (bone, bone marrow, fat, muscle, and skin). All hidden layers were equipped with the Rectified Linear Unit (ReLU) activation function. All the VGG16-Net models were trained and tested using the same workstation to train the CNN denoising models. We initialized all the weight in the models with a similar seed value to ensure fair training. We trained the models with 1000 epochs. The classifiers training was also done in mini-batches, with a batch size of 16 images, to fit our GPU's memory capacity. We defined the cross-categorical entropy as the training loss function and Adam (learning rate = 1.0×10^{-4}) as the training optimizer. We implemented weight decay (L_2 penalty multiplier set to 5.0×10^{-4}) regularizers for all convolutional layers and dropout regularizers for the last two fully connected layers (dropout ratio set to 0.2) as described in the reference [41].

V. RESULTS AND DISCUSSION

After training all of the defined CNN denoiser models, we then applied them to denoise a set of testing images. Comparisons of the denoised images for the bone tissue are shown in Fig. 9. More image comparisons for the other tissue types are shown in the supplementary materials (Fig. 4-8). In this section, we discussed the benefit of using the CNN models for improving the raw image quality. We also compared the results with the frame-averaging, median filter, BM3D, and DD-CDWT. We also measured the similarity between the frame-averaged images with the images that were denoised with the CNN models, median filter, BM3D, DD-CDWT, and the raw images. Furthermore, we measured and compared the processing time for these denoisers to denoise the raw images. Finally, we trained a tissue classifier for each denoiser which used the corresponding denoised image as the training dataset. We compared the changes in the classifier's accuracy relative to a classifier that was trained with the frame-averaged images.

A. Quantitative Evaluation Results

The quantitative measurements were done to compare the SNR and CNR of the denoised images. Table II shows the averaged SNR and CNR of the raw images and the

TABLE II
THE MEAN SNR AND CNR OF THE RAW IMAGES, FRAME-AVERAGED IMAGES, AND THE IMAGES THAT WERE DENOISED WITH THE CNN MODELS, MEDIAN FILTER, BM3D, AND DD-CDWT. BEST VALUE IS DENOTED WITH BOLD TEXT

Method	Quality Metrics	
	Mean SNR (dB)	Mean CNR (dB)
Raw	52.830	0.855
Frame-Averaged	55.858	3.022
Median Filter	58.964	2.598
BM3D	56.338	2.228
DD-CDWT	57.244	2.105
UNet-MSE 1	**61.759**	2.656
UNet-MSE 2	60.027	2.858
UNet-MSE 3	57.923	2.889
UNet-WGAN 1	55.087	1.582
UNet-WGAN 2	53.677	3.419
UNet-WGAN 3	57.281	2.922
ResNet-MSE 1	56.019	**3.566**
ResNet-MSE 2	57.127	2.972
ResNet-MSE 3	57.038	2.899
ResNet-WGAN 1	60.629	2.570
ResNet-WGAN 2	57.299	2.741
ResNet-WGAN 3	57.830	1.999

denoised images using the denoisers investigated in this paper. As the reference method, the frame-averaging effectively improves the image quality of the raw images. However, the frame-averaged images have less SNR compared to almost all other denoised images. The frame-averaging method calculates the mean individual pixel intensity over the temporal domain (frames) and maintains the speckle information in the image. Meanwhile, the other denoising methods effectively reduce the noise (including speckle) in the spatial domain. Here, the SNR is calculated based on the mean and noise in the spatial domain. Nevertheless, the frame-averaged images have higher CNR compared to almost all of the denoised images.

Most of the deep learning methods also improved the SNR of the raw images better than frame-averaging method, except the UNet-WGAN 1 and UNet-WGAN 2. Moreover, the UNet-WGAN 2 and ResNet-MSE 1 improved the CNR of the raw images better than frame-averaging method. The CNN models learned to denoise images through convolutional spatial filters that minimize the loss function. Similar to the median filter, the convolutional filters also consider the neighboring pixels (kernel) for each individual pixel in the image. Additionally, the deep learning regularizers give better generalization to reduce the noise without over-smoothing the speckle pattern and edge details.

Investigation on the generalization capacity of the CNN models has also been conducted by variation of training data sizes. Among the deep learning models with MSE loss, the UNet-MSE models have lower SNR when trained with a higher number of training sizes. Meanwhile, although there is a slight decrease of CNR between the ResNet-MSE 2 and ResNet-MSE 3, the ResNet-MSE models tend to have higher SNR with higher number of training sizes. We also observed that the UNet-MSE models tend to have higher CNR with higher number of training sizes. In contrast, lower CNR is noticeable with higher number of training sizes for the

TABLE III

THE SIMILARITY METRICS OF THE FRAME-AVERAGED IMAGES WITH THE IMAGES THAT WERE DENOISED WITH THE CNN MODELS, MEDIAN FILTER, BM3D, AND DD-CDWT IMAGES. BEST VALUE IS DENOTED WITH BOLD TEXT

Method	Similarity Metrics		
	Mean EPI	Mean PSNR (dB)	Mean SSIM
Median Filter	0.267	34.950	0.979
BM3D	0.466	34.347	0.973
DD-CDWT	0.473	34.404	0.971
UNet-MSE 1	0.529	36.419	0.986
UNet-MSE 2	0.512	37.136	0.986
UNet-MSE 3	0.577	38.906	**0.992**
UNet-WGAN 1	0.542	35.051	0.988
UNet-WGAN 2	0.572	32.195	0.986
UNet-WGAN 3	**0.677**	39.984	**0.992**
ResNet-MSE 1	0.665	39.324	**0.992**
ResNet-MSE 2	0.671	**40.094**	**0.992**
ResNet-MSE 3	0.667	40.033	**0.992**
ResNet-WGAN 1	0.558	33.888	0.986
ResNet-WGAN 2	0.615	32.743	0.989
ResNet-WGAN 3	0.629	30.466	0.989

ResNet-MSE models. On the other hand, we observed a non-linear relation between variation of training data size with both SNR and CNR for all deep learning models with WGAN loss.

Although all of the denoiser performed well in improving the image's SNR and CNR, the image quality measurements were insufficient in quantifying the extent to which denoiser images preserved better structural details, especially for comparing between the deep learning models with WGAN loss. Additional comparison metrics have been done to show the averaged similarity between the frame-averaged images with the images that were denoised with the CNN models, median filter, BM3D, DD-CDWT, and the raw images, respectively. The measurements considered the image's edge (EPI), structural (SSIM), and intensity (PSNR) preservation (or similarity) concerning the frame-averaged images. These metrics give clearer information about the preservation of the tissue structural information, which was previously inferred from the frame-averaged images. The measurement results are shown in Table III.

The results show that the median filter improved the SNR better than the frame-averaging method. However, although the median filter reduced the speckle noise, the median filtered images have the lowest EPI compared to the other denoiser, which indicates the loss of sharp edge details. Meanwhile, the deep learning methods have higher EPI and SSIM than the median filter, indicating better preservation of the sharp edge details. Most of the deep learning denoised images have higher PSNR than the median filtered, BM3D, and DD-CDWT denoised images, except those denoised by the UNet-WGAN 2, ResNet-WGAN 1, ResNet-WGAN 2, and ResNet-WGAN 3. The similarity measurement also shows that the BM3D and DD-CDWT denoised images have higher EPI, which better preserved the edge details than the median filter. These methods still kept some residual noise on the resulting image, leading to lower PSNR and SSIM than the median

filter. The SSIM of the BM3D and DD-CDWT denoised images were lower than the deep learning denoising methods.

The image comparisons in Fig. 9 show that the deep learning models with the MSE loss respond to uncertainty with smoothing (blurring) [30], [42]. Although this problem could be solved by increasing the number of training data sizes. The WGAN based model performed better generalization by keeping the speckle noise as one feature to distinguish between the reference and generated images. Therefore, the WGAN-based models kept a small amount of artificial speckle noise in the generated images.

The results also show that the image quality and similarity were slightly higher for the ResNet-based models than UNet-based models. However, since the image quality and similarity differences between the deep learning models are relatively small, it is difficult to conclude the best model. We believe that these image quality and similarity metrics are insufficient to measure the difference between the deep learning models. Further qualitative measurements are needed to visually inspect the tissue's complex anatomical structures preserved by the proposed deep learning methods.

B. Qualitative Evaluation Results

Evaluation of the denoised image quality based on the quantitative metrics alone may be insufficient to show the visual improvement of the denoised images. Additional qualitative evaluations to show the performance of the CNN models were also conducted. Visual inspection of the denoised images was conducted by six experts. The experts were shown 17 image versions of the same image as described in Table I. The image file names were number coded to hide the corresponding denoising method, which was used to denoise the image before being shown to experts. The experts were asked to evaluate the sharpness, contrast, and noise level of the images as explained in section III.E. The experts survey results for the image sharpness, contrast, and noise level are shown in Figs 10, 11, and 12, respectively.

The survey shows that the experts voted the frame-averaged images to have the highest sharpness. In contrast, the median filter method has been voted to show a low preservation of sharpness. The UNet-MSE 1 and 2 are also voted almost similarly with the median filter. Figs. 9f and 9g confirmed that the images denoised with the UNet-MSE 1 and 2 are over-smoothed. On the other hand, UNet-WGAN 1 and ResNet-WGAN 1 were partly voted to preserve the image sharpness. Although Figs. 9i and 9o show sharp images, loss of tissue-specific textural features are noticeable.

In the second part of the survey, we asked the experts to compare the perceived contrast of the images. The experts voted that the UNet-MSE 1 and 2 have low contrast, even lower than the median filter. Similar to the sharpness evaluation, the UNet-WGAN 1 and ResNet-WGAN 1 were also partly voted to preserve the image contrast. Almost similar performance was observed between the frame-averaging methods and the ResNet-WGAN 2 models in the eyes of the experts.

The noise level evaluation showed that the UNet-WGAN 3 model reduced the noise as well as the frame-averaging
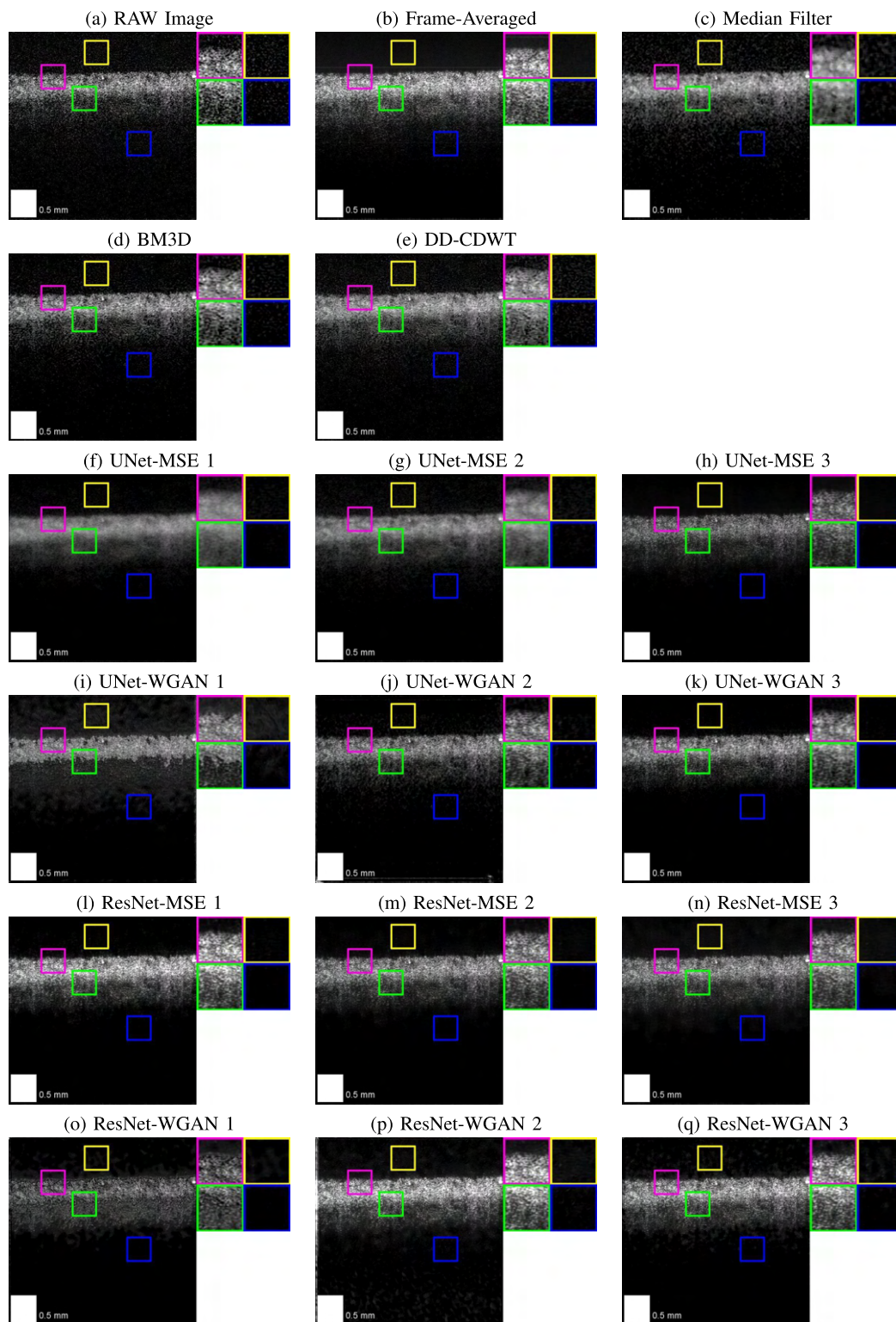
Fig. 9. Image comparison for bone tissue. From top left to bottom right, starting with the raw-image (a), frame-averaged image (b), and the raw images that were denoised using the median filter (c), BM3D (d), DD-CDWT (e), UNet-MSE 1 (f), UNet-MSE 2 (g), UNet-MSE 3 (h), UNet-WGAN 1 (i), UNet-WGAN 2 (j), UNet-WGAN 3 (k), ResNet-MSE 1 (l), ResNet-MSE 2 (m), ResNet-MSE 3 (n), ResNet-WGAN 1 (o), ResNet-WGAN 2 (p), and ResNet-WGAN 3 (q). The colored boxes show the zoomed version of regions inside the images. Yellow box shows the background region above the tissue surface. Magenta box shows the surface region of the tissue. Green box shows the region inside the tissue with high signal. Blue box shows the background region with low to no signal inside the tissue.
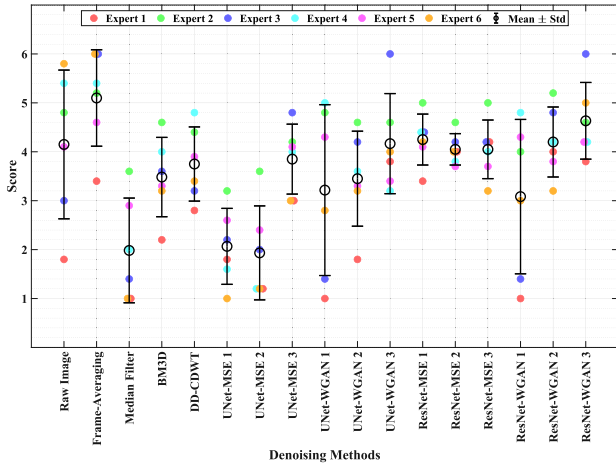
Fig. 10. Error bar plot represents the survey results for the sharpness of the raw and denoised images. The mean scores are indicated in black circles, with the whiskers indicating one standard deviation of the scores (equally up and down). Individual scores from each expert are shown in dots with the same color. Scores with similar values are shown side by side horizontally to avoid overlapping.

Fig. 12. Error bar plot represents the survey results for the noise level of the raw and denoised images. The mean scores are indicated in black circles, with the whiskers indicating one standard deviation of the scores (equally up and down). Individual scores from each expert are shown in dots with the same color. Scores with similar values are shown side by side horizontally to avoid overlapping.
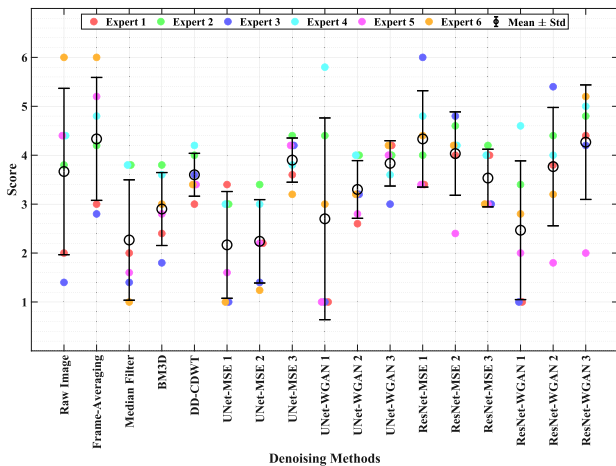
Fig. 11. Error bar plot represents the survey results for the contrast of the raw and denoised images. The mean scores are indicated in black circles, with the whiskers indicating one standard deviation of the scores (equally up and down). Individual scores from each expert are shown in dots with the same color. Scores with similar values are shown side by side horizontally to avoid overlapping.

method. The UNet-WGAN 1 only received two votes, which indicates that the denoised images have relatively higher noise level, confirmed in Fig. 9i. The images showed that the denoiser completely hid the textural features in the tissue. Other than that, the median filtered and BM3D has almost similar noise level as the raw images.

In summary, it is difficult to conclude which denoising methods performed best due to the high variation of scores among the experts. It is because the experts work with different medical imaging modalities, including OCT for ophthalmology and ultrasound imaging.

C. Improvement in Classifier Accuracy

We collected 700 additional OCT image pairs to demonstrate the effect of image denoising on a tissue classifier's accuracy. Each image pair consists of raw and frame-averaged images. We denoised the raw images for each image pair or scan location using each denoisers investigated in this study separately. However, we only investigated the deep learning models which were trained with 7000 images. Images patches were extracted from the denoised images and used as input for the tissue classifier. We trained a tissue classifier for each denoising method and used the corresponding denoised image patches as the training dataset. In total, nine VGG16-Net tissue classifiers were trained separately. After training the models, we tested them with the test dataset consisting of 700 denoised OCT image patches. A detailed explanation regarding the training process is explained in section IV-F.

The results in Table IV show that all of the denoiser increased the accuracy of the tissue classifier except the median filtering method. The tissue classifier trained with raw images has an accuracy of 86.43 %. As a reference, the tissue classifier trained with frame-averaged images has an accuracy of 91.29 %, the highest accuracy compared to the other tissue classifiers. Although the frame-averaged image still contained an element of the speckle noise, they also contained more information about the tissue structure, gathered from several image frames, compared to single raw denoised images. Therefore, the tissue classifier trained with frame-averaged images could learn more features than that trained with single raw denoised images. However, the tissue classifiers trained using deep learning denoised images have the closest accuracies to the tissue classifier trained using the frame-averaged images. In comparison, the ResNet models improved the tissue classifier's accuracy more than the UNet models. This is because the ResNet models reduced the speckle noise better than the UNet models. Therefore, the tissue classifier could focus more on the signal-carrying speckle to differentiate tissue type.

Here, it is also shown that the MSE-based deep learning models have higher accuracy compared to the WGAN-based models. This discrepancy appears because the WGAN

TABLE IV
THE AVERAGE ACCURACY OF THE CLASSIFIER TRAINED USING THE RAW IMAGES, FRAME-AVERAGED IMAGES, AND THE IMAGES THAT IS DENOISED USING BM3D, DD-CDWT, UNET-MSE 3, UNET-WGAN 3, RESNET-MSE 3, AND RESNET-WGAN 3. BEST VALUE IS DENOTED WITH BOLD TEXT

Denoising Method	Classifier Average Accuracy (%)
Raw	86.43
Frame-Averaging	**91.29**
Median Filter	85.57
BM3D	88.86
DD-CDWT	88.71
UNet-MSE 3	90.83
UNet-WGAN 3	89.73
ResNet-MSE 3	91.03
ResNet-WGAN 3	89.86

TABLE V
THE AVERAGE TIME REQUIRED TO DENOISE THE OCT IMAGES USING THE FRAME-AVERAGING, BM3D, DD-CDWT, UNET-MSE, UNET-WGAN, RESNET-MSE, AND RESNET-WGAN. NOTE THAT DEEP LEARNING ALGORITHMS WERE RUN USING GPU. THE MEASUREMENT ALSO INCLUDES ACQUISITION TIME OF 300 RAW IMAGES FOR THE FRAME-AVERAGING METHOD. BEST VALUE IS DENOTED WITH BOLD TEXT

Denoising Method	Average Time (s)
Single Raw	**0.009**
Frame-Averaging	2.731
Median Filter	0.010
BM3D	2.495
DD-CDWT	0.202
UNet	0.084
ResNet	0.078

generated more artificial noise, lowering the image quality and similarity, as discussed in the previous section.

The BM3D and DD-CDWT methods also increased the tissue classifier's accuracy. However, the increases are lower than the frame-averaging and deep learning because of the residual noise in the denoised images. On the other hand, the median filter even reduced the tissue classifier's accuracy. These results indicate that the deep learning-based methods could replace the classical digital filtering methods without significantly altering the structural information in the image for tissue classification. Nevertheless, further investigation is needed to optimize the tissue classifier's performance by training and testing with more data samples.

D. Processing Time

One aim of this work was to find a fast algorithm for denoising the OCT images. We evaluated the performance of the trained models on our OCT computer with the following specifications: 2.8 GHz Intel Core i7 processor, 16 GB 1867 MHz DDR3 memory, and equipped with a GPU NVIDIA GTX 1050Ti. We compared the average processing time of the proposed denoising methods. However, the processing performance comparison of the CNN models was done between UNet and Resnet only, as the generator in both MSE and WGAN based loss model. The results are given in Table V. Our OCT system required 9 msec to acquire a single B-Scan (raw) image. Therefore, the frame-averaging was the longest denoising method which required 300 raw images. The median filter was the fastest algorithm as it only calculates a small-sized median kernel over the input image. However, this achievement must be considered alongside the median filter's image similarity result, which erased information over the image. In the second position, the deep learning methods denoised the raw image faster than the BM3D and DD-CDWT methods. The deep learning based models were faster since they could be run in parallel on the GPU, whereas the classical digital filtering methods were run on the CPU. Furthermore, among the deep learning methods, the ResNet models denoised a single raw image in only 0.078 s, which is faster than the UNet-based models that denoised a single raw image in 0.084 s.

Our findings confirm the possibility of achieving real-time image denoising for our smart laser osteotomy approach. All of the deep learning models denoise the raw image below 90 msec. Additionally, the tissue classification using VGG16-Net needs an average time of 0.034 s. This leads to a total processing time of 0.112 s for the acquisition, denoising with ResNet, and classification steps, which is slightly slower than the optimum repetition rate of the ablation laser. Our previous experiment showed that the optimum pulse repetition rate of our laser ablation was 10 Hz (100 msec per pulse) [43].

VI. CONCLUSION AND FUTURE WORKS

This work demonstrates the ability of deep learning methods to mimic the frame-averaging method for denoising OCT images of five normal tissue types. The deep learning methods produced better image quality and similarity than the classical digital filtering methods. Even though the median filter could increase the SNR and CNR better than some of the deep learning methods, it failed to maintain the structural information of the image. Furthermore, the processing speed of the deep learning-based method was also faster than the BM3D, DD-CDWT, and frame-averaging methods. The quantitative and qualitative experiment results suggest that the deep learning methods are a feasible alternative to the frame-averaging method for real-time OCT image denoising, a necessary subprocess for our smart laser osteotomy approach.

Moreover, we also showed that denoising the OCT image increased the tissue classifier's accuracy. The frame-averaging method improved the tissue classifier's accuracy better than the other denoising methods. Furthermore, the tissue classifier has better accuracy when trained using the images that are denoised by deep learning methods than the classical digital filter methods. It proves that the deep learning methods could mimic the frame-averaging method better than the classical digital filter methods.

In the future, we will integrate the deep learning denoising method into our OCT device for monitoring laser ablation in real-time. We are aware that integrating the tissue classifier and image denoising problem will further increase the processing time. The average total time for acquisition, denoising, and

classification steps is slightly slower than the optimum repetition rate of the ablation laser. Further study will involve optimization of the tissue classifier for better accuracy and faster prediction time. One of the optimizations includes using a faster CPU and GPU. Other than that, we also plan to use the classification accuracy directly as one of the loss functions for deep learning denoising models. Thus, reducing the processing time for both denoising images and predicting tissue type simultaneously seem feasible.

References

[1] K.-W. Baek et al., "A comparative investigation of bone surface after cutting with mechanical tools and Er: YAG laser," Lasers Surg. Med., vol. 47, no. 5, pp. 426–432, Jul. 2015.

[2] A. Trompeter, J. Dabis, O. Templeton-Ward, A. E. Lacey, and B. Narayan, "The history, evolution and basic science of osteotomy techniques," Strategies Trauma Limb Reconstruction, vol. 12, no. 3, pp. 169–180, Nov. 2017.

[3] S. Kondo et al., "Thermological study of drilling bone tissue with a high-speed drill," Neurosurgery, vol. 46, no. 5, pp. 1162–1168, 2000.

[4] S. Stübinger, "Advances in bone surgery: The Er: YAG laser in oral surgery and implant dentistry," Clin., Cosmetic Investigational Dentistry, vol. 2, pp. 47–62, Jun. 2010.

[5] K.-W. Baek et al., "Clinical applicability of robot-guided contact-free laser osteotomy in cranio-maxillo-facial surgery: In-vitro simulation and in-vivo surgery in minipig mandibles," Brit. J. Oral Maxillofacial Surg., vol. 53, no. 10, pp. 976–981, Dec. 2015.

[6] H. Abbasi, G. Rauter, R. Guzman, P. C. Cattin, and A. Zam, "Laser-induced breakdown spectroscopy as a potential tool for autocarbonization detection in laserosteotomy," J. Biomed. Opt., vol. 23, no. 7, p. 1, Mar. 2018.

[7] R. Kanawade et al., "Qualitative tissue differentiation by analysing the intensity ratios of atomic emission lines using laser induced breakdown spectroscopy (LIBS): Prospects for a feedback mechanism for surgical laser systems," J. Biophotonics, vol. 8, nos. 1–2, pp. 153–161, Jan. 2015.

[8] F. Mehari et al., "Investigation of laser induced breakdown spectroscopy (LIBS) for the differentiation of nerve and gland tissue—A possible application for a laser surgery feedback control mechanism," Plasma Sci. Technol., vol. 18, no. 6, pp. 654–660, Jun. 2016.

[9] H. N. Kenhagho, G. Rauter, R. Guzman, P. C. Cattin, and A. Zam, "Optoacoustic tissue differentiation using a Mach–Zehnder interferometer," IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 66, no. 9, pp. 1435–1443, Sep. 2019.

[10] N. Kenhagho et al., "Characterization of ablated bone and muscle for long-pulsed laser ablation in dry and wet conditions," Materials, vol. 12, no. 8, p. 1338, Apr. 2019.

[11] E. Bay, A. Douplik, and D. Razansky, "Optoacoustic monitoring of cutting efficiency and thermal damage during laser ablation," Lasers Med. Sci., vol. 29, no. 3, pp. 1029–1035, May 2014.

[12] V. Periyasamy, C. Özsoy, M. Reiss, X. L. Deán-Ben, and D. Razansky, "In vivo optoacoustic monitoring of percutaneous laser ablation of tumors in a murine breast cancer model," Opt. Lett., vol. 45, no. 7, pp. 2006–2009, 2020.

[13] M. E. Brezinski et al., "Optical coherence tomography for optical biopsy," Circulation, vol. 93, no. 6, pp. 1206–1213, 1996.

[14] A. Hamidi, Y. A. Bayhaqi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Long-range optical coherence tomography with extended depth-of-focus: Avisual feedback system for smart laser osteotomy," Biomed. Opt. Exp., vol. 12, no. 4, pp. 2118–2133, Apr. 2021.

[15] J. Schmitt, S. Xiang, and K. Yung, "Speckle in optical coherence tomography," J. Biomed. Opt., vol. 4, no. 1, pp. 95–105, 1999.

[16] W. Wu, O. Tan, R. R. Pappuru, H. Duan, and D. Huang, "Assessment of frame-averaging algorithms in oct image analysis," Ophthalmic Surg. Lasers Imag. Retina, vol. 44, no. 2, pp. 75–168, 2013.

[17] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[18] S. Huang, C. Tang, M. Xu, Y. Qiu, and Z. Lei, "BM3D-based total variation algorithm for speckle removal with structure-preserving in OCT images," Appl. Opt., vol. 58, no. 23, pp. 6233–6243, Aug. 2019.

[19] I. W. Selesnick, "The double-density dual-tree DWT," IEEE Trans. Signal Process., vol. 52, no. 5, pp. 1304–1314, May 2004.

[20] H. Liu, S. Lin, C. Ye, D. Yu, J. Qin, and L. An, "Using a dual-tree complex wavelet transform for denoising an optical coherence tomography angiography blood vessel image," OSA Continuum, vol. 3, no. 9, pp. 2630–2645, Sep. 2020.

[21] D. Perdios, A. Besson, M. Arditi, and J.-P. Thiran, "A deep learning approach to ultrasound image recovery," in Proc. IEEE Int. Ultrason. Symp. (IUS), Sep. 2017, pp. 1–4.

[22] Q. Yang et al., "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," IEEE Trans. Med. Imag., vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[23] K. J. Halupka et al., "Retinal optical coherence tomography image enhancement via deep learning," Biomed. Opt. Exp., vol. 9, no. 12, pp. 6205–6221, Dec. 2018.

[24] Y. Huang, N. Zhang, and Q. Hao, "Real-time noise reduction based on ground truth free deep learning for optical coherence tomography," Biomed. Opt. Exp., vol. 12, no. 4, pp. 2027–2040, Apr. 2021.

[25] Z. Mao et al., "Deep learning based noise reduction method for automatic 3D segmentation of the anterior of lamina cribrosa in optical coherence tomography volumetric scans," Biomed. Opt. Exp., vol. 10, no. 11, pp. 5832–5851, 2019.

[26] B. Baumann et al., "Signal averaging improves signal-to-noise in oct images: But which approach works best, and when?" Biomed. Opt. Exp., vol. 10, no. 11, pp. 5755–5775, 2019.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention. Cham, Switzerland: Springer, 2015, pp. 234–241.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in Proc. Eur. Conf. Comput. Vis. Cham: Switzerland: Springer, 2016, pp. 630–645.

[29] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in Proc. 34th Int. Conf. Mach. Learn. (ICML), 2017, pp. 214–223.

[30] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 105–114.

[31] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein gans," in Proc. 31st Int. Conf. Neural Inf. Process. Syst., 2017, pp. 5769–5779.

[32] J. Sarvaiya, S. Patnaik, and S. Bombaywala, "Image registration by template matching using normalized cross-correlation," in Proc. Int. Conf. Adv. Comput., Control, Telecommun. Technol. (ACT), 2009, pp. 819–822.

[33] S. Makita, M. Miura, S. Azuma, T. Mino, T. Yamaguchi, and Y. Yasuno, "Accurately motion-corrected Lissajous oct with multi-type image registration," Biomed. Opt. Exp., vol. 12, no. 1, pp. 637–653, 2021.

[34] MATLAB. Registering an Image Using Normalized Cross-Correlation. [Online]. Available: https://www.mathworks.com/help/images/registering-an-image-using-normalized-cross-correlation.html

[35] F. Chollet et al. (2015). Keras. [Online]. Available: https://keras.io

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, 2015, pp. 1–15.

[37] J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[38] S. Luo, J. Yang, Q. Gao, S. Zhou, and C. A. Zhan, "The edge detectors suitable for retinal oct image segmentation," J. Healthc. Eng., vol. 2017, p. 3978410, 2017.

[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[40] F. Sattar, L. Floreby, G. Salomonsson, and B. Lövström, "Image enhancement based on a nonlinear multiscale method," IEEE Trans. Image Process., vol. 6, no. 6, pp. 888–895, Jun. 1997.

[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. 3rd Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, 2015, pp. 1–14.

[42] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," 2016, arXiv:1603.08155.

[43] L. M. B. Bernal et al., "Optimizing controlled laser cutting of hard tissue (bone)," Automatisierungstechnik, vol. 66, no. 12, pp. 1072–1082, Dec. 2018.

Deep Learning Models Comparisons

The publications in this chapter explored and compared several deep learning models. Offline experiments were conducted with the aim of finding the best possible deep learning models capable of identifying several tissue types. It is also important to note that the image denoising in these two publications was done using the frame-averaging method.

5.1 Tissue Characterization with Deep Learning: Preliminary Study

The first paper evaluated two deep neural network models to identify three tissue types: bone, fat, and muscle. The first neural network model was outlined by Debdoot et al. [118], with two fully connected layers. For the second model, the deep neural network was engineered to accept two inputs. The first input was the intensity-based image patch, while the second input was the A-Scan profile. The deep neural network model was then engineered into a decision-tree-like model that separates textures of heterogeneous (muscle) and homogeneous (bone and fat) tissues. The results of the second model showed $98.66\,\%$ accuracy, while the first model achieved only $82.4\,\%$ accuracy.

Publication: Y. A. Bayhaqi, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Neural network in tissue characterization of Optical Coherence Tomography (OCT) image for smart laser surgery: preliminary study," *Proc. SPIE*, Vol. 11044, in *Third International Seminar on Photonics, Optics, and Its Applications 2018*, 2018.

Neural network in tissue characterization of Optical Coherence Tomography (OCT) image for smart laser surgery: preliminary study

Yakub A. Bayhaqi, Alexander Navarini, Georg Rauter, Philippe C. Cattin, Azhar Zam

SPIE.

Neural Network in Tissue Characterization of Optical Coherence Tomography Images for Smart Laser Surgery: A Preliminary Study

Yakub A. Bayhaqi[*a], Alexander Navarini[b], Georg Rauter[c], Philippe C. Cattin[d], and Azhar Zam[†a]

[a]Biomedical Laser and Optics Group (BLOG), Dept. of Biomedical Engineering, University of Basel, Gewerbestrasse 14, 4123 Allschwil, Switzerland
[b]Dermatology, Dept. of Biomedicine, University of Basel, Hebelstrasse 20, 4031 Basel, Switzerland
[c]Bio-Inspired RObots for MEDicine-Lab (BIROMED-Lab), Dept. of Biomedical Engineering, University of Basel, Gewerbestrasse 14, 4123 Allschwil, Switzerland
[d]Center for medical Image Analysis and Navigation (CIAN), Dept. of Biomedical Engineering, University of Basel, Gewerbestrasse 14, 4123 Allschwil, Switzerland

ABSTRACT

The aim of this study is to develop an automatic tissue characterization system, based on Optical Coherence Tomography (OCT) images, for smart laser surgery. OCT is rapidly becoming the method of choice for investigating thin tissues or subsurface imaging. In smart laser surgery, OCT could be used to indicate which tissue is being irradiated, thereby preventing the laser from ablating critical tissue such as nerves and veins. Automatic tissue characterization based on the OCT images should be sufficient to give feedback to the laser control. In this study, two main neural networks were trained to classify texture and optical attenuation of three different tissues (bone, fat, and muscle). One neural network texture classifier was trained to differentiate between patterned and patternless images. The other neural network was trained to classify patternless images based on their attenuation profile. The two neural networks were stacked as a binary tree. The ability of this hybrid deep-learning approach to characterize tissue was evaluated for accuracy in classifying OCT images from these three different tissues. The overall (averaged) accuracy was 82.4% for the texture-based network and 98.0% for the attenuation-based (A-Scan) network. The fully connected layer of the neural network achieved 98.7% accuracy. This method shows the ability of the neural network to learn feature representation from OCT images and offers a feasible solution to the challenge of heuristic independent tissue characterization for histology and use in smart laser surgery.

Keywords: Neural Network, Laser Surgery, Tissue Characterization, Optical Coherence Tomography

1. INTRODUCTION

Laser osteotomy is the surgical procedure of cutting bone using a laser. The medical application of this emerging technology is of increasing interest to surgeons, as it overcomes many of the disadvantages of using conventional burrs and saws. Those disadvantages include extensive heat deposition, broadening of cuts, mechanical trauma, metal shaving depositions, and bacterial contamination. The MIRACLE (Minimally Invasive Robot-Assisted Computer-guided LaserosteotomE) project investigates the use of computers and robots to guide minimaly invasive laser osteotomes. One goal of the project is to ensure that laser ablation of the targeted tissue is monitored to avoid cutting important tissues, especially vital ones such as nerves and veins. Thus, the problem of resolving tissue characterization becomes a critical area of study. In a previous study, tissue characterization was based on physical properties like measured acoustic shock waves and Laser-Induced Breakdown Spectroscopy.[1,2] This study focuses on the use of Optical Coherence Tomography (OCT) to distinguish between tissues because of its advantage of providing information in the form of images to monitor the depth of laser ablation. Previous studies have also shown how OCT can help to monitor the depth of laser ablation.[3] Furthermore, this study focuses on the use of neural networks to characterize tissues based on OCT images.

---

[*]yakub.bayhaqi@unibas.ch; phone +41 (0)61 207 54 61; dbe.unibas.ch
[†]azhar.zam@unibas.ch; phone +41 (0)61 207 54 60; dbe.unibas.ch

1.1 Smart laser surgery

The basic idea of the smart laser surgery system proposed in this study is to use OCT images to monitor tissue anatomy at subsurface level before laser ablation. An artificial neural network was trained to differentiate tissue based on the OCT images and can then determine if the ablating laser should stop or continue. A schematic diagram of this system is shown in Figure 1. In this study, we evaluate the neural networks' ability to differentiate or characterize OCT images to support the feedback system of the ablation laser.

1.2 Optical Coherence Tomography imaging

OCT is rapidly becoming the preferred method for *in vivo* investigation of thin tissues or subsurface imaging. Subsurface imaging, using ultrasonic and optical imaging modalities, is a clinically viable alternative for real-time high resolution *in situ* investigation of soft tissue lesions in critical organs. OCT is an optical diffraction tomography principle-based imaging modality that relies on sensing speckles formed due to the coherent interference of photons backscattered from tissues irradiated with a broadband light source.[4] OCT is suitable for use in laser surgery because it is a high resolution and non-invasive imaging compared to ultrasound and histology.[5,6] Although the maximum imaging range of the OCT system is limited to a few millimeters, this depth is enough to examine small ablation craters. Ablation craters created with an Er:YAG laser typically has an ablation rate of 0.4 mm/s.[3]
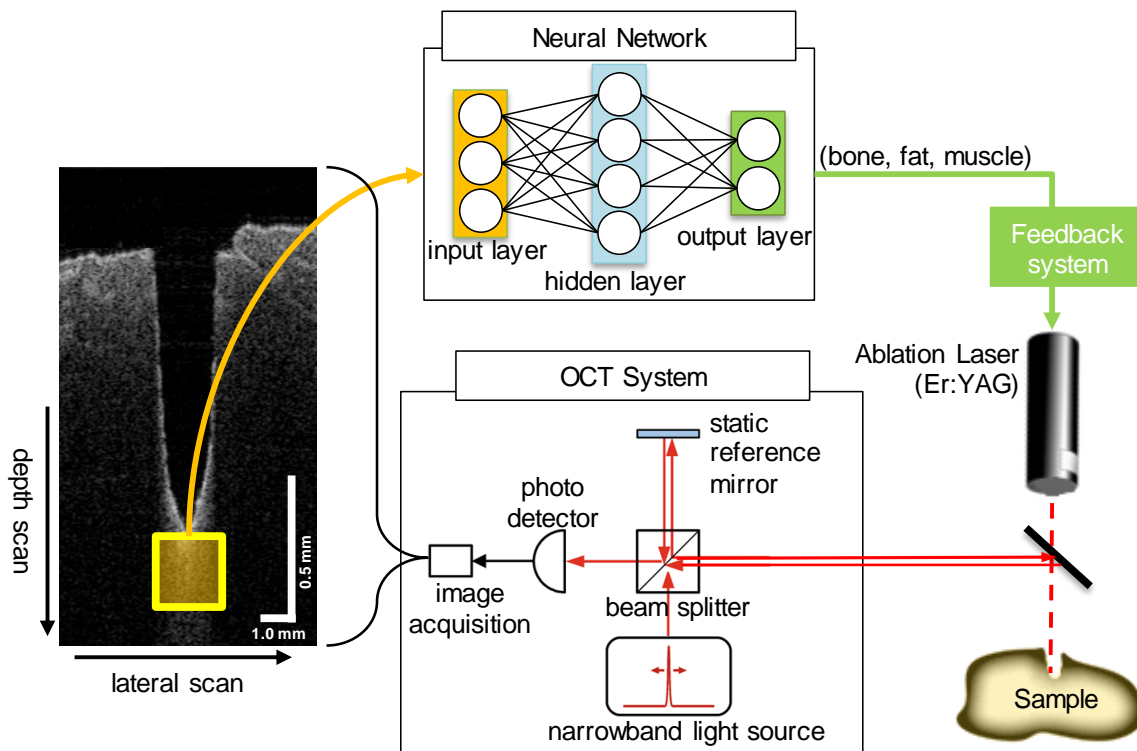
Figure 1. Schematic diagram of the smart laser surgery. A region of interest from the OCT images is used as the input layer of the neural network to classifies the tissue type at the ablation position, which output is used as a feedback system for the ablation laser to stop or continue the ablation.

1.3 Artificial neural network for smart laser surgery

In image pattern recognition and computer vision, neural networks are known for their ability to classify, retrieve, detect and segment images. In some applications, neural networks are used to correct and caption images. Image correction or denoising is very useful in OCT imaging because the neural network creates a better resolution of the speckled image by learning to reconstruct an image based on a ground truth image.[7]

To date, tissue characterization with OCT has usually been based on texture features, but for bone and fat tissues with fewer visible structures in the required size range, OCT may be unsuitable. However, an examination of structure-poor OCT images reveals that they frequently display differentiated attenuation profiles.[8] In this study, we propose combining the texture and attenuation features of OCT images.

2. MATERIAL AND METHOD

2.1 Optical Coherence Tomography system

The OCT system used in this study was a custom Axsun swept source OCT system with a center wavelength (λ_0) of 1060 nm and a sweep rate of 100 kHz. This system was used to obtain subsurface images of the femur bone, fat, and muscle tissues of a pig. For each tissue sample, OCT B-Scan images were taken in 100 different locations. The locations were spaced out with a gap of 1 mm between them and randomly taken from different viewing angles. The image size acquired from this OCT system was 1024 pixels (1.6 mm) deep x 300 pixels (7.5 mm) wide. In each location, 300 image samples were acquired, which were then averaged to eliminate the speckle of the OCT. In total, from all tissue types, 300 averaged B-Scan images were used for analysis. The OCT system produces a 32-bit grayscale TIFF image. Contrast enhancement of the images was done using the contrast enhancement feature of ImageJ software, Version 1.52d.
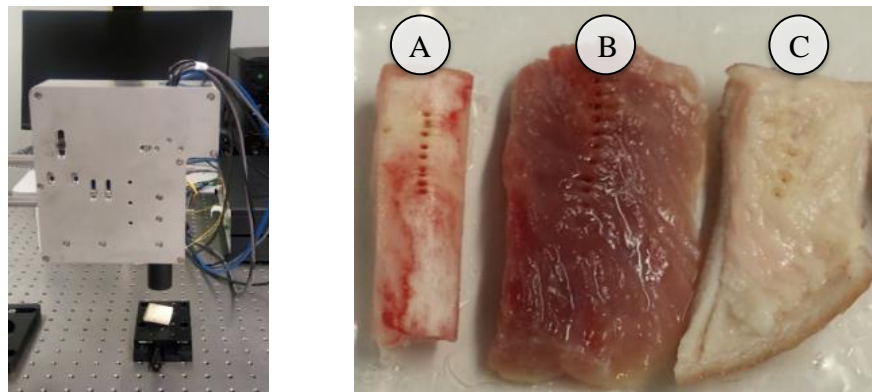
Figure 2. Axsun swept source OCT system (left) and pig femur bone (A), muscle (B), and fat (C) tissue samples (right).
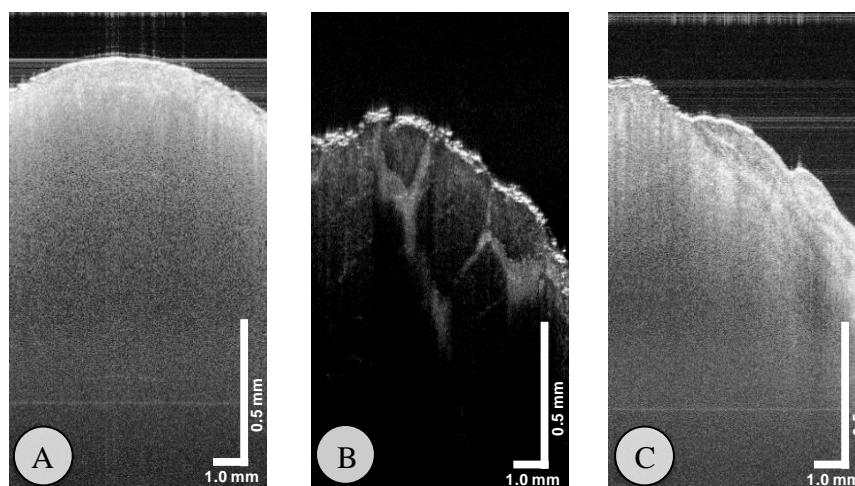
Figure 3. Example of the enhanced OCT images with 1024 pixels (1.6 mm) size in deep x 300 pixels (7.5 mm) size in wide for the pig femur bone (A), muscle (B), and fat (C) tissue samples.

2.2 Neural network architecture

The first neural network model was outlined by Debdoot et al.[4] In his paper, he proposed two denoising autoencoder layers to denoise the speckled appearance of OCT images. The input layer for the neural network was the image patch with pixel size 36 x 36, selected at the center position. The first autoencoder consisted of 400 neurons and the second autoencoder consisted of 100 neurons. The output layer of the network was the classification layer of the three different types of tissues.
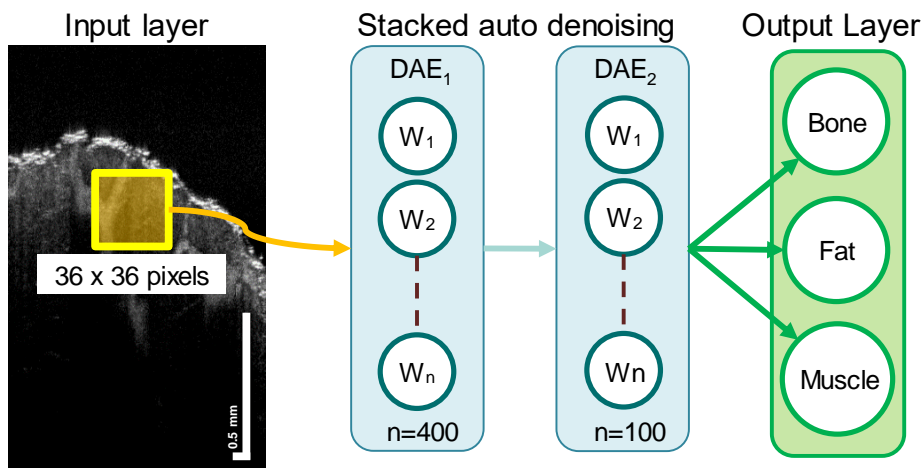
Figure 4. The first model of neural network architecture for texture-based classification.
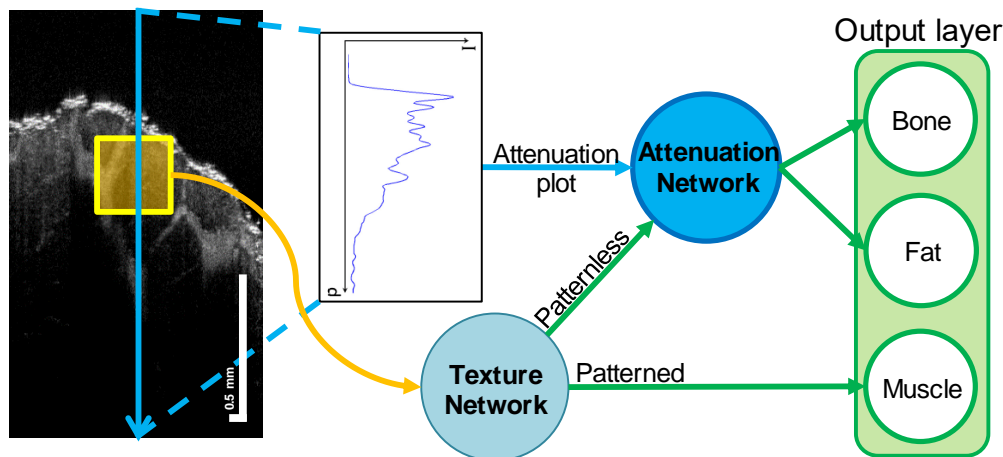
Figure 5. Combination model of the neural network architecture using the texture and attenuation features.

Another feature of the OCT image was the attenuation profile of tissue. A network to interpret the attenuation profile was originally implemented by Giovanni et al. to differentiate between atherosclerotic tissue in intravascular optical coherence tomography (IV-OCT).[8] In their study, the researchers made use of the backscattering and attenuation features of OCT images. This feature was then combined with the previous neural network model in a decision-tree model. The design of the combination model can be seen in Figure 5. In the combination model, the two networks were trained independently. The first network was trained to differentiate between patternless (texture-poor) images, such as bone and fat, and patterned images, such as muscle. Using the patternless output image from the first network, the second network was trained to differentiate between bone and fat based on their attenuation profiles.

Both neural networks were trained during separate periods. For each network, the autoencoder was also trained independently, followed by the training of the softmax layer. The denoising autoencoder used in the intermediate layer

employed sparse learning to pre-train the neural network that tries to compress and encode the cardinal representation of a pattern.[7] The autoencoder layer was trained with parameters of 2000 epochs, 0.004 L2 weight regularization, 4 regularizations and 0.15 proportion of sparsity; the softmax layer was trained with parameters of 1000 epochs by using the mean squared error loss function. The 300 OCT images for all the tissues were divided evenly for the training and testing. The training of all the neural networks was done in MATLAB R2017b software.

3. RESULT AND DISCUSSION

The ability of our neural networks to characterize tissue was evaluated following a leave-one-out 10-fold cross validation. Table 1 shows the result of the first model of the neural network, with an overall average accuracy of 82.4%. This result indicates that the texture-based learning was inaccurate and insufficient for differentiating bone and fat because of the similarity in texture, compared to muscle tissue, which was clearly classified. However, combined with the attenuation profile, the second model of the neural network performed better in terms of classifying bone and fat. The overall accuracy for the combination model was 98.66%, as shown in Table 2.

Table 1. Confusion matrix of the first neural network model.

Input Data	Classified as Bone	Classified as Fat	Classified as Muscle	Accuracy
Bone	39	2	9	78
Fat	15	35	0	70
Muscle	0	0	50	100
			Average Accuracy	82.66

Table 2. Confusion matrix of the hybrid neural network model.

Input Data	Classified as Bone	Classified as Fat	Classified as Muscle	Accuracy (%)
Bone	48	1	1	96
Fat	0	50	0	100
Muscle	0	0	50	100
			Average Accuracy	98.66

This result points to the possibility of using neural networks to characterize tissue during laser surgery. However, the data used in this study (in terms of OCT images) was limited, thus, further studies are needed to train and test the neural networks using thousands of OCT images and more tissue types to be classified. Further studies of another neural network model should also be considered.

4. CONCLUSION

This preliminary study shows that artificial neural networks can be used to differentiate OCT images of bone, fat, and muscle. The method has been experimentally demonstrated with *in situ* tests on animal tissue. The output of the fully connected layer in this study was 98.66 % accurate. Combining pattern and attenuation profile learning improves accuracy, particularly when homogenous tissues have no distinct pattern. It would be worth exploring other artificial neural network architectures commonly used in image recognition. Further studies are also needed to investigate how the system would work in real conditions, particularly the real-time feedback to the laser system and the characterization of more heterogeneous tissues.

ACKNOWLEDGMENT

REFERENCES

[1] Nguendon, H.K., Faivre, N., Meylan, B., Shevchik, S., Rauter, G., Guzman, R., Cattin, P.C., Wasmer, K. and Zam, A. "Characterization of ablated porcine bone and muscle using laser-induced acoustic wave method for tissue differentiation," Proc. SPIE 10417, 104170N (2017).

[2] Abbasi, H., Rauter, G., Guzman, R., Cattin, P.C. and Zam, A., "Differentiation of femur bone from surrounding soft tissue using laser-induced breakdown spectroscopy as a feedback system for Smart Laserosteotomy," Proc. SPIE 10685, 1068519 (2018).

[3] Bernal, L. M. B., Schmidt, I.T., Vulin, N., Widmer, J.,Snedeker, J.G., Cattin, P.C., Zam, A. and Rauter, Georg, R., "Optimizing controlled laser cutting of hard tissue (bone)," accepted to be published in Automatisierungstechnik (2018).

[4] Sheet, D., Karri, S. P. K., Katouzian, A., Navab, N., Ray, A. K. and Chatterjee J., "Deep learning of tissue specific speckle representations in optical coherence tomography and deeper exploration for in situ histology," IEEE 12th Int. Symp. Biomed. Imaging, 777-780, (2015).

[5] Tearney, G.J., Brezinski, M.E., Boppart, S.A., Bouma, B.E., Weissman, N., Southern, J.F., Swanson, E.A., Fujimoto, J.G., "Catheter-based optical imaging of a human coronary artery," Circulation 94, 3013 (1996).

[6] Brezinski, M.E., Tearney, G.J., Bouma, B.E., Izatt, J.A., Hee, M.R., Swanson, E.A., Southern, J.F. and Fujimoto, J.G., "Optical coherence tomography for optical biopsy. Properties and demonstration of vascular pathology," Circulation 93, 1206–1213 (1996).

[7] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P., "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," J. Mach. Learn. Res., vol. 11, 3371-3408 (2010).

[8] Ughi, G. J., Adriaenssens, T., Sinnaeve, P., Desmet, W., and D'hooge, J., "Automated tissue characterization of in vivo atherosclerotic plaques by intravascular optical coherence tomography images," Biomed. Opt. Express 4, 1014-1030 (2013).

5.2 Deep Learning Model Comparison

The feasibility shown in the first paper was extended to the next publication, with the same aim of finding the best possible deep learning models. Here, the comparison focused on models that have been proven for use in computer vision. The models include VGG [119], ResNet [108], and DenseNet [109]. They were evaluated to identify five tissue types, namely bone, bone marrow, fat, muscle, and skin tissues. Different input features (the intensity-based image patch, A-Scan profile, and attenuation map) were also combined. The attenuation map was extracted using the depth-resolved attenuation coefficient extraction method (see Section 3.1.1). The results show that the DenseNet161 model, with combined intensity-based and attenuation map as the input, has the highest average accuracy of 94.85 % and F1-score of 94.67 %.

Publication: Y. A. Bayhaqi, A. Hamidi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Deep learning models comparison for tissue classification using optical coherence tomography images: toward smart laser osteotomy", *OSA Continuum*, Vol. 4, No. 9, pp. 2510-2526, 2021.

Deep learning models comparison for tissue classification using optical coherence tomography images: toward smart laser osteotomy

YAKUB A. BAYHAQI,[1,4] Ⓘ ARSHAM HAMIDI,[1] Ⓘ FERDA CANBAZ,[1] Ⓘ ALEXANDER A. NAVARINI,[2] PHILIPPE C. CATTIN,[3] Ⓘ AND AZHAR ZAM[1,5] Ⓘ

[1]*Biomedical Laser and Optics Group (BLOG), Department of Biomedical Engineering, University of Basel, 4123 Allschwil, Switzerland*
[2]*Digital Dermatology Group, Department of Biomedical Engineering, University of Basel, 4123 Allschwil, Switzerland*
[3]*Center for medical Image Analysis and Navigation (CIAN), Department of Biomedical Engineering, University of Basel, 4123 Allschwil, Switzerland*
[4]*yakub.bayhaqi@unibas.ch*
[5]*azhar.zam@unibas.ch*

Abstract: We compared deep learning models as a basis for OCT image-based feedback system for smart laser osteotomy. A total of 10,000 OCT image patches were acquired ex-vivo from pig's bone, bone marrow, fat, muscle, and skin tissues. We trained neural network models using three different input features (the texture, intensity profile, and attenuation map). The comparison shows that the DenseNet161 model with combined input has the highest average accuracy of 94.85% and F1-score of 94.67%. Furthermore, the results show that our method improved the accuracy of the models and the feasibility of identifying tissue types from OCT images.

1. Introduction

Bone surgery (osteotomy) has been practiced for treatment of various bone disorders in centuries. The instruments used in such surgeries have not changed much over time, such as saws, drills, chisels, and hammers). The primary mechanism generally is by applying mechanical stress to the bone surface until the instrument break the bone, producing high pressure, friction, and vibration [1]. As a result, there is a tendency to remove more tissue than necessary [2,3]. Additionally, the contact between tissue and the tool's surface increases the risk of bacterial contamination [4].

On the other hand, laser bone surgery (laser osteotomy) is an emerging technique that promises to overcome the limitations of using conventional mechanical tools. Laser osteotomy is a contactless intervention tool that is capable of delivering a precise cut, which, in turn, reduces tissue loss during the ablation procedure, and supports faster tissue healing [4]. The non-contact application of the laser osteotomy reduces the potential risk of bacterial contamination [5]. Moreover, the photothermal effect after absorption of a laser (such as Er:YAG laser) beam causes micro explosions and breakup of bacteria. This effect subsequently leads to bacterial death in the ablation spot [6–8].

Furthermore, by embedding the laser in a robotic arm, small complex shapes, such as a dovetail, diamond, or circle, are made possible. [2]. However, similar to mechanical tools, laser osteotomy may also cause collateral damage if critical tissues such as nerve and bone marrow are not avoided. Thus, a feedback mechanism capable of distinguishing tissue types during ablation becomes a critical focus for research. Several studies have explored tissue type identification based on physical feedback properties. Non-invasive methods such as Raman [9–11], autofluorescence

[12–16], and diffuse reflectance spectroscopy [17–19] have been demonstrated to have a high sensitivity to differentiate tissue type. Nevertheless, these techniques work in point measurement, thus, losing the tissue's margin or morphological information. Other methods such as acoustic shock waves and laser-induced breakdown spectroscopy also have the potential to differentiate tissue types [20–25]. However, these methods distinguish tissue type only after an ablation pulse has been applied, thus, increase the risk of collateral damage during laser ablation.

Our project proposed a non-destructive tissue identification system based on Optical Coherence Tomography (OCT) imaging. OCT is the most applicable tool for margin detection and also provides morphological information [26]. OCT offers an alternative approach to real-time tissue classification, It is rapidly becoming the preferred method for real-time in-vivo investigation of thin tissues and subsurface imaging because of its high resolution and non-invasive nature [26–28]. Several studies combining OCT and laser ablation treatment have reported experiments demonstrating OCT-guided laser surgical concepts for clinical application. The combinations reportedly increased resection accuracy and precision for brain tumors and blood coagulation [29–31]. This imaging system implemented an intra-operative diagnosis by providing positional feedback and tissue specificity during surgery. In our approach, we extend the application of the OCT-guided laser surgical concepts for smart laser osteotomy. In this manner, we focused more on discriminating between the bone tissue and the surrounding tissues. We wanted to avoid damaging the nerve and bone marrow (inside the bone), which may lead to complications if it is accidentally cut. Other than that, the feedback system will help to prevent any damage to surrounding tissue (such as muscle and skin) due to accidental patient movement.

Tissue characterization or classification has most recently been the focus of research in the medical application of OCT. Several machine learning approaches, such as random forest [32–34] and support vector machine (SVM) [35,36], have been demonstrated to achieved average accuracy of 80.37% and 96.80%, respectively, for tissue characterization of atherosclerotic plaques (fibrous, calcific, and lipid-rich). These methods involved the attenuation coefficient, statistical and geometrical features of the image. Although, these methods have shown good accuracy for automatic tissue classification. The complicated feature extraction of the image increased the prediction time, which is a critical point for real-time laser ablation monitoring.

As a subset of machine learning, deep learning is developing linearly to the advances of computation technology in recent years. The introduction of the convolutional neural network (CNN) bring a significant breakthrough for automatic medical image analysis and recognition. CNN is often used to classify, retrieve, correct, and segment medical images. In the field of OCT image classification, CNN has been demonstrated to achieve similar or even better classification accuracy and sensitivity with the classical machine learning methods. Previously, Abdolmanafi et al. used the AlexNet and VGG19 models for the intima and media tissue classification [37]. These models were investigated to have an average accuracy of 96.0% and 98.0%, respectively. Furthermore, Gessert et al. reported that the ResNet50V2 and DenseNet121 models have an average accuracy of 91.3% and 91.0%, respectively, to classify the plaque area as lipid tissue, fibrous tissue, and calcified tissue [38]. Although deep learning model required an extensive amount of training data and more extended training time. Nevertheless, the deep learning model performs faster image classification by skipping the feature extraction process, consequently enabling the implementation of real-time tissue classification.

In this paper, we compared well-established CNN models to classify several healthy tissue types, the first application of its kind to the best of our knowledge. We focused on distinguishing the tissue types most likely to be encountered during laser osteotomy, such as bone, bone marrow, fat, muscle, and skin. As a preliminary experiment, we investigated the CNN's ability to classify tissues in the absence of the ablation laser. In the future, we foresee integrating the CNN model for real-time tissue differentiation during laser osteotomy.

2.   Materials and methods

The ultimate aim of the smart laser osteotomy is to provide feedback on the tissue types which would be encountered during laser ablation to avoid cutting critical tissues. Here, we use an OCT imaging system to monitor tissue anatomy at the subsurface level during laser ablation. A patch image taken from the ablation area is used as input of the CNN model to discriminate between tissue types. The output of the CNN model provides feedback for the ablation laser to either stop or continue ablating. Our smart laser osteotomy concept is illustrated in Fig. 1. This paper aimed to find the CNN model with high accuracy in distinguishing tissue type and short prediction time. The output of the CNN will be used as feedback to close or open an optical shutter for controlling the ablation laser (e.g Er:YAG or Nd:YAG Laser). However, as a preliminary experiment, we investigated the CNN's ability to classify tissue in the absence of the ablation laser.
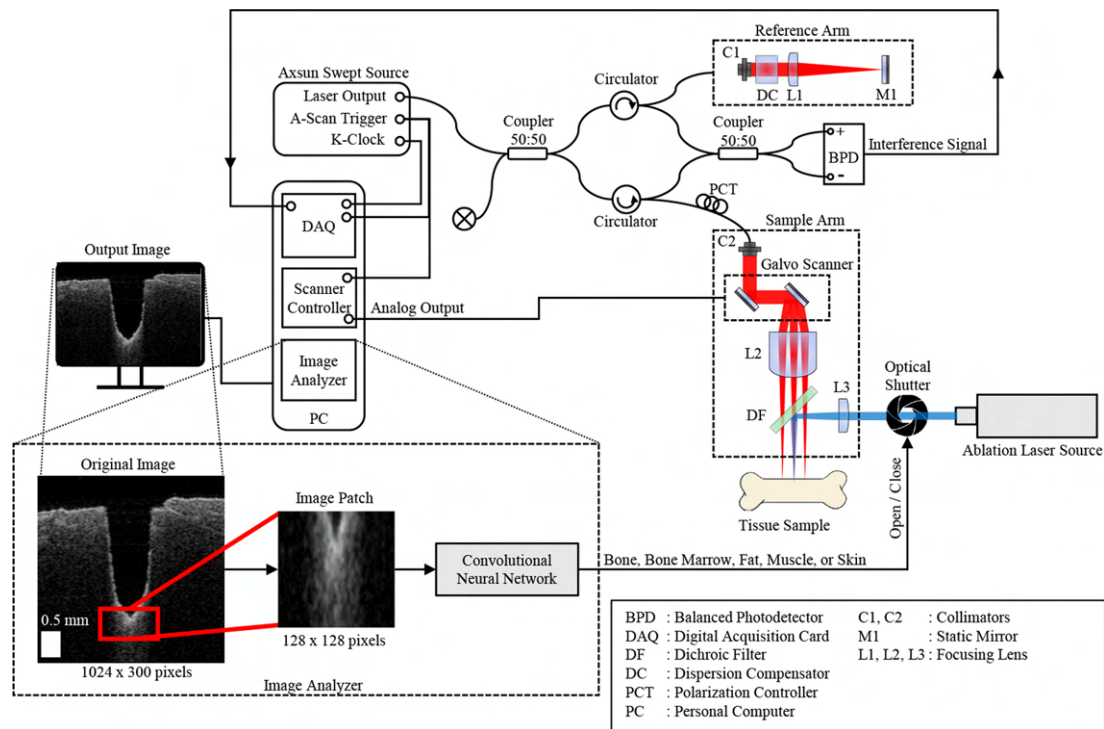
Fig. 1. Schematic of the proposed OCT-based smart laser surgery system. We used a Fourier-Domain OCT with an Axsun swept-source laser. The OCT laser (red line) is coupled with an ablation laser (blue line) by a dichroic filter. The OCT images are streamed to monitor the ablation process. A region of interest (image patch) from the OCT image is selected on the ablation spot. We trained a convolutional neural network model to identify tissue type based on the extracted image patch. The convolutional neural network's output provides feedback to an optical shutter and controls the ablation laser to either stop or continue ablation.

2.1.   OCT imaging system

OCT image acquisition is the starting point of the proposed smart laser surgery system. We used a custom OCT system equipped with an Axsun swept-source laser. The laser source operated at a central wavelength (λ_c) of 1046 nm, a spectral bandwidth of 112.15 nm, and an A-scan line rate of 100.16 kHz. The imaging range of the OCT system was equal to 3.6 mm in the air. The OCT system provided a field of view of 4.8×4.8 mm^2. The image size acquired with this OCT system was 1024 pixels in height and 300 pixels in width. The sensitivity (SNR_{max}) of our OCT

system was 96.46 dB. The corresponding lateral and axial resolutions were 44μm and 10μm, respectively.

During acquisition, frame-averaging method was used to acquire high-quality images by averaging several image frames from a single sample location on the tissue. This method reduced the noise originating from random interference signals and natural bandpass filter problems while still preserving the speckle signal as the primary information carrier [39]. We ensured that the tissues were statically placed during image acquisition to avoid motion artifacts (a potential disadvantage of the averaging method).

2.2. Image patch extraction

The next step of our smart laser osteotomy was image patch extraction from OCT image. The patch was selected in such a way to represent a region of interest where a destructive laser pulse would be applied and used as input to the CNN to identify tissue types. We defined the ablation spot as always in the lateral center of the image. Vertical Canny edge detection method was used to trace the tissue surface in axial direction because of its simplicity and low sensitivity to noise [40,41]. After getting the ablation spot's lateral and axial location, a square 128 x 128 pixels image patch was extracted with the ablation spot as the top center of the image patch. We defined the image patch as the texture feature and used it as the input of the CNN model.

2.3. Neural network frameworks

This study aimed to find a CNN model capable of accurately and efficiently classifying tissue type based on the extracted image patch. We evaluated well-established CNN models to find the best model with the highest accuracy and fastest processing speed. In this study, we use four main primary CNN models and define them as the base models. The first base model was the AlexNet, developed by Alex et al. [42]. The second base model was the deep Visual Geometry Group model (VGG), used by Simonyan & Zisserman for image-based object recognition [43]. The third base model was the deep Residual Network model (ResNet), shown by He et al. [44]. The last base model was the deep Densely Connected Network model (DenseNet), proposed by Huang et al. [45].

We also exploited the variants for each CNN base model except for the AlexNet model. We evaluate the VGG based model with the variance of 11, 13, 16, and 19 -layers (VGG11, VGG13, VGG16, and VGG19). Additionally, we evaluated the VGG model with shallower depth by removing the last two convolutional blocks (we defined it as VGG-3Block). The VGG-3Block framework was similar to VGG13, but we used only the first three convolutional blocks and directly connected the convolutional layer's output to the fully connected layer. Additionally, we also evaluated the effect of removing the fully connected layer (defined as VGG16-A) or adding two more fully connected layers (defined as VGG16-B) at the end of the VGG16. Furthermore, we evaluated the ResNetV2 model with the variance of 50, 101, and 152 residual layers depth (ResNet50V2, ResNet101V2, and ResNet152V2). Finally, we evaluated the DenseNet model with variance of 121, 161, 169, and 201 dense layers (DenseNet121, DenseNet161, DenseNet169, and DenseNet201).

2.4. Feature fusion

Intuitively, we could train the CNN models directly to classify tissue type based on the texture feature alone. However, previous studies suggest additional features to support the texture feature could improve the tissue classification accuracy. Rico-Jimenez et al. proposed using a single A-line (defined as intensity profile feature) to characterize atherosclerotic plaques for a faster prediction process and reported to achieve an average classification accuracy of 85% [46]. In this study, we believed that combining the texture and the profile features will improve the accuracy of the CNN models. Therefore, our first approach was combining the texture feature with the

intensity profile feature which was extracted from the middle of the image patch. The extraction and combination process for the first approach is illustrated in Fig. 2. In the beginning, we independently trained two CNN models. We trained the first model (CNN Texture model) to classify tissue type by using the 2-dimensional texture feature as input. Concurrently, we trained the second model (CNN Profile model) by using a 1-dimensional intensity profile as input. Both models were constructed using the same base model architecture. However, for the second model, we changed the input size of the network to accept 1-dimensional input and used the 1-dimensional convolutional layers instead of 2-dimensional layers. After training both CNN models, we fused them to concatenate each output to an intermediate layer. The intermediate layer was a fully-connected layer with 4096 neurons. Finally, we retrained the combination to fine-tune the intermediate layer. We defined this combination as combination-A.
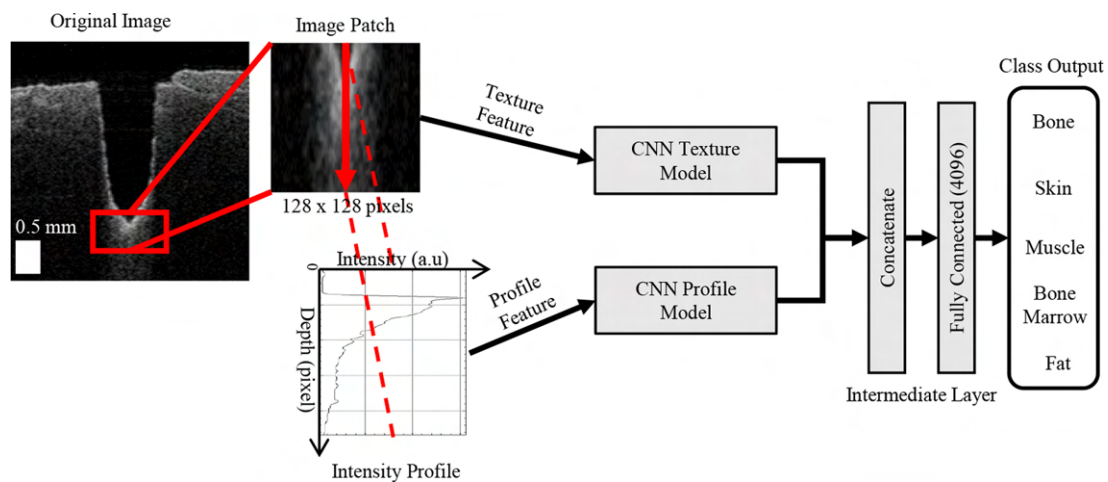
Fig. 2. Combination-A. The CNN texture and CNN profile models were trained separately. The first model was trained to identify tissue type based on the texture feature. The second CNN model was trained to identify tissue type based on the intensity profile feature in the middle of the image patch. Both of the CNN model's outputs were then fused with a concatenation layer. The fused feature was connected to a fully connected layer of 4096. At the end layer, a sigmoid activation function was used to identify the tissue type.

Furthermore, Ughi et al. combined the texture and attenuation coefficient features for atherosclerotic plaque tissue characterization [32]. We also believed that the combination would improve the accuracy of the CNN models. Therefore, in our second approach, we combined the texture feature with a pixel-wise attenuation coefficient map. The extraction of the depth-resolved attenuation coefficient map was proposed by Vermeer et al. to have a pixel-wise attenuation coefficient map from the image [47]. We reconstructed the attenuation map before extracting the image patch. Then, we extracted the attenuation coefficient map patch at the same location as the texture image patch. These two patches were then fused into a 2-channel image and used as the CNN model's input. Figure 3 illustrates the processing of the second approach. We defined this combination as combination-B.

2.5. Attenuation coefficient map reconstruction

The attenuation coefficient maps were reconstructed using the model of light transmission which has been introduced in the work of Vermeer et al. [47]. The light transmission model of an A-Scan is defined with

$$I(z) \cong \beta \, L_0 \, \mu_b(z) \, e^{-2 \int_0^z \mu_a(u) du}, \tag{1}$$

where $I(z)$ is the intensity signal detected at depth z, β defines the analog to digital conversion factor of the digitizer and the detector's quantum efficiency, L_0 is the source light power (Wm^{-2}),
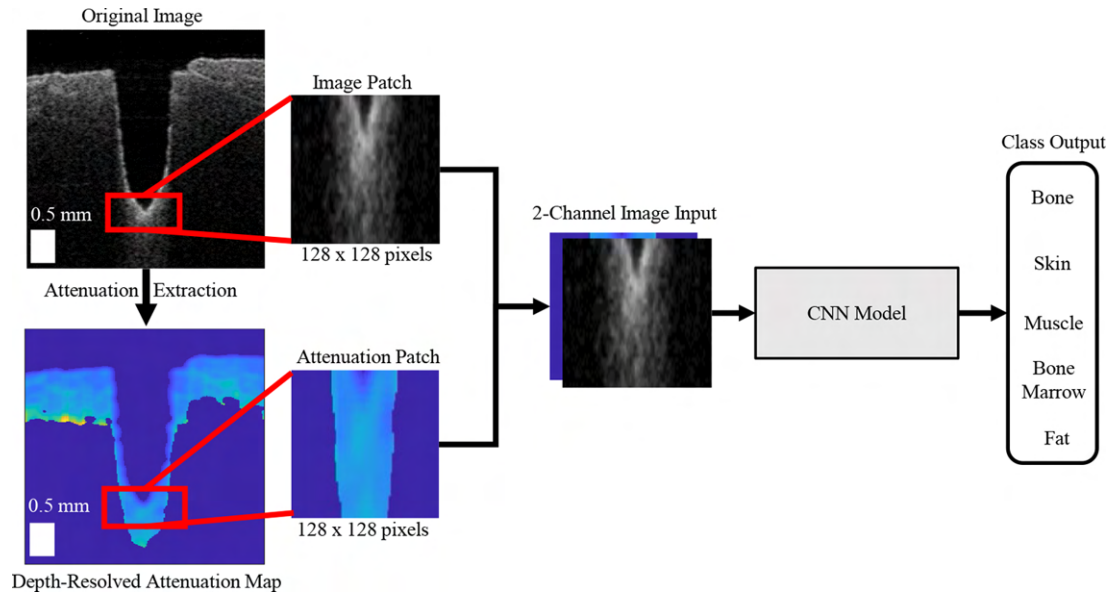
Fig. 3. Combination-B. The CNN model was trained to differentiate tissue type based on 2-channel image input. This 2-channel image was constructed by combining the image patch and the attenuation patch. The attenuation patch was extracted from the attenuation map at same location with image patch.

μ_a denote the attenuation coefficient, and μ_b denote the backscatter coefficient. We assume that the backscatter coefficient is a fixed fraction α of the attenuation coefficient ($\mu_b = \alpha\mu_a$) and the $I(\infty) = 0$. The integral of $I(z)$ from z to infinity (∞) is then given by

$$\int_z^\infty I(u)du \cong -\frac{\alpha\beta L_0}{2}e^{-2\int_0^u \mu(v)dv} + C\Big|_z^\infty$$

$$= -\frac{I(u)}{2\mu_a(u)} + C\Big|_z^\infty \qquad (2)$$

$$= -\frac{I(z)}{2\mu_a(z)}.$$

We can rewrite Eq. (2) to get the estimated attenuation coefficient as

$$\mu_a \cong \frac{I(z)}{2\int_z^\infty I(u)du}. \qquad (3)$$

The pixel-wise form of the estimated attenuation coefficient is given by

$$\mu_a[i,j] \approx \frac{I[i,j]}{2\delta \sum_{z=j+1}^\infty I[i,z]}, \qquad (4)$$

where δ denotes the axial pixel spacing (mm/pixel), i denotes the A-Scan index (horizontal index) over the image, and j denotes the index over the depth of i-th A-Scan (vertical index). Furthermore, the image $I[i,j]$ is first denoised with Gaussian filter to reduce the speckle noise.

2.6. Image data collection

The tissue samples were taken ex-vivo from five individual pigs. For each pig, we took tissue samples from the femur bone. Figure 4 illustrates the femur bone anatomical structure of a pig.

The OCT images were acquired from five different tissue types (bone, bone marrow, fat, muscle, and skin). We prepared 400 tissue samples for each tissue type from random locations around the diaphysis and epiphysis areas. Figure 5 shows examples of the tissue sample. A B-Scan image was acquired for each tissue sample. All of the B-Scan images were enhanced using frame-averaging over 300 frames per image. Therefore, 2000 OCT images were acquired for each pig. In total, we acquired 10000 OCT scan images. Furthermore, the classification was done by using the image patch as the input. We extracted an image patch for each OCT image. The image patch extraction is explained in section 2.2.
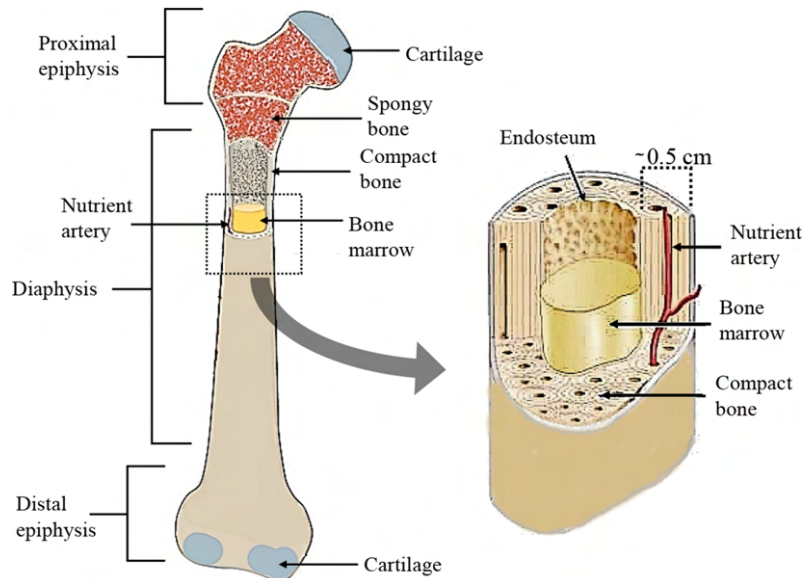
Fig. 4. Illustration of the femur bone anatomical structure of a pig (left). Cross-cut example of the bone on the diaphysis area is illustrated on the right image.

In the experiment, we used leave-one-out of five pigs (5-fold) cross-validation to measure the performance of the CNN models. In one fold, we trained the CNN model with patch images from four animals (8000 image patches) and tested it with one other animal (2000 image patches).

2.7. Models training

We implemented the CNN models in Keras python implementation, with a TensorFlow backend [48]. The training workstation was equipped with an Intel Xeon E5620 processor and two NVIDIA GTX 1080 Ti GPUs. All models were trained with 1000 epochs and a batch size of 32 samples to fit the GPU's memory capacity. We defined cross-categorial entropy as the training loss function and Adam (learning rate = 1.0×10^{-4}) as the training optimizer. We evaluated the classification performance of the CNN models in terms of average cross-validated accuracy and F1-score. The F1-score is a measure of the model's average accuracy for each class (tissue type). The F1-score of a class can be calculated from

$$\%F_1 - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \times 100\,\%, \tag{5}$$

where TP is the number of correctly labelled samples in current class, FP is the number of incorrectly labelled samples as belonging to the current class, and, FN is the number of incorrectly labelled samples as belonging to other class [49].
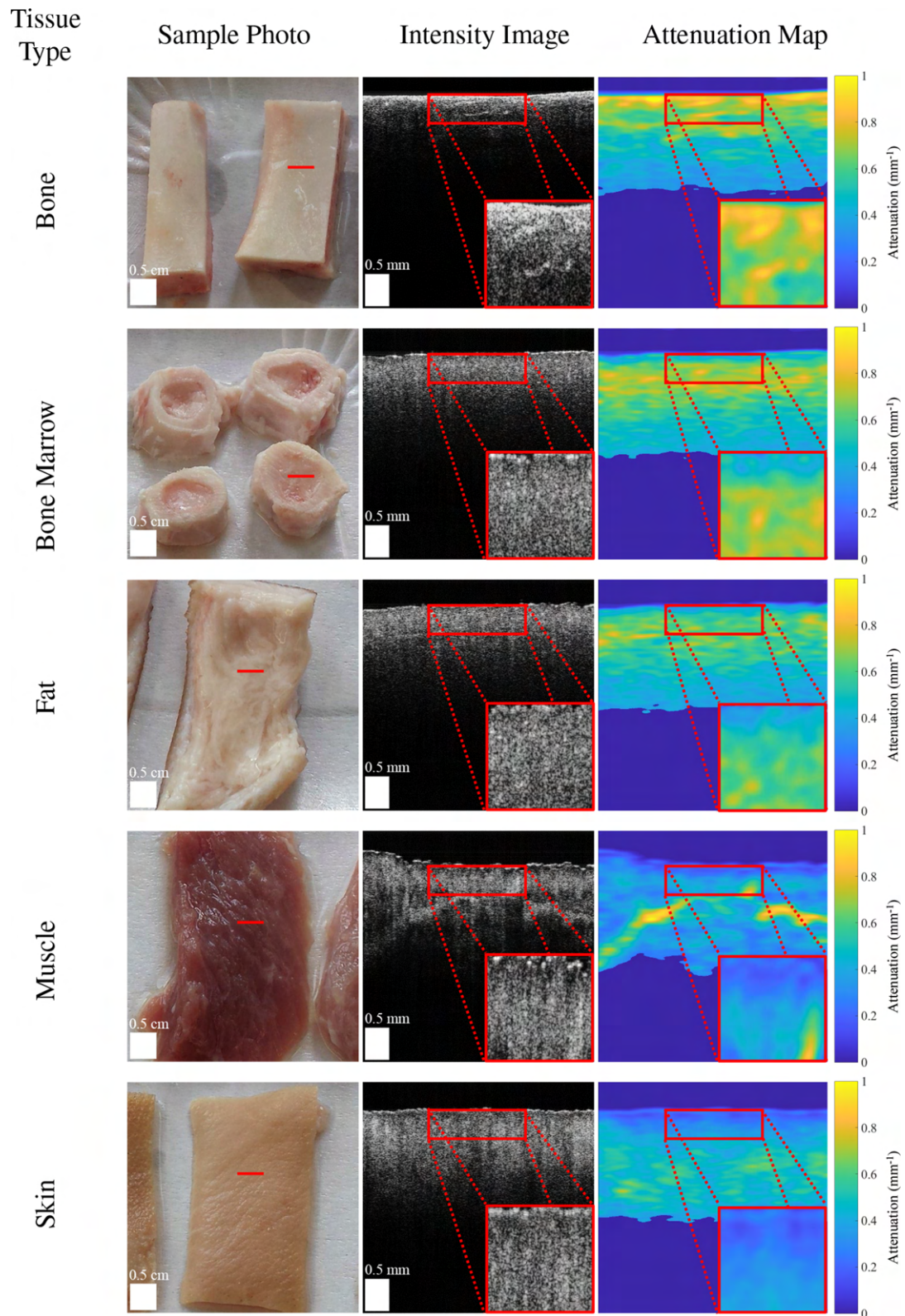
Fig. 5. The left column shows examples of the bone, bone marrow, fat, muscle, and skin tissue samples used in our experiment. The corresponding OCT images (middle column) were scanned on the red line for each tissue sample. The last column (right) shows the attenuation maps reconstructed from the OCT images. The image patches (red box) were taken on the surface of the tissue and used to train the CNN models.

3. Results and discussion

3.1. Classification performance

Our first experiment was to find the optimal CNN models (VGG, ResNet, and DenseNet) by using different number of layers except for the AlexNet model because the VGG model has almost the same configuration as the AlexNet with more convolutional layers and smaller filter sizes. Here the CNN models were trained with the texture feature only. Specifically for the VGG based models, the results show that the top fully connected layers play an important part in classification accuracy. The VGG model has less accuracy and F1-score when we removed the fully connected layer (VGG16-A). However, we observed similar classification performance when we use more fully connected layer (VGG16 and VGG16-B). Other than that, the number of convolutional layer blocks also plays an important part in the VGG model's performance. The VGG-3block model has even less accuracy and F1-score in comparison with the AlexNet. The VGG19 model has the best performance among the VGG models. Furthermore, the results for the ResNet models show that the accuracy and F1-score are higher when we used more number of layers. Finally, we observed that the DenseNet121 model performed best than the other models. However, we also observed a relatively small difference in accuracy and F1-score between the DenseNet models. It is difficult to conclude the relation between the number of layers and the performance improvement of the DenseNet models. Table 1 and 2 further show the accuracy and F1-score of the AlexNet and DenseNet models. The accuracy and F1-score for the other models can be found in Appendix A (Table 5 and 6).

Table 1. Comparison of the average ± standard deviation of accuracy for the AlexNet and DenseNet models trained with the texture, profile, attenuation, and combinations of features (Combination A and B). The highest average accuracies are highlighted in bold.

Model	Accuracy (%)				
	Input Feature				
				Combination	
	Texture	Profile	Attenuation	A	B
AlexNet	89.22 ± 0.35	86.86 ± 0.68	88.65 ± 0.91	92.29 ± 0.16	93.39 ± 0.43
DenseNet121	**91.52 ± 0.28**	**87.83 ± 0.41**	**89.76 ± 0.22**	94.45 ± 0.46	94.57 ± 0.91
DenseNet161	91.04 ± 0.18	87.68 ± 0.49	89.55 ± 0.78	**94.70 ± 0.42**	**94.85 ± 0.88**
DenseNet169	91.30 ± 0.49	87.77 ± 0.42	89.61 ± 0.35	94.20 ± 1.09	94.42 ± 0.37
DenseNet201	91.04 ± 0.58	87.16 ± 0.15	89.68 ± 0.67	93.90 ± 0.55	94.13 ± 0.31

Table 2. Comparison of the average ± standard deviation of F1-score for the AlexNet and DenseNet models trained with the texture, profile, attenuation, and combinations of features (Combination A and B). The highest average F1-scores are highlighted in bold.

Model	F1-Score (%)				
	Input Feature				
				Combination	
	Texture	Profile	Attenuation	A	B
AlexNet	89.31 ± 0.29	86.95 ± 0.66	88.87 ± 0.95	92.35 ± 0.58	93.43 ± 0.14
DenseNet121	**91.60 ± 0.37**	**87.90 ± 0.42**	**89.75 ± 0.07**	94.43 ± 0.44	94.54 ± 0.20
DenseNet161	91.16 ± 0.11	87.76 ± 0.62	89.72 ± 0.44	**94.79 ± 0.24**	**94.67 ± 0.35**
DenseNet169	91.31 ± 0.76	87.85 ± 0.44	89.74 ± 0.18	94.24 ± 0.20	94.40 ± 0.28
DenseNet201	91.12 ± 0.26	87.22 ± 0.15	89.74 ± 0.73	93.91 ± 0.31	94.48 ± 0.30

In the second experiment, we also discussed the classification performance of the CNN models with three different input features (texture, intensity profile, and attenuation map feature). Examples of the attenuation maps are shown in Fig. 5. The results show that all the CNN models achieved higher accuracy by learning the texture feature than learning the profile or attenuation features. We also observed that the CNN models trained with the attenuation feature have higher accuracy and F1-score than those trained with the profile feature. Although we may discriminate tissue type based on the average attenuation coefficient, the standard deviations indicate noises that reduce the classification performance (see Table 3). Furthermore, the results prove the improvement of accuracy by combining the texture and profile or attenuation features. We observed that combination-B improves the models' performance better than combination-A. The combination-B increases the accuracy and F1-score of all texture feature-based models by an average of 3.59 % and 3.60 %, respectively. On the other hand, the combination-A increases the accuracy and F1-score of all texture feature-based models by an average of 3.00 % and 2.98 %, respectively, which proves that the attenuation map feature discriminates better than the profile feature. These accuracy improvements happened because the CNN models learn the discriminant between bone marrow and fat better when trained with the profile or attenuation features than the texture feature. As an example, Fig. 6 illustrates the benefit of using the combinations to improve the DenseNet121 models.

Table 3. The average ± standard deviation of the tissues' attenuation coefficient. The average attenuation coefficient was measured based on the reconstructed attenuation coefficient map patches

Tissue Type	Attenuation Coefficient (mm^{-1})
Bone	0.995 ± 0.065
Bone Marrow	0.933 ± 0.179
Fat	0.866 ± 0.107
Muscle	0.849 ± 0.167
Skin	0.213 ± 0.116

3.2. Computation Performance

The computation performances of the models were evaluated for the average prediction time to classify the image patches. We tested the computational performance of the models on the same workstation that we used to train the models. The prediction time measurements were the average prediction time of five prediction runs for each model. With the texture feature alone as the input, both the AlexNet and VGG based models predicted the tissue type in less than 40 msec. We observed that the computation time of the VGG based models does not significantly increase along with the increased number of layers. Furthermore, the ResNet and DenseNet models predicted the tissue type in around 40~70 msec. Here, the ResNet and DenseNet models' prediction time increased along with the increasing number of layers. The AlexNet model predicted the tissue type faster than the models because it used fewer layers and less network complexity (parameters).

We also measured the prediction time when using the profile and attenuation features as the models' input. The models predicted the tissue type faster when we used the profile feature than the texture and attenuation features as input because of the smaller number of parameters. Additional delay for attenuation map extraction also increased the prediction time when we used the attenuation feature as the input. The attenuation map extraction delayed the prediction time by 18.76 ± 1.36 msec. On the other hand, there is no specific processing time for extracting the intensity profile. Furthermore, the previous section demonstrated that combining the texture

(a) Texture Feature

(b) Profile Feature

(c) Attenuation Feature

(d) Combination A

(e) Combination B
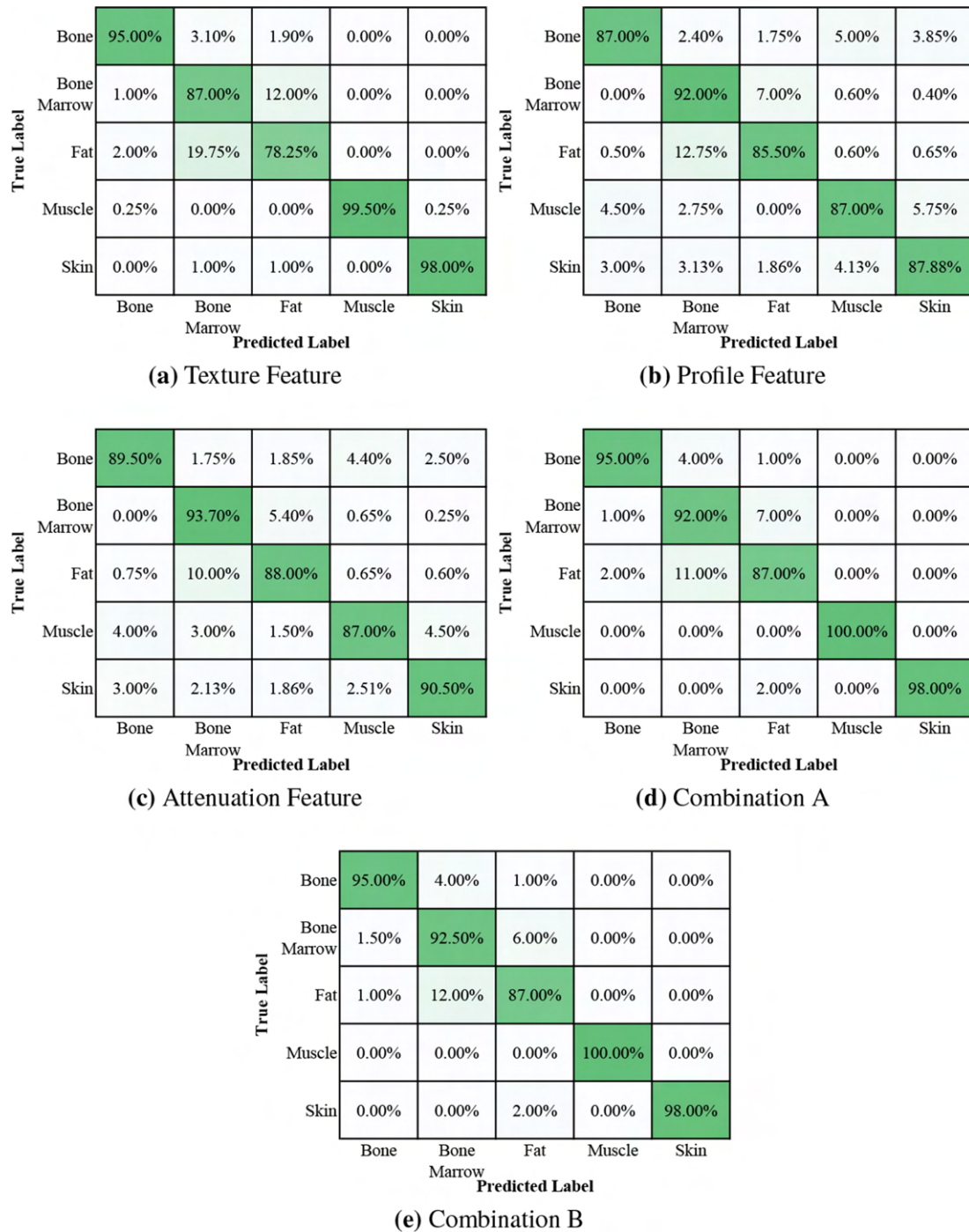
Fig. 6. The test confusion matrix of the DenseNet121 models that were trained with the texture feature (a), the profile feature (b), and the attenuation of both features (c). The model trained with texture feature have lower accuracy in classifying bone marrow and fat. On the other hand, the profile and attenuation feature discriminate better for the bone marrow and fat. Therefore, the model has higher accuracy in the combination A and B (d) and (e).

feature with the profile or attenuation features significantly increased the classification accuracy and sensitivity. In principle, the prediction time for both combination input models (combination A and B) were similar to those trained with the texture feature alone. However, in combination-B, the extraction of the attenuation map put an additional delay in the prediction time. Table 4 show the computation performance of the AlexNet and DenseNet models. The computation performance for the other models can be found in Appendix A (Table 7).

Table 4. Comparison of the average ± standard deviation of the prediction time for the AlexNet and DenseNet models trained with the proposed features. The prediction time includes the attenuation coefficient map extraction time for the combination-B model and the model with attenuation maps input. The attenuation coefficient map extraction takes 18.76 ± 1.36 msec. The fastest computation performances are highlighted in bold.

Model	Prediction Time (*msec*)				
	Input Feature				
				Combination	
	Texture	Profile	Attenuation	A	B
AlexNet	**34.26 ± 2.88**	**29.89 ± 8.43**	**52.60 ± 6.17**	**34.40 ± 2.52**	**53.52 ± 1.11**
DenseNet121	52.73 ± 1.89	57.25 ± 2.83	71.84 ± 2.27	53.37 ± 2.16	71.49 ± 6.71
DenseNet161	59.44 ± 3.43	57.25 ± 3.52	78.49 ± 7.86	59.23 ± 3.62	78.20 ± 3.23
DenseNet169	62.78 ± 7.67	58.70 ± 6.12	82.52 ± 3.53	63.62 ± 4.32	81.54 ± 4.56
DenseNet201	64.98 ± 2.68	56.16 ± 3.85	84.34 ± 6.78	63.61 ± 2.98	83.74 ± 3.69

All models' computation performances were below 90 msec, excluding the image acquisition and frame-averaging. Our results are faster than the reference machine learning methods [32–36]. These results show the possibility of achieving real-time tissue classification with the optimum pulse repetition rate of our laser ablation was 10 Hz (100 msec per pulse) [27]. Nevertheless, the results show that the AlexNet model predicted the tissue type fastest than the other models even when using the combined features as the input. However, we prioritize more on the classification accuracy in this study. Therefore, within the pulse repetition rate window, the DenseNet161 with the texture and attenuation feature combination input would be the best option to predict the tissue type.

4.   Conclusions and outlook

We demonstrated that the classification accuracy was significantly increased by combining the texture feature with the intensity profile or attenuation map features. Combining the texture feature with the attenuation map improved the classification accuracy by an average of 3.59%. Meanwhile, the combination with the profile features improved the classification accuracy by an average of 3.00%. The difference in accuracy's improvement between both combinations was relatively small, especially between the DenseNet models. However, the attenuation feature extraction delays the prediction time by 18.76 msec. Therefore, in practice, the combination between the texture and intensity profile is preferable to the combination between the texture and attenuation map.

The results for the combination between the texture and intensity profile show that DenseNet models have higher accuracy than the other models. Specifically, the DenseNet161 model has the highest accuracy compared with the other models. However, the complexity of the DenseNet161 model also increased the prediction time. Although the AlexNet model has 1.46% less accuracy than the DenseNet161, this model's prediction time is ~1.43 times faster. If low computation resource is available, such as in an embedded system, the AlexNet model would be a better choice than the DenseNet161 model. For our smart laser surgery, the computation time required by all models is still carried below the optimum ablation laser pulse rate, which suggests the

application for real-time feedback. This study suggests that the DenseNet161 with the texture and attenuation feature combination input would be the best option to predict the tissue type.

The frame-averaging image enhancement will delay the feedback system by ~900 msec for acquiring 300 frames (B-Scan) per image in this study. We would consider using a smaller number of frames for a faster feedback system in the real-time implementation. However, the reconstructed images will have lower image quality compared to the images used in this study. Consequently, the CNN may have lower accuracy if trained using the images with lower image quality. Several studies suggested faster OCT image denoising, such as block-matching 3D (BM3D) [50], double-density dual-tree complex wavelet transform (DD-CDWT) [51], or deep learning denoising [52] methods, as an alternative to the frame-averaging method. These methods are demonstrated to be able to reconstruct a high-quality image from a single raw image. In such a way, we could directly train the CNN using the denoised raw images to identify tissue type. However, there is a possibility that the denoising method will alter the important information (such as textural and attenuation features) for tissue classification. Further investigation is needed to ensure that the CNN performs similarly with this study.

Furthermore, we are aware that implementing this method and incorporating it into a laser ablation system remains challenging. One of the challenges is that the tissue will experience an increase in temperature (heating up) during microsecond ablation. The area of interest for the tissue classification would experience heat transfer from the ablated tissue in the focal spot area. This process happens in a fraction of a second before the ablated tissue in the focal spot area explodes or evaporates and is confined to the laser's short pulse duration. Therefore, this indirect heating process can be considered slow heating.

Additionally, the optical properties of the tissue (such as refractive index, absorption coefficient, and scattering coefficient) would change due to the heating process [18]. These changes will affect the tissue classifier's (CNN) accuracy. Further studies are needed to determine how the CNN performs during laser ablation. We planned to train and test the performance of the CNN models with two approach conditions. In the first approach, we will train and test the CNN modes using the images that are collected in controlled temperature conditions. For example, we would collect the OCT images while heating the tissue in every 5 °C increment up to 100 °C. The second approach is by directly collecting OCT images during laser ablation and use them for training and test the CNN models.

Finally, for in-vivo experiments, we will also face challenges such as bleeding and tissue debris from the ablated tissue, which might induce artifacts to the OCT image. Therefore, we will also investigate the feasibility of integrating a cooling system such as pressurized air and/or water irrigation with laser ablation in our future work. The cooling system will be helpful to maintain the tissue temperature and cleaned the ablation area from bleeding and tissue debris.

Appendix A: Tables on accuracy, F1-score, and computation time for all models

Table 5. The average ± standard deviation of accuracy for all models. The highest average accuracies are highlighted in bold.

Model	Accuracy (%)				
	Input Feature				
				Combination	
	Texture	Profile	Attenuation	A	B
AlexNet	89.22 ± 0.35	86.86 ± 0.68	88.65 ± 0.91	92.29 ± 0.16	93.39 ± 0.43
VGG 3Block	87.37 ± 0.13	83.26 ± 0.24	85.06 ± 0.26	92.10 ± 0.81	92.93 ± 0.35
VGG11	89.57 ± 0.22	84.38 ± 0.67	86.46 ± 0.52	92.28 ± 0.28	93.51 ± 0.27
VGG13	90.14 ± 0.63	85.61 ± 0.56	87.68 ± 0.76	93.26 ± 0.35	94.01 ± 0.60
VGG16-A	87.13 ± 0.18	82.47 ± 0.44	84.34 ± 1.01	90.85 ± 0.19	91.87 ± 0.32
VGG16	90.86 ± 0.58	87.72 ± 0.88	89.64 ± 0.59	92.86 ± 0.41	94.07 ± 0.74
VGG16-B	90.83 ± 0.70	87.63 ± 0.21	89.52 ± 0.65	93.59 ± 0.37	94.05 ± 0.61
VGG19	91.27 ± 0.81	87.51 ± 0.71	89.27 ± 1.04	93.56 ± 0.28	94.42 ± 0.22
ResNet50V2	90.78 ± 0.36	87.07 ± 0.98	89.25 ± 1.08	93.89 ± 0.56	94.06 ± 0.23
ResNet101V2	91.31 ± 0.84	87.51 ± 1.15	89.55 ± 0.15	93.92 ± 0.22	94.25 ± 0.44
ResNet152V2	91.52 ± 0.99	87.57 ± 0.66	89.54 ± 0.22	94.16 ± 0.61	94.32 ± 0.74
DenseNet121	**91.52 ± 0.28**	**87.83 ± 0.41**	**89.76 ± 0.22**	94.45 ± 0.46	94.57 ± 0.91
DenseNet161	91.04 ± 0.18	87.68 ± 0.49	89.55 ± 0.78	**94.70 ± 0.42**	**94.85 ± 0.88**
DenseNet169	91.30 ± 0.49	87.77 ± 0.42	89.61 ± 0.35	94.20 ± 1.09	94.42 ± 0.37
DenseNet201	91.04 ± 0.58	87.16 ± 0.15	89.68 ± 0.67	93.90 ± 0.55	94.13 ± 0.31

Table 6. The average ± standard deviation of F1-score for all models. The highest average F1-scores are highlighted in bold.

Model	F1-Score (%)				
	Input Feature				
				Combination	
	Texture	Profile	Attenuation	A	B
AlexNet	89.31 ± 0.29	86.95 ± 0.66	88.87 ± 0.95	92.35 ± 0.58	93.43 ± 0.14
VGG 3Block	87.39 ± 0.11	83.41 ± 0.24	85.25 ± 0.28	92.39 ± 0.71	93.00 ± 0.49
VGG11	89.64 ± 0.22	84.46 ± 0.69	86.57 ± 0.23	92.36 ± 0.57	93.61 ± 0.72
VGG13	90.22 ± 0.93	85.78 ± 0.53	87.69 ± 0.31	93.34 ± 0.48	94.05 ± 0.52
VGG16-A	87.26 ± 0.17	82.64 ± 0.39	84.47 ± 0.17	90.91 ± 0.65	91.99 ± 0.24
VGG16	91.06 ± 0.61	87.81 ± 0.85	89.65 ± 0.28	92.96 ± 0.27	94.12 ± 0.31
VGG16-B	90.93 ± 0.71	87.74 ± 0.21	89.58 ± 0.19	93.62 ± 0.34	94.53 ± 0.89
VGG19	91.34 ± 0.81	87.57 ± 0.71	89.32 ± 0.08	93.60 ± 1.04	94.47 ± 0.32
ResNet50V2	90.89 ± 0.27	87.14 ± 0.97	89.37 ± 1.14	93.90 ± 0.28	94.13 ± 0.28
ResNet101V2	91.38 ± 0.83	87.59 ± 1.18	89.64 ± 0.15	93.95 ± 0.25	94.32 ± 0.58
ResNet152V2	91.58 ± 0.96	87.60 ± 0.65	89.60 ± 0.22	94.19 ± 0.51	94.35 ± 0.31
DenseNet121	**91.60 ± 0.37**	**87.90 ± 0.42**	**89.75 ± 0.07**	94.43 ± 0.44	94.54 ± 0.20
DenseNet161	91.16 ± 0.11	87.76 ± 0.62	89.72 ± 0.44	**94.79 ± 0.24**	**94.67 ± 0.35**
DenseNet169	91.31 ± 0.76	87.85 ± 0.44	89.74 ± 0.18	94.24 ± 0.20	94.40 ± 0.28
DenseNet201	91.12 ± 0.26	87.22 ± 0.15	89.74 ± 0.73	93.91 ± 0.31	94.48 ± 0.30

Table 7. Comparison of the average ± standard deviation of the prediction time for all models. The prediction time includes the attenuation coefficient map extraction time for the combination-B model and the model with attenuation maps input. The attenuation coefficient map extraction takes 18.76 ± 1.36 msec. The fastest computation performances are highlighted in bold.

Model	Prediction Time (*msec*)				
	Input Feature				
	Texture	Profile	Attenuation	Combination	
				A	B
AlexNet	**34.26 ± 2.88**	29.89 ± 8.43	**52.60 ± 6.17**	**34.40 ± 2.52**	**53.52 ± 1.11**
VGG-3Block	38.59 ± 1.13	30.52 ± 2.32	57.08 ± 2.91	38.09 ± 7.15	56.87 ± 6.51
VGG11	35.37 ± 2.96	28.65 ± 7.55	55.38 ± 2.37	35.89 ± 2.35	54.63 ± 0.99
VGG13	37.00 ± 2.78	35.88 ± 2.30	56.55 ± 3.13	37.67 ± 2.98	55.89 ± 2.25
VGG16-A	38.44 ± 8.11	33.61 ± 1.54	56.99 ± 1.73	38.68 ± 2.48	56.29 ± 1.08
VGG16	37.77 ± 1.76	**27.80 ± 2.04**	56.17 ± 2.86	38.47 ± 3.32	57.33 ± 4.43
VGG16-B	38.63 ± 2.93	34.81 ± 2.59	57.51 ± 1.92	39.25 ± 2.16	57.78 ± 0.35
VGG19	38.55 ± 2.78	42.63 ± 0.67	57.86 ± 0.81	39.51 ± 3.53	59.80 ± 2.38
ResNet50V2	44.41 ± 1.43	41.85 ± 2.76	64.00 ± 2.47	44.57 ± 6.86	63.17 ± 2.85
ResNet101V2	51.11 ± 2.58	58.26 ± 3.55	70.26 ± 2.83	51.07 ± 2.43	69.87 ± 4.52
ResNet152V2	60.80 ± 2.47	50.08 ± 1.82	79.16 ± 3.71	60.48 ± 6.74	79.56 ± 1.60
DenseNet121	52.73 ± 1.89	57.25 ± 2.83	71.84 ± 2.27	53.37 ± 2.16	71.49 ± 6.71
DenseNet161	59.44 ± 3.43	57.25 ± 3.52	78.49 ± 7.86	59.23 ± 3.62	78.20 ± 3.23
DenseNet169	62.78 ± 7.67	58.70 ± 6.12	82.52 ± 3.53	63.62 ± 4.32	81.54 ± 4.56
DenseNet201	64.98 ± 2.68	56.16 ± 3.85	84.34 ± 6.78	63.61 ± 2.98	83.74 ± 3.69

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time for further continuation of the study. However, the data may be obtained from the authors upon reasonable request.

References

1. J. Dabis, O. Templeton-Ward, A. E. Lacey, B. Narayan, and A. Trompeter, "The history, evolution and basic science of osteotomy techniques," Strateg. Trauma Limb Reconstr. **12**(3), 169–180 (2017).
2. K.-W. Baek, W. Deibel, D. Marinov, M. Griessen, M. Dard, A. Bruno, H.-F. Zeilhofer, P. Cattin, and P. Juergens, "A comparative investigation of bone surface after cutting with mechanical tools and er:yag laser," Lasers Surg. Med. **47**(5), 426–432 (2015).
3. S. Kondo, Y. Okada, H. Iseki, T. Hori, K. Takakura, A. Kobayashi, and H. Nagata, "Thermological study of drilling bone tissue with a high-speed drill," Neurosurgery **46**(5), 1162–1168 (2000).
4. S. Stübinger, "Advances in bone surgery: the er:yag laser in oral surgery and implant dentistry," Clin., Cosmet. Invest. Dent. **2**, 47–62 (2010).
5. C. Duverney, H. Abbasi, M. Berkelaar, K. Pelttari, P. C. Cattin, A. Barbero, A. Zam, and G. Rauter, "Sterile tissue ablation using laser light–system design, experimental validation, and outlook on clinical applicability," J. Med. Devices **15**(1), 011104 (2021).
6. M. Wilson, "Photolysis of oral bacteria and its potential use in the treatment of caries and periodontal disease," J. Appl. Bacteriol. **75**(4), 299–306 (1993).
7. Y. Ando, A. Aoki, H. Watanabe, and I. Ishikawa, "Bactericidal effect of erbium yag laser on periodontopathic bacteria," Lasers Surg. Med. **19**(2), 190–200 (1996).
8. H. Yamaguchi, K. Kobayashi, R. Osada, E.-i. Sakuraba, T. Nomura, T. Arai, and J. Nakamura, "Effects of irradiation of an erbium: Yag laser on root surfaces," J. Periodontol. **68**(12), 1151–1155 (1997).
9. N. M. Ralbovsky and I. K. Lednev, "Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning," Chem. Soc. Rev. **49**(20), 7428–7453 (2020).

OSA *CONTINUUM*

10. E. Cordero, I. Latka, C. Matthäus, I. Schie, and J. Popp, "In-vivo raman spectroscopy: from basics to applications," J. Biomed. Opt. **23**(07), 1 (2018).

11. P. C. Ashok, M. E. Giardini, K. Dholakia, and W. Sibbett, "A raman spectroscopy bio-sensor for tissue discrimination in surgical robotics," J. Biophotonics **7**(1-2), 103–109 (2014).

12. F. Stelzle, M. Rohde, M. Riemann, N. Oetter, W. Adler, K. Tangermann-Gerk, M. Schmidt, and C. Knipfer, "Autofluorescence spectroscopy for nerve-sparing laser surgery of the head and neck–the influence of laser-tissue interaction," Lasers Med. Sci. **32**(6), 1289–1300 (2017).

13. F. Stelzle, C. Knipfer, W. Adler, M. Rohde, N. Oetter, E. Nkenke, M. Schmidt, and K. Tangermann-Gerk, "Tissue discrimination by uncorrected autofluorescence spectra: A proof-of-principle study for tissue-specific laser surgery," Sensors **13**(10), 13717–13731 (2013).

14. A. Zam, J. Franke, and M. Merklein, *Optical Tissue Differentiation for Sensor-controlled Tissue-specific Laser Surgery* (Meisenbach, 2011).

15. R. Gunaratne, J. Goncalves, I. Monteath, R. Sheh, M. Kapfer, R. Chipper, B. Robertson, R. Khan, D. Fick, and C. N. Ironside, "Wavelength weightings in machine learning for ovine joint tissue differentiation using diffuse reflectance spectroscopy (drs)," Biomed. Opt. Express **11**(9), 5122–5131 (2020).

16. F. Fanjul-Vélez, S. Pampín-Suárez, and J. L. Arce-Diego, "Application of classification algorithms to diffuse reflectance spectroscopy measurements for ex vivo characterization of biological tissues," Entropy **22**(7), 736 (2020).

17. R. Gunaratne, I. Monteath, J. Goncalves, R. Sheh, C. N. Ironside, M. Kapfer, R. Chipper, B. Robertson, R. Khan, and D. Fick, "Machine learning classification of human joint tissue from diffuse reflectance spectroscopy data," Biomed. Opt. Express **10**(8), 3889–3898 (2019).

18. A. Zam, F. Stelzle, K. Tangermann-Gerk, W. Adler, E. Nkenke, F. W. Neukam, M. Schmidt, and A. Douplik, "In vivo soft tissue differentiation by diffuse reflectance spectroscopy: preliminary results," Phys. Procedia **5**, 655–658 (2010).

19. F. Stelzle, K. Tangermann-Gerk, W. Adler, A. Zam, M. Schmidt, A. Douplik, and E. Nkenke, "Diffuse reflectance spectroscopy for optical soft tissue differentiation as remote feedback control for tissue-specific laser surgery," Lasers Surg. Med. **42**(4), 319–325 (2010).

20. H. Nguendon Kenhagho, G. Rauter, R. Guzman, P. C. Cattin, and A. Zam, "Optoacoustic tissue differentiation using a mach-zehnder interferometer," IEEE Transactions on Ultrason. Ferroelectr. Freq. Control. **66**(9), 1435–1443 (2019).

21. E. Bay, A. Douplik, and D. Razansky, "Optoacoustic monitoring of cutting efficiency and thermal damage during laser ablation," Lasers Med. Sci. **29**(3), 1029–1035 (2014).

22. V. Periyasamy, C. Özsoy, M. Reiss, X. L. Deán-Ben, and D. Razansky, "In vivo optoacoustic monitoring of percutaneous laser ablation of tumors in a murine breast cancer model," Opt. Lett. **45**(7), 2006–2009 (2020).

23. H. Abbasi, L. M. B. Bernal, A. Hamidi, A. Droneau, F. Canbaz, R. Guzman, S. L. Jacques, P. C. Cattin, and A. Zam, "Combined nd:yag and er:yag lasers for real-time closed-loop tissue-specific laser osteotomy," Biomed. Opt. Express **11**(4), 1790–1807 (2020).

24. R. Kanawade, F. Mahari, F. Klämpfl, M. Rohde, C. Knipfer, K. Tangermann-Gerk, W. Adler, M. Schmidt, and F. Stelzle, "Qualitative tissue differentiation by analysing the intensity ratios of atomic emission lines using laser induced breakdown spectroscopy (libs): prospects for a feedback mechanism for surgical laser systems," J. Biophotonics **8**(1-2), 153–161 (2015).

25. F. Mehari, M. Rohde, C. Knipfer, R. Kanawade, F. Klämpfl, W. Adler, N. Oetter, F. Stelzle, and M. Schmidt, "Investigation of laser induced breakdown spectroscopy (libs) for the differentiation of nerve and gland tissue–a possible application for a laser surgery feedback control mechanism," Plasma Sci. Technol. **18**(6), 654–660 (2016).

26. M. E. Brezinski, G. J. Tearney, B. E. Bouma, J. A. Izatt, M. R. Hee, E. A. Swanson, J. F. Southern, and J. G. Fujimoto, "Optical coherence tomography for optical biopsy," Circulation **93**(6), 1206–1213 (1996).

27. L. M. B. Bernal, I. T. Schmidt, N. Vulin, J. Widmer, J. G. Snedeker, P. C. Cattin, A. Zam, and G. Rauter, "Optimizing controlled laser cutting of hard tissue (bone)," at - Autom. **66**(12), 1072–1082 (2018).

28. A. Hamidi, Y. A. Bayhaqi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Long-range optical coherence tomography with extended depth-of-focus: avisual feedback system for smart laser osteotomy," Biomed. Opt. Express **12**(4), 2118–2133 (2021).

29. N. Katta, A. D. Estrada, A. B. McElroy, A. Gruslova, M. Oglesby, A. G. Cabe, M. D. Feldman, R. D. Fleming, A. J. Brenner, and T. E. Milner, "Laser brain cancer surgery in a xenograft model guided by optical coherence tomography," Theranostics **9**(12), 3555–3564 (2019).

30. Y. Fan, B. Zhang, W. Chang, X. Zhang, and H. Liao, "A novel integration of spectral-domain optical-coherence-tomography and laser-ablation system for precision treatment," Int. J. Comput. Assist. Radiol. Surg. **13**(3), 411–423 (2018).

31. F.-Y. Chang, M.-T. Tsai, Z.-Y. Wang, C.-K. Chi, C.-K. Lee, C.-H. Yang, M.-C. Chan, and Y.-J. Lee, "Optical coherence tomography-guided laser microsurgery for blood coagulation with continuous-wave laser diode," Sci. Rep. **5**(1), 16739 (2015).

32. G. J. Ughi, T. Adriaenssens, P. Sinnaeve, W. Desmet, and J. D'Hooge, "Automated tissue characterization of in vivo atherosclerotic plaques by intravascular optical coherence tomography images," Biomed. Opt. Express **4**(7), 1014–1130 (2013).

33. S. Liu, Y. Sotomi, J. Eggermont, G. Nakazawa, S. Torii, T. Ijichi, Y. Onuma, P. W. Serruys, B. P. F. Lelieveldt, and J. Dijkstra, "Tissue characterization with depth-resolved attenuation coefficient and backscatter term in intravascular optical coherence tomography images," J. Biomed. Opt. **22**(09), 1–16 (2017).

34. L. S. Athanasiou, C. V. Bourantas, G. Rigas, A. I. Sakellarios, T. P. Exarchos, P. K. Siogkas, A. Ricciardi, K. K. Naka, M. I. Papafaklis, L. K. Michalis, F. Prati, and D. I. Fotiadis, "Methodology for fully automated segmentation and plaque characterization in intracoronary optical coherence tomography images," J. Biomed. Opt. **19**(2), 026009 (2014).

35. J. Yang, B. Zhang, H. Wang, F. Lin, Y. Han, and X. Liu, "Automated characterization and classification of coronary atherosclerotic plaques for intravascular optical coherence tomography," Biocybern. Biomed. Eng. **39**(3), 719–727 (2019).

36. R. Shalev, D. Nakamura, S. Nishino, A. Rollins, H. Bezerra, D. Wilson, and S. Ray, "Automated volumetric intravascular plaque classification using optical coherence tomography," AI Magazine **38**(1), 61–72 (2017).

37. A. Abdolmanafi, L. Duong, N. Dahdah, I. R. Adib, and F. Cheriet, "Characterization of coronary artery pathological formations from oct imaging using deep learning," Biomed. Opt. Express **9**(10), 4936–4960 (2018).

38. N. Gessert, M. Lutz, M. Heyder, S. Latus, D. M. Leistner, Y. S. Abdelwahed, and A. Schlaefer, "Automatic plaque detection in ivoct pullbacks using convolutional neural networks," IEEE Transactions on Med. Imaging **38**(2), 426–434 (2019).

39. J. Schmitt, S. Xiang, and K. Yung, "Speckle in optical coherence tomography," J. Biomed. Opt. **4**(1), 95–00 (1999).

40. J. Canny, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis Mach. Intell. **PAMI-8**(6), 679–698 (1986).

41. S. Singh and R. Singh, "Comparison of various edge detection techniques," in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, (2015), pp. 393–396.

42. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Commun. ACM **60**(6), 84–90 (2017).

43. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015*, Y. Bengio and Y. LeCun, eds. (2015).

44. K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision - ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, eds. (Springer International Publishing, 2016), pp. 630–645.

45. G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), pp. 2261–2269.

46. J. J. Rico-Jimenez, D. U. Campos-Delgado, M. Villiger, K. Otsuka, B. E. Bouma, and J. A. Jo, "Automatic classification of atherosclerotic plaques imaged with intravascular OCT," Biomed. Opt. Express **7**(10), 4069–4085 (2016).

47. K. A. Vermeer, J. Mo, J. J. A. Weda, H. G. Lemij, and J. F. de Boer, "Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography," Biomed. Opt. Express **5**(1), 322–337 (2014).

48. F. Chollet, "Keras," (2015). https://keras.io.

49. D. Hand and P. Christen, "A note on using the f-measure for evaluating record linkage algorithms," Stat. Comput. **28**(3), 539–547 (2018).

50. S. Huang, C. Tang, M. Xu, Y. Qiu, and Z. Lei, "Bm3d-based total variation algorithm for speckle removal with structure-preserving in oct images," Appl. Opt. **58**(23), 6233–6243 (2019).

51. H. Liu, S. Lin, C. Ye, D. Yu, J. Qin, and L. An, "Using a dual-tree complex wavelet transform for denoising an optical coherence tomography angiography blood vessel image," OSA Continuum **3**(9), 2630–2645 (2020).

52. K. J. Halupka, B. J. Antony, M. H. Lee, K. A. Lucy, R. S. Rai, H. Ishikawa, G. Wollstein, J. S. Schuman, and R. Garnavi, "Retinal optical coherence tomography image enhancement via deep learning," Biomed. Opt. Express **9**(12), 6205–6221 (2018).

CHAPTER 6

Real-Time Tissue Characterization

After identifying DenseNet as a promising deep learning model for tissue classification, the publication presented in this chapter reports on an attempt to demonstrate, for the first time, online tissue differentiation during laser ablation. This paper focused on differentiating between bone and bone marrow tissues. The laser ablation was designed to cut the bone and stop whenever the tissue classifier detected bone marrow. Three main challenges have been addressed in this thesis. The first challenge was the primary beam size of the Er:YAG ablation laser, which created a big crater hole. The hole was bigger than the lateral resolution of the OCT. Thus, for example, the OCT would only scan a small area at the crater's center. Here, we proposed multiple scanning locations and determined the tissue type based on the majority vote of the detected tissue type from each location. However, the scanning locations were difficult to set since the Er:YAG laser beam shape used in this thesis is near Gaussian-shaped, and results in a non-rectangular crater shape. The tracked patch location at the angled crater wall was unable to cover the tissue completely. The OCT-based feedback system should monitor tissue type over the crater area; thus, a top-hat shaped Er:YAG would be ideal for laser osteotomy.

The second challenge was posed by the changes in the tissue's optical properties due to laser ablation, which is mostly caused by the photothermal interaction of light with tissue [120]; thus, changes in a tissue's temperature affect its optical properties and, consequently, alter the tissue's structure appearance. Moreover, changes in intensity on marrow images were observed, likely due to the presence of liquid material at the surface. The liquid appears as the effect of the high-pressure explosion at the ablation center that squeezes the surrounding bone marrow and discharges liquid materials [120], e.g., water and liquid fat. They then diffuse and cover the bone marrow surface. These conditions were addressed in this publication by training the DenseNet model with training data collected while ablating with laser.

The last challenge that has been addressed was the water spray problem. Er:YAG laser is highly efficient in water-cooled conditions. Therefore, ideally, during laser osteotomy, a water spray is necessary to rehydrate the target tissue material so that water-mediated explosions occur. However, the water droplets at the tissue surface will deteriorate the

OCT image and, consequently, disturb the patch tracker. In this publication, a solution was proposed whereby water irrigation was only delivered at a time window while pausing both the ablation and tissue detection system.

This publication demonstrated the feasibility of a real-time feedback system using OCT and deep learning for laser osteotomy. The feedback system offers a completely noninvasive way of monitoring laser ablation. For the hole-ablation experiments, the average maximum depth of perforation and volume loss was $0.216\,\mathrm{mm}$ ($\pm\,0.140\,\mathrm{mm}$) and $0.077\,\mathrm{mm}^3$ ($\pm\,0.076\,\mathrm{mm}^3$), respectively. The average maximum depth of perforation and volume loss for the line-ablation experiments was $0.645\,\mathrm{mm}$ ($\pm\,0.291\,\mathrm{mm}$) and $0.878\,\mathrm{mm}^3$ ($\pm\,0.643\,\mathrm{mm}^3$), respectively.

Publication: Y. A. Bayhaqi, A. Hamidi, A. Navarini, P. C. Cattin, Ferda Canbaz, and A. Zam, "Real-time closed-loop tissue-specific laser osteotomy using deep-learning-assisted optical coherence tomography," *Biomedical Optics Express* Vol. 14, No. 6, pp. 2986–2986 (2023).

Real-time closed-loop tissue-specific laser osteotomy using deep-learning-assisted optical coherence tomography

YAKUB. A. BAYHAQI,[1] ![ID] ARSHAM HAMIDI,[1] ![ID] ALEXANDER A. NAVARINI,[2] PHILIPPE C. CATTIN,[3] ![ID] FERDA CANBAZ,[1] ![ID] AND AZHAR ZAM[1,4,5,*] ![ID]

[1] *Biomedical Laser and Optics Group (BLOG), Department of Biomedical Engineering, University of Basel, 4123 Allschwil, Switzerland*
[2] *Digital Dermatology Group, Department of Biomedical Engineering, University of Basel, 4123 Allschwil, Switzerland*
[3] *Center for medical Image Analysis and Navigation (CIAN), Department of Biomedical Engineering, University of Basel, 4123 Allschwil, Switzerland*
[4] *Division of Engineering, New York University Abu Dhabi, Abu Dhabi, 129188, United Arab Emirates*
[5] *Tandon School of Engineering, New York University, Brooklyn, NY, 11201, USA*
*azhar.zam@nyu.edu

Abstract: This article presents a real-time noninvasive method for detecting bone and bone marrow in laser osteotomy. This is the first optical coherence tomography (OCT) implementation as an online feedback system for laser osteotomy. A deep-learning model has been trained to identify tissue types during laser ablation with a test accuracy of 96.28 %. For the hole ablation experiments, the average maximum depth of perforation and volume loss was 0.216 mm and 0.077 mm^3, respectively. The contactless nature of OCT with the reported performance shows that it is becoming more feasible to utilize it as a real-time feedback system for laser osteotomy.

1. Introduction

Over the last decades, laser technologies for cutting bones have developed rapidly. Lasers produce a cleaner surface cut, cause fewer mechanical vibrations, offer more flexible cutting geometries, cause fewer material losses, and show less microorganism contamination than conventional saws, drills, or burs [1–7]. Laser osteotomy, especially with *erbium-dopped yttrium aluminium garnet* (Er:YAG) laser, is ideal for efficient bone ablation with very little carbonization [8–13]. However, similarly to mechanical tools, there is an inherent risk of collateral damage to surrounding tissue, such as bone marrow and nerves. This remains a challenge in laser osteotomy, especially over the ablation direction. In osteotomy with mechanical tools, physical feedback (i.e., haptic feedback) helps surgeons to stop cutting as soon as some damage is done to the bone marrow so that the risk of cutting nerves is avoided. Controlling the laser osteotomy is also difficult. Ideally, a surgeon plans the ablation site and depth based on the patient's computed tomography (CT) data. The ablation depth could be simply calculated from the ablation rate per pulse. Nevertheless, the accuracy of such a process is limited to the resolution of the CT and the ablation rate calculation. The ablation rate is influenced by multiple factors, such as the shape of the laser beam, the energy per pulse, the cooling system, the tissue density, and the water content [13–15].

One solution to this problem is to integrate a closed-loop control system that would make it possible to safely perform laser osteotomy without damaging critical internal tissues. Researchers have sought to address this issue throughout the last decade. Such a control mechanism aims to identify tissue type below the laser-ablated incision during the cutting process. The ablation process would be terminated as soon as soft tissue (such as bone marrow or nerves) is detected.

Such a detection system can be achieved by analyzing the secondary light and the acoustic emission produced by the laser's interaction with tissue or by integrating the ablation laser with additional diagnostic modalities.

Analysis of secondary light emissions includes random lasing (RL) [16] and laser-induced breakdown spectroscopy (LIBS) [17–21], which have been demonstrated to be promising methods for differentiating tissue types. It has been shown that RL has nanoscale sensitivity to small structural changes [16]. LIBS could run under operating room light and has a very high accuracy for tissue differentiation at an atomic scale [18,19,22]. Even more, LIBS could also be used for carbonization detection [23], preventing permanent damage to the bone. But each of these methods also has its own disadvantages. RL depends on a laser dye that might harm biological tissue, while LIBS has limitations in reaching a deep cut and is more accurate for characterizing tissue surfaces. Furthermore, research in analyzing the acoustic signals emitted by photoablation (ablative optoacoustic techniques) has also been demonstrated to have comparable accuracy [24–28]. These methods require, however, a high-energy pulsed laser to produce a measurable acoustic signal; thus, damage to the tissue is almost unavoidable.

Another interesting method for analyzing the effects of photoablation was presented by Rupprecht *et al.* [29,30]. Their approach combines analyzing both secondary light emissions and acoustic signals (combined pyrolysis-photoacoustic). Here, the light emission from the pyrolysis process in the ablation zone is observed with a photodiode. At the same time, a piezoelectric accelerometer is used to measure the generated acoustic signal. The measured signals were used to interrupt the laser beam as soon as threshold values were reached. Additionally, we could also use the pyrolysis signal for detecting bone carbonization [31]. However, this method inherits similar drawbacks from the previously mentioned methods. A high-energy pulsed laser is needed to generate the pyrolysis light. Even more, the piezoelectric accelerometer needs to be mounted to the bone, which causes unnecessary damage.

Contactless laser ablation can be integrated with additional nonimaging and nonablating diagnostic modalities. Raman spectroscopy [32,33] provides information regarding the molecular bonds of tissue and has a high accuracy comparable to LIBS. However, a few seconds of integration time is required to get an observable signal [32], which makes it unsuitable for a real-time feedback system. Another alternative, autofluorescence [34,35] and diffuse reflectance spectroscopy [36–39], may be used to differentiate tissue types with a faster processing time. But these techniques work as a point measurement, thus lacking structural information about the tissue anatomy over the depth.

Among the methods previously mentioned, to date, only LIBS [21] and the combined pyrolysis-photoacoustic [29,30] methods have been experimentally tested for use in a real-time closed-loop control system for laser osteotomy. However, although both methods have high accuracy, these methods distinguish tissue type only after an ablation pulse has been applied, which increases the risk of collateral damage during laser ablation. A further drawback of these approaches is that they can only detect tissue transitions after crossing the tissue boundary. Damage to critical structures is thus almost unavoidable.

We proposed a contactless and ablation-free optical coherence tomography (OCT) imaging system for monitoring the laser ablation process. This emerging imaging technology is analogous to ultrasound imaging that performs noninvasive cross-sectional tomography using light propagation in media and interference phenomena. The echo time delay of the back-reflected or backscattered light from the tissue's internal microstructures is measured using interferometry of partially coherent light. It has been demonstrated as a viable alternative to ultrasound for real-time, high-resolution, and in situ investigations of thin tissue structures [40–42]. This noninvasive imaging modality depicts not only a topological profile of the examined surface but also displays the subsurface structures [43,44]. It has become a standard technique in ophthalmology for

diagnosing retinal diseases [45,46]. Additionally, over the last decades, OCT has become an alternative to ultrasound for intravenous imaging [47–51].

OCT has been employed in several studies to obtain subsurface information from different tissue types and has been used as guidance for laser-ablation surgery, including in tumor resection [52,53], cochlear implant surgery [54,55], and bone osteotomy [56–58]. Increases in resection accuracy and precision for brain tumors and blood coagulation have been reported as advantages of utilizing this technology [52,53,59]. One study showed that OCT helps to shape the endosteal layer's curvature for cochlear implantation with a mean absolute accuracy of around 20 μm [54]. Moreover, some literature has demonstrated that OCT could also be used to monitor the relative temperature of the ablated bone surface, enabling the prevention of bone carbonization [60,61]. Specifically, OCT has been tested as a noncontact real-time feedback system for monitoring the ablation depth of a laser in controlled bone osteotomy [56–58]. Nevertheless, these approaches mainly focus on measuring the position of the target tissue (depth control).

In this work, we utilized OCT to guide the laser osteotomy by detecting tissue type based on the OCT image and providing real-time feedback. Specifically, our proposed method takes advantage of deep learning algorithm to identify the tissue types because it could provide an efficient and accurate way to differentiate tissue. Thus supporting our real-time feedback system. To the best of our knowledge, this is the first approach to provide the basic framework for a real-time closed-loop tissue-specific feedback system based on OCT. In the present study, we focused on testing the performance of the feedback system in differentiating bone and bone marrow in real-time. The experiments were done on fresh porcine femur bones with bone marrow inside. The primary objective of the tissue-specific feedback system was to cut the bone and to stop the laser ablation whenever bone marrow was detected. In the future, we foresee optimizing the feedback system for more precise predictions and more tissue types.

2.   Methodology

2.1.   OCT closed-loop controlled-laser-ablation setup

An Er:YAG (Syneron LiteTouch, Israel) ablation laser was used for our laser-osteotomy study because of its efficient ablation rates and low carbonization effect [8–13]. The ablation laser's wavelength (2.94 μm) is strongly absorbed by both water and hydroxyapatite, which are the main components of bone. The absorption of such laser light leads to photothermal ablation [1,10]. The irradiated area consequently experiences heat transfer and an increase in pressure. The tremendous build-up of pressure in a fraction of microseconds induces the explosion of tissue material at the focal-spot area. Such heat transfer may rapidly increase the temperature in the surrounding area and lead to carbonization outside the focal spot. A water spray and pressurized air are usually used to keep the temperature of tissue below the carbonization threshold. Rehydrating the tissue also increases its water content and improves the ablation rate.

The closed-loop controlled-laser osteotomy aims to provide detection of the tissue types encountered during laser ablation to avoid cutting critical tissues. Our concept of OCT closed-loop controlled-laser ablation is illustrated in Fig. 1(a). The Er:YAG laser beam was focused on the bone surface by a CaF$_2$ lens (L3) with a focal length of f = 75 mm. The ablation laser and the OCT system were coaxially coupled with the help of a dichroic filter (DF) [Advanced Osteotomy Tools II-VI 80048151, Switzerland]. A sapphire window with a 2 mm thickness was placed in front of the filter to protect the optical components from debris and water droplets splashing around during ablation. Additionally, pressurized air was used to deflect the debris trajectory to prevent bone debris from accumulating on the sapphire window, which could reduce the incident energy of the laser and deteriorate the OCT image.

We used a custom long-range Fourier-domain OCT system with an extended depth of focus for deep-ablation monitoring. It was equipped with a laser source (Insight Photonic Solution, Inc., Lafayette, Co, USA) with a central wavelength of 1310 nm, a spectral bandwidth of 61.5 nm, and
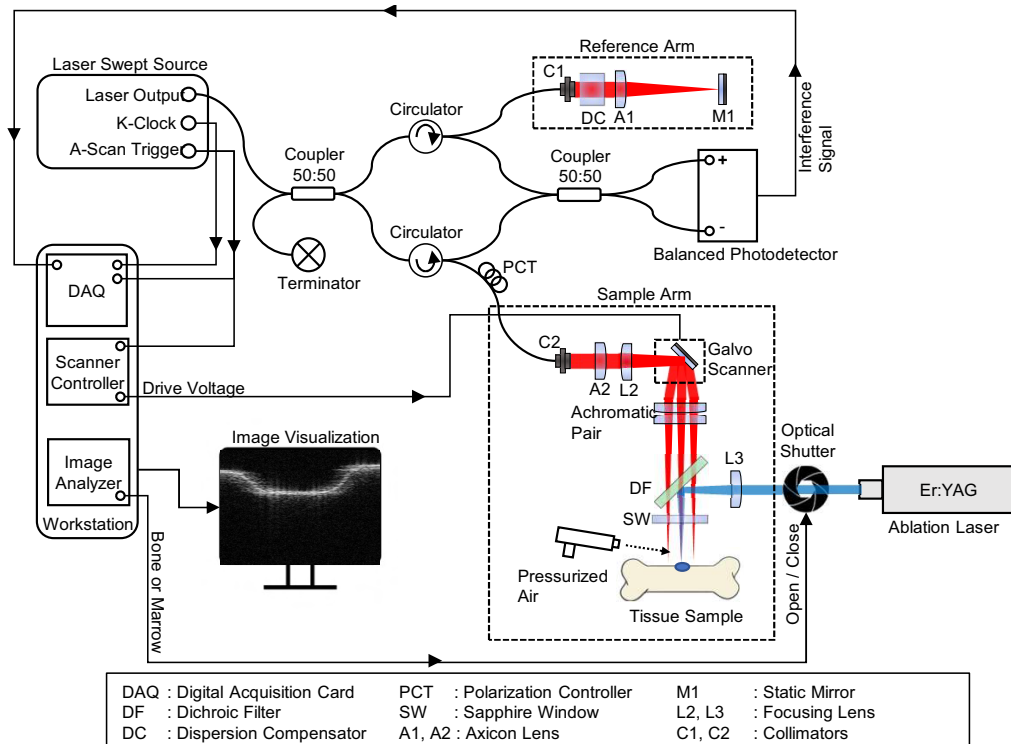
(a)
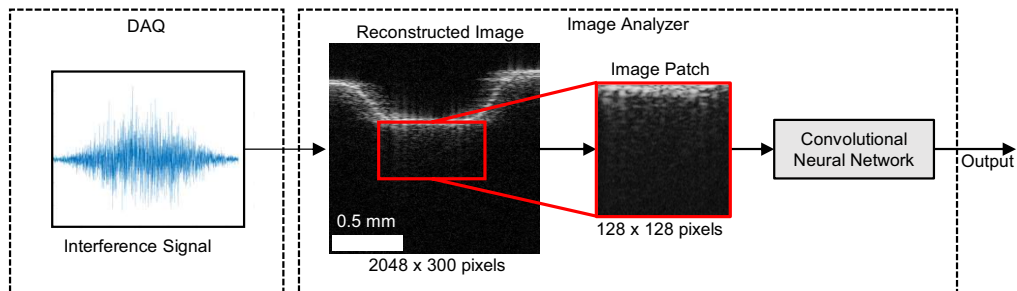
(b)

Fig. 1. (a) Schematic of the proposed OCT-based closed-loop controlled-laser osteotomy. We used a custom long-range Fourier-domain OCT with a Bessel-like beam (BLB) optical setup [39]. A dichroic filter coupled the OCT (red line) with an ablation laser (blue line). The OCT images were analyzed (image analyzer) to identify the ablated tissue type. The output provided feedback to an optical shutter that controlled the ablation laser to either stop or continue ablation. (b) The image analyzer worked based on an image patch. A region of interest (image patch) from the OCT image was selected on the ablation spot. We trained a convolutional neural-network model to identify the tissue type based on the extracted image patch.

an A-scan line rate of 104.17 kHz. The OCT system had an imaging range of 26.2 mm in the air. The long-range imaging ability was achieved using a Bessel-like beam (BLB) optical setup, as explained in our previous publication [42]. The reconstructed B-scan images from the OCT system had the dimensions of 2048 pixels (26.2 mm) in the axial and 300 pixels (1.5 mm) in the lateral direction. The corresponding lateral and axial resolutions were 26 μm and 18 μm, respectively. The sensitivity or the maximum signal to noise ratio (SNR_{max}) of the OCT system was 110 dB. The OCT system was controlled with a workstation that was equipped with an Intel Core i9-7900X

central processing unit (CPU) and an NVIDIA GTX 1080Ti graphics processing unit (GPU). The interference signal was detected by a balanced photodetector (Thorlabs PDB48xC-AC, Germany) and digitized by a PCIe Waveform Digitizer (Alazartech ATS9373, Canada). OCT images were reconstructed from the interference signal through several preprocessing steps such as Direct Current (DC) subtraction, windowing, zero padding, and inverse Fourier transformation. All of this signal processing and image reconstructions was calculated on the GPU using the compute unified device architecture (CUDA) library for fast parallel processing [62].

The OCT images were streamed to monitor the ablation process. The detection system worked based on an image patch taken from the ablation area (see Fig. 1(b)). The patch was used as a deep-learning model input to discriminate between tissue types. In this paper, the deep-learning model was trained to differentiate between bone and bone marrow (see Section 2.3). The model's output provided feedback to an optical shutter (Thorlabs SH1, Germany) for indirectly controlling the Er:YAG ablation laser.

2.2.   OCT scan pattern

The Er:YAG laser circular beam at the surface of the bone tissue had an estimated diameter of ~1.0 mm with beam quality factor $M^2 = 22$. Meanwhile, the lateral resolution of the OCT image was 20 μm, so a significant area of ablation was not covered by a single B-Scan image, see Fig. 2. Our proposed approach for detecting tissue type was based on detecting three image patches from three different scan locations (images) over the monitored ablation hole. The number of locations was chosen as a trade-off between acquisition time and coverage of the ablation area. This scanning mechanism is illustrated in Fig. 2. In total, nine B-scan frames were acquired for every detection step, which corresponds to an acquisition rate of 28.94 Hz (34.56 msec). For each scan location, the image was frame-averaged from three consecutive B-Scan frames to obtain better quality images. The tissue type of the monitored ablation crater was determined by detecting three image patches from three scan locations simultaneously and taking the majority voting of the detection results.
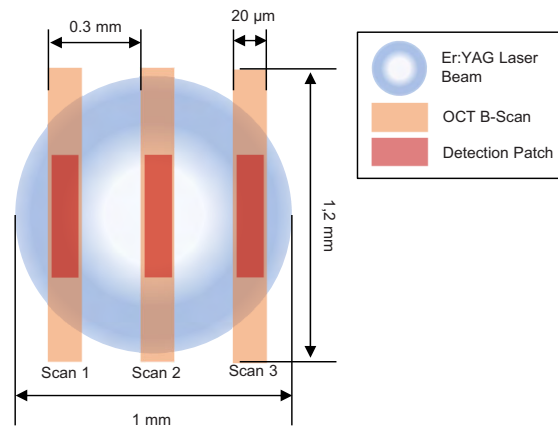
Fig. 2. Top-view illustration of the OCT scan pattern (orange) to cover the bigger area of the Er:YAG laser beam. The color gradation from white (higher) to blue (lower) indicates the intensity (fluence) of the laser beam. Our detection approach involved three detection locations (red) that corresponded to analyzing three OCT B-scan images simultaneously.

2.3.   Tissue differentiation

Tissue detection or classification has most recently been the research focus in clinical applications of OCT, such as retinal diseases [63], cancer [64], or atherosclerosis plaque detection [47].

Conventionally, tissue detection has usually been done by discriminating the feature representation (such as attenuation coefficient and textural features) of the tissue in the OCT image. Discrimination using machine learning algorithms, such as random forest (RF) [47], principle component analysis (PCA), and support vector machine (SVM) [64] have been demonstrated to be useful and have shown good accuracy for automatic tissue differentiation. However, the complicated feature extraction of the image increased the prediction time, which is a critical point for a real-time feedback system. As an illustration, the method proposed by Ughi *et al.* [63] in 2013 took about 2 seconds to detect the atherosclerotic plaque using the attenuation estimation and RF classification algorithm. Another example was done by Müller *et al.* [64] in 2021. They needed 1.6 seconds to automatically classify tissue in brain metastases using textural features and SVM.

Currently, one of the most popular methods for image recognition is by using deep learning algorithms, specifically the convolutional neural network (CNN). As one of the deep-learning models, CNN has been known to have better accuracy and efficiency for object detection on an image than other classical machine-learning methods or even other traditional deep-learning models such as multi-layer perceptron (MLP) [65,66]. The convolutional layer of CNN operates as spatial filters that extract high-level features such as edges from the input image.

In OCT research, CNN was proposed as an alternative for real-time image recognition because of the non-complicated or straightforward feature extraction for recognizing the patterns of specific tissues in OCT images. In 2017, Roy *et al.* developed ReLayNet [67] which is able to segment retinal layers in 10 msec. Followed by Borkovkina *et al.* in 2020 [68], they have successfully accelerated the segmentation of retinal layers to only 3.5 msec.

Moreover, in our previous experiments [69], CNN models such as VGG [70], ResNet [71], and DenseNet [72] were demonstrated to have an accuracy of more than 90 % to differentiate five tissue types with an inference time of less than 65 msec. We found that DenseNet has better accuracy compared to the other tested models. Specifically, we have demonstrated that this model had an accuracy of 91.52 % with an inference time of 52.73 msec. Therefore we chose DenseNet (DenseNet121) as the tissue-classifier model for the present experiments.

In this work, we modified the DenseNet model output to recognize bone and bone marrow. The input of the DenseNet model is an image patch that was selected in such a way as to represent a region of interest where a destructive laser pulse would be applied. Technically, the ablation spot is fixed in the lateral center of the B-scan image. Before conducting the experiments, the OCT was aligned such that the B-scan image is centered to the hole ablation, as illustrated in Fig. 1(b). The tissue surface in the B-scan image was detected using the vertical Canny edge detection method because of its simplicity and low sensitivity to noise [73,74]. A square of 128 pixels × 128 pixels area was extracted from the ablation spot and defined as the input patch for the DenseNet model. Examples of the image patches are shown in Fig. 4. Finally, the model had two outputs with sigmoid activation function, which gave the tissue-type prediction (bone or bone marrow).

2.4. Training, testing, and online inference

The data sets used to train the DenseNet model were taken from ex-vivo pig-bone samples. The samples were acquired from a local butcher. Specifically, the samples were taken from the middle section of the femur bone and consisted of the compact bone as the outer layer and the bone marrow inside. The muscle and connective layer were removed before the experiments. For fully supervised training, the bone image patches were taken from the bone surface, while the bone marrow image patches were taken from the sample's side, as illustrated in Fig. 3(a) (blue line for bone and red line for bone marrow).

There are, however, some challenges for tissue classifiers during laser ablation. The tissue experiences an increase in temperature (heating up) during microsecond ablation. This will

(a)   (b)
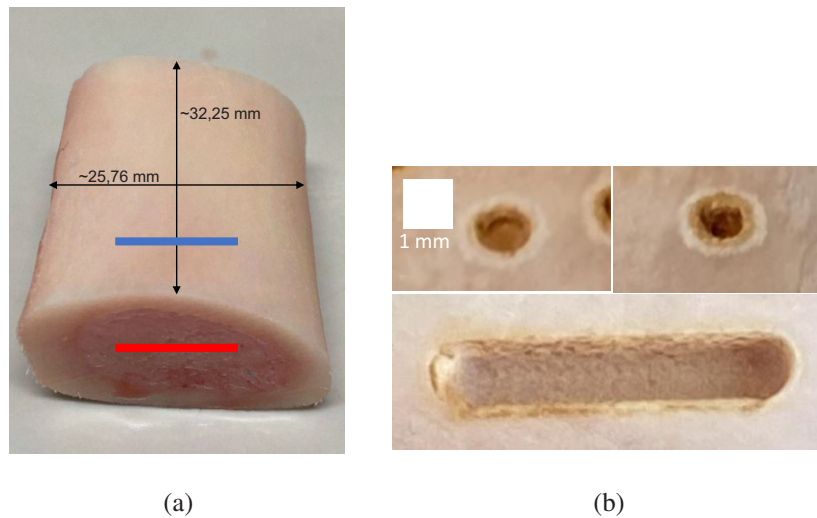
Fig. 3. (a) Example of a pig femur bone sample used in the experiment with bone marrow inside. The OCT B-scan images, used for deep learning training, were scanned at the blue line for bone tissue and at the red line for bone marrow. The average thickness of the bone layer was 2.5 mm. (b) Example of the dehydrated bone after receiving 12-18 laser pulses. The top images show bone dehydration with hole ablations. The bottom image shows bone dehydration with a line ablation. The hole-ablated bone experiences more dehydration to carbonization (indicated with a darker color) than the line-ablated bone, because it better distributed the heat from the laser (See section Section 2.5).

(a) Bone without ablation   (b) Bone marrow without ablation

(c) Bone with ablation   (d) Bone marrow with ablation

Fig. 4. Examples of OCT B-scan images collected in the experiment. Bone (a) and bone marrow (b) images without laser ablation. Bone (c) and bone marrow (d) images during laser ablation with 200 mJ. The image of bone during ablation shows an increase in intensity due to water dehydration. By contrast, the image of bone marrow shows a decrease in intensity due to light absorption by liquid materials at the surface.

induced dehydration to the bone as shown in Fig. 3(b). As a consequence, the optical properties of the tissue (such as the refractive index, absorption coefficient, and scattering coefficient) change [35,37]. These changes affect the textural information in the OCT images and may reduce the tissue classifier's accuracy. For these reasons, we trained the classifier using image patches from four different conditions. The first patches were taken without applying the laser. The other patches were taken while applying the ablation pulses with the energy of 200, 150, and 110 mJ, respectively.

Five pigs, with three samples from each pig, were used to train and test the classifier. As explained in the previous paragraph, image patches from each sample were taken from four ablation conditions, with 1000 image patches per ablation condition. Thus, a total of 8000 image patches consisting of 4000 image patches of bone tissue and 4000 image patches of bone marrow were collected for each sample. The samples were separated into six samples from two pigs for training, three samples from one pig for validation, and six samples from two pigs for testing. In other w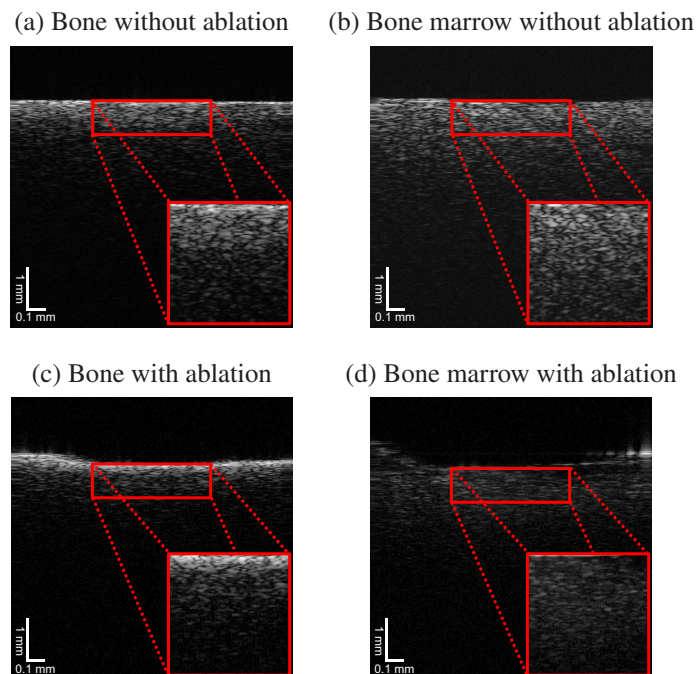ords, the data sets were separated into 48,000 (24,000 per tissue) patches for training, 24,000 (12,000 per tissue) for validations, and 48,000 (24,000 per tissue) for testing the DenseNet model.

The training of the DenseNet model was implemented using the Pytorch Deep Learning framework [75]. We trained it on an NVIDIA DGX A100 workstation equipped with NVIDIA A100 GPUs, which enabled us to perform parallel computations to speed up the training process. The model was trained with 1000 epochs and a batch size of 32 samples. We defined cross entropy as the training loss function and Adam (learning rate $= 1.0 \times 10^{-4}$) as the training optimizer. Furthermore, we performed data augmentation to the patch image during training by random small shifts and horizontal flips.

The inference of the DenseNet model was done on the OCT workstation and embedded as one of the processing pipelines. After training the DenseNet model in Pytorch, we converted the model to the Open Neural Network Exchange (ONNX) format [76], which can be read by the NVIDIA TensorRT [77] library for parallel deep-learning inference in our OCT's GPU. This inference mechanism performed faster than previously reported inference mechanisms [69]. TensorRT ran the model inference in a parallel programming manner inside the NVIDIA GPU and used the GPU's memory to hold the input patch of the model. In our system, the OCT signals were processed in the GPU. Thus, the reconstructed images already resided in the GPU's memory. Copying the OCT-reconstructed image to the DenseNet-model input was also done in the GPU. We thus avoided unnecessary memory copy between the CPU and the GPU. The inference of our DenseNet model took an average time of 11.96 msec by using this mechanism.

2.5. Ablation experiments

The experiments to evaluate the real-time feedback system were split into two parts. In the first experiment, the Er:YAG laser was used to drill a hole starting at the bone surface and was stopped when the bone marrow was reached. The ablation laser was set to send laser pulses at the energy of 200 mJ per pulse with a repetition rate of 4 Hz. During ablation, pressurized air was pointed to the ablation site from the side of the ablation direction (see Fig. 1(a)).

Pressurized air of 2 bar was applied to gently sweep the debris from the ablation site and prevent it from flying up to the sapphire window. Even though the air could cool down the tissue, it did not optimally prevent the tissue from carbonizing, which could appear after a few dozen laser pulses. We therefore paused the ablation laser, tissue detection, and pressurized air, and then manually sprayed water on the ablation spot whenever carbonization became visible (it was usually identifiable by blackening on the surface). Moreover, before continuing ablation, we reapplied the pressurized air to sweep the remaining sprayed water, which could deteriorate or hinder the tissue texture in the OCT images.

In the second experiment, we used a motorized linear stage to move the sample and make a 0.5 mm long line cut. The stage was set to move at a constant speed of 1 mm/s. The repetition rate and energy of the ablation laser were set similarly to the hole-ablation experiment. The OCT B-scan direction was set to be perpendicular to the ablation line. Despite that, the parallel scan was superior, and the OCT-scanning size was larger than the laser-beam size. This would result in a wavy cut, where a small-sized bone (smaller than the B-scan) will be considered as bone marrow and will not be ablated by the laser. Since peripheral heat is distributed better in line ablation than in hole ablation, carbonization is observed less often. It consequently leads to a higher ablation rate [13]. Since carbonization may nevertheless appear, we still applied similar pausing, water spraying, and air-blowing steps whenever it was noticeable. However, we only paused ablation at every endpoint of the line.

2.6.   Performance evaluation

The ablated samples were evaluated radiographically with a micro computed tomography (micro-CT) (Bruker SkyScan 1275, Belgium) to obtain three-dimensional geometrical information of the ablation crater shape. The micro-CT images had a dimension of 1944 pixels in height and 1944 pixels in width. The voxel spacing of the image was 16 μm equally for both axes. The radiographical images were taken before and after cutting the sample with the laser (pre- and post-ablation). Both images were registered and overlaid to precisely identify the boundary between bone and bone marrow. We used the registration estimator application in Matlab with the monomodal intensity-based rigid registration techniques [78] to estimate the shift between the pre- and post-ablation CT images.

The evaluation of the feedback system was performed by measuring the volume and maximum depth of collateral damage (perforation) to the bone marrow. The ablation samples used for evaluation were different from the samples for training the deep-learning model. A total of six new samples were used, three of them for hole ablation and the rest for line-ablation experiments. Here, ten hole- or line-ablation experiments were made for each sample. We defined the maximum depth of perforation as the depth from the bone-bone marrow border to the bottom of the crater (illustrated as a white arrow in Fig. 6(i)). The maximum perforation depth was selected over the entire micro-CT image slice that covered the ablation crater. Furthermore, we measured the volume of the collateral damage by segmenting (with a threshold method) the damaged area on the overlaid image and summing up the segmented areas over the whole micro-CT image slices that covered the ablation crater. Similarly, the line ablation was also evaluated over the entire micro-CT image slices covering the ablation line.

3.   Results and discussion

3.1.   Deep-learning training and inference

Training the Densenet model took ~6.5 hours to complete with a training accuracy of 99.67 % and a validation accuracy of 96.52 %. Furthermore, the test results of the DenseNet model showed an accuracy of 96.28 %. In our previous study, the attenuation coefficient profile, which was explicitly extracted using a depth-resolved method, improved the detection accuracy. However, the time to extract this profile delayed the detection time. Other than that, we believed that our model also tried to implicitly extract the attenuation coefficient profile, but it is difficult to visualize filters over all the layers in the DenseNet model. Finally, in our current approach, both image acquisition and inference with the DenseNet model (detection time) took only 45.96 msec, which was faster than the laser's repetition rate of 4 Hz. Our feedback system can therefore be considered a real-time feedback system.

3.2. Ablation evaluation

Figure 5(a) shows an axial slice of the micro-CT image for the hole-ablation experiment. The closed-loop controlled laser system resulted in an automatic interruption of the laser beam as soon as bone marrow was detected. However, a small perforation on the bone marrow is noticeable. Statistically, the average maximum depth of perforation in 30 holes was 0.216 mm ($\pm$ 0.140 mm), and the deepest measured bone marrow perforation was 0.910 mm. The average volume loss of bone marrow was 0.077 mm^3 ($\pm$ 0.076 mm^3) with a maximum of 0.299 mm^3. Figure 6(h) shows a sagittal slice of the CT image for the line-ablation experiment. Similarly, we also observed the perforation of the bone marrow. The average maximum depth of perforation in 30 ablation lines was 0.645 mm ($\pm$ 0.291 mm) with a deepest measured perforation of 1.778 mm. The average volume loss of bone marrow was 0.878 mm^3 ($\pm$ 0.643 mm^3) with a maximum of 2.269 mm^3. Tables 1 and 2 show more detail of the experiment results.

Fig. 5. An example for the hole-ablation experiment (a) with a zoomed-in image of a hole (b) and its maximum intensity projection from the micro-CT image (c). A micro-CT axial slice image for post-ablation is shown in (d) with a zoomed-in image (e) over a hole. The axial slice was taken over the blue line at (a). Similarly, image (f) shows the overlaid CT images between pre- and post-ablation of the axial slice with the zoomed version (g). Image (h) illustrates a segmented region of the bone marrow perforation (transparent red) to measure the bone marrow volume loss. Lastly, image (i) illustrates the measurement of the maximum depth perforation (white arrow) of the bone marrow.

3.3. Discussion

Ideally, clean bone surfaces free of carbonization are expected when using Er:YAG laser osteotomy if adequate water cooling is applied. In our experiments, continuous water cooling was impossible because it would have deteriorated the OCT images and limited the detection accuracy of our tissue classifier. Therefore, we manually paused the ablation and sprayed water on the ablation spot whenever carbonization was noticeable. The carbonized part of the bone quickly evaporated after rehydration and reablation sequentially. This sequence was stopped once the bone marrow was reached to prevent further damage or perforation to the bone marrow. We, therefore, expected a trace amount of carbonization. Figure 5(d) shows that a noticeable amount of carbonization appeared on the walls of the ablation craters.

In the line-ablation experiments, carbonization was less than in the hole-ablation experiments (see Fig. 6(b)). This is because the thermal dissipation was distributed better along the ablation line. However, the perforation of bone marrow in the line-ablation experiments was more than in the hole-ablation experiments. This is mainly due to the overlapping conditions at the periphery
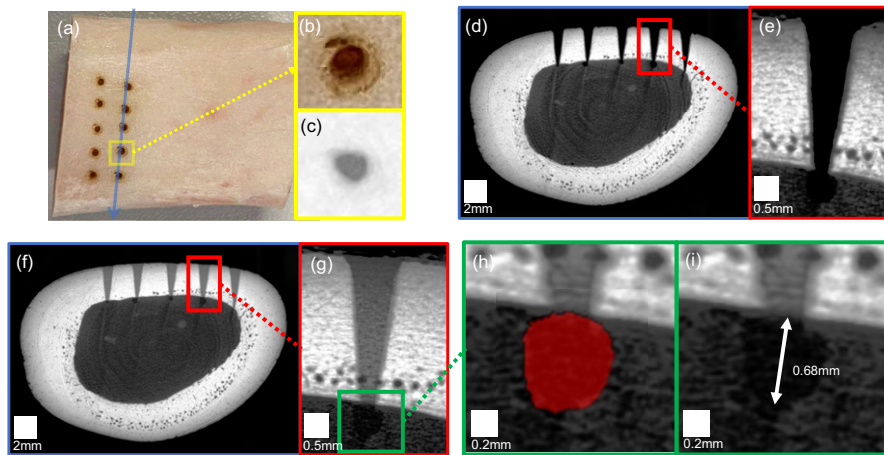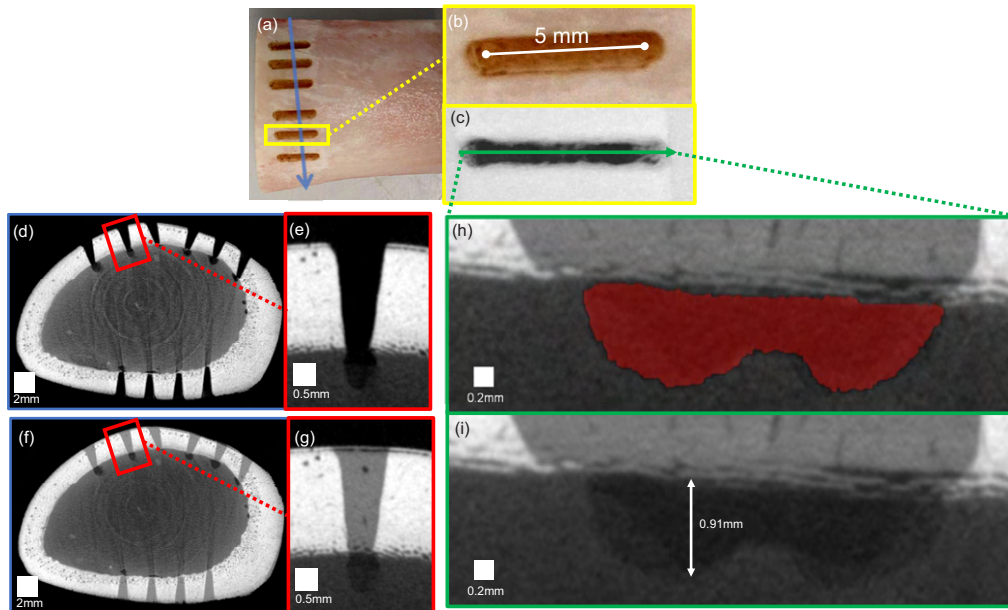
Fig. 6. An example result from the line-ablation experiment (a) with a zoomed image of the line cut (b) and its maximum intensity projection from the micro-CT images (c). A micro-CT axial slice image for post-ablation is shown in (d) with a zoomed-in image (e) over a hole. The axial slice was taken over the blue line at (a). Similarly, image (f) shows the overlaid CT images between pre- and post-ablation of the axial slice with the zoomed version (g). Image (h) shows the micro-CT sagittal slice image of the cut line and illustrates a segmented region of the bone marrow perforation (transparent red) to measure the bone marrow volume loss. Lastly, image (i) illustrates the measurement of the maximum depth perforation (white arrow) of the bone marrow.

Table 1. The measurement results for maximum perforation depth and volume loss from *N* = 30 hole-ablation experiments. The results were also compared with the pyrolysis-photoacoustic method [29,30]. Note that only the maximum perforation measurements from hole-ablation experiments are available from the reference method. The data were recalculated based on the average bone thickness of 2.9 mm.

Methods	Our method (Hole ablation)		Pyrolysis-photoacoustic [29,30]
	N = 30		*N* = 98
	Max perforation	Volume loss	Max perforation
	(mm)	(mm^3)	(mm)
Min	0.021	0.001	0.000
Max	0.910	0.299	0.512
Mean	0.216	0.077	0.065
Median	0.178	0.055	0.015
Std dev	0.140	0.076	0.088

of the laser beam. The intensity profile of our laser beam was close to Gaussian, with higher intensity in the middle and lower intensity on the periphery. The energy on the periphery is insufficient to ablate the bone significantly, but it is enough to ablate the bone marrow. A tophat-beam intensity profile could be ideal for osteotomy. During hole ablation, this peripheral intensity was mostly eliminated by the wall of the crater. By contrast, this peripheral beam may still ablate the adjacent point over the line during line ablation. Additionally, the OCT detection

Table 2. The measurement results for maximum perforation depth and volume loss from _N_ = 30 line-ablation experiments.

	Line ablation _N_ = 30	
	Max perforation (mm)	Volume loss (mm^3)
Min	0.178	0.110
Max	1.778	2.269
Mean	0.645	0.878
Median	0.572	0.777
Std dev	0.291	0.643

locations were limited to the primary laser beam and could not cover the whole beam (primary and peripheral beam). Although it may be beneficial in faster bone cutting time, the damage to the bone marrow was more. Additionally, the water content in the bone marrow is higher than in bone. Bone material consists approximately 13 % water [79], while marrow tissue consists approximately between 15 - 40 % water [80]. Thus with a constant laser-pulse energy, the ablation rate in bone marrow is higher than in bone. As a result, bone marrow is easily ablated.

All experiments showed a slight perforation of bone marrow. This indicates that the deep learning always stopped the ablation whenever the bone marrow was reached and that no premature stops were made. The perforations were expected because the deep-learning model was trained based on image patches containing only a single tissue type. It was insensitive to multilayered conditions. It therefore only stopped the laser when the image patch only contained bone marrow.

Multilayered tissue detection (segmentation) would provide more accurate detection, especially at bone-bone marrow interface areas. Unfortunately, our OCT image contrast was insufficient for multilayered detection. The bone tissue scattered and absorbed most of the OCT light. Furthermore, the small signal coming from the bone marrow structure was reduced by multiple scatterings. There was therefore a large contrast difference between bone and bone marrow on multilayered tissue with bone on top. As a result, the bone marrow structure intensity appeared to be weak in the OCT images (see Fig. 7(a)) and detecting it was difficult. Figure 7(b) shows the OCT image when ablation had already reached the bone marrow. The bone marrow texture (speckle) is blurred due to frame averaging during the ablation process, which is induced by the motion of liquid materials at the surface. The liquid appears as the effect of the high-pressure explosion at the ablation center that squeezes the surrounding bone marrow and discharges liquid materials, e.g., water and liquid fat [14]. They then diffuse and cover the bone marrow surface. Furthermore, multilayered detection of bone and bone marrow on the OCT image would provide more precise control through predicting the remaining depth of the bone needed to be cut. Ideally, with such predictions, we could slow down the ablation rate or reduce the laser energy such that the cut would stop right before it touches the bone marrow. However, our laser safety mechanism prevents us from having control over increasing or reducing the laser pulse energy (ablation rate). This mechanism requires the user to stop or pause the laser from pulsing before changing any parameter of the laser pulse (repetition rate and energy). The pulsing of the laser may be continued once the repetition rate and energy are confirmed. This pausing mechanism delays the laser pulse up to around 30 seconds. Therefore, during the experiments, the laser energy was kept constant. Ablation to the bone marrow was, therefore, almost unavoidable when the bone thickness was less (still detected as bone) than the amount that would be ablated by the next laser pulse.

To summarize, the proposed real-time feedback system demonstrated to have a comparable performance to the combined pyrolysis-photoacoustic method [29,30] for hole ablation. The reference method used laser ablation alone both for ablation and as the detection light source.
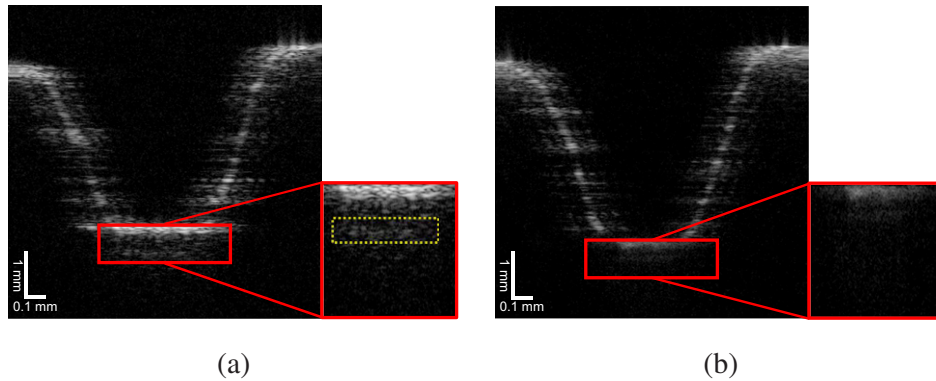
(a)   (b)

Fig. 7. OCT images during ablation with patches for tissue detection. The left image (a) shows an OCT image with two layers of tissue, bone on top and bone marrow on the bottom (dashed yellow box). The bone marrow structure intensity appears to be weak due to the high absorption of light in the bone marrow tissue. The right image (b) shows an OCT image when ablation had already reached the bone marrow.

In this situation, the detection mechanism stops together with the ablation laser once the laser reaches the bone marrow, which makes the closed-loop scenario unrepeatable. This means that the ablation and detection mechanism can only start at bone. A restart mechanism is needed once the laser beam reaches the bone marrow, after which neither ablation nor tissue-type detection can continue. The detection system has to be manually set to the default tissue type for each ablation point. By contrast, our OCT detection system runs independently from the ablation laser, which eliminates the redundant restart mechanism. And although our results show higher perforation, this approach offers a noninvasive way to monitor the ablation process. It eliminates unnecessary damage due to sensor mounting in the pyrolysis-photoacoustic method. The perforation, in our case, is also higher since we have a higher ablation rate of the Er:YAG laser. Even though perforation could not be perfectly solved with our method, we believe that it could be used as a safety feature on a laser-osteotomy system.

4.   Conclusion and outlook

The first real-time feedback using OCT for laser osteotomy was demonstrated and performed comparable to the reference method [29,30]. The feedback system offers a completely noninvasive way of monitoring laser ablation. For the hole-ablation experiments, the average maximum depth of perforation and volume loss were 0.216 mm ($\pm$ 0.140 mm) and 0.077 mm^3 ($\pm$ 0.076 mm^3), respectively. On the other hand, the average maximum depth of perforation and volume loss for the line-ablation experiments were 0.645 mm ($\pm$ 0.291 mm) and 0.878 mm^3 ($\pm$ 0.643 mm^3), respectively. These results are also comparable to osteotomy with a drill that was reported to have a mean perforation of 0.660 mm [81].

These results show the feasibility of using OCT as a feedback system for laser osteotomy in the operating room. This paper provides the basic framework for tissue-specific laser osteotomy using deep-learning-assisted optical coherence tomography. Several optimizations could still be undertaken or combined with the system, such as carbonization detection, depth control, predefined ablation rate, multilayered tissue detection, and real-time control of the laser energy.

Carbonization remains a challenge in using OCT as the guidance in laser osteotomy because continuous water irrigation reduces the tissue-detection accuracy. Our pause-sequence method may partly solve this problem, but it would be better if one could detect any carbonization and sync the result to an irrigation controller. Synchronization between OCT detection and a water spray may improve this method's effectiveness.

Furthermore, several other aspects could still be optimized to improve the precision of detection and reduce the perforation depth. In ideal clinical conditions, a surgeon could plan the ablation site and depth based on a patient's CT data. The OCT system could be used as a depth-control mechanism during ablation. Concurrent tissue detection could be used as a second parameter to support the depth control in stoping the laser whenever it accidentally touches bone marrow. Nevertheless, a registration mechanism between the CT and OCT images is essential to pinpointing the surface reference for accurate depth measurements. In the future, we need to investigate the implementation of such a method to register the OCT image position relative to the CT image position and track the ablation site in real time. Additionally, this paper only demonstrated and validated the performance of real-time tissue detection using deep learning. In this way, we avoided bias in validating the performance between depth control and tissue detection.

In future work, we will also investigate multilayered tissue detection (segmentation) from OCT images. It should improve the precision of cuts by providing a prediction of the remaining bone thickness. Then control over the laser pulse energy and repetition rate could be used to slow down the ablation rate whenever the laser pulse is predicted to cut bone marrow. Such a prediction could be implemented using a Kalman filter to predict the trajectory of the ablation after the next laser pulse [82]. Nevertheless, a sufficient number of labeled ground-truth images will be needed to train the deep-learning (segmentation) model. It is a challenge to label such ground-truth images, which may involve a trained OCT image expert. Even more, it will be a challenge to label images taken during laser ablation, which alters the tissue texture in OCT images.

Disclosures. The authors declare no conflicts of interest.

Data availability. The data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. N. Jowett, W. Wöllmer, R. Reimer, J. Zustin, U. Schumacher, P. W. Wiseman, A. M. Mlynarek, A. Böttcher, C. V. Dalchow, B. B. Lörincz, R. Knecht, and R. J. D. Miller, "Bone ablation without thermal or acoustic mechanical injury via a novel picosecond infrared laser (PIRL)," Otolaryngol. Neck Surg. **150**(3), 385–393 (2014).
2. Y.-M. Lee, R. Y. Tu, A. Chiang, and Y.-C. Huang, "Average-power mediated ultrafast laser osteotomy using a mode-locked Nd:YVO$_4$ laser oscillator," J. Biomed. Opt. **12**(6), 060505 (2007).
3. B. Girard, K. Franjic, M. Cloutier, D. Wilson, C. Clokie, B. Wilson, and R. Miller, "Bone surgery with femtosecond laser compared to mechanical instruments: Healing studies," Proc. SPIE **6261**, 62612J (2006).
4. M. Ivanenko, M. Werner, S. Afilal, M. Klasing, and P. Hering, "Ablation of hard bone tissue with pulsed Co$_2$ lasers," Med. Laser Appl. **20**(1), 13–23 (2005).
5. R. Mauceri, V. Panzarella, L. Maniscalco, A. Bedogni, M. Licata, A. Albanese, F. Toia, E. Cumbo, G. Mazzola, O. Di Fede, and G. Campisi, "Conservative surgical treatment of bisphosphonate-related osteonecrosis of the jaw with Er,Cr:YsGg laser and platelet-rich plasma: a longitudinal study," BioMed Res. Int. **2018**, 1–10 (2018).
6. Y. Ohsugi, A. Aoki, K. Mizutani, S. Katagiri, M. Komaki, M. Noda, T. Takagi, S. Kakizaki, W. Meinzer, and Y. Izumi, "Evaluation of bone healing following Er:YAG laser ablation in rat calvaria compared with bur drilling," J. Biophotonics **12**(3), e201800245 (2019).
7. D. G. Panduric, I. B. Juric, S. Music, K. Molcanov, M. Susic, and I. Anic, "Morphological and ultrastructural comparative analysis of bone tissue after Er:YAG laser and surgical drill osteotomy," Photomed. Laser Surg. **32**(7), 401–408 (2014).
8. K.-U. Lewandrowski, C. Lorente, K. T. Schomacker, T. J. Fiotte, J. W. Wilkes, and T. F. Deutsch, "Use of the Er:YAG laser for improved plating in maxillofacial surgery: Comparison of bone healing in laser and drill osteotomies," Lasers Surg. Med. **19**(1), 40–45 (1996).
9. G. M. Peavy, L. Reinisch, J. T. Payne, and V. Venugopalan, "Comparison of cortical bone ablations by using infrared laser wavelengths 2.9 to 9.2 micrometer," Lasers Surg. Med. **25**(5), 421–434 (1999).
10. S. Stübinger, "Advances in bone surgery: the Er:YAG laser in oral surgery and implant dentistry," Clin. cosmetic and investigational dentistry **2**, 47–62 (2010).

11. H. Abbasi, L. Beltran, G. Rauter, R. Guzman, P. C. Cattin, and A. Zam, "Effect of cooling water on ablation in Er:YAG laserosteotome of hard bone," in *Third International Conference on Applications of Optics and Photonics*, vol. 10453 (SPIE, 2017).

12. L. Beltran, H. Abbasi, G. Rauter, N. Friederich, P. Cattin, and A. Zam, "Effect of laser pulse duration on ablation efficiency of hard bone in microseconds regime," Proc. SPIE **10453**, 104531S (2017).

13. L. Beltrán Bernal, F. Canbaz, A. Droneau, N. Friederich, P. Cattin, and A. Zam, "Optimizing deep bone ablation by means of a microsecond Er:YaG laser and a novel water microjet irrigation system," Biomed. Opt. Express **11**(12), 7253 (2020).

14. A. Vogel and V. Venugopalan, "Mechanisms of pulsed laser ablation of biological tissues," Chem. Rev. **103**(2), 577–644 (2003).

15. S. R. Visuri, J. T. Walsh Jr., and H. A. Wigdor, "Erbium laser ablation of dental hard tissue: Effect of water cooling," Lasers Surg. Med. **18**(3), 294–300 (1996).

16. M. Hohmann, D. Dörner, F. Mehari, C. Chen, M. Späth, S. Müller, H. Albrecht, F. Klämpfl, and M. Schmidt, "Investigation of random lasing as a feedback mechanism for tissue differentiation during laser surgery," Biomed. Opt. Express **10**(2), 807–816 (2019).

17. R. Kanawade, F. Mahari, F. Klämpfl, M. Rohde, C. Knipfer, K. Tangermann-Gerk, W. Adler, M. Schmidt, and F. Stelzle, "Qualitative tissue differentiation by analysing the intensity ratios of atomic emission lines using laser induced breakdown spectroscopy (libs): prospects for a feedback mechanism for surgical laser systems," J. Biophotonics **8**(1-2), 153–161 (2015).

18. F. Mehari, M. Rohde, C. Knipfer, R. Kanawade, F. Klämpfl, W. Adler, F. Stelzle, and M. Schmidt, "Laser induced breakdown spectroscopy for bone and cartilage differentiation - ex vivo study as a prospect for a laser surgery feedback mechanism," Biomed. Opt. Express **5**(11), 4013–4023 (2014).

19. M. Rohde, F. Mehari, F. Klämpfl, W. Adler, F.-W. Neukam, M. Schmidt, and F. Stelzle, "The differentiation of oral soft- and hard tissues using laser induced breakdown spectroscopy - a prospect for tissue specific laser surgery," J. Biophotonics **10**(10), 1250–1261 (2017).

20. K. Henn, G. G. Gubaidullin, J. Bongartz, J. Wahrburg, H. Roth, and M. Kunkel, "A spectroscopic approach to monitor the cut processing in pulsed laser osteotomy," Lasers Med. Sci. **28**(1), 87–92 (2013).

21. H. Abbasi, L. M. Beltrán Bernal, A. Hamidi, A. Droneau, F. Canbaz, R. Guzman, S. L. Jacques, P. C. Cattin, and A. Zam, "Combined nd:yag and Er:YAG lasers for real-time closed-loop tissue-specific laser osteotomy," Biomed. Opt. Express **11**(4), 1790 (2020).

22. F. Mehari, M. Rohde, R. Kanawade, C. Knipfer, W. Adler, F. Klämpfl, F. Stelzle, and M. Schmidt, "Investigation of the differentiation of ex vivo nerve and fat tissues using laser-induced breakdown spectroscopy (libs): Prospects for tissue-specific laser surgery," J. Biophotonics **9**(10), 1021–1032 (2016).

23. H. Abbasi, G. Rauter, R. Guzman, P. C. Cattin, and A. Zam, "Laser-induced breakdown spectroscopy as a potential tool for autocarbonization detection in laserosteotomy," J. Biomed. Opt. **23**(7), 1–7 (2018).

24. B. Lengenfelder, F. Mehari, M. Hohmann, M. Heinlein, E. Chelales, M. J. Waldner, F. Klämpfl, Z. Zalevsky, and M. Schmidt, "Remote photoacoustic sensing using speckle-analysis," Sci. Rep. **9**(1), 1057 (2019).

25. B. Lengenfelder, F. Mehari, M. Hohmann, C. Löhr, M. J. Waldner, M. Schmidt, Z. Zalevsky, and F. Klämpfl, "Contact-free endoscopic photoacoustic sensing using speckle analysis," J. Biophotonics **12**(12), e201900130 (2019).

26. H. Nguendon Kenhagho, G. Rauter, R. Guzman, P. C. Cattin, and A. Zam, "Optoacoustic tissue differentiation using a mach-zehnder interferometer," IEEE Trans. Ultrason., Ferroelect., Freq. Contr. **66**(9), 1435–1443 (2019).

27. H. Nguendon Kenhagho, S. Shevchik, F. Saeidi, N. Faivre, B. Meylan, G. Rauter, R. Guzman, P. Cattin, K. Wasmer, and A. Zam, "Characterization of ablated bone and muscle for long-pulsed laser ablation in dry and wet conditions," Materials **12**(8), 1338 (2019).

28. H. Nguendon Kenhagho, F. Canbaz, T. E. Gomez Alvarez-Arenas, R. Guzman, P. Cattin, and A. Zam, "Machine learning-based optoacoustic tissue classification method for laser osteotomes using an air-coupled transducer," Lasers Surg. Med. **53**(3), 377–389 (2021).

29. S. Rupprecht, K. Tangermann-Gerk, J. Wiltfang, F. W. Neukam, and A. Schlegel, "Sensor-based laser ablation for tissue specific cutting: an experimental study," Lasers Med. Sci. **19**(2), 81–88 (2004).

30. S. Rupprecht, K. Tangermann, P. Kessler, F. W. Neukam, and J. Wiltfang, "Er:yag laser osteotomy directed by sensor controlled systems," J. Cranio-Maxillofacial Surg. **31**(6), 337–342 (2003).

31. A. S. Novikov, I. Usenov, D. Schweda, P. Caffier, B. Limmer, V. Artyushenko, and H. J. Eichler, "Pyrometry with flexible infrared fibers for temperature-controlled laser surgery," Biomed. Opt. Express **13**(2), 744–760 (2022).

32. L. A. Reisner, B. W. King, M. D. Klein, G. W. Auner, and A. K. Pandya, "A prototype biosensor-integrated image-guided surgery system," Int. J. Med. Robotics Comput. Assist. Surg. **3**(1), 82–88 (2007).

33. P. C. Ashok, M. E. Giardini, K. Dholakia, and W. Sibbett, "A raman spectroscopy bio-sensor for tissue discrimination in surgical robotics," J. Biophotonics **7**(1-2), 103–109 (2014).

34. F. Stelzle, C. Knipfer, W. Adler, M. Rohde, N. Oetter, E. Nkenke, M. Schmidt, and K. Tangermann-Gerk, "Tissue discrimination by uncorrected autofluorescence spectra: A proof-of-principle study for tissue-specific laser surgery," Sensors **13**(10), 13717–13731 (2013).

35. A. Zam, "Optical tissue differentiation for sensor-controlled tissue-specific laser surgery," doctoral thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (2011).

36. R. Gunaratne, I. Monteath, J. Goncalves, R. Sheh, C. N. Ironside, M. Kapfer, R. Chipper, B. Robertson, R. Khan, and D. Fick, "Machine learning classification of human joint tissue from diffuse reflectance spectroscopy data," Biomed. Opt. Express **10**(8), 3889–3898 (2019).

37. A. Zam, F. Stelzle, K. Tangermann-Gerk, W. Adler, E. Nkenke, F. W. Neukam, M. Schmidt, and A. Douplik, "In vivo soft tissue differentiation by diffuse reflectance spectroscopy: preliminary results," Phys. Procedia **5**, 655–658 (2010).

38. F. Stelzle, A. Zam, W. Adler, K. Tangermann-Gerk, A. Douplik, E. Nkenke, and M. Schmidt, "Optical nerve detection by diffuse reflectance spectroscopy for feedback controlled oral and maxillofacial laser surgery," J. Transl. Med. **9**(1), 20 (2011).

39. F. Stelzle, K. Tangermann-Gerk, W. Adler, A. Zam, M. Schmidt, A. Douplik, and E. Nkenke, "Diffuse reflectance spectroscopy for optical soft tissue differentiation as remote feedback control for tissue-specific laser surgery," Lasers Surg. Med. **42**(4), 319–325 (2010).

40. M. E. Brezinski, G. J. Tearney, B. E. Bouma, J. A. Izatt, M. R. Hee, E. A. Swanson, J. F. Southern, and J. G. Fujimoto, "Optical coherence tomography for optical biopsy," Circulation **93**(6), 1206–1213 (1996).

41. J. G. Fujimoto, C. Pitris, S. A. Boppart, and M. E. Brezinski, "Optical coherence tomography: An emerging technology for biomedical imaging and optical biopsy," Neoplasia **2**(1-2), 9–25 (2000).

42. A. Hamidi, Y. A. Bayhaqi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Long-range optical coherence tomography with extended depth-of-focus: avisual feedback system for smart laser osteotomy," Biomed. Opt. Express **12**(4), 2118–2133 (2021).

43. J. Fujimoto and W. Drexler, *Introduction to Optical Coherence Tomography* (Springer Berlin Heidelberg, 2008), pp. 1–45.

44. O. Müller, S. Donner, T. Klinder, R. Dragon, I. Bartsch, F. Witte, A. Krüger, A. Heisterkamp, and B. Rosenhahn, "Model based 3d segmentation and oct image undistortion of percutaneous implants," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, G. Fichtinger, A. Martel, and T. Peters, eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011), pp. 454–462.

45. D. C. Fernandez, "Delineating fluid-filled region boundaries in optical coherence tomography images of the retina," IEEE Trans. Med. Imaging **24**(8), 929–945 (2005).

46. C. A. Puliafito, M. R. Hee, C. P. Lin, E. Reichel, J. S. Schuman, J. S. Duker, J. A. Izatt, E. A. Swanson, and J. G. Fujimoto, "Imaging of macular diseases with optical coherence tomography," Ophthalmology **102**(2), 217–229 (1995).

47. G. J. Ughi, T. Adriaenssens, P. Sinnaeve, W. Desmet, and J. D'hooge, "Automated tissue characterization of in vivo atherosclerotic plaques by intravascular optical coherence tomography images," Biomed. Opt. Express **4**(7), 1014–1030 (2013).

48. S. Liu, Y. Sotomi, J. Eggermont, G. Nakazawa, S. Torii, T. Ijichi, Y. Onuma, P. W. Serruys, B. P. F. Lelieveldt, and J. Dijkstra, "Tissue characterization with depth-resolved attenuation coefficient and backscatter term in intravascular optical coherence tomography images," J. Biomed. Opt. **22**(09), 1 (2017).

49. L. S. Athanasiou, C. V. Bourantas, G. Rigas, A. I. Sakellarios, T. P. Exarchos, P. K. Siogkas, A. Ricciardi, K. K. Naka, M. I. Papafaklis, L. K. Michalis, F. Prati, and D. I. Fotiadis, "Methodology for fully automated segmentation and plaque characterization in intracoronary optical coherence tomography images," J. Biomed. Opt. **19**(2), 026009 (2014).

50. J. Yang, B. Zhang, H. Wang, F. Lin, Y. Han, and X. Liu, "Automated characterization and classification of coronary atherosclerotic plaques for intravascular optical coherence tomography," Biocybern. Biomed. Eng. **39**(3), 719–727 (2019).

51. J. J. Rico-Jimenez, D. U. Campos-Delgado, M. Villiger, K. Otsuka, B. E. Bouma, and J. A. Jo, "Automatic classification of atherosclerotic plaques imaged with intravascular oct," Biomed. Opt. Express **7**(10), 4069–4085 (2016).

52. N. Katta, A. D. Estrada, A. B. McElroy, A. Gruslova, M. Oglesby, A. G. Cabe, M. D. Feldman, R. D. Fleming, A. J. Brenner, and T. E. Milner, "Laser brain cancer surgery in a xenograft model guided by optical coherence tomography," Theranostics **9**(12), 3555–3564 (2019).

53. Y. Fan, B. Zhang, W. Chang, X. Zhang, and H. Liao, "A novel integration of spectral-domain optical-coherence-tomography and laser-ablation system for precision treatment," International Journal of Computer Assisted Radiology and Surgery **13**(3), 411–423 (2018).

54. Y. Zhang, T. Pfeiffer, M. Weller, W. Wieser, R. Huber, J. Raczkowsky, J. Schipper, H. Wörn, and T. Klenzner, "Optical coherence tomography guided laser cochleostomy: Towards the accuracy on tens of micrometer scale," BioMed Res. Int. **2014**, 1–10 (2014).

55. J. D. Díaz, D. Kundrat, K.-F. Goh, O. Majdani, and T. Ortmaier, *Towards Intra-operative OCT Guidance for Automatic Head Surgery: First Experimental Results* (Lecture Notes in Computer Science, 2013), p. 347–354.

56. A. Fuchs, M. Schultz, A. Krüger, D. Kundrat, J. D. Díaz, and T. Ortmaier, "Online measurement and evaluation of the Er:YAG laser ablation process using an integrated OCT system," Biomedical Engineering/Biomedizinische Technik **57**(SI-1 Track-H), 434–437 (2012).

57. J. Jivraj, C. Chen, Y. Huang, J. Ramjist, Y. Lu, B. Vuong, X. Gu, and V. X. D. Yang, "Smart laser osteotomy: integrating a pulsed 1064nm fiber laser into the sample arm of a fiber optic 1310nm OCT system for ablation monitoring," Biomed. Opt. Express **9**(12), 6374–6387 (2018).

58. B. Y. Leung, P. J. Webster, J. M. Fraser, and V. X. Yang, "Real-time guidance of thermal and ultrashort pulsed laser ablation in hard tissue using inline coherent imaging," Lasers Surg. Med. **44**(3), 249–256 (2012).

59. F.-Y. Chang, M.-T. Tsai, Z.-Y. Wang, C.-K. Chi, C.-K. Lee, C.-H. Yang, M.-C. Chan, and Y.-J. Lee, "Optical coherence tomography-guided laser microsurgery for blood coagulation with continuous-wave laser diode," Sci. Rep. **5**(1), 16739 (2015).

60. H. H. Müller, L. Ptaszynski, K. Schlott, C. Debbeler, M. Bever, S. Koinzer, R. Birngruber, R. Brinkmann, and G. Hüttmann, "Imaging thermal expansion and retinal tissue changes during photocoagulation by high speed OCT," Biomed. Opt. Express **3**(5), 1025–1046 (2012).

61. A. Hamidi, Y. A. Bayhaqi, F. Canbaz, A. Navarini, P. C. Cattin, and A. Zam, "Imaging photothermal-induced expansion of bone during laser osteotomy by phase-sensitive OCT: preliminary results," Proc. SPIE **11359**, 113590K (2020).

62. "CUDA Toolkit Documentation," https://docs.nvidia.com/cuda/. Accessed: 2020-09-30.

63. D. M. S. Barros, J. C. C. Moura, C. R. Freire, A. C. Taleb, R. A. M. Valentim, and P. S. G. Morais, "Machine learning applied to retinal image processing for glaucoma detection: review and perspective," BioMedical Engineering OnLine **19**(1), 20 (2020).

64. J. Möller, A. Bartsch, M. Lenz, I. Tischoff, R. Krug, H. Welp, M. R. Hofmann, K. Schmieder, and D. Miller, "Applying machine learning to optical coherence tomography images for automated tissue classification in brain metastases," Int. J. Comput. Assist. Radiol. Surg. **16**(9), 1517–1526 (2021).

65. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016). http://www.deeplearningbook.org.

66. O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Convolutional Neural Networks* (Springer International Publishing, 2022), pp. 533–577.

67. A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," Biomed. Opt. Express **8**(8), 3627–3642 (2017).

68. S. Borkovkina, A. Camino, W. Janpongsri, M. V. Sarunic, and Y. Jian, "Real-time retinal layer segmentation of OCT volumes with GPU accelerated inferencing using a compressed, low-latency neural network," Biomed. Opt. Express **11**(7), 3968–3984 (2020).

69. Y. A. Bayhaqi, A. Hamidi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Deep learning models comparison for tissue classification using optical coherence tomography images: toward smart laser osteotomy," OSA Continuum **4**(9), 2510–2526 (2021).

70. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015*, Y. Bengio and Y. LeCun, eds. (2015).

71. K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision - ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, eds. (Springer International Publishing, 2016), pp. 630–645.

72. G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), pp. 2261–2269.

73. J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-8**(6), 679–698 (1986).

74. S. Luo, J. Yang, Q. Gao, S. Zhou, and C. A. Zhan, "The edge detectors suitable for retinal OCT image segmentation," J. Healthcare Eng. **2017**, 1–13 (2017).

75. A. Paszke, S. Gross, and F. Massa, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, (Curran Associates, Inc., 2019), pp. 8024–8035.

76. "Open Neural Network Exchange," https://onnx.ai/. Accessed: 2020-09-30.

77. "NVIDIA TensorRT," https://developer.nvidia.com/tensorrt. Accessed: 2020-09-30.

78. "Matlab: Registration Estimator App," https://ch.mathworks.com/help/images/approaches-to-registering-images.html. Accessed: 2020-09-30.

79. M. Forrer, M. Frenz, V. Romano, H. Altermatt, H. Weber, A. Silenok, M. Istomyn, and V. Konov, "Bone-ablation mechanism using co 2 lasers of different pulse duration and wavelength," Appl. Phys. B **56**(2), 104–112 (1993).

80. T. Aoki, S. Yamaguchi, S. Kinoshita, Y. Hayashida, and Y. Korogi, "Quantification of bone marrow fat content using iterative decomposition of water and fat with echo asymmetry and least-squares estimation (ideal): reproducibility, site variation and correlation with age and menopause," The Br. J. Radiol. **89**(1065), 20150538 (2016).

81. Y. Hsu, S. Lee, and H. Lin, "A modular mechatronic system for automatic bone drilling," Biomed. Eng. - Appl. Basis Commun. **13**(04), 168–174 (2001).

82. Kalman filtered depth prediction using Optical Coherence Tomography for laser bone cutting, Zenodo, 2021, https://doi.org/10.5281/zenodo.4922992.

Discussion and Conclusion

7.1 Discussion

The work of this thesis aimed to develop a smart monitoring system for laser osteotomy so that critical tissues, such as nerves, could be avoided. The use of an OCT imaging system for monitoring and feedback was proposed because it could provide a high-resolution image slice over the ablation site, in real time. In this thesis, a "smart" feedback system, rephrased as "smart laser osteotomy," was emphasized by developing a deep learning-based feedback system. Specifically, deep neural network models were used for denoising the OCT images and for tissue classification based on the OCT images. The last three chapters aimed to demonstrate the results of using deep learning methods for the smart laser osteotomy system. This section discusses the results of the publications in each chapter.

The scheme of the feedback system was divided into three steps or subprocesses (see Figure 3.7), (1) acquisition and denoising, (2) patch tracking, and (3) tissue classification. For the first step of acquiring images, the SS-OCT was selected for its acquisition performance, which best supports our real-time feedback system. It is necessary to discuss the swept source laser used in this thesis. Two SS-OCT devices with different light sweeping sources were used: a short cavity laser and an Akinetic swept laser. Images taken from the OCT system with the short cavity laser source, Axsun Technologies, were used in our offline experiments (Chapters 4 and 5). At the same time, a long-range OCT setup with an Akinetic swept laser, Insight Photonic Solution, was being developed. This second OCT setup was intended to provide an imaging range of more than $1\,cm$, which is helpful in monitoring deep bone ablation and was used for the last publication (Chapter 6).

Furthermore, the first step also includes the OCT image denoising process. Two publications (Chapter 4) demonstrated our experiment with using deep learning models to resemble the frame-averaging method and prove that image denoising improves the accuracy of the tissue classifier. Although the results show the superiority of the frame-averaging method, the deep learning models provide faster image denoising than the frame-averaging method and other classical denoising filters. The results suggested that

the supervised learning models, U-Net and ResNet, were sufficient for real-time OCT image denoising.

The second subprocess was patch tracking. A simple Canny edge detection method was sufficient to track the surface of the tissue in the OCT image. Two pieces of information could be extracted from the surface detection result. The first was the ablation crater location. The feedback system was supposed to identify tissue type at the center of the ablation crater. Therefore. surface detection, in practice, was used to track the crater surface, which gradually gets deeper as the laser pulses continue to ablate. The second was the relative depth of the ablation crater. The relative depth is defined by the distance from the border surface to the bottom of the crater. Several studies have been done on using this information for laser osteotomy. The crater depth information can, over time, even be used to predict the trajectory of the cut. While this thesis work was conducted, a trajectory prediction based on the Kalman filter [121] was investigated, and the results were published in a conference paper (see Appendix).

The last subprocess was tissue classification, an essential part of this project. Two experiments were conducted to find the best deep neural network model and the results were published (see Chapter 5). The DenseNet model was deemed sufficient for tissue classification, with an accuracy of 94.85 % for the offline experiments. It is important to note that the experiments in Chapters 4 and 5 were conducted offline, without laser ablation. This stage was followed by the implementation of the real-time feedback system during laser ablation in Chapter 6.

An overview of the challenges encountered during real-time feedback system application is outlined in Chapter 6. It is worth discussing some limitations of the proposed solutions. One challenge was bone carbonization. An optimal ablation rate with the Er:YAG laser could only be reached by spraying water at the ablation site. Unfortunately, the water particles deteriorate the OCT image and disturb the crater tracking system. A simple pausing mechanism was proposed; however, this method did not optimally solve the problem. Carbonization on the bone surface could still appear. During the laser cutting process, the carbonized part of the bone will evaporate after sequentially rehydrated and reablated. However, when bone marrow is detected, the pausing mechanism is completely stopped, leaving the residual carbonized bone untouched. Therefore, synchronizing OCT detection and water spray may improve the effectiveness of this method.

Another limitation of the current approach is the image patch-based detection system. The deep learning tissue classifier requires the patch be fully filled with the tissue structure. thus, the stop condition only happens when the target surface is fully exposed to bone marrow. To solve this problem, we may have to change the detection method to a multilayered one, such as segmenting the tissue type on the image layer by layer. However, it will be difficult to get a labeled dataset, especially with the ablation and temperature change conditions.

7.2 Future Direction of Research

The use of OCT imaging and a deep learning method to analyze the image offers additional benefits for realizing smart laser osteotomy, compared to other feedback modalities. Publications throughout the thesis project have partly demonstrated this claim. The last paper in particular, as the ultimate result, showed real-time application of the feedback system. However, several other aspects could still be optimized to improve the precision of detection and reduce the perforation depth. In this section, two main directions of future research are considered.

The first main research direction is to improve the feedback mechanism. Ideally, a surgeon could plan the ablation site and depth based on a patient's Computed Tomography (CT) data. The OCT system could be used as a depth-control mechanism during ablation. Concurrently, patch-based tissue detection could be used as a second parameter to support depth control, which in turn could be used to stop the laser whenever it touches bone marrow. In this situation, a registration mechanism between the CT and OCT images would be essential for pinpointing the surface reference for accurate depth measurements; thus, a study on real-time image registration between the OCT image and CT image is an interesting topic for future research.

Alternatively, segmentation of the tissue layer will give more precise information about the target and the surrounding tissue geometries. The layer-by-layer segmentation result can be used to estimate the remaining target (bone) thickness. Then, by controlling the Er:YAG laser pulse energy and repetition rate, one could slow down the ablation rate whenever the laser pulse is predicted to cut bone marrow. Depth prediction can be implemented using a Kalman filter [121] that predicts the trajectory of the ablation for the next laser pulse based on the history of the ablation rate from the previous laser pulse (see Appendix). More interestingly, the real-time application of deep-learning-based image denoising would further improve segmentation accuracy. It would also enable the detection of more than two tissue types.

This mechanism is worth investigating in future research. However, it will be challenging to label a sufficient number of ground-truth images, especially images taken during laser ablation, to train the deep learning (segmentation) model. While the laser cut could be more precise by controlling the laser energy and repetition rate, unfortunately, the Er:YAG laser used in this thesis project did not have the interface to control pulse energy or repetition rate in real-time. Therefore, a real-time control interface for the Er:YAG laser is compulsory for future research.

The second main future research direction is related to the Er:YAG laser beam shape and irrigation strategy. Feedback system accuracy could be improved by engineering the Er:YAG laser beam shape from a near Gaussian shape to a top hat beam shape. The OCT could precisely cover the ablation area with the top hat beam shape. It will eliminate the overlapping area (hotspot) problem for the line ablation (see Section 3.3 in the last publication). Additionally, the OCT-guided laser osteotomy could also be improved by optimizing the irrigation strategy. There is a conflict between the need for water irrigation to minimize the carbonization effect and increase the ablation efficiency of the Er:YAG

laser, on the one hand, and the role of water droplets in deteriorating the OCT image and reducing the feedback system's accuracy, on the other. In the future, research to synchronize the OCT detection and water spray is necessary to improve the effectiveness of the OCT-based feedback system.

Finally, this thesis project was initiated as part of the (MIRACLE) project at the University of Basel. This multidisciplinary (robotic, implant, and navigation) project endeavors to develop a minimally invasive robot-assisted computer-guided laser osteotomy. In other words, the project required a miniaturized feedback system to achieve the least invasive laser osteotomy; thus, integrating the automatic tissue characterization system into a miniaturized OCT system will be another target of future research.

7.3 Conclusion

This thesis aimed to develop an automatic tissue characterization system, emphasizing the use of deep learning for a smart laser osteotomy. Several experiments were conducted to support the claims and published in journals. In Chapter 4, a deep-learning-based OCT image denoising process was shown to improve the accuracy of a tissue classifier. The publication in Chapter 5 attempts to find the best tissue classifier in offline conditions. Here, DenseNet was selected as the best candidate. The last publication in Chapter 6 demonstrated a real-time implementation of the deep learning-based OCT-guided laser osteotomy. The average maximum perforation depth in bone marrow was only $216\,\mu$m. Further investigations are required to improve the accuracy and precision of using OCT for smart laser osteotomy. These investigations include optimizing the OCT image detection system and the ablation strategy, as described in the previous section. The results offer more evidence for replacing mechanical tools with a safer and less invasive laser osteotomy.

Appendix

Depth Prediction with Kalman Filter

Surface detection is an essential subprocess of the feedback system. The detection output is used to track the ablation location over depth. During implementation, the detection rate was found to be relatively faster than the ablation laser pulse repetition rate, thereby enabling tissue surface prediction with the Kalman filter [121]. Specifically, the prediction works by measuring the ablation rate of every ablating laser pulse, based on the history of the surface shape. The measured ablation rate can be used to predict the tissue surface shape for the next consecutive ablating laser pulse. This prediction prevents the ablation from overcutting the target tissue beyond what is planned. A study of the prediction performance was published as a conference proceeding.

Publication: Y. A. Bayhaqi, A. Hamidi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Kalman filtered depth prediction using Optical Coherence Tomography for laser bone cutting", in *Automation in Medical Engineering* (AUTOMED 2021), Zenodo, 2021.

Kalman filtered depth prediction using Optical Coherence Tomography for laser bone cutting

Y. A. Bayhaqi[1*], A. Hamidi[1], F. Canbaz[1], A. A. Navarini[2], P. C. Cattin[3], and A. Zam[1*]

[1] *Biomedical Laser and Optics Group (BLOG), Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland*

[2] *Digital Dermatology Group, Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland*

[3] *Center for medical Image Analysis and Navigation (CIAN), Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland*

[*] *Corresponding author, email: yakub.bayhaqi@unibas.ch, azhar.zam@unibas.ch*

Abstract: Laser osteotomy may benefits patients with minor damage to the bone and accelerated the healing process. However, the lack of depth control is still needed to be overcome. We proposed to use Optical Coherence Tomography imaging for depth control of the laser ablation process. The OCT image provides geometrical information over the ablation spot and enables the measurement of depth in real-time. Additionally, we demonstrated the use of Kalman filter to predict incoming pulse ablation depth and avoid excessive cut than planned. The result shows that the Kalman filter reduced the error between the planned and real ablated depth.

I. Introduction

The use of lasers in medical applications is still an interesting focus of research in recent years. Lasers have been used in several surgical procedures such as cataracts, endodontic, and tumor removal [1,2]. Particularly in bone surgery, lasers are used due to the advantages such as minor damage to the bone and accelerated the healing process [3]. However, the lack of depth control in laser still needs to be overcome. This depth control will help the surgeon avoid cutting an excessive amount of bone than planned. Several approaches have been proposed to predict the laser ablation process kinematic such as solving the forward and inverse problem [4,5]. However, these methods are impractical and complex. Here, extensive databases of the ablation process condition are needed. It will become complex when we face a heterogenous or multilayered tissue.

This paper proposed to reflect the use of Optical Coherence Tomography (OCT) imaging system to monitor the laser ablation process [4,6]. We demonstrated and extended the OCT depth measurement to predict the next incoming pulse ablation depth by employing the Kalman filter [7]. This prediction will help prevent the laser from overcutting the bone, which may arise due to the passive component's delayed response to stop the laser.

II. Material and methods

The combination setup between the Er:YAG ablation laser system (LITETOUCH by Syneron, working wavelength of 2.94 μm) and the OCT imaging system is illustrated in Fig 1. Both laser beams were combined with a dichroic mirror. An optical beam shutter (Thorlabs SH1) is placed in front of the Er:YAG laser to stop the ablation laser pulse.
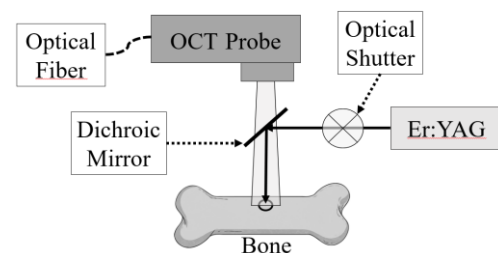
Figure 1. Combination setup between the Er:YAG ablation laser and OCT. Both laser beams were combined with a dichroic mirror.
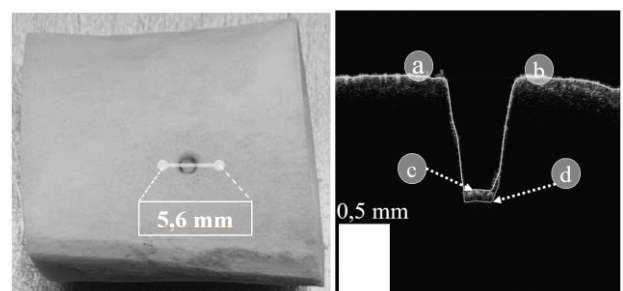
Figure 2. Example of the ablated bone (left) and the OCT image scanned over the ablated crater (right). In the OCT images, the measured depth was defined as the averaged distance between circle (a) and (b) to the bottom of the crater at circle (c). Meanwhile, circle (d) represents the predicted bone surface for the incoming laser ablation pulse.

B-scan image frames were streamed with a custom made OCT system equipped with an Axsun swept laser source (λ_0 = 1060 nm, $\Delta\lambda$ = 100 nm, and sweep rate = 100 kHz). The

corresponding lateral and axial resolutions were 44 μm and 10 μm, respectively. The system enables image acquisition with a size of 1024 x 400 pixels (3.6 mm depth x 5.6 mm lateral) and previews the ablation process in real-time. We set the OCT system's acquisition rate to 8.5 frames per second after averaging 20 frames to provide a good image quality.

The bone surface was extracted from the OCT image by employing the Canny edge detection algorithm [8]. The measured depth was defined as the relative distance between the adjacent surface and the bottom of the crater (illustrated on the right side of Fig. 2). We then used the depth measurement as feedback to close the optical beam shutter before the measured or predicted depth reached the targeted depth.

I.I. Kalman filter depth prediction

Kalman filter produces estimates of unknown variables that tend to be more accurate than those based on a single measurement alone. This filter has been used in numerous technologies, particularly in distance measurement sensors [7]. The Kalman filter used in the experiment was based on the constant velocity model, which assumes that the velocity is constant during a sampling interval and expressed as

$$\begin{bmatrix} x_t \\ v_t \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} x_{t-1} + w_t \tag{1}$$

where x_t, v_t and w_t are the predicted actual state of depth, ablation rate, and process noise at time t, respectively. This model has been used in many applications because of its versatility, effectiveness, and simplicity.

I.I. Experiment setup

The tissues used in this experiment were porcine femur bone samples. We tested the feedback system for cutting the bone with the targeted depth of 703.13 μm and 1406.25 μm. We also compared the result when we employ the Kalman filter for prediction. The laser ablation energy was set to 450 mJ per pulse. The bone was ablated in repetition rates of 1 Hz, 5 Hz, and 10 Hz, respectively.

III. Results and discussion

The results of the experiment are shown in Table 1. In the absence of the Kalman filter, the feedback system always passes the targeted depth. These errors mainly due to the communication delay of the optical shutter. Furthermore—when the Er:YAG laser repetition rate was set to 10 Hz—the slower acquisition rate of the OCT also contributes to error, because it may miss an ablation pulse measurement.

Furthermore, the results show that the Kalman filter significantly reduces the error of ablation. In a repetition rate of 1 Hz, the filter stopped the ablation even before the targeted depth (premature stop), which indicates that the incoming pulse will ablate deeper than the targeted depth. However, similarly to the absence of the Kalman filter, the feedback system exceeds the targeted depth when the laser repetition rate was set to 5 Hz and 10 Hz.

Table 1. Final measured depth after stopping the ablation using the proposed feedback system. A negative error value indicates premature stop.

Kalman Filter	Target Depth (μm)	Repetition rate	Measured Depth (μm)	Prediction Error (μm)
Off	703.13	1 Hz	712.828 ± 10.616	9.698 ± 10.616
		5 Hz	738.422 ± 12.909	35.292 ± 12.909
		10 Hz	774.492 ± 13.977	71.362 ± 13.977
	1406.25	1 Hz	1413.703 ± 12.767	7.453 ± 12.767
		5 Hz	1436.906 ± 13.448	30.656 ± 13.448
		10 Hz	1469.813 ± 8.114	63.563 ± 8.114
On	703.13	1 Hz	691.523 ± 4.313	-11.607 ± 4.313
		5 Hz	716.484 ± 16.704	13.354 ± 16.704
		10 Hz	729.352 ± 23.187	26.222 ± 23.187
	1406.25	1 Hz	1383.539 ± 18.626	-22.711 ± 18.626
		5 Hz	1414.898 ± 27.777	8.648 ± 27.777
		10 Hz	1437.188 ± 23.734	30.938 ± 23.734

IV. Conclusions and future work

We demonstrated the use of OCT to monitor the laser ablation process and the Kalman filter to predict incoming ablation depth to reduced cutting an excessive amount of bone than planned. However, further investigation is needed to demonstrate the benefit of using the Kalman filter for deeper ablation prediction. In the future, we will investigate more in finding a better kinematic model of the Kalman filter. We will also explore how to overcome the response delay problem (e.g., increasing our OCT system's acquisition rate and defining an offset to the target depth).

REFERENCES
[1] B. Azadgoli and R. Y. Baker, "Laser applications in surgery," *Ann Transl Med*, vol. 4, no. 23, pp. 452-452, 2016, doi: 10.21037/atm.2016.11.51.
[2] E. Khalkhal, M. Rezaei-Tavirani, M. R. Zali, and Z. Akbari, "The Evaluation of Laser Application in Surgery: A Review Article," *J Lasers Med Sci,* vol. 10, no. Suppl 1, pp. S104-S111, Fall 2019, doi: 10.15171/jlms.2019.S18.
[3] S. Stübinger, S. Ghanaati, B. Saldamli, C. J. Kirkpatrick, and R. Sader, "Er:YAG laser osteotomy: preliminary clinical and histological results of a new technique for contact-free bone surgery," *Eur Surg Res*, vol. 42, no. 3, pp. 150-6, 2009, doi: 10.1159/000197216.
[4] L. M. B. Bernal *et al.*, "Optimizing controlled laser cutting of hard tissue (bone)," *at - Automatisierungstechnik,* vol. 66, no. 12, pp. 1072-1082, 2018, doi: doi:10.1515/auto-2018-0072.
[5] U. Nahum, A. Zam, and P. C. Cattin, "Bone Reconstruction and Depth Control During Laser Ablation," in *Computational Methods and Clinical Applications in Musculoskeletal Imaging*, Cham, T. Vrtovec, J. Yao, G. Zheng, and J. M. Pozo, Eds., 2019// 2019: Springer International Publishing, pp. 126-135.
[6] A. Hamidi *et al.*, "Long-range optical coherence tomography with extended-depth-of-focus: A visual feedback system for smart laser osteotomy," *Bio. Opt. Express,* 2021 (Accepted to be published)
[7] B. Ekstrand, "Some Aspects on Filter Design for Target Tracking," *Journal of Control Science and Engineering,* vol. 2012, p. 870890, 2012/05/08 2012, doi: 10.1155/2012/870890.
[8] S. Luo, J. Yang, Q. Gao, S. Zhou, and C. a. A. Zhan, "The Edge Detectors Suitable for Retinal OCT Image Segmentation," *J Healthc Eng,* vol. 2017, pp. 3978410-3978410, 2017, doi: 10.1155/2017/3978410.

Bibliography

[1] K. Rönn, N. Reischl, E. Gautier, and M. Jacobi, "Current surgical treatment of knee osteoarthritis," Arthritis **2011** (2011).

[2] D. R. Wenger and J. Bomar, "Human hip dysplasia: evolution of current treatment concepts," J. Orthop. Sci. **8**, 264–271 (2003).

[3] K. C. Kose, O. Bozduman, A. E. Yenigul, and S. Igrek, "Spinal osteotomies: indications, limits and pitfalls," EFORT Open Rev. **2**, 73 – 82 (2017).

[4] L. J.-H. Kim Ki-Tack, Park Kyoung-Jun, "Osteotomy of the spine to correct the spinal deformity," Asian Spine J **3**, 113–123 (2009).

[5] S. R. Aziz, "Simon p. hullihen and the origin of orthognathic surgery," J. oral maxillofacial surgery **62**, 1303–1307 (2004).

[6] H. J. Seo and Y.-K. Choi, "Current trends in orthognathic surgery," Arch. Craniofacial Surg. **22**, 287–295 (2021).

[7] S. Lustig, F. Khiami, P. Boyer, Y. Catonne, G. Deschamps, and P. Massin, "Posttraumatic knee osteoarthritis treated by osteotomy only," Orthop. & Traumatol. Surg. & Research **96**, 856–860 (2010).

[8] A. Trompeter, J. Dabis, O. Templeton-Ward, A. E. Lacey, and B. Narayan, "The history, evolution and basic science of osteotomy techniques," Strateg. Trauma Limb Reconstr. **12**, 169–180 (2017).

[9] L. Gao, H. Madry, D. V. Chugaev, M. Denti, A. Frolov, M. Burtsev, N. Magnitskaya, V. Mukhanov, P. Neyret, L. N. Solomin, and et al., "Advances in modern osteotomies around the knee," J. Exp. Orthop. **6** (2019).

[10] N. Davis, A. Curry, A. K. Gambhir, H. Panigrahi, C. R. C. Walker, E. G. L. Wilkins, M. A. Worsley, and P. R. Kay, "Intraoperative bacterial contamination in operations for joint replacement," The J. Bone Jt. Surgery. Br. volume **81-B**, 886–889 (1999).

[11] M. Takamoto, M. Takechi, K. Ohta, Y. Ninomiya, S. Ono, H. Shigeishi, M. Tada, and N. Kamata, "Risk of bacterial contamination of bone harvesting devices used for autogenous bone graft in implant surgery," Head & Face Med. **9** (2013).

[12] F. J. Manzano-Moreno, F. J. Herrera-Briones, M. Linares-Recatala, F. M. Ocaña-Peinado, C. Reyes-Botella, and M. F. Vallecillo-Capilla, "Bacterial contamination levels of autogenous bone particles collected by 3 different techniques for harvesting intraoral bone grafts," J. Oral Maxillofacial Surg. **73**, 424–429 (2015).

[13] K. W. Baek, W. Deibel, D. Marinov, M. Griessen, A. Bruno, H.-F. Zeilhofer, P. Cattin, and P. Juergens, "Clinical applicability of robot-guided contact-free laser osteotomy in cranio-maxillo-facial surgery: in-vitro simulation and in-vivo surgery in minipig mandibles," Br. J. Oral Maxillofacial Surg. **53**, 976–981 (2015).

[14] K. W. Baek, W. Deibel, D. Marinov, M. Griessen, M. Dard, A. Bruno, H.-F. Zeilhofer, P. Cattin, and P. Juergens, "A comparative investigation of bone surface after cutting with mechanical tools and er:yag laser," Lasers Surg. Med. **47**, 426–432 (2015).

[15] K. W. Baek, M. Dard, H. Zeilhofer, P. C. Cattin, and P. Juergens, "Comparing the bone healing after cold ablation robot-guided er:yag laser osteotomy and piezoelectric osteotomy—a pilot study in a minipig mandible," Lasers Surg. Med. **53**, 291–299 (2021).

[16] M. Hohmann, D. Dörner, F. Mehari, C. Chen, M. Späth, S. Müller, H. Albrecht, F. Klämpfl, and M. Schmidt, "Investigation of random lasing as a feedback mechanism for tissue differentiation during laser surgery," Biomed. Opt. Express **10**, 807–816 (2019).

[17] Q. Song, S. Xiao, Z. Xu, J. Liu, X. Sun, V. Drachev, V. M. Shalaev, O. Akkus, and Y. L. Kim, "Random lasing in bone tissue," Opt. Lett. **35**, 1425–1427 (2010).

[18] R. Kanawade, F. Mahari, F. Klämpfl, M. Rohde, C. Knipfer, K. Tangermann-Gerk, W. Adler, M. Schmidt, and F. Stelzle, "Qualitative tissue differentiation by analysing the intensity ratios of atomic emission lines using laser induced breakdown spectroscopy (libs): prospects for a feedback mechanism for surgical laser systems," J. Biophotonics **8**, 153–161 (2015).

[19] M. Rohde, F. Mehari, F. Klämpfl, W. Adler, F.-W. Neukam, M. Schmidt, and F. Stelzle, "The differentiation of oral soft- and hard tissues using laser induced breakdown spectroscopy – a prospect for tissue specific laser surgery," J. Biophotonics **10**, 1250–1261 (2017).

[20] F. Mehari, M. Rohde, C. Knipfer, R. Kanawade, F. Klämpfl, W. Adler, N. Oetter, F. Stelzle, and M. Schmidt, "Investigation of laser induced breakdown spectroscopy (LIBS) for the differentiation of nerve and gland tissue—a possible application for a laser surgery feedback control mechanism," Plasma Sci. Technol. **18**, 654–660 (2016).

[21] H. Abbasi, L. M. Beltrán Bernal, A. Hamidi, A. Droneau, F. Canbaz, R. Guzman, S. L. Jacques, P. C. Cattin, and A. Zam, "Combined nd:yag and er:yag lasers for

real-time closed-loop tissue-specific laser osteotomy," Biomed. Opt. Express **11**, 1790 (2020).

[22] B. Lengenfelder, F. Mehari, M. Hohmann, M. Heinlein, E. Chelales, M. J. Waldner, F. Klämpfl, Z. Zalevsky, and M. Schmidt, "Remote photoacoustic sensing using speckle-analysis," Sci. Reports **9** (2019).

[23] B. Lengenfelder, F. Mehari, M. Hohmann, C. Löhr, M. J. Waldner, M. Schmidt, Z. Zalevsky, and F. Klämpfl, "Contact-free endoscopic photoacoustic sensing using speckle analysis," J. Biophotonics **12** (2019).

[24] N. Kenhagho, Shevchik, Saeidi, Faivre, Meylan, Rauter, Guzman, Cattin, Wasmer, Zam, and et al., "Characterization of ablated bone and muscle for long-pulsed laser ablation in dry and wet conditions," Materials **12**, 1338 (2019).

[25] H. Nguendon Kenhagho, F. Canbaz, T. E. Gomez Alvarez-Arenas, R. Guzman, P. Cattin, and A. Zam, "Machine learning-based optoacoustic tissue classification method for laser osteotomes using an air-coupled transducer," Lasers Surg. Med. **53**, 377–389 (2021).

[26] L. A. Reisner, B. W. King, M. D. Klein, G. W. Auner, and A. K. Pandya, "A prototype biosensor-integrated image-guided surgery system," The Int. J. Med. Robotics Comput. Assist. Surg. **3**, 82–88 (2007).

[27] P. C. Ashok, M. E. Giardini, K. Dholakia, and W. Sibbett, "A raman spectroscopy bio-sensor for tissue discrimination in surgical robotics," J. Biophotonics **7**, 103–109 (2014).

[28] F. Stelzle, C. Knipfer, W. Adler, M. Rohde, N. Oetter, E. Nkenke, M. Schmidt, and K. Tangermann-Gerk, "Tissue discrimination by uncorrected autofluorescence spectra: A proof-of-principle study for tissue-specific laser surgery," Sensors **13**, 13717–13731 (2013).

[29] A. Zam, "Optical tissue differentiation for sensor-controlled tissue-specific laser surgery," doctoralthesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (2011).

[30] R. Gunaratne, I. Monteath, J. Goncalves, R. Sheh, C. N. Ironside, M. Kapfer, R. Chipper, B. Robertson, R. Khan, D. Fick, and et al., "Machine learning classification of human joint tissue from diffuse reflectance spectroscopy data," Biomed. Opt. Express **10**, 3889 (2019).

[31] A. Zam, F. Stelzle, K. Tangermann-Gerk, W. Adler, E. Nkenke, F. W. Neukam, M. Schmidt, and A. Douplik, "In vivo soft tissue differentiation by diffuse reflectance spectroscopy: preliminary results," Phys. Procedia **5**, 655–658 (2010). Laser Assisted Net Shape Engineering 6, Proceedings of the LANE 2010, Part 2.

[32] F. Stelzle, A. Zam, W. Adler, K. Tangermann-Gerk, A. Douplik, E. Nkenke, and M. Schmidt, "Optical nerve detection by diffuse reflectance spectroscopy for feedback controlled oral and maxillofacial laser surgery," J. Transl. Med. **9**, 20 (2011).

[33] F. Stelzle, K. Tangermann-Gerk, W. Adler, A. Zam, M. Schmidt, A. Douplik, and E. Nkenke, "Diffuse reflectance spectroscopy for optical soft tissue differentiation as remote feedback control for tissue-specific laser surgery," Lasers Surg. Med. **42**, 319–325 (2010).

[34] P. L. Müller, S. Wolf, R. Dolz-Marco, A. Tafreshi, S. Schmitz-Valckenberg, and F. G. Holz, *Ophthalmic Diagnostic Imaging: Retina* (Springer International Publishing, Cham, 2019), pp. 87–106.

[35] J. Fujimoto and W. Drexler, *Introduction to Optical Coherence Tomography* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008), pp. 1–45.

[36] Q. Li, S. Li, Z. He, H. Guan, R. Chen, Y. Xu, T. Wang, S. Qi, J. Mei, and W. Wang, "DeepRetina: Layer Segmentation of Retina in OCT Images Using Deep Learning," Transl. Vis. Sci. & Technol. **9**, 61–61 (2020).

[37] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Y. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, and A. Shi, "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell **172**, 1122–1131.e9 (2018).

[38] Z. Mishra, A. Ganegoda, J. Selicha, Z. Wang, S. R. Sadda, and Z. Hu, "Automated retinal layer segmentation using graph-based algorithm incorporating deep-learning-derived information," Sci. Reports **10**, 9541 (2020).

[39] J. Yoon, J. Han, J. I. Park, J. S. Hwang, J. M. Han, J. Sohn, K. H. Park, and D. D.-J. Hwang, "Optical coherence tomography-based deep-learning model for detecting central serous chorioretinopathy," Sci. Reports **10**, 18852 (2020).

[40] N. Gessert, M. Lutz, M. Heyder, S. Latus, D. M. Leistner, Y. S. Abdelwahed, and A. Schlaefer, "Automatic plaque detection in ivoct pullbacks using convolutional neural networks," IEEE Trans. on Med. Imaging **38**, 426–434 (2019).

[41] A. Abdolmanafi, L. Duong, N. Dahdah, I. R. Adib, and F. Cheriet, "Characterization of coronary artery pathological formations from oct imaging using deep learning," Biomed. Opt. Express **9**, 4936–4960 (2018).

[42] N. Katta, A. D. Estrada, A. B. McElroy, A. Gruslova, M. Oglesby, A. G. Cabe, M. D. Feldman, R. D. Fleming, A. J. Brenner, and T. E. Milner, "Laser brain cancer surgery in a xenograft model guided by optical coherence tomography," Theranostics **9**, 3555–3564 (2019).

[43] Y. Fan, B. Zhang, W. Chang, X. Zhang, and H. Liao, "A novel integration of spectral-domain optical-coherence-tomography and laser-ablation system for precision treatment," Int. J. Comput. Assist. Radiol. Surg. **13**, 411–423 (2018).

[44] Y. Zhang, T. Pfeiffer, M. Weller, W. Wieser, R. Huber, J. Raczkowsky, J. Schipper, H. Wörn, and T. Klenzner, "Optical coherence tomography guided laser cochleostomy: Towards the accuracy on tens of micrometer scale," BioMed Research Int. **2014**, 1–10 (2014).

[45] J. D. Díaz, D. Kundrat, K.-F. Goh, O. Majdani, and T. Ortmaier, *Towards Intra-operative OCT Guidance for Automatic Head Surgery: First Experimental Results* (Lecture Notes in Computer Science, 2013), p. 347–354.

[46] A. Fuchs, M. Schultz, A. Krüger, D. Kundrat, J. D. Díaz, and T. Ortmaier, "Online measurement and evaluation of the er:yag laser ablation process using an integrated oct system," Biomed. Eng. / Biomedizinische Tech. **57**, 434–437 (2012).

[47] J. Jivraj, C. Chen, Y. Huang, J. Ramjist, Y. Lu, B. Vuong, X. Gu, and V. X. D. Yang, "Smart laser osteotomy: integrating a pulsed 1064nm fiber laser into the sample arm of a fiber optic 1310nm oct system for ablation monitoring," Biomed. Opt. Express **9**, 6374–6387 (2018).

[48] B. Y. Leung, P. J. Webster, J. M. Fraser, and V. X. Yang, "Real-time guidance of thermal and ultrashort pulsed laser ablation in hard tissue using inline coherent imaging," Lasers Surg. Med. **44**, 249–256 (2012).

[49] T.-A. Wang, M.-C. Chan, H.-C. Lee, C.-Y. Lee, and M.-T. Tsai, "Ultrahigh-resolution optical coherence tomography/angiography with an economic and compact supercontinuum laser," Biomed. Opt. Express **10**, 5687–5702 (2019).

[50] L. Bernstein, A. Ramier, J. Wu, V. D. Aiello, M. J. Béland, C. P. Lin, and S.-H. Yun, "Ultrahigh resolution spectral-domain optical coherence tomography using the 1000-1600 nm spectral band," Biomed. Opt. Express **13**, 1939–1947 (2022).

[51] J. A. Izatt and M. A. Choma, *Theory of Optical Coherence Tomography* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008), pp. 47–72.

[52] S. Aumann, S. Donner, J. Fischer, and F. Müller, *Optical Coherence Tomography (OCT): Principle and Technical Realization* (Springer International Publishing, Cham, 2019), pp. 59–85.

[53] J. F. de Boer, B. Cense, B. H. Park, M. C. Pierce, G. J. Tearney, and B. E. Bouma, "Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography," Opt. Lett. **28**, 2067–2069 (2003).

[54] M. A. Choma, M. V. Sarunic, C. Yang, and J. A. Izatt, "Sensitivity advantage of swept source and fourier domain optical coherence tomography," Opt. Express **11**, 2183–2189 (2003).

[55] R. Leitgeb, C. K. Hitzenberger, and A. F. Fercher, "Performance of fourier domain vs. time domain optical coherence tomography," Opt. Express **11**, 889–894 (2003).

[56] V. Jayaraman, J. Jiang, B. Potsaid, M. Robertson, P. J. S. Heim, C. Burgner, D. John, G. D. Cole, I. Grulkowski, J. G. Fujimoto, A. M. Davis, and A. E. Cable, *VCSEL Swept Light Sources* (Springer International Publishing, Cham, 2015), pp. 659–686.

[57] M. Minneman, J. Ensher, M. Crawford, M. Bonesi, B. Zabihian, P. Boschert, E. Hoover, D. Derickson, B. E. Applegate, T. Milner, and W. Drexler, *Akinetik Swept Sources* (Springer International Publishing, Cham, 2015), pp. 687–739.

[58] B. Hofer, B. Povazˇay, B. Hermann, S. M. Rey, V. Kajić, A. Tumlinson, K. Powell, G. Matz, and W. Drexler, "Artefact reduction for cell migration visualization using spectral domain optical coherence tomography," J. Biophotonics **4**, 355–367 (2011).

[59] S. Moon, S.-W. Lee, and Z. Chen, "Reference spectrum extraction and fixed-pattern noise removal in optical coherence tomography," Opt. Express **18**, 24395–24404 (2010).

[60] D. L. Marks, A. L. Oldenburg, J. J. Reynolds, and S. A. Boppart, "Autofocus algorithm for dispersion correction in optical coherence tomography," Appl. Opt. **42**, 3038–3046 (2003).

[61] M. Duelk and K. Hsu, *SLEDs and Swept Source Laser Technology for OCT* (Springer International Publishing, Cham, 2015), pp. 527–561.

[62] H.-C. Lee, J. J. Liu, Y. Sheikine, A. D. Aguirre, J. L. Connolly, and J. G. Fujimoto, "Ultrahigh speed spectral-domain optical coherence microscopy," Biomed. Opt. Express **4**, 1236–1254 (2013).

[63] K. Zhang and J. U. Kang, "Real-time 4d signal processing and visualization using graphics processing unit on a regular nonlinear-k fourier-domain oct system," Opt. Express **18**, 11772–11784 (2010).

[64] Y. Jian, K. Wong, and M. V. Sarunic, "Graphics processing unit accelerated optical coherence tomography processing at megahertz axial scan rate and high resolution video rate volumetric rendering," J. Biomed. Opt. **18**, 026002 (2013).

[65] W. Wieser, B. R. Biedermann, T. Klein, C. M. Eigenwillig, and R. Huber, "Multi-megahertz oct: High quality 3d imaging at 20 million a-scans and 4.5 gvoxels per second," Opt. Express **18**, 14685–14704 (2010).

[66] J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," J. Biomed. Opt. **4**, 95 – 105 (1999).

[67] M. Kobayashi, H. Hanafusa, K. Takada, and J. Noda, "Polarization-independent interferometric optical-time-domain reflectometer," J. Light. Technol. **9**, 623–628 (1991).

[68] B. Baumann, C. W. Merkle, R. A. Leitgeb, M. Augustin, A. Wartak, M. Pircher, and C. K. Hitzenberger, "Signal averaging improves signal-to-noise in oct images: But which approach works best, and when?" Biomed. Opt. Express **10**, 5755–5775 (2019).

[69] M. Pircher, E. Götzinger, R. A. Leitgeb, A. F. Fercher, and C. K. Hitzenberger, "Speckle reduction in optical coherence tomography by frequency compounding," J. Biomed. Opt. **8**, 565 – 569 (2003).

[70] S. Huang, C. Tang, M. Xu, Y. Qiu, and Z. Lei, "Bm3d-based total variation algorithm for speckle removal with structure-preserving in oct images," Appl. Opt. **58**, 6233–6243 (2019).

[71] S. Chitchian, M. A. Fiddy, and N. M. Fried, "Denoising during optical coherence tomography of the prostate nerves via wavelet shrinkage using dual-tree complex wavelet transform," J. Biomed. Opt. **14**, 014031 (2009).

[72] A. Ozcan, A. Bilenca, A. E. Desjardins, B. E. Bouma, and G. J. Tearney, "Speckle reduction in optical coherence tomography images using digital filtering," J. Opt. Soc. Am. A **24**, 1901–1910 (2007).

[73] B. Chong and Y.-K. Zhu, "Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified bm3d filter," Opt. Commun. **291**, 461–469 (2013).

[74] K. J. Halupka, B. J. Antony, M. H. Lee, K. A. Lucy, R. S. Rai, H. Ishikawa, G. Wollstein, J. S. Schuman, and R. Garnavi, "Retinal optical coherence tomography image enhancement via deep learning," Biomed. Opt. Express **9**, 6205–6221 (2018).

[75] Y. Huang, N. Zhang, and Q. Hao, "Real-time noise reduction based on ground truth free deep learning for optical coherence tomography," Biomed. Opt. Express **12**, 2027–2040 (2021).

[76] S. Chang and A. K. Bowden, "Review of methods and applications of attenuation coefficient measurements with optical coherence tomography," J. Biomed. Opt. **24**, 090901 (2019).

[77] T. van Leeuwen, D. Faber, and M. Aalders, "Measurement of the axial point spread function in scattering media using single-mode fiber-based optical coherence tomography," IEEE J. Sel. Top. Quantum Electron. **9**, 227–233 (2003).

[78] G. van Soest, T. P. M. Goderie, E. R. M.D., S. K. M.D., A. G. J. L. H. van Leenders M.D., N. G. M.D., S. van Noorden, T. O. M.D., B. E. Bouma, G. J. Tearney,

W. Oosterhuis, P. W. S. M.D., and A. F. W. van der Steen, "Atherosclerotic tissue characterization in vivo by optical coherence tomography attenuation imaging," J. Biomed. Opt. **15**, 011105 (2010).

[79] K. A. Vermeer, J. Mo, J. J. A. Weda, H. G. Lemij, and J. F. de Boer, "Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography," Biomed. Opt. Express **5**, 322–337 (2014).

[80] J. Liu, N. Ding, Y. Yu, X. Yuan, S. Luo, J. Luan, Y. Zhao, Y. Wang, and Z. Ma, "Optimized depth-resolved estimation to measure optical attenuation coefficients from optical coherence tomography and its application in cerebral damage determination," J. Biomed. Opt. **24**, 035002 (2019).

[81] M. Almasian, N. Bosschaart, T. G. van Leeuwen, and D. J. Faber, "Validation of quantitative attenuation and backscattering coefficient measurements by optical coherence tomography in the concentration-dependent and multiple scattering regime," J. Biomed. Opt. **20**, 121314 (2015).

[82] C. Christodoulou, C. Pattichis, M. Pantziaris, and A. Nicolaides, "Texture-based classification of atherosclerotic carotid plaques," IEEE Trans. on Med. Imaging **22**, 902–912 (2003).

[83] L.-K. Soh and C. Tsatsoulis, "Texture analysis of sar sea ice imagery using gray level co-occurrence matrices," IEEE Trans. on Geosci. Remote Sens. **37**, 780–795 (1999).

[84] C. Christodoulou, S. Michaelides, and C. Pattichis, "Multifeature texture analysis for the classification of clouds in satellite imagery," IEEE Trans. on Geosci. Remote Sens. **41**, 2662–2668 (2003).

[85] J. Z. Liu, L. D. Zhang, and G. H. Yue, "Fractal dimension in human cerebellum measured by magnetic resonance imaging," Biophys. J. **85**, 4041–4046 (2003).

[86] C. He, Z. Li, J. Wang, Y. Huang, Y. Yin, and Z. Li, "Atherosclerotic plaque tissue characterization: An oct-based machine learning algorithm with ex vivo validation," Front. Bioeng. Biotechnol. **8** (2020).

[87] G. J. Ughi, T. Adriaenssens, P. Sinnaeve, W. Desmet, and J. D'hooge, "Automated tissue characterization of in vivo atherosclerotic plaques by intravascular optical coherence tomography images," Biomed. Opt. Express **4**, 1014–1030 (2013).

[88] M. Ding, S.-y. Pan, J. Huang, C. Yuan, Q. Zhang, X.-l. Zhu, and Y. Cai, "Optical coherence tomography for identification of malignant pulmonary nodules based on random forest machine learning algorithm," PLOS ONE **16**, 1–15 (2022).

[89] J. Möller, A. Bartsch, M. Lenz, I. Tischoff, R. Krug, H. Welp, M. R. Hofmann, K. Schmieder, and D. Miller, "Applying machine learning to optical coherence

tomography images for automated tissue classification in brain metastases," Int. J. Comput. Assist. Radiol. Surg. **16**, 1517–1526 (2021).

[90] M. Sharifi, T. Khatibi, M. H. Emamian, S. Sadat, H. Hashemi, and A. Fotouhi, "Development of glaucoma predictive model and risk factors assessment based on supervised models," BioData Min. **14**, 48 (2021).

[91] L. Breiman, "Random forests," Mach. Learn. **45**, 5–32 (2001).

[92] M. Awad and R. Khanna, *Support Vector Machines for Classification* (Apress, Berkeley, CA, 2015), pp. 39–66.

[93] O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Fundamentals of Artificial Neural Networks and Deep Learning* (Springer International Publishing, Cham, 2022), pp. 379–425.

[94] S. Haykin, *Neural Networks and Learning Machines*, Pearson International Edition (Pearson, 2009).

[95] R. Eldan and O. Shamir, "The power of depth for feedforward neural networks," in *Annual Conference Computational Learning Theory,* (2015).

[96] M. Telgarsky, "benefits of depth in neural networks," in *29th Annual Conference on Learning Theory,* vol. 49 of *Proceedings of Machine Learning Research* V. Feldman, A. Rakhlin, and O. Shamir, eds. (Curran Associates, Inc., Columbia University, New York, New York, USA, 2016), pp. 1517–1539.

[97] J. Patterson and A. Gibson, *Deep Learning: A Practitioner's Approach* (O'Reilly Media, Inc., 2017), 1st ed.

[98] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics,* vol. 15 of *Proceedings of Machine Learning Research* G. Gordon, D. Dunson, and M. Dudík, eds. (PMLR, Fort Lauderdale, FL, USA, 2011), pp. 315–323.

[99] O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Fundamentals of Artificial Neural Networks and Deep Learning* (Springer International Publishing, Cham, 2022), pp. 379–425.

[100] H. Izadkhah, "Chapter 4 - basic structure of neural networks," in *Deep Learning in Bioinformatics,* H. Izadkhah, ed. (Academic Press, 2022), pp. 67–93.

[101] M. D. Zeiler, "Adadelta: An adaptive learning rate method," ArXiv **abs/1212.5701** (2012).

[102] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings,* Y. Bengio and Y. LeCun, eds. (2015).

[103] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature **323**, 533–536 (1986).

[104] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," The J. Physiol. **148**, 574–591 (1959).

[105] O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Convolutional Neural Networks* (Springer International Publishing, Cham, 2022), pp. 533–577.

[106] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE **86**, 2278–2324 (1998).

[107] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems,* vol. 25 F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds. (Curran Associates, Inc., 2012).

[108] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* (2016), pp. 770–778.

[109] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* (2017), pp. 2261–2269.

[110] D. Fantarella and L. Kotlow, "The 9.3-μm co2 dental laser: Technical development and early clinical experiences," Laser Dent **22**, 10–27 (2014).

[111] K.-U. Lewandrowski, C. Lorente, K. T. Schomacker, T. J. Fiotte, J. W. Wilkes, and T. F. Deutsch, "Use of the er:yag laser for improved plating in maxillofacial surgery: Comparison of bone healing in laser and drill osteotomies," Lasers Surg. Med. **19**, 40–45 (1996).

[112] L. M. B. Bernal, I. T. Schmidt, N. Vulin, J. Widmer, J. G. Snedeker, P. C. Cattin, A. Zam, and G. Rauter, "Optimizing controlled laser cutting of hard tissue (bone)," at - Autom. **66**, 1072–1082 (2018).

[113] S. Lu, C. Y.-l. Cheung, J. Liu, J. H. Lim, C. K.-s. Leung, and T. Y. Wong, "Automated layer segmentation of optical coherence tomography images," IEEE Trans. on Biomed. Eng. **57**, 2605–2608 (2010).

[114] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015,* N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds. (Springer International Publishing, Cham, 2015), pp. 234–241.

[115] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," IEEE Trans. Image Process. **16**, 2080–2095 (2007).

[116] I. W. Selesnick, "The double-density dual-tree dwt," IEEE Trans. Signal Process. **52**, 1304–1314 (2004).

[117] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Commun. ACM **63**, 139–144 (2020).

[118] D. Sheet, S. P. K. Karri, A. Katouzian, N. Navab, A. K. Ray, and J. Chatterjee, "Deep learning of tissue specific speckle representations in optical coherence tomography and deeper exploration for in situ histology," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI),* (2015), pp. 777–780.

[119] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," (Computational and Biological Learning Society, 2015), pp. 1–14.

[120] A. Vogel and V. Venugopalan, "Mechanisms of pulsed laser ablation of biological tissues," Chem. Rev. **103**, 577–644 (2003).

[121] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Eng. **82**, 35–45 (1960).

Personal Info

Address:
Innere Margarethenstasse
24A Basel 4051
Switzerland

Date of Birth: 11 May 1989

Phone:
(+41) 0766450647

Email:
yakub.rouge@gmail.com

Nationality:
Indonesia

Languages

English:
Professional working
Proficiency

Bahasa:
Native Language

German:
Elementary Proficiency

Strengths

- Teamwork
- Creativity
- Analytical Skill

Yakub Bayhaqi

Medical Image Processing Software Engineer

Ph.D. candidate -Biomedical Engineering

Objectives:

An experienced software engineer for biomedical applications with a focus on high-performance computing. I am seeking an opportunity to solve challenges in real-time automation systems for medical applications where my advanced skills, education, extensive training, and many years of experience can be fully utilized.

Summary:

- Experienced Software Developer with Biomedical Engineering and Computer Science background.
- Demonstrated work experience in medical imaging devices.
- Extensive experience in C++ GUI programming using Qt and CUDA for real-time image processing and feedback system.
- Work experience with a multidisciplinary team including laser physicists, robotic engineers, and physicians with the aim of minimally invasive surgery.

Professional Experience:

2018 – 2022

Researcher: Ph.D.   University of Basel, Switzerland

Relevant Research Projects:
- Real-Time Tissue characterization based on Optical Coherence Tomography images using Deep Learning for laser bone surgery.
- Real-time Acquisition and Visualization of Optical Coherence Tomography image in Virtual Reality.
- Optoacoustic Tissue Classification for Laser Osteotomes Using Mahala Nobis Distance-Based Method.
- Laser-Induced Breakdown Spectroscopy Combined with Artificial Neural Network for Pre-Carbonization Detection in Laser osteotomy.

2010 – 2018

Laboratory Assistant   University of Indonesia, Indonesia

Tasks:
- Research assistant in radiation therapy dosimetry, simulation, and radiograph image processing.
- Medical Radiation Safety Compliance Tester by The National Atomic Energy Agency of Indonesia.

2010 – 2011

Medical Physicist   Bina Centra Pertamina Hospital, Jakarta, Indonesia

Tasks:
- Medical Imaging Devices Quality Assurance and Calibration (Computed Tomography, MRI, and Ultrasound).

Education

2018 – Now: Ph.D. (Final Year) in Department of Biomedical Engineering, Faculty of Medicine, University of Basel.

- Thesis topic: "Real-Time Tissue characterization based on OCT images using machine learning for bone laser surgery." (Two published Journal)
- Planned Thesis defense, March 2023.

2012 – 2015: Master of Computer Science in high-performance computing at Faculty of Computer Science, University of Indonesia.

- Thesis title: "Investigation of Optimum Parameter for Photon Radiotherapy Monte Carlo Dose Calculation Method in GPU and Cluster MPI Computation Environment."

2007 – 2012: Bachelor of Science in Medical Physics, Department of Physics, University of Indonesia.

Awards

- Best Paper Contribution Silver Award at AUTOMED-Automation in Medical Engineering 2021
  Title:"Kalman filtered depth prediction using Optical Coherence Tomography for laser bone cutting".
- Best Paper and Oral Presentation at 3rd International Seminar on Photonics, Optics and its Applications 2018
  Title: "Neural Network in Tissue Characterization of Optical Coherence Tomography Images for Smart Laser Surgery: A Preliminary Study".

Technical Programming Language Skills

- Native Programming: **C++ with Object-Oriented Programming**
- GUI Programming: **Qt and MFC**
- Analytical Programming: **Python and MATLAB**
- Parallel Programming: **CUDA, OpenCL, MPI, and OpenMP**
- Image Processing, 3D Rendering, and Virtual Reality: **OpenCV, OpenGL, DirectX, and Unity**
- Data Science: **Tensorflow, Pytorch, and TensorRT**
- Embedded System: **Raspberry, Arduino, and Nvidia Jetson**

LIST OF PUBLICATIONS

Journal Papers

Published

- **Y. A. Bayhaqi**, A. Hamidi, A. Navarini, P. C. Cattin, Ferda Canbaz, and A. Zam, "Real-time closed-loop tissue-specific laser osteotomy using deep-learning-assisted optical coherence tomography," *Biomedical Optics Express* Vol. 14, No. 6, pp. 2986–2986 (2023).

- **Y. A. Bayhaqi**, A. Hamidi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Deep-Learning-Based Fast Optical Coherence Tomography (OCT) Image Denoising for Smart Laser Osteotomy", *IEEE Transactions on Medical Imaging*, Vol. 41, No. 10, pp. 2615-2628 (2022).

- **Y. A. Bayhaqi**, A. Hamidi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Deep learning models comparison for tissue classification using optical coherence tomography images: toward smart laser osteotomy", *OSA Continuum*, Vol. 4, No. 9, pp. 2510-2526 (2021).

- A. Hamidi, **Y. A. Bayhaqi**, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Long-range optical coherence tomography with extended depth-of-focus: a visual feedback system for smart laser osteotomy", *Biomedical Optics Express*, Vol. 12, No. 4, pp. 2118–2133 (2021).

Conference Proceedings

- **Y. A. Bayhaqi**, A. Hamidi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Kalman filtered depth prediction using Optical Coherence Tomography for laser bone cutting", in *Automation in Medical Engineering* (AUTOMED 2021), Zenodo (2021). [Best Application Paper Silver Award]

- A. Hamidi, **Y. A. Bayhaqi**, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam "Observation of controlled temperature changes of bone by phase-sensitive optical coherence tomography", *Proc. SPIE*, Vol. 11924, in *Optical Coherence Imaging Techniques and Imaging in Scattering Media IV* (2021).

137

- A. Hamidi, **Y. A. Bayhaqi**, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam "Imaging photothermal-induced expansion of bone during laser osteotomy by phase-sensitive OCT: preliminary results", *Proc. SPIE*, VOl. 11359, in *Biomedical Spectroscopy, Microscopy, and Imaging* (2020).

- F. Canbaz, H. Abbasi, **Y. A. Bayhaqi**, P. C. Cattin, A. Zam, "Laser-Induced Breakdown Spectroscopy Combined with Artificial Neural Network for Pre-carbonization Detection in Laserosteotomy", in *MESROB Conf. book, Springer* (2020). [Best Application Paper Bronze Award]

- H. N. Kenhagho, **Y. A. Bayhaqi**, Canbaz, F., R. Guzman, R., T. E. G. Alvarez-Arenas, P. C. Cattin, and A. Zam, "Optoacoustic Tissue Classification for Laser Osteotomes Using Mahalanobis Distance-Based Method", in *MESROB Conf. book, Springer* (2020).

- **Y. A. Bayhaqi**, G. Rauter, A. A. Navarini, P. C. Cattin, and A. Zam, "Fast optical coherence tomography image enhancement using deep learning for smart laser surgery: preliminary study in bone tissue", *Proc. SPIE*, Vol. 11207, in *Fourth International Conference on Applications of Optics and Photonics*, (2019).

- **Y. A. Bayhaqi**, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Neural network in tissue characterization of Optical Coherence Tomography (OCT) image for smart laser surgery: preliminary study," *Proc. SPIE*, Vol. 11044, in *Third International Seminar on Photonics, Optics, and Its Applications 2018* (2018).

Oral Presentations

- **Y. A. Bayhaqi**, A. Hamidi, F. Canbaz, A. A. Navarini, P. C. Cattin, and A. Zam, "Kalman filtered depth prediction using Optical Coherence Tomography for laser bone cutting", in *Automation in Medical Engineering* (AUTOMED 2021), Basel, Switzerland 8 June 2021.

- **Y. A. Bayhaqi**, A. Hamidi, A. A. Navarini, P. C. Cattin, and A. Zam, "Monte Carlo investigation of deep learning tissue classification performance in OCT-based smart laser bone surgery", in *Photonics West: BIOS*, San Francisco, United State, 9 March 2020.

- **Y. A. Bayhaqi**, G. Rauter, A. A. Navarini, P. C. Cattin, and A. Zam, "Fast optical coherence tomography image enhancement using deep learning for smart laser surgery: preliminary study in bone tissue", in *Fourth International Conference on Applications of Optics and Photonics* (AOP 2019), Lisbon, Portugal, 1 June 2019.

- **Y. A. Bayhaqi**, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Toward Tissue Differentiation from OCT Images using Deep Learning for Smart Laser Surgery", *5th DBE Summer School 2019*, Riederalp, Switzerland, 16 June 2019.

- **Y. A. Bayhaqi**, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Neural network in tissue characterization of Optical Coherence Tomography (OCT) image for smart laser surgery: preliminary study," in *Third International Seminar on Photonics, Optics, and Its Applications 2018*, Surabaya, Indonesia, 1 August 2018.

- **Y. A. Bayhaqi**, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Neural Network Based Tissue Characterization of Optical Coherence Tomography (OCT) Image for Smart Laser Surgery", in *4th DBE MIRACLE Summer School 2018*, Engelberg, Switzerland, 8 June 2018.

Poster Presentations

- **Y. A. Bayhaqi**, A. Hamidi, A. A. Navarini, F. Canbaz, P. C. Cattin, and A. Zam, "Real-Time Closed-Loop Controlled Laser Osteotomy using Optical Coherence Tomography", in *8th DBE Research Day*, Zentrum für Lehre und Forschung, Basel, Switzerland, 30 August 2022.

- **Y. A. Bayhaqi**, A. Hamidi, A. A. Navarini, F. Canbaz, P. C. Cattin, and A. Zam, "Toward Image Guided Bone Surgery with Optical Coherence Tomography (OCT)", in *7th DBE Research Day*, Zentrum für Lehre und Forschung, Basel, Switzerland, 7 September 2021.

- **Y. A. Bayhaqi**, A. Hamidi, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Automatic Tissue Characterization for Smart Laser Surgery Using Optical Coherence Tomography (OCT) Image", in *Biophotonics and Imaging Graduate Summer School* (BIGSS), Online, 29 August 2020.

- **Y. A. Bayhaqi**, A. Hamidi, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Toward Smart-Laser Osteotomy Guided with Optical Coherence Tomography Image", in *6th DBE Research Day*, Zentrum für Lehre und Forschung, Basel, Switzerland, 8 September 2020.

- **Y. A. Bayhaqi**, A. Hamidi, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Toward Tissue Differentiation from OCT Images using Deep Learning for Smart Laser Surgery", in *Yearly meeting of the Biomedical Photonics Network (BMPN), Progress in Biomedical Photonics*, Bern, Switzerland, 10 December 2019.

- **Y. A. Bayhaqi**, A. Hamidi, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Toward Tissue Differentiation from OCT Images using Deep Learning for Smart Laser Surgery", in *5th DBE Research Day*, Zentrum für Lehre und Forschung, Basel, Switzerland, 28 August 2019.

- **Y. A. Bayhaqi**, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, ""Automatic Tissue Characterization from Optical Coherence Tomography (OCT) Images for Smart Laser Surgery", in *Eucor-GRACE Crossborder Networking Event*, La Bresse, France, 29 April 2019.

- **Y. A. Bayhaqi**, A. Hamidi, A. A. Navarini, G. Rauter, P. C. Cattin, and A. Zam, "Automatic Tissue Characterization of Optical Coherence Tomography (OCT) Image for Smart Laser Surgery", in *4th DBE Research Day*, Zentrum für Lehre und Forschung, Basel, Switzerland, 5 September 2018.