Fully Convolutional Mandible Segmentation on a valid Ground-Truth Dataset

Jan Egger [a,b,c], Birgit Pfarrkirchner [a,b], Christina Gsaxner [a,b], Lydia Lindner [a,b]
Dieter Schmalstieg [a], Jürgen Wallner [b,d]

[a] Institute for Computer Graphics and Vision, TU Graz, Inffeldgasse 16, 8010 Graz, Austria
[b] Computer Algorithms for Medicine Laboratory, Graz, Austria
[c] BioTechMed, Krenngasse 37/1, 8010 Graz, Austria
[d] Dept. of Oral and Maxillofacial Surgery, Med Uni Graz, Auenbruggerplatz 5/1, 8036 Graz, Austria

Introduction

Deep learning [1] with neural networks is an increasingly important topic for research and economic purposes. Software giants use deep networks for the development of their latest technological gadgets. Daily examples are Facebook's face detection, Apple's speech recognition Siri or Google Translate, which all comprise deep learning algorithms [2].

The motivation of this contribution is to utilize deep learning networks for medical image processing and analysis, and create a more reliable ground-truth. In particular, the aim was to implement convolutional neural networks (CNNs) as well as to train and test them with computed tomography images from the clinical routine in order to enable an automatic segmentation of the lower jawbone.

Methods

The deep learning implementations of this work comprise classification as well as segmentation networks. The idea is to mark out the images, which show parts of the lower jawbone (mandible) in head and neck CT data sets, with a trained classification net and to provide those slices to the segmentation networks. The reason for this two-step implementation is that many CT slices occur, which don't display the anatomical region of interest. Hence, various classification and segmentation networks were implemented as well as trained and tested with the deep learning framework TensorFlow [3] and its higher level application programming interfaces (API). The results show that the automatic segmentation of the mandible works adequately for the available CT datasets.

During this work, classification networks with various net topologies were trained with four different sized datasets. Each dataset contained a diverse number of images, as there were different augmentation methods applied [4]. The first image set involved the initial CT images (1680 slices), the second one was enlarged with noisy images (6720 slices) and the third one with affine transformed ones (13440 slices). Dataset four covered both data augmentation types (18480 slices). The affine transformations were applied separately from each other (for each slice separately).

The implementation of the deep networks was conducted with TensorFlow and its high-level API TF-Slim [5]. Again, the Python interface of TensorFlow was utilized. The realized segmentation method follows the upsampling principle presented by Long et al. [6] in their Fully Convolutional Networks for Semantic Segmentation contribution as well as the contribution of Pakhomov et al. [7]. As already outlined, Long et al. recommended a three-step training principle of a fully convolutional network. Figure 1 illustrates the workflow of the model implementations.

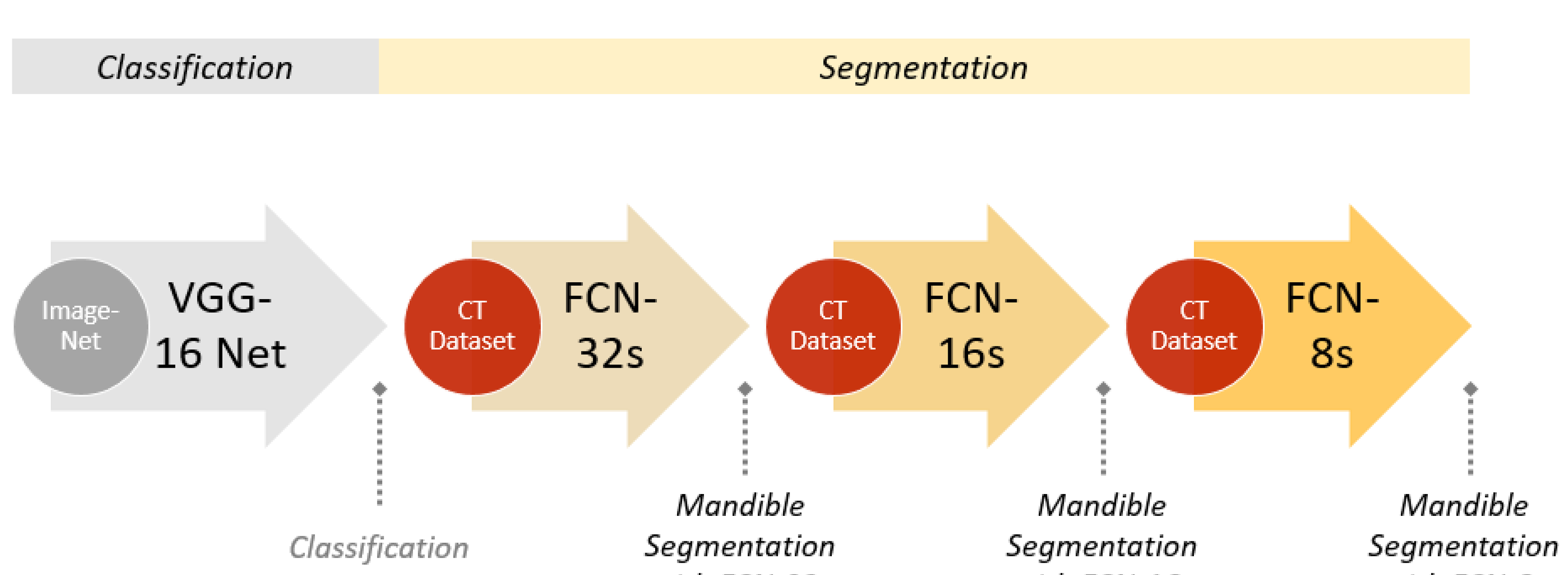

Figure 1 – Workflow of the segmentation network implementations. The classification part was provided by the TF-Slim library, whereas the segmentation part was trained with the CT datasets during this work.

Results

Figure 2 displays classified images (50x50) and their predicted probabilities. The class predictions were accomplished with the best performing classification model. This trained network delivered an accuracy of one for the training dataset, whilst the test accuracy had a value of 0.9877.

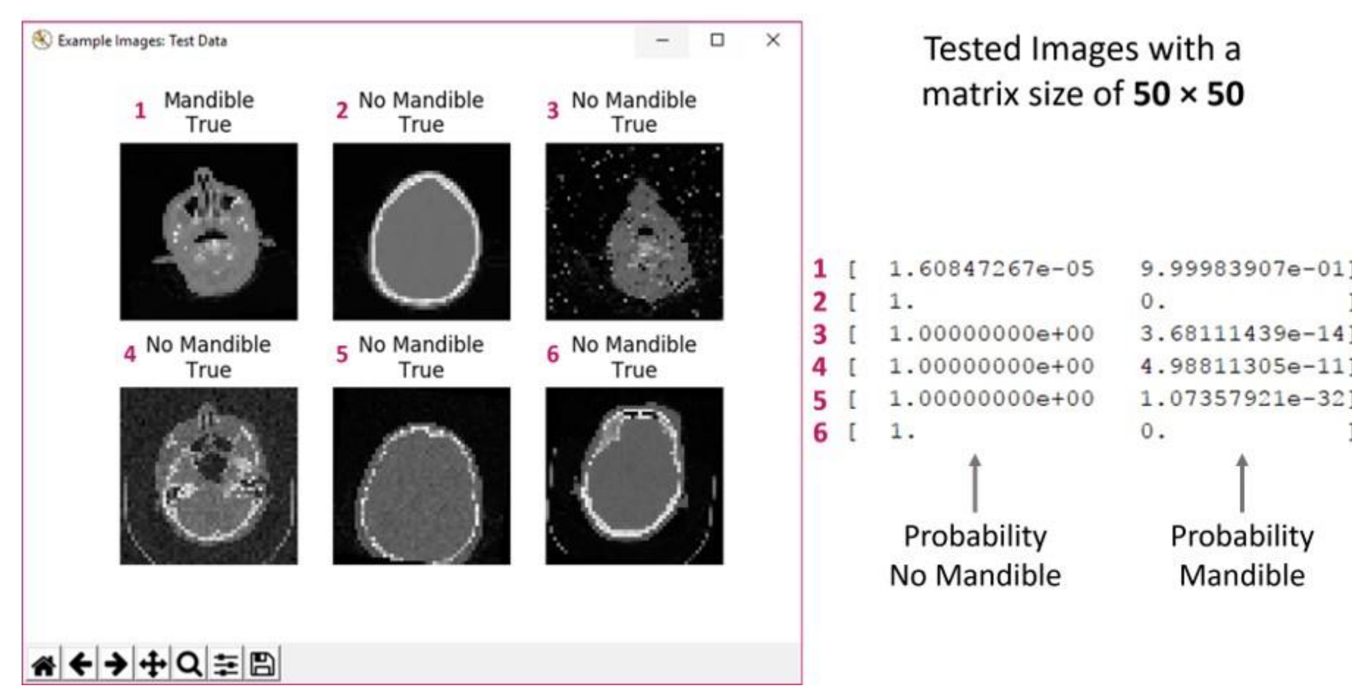

Figure 2 – Test images (50x50) and their predicted classes. The network exhibited a topology of six convolutional and six max-pooling layers. Training was accomplished with the largest sized dataset.

Figure 3 shows a CT slice, its ground truth and the predicted probabilities of the three network architectures: FCN-32s, the FCN-16s and the FCN-8s. The networks used for those predictions were trained with the largest dataset and the original image sizes. The figure indicates that the predictions improve with the involvement of information of the VGG-16 model. The segmentations of the FCN-32s architecture seem to be awkward, whilst the FCN-8s predictions are smoother.

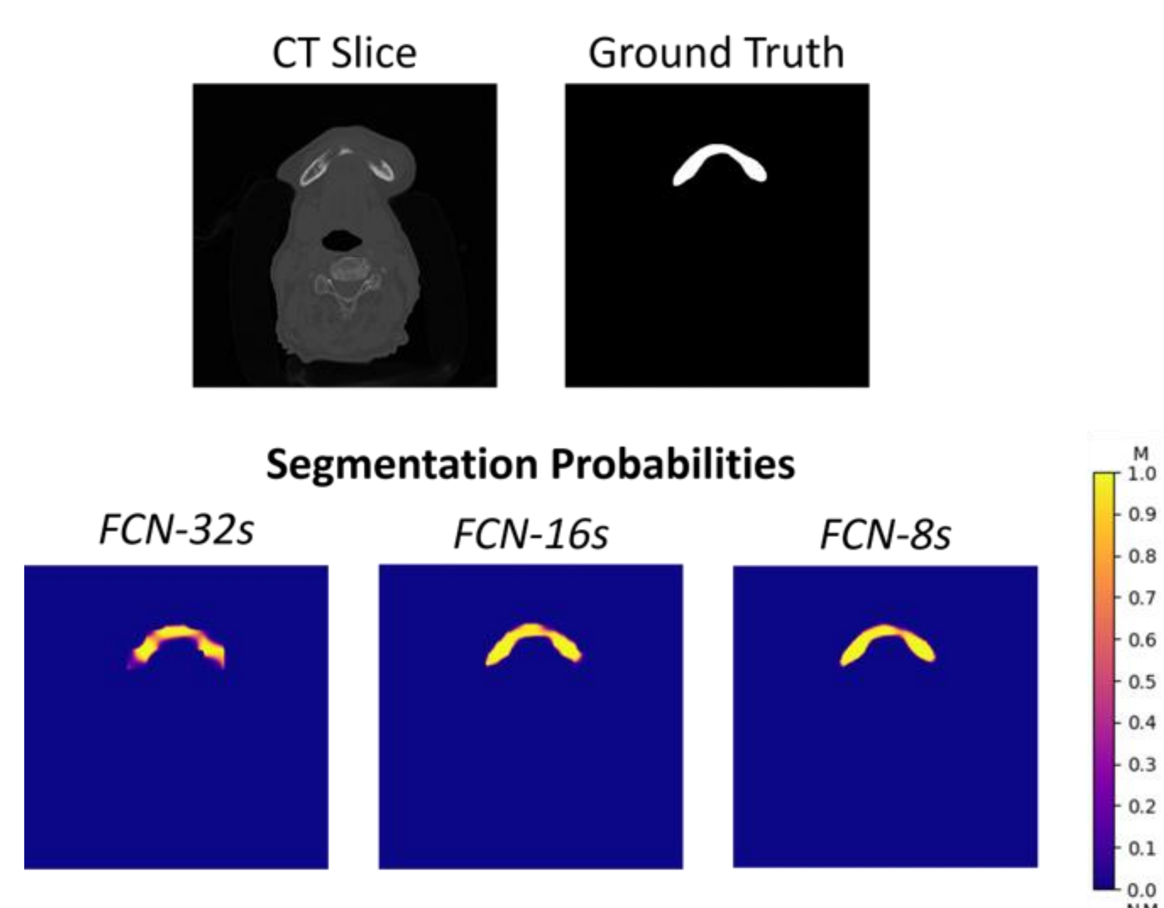

Figure 3 – Depiction of a CT slice, its ground truth and the predicted probability maps. The maps were forecasted with the networks trained with dataset IV. The brighter the voxels, the more likely they are part of the mandible (M), whilst the blue color implies that there is probably no mandible (NM) appearing.

Conclusions

A MeVisLab [1, 2] network and a macro module were generated to process and enlarge the head-neck CT datasets during this contribution. Moreover, the ultimate objective was to implement deep networks, which permit an automatic segmentation of the mandible. Therefore, classification networks were trained in order to distinguish whether a slice comprises the lower jawbone or not and consequently, segmentation networks computed the algorithmic demarcations within these slices. All networks were trained and tested with images exported by the MeVisLab realizations.

Acknowledgements

The work received funding from BioTechMed-Graz in Austria, the 6th Call of the Initial Funding Program from the F&T-Haus at TU Graz, the TU Graz Lead Project (Mechanics, Modeling and Simulation of Aortic Dissection) and the Austrian Science Fund (FWF) KLI 678-B31: "enFaced: Virtual and Augmented Reality Training and Navigation Module for 3D-Printed Facial Defect Reconstructions" (Principal Investigators: Jan Egger & Jürgen Wallner).

References

1. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, 521(7553):436-444, May 2015.
2. N. Jones, "Computer science: The learning machines," Nature, 505(7482):146-148, Jan. 2014.
3. M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," CoRR, abs/1603.04467, 2016.
4. B. Pfarrkirchner et al., "Lower jawbone data generation for deep learning tools under MeVisLab," SPIE Medical Imaging Conference, Paper 10578-96, 2018.
5. S. Guadarrama and N. Silberman, "TensorFlow-Slim," Available: https://github.com/tensorflow/tensorflow/tree/master/tensorflow/ contrib/slim, 2017. (Last access 27.09.2017).
6. J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," CVPR, June 2015.
7. D. Pakhomov et al., "Deep Residual Learning for Instrument Segmentation in Robotic Surgery," arXiv preprint arXiv:1703.08580, 2017.
8. J. Egger et al., "HTC Vive MeVisLab integration via OpenVR for medical applications," PLoS ONE 12(3): e0173972, 2017.
9. J. Egger et al., "Integration of the OpenIGTLink Network Protocol for Image-Guided Therapy with the Medical Platform MeVisLab," Int J Med Robot. 8(3):282-90. 2012.

40th International Conference of the IEEE Engineering in Medicine and Biology Society
EMBC '18
July 17-21
Honolulu, HI

Hawaii Convention Center
Honolulu, HI, United States
17 July - 21 July 2018