Jirapong Manit*, Achim Schweikard, and Floris Ernst

Deep convolutional neural network approach for forehead tissue thickness estimation

Abstract: In this paper, we presented a deep convolutional neural network (CNN) approach for forehead tissue thickness estimation. We use down sampled NIR laser backscattering images acquired from a novel marker-less near-infrared laser-based head tracking system, combined with the beam's incident angle parameter. These two-channel augmented images were constructed for the CNN input, while a single node output layer represents the estimated value of the forehead tissue thickness. The models were – separately for each subject – trained and tested on datasets acquired from 30 subjects (high resolution MRI data is used as ground truth). To speed up training, we used a pre-trained network from the first subject to bootstrap training for each of the other subjects. We could show a clear improvement for the tissue thickness estimation (mean RMSE of 0.096 mm). This proposed CNN model outperformed previous support vector regression (mean RMSE of 0.155 mm) or Gaussian processes learning approaches (mean RMSE of 0.114 mm) and eliminated their restrictions for future research.

Keywords: Forehead Skin, Tissue Thickness, Deep Convolutional Neural Network, Regression, Head Tracking, Near-infrared Laser, Backscatter

https://doi.org/10.1515/cdbme-2017-0022

1 Introduction

Patient head motion during treatment is a critical issue in cranial radiotherapy. As the fact that inaccurate delivery of a prescribed radiation dose to the target volume may decrease the efficacy of the treatments and cause unwanted side effect, immobilisation devices such as thermoplastic masks were introduced to overcome the problem. However, these methods are not yet the optimum solution due to many restrictions and drawbacks, e.g., complicated setup, patient discomfort, lack of reusability and inaccuracies when used over multiple treatment fractions [1].

Instead of prohibiting the patient from moving, we have proposed a novel marker-less near-infrared laser base system to tack the patient's head [2]. This tracking system utilises an 850nm laser beam to obtain the 3D forehead geometry, and uses a machine learning algorithm to estimate the underlying tissue thickness (see **Figure**). The radiotherapy system then can adjust the beam trajectory and compensate for the patient's motion. Nevertheless, the performance of this tracking system critically depends on the accuracy of the forehead tissue thickness estimation.

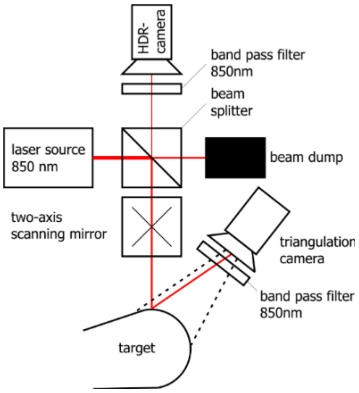

Figure 1: Components and configuration of the NIR laser-based head tracking system. The 850nm laser is projected onto the subject's forehead through the x-y scanning mirrors. The triangulation camera is used to estimate the laser spot in 3D space, while the HDR-camera captures backscatter from the NIR laser.

***Corresponding author: Jirapong Manit:** Institute for Robotics and Cognitive Systems, Graduate School for Computing in Medicine and Life Sciences, University of Lübeck, Germany, e-mail: manit@rob.uni-luebeck.de
Achim Schweikard, Floris Ernst: Institute for Robotics and Cognitive Systems, University of Lübeck, Germany, e-mail: schweikard@rob.uni-luebeck.de, ernst@rob.uni-luebeck.de

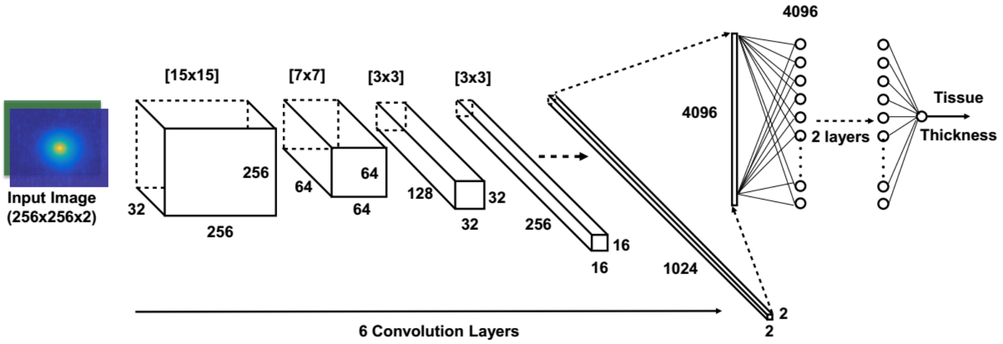

Figure 2: Proposed CNN architecture: it consists of six convolution layers and three fully connected layers for regression. The value from the output node represents the estimated tissue thickness in millimetre.

Our previous research showed that with support vector machine regression (SVR) or Gaussian processes (GP) estimation by using combined features (image data, incident angle and neighbourhood features), it is possible to determine tissue thickness with RMSE of 0.217 mm and 0.1140 mm, respectively [3]. These previous learning methods, however, require the user, to manually define the learning features [3,4], which may not be good enough for this application. Although the GP algorithm has shown good performance, its complexity of $O(n^3)$ limited the training on training sets with a large size.

Recently, a powerful artificial intelligent algorithm called *deep learning* has been introduced for solving computer vision problems. The main advantage of this algorithm compared to classical approaches is the ability to automatically learn how to extract image features. In 2014, *Simonyan, et al.* proposed a simple, and yet effective convolutional neural network (CNN) named VGGnet [5]. It could perform image classification task with 84.0% accuracy. Hence, if we could apply this CNN model to create an optimal feature extractor, the tissue thickness estimation accuracy should also be improved

In this paper, we study the feasibility of employing a deep learning approach for tissue thickness estimation by using the full backscatter image as the model input, and comparing the training results with our previous algorithms.

2 Deep learning architecture

Many different deep learning architectures were proposed so far to solve various types of problems (both supervised and unsupervised learning). Most of them are designed for the applications of image classification, object recognition, and natural language processing, where the output layer consists of several nodes representing the probability of each label. The output of our problem is, in contrast, only one floating point number denoting the estimated tissue thickness in millimetre. Therefore, we chose to study the feasibility of employing deep learning to solve a regression problem by using a simple and straightforward model such as VGGnet.

The architecture of the CNN model used in this study is shown in **Figure 2**. The model was slightly modified from the concept of VGGnet to receive input images with a dimension of $256 \times 256 \times 2$. It consists of six convolutional layers working as feature extractors. The kernel sizes of the first and the second convolutional layers are 15×15 and 7×7, respectively, while the size from the third layer on is set to 3×3. Each layer is followed by a 2×2 subsampling layer. The dimension of the output image from these layers is $2 \times 2 \times 1024$, then converted into 1×4096 for feedforwarding to the next regression input layer.

The regression layers are three fully-connected layers with 4096 hidden nodes, and they use $\tanh$ as their activation function. Since the output of this model represents the estimated thickness value in millimetre, the final output layer is only a single node without any activation function.

During the training phase, dropout regulation was applied for the regression layers to prevent overfitting and enhanced training speed ($p = 0.2$ for the first layer and $p = 0.5$ for the second and third layers).

3 Methodology

3.1 Data acquisition

To evaluate and compare the tissue thickness estimation accuracy of this CNN model, the dataset we used in this experiment was the same dataset that was acquired from the experiment in [3]. This dataset was collected by the laser-based head tracking system from 30 healthy subjects; 16

males and 14 females, aged from 24 to 65 years. The information of each subject consisted of:

1. A 3D point cloud represented the scanning points where the laser projected onto the subject's forehead.
2. The corresponding 900×900 pixel NIR laser backscattering images captured from the HDR-camera.
3. The corresponding incident angles between the laser direction and the normal vector of the forehead surface at the projected location.
4. The thickness of the underlying tissue extracted from their MRI ground truth (acquired with a resolution of 0.1 mm)

3.2 Data pre-processing

An important property of the NIR backscattering image is that the intensity values of the region around the laser spot centre could go up to 65,535, while the area around the spot will contain values ranged from 100 to 1,000. The results of simulations in [4] showed that the values in the region surrounding the spot also have a significant relation to the tissue thickness. Hence, this extreme difference in magnitude of the pixel values could degrade the quality of the information during the network training phase.

To prevent this dominating phenomenon, the intensity values of the input images were subjected to rescaling by applying an operator $20 \cdot \log(I)$, where I is the original intensity value. With this transformation, the intensity dynamic range was shortened, emphasising the details lying around the central specular reflection (see **Figure 3**).

The transformed images were then downscaled to the desired dimension by using bicubic interpolation, and finally an additional layer was inserted, which contains the corresponding laser incident angle, forming the augmented input image ($256 \times 256 \times 2$ pixels).

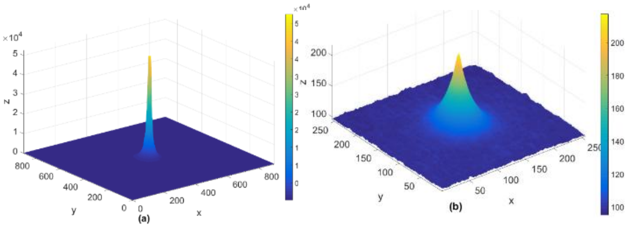

Figure 3: A comparison between (a) an original 900x900 pixels image and (b) its logarithmically rescaled result (256x256 pixels). The transformed image emphasises the details lying on the area directly around the NIR laser spot centre.

3.3 Model training and validation

The performance of the proposed CNN model was tested by using leave-one-out validation. The number of elements in the training set for each fold was set to 80% of the total data by randomly selecting based on the uniform distribution. The remaining part was then labelled as the validation set. Due to some corrupted images during acquisition, the total number of laser point images for each subject was not identical. The estimation error was evaluated by the differences between the CNN output value and the thickness provided from the MRI ground truth.

It is well known that training deep learning models may take a long time before converging. Our experiment aimed to train 900 CNN models, which would take exceedingly long. Hence, we used a pre-training model to speed up the training. We manually selected the CNN model from the first fold of Subject5 as the pre-training network for the other subjects' training, while every testing on Subject5 was done using an untrained network.

The training parameters in this experiment were *learning rate* = 0.001, *momentum* = 0.05 and *weight decay* = 0.001. It was set to run for at least 100 epochs, then continued until either of the two termination criteria was met: 1) 800th epoch was reached or 2) The validation error was less than or equal to 0.090 mm.

The CNN model was implemented using the Torch library on a PC with an NVIDIA GTX 980 GPU and an Intel® Core™ i5 CPU (3.40 GHz), running Ubuntu 14.04.

4 Results and discussion

The results from the entire testing models is illustrated in the two bar plots of training error and validation error, as shown in **Figure 4**. The mean training and validation RMSE of the entire testing was 0.096 ±0.028 and 0.077 ±0.012, respectively. Comparing this accuracy with the previous results in **Table 1**, the CNN model clearly outperformed both SVM and GP learning with All features. When considering only the learning from image and angle features (no neighbourhood data), we can gain an improvement by 48.78%.

The examples of the training progress on the untrained and pre-training network are shown in **Figure 5**. We saw that the validating error of the pre-trained networks (Subject5), started from a lower value (0.211 mm) than the untrained network's initial point (1.893 mm). Then the errors of both networks continuously decreased and reached the termination condition at the point where the validation errors went below

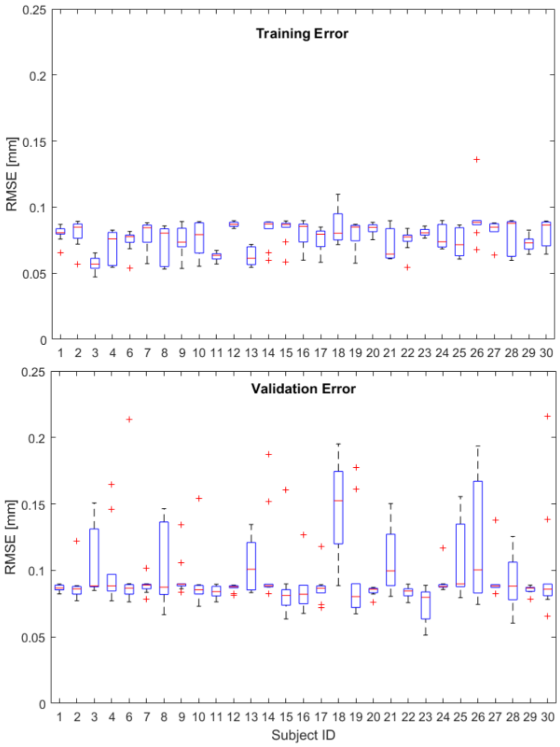

Figure 4: The bar plot of training and validating RMSE from every folds and subjects. The average RMSE of the training errors is 0.077 mm, while it is 0.096 mm for validation.

0.090 mm. However, their validation errors still tended to be slightly decrease when training continued. The fastest learning on the pre-training model was likely to take less then 350 epochs to reach the terminate condition, while the untrained model needed more than 500 epochs. These results confirmed that using a pre-training model for a dataset from a different subject to enhance the model convergence is possible.

Overfitting models also still occurred in the learning of 22 subjects, and they were 19.26% of the total testing. The graph of the Subject#8's 5th fold training in **Figure 5** demonstrates this overfitting situation. At the early training epochs, both errors descended in the same manner, but then the validation error began to stop improving after the 200th epoch, although the training error slowly moved to below 0.1 mm.

Table 1: RMSE in millimetre of the estimation produced by the proposed SVM, GP and CNN. The learning features are: region-of-interest (ROI), incident Angles and neighbourhood data (NBH).

SVM			GP			CNN
ROI	*ROI*	*NBH*	*ROI*	*Angle*	*NBH*	
0.225	0.217	0.155	0.207	0.196	0.119	0.096

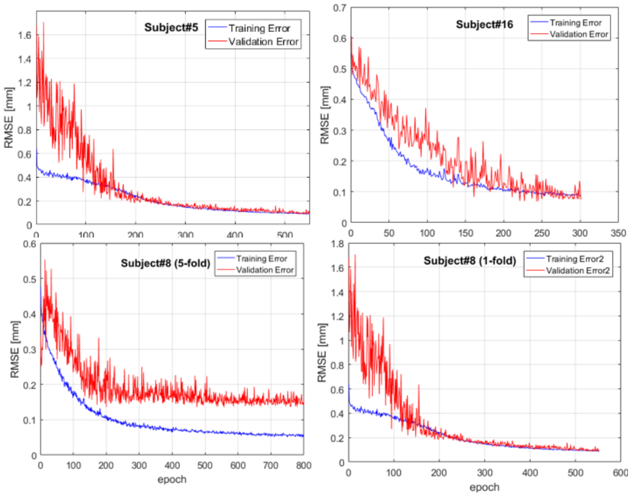

Figure 5: Graph plots show the learning progress of the untrained network (Subject#5), pre-training network (Subject#16), an overfitting model (Subject#8's 5th fold) and a good model (Subject#8 1st fold).

5 Conclusion

We have applied the convolutional neural network for forehead tissue thickness estimation by using the full NIR laser backscattering images. The intra-subject validation results show clear improvements when compared to all previously used learning approaches. While employing the pre-training network can also bootstrap the training, overfitting is still present in several cases. The results of this experiment also demonstrate that CNN can solve regression problems as well as recognition and classification.

In the future, we will enhance the CNN model to be able to consider the laser point neighbourhood relations, and investigate ways to eliminate overfitting.

Author's Statement

Research funding: This work was supported by Varian Medical Systems, Inc., by the Graduate School for Computing in Medicine and Life Science [DFG GSC 235/1], and the Business Development and Technology Transfer Corporation of Schleswig-Holstein (WTSH) [123 16 014]. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent is not applicable. Ethical approval: The conducted research is not related to either human or animals use.

References

[1] Sharp L, Lewin F, Johansson H, Payne D, Gerhardsson A, Rutqvist R E, Randomized trial on two types of thermoplastic

masks for patient immobilization during radiation therapy for head-and-neck cancer, International Journal of Radiation Oncology*Biology*Physics, Volume 61, Issue 1, January 2005, Pages 250-256, ISSN 0360-3016

[2] Stüber P, Wagner B, Wissel T, Bruder R, Schweikard A and Ernst F. An Approach to Improve Accuracy of Optical Tracking Systems in Cranial Radiation Therapy. Muacevic A, Adler JR, eds. *Cureus*. 2015;7(1):e239. doi:10.7759/cureus.239. Institute of Medicine (US). Washington: The Institute 1999.

[3] Wissel T, Stüber P, Wagner B, Bruder R, Erdmann C, Christin-Sophie Deutz, Sack B, Manit J, Schweikard A and Ernst F, Enhanced Optical Head Tracking for Cranial Radiotherapy: Supporting Surface Registration by Cutaneous Structures (2016), in: International Journal of Radiation Oncology, Biology, Physics, 95:2(810-817)

[4] Wissel T, Bruder R, Schweikard A, Ernst F. Estimating soft tissue thickness from light-tissue interactions—a simulation study. *Biomedical Optics Express*. 2013;4(7):1176-1187. doi:10.1364/BOE.4.001176.

[5] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).