IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 October 2022, accepted 9 November 2022, date of publication 30 November 2022, date of current version 7 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3225651

RESEARCH ARTICLE

Bone Ablation Depth Estimation From Er:YAG Laser-Generated Acoustic Waves

CARLO SEPPI[1,2], ANTAL HUCK[1,2], ARSHAM HAMIDI[1,3], EVA SCHNIDER[1,2], MASSIMILIANO FILIPOZZI[1,2], GEORG RAUTER[1,4], (Member, IEEE), AZHAR ZAM[1,3,5,6], AND PHILIPPE C. CATTIN[1,2], (Member, IEEE)

[1]Department of Biomedical Engineering, University of Basel, 4001 Basel, Switzerland
[2]Center for Medical Image Analysis and Navigation (CIAN), 4123 Allschwil, Switzerland
[3]Biomedical Laser and Optics Group (BLOG), 4123 Allschwil, Switzerland
[4]Bio-Inspired Robots for Medicine-Laboratory (BIROMED-Lab), 4123 Allschwil, Switzerland
[5]Division of Engineering, New York University Abu Dhabi (NYUAD), Abu Dhabi, United Arab Emirates
[6]Department of Biomedical Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11201, USA

Corresponding author: Carlo Seppi (carlo.seppi@unibas.ch)

This work was supported by the Werner Siemens Foundation through the Minimally Invasive Robot-Assisted Computer-Guided LaserosteotomE (MIRACLE) Project.

ABSTRACT Using a laser for cutting bones instead of the traditional saws improves a patient's healing process. Additionally, the laser has the potential to reduce the collateral damage to the surrounding tissue if appropriately applied. This can be achieved by building additional sensing elements besides the laser itself into an endoscope. To this end, we use a microsecond pulsed Erbium-doped Yttrium Aluminium Garnet (Er:YAG) laser to cut bones. During ablation, each pulse emits an acoustic shock wave that is captured by an air-coupled transducer. In our research, we use the data from these acoustic waves to predict the depth of the cut during the ablation process. We use a Neural Network (NN) to estimate the depth, where we use one or multiple consecutive measurements of acoustic waves. The NN outperforms the base-line method that assumes a constant ablation rate with each pulse to predict the depth. The results are evaluated and compared against the ground-truth depth measurements from Optical Coherence Tomography (OCT) images that measure the depth in real-time during the ablation process.

INDEX TERMS Acoustic feedback, depth control, laser ablation, neural network.

I. INTRODUCTION

Reducing the trauma of a patient during surgical procedures is paramount in improving the post-surgical standard of living. Consequently, minimally invasive alternatives to common interventions are a highly researched topic [1], [2]. One line of inquiry is the replacement of mechanical tools with laser-based ablation [3], [4], [5], [6], [7].

When the tissue is exposed to microsecond pulsed Er:YAG laser light, the water in the tissue heats up, vaporizes, and the expansion causes micro-explosions that ablate a small part of the tissue [8]. This process emits an acoustic wave that is captured by a transducer [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Ahsan Khandoker.

In contrast to classical mechanical cutting tools, laser ablation does not provide direct haptic feedback on the progress of the cut to the surgeon. Furthermore, the laser system occludes the cutting location and impedes visual inspection. With the development of new tools that assist the surgeon in monitoring the depth of the cut, damaging sensitive tissue can be avoided.

A classical method to measure the cutting depth would be using an OCT [10], [11], [12], [13]. However, it can be challenging to integrate an OCT in combination with an Er:YAG laser for minimally invasive surgery. To this end, we propose an acoustic depth measurement technique that uses the acoustic wave created during the ablation process to determine the depth of the cut.

When the location of the ablation source is known, the depth of the cut can be estimated via various approaches. One

approach is to triangulate the source [14], [15], [16]. However, this approach requires multiple transducers to determine the source location. A different approach uses acoustic waves in a 2D simulation to detect the source position using just two transducers [17]. The main limitation of this approach is that the exact acoustic wave generator (form and the frequency composition of the source) has to be known. In another 2D simulation, the source position and its form were reconstructed in an unknown surrounding [18]. However, many transducers are required for the reconstruction, making this approach unsuitable for minimally invasive surgery.

There are several ways to determine the distance between an acoustic source and the transducer. One option is to estimate the time-of-flight (ToF). For example, in [19], the authors used auto-correlation between the transmitted and echoed signal. In [20], they exploited the phase shift between transmitting and measured signals. However, in our application, the source signal shape is unknown, and the signal may even change for different depths. In [21] and [22], the authors showed that the ToF correlates with the signal's distance and decay when ablating tissue with an Er:YAG laser and in [23], they used ToF to estimate the depth of the ablation with a Neodymium-doped Yttrium Aluminium Garnet (Nd:YAG) laser. However, the ToF option has a significant drawback. It requires the distance between the transducer and the bone surface to be constant, and the medium velocity of the ablated tissue must be known. As this can not be guaranteed in our envisioned application of robotic-guided laser osteotomy, we focus on different approaches to estimate the depth of the laser cut.

Since the use of Neural Networks [24] is well established in medical imaging [25], [26], [27], [28], speech, and signal processing [29], [30], [31], [32], we aim to estimate the ablation depth by interpreting the signal from one single air-coupled transducer with a neural network. We compare two approaches: The CA is used as a base-line approach, where we assume that the ablation rate per pulse is constant, and therefore, the depth is proportional to the number of ablation shots. The second approach uses one or multiple consecutive measurements of the acoustic waves during the ablation process. These acoustic waves are then used as input for an NN to predict the depth of the laser cut.

The goal of this work is to analyze the acoustic wave and to prove that there is depth information in the acoustic wave produced during the ablation process of the bone using the Er:YAG laser. In addition, we can show that one transducer is sufficient to measure the depth of the laser cut, simplifying the complexity of the setup in future work.

II. MATERIAL AND SETUP

An Er:YAG laser (Syneron Candela, litetouch LI-FG0001A) with an energy of 153 mJ, a wavelength of 2940 nm, with a repetition rate of 10 Hz, is used for ablating the bone. The ablation process emits an acoustic wave. A CaF_2 mirror diverts a small percentage of the laser beam to a PbSe photodiode (PbSe Fixed Gain Detector, PDA20H, $1500 - 4800$ nm), to trigger the acoustic measurement. The wideband transducer,[1] with a frequency range of $100 - 1000$ kHz and a resonant frequency at 650 kHz, measures acoustic signals with a sample rate of 7.8125 MHz. It is placed at a distance of approximately 5 cm to the bone surface. The setup is displayed on the left of Fig. 1. We note that due to the limited acquisition rate of our setup, and the high repetition rate of the laser, we can only measure the acoustic wave of every second laser pulse.

A custom-made dichroic filter reflects the wavelength of the Er:YAG laser and transmits the OCT's wavelength, therefore, integrating the Er:YAG laser into the OCT system in a co-axial configuration. Consequently, we can monitor the ablation depth with a long-range Bessel-like beam OCT system [33] in real-time. The OCT has an imaging half-range of 22.21 mm in air and a field-of-view of 4.2 mm. The OCT system uses a swept-source laser (Insight Photonic Solution, Inc., Lafayette, Co, USA), with a scan rate of 104.17 kHz, a central wavelength of 1288.82 nm, and spectral bandwidth of 61.5 nm.

We used 13 cow femur bones as ablation material, bought in a local grocery store. The height of the bone varied between 2.4 cm to 2.9 cm. Muscle, fat, bone marrow, and tendons were carefully removed from the hard bone. To avoid dehydration of the bone, we submerged the samples in water between experiments.

We conducted our experiment by ablating a maximum of nine holes in each bone, each reaching a depth of up to 3.5 mm. The ablation process was stopped when we noticed that the cutting depth stagnated, i.e. if the bone started carbonizing. Bone carbonizes when insufficient water is in the tissue; hence, no micro-explosions remove the tissue, and the laser energy accumulates heating up the tissue. An exemplary bone is shown on the right of Fig. 1.

Since the frame rate of the OCT is 173.6167 Hz and the repetition rate of the laser is 10 Hz, we have a fixed number of OCT frames between the laser pulses. Therefore, we can align the OCT frames to the corresponding acoustic waves emitted during bone ablation. To generate a ground truth depth for each measured ablation wave, we locate the first pulse of the ablation in the OCT image stream. Then we labeled the 2443 OCT images by marking the bone's edge and the cut's end (see Fig. 2). We deduced the depth of the cut using the pixel resolution of the OCT of 10.86 μm.

III. METHOD

To assess the performance of our approaches, we divided our data into three mutually disjoint subgroups: training data, validation data, and testing data. Each bone was only part of one of these groups. Five-fold cross-validation was conducted to demonstrate the performance of the approach.

We investigated two approaches: (1) the first approach CA assumes that the ablation rate of the laser is constant, and

[1] PHYSICAL ACOUSTICS WSα SNAK28 6.9.2022, physicalacoustics.com/content/literature/sensors/Model_WSa.pdf

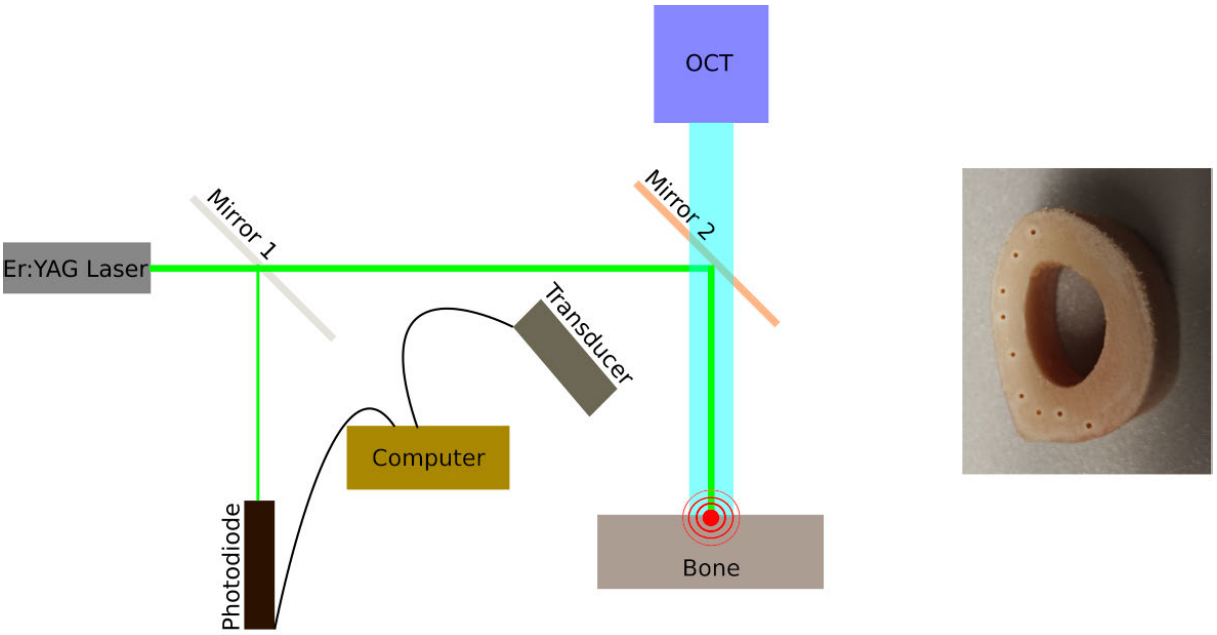


FIGURE 1. (Left) the setup of our experiment. Laser light ablates the bone while the air-coupled transducer measures the acoustic wave. The laser beam gets split at mirror 1, and a small percentage of the laser beam is redirected to the photodiode. The photodiode triggers the acquisition of the acoustic wave with the transducer (distance to the bone: 5 cm at a 45° angle to the ablation). Further, the laser light is diverted with mirror 2 (a custom-made dichroic filter that reflects the wavelength of the Er:YAG laser and transmits the OCT's wavelength) onto the bone for the ablation. At the same time, the OCT measures the depth of the ablation. (Right) an exemplary bone with nine holes after the ablation process.

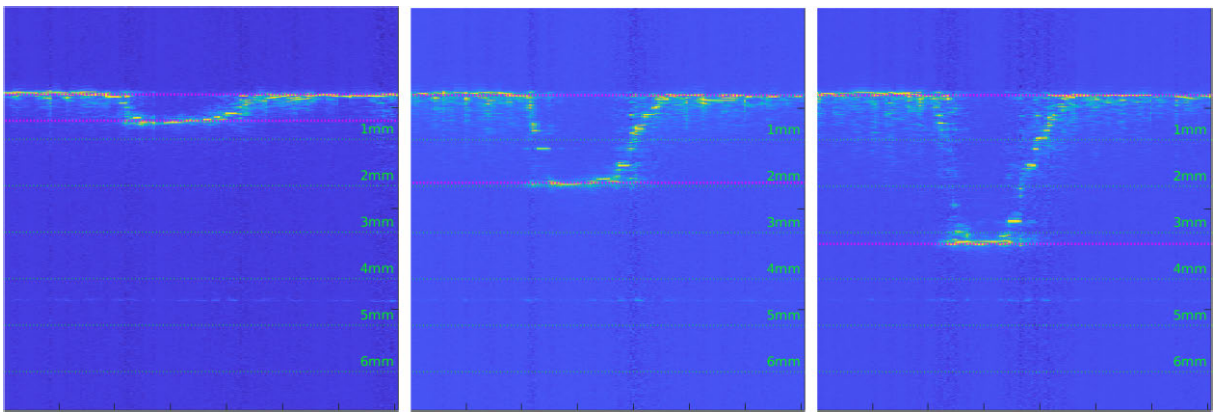


FIGURE 2. Exemplary OCT images, where we manually labeled the image after 10, 35, and 60 ablations. These corresponds to the depth of 0.5 mm, 1.9 mm, and 3.1 mm.

therefore when counting the number of shots, it can predict the depth of the laser cut. The ablation rate is estimated by the median value of the ablation rate of the training and validation data, and its performance is evaluated on the testing data. (2) In the second approach, we use a NN to predict the depth using the acoustic ablation waves.

A. NEURAL NETWORK (NN)

As the distance between the surface of the bone and the transducer can vary, all approaches using the ToF to estimate the cutting depth are of limited use. Furthermore, we wish to investigate if the depth information of a cut is embedded in other parts of the acoustic signal (excluding the ToF); hence, we removed this information by cropping the signal as follows: First, we identified the maximum absolute value of the first 500 sample points (the time window of 64 μs). Then we used 1.5 times this value as a noise threshold to remove the first part of the signal that only contained noise and no signal from the ablation (see Fig. 3, top left): We removed all sample points before the absolute value of the acoustic wave that first exceeded this threshold (see Fig. 3, top right, red dot) and solely used the window after that point (see Fig. 3, bottom left). In the final step, we normalized the area of interest of the acoustic wave w with:

$$w = \frac{w^* - \bar{w}}{\sigma_w} \tag{1}$$

where $\bar{w}$ is the mean value and σ_w is the standard deviation of the preprocessed acoustic wave w^* (see Fig. 3, bottom right).

To reduce over-optimization during training, we applied non-linear scaling of the wave amplitudes as data augmentation: Each sample point of the acoustic wave w was scaled by the formula

$$\hat{w} = w \cdot (1 + \lambda \cdot \exp[|w|]) \tag{2}$$

with λ a uniformly distributed random scaling factor.

As an input for the NN, we tried three different variants; namely, one, five, or ten consecutively measured acoustic signals as the input to the NN_1, NN_5, and NN_{10}, respectively. The Output of the NNs was a single value corresponding to the depth of the laser cut in [mm] from the latest measurement from the input. We trained the network using the *Mean Square Error Loss* (MSELoss):

$$\mathcal{L}_{MSE}(x, y) = \frac{1}{N_B} \sum_{i=1}^{N_B} (x_i - y_i)^2 \tag{3}$$

where $x = (x_1, \ldots, x_{N_B})$ is the output of the NN and $y = (y_1, \ldots, y_{N_B})$ is the label, that we labeled manually using the OCT images (see Section II and Fig. 2), and N_B is the batch size.

B. HYPERPARAMETER SEARCH

We use a hyperparameter search [34], [35] approach with five consecutive acoustic waves to find an initial network with the following search constraints, as visualized in Table 1. The input size, which correspond to the number of sample points, varies between 2000, 3000, ..., 7000. The numbers of convolutional and fully connected layers vary between 1 and 9. The parameters of the convolution layers can get the following values: Each output channel can allocate a value of 2^n, where n varies between $n = 1, \ldots, 8$,, the kernel size varies between 2 and 5, and the stride is 1 or 2. In addition, the maxpool kernel is 2 or 3 with a stride of 1 or 2. Batch normalization is applied randomly before any layer, and dropout is randomly applied to the fully connected layers with a dropout rate between 0 and 1. The number of neurons of the fully connected layer is $\min(2^n, out_{\text{CNN}})$, where n varies between $n = 3, \ldots, 11$ and out_{CNN} is the number of neurons after the flattened output of the last convolutional layer. We use the Adam optimizer with a learning rate of 10^{-n}, where n varies between $n = 2, \ldots, 5$. The maximum value λ of (2), varies between $0, 0.1, \ldots, 1$.

C. IMPLEMENTATION DETAILS

We implemented the network using the PyTorch[2] [36] framework and trained the networks on an NVIDIA Tesla V100 DGXS 16 GB. As a result of the hyperparameter search, a well-performing architecture is shown in Fig. 4, which uses 2000 sample points as the input size, namely a time window of 256 μs, for the NN. It has 5 convolutional layers followed by 8 fully connected layers with 1024 neurons. The Adam optimizer had a learning rate of 10^{-3}, a batch size of $N_B = 32$, and the data augmentation parameter λ in (2) was a random value between -0.5 and 0.5. In addition, batch normalization uses a momentum of 0.1 and eps=1e-05, while the maxpool layer had a padding of 0 with dilation of 1. The Dropout layer had the parameter inplace=False, and the Linear layers bias was set to true. The code is published on our GitLab page.[3]

TABLE 1. Overview of the parameter used in the hyperparameter search. Batch normalization is randomly applied before the convolutional and fully connected layers, and the dropout layers are randomly applied on the fully connected layers. out_{CNN} is the number of neurons after the flattened output of the last convolutional layer.

Input size (# of sample points):	$\{2000, 3000, \ldots, 7000\}$
Convolutional Layer	
Number of layers:	$\{1, \ldots, 9\}$
Output channels:	2^n, with $n \in \{1, \ldots, 8\}$
Kernel size:	$\{2, 3, 4, 5\}$
Stride:	$\{1, 2\}$
Maxpool layer	
Kernel size:	$\{2, 3\}$
Stride:	$\{1, 2\}$
Fully connected layer	
Number of layers:	$\{1, \ldots, 9\}$
Neurons:	$\min(2^n, out_{\text{CNN}})$, with $n \in \{3, \ldots, 11\}$
Dropout:	$p \in \mathcal{U}([0.1, 0.8])$
Optimizer:	Adam
learning rate:	10^{-n}, with $n \in \{2, 3, 4, 5\}$
Data augmentation:	$\max\|\lambda\| \in \{0, 0.1, \ldots, 1\}$

IV. RESULTS

We evaluated the accuracy of our models through five-fold cross-validation, which included splitting the data into five mutually disjoint subsets and omitting one subset during training for unseen forward passes during testing.

We assumed a constant ablation rate in the first approach CA. The median error over all the five-fold cross-validation sets was 0.13 mm and the distance $\mathcal{B}$ between the 25th percentile and the 75th percentile was 0.163 mm, as shown in Table 2. In Fig. 5, we visualize a box plot that shows the distribution of the deviating distance between the ground truth value and the output. As shown in Table 2, we see that the average box length was 0.271 mm with an average median value of -0.024 mm. At the interval between $\mathcal{I} = [3.25, 3.5]$mm the median value was -0.076 mm with a box length of 0.396 mm. We note that the number of shots was counted from the start of the ablation process to determine the depth of the cut.

In the second approach, we used one, five, or ten consecutively measured acoustic waves as the input for the NN. We trained the network on the training data and used the best-performing network on the validation data to test the performance of the testing data. In Table 2, we give a detailed description of the results. The median error was 0.174 mm, 0.130 mm, and 0.092 mm for NN_1, NN_5, and NN_{10}, respectively. In the box plots of Fig. 5, we visualize the difference

[2]May 2022, pytorch.org

[3]https://gitlab.com/cian.unibas.ch/ablaiton-depth-estimation

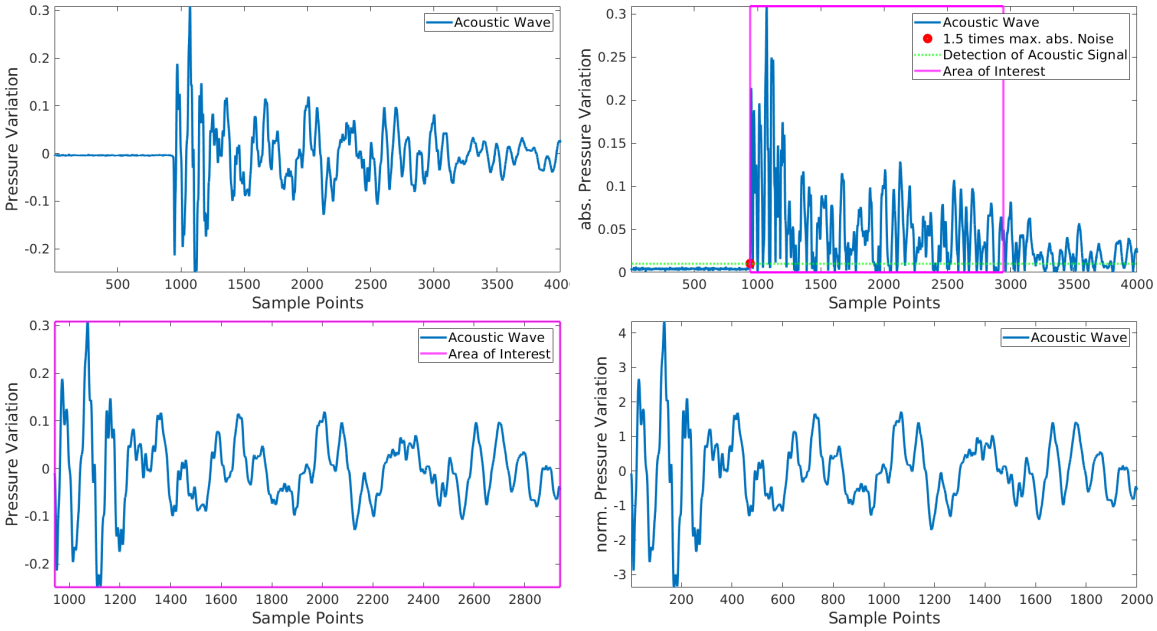


FIGURE 3. (Top left) measured acoustic wave with a sample rate of $7.8125\,\text{MHz}$. (Top right) multiply the maximum absolute value of the 500 first sample points (time window of $64\,\mu\text{s}$) by 1.5 (green). Remove all the data before the intersection between the green line and the acoustic signal from the ablation (redpoint). The 2000 sample points (time window of $256\,\mu\text{s}$) after the red dot are used for the further processing step (magenta). (Bottom left) the acoustic wave is visualized, where the time-of-flight is removed. (Bottom right) Normalized acoustic wave in the area of interest, which we use as input for the NN.

```
(0): Input(size=N×2000), BatchNorm1d( )
(1a): Conv1d(out_c=8, k_size=4, s=1)
(1b): MaxPool1d(k_size=3, s=2)
(2a): Conv1d(out_c=32, k_size=4, s=2)
(2b): MaxPool1d(k_size=3, s=2)
(3a): Conv1d(out_c=64, k_size=4, s=1)
(3b): MaxPool1d(k_size=2, s=1)
(4a): Conv1d(out_c=256, k_size=5, s=2)
(4b): BatchNorm1d( )
(4c): MaxPool1d(k_size=3, s=2)
(5a): Conv1d(out_c=256, k_size=4, s=2)
(5b): BatchNorm1d( )
(5c): MaxPool1d(k_size=2, s=1)
.....................................................
(6): Flattening(size=6912)
(7a): Dropout(p=0.48536), Linear(out_f=1028)
(7b): BatchNorm1d( )
(8a): Linear(out_f=1028)
(8b): BatchNorm1d( )
(9a): Linear(out_f=1028)
(9b): BatchNorm1d( )
(10): Dropout(p=0.54307), Linear(out_f=1028)
(11a): Dropout(p=0.43455), Linear(out_f=1028)
(11b): BatchNorm1d( )
(12a): Dropout(p=0.56238), Linear(out_f=1028)
(12b): BatchNorm1d( )
(13a): Dropout(p=0.19907), Linear(out_f=1028)
(13b): BatchNorm1d( )
(14): Dropout(p=0.20917), Linear(out_f=1028)
(15): Linear(out_f=1)
    ⇒ Output for MSELoss
```

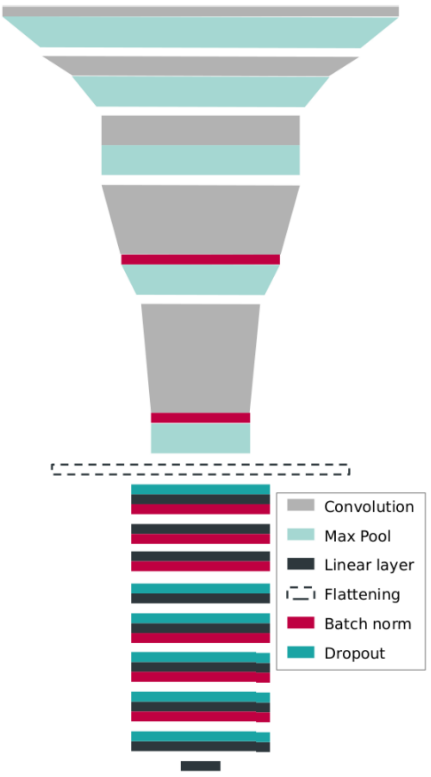


FIGURE 4. Detailed description of the neural network (left) and graphical visualization (right). It has 5 convolutional layers, followed by 8 fully connected layers. We chose a batch size of $N_B = 32$ and used the Adam optimizer. We note, that *k_size* represents the kernel size, *s* is the stride, *out_c* are the channel out, and *out_f* are the number output feature.

between the ground truth value and the label's output. Here, the mean length of the boxes $\mathcal{B}$ were 0.336 mm, 0.253 mm, and 0.200 mm for NN_1, NN_5, and NN_{10}, with a mean median value of -0.010 mm, -0.036 mm, and -0.027 mm, respectively. The largest errors were located at the interval $\mathcal{I}$ (in the last box plot of each subfigure) with a median value of

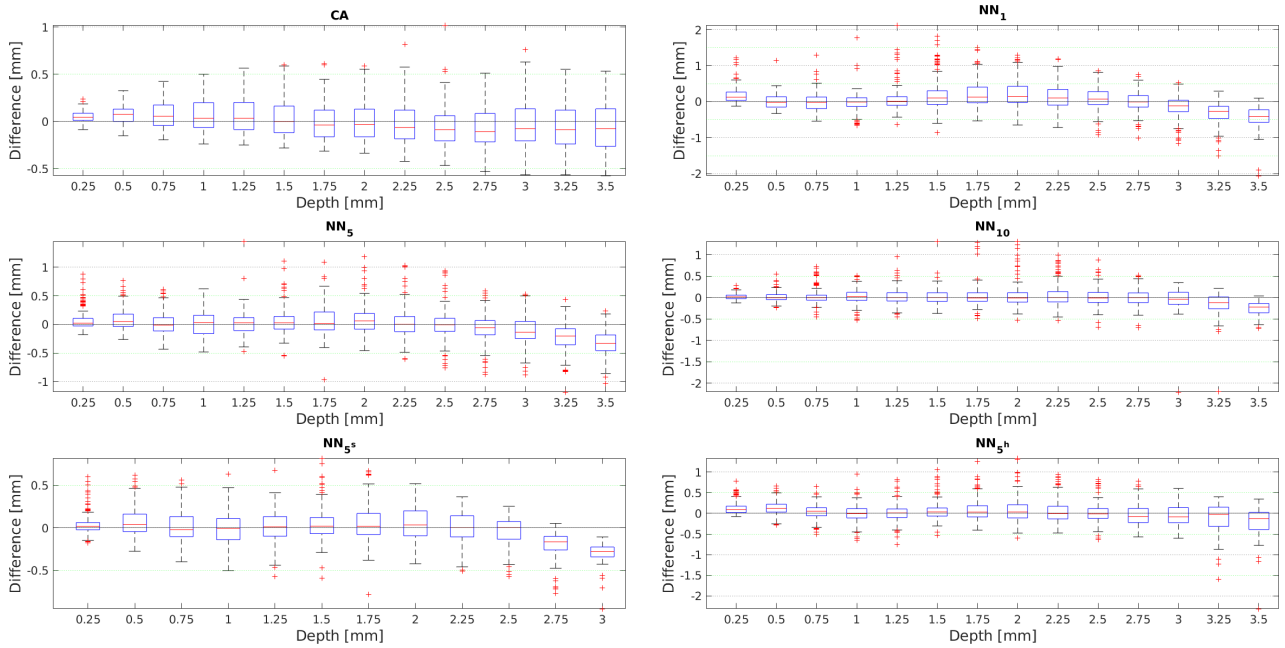


FIGURE 5. The box plots visualize the difference between the output and ground truth depth. CA is the method where we assume a constant ablation rate. NN_1, NN_5, and NN_{10} are the NN approaches with one, five, or ten consecutive acoustic measurements, respectively. NN_{5^s} and NN_{5^h} use five consecutive acoustic measurements as input. NN_{5^s} data do not exceed the depth of 2.8 mm, and NN_{5^h} uses some label augmentation. Each value on the x-axis corresponds to an interval, e.g., 0.25 mm represents the interval between 0 and 0.25 mm.

-0.407 mm, -0.332 mm, and -0.222 mm. The box length $\mathcal{B}$ was 0.349 mm, 0.275 mm, and 0.218 mm for NN_1, NN_5, and NN_{10}, respectively.

V. DISCUSSION

We compared two approaches: CA, which assumes a linear estimation of the depth and a NN with either one (NN_1), five (NN_5), or ten (NN_{10}) consecutively measured acoustic waves as an input. In the box plots in Fig. 5, we observed that CA has fewer outliers than the NNs. The median value is close to 0 for all intervals, and even within the interval $\mathcal{I}$ ($Distance_I$), the median value was close to 0. The NNs had more outliers, and the median value between the distance of the output and the ground truth depth shifts into the negative with increasing depth. Hence, it underestimated the depth of the cut, especially in the interval $\mathcal{I}$, as can be seen in Table 2 and Fig. 5. However, the box length $\mathcal{B}$ was smaller for the NN than the CA for the interval $\mathcal{I}$. We further observed that CA outperforms the NN_1 and had a similar performance to NN_5. The best performing network was NN_{10}.

To further investigate the underestimation of the depth, we retrained the Network NN_5 with data that reaches a maximum depth of 2.8 mm (NN_{5^s}). We observe in Fig. 5 that both NN_5 and NN_{5^s} had a significantly larger error in the reported interval $\mathcal{I}$ and [2.75, 3]mm, respectively. This is also reflected in Table 2, where the median value at $Distance_I$ were -0.332 mm and -0.278 mm, with a box length of 0.275 mm and 0.117 mm, respectively. Therefore, we assume that underestimating the depth at the end is due to the lack of training data.

TABLE 2. In the top row, the error, in the centre row, the mean distance over all intervals between the estimation and the label from the box-plots in Fig. 5, and in the bottom row, the distance of the last interval from the box-plots $Distance_I$ are described. We show the median, 25th percentile (PCTL), 75th PCTL value, and the distance $\mathcal{B}$, which is the difference between the 25th and 75th PCTL, of all the testing data from the cross-validation. CA is the method where we assume a constant ablation rate. NN_1, NN_5, and NN_{10} are the approach with the NN that uses one, five, or ten consecutive acoustic measurements. NN_{5^s} and NN_{5^h} use 5 consecutive acoustic signals as input. NN_{5^s} data do not exceed 2.8 mm depth, and therefore $Distance_I$ represents the interval 2.75 mm − 3 mm, while all the others represent the interval of $\mathcal{I} = [3.25, 3.5]$mm. NN_{5^h} was trained with some label augmentation.

mm	Median Value	25th PCTL	75th PCTL	distance $\mathcal{B}$
Error				
CA	0.130	0.065	0.228	0.163
NN_1	0.174	0.081	0.326	0.245
NN_5	0.130	0.060	0.252	0.192
NN_{10}	**0.092**	**0.040**	**0.189**	**0.149**
NN_{5^s}	0.110	0.047	0.213	0.166
NN_{5^h}	0.135	0.065	0.251	0.187
Distance				
CA	−0.024	−0.139	0.132	0.271
NN_1	−0.010	−0.1641	0.172	0.336
NN_5	−0.036	−0.158	0.096	0.253
NN_{10}	−0.027	**−0.120**	**0.080**	**0.200**
NN_{5^s}	−0.026	−0.123	0.083	0.206
NN_{5^h}	**0.003**	−0.136	0.143	0.278
$Distance_I$				
CA	**−0.076**	**−0.263**	0.133	0.396
NN_1	−0.407	−0.571	−0.222	0.349
NN_5	−0.332	−0.460	−0.185	0.275
NN_{10}	−0.222	−0.356	−0.138	0.218
NN_{5^s}	−0.278	−0.340	−0.224	**0.117**
NN_{5^h}	−0.127	−0.392	**0.023**	0.415

To this end, we retrained the network with label augmentation. Specifically, we added a random value r to the depth during the training of the network for all depths exceeding

3 mm. The value r is the absolute random number from the normal distribution $\mathcal{N}(0, 0.5)$, with the mean value of 0 mm and the standard deviation of 0.5 mm. We augmented the label as described above, leading to overestimating the depth during the network training. In Fig. 5 we can see that this strategy counteracts the general underestimation of the depth of the network at the last intervals. This is also reflected in Table 2, where the median value of $Distance_I$ was reduced from -0.332 mm to -0.127 mm. However, the box length $\mathcal{B}$ was increased to the value 0.275 mm to 0.415 mm. Therefore, augmenting the data improves the median accuracy but produces a higher output fluctuation.

VI. CONCLUSION

The experiments show that assuming a constant ablation rate already leads to good depth estimations of the cut by only counting the number of shots. This assumption, however, is limited to shallow cuts and does not hold for deep bone ablations that need a cooling system [37]. These cuts can reach a depth of up to 3 cm. Moreover, the number of shots must be maintained for a valid estimation. This is not stateless, meaning it needs all the information since the beginning of the ablation to estimate the depth. Therefore, it is not fail-safe since it may cause loss of depth information in case the number of the previous ablation gets lost.

It is essential for medical devices to continue working, even when a power failure occurs and all prior information is lost. Hence, we opt for a stateless method to ensure a fail-safe device. In this regard, the proposed approach with the NN is stateless (almost no prior information of previous ablations is needed) and, therefore, advantageous as it uses one or multiple consecutive acoustic waves as an input to predict the depth of the cut and does not need all information from the beginning of the ablation. Therefore, it is fail-safe and can predict the depth after only a few laser pulses. The NN approach has comparable accuracy but slightly more outliers due to its statelessness and sensitivity. The performance improved with an increasing number of consecutive acoustic waves used as input. Too many consecutive acoustic waves are disadvantageous because multiple acoustic waves are needed to estimate the depth accurately, increasing the risk of cutting the hole too deep and damaging sensitive tissue behind the bone.

In this work, we demonstrate the possibility of predicting the depth of a laser-ablated hole by analyzing the acoustic shock waves captured by a single transducer. The results encourage further investigations into the depth estimation during the laser ablation of tissue using acoustic waves. Our experiments were performed in a dry environment, and for future work, we plan depth estimation during a similar setup in wet conditions. Irrigation during laser ablation allows deeper cuts yet presents further challenges in combination with OCT systems. One of the challenges facing the irrigation system is that water accumulates in the hole, distorting depth measurement. In addition, the debris and water droplets pollute the OCT's protective window, reducing the image's contrast. An important factor may be the heterogeneity and the age of the bone influence the ablation process and the prediction of the depth using acoustic waves, which needs to be investigated.

ACKNOWLEDGMENT

We acknowledge the funding by the Werner Siemens Foundation through the MIRACLE (Minimally Invasive Robot-Assisted Computer-guided LaserosteotomE) project. We are grateful to Peter Zimmermann (Institute of Anatomy, University Basel) for helping prepare the bone specimens.

REFERENCES

[1] J. C. Hu, X. Gu, S. R. Lipsitz, M. J. Barry, A. V. D'Amico, A. C. Weinberg, and N. L. Keating, "Comparative effectiveness of minimally invasive vs open radical prostatectomy," *JAMA*, vol. 302, no. 14, pp. 1557–1564, 2009.

[2] J. D. Luketich, M. Alvelo-Rivera, P. O. Buenaventura, N. A. Christie, J. S. McCaughan, V. R. Litle, P. R. Schauer, J. M. Close, and H. C. Fernando, "Minimally invasive esophagectomy: Outcomes in 222 patients," *Ann. Surg.*, vol. 238, no. 4, p. 486, 2003.

[3] D. D. Lo, M. A. Mackanos, M. T. Chung, J. S. Hyun, D. T. Montoro, M. Grova, C. Liu, J. Wang, D. Palanker, A. J. Connolly, M. T. Longaker, C. H. Contag, and D. C. Wan, "Femtosecond plasma mediated laser ablation has advantages over mechanical osteotomy of cranial bone," *Lasers Surg. Med.*, vol. 44, no. 10, pp. 805–814, Dec. 2012.

[4] S. R. Visuri, J. T. Walsh, and H. A. Wigdor, "Erbium laser ablation of dental hard tissue: Effect of water cooling," *Lasers Surg. Med.*, vol. 18, no. 3, pp. 294–300, 1996.

[5] 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.*, vol. 47, no. 5, pp. 426–432, Jul. 2015.

[6] K.-W. Baek, M. Dard, H.-F. 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.*, vol. 53, no. 3, pp. 291–299, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/lsm.23281, doi: 10.1002/lsm.23281.

[7] M. Augello, W. Deibel, K. Nuss, P. Cattin, and P. Jürgens, "Comparative microstructural analysis of bone osteotomies after cutting by computer-assisted robot-guided laser osteotome and piezoelectric osteotome: An in vivo animal study," *Lasers Med. Sci.*, vol. 33, no. 7, pp. 1471–1478, Sep. 2018.

[8] H. Kang, I. Rizoiu, and A. Welch, "Hard tissue ablation with a spray-assisted mid-ir laser," *Phys. Med. Biol.*, vol. 52, no. 24, p. 7243, 2007.

[9] H. K. N. Kenhagho, G. Rauter, R. Guzman, P. C. Cattin, and A. A. Zam, "Comparison of acoustic shock waves generated by micro and nanosecond lasers for a smart laser surgery system," *Proc. SPIE*, vol. 10484, Feb. 2018, Art. no. 104840P.

[10] A. Fuchs, M. Schultz, A. Krüger, D. Kundrat, J. Díaz Díaz, and T. Ortmaier, "Online measurement and evaluation of the Er: YAG laser ablation process using an integrated OCT system," *Biomed. Eng. Biomedizinische Technik*, vol. 57, pp. 1–4, Jan. 2012.

[11] 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," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds. Berlin, Germany: Springer, 2013, pp.347–354.

[12] 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.*, vol. 2014, pp. 1–10, Sep. 2014.

[13] 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," Jun. 2021, doi: 10.5281/zenodo.4922992.

[14] B.-W. Jang and C.-G. Kim, "Acoustic emission source localization in composite stiffened plate using triangulation method with signal magnitudes and arrival times," *Adv. Compos. Mater.*, vol. 30, no. 2, pp. 149–163, Mar. 2021.

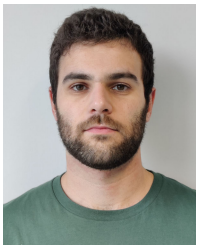
[15] B.-W. Jang, Y.-G. Lee, J.-H. Kim, Y.-Y. Kim, and C.-G. Kim, "Real-time impact identification algorithm for composite structures using fiber Bragg grating sensors," *Struct. Control Health Monitor.*, vol. 19, no. 7, pp. 580–591, Nov. 2012.
[16] B.-W. Jang and C.-G. Kim, "Impact localization of composite stiffened panel with triangulation method using normalized magnitudes of fiber optic sensor signals," *Compos. Struct.*, vol. 211, pp. 522–529, Mar. 2019.
[17] 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*, T. Vrtovec, J. Yao, G. Zheng, and J. M. Pozo, Eds. Cham, Switzerland: Springer, 2019, pp. 126–135.
[18] U. Nahum, C. Seppi, and P. C. Cattin, "Joint inverse medium and optimal control problem for acoustic waves," in *Proc. Platform Adv. Sci. Comput. Conf.*, Jun. 2019, p. 8.
[19] D. Marioli, C. Narduzzi, C. Offelli, D. Petri, E. Sardini, and A. Taroni, "Digital time-of-flight measurement for ultrasonic sensors," *IEEE Trans. Instrum. Meas.*, vol. 41, no. 1, pp. 93–97, Feb. 1992.
[20] X. Chen, J. Xu, H. Chen, H. Ding, and J. Xie, "High-accuracy ultrasonic rangefinders via pMUTs arrays using multi-frequency continuous waves," *J. Microelectromech. Syst.*, vol. 28, no. 4, pp. 634–642, Aug. 2019.
[21] H. N. Kenhagho, F. Canbaz, R. Guzman, P. Cattin, and A. Zam, "Miniaturized optoacoustic feedback sensor for smart laser osteotome: Fiber-coupled Fabry–Pérot etalon sensor," *Sens. Actuators A, Phys.*, vol. 317, Jan. 2021, Art. no. 112394.
[22] H. N. Kenhagho, I. Sugiarto, R. Guzman, P. Cattin, and A. Zam, "Contact-free Crater depth monitoring using measured acoustic shock waves for smart laser surgery applications: Preliminary result," in *Proc. Int. Conf. Radar, Antenna, Microw., Electron., Telecommun. (ICRAMET)*, Oct. 2019, pp. 118–121.
[23] F. J. Landa, X. L. Deán-Ben, F. M. de Espinosa, and D. Razansky, "Non-contact monitoring of incision depth in laser surgery with air-coupled ultrasound transducers," *Opt. Lett.*, vol. 41, no. 12, pp. 2704–2707, Jun. 2016.
[24] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[25] S. N. Deepa, "A survey on artificial intelligence approaches for medical image classification," *Indian J. Sci. Technol.*, vol. 4, no. 11, pp. 1583–1595, Nov. 2011.
[26] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *Proc. 13th Int. Conf. Control Autom. Robot. Vis. (ICARCV)*, Dec. 2014, pp. 844–848.
[27] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 31–40, Jan. 2016.
[28] E. Schnider, A. Horváth, G. Rauter, A. Zam, M. Müller-Gerbl, and P. C. Cattin, "3D segmentation networks for excessive numbers of classes: Distinct bone segmentation in upper bodies," in *Machine Learning in Medical Imaging*, M. Liu, P. Yan, C. Lian, and X. Cao, Eds. Cham, Switzerland: Springer, 2020, pp. 40–49.
[29] A. S. Miller, B. H. Blott, and T. K. Hames, "Review of neural network applications in medical imaging and signal processing," *Med. Biol. Eng. Comput.*, vol. 30, no. 5, pp. 449–464, Sep. 1992.
[30] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Aug. 2015.
[31] Y. Han and K. Lee, "Convolutional neural network with multiple-width frequency-delta data augmentation for acoustic scene classification," in *Proc. IEEE AASP Challenge Detection Classification Acoustic Scenes Events*, Jun. 2016, pp. 1–4. [Online]. Available: https://dcase.community/documents/challenge2016/technical_reports/DCASE2016_Lee_1034.pdf
[32] H. Kamper, W. Wang, and K. Livescu, "Deep convolutional acoustic word embeddings using word-pair side information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 4950–4954.
[33] A. Hamidi, Y. A. Bayhaqi, F. Canbaz, A. A. Navarini, P. Cattin, and A. Zam, "Long-range optical coherence tomography with extended depth-of-focus: A visual feedback system for smart laser osteotomy," *Biomed. Opt. Exp.*, vol. 12, no. 4, pp. 2118–2133, 2021.
[34] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6765–6816, 2017.
[35] J. Bergstra, D. Yamins, and D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 115–123.
[36] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2019, pp. 8024–8035.
[37] L. M. B. Bernal, F. Canbaz, A. Droneau, N. F. Friederich, P. C. 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. Exp.*, vol. 11, no. 12, pp. 7253–7272, 2020.

CARLO SEPPI received the degree in mathematics from the Mathematical Institute, University of Basel, Basel, Switzerland, in 2009, the B.Sc. and M.Sc. degrees in mathematics, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Planning and Navigation Group, Department of Biomedical Engineering, University of Basel. He worked as a Teaching Assistant with the Mathematical Institute during his studies. He started his Swiss Civil Service at the Department of Biomedical Engineering, University of Basel, in 2016. He helped to develop novel visualization methods for nuclear medicine to detect sentinel lymph nodes. His project involves solving inverse problems in medical applications: e.g., finding the position of sentinel lymph nodes using a pinhole collimator (reconstructing a 99mTc source using a single image of the detector), reconstruction of the medium velocity and the acoustic source with the help of the Helmholtz equation, and extracting information of the acoustic wave during the ablation of tissue.

ANTAL HUCK was born in Basel, Switzerland. He received the B.Sc. and M.Sc. degrees in mathematics from the University of Basel, Switzerland, with a focus on numerical analysis, and the Ph.D. degree from the Faculty of Medicine, University of Basel, in 2019, working on the research project "Segmentation of the Gray and White Matter in the Human Spinal Cord." Since then, he has been a Postdoctoral Researcher with the Group of Prof. Philippe Cattin, Department of Biomedical, University of Basel.

ARSHAM HAMIDI received the B.Sc. degree in physics from the University of Arak, Iran, in 2014, and the M.Sc. degree in photonics from the Institute of Advanced Studies in Basic Science (IASBS), Iran, in 2017. He is currently pursuing the Ph.D. degree with the Biomedical Laser and Optics Group (BLOG), Department of Biomedical Engineering, University of Basel, Switzerland. His main research interests include optical coherence tomography (OCT), biomedical imaging, and smart laser osteotomy.

EVA SCHNIDER received the M.Sc. degree in mathematics from the University of Basel, Switzerland, in 2018, with a thesis on spectral decomposition as a regularization method for inverse problems involving the acoustic wave equation in breast tumor detection. She is currently pursuing the Ph.D. degree with the MIRACLE Planning and Navigation Group of Prof. Philippe Cattin, Department of Biomedical Engineering, University of Basel. Her thesis focuses on the development of machine learning algorithms for distinct bone segmentation in CT images. The results of which will then be used in the group's virtual reality surgical planning systems. She has Interned at Novartis in Basel, and at Google Research, Zürich and London, working on clinical trial data embeddings, machine perception, and on fairness aspects of machine learning for time series predictions. Her research interests include medical image segmentation, neural network architectures for 3-D image modalities, and fairness for machine learning.



MASSIMILIANO FILIPOZZI received the B.Sc. degree in aerospace engineering and the M.Sc. degree in mechatronic engineering from the Politecnico di Torino, Italy, in 2016 and 2019, respectively. In 2018, he spent six months of his studies at the Universidad Nacional de Cordoba, Argentina. He started to work for his master's thesis at the end of 2018 at the Planning and Navigation Group of the MIRACLE project, Department of Biomedical Engineering, University of Basel. His activity focused on an opto-mechanical position sensor (ASTRAS) for medical tracking systems. Since August 2019, he has started his Ph.D. in the same department continuing the investigation and advancement in the sensor tracking system technology for minimally invasive surgery.

GEORG RAUTER (Member, IEEE) received the degree in mechanical engineering from TU-Graz, the degree in mathematical and mechanical modeling from MATMECA, Bordeaux, and the Ph.D. degree in robotics from ETH Zurich. From 2014 to 2016, he was a Postdoc in rehabilitation robotics at ETH Zurich, University of Southern California, and the University of Zurich. In 2016, he commercialized the gait rehabilitation robot the FLOAT in collaboration with the company Lutz Medical Engineering and the Spinal Cord Injury Center at the Balgrist, Zurich. Since May 2016, he has been heading the BIROMED-Laboratory, as an Assistant Professor of medical robotics and mechatronics at the Department of Biomedical Engineering, University of Basel. Since March 2022, he has been continuing to head the BIROMED-Laboratory, as an Associate Professor of surgical robotics. He has been the Chair of IFToMM Switzerland, since 2019, and an Associate Member of the National Centre of Competence in Research (NCCR) Robotics, since 2020.

AZHAR ZAM received the B.Sc. degree in medical physics from the University of Indonesia, Depok City, Indonesia, in 2004, the M.Sc. degree in biomedical engineering from the University of Lübeck, Lübeck, Germany, in 2007, and the Ph.D. degree in advanced optical technologies from Friedrich-Alexander University Erlangen–Nuremberg, Germany, with the focus on optical feedback for tissue-specific laser surgery, in 2011. He held a research positions at the University of Waterloo, Waterloo, ON, Canada; the University of Galway, Galway, Ireland; the Toronto Metropolitan University, Toronto, ON, Canada; and the University of California at Davis, Davis, CA, USA, before he joined the Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland, in 2016, as an Assistant Professor, where he founded and leads the Biomedical Laser and Optics Group (BLOG). Since 2022, he has been an Associate Professor of bioengineering at the Division of Engineering, New York University Abu Dhabi (NYUAD). He is also an Associated Faculty with the Department of Biomedical Engineering, Tandon School of Engineering, New York University, Brooklyn, NY, USA. He has authored over 90 peer-reviewed articles, book chapters, books, and patents. His main research interests include the development of smart devices for medical therapy, diagnostics, and monitoring using novel optical technologies, which include smart laser surgery, optical coherence tomography (OCT), photoacoustics, biomedical spectroscopy, AI-aided optical diagnostics and imaging, optical-based smart wearable biosensors, and miniaturized systems.

PHILIPPE C. CATTIN (Member, IEEE) was born in Switzerland, in 1967. He received the B.Sc. degree from the University of Applied Science, Brugg/Windisch, in 1991, the M.Sc. degree in computer science and the Ph.D. degree in robotics from ETH Zurich, Switzerland, in 1995 and 2003, respectively. From 2003 to 2007, he was a Postdoctoral Fellow with the Computer Vision Laboratory, ETH Zurich. In 2007, he became an Assistant Professor at the University of Basel and promoted to an Associate Professor, in 2015 and to a Full Professor, in 2019. He is the Founder of the Center for Medical Image Analysis and Navigation (CIAN), Medical Faculty of the University of Basel. He is also the Founding Head and the Head of the Department of Biomedical Engineering, University of Basel. He was a Research Fellow at the Brigham and Women's Hospital, Boston, MA, USA, in 2017. His research interests include medical image analysis, image-guided therapy, robotics-guided laser osteotomy, and virtual reality. As a Principal Investigator, he has finished many projects in these areas and published over 250 papers, patents, and book chapters. He is also the founder of three spin-off companies and licensed his patents and software to medical device companies.