Uncertainty estimation in medical image registration

Elena Serkova
Norwegian University of Science and Technology
Trondheim, Trøndelag, Norway
elenase@stud.ntnu.no

CCS CONCEPTS

• **Computing methodologies → Matching**.

KEYWORDS

image registration, deep learning ensemble, medical imaging, uncertainty estimation, model confidence

ACM Reference Format:
Elena Serkova. 2023. Uncertainty estimation in medical image registration. In *Proceedings of September 20–22, 2023 (10th ACM Celebration of Women in Computing womENcourage).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 ABSTRACT

Medical imaging, or radiology, is the process of recreating accurate graphical representations of a human body. Image processing became an essential part of diagnosing, monitoring and treatment. Imaging technologies in medicine developed significantly in the last decades with an advance in computer technologies. Apart from static two-dimensional images, it became possible to reconstruct three-dimensional images, capture and track motion or record a video. The most common imaging techniques are X-ray, computerized tomography (CT), magnetic resonance imaging (MRI) and ultrasound [2].

Though radiology improves the accuracy of diagnosis and positively affects the treatment process overall, acquired images are still prone to distortions because of organ movements and equipment errors [2]. Moreover, the analysis of images of different modalities or serial images is complicated as human tissues and organs are not static. Therefore, it was necessary to implement a solution to align such depicted objects in one coordinate space.

Two or more images are aligned in the process of registration. Medical image registration is a process of combining and aligning several images, for example, scans of the same person taken from different modalities or at different times, in one coordinate system [2]. Clinical applications of medical image registration include diagnosis, post-operative assessment, monitoring and image-guided treatment delivery. Moreover, image registration can be performed at the preprocessing stage to enhance further object detection, segmentation or classification. As manual image registration is time-consuming and susceptible to human errors, automated image registration has become widely addressed in recent years.

In the simplest case of aligning two images, image registration solutions consider one image fixed and another - moving. The moving image is warped to match the fixed one. The rules for warping the image are expressed in a form of a deformation field matrix. The core process of image registration is the search for an optimal deformation field.

Though traditional methods of automatic image registration can perform this task with good accuracy, they are still time-consuming. Implementing deep learning methods led to a significant increase in algorithmic performance for medical image registration [2]. However, many deep learning models are not derived from a probabilistic framework and therefore cannot estimate uncertainty. As machine learning models can be unstable in image processing tasks, uncertainty estimation can be used to quantify the risks of failures and prevent failures from happening. Uncertainty quantification measures the reliability of the output of the deep learning model [1]. Moreover, uncertainty quantification has also the potential to improve clinicians'trust.

The analysis of existing deep learning registration frameworks and models showed that there are models with good performance and high accuracy. However, the majority of them do not provide confidence and uncertainty data. This drawback limits their practical usage in critical fields.

This Master's thesis project provides an overview of uncertainty sources in medical images and estimation methods. Moreover, the uncertainty estimation methods were assessed from the point of suitability for image registration models. Uncertainty describes the level of confidence of a model in the predictions [4]. While is impossible to create a model which is absolutely confident, understanding the uncertainty allows to evaluate model performance better and determine if the model is overconfident in its predictions. For example, automated decision-making can be implemented in clinical practice as a process involving humans for regions where the model is uncertain. Such an approach would require two outputs, result and confidence value.

Deep learning models deal with two types of uncertainty, epistemic and aleatoric. Epistemic uncertainty described uncertainty of the model caused by a lack of training data. For example, the dataset with different training and test parts causes high epistemic uncertainty. Aleatoric uncertainty is caused by the uncertainty of the data, for example, due to noise. The sum of these two uncertainties is called predictive uncertainty [1].

Uncertainty quantification is a major challenge for deep learning. However, there are methods that show good performance: deep Bayesian neural networks, sparse Gaussian processes, Monte Carlo dropout, Deep Ensembles, and Deep Bayesian Active Learning [4]. Most of the methods require changing the model structure. Therefore, it was decided to apply ensemble methods to existing models

10th ACM Celebration of Women in Computing womENcourage, Trondheim, Norway,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

that showed good performance.

This Master's thesis project presents an implementation of two methods for estimating the uncertainty of existing models. Both methods are based on ensembles of models. The first method implements a bagging ensemble of deep-learning models, compares the predictions of the models and evaluates the disagreement between the models based on standard deviation. After that, the regions of the image with high disagreement values in several adjacent pixels are marked as uncertain and can be passed to a human expert for correction. It is possible to set a level of uncertainty and number of adjacent pixels to mark a region as uncertain.

The second method utilizes the anatomical segmentations of fixed and moving images. Similar to the first approach, a bagging ensemble of deep-learning models is trained on a dataset of fixed and moving images. After that, a test pair of images is registered with all models of the ensemble and resulting deformations are also applied to the segmentation of a moving image [3]. The deformed segmentation is compared to the fixed image segmentation. The averaged deformation vectors are considered as a final output. The disagreements between warped segmentations are used to calculate the confidence of the ensemble. Based on confidence data, final output and fixed image segmentation, the ensemble calibration is assessed [5]. The regions, where the ensemble shows under or overconfidence are passed to a human expert for correction.

The output of implemented methods is a pair of images: an aggregated output warped moving image and a mask for regions with high uncertainty. Both methods showed good performance and potential to contribute to implementing deep learning models into real clinical practice. Obtained datasets with warped images and uncertainty maps will be used in further assessment with expert radiologists.

REFERENCES

[1] Riccardo Barbano, Simon Arridge, Bangti Jin, and Ryutaro Tanno. 2022. Chapter 26 - Uncertainty quantification in medical image synthesis. In *Biomedical Image Synthesis and Simulation*, Ninon Burgos and David Svoboda (Eds.). Academic Press, 601–641. https://doi.org/10.1016/B978-0-12-824349-7.00033-5

[2] Xiang Chen, Andres Diaz-Pinto, Nishant Ravikumar, and Alejandro F Frangi. 2021. Deep learning in medical image registration. *Progress in Biomedical Engineering* 3, 1 (feb 2021), 012003. https://doi.org/10.1088/2516-1091/abd37c

[3] M.A. Ganaie, Minghui Hu, A.K. Malik, M. Tanveer, and P.N. Suganthan. 2022. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence* 115 (oct 2022), 105151. https://doi.org/10.1016/j.engappai.2022.105151

[4] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu. 2022. A Survey of Uncertainty in Deep Neural Networks. arXiv:2107.03342 [cs.LG]

[5] Lucy Nwosu, Xiangfang Li, Lijun Qian, Seungchan Kim, and Xishuang Dong. 2022. Calibrated bagging deep learning for image semantic segmentation: A case study on COVID-19 chest X-ray image. *PLOS ONE* 17 (11 2022), e0276250. https://doi.org/10.1371/journal.pone.0276250