TOWARDS A ROBUST DEEP CT-ULTRASOUND REGISTRATION METHOD FOR IMAGE GUIDANCE IN LIVER TUMOR ABLATION

Sun Y.,[*†] Moelker A.,[†] Niessen W.J.,[*†‡] van Walsum T.[*†]

[*]Department of Medical Informatics, Erasmus MC, Rotterdam, the Netherlands,
[†]Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, the Netherlands,
[‡]Department of Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft, the Netherlands

Abstract

Multi-modal registration, especially CT/MR to ultrasound (US), is still a challenge, as conventional similarity metrics such as mutual information do not match the imaging characteristics of ultrasound. The main motivation for this work is to investigate whether a deep learning network (DVNet) can be used to directly estimate the displacement between a pair of multi-modal image patches, without explicitly performing similarity metric and optimizer, the two main components in a registration framework. The proposed DVNet is a fully convolutional neural network and is trained using a large set of artificially generated displacement vectors (DVs). The DVNet was evaluated on mono- and simulated multi-modal data, as well as real CT and US liver slices (selected from 3D volumes). The results show that the DVNet is quite robust on the single- and multi-modal (simulated) data, but does not work yet on the real CT and US images.

Figure 1: ***An example of a CT volume aligned in the US space*** *The yellow crossing lines locate a landmark annotated by clinical experts.*

Introduction

- Real-time overlay of CT in ultrasound (US) may improve image guidance in liver tumor ablation, as it allows the clinicians to accurately visualize the target
- Alignment of CT and Ultrasound (US) volumes is the prerequisite
- There is still a room to improve the state-of-the-art multi-modality metric for CT-US registration
- We proposed a block-matching based method before for the alignment of CT and US scans: first, estimating counterparts (point pairs) in ct and us images; second, point-based registration is performed to get the final transformation
- In this work, we focused on using deep learning methods to replace the block matching part - directly estimating the displacement between a pair of multi-modal image patches, without implementing an explicit multi-modality metric

Method

- Using a fully convolutional neural network: DVNet
- Loss function: Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y|$$

- Data preparation: align CT and US with landmarks (annotated by clinical experts); synthesizing displacement vectors (DVs)

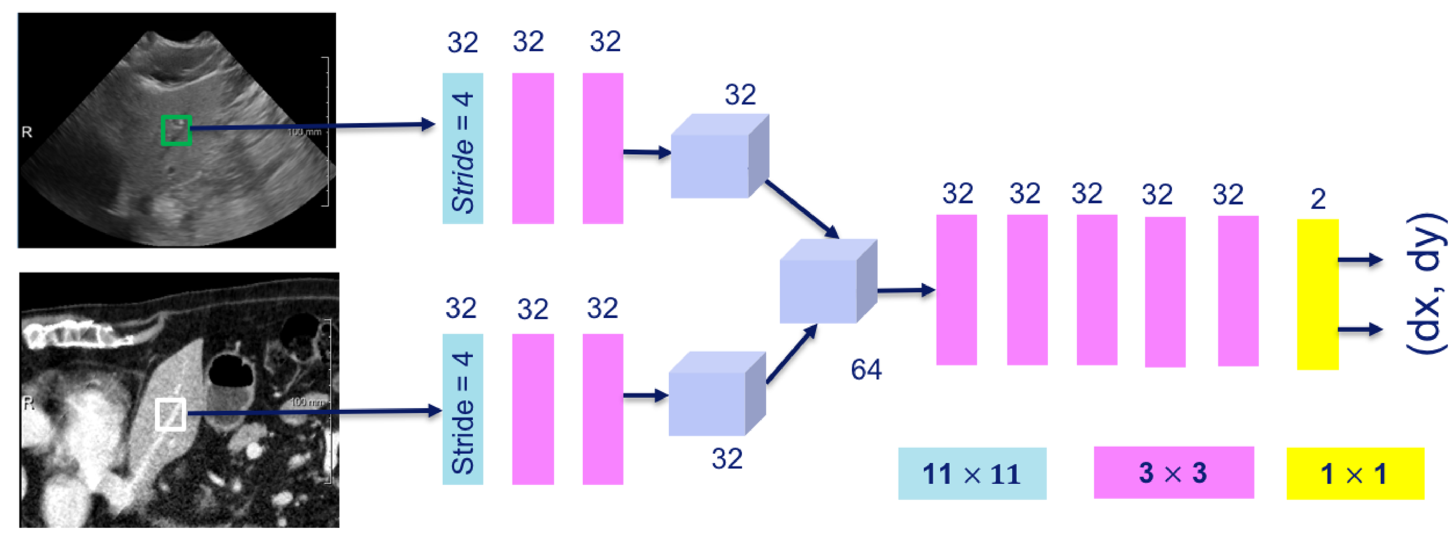

Figure 2: ***DVNet.*** *The architecture of the proposed DVNet. There are 101,026 parameters in total.*

Results

- Data: 120 / 60 slices from 30 pairs of CT and US volumes (9 more augmentation for each slice)
- Mono-modality (either CT or US): the final error on the test data is about 1 mm after 27k iterations
- Simulated multi-modality (CT - αCT + (1-α) ∇|CT|): the final error on the test data is about 1 mm after 55k iterations
- Inaccurate ground truth: We disturb the ground truth DVs with a random value between -4(-8) and 4(8). The results show that the DVNet could still learn the right displacements, for both single-modality and (simulated) multi-modality.
- Full CT-US registration: work-in-progress

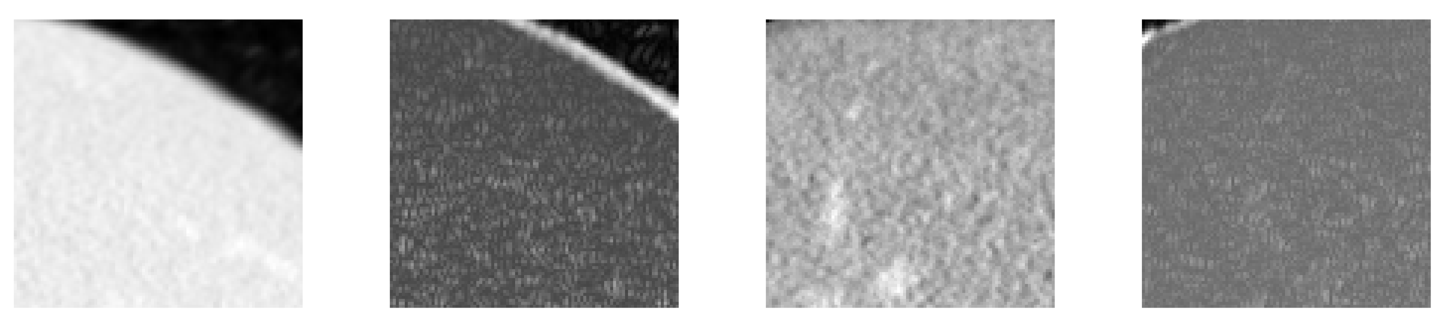

Figure 3: ***Examples of simulated multi-modality data.*** *The first and third column are the fixed CT patches and the other two columns are the moving simulated data.*

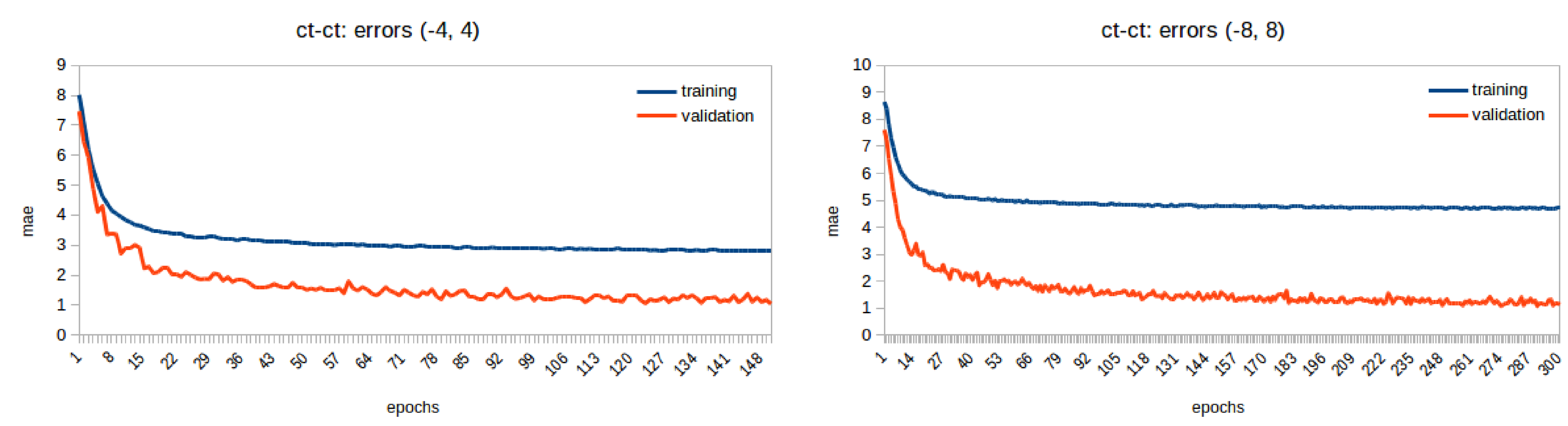

Figure 4: ***Loss plots with inaccurate ground truth.*** *The left plot is with disturbance (-4, 4) and the right one is with (-8, 8). The inaccurate ground truth affects the training loss (increased with the average value of d), but that the testing loss (on uncorrupted data) is only slightly larger than when training with accurate ground truth (about 1 pixel after 500 epochs).*

Discussion

- The current method is not feasible for the CT-US problem?

Contact: y.sun@erasmusmc.nl