Sub-second speed 4D-CT image registration using deep learning

Thomas van der Meulen[1], Oscar Pastor-Serrano[1], Steven Habraken[2], Zoltán Perkó[1]
[1]Delft University of Technology, Department of Radiation Science and Technology, Delft, The Netherlands
[2]Erasmus University Medical Center, Department of Radiotherapy, Rotterdam, The Netherlands

Fast image registration between 4D-CT lung phases

- Fast image registration is crucial in several radiotherapy tasks, and especially for treating moving targets.
- We train a model predicting the deformation vector field (DVF) $\vec{\phi}$ between fixed and moving phases of 4D-CT lung scans.
- Such model can be used for dose accumulation, contour propagation or future real-time adaptive treatments.

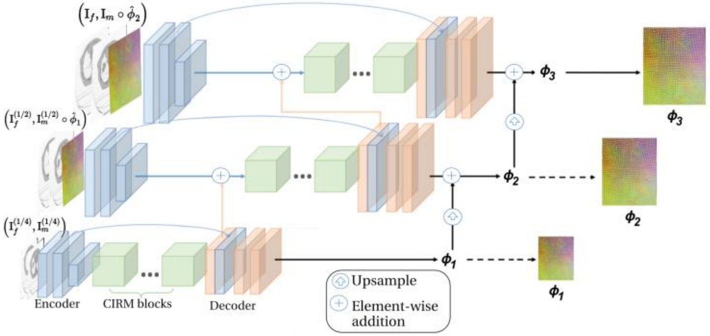

Figure 1. Schematic overview of the Laplacian Pyramid model [1]. Three identical sub-networks predict DVFs at increasingly finer grids. Stacking them yields superior accuracy.

DVF prediction by unsupervised Laplacian network

- An unsupervised Laplacian pyramid network with three sub-networks was trained [1]. The sub-networks predict DVFs at 1/4, 1/2 and full resolution respectively, in a coarse-to-fine order.
- The DVFs are up-sampled and elementwise combined to obtain the final DVF between the scans.
- Model is unsupervised as it trains without the ground truth DVFs, only by comparing the fixed and moving scans.
- The DVF was used for image deformation and contour propagation to test the trained model's speed and accuracy.

Patient data for training and testing

- The model was trained on 65 lung 4D-CTs and validated on another 7 4D-CTs. Each 4D-CT had 10 phases, with size 80x256x256 and 3 x 1.94 x 1.94 mm^3 resolution.
- Predictions were tested in 5 4D-CTs (with 10 phases each) not used in training [2].
- Target Registration Error $TRE = \left\|\vec{\phi}_i + \vec{x}_i^f - \vec{x}_i^m\right\|_2$ of 300 landmarks between 0% and 50% phases were determined; and mean absolute HU errors and grid folding between all phases were calculated. Values represent mean ± standard deviations.

Table 1. Average Target Registration Error In mm for 300 landmarks between 0 and 50% phase for five scans from the DIRLab dataset [2].

Scan	Original TRE in mm	Deformed TRE in mm
1	10.9 ± 7.0	2.9 ± 1.9
2	11.0 ± 7.4	2.2 ± 2.0
3	15.0 ± 9.0	3.1 ± 2.8
4	7.9 ± 4.0	2.5 ± 2.0
5	7.3 ± 6.3	2.8 ± 2.8
Mean	10.4 ± 6.7	2.7 ± 2.3

Results

- The DVF is predicted by the model in 24 ± 4 ms using a NVIDIA Tesla V100S GPU.
- The mean absolute HU error reduces from 23.4 ± 10.5 HU to 15.4 ± 5.4 HU, and the average TRE is reduced from 10.4 ± 6.7 mm to 2.7 ± 2.3 mm after deformation.
- Low number of voxels exhibit grid folding in the DVF. On average, 1.3 ± 6.5 voxels with a negative Jacobian determinant were seen.

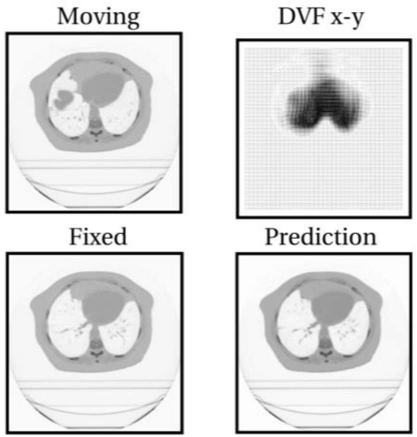

Figure 2: Visual comparison of moving, fixed, predicted and DVF in X-Y direction between the 0% to 50% phase.

Conclusions and further plans.

- The Laplacian pyramid network allows for fast and accurate prediction of the DVF between phases in around 20 milliseconds.
- TRE comparable with other deeplearning methods which ranged between 3.7 and 1.1 mm but have longer computational time from seconds to minutes.
- Future plans to use the DVF to perform interplay dose calculation and compare with dose distribution with the dose distribution obtained using a DVF from clinical registration software.

References

1. Tony C. W. Mok, Albert C. S. Chung, "Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks" (TMICCAI, 2020, arXiv:2006.16148.)
2. Castillo E, Castillo R, Martinez J, Shenoy M, Guerrero T., " Four-dimensional deformable image registration using trajectory modeling." (2009, Phys Med Biol 55 305-327.)