

Automatic Facial Landmark Detection for Neurosurgical Mixed Reality Applications in MRI and CT scans using Deep Learning

Mathijs de Boer¹

M.DEBOER-41@UMCUTRECHT.NL

Lambertus W. Bartels¹

W.BARTELS@UMCUTRECHT.NL

Pierre A.J.T Robe²

P.ROBE@UMCUTRECHT.NL

Tristan P.C. van Doormaal^{2,3}

T.P.C.VANDOORMAAL@UMCUTRECHT.NL

¹ *Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands*

² *Department of Neurosurgery, University Medical Center Utrecht, Utrecht, The Netherlands*

³ *Department of Neurosurgery, University Hospital Zürich, Zürich, Switzerland*

Abstract

Mixed reality (MxR) has the potential to revolutionize the way neurosurgical interventions are performed. However, the use of MxR in the operating room (OR) introduces new challenges, such as the registration of preoperative images to the patient. This paper presents a deep learning method for automatic detection of facial landmarks in Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scans, which can be used for image-to-patient registration in MxR applications. The method achieves a mean error of $4.02(\pm 2.65)$ mm in $3.55(\pm 1.53)$ seconds on a CPU. Apart from the nasion, no statistically significant differences were found between the performance of the method between the CT and MRI scans.

Keywords: Deep Learning, Facial Landmark Detection, Geometry, Mixed Reality, Neurosurgery

1. Introduction

Mixed reality (MxR) has the potential to enhance neurosurgical interventions by providing additional spatial information. However, the use of MxR in the operating room (OR) also introduces new challenges. To effectively use MxR for neuronavigation, the preoperative images of the patient need to be registered to the patient.

Facial landmarks could be used to perform the image-to-patient registration. The process would involve identifying the location of the same landmarks on the patient and in the preoperative images. Because preoperative imaging may be in the form of Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) scans, the automatic method should be able to predict the landmarks in scans from both modalities.

In this paper, we present a method for the automatic detection of facial landmarks in head CT and MRI scans using deep learning.

2. Methods

Data The skin in 262 scans (125 CT, 137 MRI) was automatically segmented using nnU-Net models [Isensee et al. \(2021\)](#); [De Boer et al. \(2024\)](#). The segmentations were converted to surface models and manually annotated with 7 facial landmarks. These landmarks

being the nasion, left and right medial and lateral canthus, and left and right pre-auricle. The landmarks do not shift significantly between the moment the scan is made and patient arrival in the OR, which is why they were chosen. Finally, the labeled meshes were split into training and testing sets, using an 80:20 split, respectively. At train time, the train set was further split into a training and validation set with an 80:20 split.

Pipeline Each mesh was reprojected to a 1024 by 512 pixel equirectangular representation using a variation of the method described by [Torres et al. \(2021\)](#). Our main change was the use of a spherical projection, instead of a cylindrical one, which allowed for a simpler approach to back-projecting the predicted 2D coordinates to 3D coordinates and does not require setting vertical sampling bounds. The landmarks were also projected to 2D coordinates on the spherical map.

We used a neural network, based on the one by [Kerfoot et al. \(2021\)](#), to predict 2D coordinates of the landmarks. For training this network, the mean absolute error was modified to measure the angular error of the predicted and ground truth 2D coordinates. Finally, predicted 2D coordinates were then converted back to 3D coordinates using the inverse of the spherical projection method.

Evaluation Euclidian distances between the predicted and ground truth 3D landmarks were measured. The errors were not normally distributed, so a Mann-Whitney U test was used to compare the errors between the CT and MRI scans. A Šidák corrected p-value of 0.0073 was used as the significance threshold.

To assess feasibility of the method in time-critical situations, the time it took to fully process each mesh was measured. All predictions were performed with an Intel Core i9-9970X CPU. No GPU was used for predictions, as the time required to transfer the model and data to GPU memory was more than the additional time it would take to predict on the CPU.

3. Results

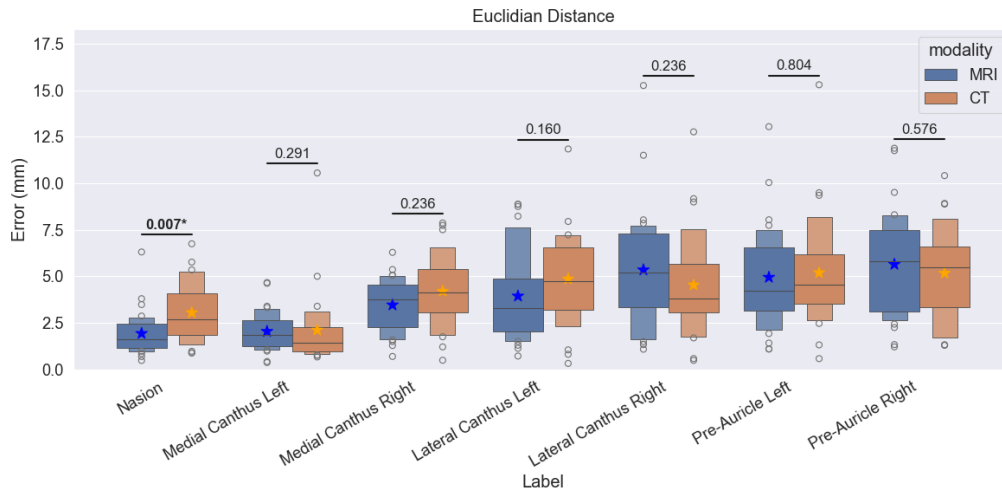
The mean Euclidian distances for all predicted and ground truth landmarks was $4.02(\pm 2.65)$ mm. The nasion error showed a statistically significance between MR and CT scans, no other statistically significant differences were found, see [Figure 1](#). The mean time to predict was $3.55(\pm 1.53)$ seconds.

4. Discussion

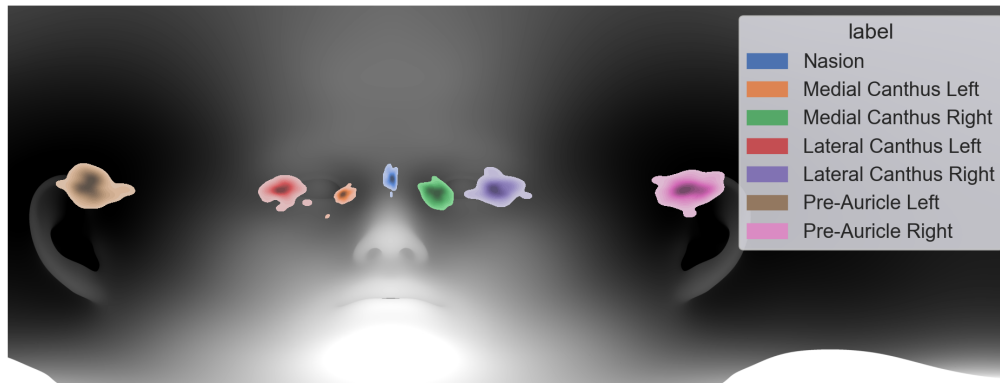
Our results show that the proposed method is able to quickly predict the facial landmarks with a clinically acceptable error. The method is also able to predict the landmarks accurately in both CT and MRI scans.

The nasion showed a significant difference between the CT and MRI scans, albeit not a strong one. It is possible that this is due to intra-observer variability and familiarity bias, as the CT scans were labeled first. [Figure 1\(b\)](#) shows that the nasion displays a larger vertical error distribution, which supports this theory. The method is also able to predict the landmarks in a short amount of time, which is important for its potential use in emergency situations.

While the method shows promise, further refinement and integration into a larger pipeline for neuronavigation are needed. Finally, the effects of inaccuracies on the overall registration accuracy need to be evaluated.



(a) Euclidian distances. Each landmark is marked with a star for the mean value, and a p-value for the Mann-Whitney U test between the MRI and CT values. Significance is marked with bold text and an asterisk.



(b) Distribution of errors. The density of the distribution is visualized using desaturation.

Figure 1: Plots showing the results of the method.

Acknowledgments

We would like to thank Jesse van Doormaal and Tessa Kos for their inputs during the development of the method. We would also like to thank the Hanarth Foundation for providing funding for the MISTICAL project. Finally, we thank The Blender Foundation for releasing the skin mesh we used for visualization under a CC-0 License.

References

- Mathijs De Boer, Tessa M. Kos, Tim Fick, Jesse A. M. Van Doormaal, Elisa Colombo, Hugo J. Kuijf, Pierre A. J. T. Robe, Luca P. Regli, Lambertus W. Bartels, and Tristan P. C. Van Doormaal. NnU-Net versus mesh growing algorithm as a tool for the robust and timely segmentation of neurosurgical 3D images in contrast-enhanced T1 MRI scans. *Acta Neurochirurgica*, 166(1):92, February 2024. ISSN 0942-0940. doi: 10.1007/s00701-024-05973-8. URL <https://link.springer.com/10.1007/s00701-024-05973-8>.
- Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, February 2021. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-020-01008-z. URL <http://www.nature.com/articles/s41592-020-01008-z>.
- Eric Kerfoot, Carlos Escudero King, Tefvik Ismail, David Nordsletten, and Renee Miller. Estimation of Cardiac Valve Annuli Motion with Deep Learning. In Esther Puyol Anton, Mihaela Pop, Maxime Sermesant, Victor Campello, Alain Lalande, Karim Lekadir, Avan Suinesiaputra, Oscar Camara, and Alistair Young, editors, *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, volume 12592, pages 146–155. Springer International Publishing, Cham, 2021. ISBN 978-3-030-68106-7 978-3-030-68107-4. doi: 10.1007/978-3-030-68107-4_15. URL https://link.springer.com/10.1007/978-3-030-68107-4_15. Series Title: Lecture Notes in Computer Science.
- Helena R. Torres, Pedro Morais, Anne Fritze, Bruno Oliveira, Fernando Veloso, Mario Rudiger, Jaime C. Fonseca, and Joao L. Vilaca. Anthropometric Landmark Detection in 3D Head Surfaces Using a Deep Learning Approach. *IEEE Journal of Biomedical and Health Informatics*, 25(7):2643–2654, July 2021. ISSN 2168-2194, 2168-2208. doi: 10.1109/JBHI.2020.3035888. URL <https://ieeexplore.ieee.org/document/9248587/>.