Investigating Recurrent Neural Networks for OCT A-scan Based Tissue Analysis

C. Otte; S. Otte; L. Wittig; G. Hüttmann; C. Kugler; D. Drömann; A. Zell; A. Schlaefer

1 Institute of Medical Technology, Hamburg University of Technology, Hamburg, Germany; 2 Cognitive Systems Group, Computer Science Department, University of Tübingen, Tübingen, Germany; 3 Medical Clinic III, University Hospital Schleswig Holstein, Luebeck, Germany; 4 Institute of Biomedical Optics, University of Luebeck, Luebeck, Germany; 5 Medical Clinic III, Department of Thoracic Surgery, LungenClinic Grosshansdorf, Grosshansdorf, Germany

Keywords
Optical Coherence Tomography, biopsy guidance, optical soft tissue classification, identification of pulmonary nodules, Long Short Term Memory, recurrent neural nets

Summary
Objectives: Optical Coherence Tomography (OCT) has been proposed as a high resolution image modality to guide transbronchial biopsies. In this study we address the question, whether individual A-scans obtained in needle direction can contribute to the identification of pulmonary nodules. Methods: OCT A-scans from freshly resected human lung tissue specimen were recorded through a customized needle with an embedded optical fiber. Bidirectional Long Short Term Memory networks (BLSTMs) were trained on randomly distributed training and test sets of the acquired A-scans. Patient specific training and different pre-processing steps were evaluated. Results: Classification rates from 67.5% up to 76% were archived for different training scenarios. Sensitivity and specificity were highest for a patient specific training with 0.87 and 0.85. Low pass filtering decreased the accuracy from 73.2% on a reference distribution to 62.2% for higher cutoff frequencies and to 56% for lower cutoff frequencies. Conclusion: The results indicate that a grey value based classification is feasible and may provide additional information for diagnosis and navigation. Furthermore, the experiments show patient specific signal properties and indicate that the lower and upper parts of the frequency spectrum contribute to the classification.

1. Introduction
Lung cancer has the highest mortality rate of all cancers in male patients [1]. The survival rate highly depends on an early diagnosis. Gold-standard is the histological examination of tissue samples obtained by transthoracic biopsies (TTNA) and transbronchial biopsy (TBNAs). While the TTNA has a pneumothorax rate of 25% and a diagnostic yield of up to 90%, the TBNAs has a pneumothorax rate of only 1.5%, but also only a 70% chance to confirm a diagnosis [2]. Usually the guidance of TBNA is assisted by endobronchial ultrasound (EBUS) which improved diagnostic yield for larger solitary pulmonary nodules (SPNs). For SPNs < 2 cm the yield is still low [3]. A method to confirm the placement of the needle before taking the biopsy would be desirable. Optical Coherence Tomography (OCT) has been proposed as a high resolution imaging to guide transbronchial biopsies [4]. Previous ex vivo studies closely matched OCT image data to histologic images, demonstrating details at the level of mucosal layers, glands, alveoli, and respiratory bronchioles [5]. Recently a variety of flexible OCT catheter and needle probes have been demonstrated and applied to standard clinical bronchoscopy [6]. However the acquisition and interpretation of high resolution 2D and 3D images in vivo is still a complicated and manual process. This motivates the key question, whether fast and automated processing of individual A-scans in front of the needle is possible and can contribute to the identification of pulmonary nodules, i.e., to classify the tissue type. On the one hand the A-scan rate of commercial OCT devices is in the range of 100 kHz and therefore less affected by movement artifacts. On the other hand typical needle tissue interactions cause a compression of the tissue in front of the needle [7]. Due to the compression structural properties as well as quantitative measures, such as the optical attenuation coefficient may change [8, 9]. This non-linear and non-uniformly behavior is typical for soft tissue deformation. Hence, the gray value pattern represents a difficult input signal for common classification methods. Particularly, it is not known where the essential information is located within the sequence. This is similar to problems in sequence classification, such as handwriting recognition [16] or
phoneme classification [14]. Bidirectional Long Short Term Memory (BLSTM) neural networks have been shown to be suitable for such classification problems, as they can handle even the raw input and usually outperform other sequence classification methods, e.g., Hidden Markov Models (HMMs) or Dynamic Time Warping (DTW) substantially [14, 16]. In contrast, approaches minimizing the empirical risk, e.g., Support Vector Machines (SVM), rely on an appropriate preprocessing of the data. Long Short-Term Memory (LSTM) networks [11–13] are a special type of recurrent neural networks containing LSTM blocks instead of or in addition to standard cells, i.e., non-linear sigmoidal cells. LSTM blocks are themselves small recurrent subnetworks and can be seen as differentiable memory cells. These subnetworks improve recurrent neural networks trained with gradient based methods. For instance, LSTMs are capable to learn even widely spaced temporal encoded patterns (up to 10,000 samples). But they are also able to deal with noisy input data and handle non-uniformly compressed sequences. Especially the latter two aspects make them interesting and promising for the sequence classification task discussed in this paper.

An advanced recurrent architecture are bidirectional LSTM networks [14] or just BLSTMs. Bidirectional recurrent neural networks such as BLSTMs process input sequences in forward and backward direction at once, while traditional recurrent networks process sequences in forward direction only. Thus, bidirectional networks are potentially able to extract more useful information from the input sequences. In this article we present details on a preliminary study that indicated the feasibility of such a classification approach [15]. Particularly, we examined patient specific aspects. The generalization strength of the classifier trained and tested on records from a single patient were evaluated regarding the sensitivity and specificity. Moreover, we investigated, which signal features mainly contribute to the classification result. Therefore the behavior on low-pass filtered signals was analyzed on a 10-fold data set. More details concerning BLSTMs and our network architecture can be found in [10].

This article is organized as follows. Sections 2 summarizes the data acquisition process and introduces the experimental setup. Then, in Section 3 the experimental results are presented and afterwards discussed in Section 4.

2. Materials and Methods

The data acquisition was performed with the Spectral Domain OCT device “Callisto” (Thorlabs) with a central wavelength of 930 nm and an acquisition rate of 1.2 kHz. The axial resolution of the device was 7 µm. A customized common path OCT probe has been used to capture the OCT data. The probe was based on an 18 gauge brachytherapy needle with embedded optical fiber. The closed needle tip was replaced by a round-tip-geometry with a central bore hole. A single mode fiber 780HP (Thorlabs) was carefully melted in a fusion splice device, to archive a slightly focusing convex profile.

Freshly resected and non-fixated human lung-tissue-specimens were obtained from lobectomies. Tumor suspicious areas with sizes of multiple centimeters were manually identified by palpation. The needle probe was manually inserted in the target area and fixated to prevent movement artifacts. Both, tumor suspicious and non-suspicious areas were punctured twice. To reduce the influence of the local speckle pattern four different OCT records were obtained from every puncture, while the needle was slightly pivoted between the records. Before the needle was removed the puncture was marked with ink. Afterwards the tumor entities were verified through pathological examination.

Each record consisted of 100 subsequent A-scans. The length of each A-scan was reduced to a depth of 200 grey values. This corresponds to a depth of 0 to 0.67 mm from the needle tip. From 15 different tissue specimens 75 records from lung tissue and 22 records from tumor tissue were acquired. The tumor data represented 13 adeno carcinoma and 2 epithel carcinoma.

We considered two general scenarios, where only data from different patients were used to classify an A-scan, and where
only data from different records – but the same patient – were used to classify an A-Scan. For the initial experiments [15], all A-Scans were randomly divided in 30 different distributions of training and test data as follows. 1) All data were partitioned such that a patient was represented either in the test or training set (PatAll). 2) The data from adenocarcinoma were partitioned such that a record was represented either in the test or training set (PatAd). 3) All data were partitioned such that a record was represented either in the test or training set (RecAll). 4) The data from adenocarcinoma were partitioned such that a record was represented either in the test or training set (RecAd). For 1 and 2 the number of patients and for 3 and 4 the number of records were divided with the ratio 3:1 to be in training set or test set. Thereby the A-scans were divided such that tumor and lung tissue were equal in quantity. BLSTM nets with a common net architecture consisting of two hidden layers (one in forward and one in backward direction) and 8 LSTM blocks were trained on the training data and applied to test data. The mean classification accuracies and the variance over 30 trials were calculated for all four distributions.

Further experiments regarding the number of LSTM blocks indicated that a net with 24 LSTM blocks can archive a reasonable tradeoff between generalization and classification accuracy [10]. Moreover the experiments indicated that merging all 100 A-scans by averaging to one profile would slightly increase the classification performance.

Based on the BLSTM 24 configuration further experiments regarding the influence of low pass filter and patient specificity were performed. Hence, for the filter experiments the averaged profiles were distributed in test and training sets 10 times with ratio 2:1. A zero phase filter with butterworth kernel of order one with successive decreasing cutoff frequency were applied to the A-scans before training the net. To investigate the patient specificity datasets from one individual patient containing 4 tumor records and 4 non-tumor records were randomly distributed 10 times with ratio 3:1. The results of the test set were compared to the unfiltered 10-fold data set regarding the sensitivity, specificity and accuracy.

All nets were trained using gradient descent with momentum term, while the gradients were calculated by using back propagation through time. Since the BLSTM architecture itself already provides good generalization capabilities compared to classical RNN’s only early stopping was used for regularization.

3. Results

The classification results of the initial experiments are shown in Figure 2. The mean accuracies over 30 distributions with the corresponding variance are pair wise visualized for training and test dataset in four different configurations. Case one (PatAll) has the lowest classification rate with 67.5% (1.4%) and 63.5% (2%) which only slightly increases for case two (PatAd) with 69.1% (1.7%) and 63.9% (4.4%). The variance is relatively high. Case three (RecAll) and four (RecAd) show a higher classification rate of 70.6% (0.74%) and 70.9% (1.1%) as well as 72% (0.8%) and 76% (1.72%) with a smaller variance.

4. Discussion

The results indicate that individual OCT A-scans can provide information to differentiate between tumor and non-tumor tissue in the lung. However, the classification
accuracies are not sufficient to form the sole basis for diagnosis, particularly for the initial experiments and without a priori knowledge about the tumor entity. Still, in context of a biopsy the additional information could help to estimate whether tumor tissue was reached by the instrument. The initial results as presented in Figure 2 indicate a significant influence of patient specific properties on the OCT signal. Additional experiments regarding to the patient specificity strongly supported this point of view. Although, the limited amount of patient specific data provides only little evidence. The results for low pass filtered A-scans indicate that particularly very high and very low frequency components contribute to the classification. This supports other studies proposing signal attenuation and speckle variance as potential signal characteristics. Attenuation and speckle noise are typically represented by relatively low and high frequency components. The generalization performance of neuronal networks highly depends on a suitable net architecture and regularization method. Our experiments were based on a net architecture commonly used in literature with early stopping as an implicit regularization technique. However, using explicit regularization, such as weight decay may further improve the generalization capabilities.

Nevertheless the limited insight into the data processing is one major drawback of neuronal networks.

5. Conclusion and Future Work

The presented approach and the initial results indicate that an automatic classification based on A-scans is feasible, although the method seems not suitable to provide a solitary diagnosis. The results presented in Figure 2 indicate a substantial influence of patient specific properties on the OCT signal. When training on data from a different record of the same patient is allowed, the accuracy increases by approximately 5%. Additional experiments regarding patient customized classification scenarios, based on the methodology in [10] confirmed the initial appearance of significant patient specific signal properties. A clinical scenario for a patient specific training would be a second intervention after an initial biopsy, if the previously collected A-scans are known to be from tumor area. Systematically signal features have been suppressed by applying a low pass filter with successively increasing cutoff frequencies to the A-scans. The classification results implied important signal features to be located in the lower and upper part of the spatial frequency spectrum.
Further investigations regarding a sufficient preprocessing of the gray values would be desirable and may improve the classification performance. Based on that further effort should be made to prove an optimal classification method for this scenario.

References