# What is the better way to train deep learning registration network to compute cardiac motion field, supervised or unsupervised?

## Deep Learning for Cardiac Motion Estimation: Supervised vs. Unsupervised Training

Huaqi Qiu, Chen Qin, Loic Le Folgoc, Benjamin Hou, Jo Schlemper, Daniel Rueckert

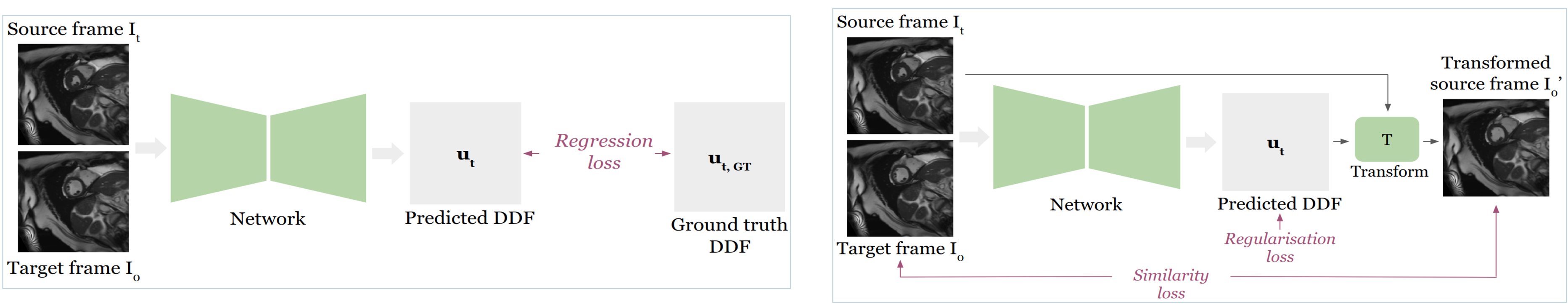
huaqi.qiu15@imperial.ac.uk

### INTRODUCTION

- Cardiac motion estimation tracks the movement of heart tissue points in time, which can be described by Dense Displacement Field (DDF)
- Computing DDF from cine MR sequence using traditional registration is **slow**
- **Fast** and more accurate deep learning registration has been proposed: using both *supervised* and *unsupervised* training[1], [2] (*supervised* = using ground truth DDF)
- **We compared the training strategies in the same task as fair as possible**

### METHODS

#### Frameworks




(a) Supervised training (b) Unsupervised training

Figure 1: Supervised and unsupervised training of deep learning image registration networks

#### Loss functions

*Supervised*

$$\mathcal{L}_{supervised} = \frac{1}{N_T}\sum_{t=1}^{N_T}\left(\frac{1}{|\Omega|}\sum_{\mathbf{p}\in\Omega}(\mathbf{u}_t(\mathbf{p}) - \mathbf{u}_{GT}(\mathbf{p}))^2\right)$$

*Unsupervised*

$$\mathcal{L}_{unsupervised} = \mathcal{L}_{MSE} + \lambda\mathcal{L}_{smooth}$$

$$\mathcal{L}_{MSE} = \frac{1}{N_T}\sum_{t=1}^{N_T}\left(\frac{1}{|\Omega|}\sum_{\mathbf{p}\in\Omega}(I'_0(\mathbf{p}) - I_0(\mathbf{p}))^2\right)$$

$$\mathcal{L}_{smooth} = \frac{1}{N_T}\sum_{t=1}^{N_T}\left(\frac{1}{|\Omega|}\sum_{\mathbf{p}\in\Omega}\sqrt{\left|\frac{\partial \mathbf{u}_t(\mathbf{p})}{\partial x}\right|^2 + \left|\frac{\partial \mathbf{u}_t(\mathbf{p})}{\partial y}\right|^2}\right)$$

#### Network architecture

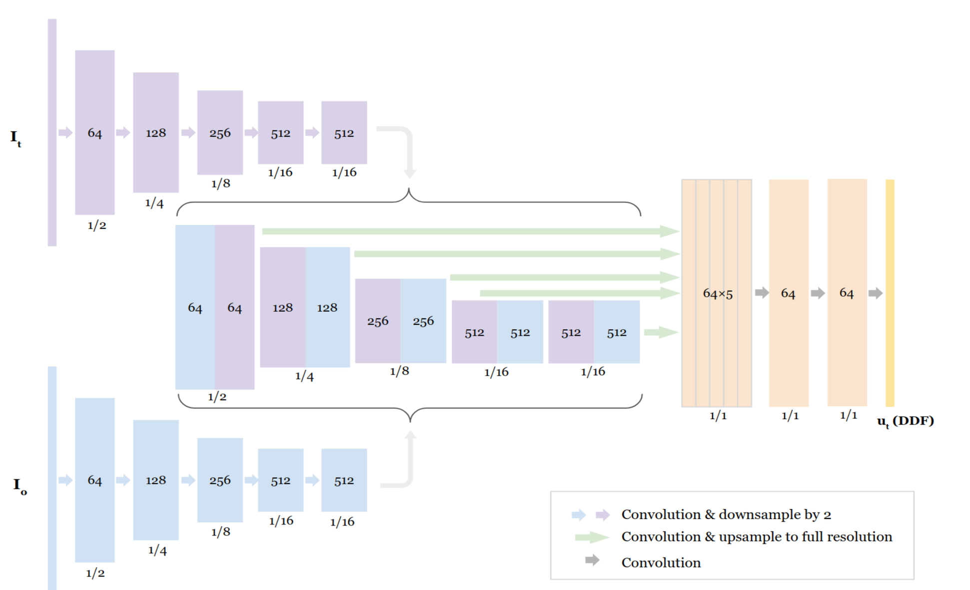


Figure 2: the dual-stream CNN-based registration network adapted from [4]

#### Generating data for supervised training

- The ground truth deformation was acquired using B-splines Free-Form Deformation (FFD) [3], a traditional registration method
- 2 input set-ups, "sup+orig." and "sup+warp." were used

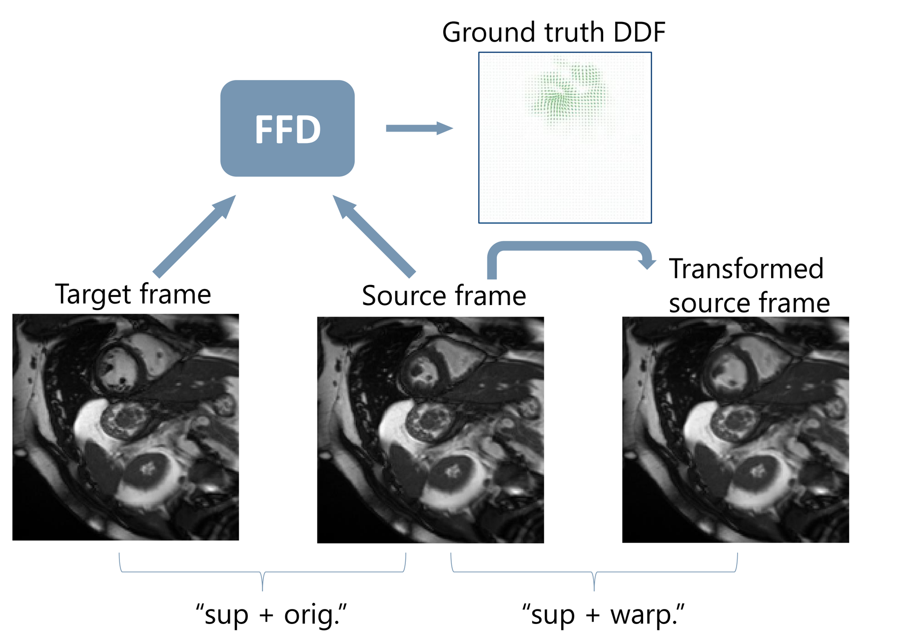



Figure 3: Illustration of how training data was generated for supervised training

### RESULTS

- Accuracy was measured by Dice score and Hausdorff Distance (HD) between the segmentation of the end-systolic (ES) frame, deformed using the estimated motion, and the segmentation of the end-diastolic (ED) frame
- Regularity is measured both in smoothness ($|\nabla|J_\phi||$) and topology preservation ($\%|J| \leq 0$) using *the Jacobian* $J_\phi(\mathbf{p}) = \nabla\phi(\mathbf{p})$

Table 1: Quantitative results - the accuracy and regularity of the estimated cardiac motion between the ED and ES frame

| Method | Dice | | | HD | | | $\lvert\nabla\lvert J_\phi\rvert\rvert$ | $\%\lvert J\rvert \leq 0$ |
|---|---|---|---|---|---|---|---|---|
| | LV | Myo | RV | LV | Myo | RV | | |
| Unreg | 0.641(0.058) | 0.322(0.086) | 0.551(0.077) | 11.40(1.40) | 8.90(2.00) | 11.50(1.90) | - | - |
| FFD | **0.941(0.049)** | **0.754(0.084)** | 0.671(0.109) | 4.52(2.33) | **4.73(1.44)** | 8.93(2.16) | 0.021(0.023) | 0.081(0.119) |
| DL(unsup) | **0.943(0.046)** | 0.740(0.077) | **0.709(0.087)** | **4.05(1.47)** | **4.62(1.25)** | 9.34(2.13) | 0.047(0.014) | 0.375(0.162) |
| DL(sup+warp.) | 0.920(0.049) | 0.735(0.080) | 0.668(0.101) | 4.61(1.11) | **4.84(1.48)** | 9.06(1.91) | 0.040(0.007) | **0.025(0.038)** |
| DL(sup+orig.) | 0.926(0.048) | 0.702(0.0801) | 0.657(0.089) | 4.41(1.35) | 5.29(1.22) | 9.22(1.88) | **0.019(0.004)** | **0.030(0.040)** |

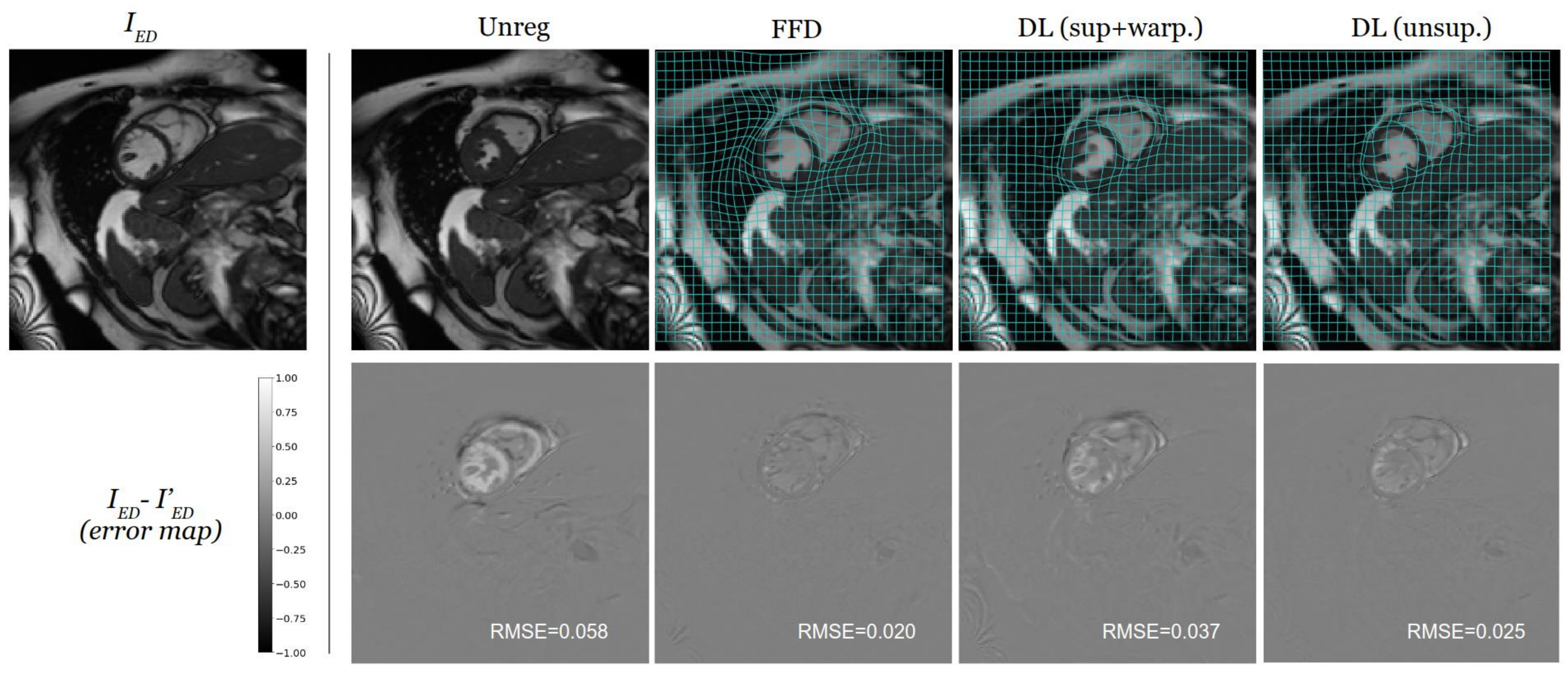



Figure 4: Visualisation of the cardiac motion estimation showing the ES frame image deformed under estimated motion (overlaid by meshgrids under the same deformation) and error between the deformed ES frame and original ED frame. ("Unreg" = unregistered)

### CONCLUSION & DISCUSSION

- Unsupervised training network results in more accurate motion estimation between the ED frame and the ES frame
- Supervised training network produces more *regular* deformation, however underestimating the motion

- **More extensive study (e.g. various networks and supervised data generation) is needed to reach more generalised conclusions**
- **Better evaluation is needed for cardiac motion estimation and deformation in general**

### REFERENCES

[1] Balakrishnan, G. et al., "VoxelMorph: a learning framework for deformable medical image registration." IEEE transactions on medical imaging 2019.
[2] Cao, X. et al. "Deformable image registration based on similarity-steered CNN regression." MICCAI 2017.
[3] Rueckert, D. et al., "Nonrigid registration using free-form deformations: application to breast MR images." IEEE transactions on medical imaging 1999.
[4] Qin, C. et al., "Joint learning of motion estimation and segmentation for cardiac MR image sequences." MICCAI 2018.



Full paper & poster