MSc Thesis Project Plan

Luca Carotenuto (luca.carotenuto@ru.nl)

September 2021

Start date:	01-09-2021
End date:	01-03-2022
External supervisor:	Guido de Jong Email: guido.dejong@radboudumc.nl
Internal supervisor:	Prof. Tom Heskes Email: tom.heskes@ru.nl
(Preliminary) project title:	Landmark detection in 3D images for Cranio-Maxillofacial Surgery using Deep Learning
Study program:	Computing Science (Data Science)

1 Problem Statement

Three-dimensional (3D) landmarks are used in various fields of medicine e.g. for the alignment of 3D radiological images, 3D scans or 3D photos. The department of Oral and Maxillofacial Surgery at Radboudumc needs to place 3D landmarks for surgical planning, follow-up and diagnostics. Placing 3D landmarks manually is a tedious process with a high degree of interobserver and intraobserver variability. Automatic landmark detection can ensure consistent and precise landmark placement. However, computer-assisted approaches need to deal with challenges such as variation in the capturing device/data source, different pre-processing techniques or varying sampling resolution.

The Radboudumc 3D lab has 3D CT-scans and 3D photos of the face and cranium in the following modalities: bony-tissue CT-scans, soft-tissue CT-scans, soft-tissue 3D photos and soft-tissue textured 3D photos. Per modality, a set of 12 to 36 clinically relevant landmarks have to be detected. The data set comprises several hundred partly annotated samples.

2 Approach

With the recent successes and the emerging field of artificial intelligence (AI) within 3D technology, the goal is to develop a neural network that automatically places landmarks in 3D photos and CT scans of faces or craniums. The final approach either fully or partly relies on AI. We try to outperform manual and other non-deep learning 3D landmark detection approaches in accuracy and time.

A deep learning approach in combination with one of the four modalities is chosen. The approach will likely either be a convolution-based approach such as MeshCNN, which attempts to mimic the convolution operation that is conventionally used on regular grid data and apply them on the edges of the mesh data. Another possible approach is a point-based method, such as PointNet [Qi et al., 2017a] that uses point clouds as a data source and takes into account the permutation invariance of points. Note, that these approaches only support classification and segmentation in their original versions and have to undergo architectural changes to support the detection of key points.

Depending on the chosen modality, several hundred annotated samples will be used to train the neural network. Data augmentation can be applied to increase the size of the training set and improve generalization.

Computing resources are provided by Radboudumc for experimentation and training of the deep learning model.

3 Literature

There is little to no literature about 3D deep learning for the application of facial landmark detection that operates directly on the 3D data source (mesh data). Instead, we rely on literature that describes 3D deep learning in more general terms or for applications that are similar to ours [Guo et al., 2020], [Bello et al., 2020].

3D deep learning is an emerging field but still lacks behind 2D computer vision. This is not only due to the shortage of large 3D data sets but also because techniques such as convolutions, that Convolutional Neural Networks (CNNs) rely on, are not directly transferable to 3D due to the unorderedness and irregularity in the data. Promising approaches are PointNet [Qi et al., 2017a] that use point clouds and consider the permutation invariance of points in the input. Variations of PointNet are PointNet++ [Qi et al., 2017b] that manage to improve classification and segmentation performance by modelling local regions through sampling and grouping or PointCNN [Li et al., 2018] that take into account the correlation between points in the local regions. An approach that operates directly on triangular mesh data is MeshCNN [Hanocka et al., 2019], which applies convolutions on edges and the four edges of their incident triangles, and pooling is applied via an edge collapse operation.

4 Plan

The following is a rough outline of the planning for my 6-month internship. The first month is mainly literature research. That includes reading into 3D deep learning and finding out which approaches exist and work well on similar applications. Moreover, I get familiar with computer graphics terminology and polygonal mesh processing [Botsch et al., 2010] including the mesh processing software MeshLab. In the second month, I familiarize myself with the code of the deep learning methods that I plan on using. I prepare the data, change the architecture of the chosen deep learning models to work for landmark detection and possibly do adaptations that benefit this task. I apply the adapted models on the Headspace dataset [Dai et al., 2020], which contains face and cranium 3D data and is available for University-based non-commercial research. This data can be used for experiments that are performed outside the Radboudumc network. In the third month, I apply the deep learning approach that works on the Headspace data on the data from Radboudumc and train the network with the provided computing resources. In the fourth month, the network with the best performance from the experiments is chosen and if necessary, trained for a final time. The results are prepared and visualized. The last two months are dedicated to the final thesis and delivering results.

There are weekly meetings in addition to spontaneous exchange in my everyday work with my external supervisor, research coordinator of the 3D Lab at Radboudumc. Moreover, I will meet monthly with my internal supervisor, professor of Artificial Intelligence at iCIS, Radboud University.

A git repository will be set up for version control and code distribution.

References

[Bello et al., 2020] Bello, S. A., Yu, S., Wang, C., Adam, J. M., and Li, J. (2020). Review: Deep learning on 3D point clouds.

[Botsch et al., 2010] Botsch, M., Kobbelt, L., Pauly, M., Alliez, P., and Lévy, B. (2010). *Polygon mesh processing*. CRC press.

[Dai et al., 2020] Dai, H., Pears, N., Smith, W., and Duncan, C. (2020). Statistical Modeling of Craniofacial Shape and Texture. *International Journal of Computer Vision*.

[Guo et al., 2020] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., and Bennamoun, M. (2020). Deep Learning for 3D Point Clouds: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Hanocka et al., 2019] Hanocka, R., Hertz, A., Fish, N., Giryes, R., Fleishman, S., and Cohen-Or, D. (2019). MeshCNN. *ACM Transactions on Graphics*.

[Li et al., 2018] Li, Y., Bu, R., Sun, M., Wu, W., Di, X., and Chen, B. (2018). PointCNN: Convolution on X-transformed points. In *Advances in Neural Information Processing Systems*.

[Qi et al., 2017a] Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017a). PointNet: Deep learning on point sets for 3D classification and segmentation. In

Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017.

[Qi et al., 2017b] Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017b). PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*.