Spatial Warping Network for 3D Segmentation of the Hippocampus in MR Images

N K Dinsdale[1], M Jenkinson[1], A IL Namburete[2]
1. Wellcome center for Integrative Neuroimaging (WIN), FMRIB, University of Oxford 2. Institute of Biomedical Engineering, University of Oxford
nicola.dinsdale@dtc.ox.ac.uk

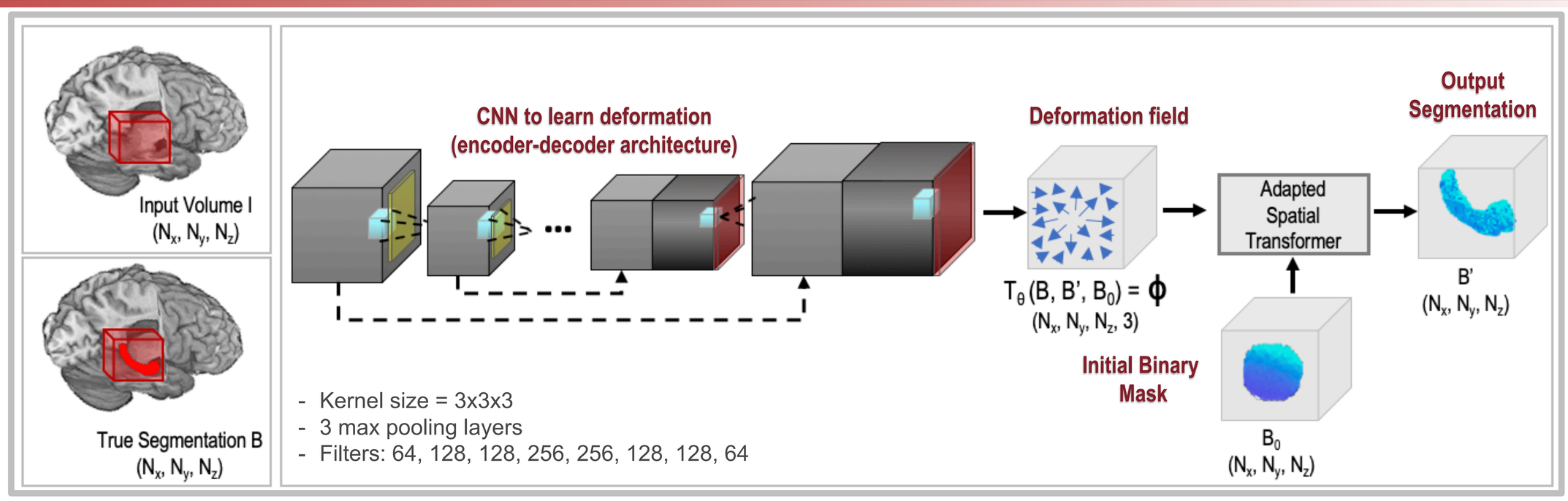
Motivation and Background

- Accurate **segmentation of the hippocampus** is vital to understand morphological changes, for instance in Alzheimer's disease.
- Extensive work has been completed using Convolutional Neural Networks (CNNs) to segment medical data, the most notable success being the U-Net architecture [1].
- **Spatial Transformer Networks** (STNs) were introduced by Jaderberg *et al.* [2] and can be used as a block within CNNs. As STNs are differentiable modules, they can be placed within the networks and trained using standard backpropagation methods.
- **In this work, the Spatial Warping Network for Segmentation (SWANS) is trained to segment the hippocampus from T1-weighted MRI scans by learning to deform an initial binary mask to produce the final segmentation.**

Method

- The network is formed of a CNN, which learns a deformation, and an **adapted spatial transformer**, which **applies the deformation to the initial binary mask to produce the output segmentation**. The deformation is applied via voxelwise multiplication such that a point (x, y, z) is mapped to (x', y', z') as (x', y', z') = (x, y, z) ◦ (a, b, c) = (xa, yb, zc).
- Loss function: $loss = \underbrace{Dice(y_{true}, y_{pred})}_{\text{Global}} + \propto \underbrace{BinaryCrossEntropy(y_{true}, y_{pred})}_{\text{Local}}$
- The network was trained on the **HarP dataset** [3] consisting of 100 adult brain volumes registered to the MNI standard space, across three different disease states: cognitively normal (CN), mild cognitive impairment (MCI) and Alzheimer's disease (AD). There was a further independent testing set of 31 volumes. A region of interest of dimensions (32, 64, 64) was taken around the hippocampus as the input volume to the network.

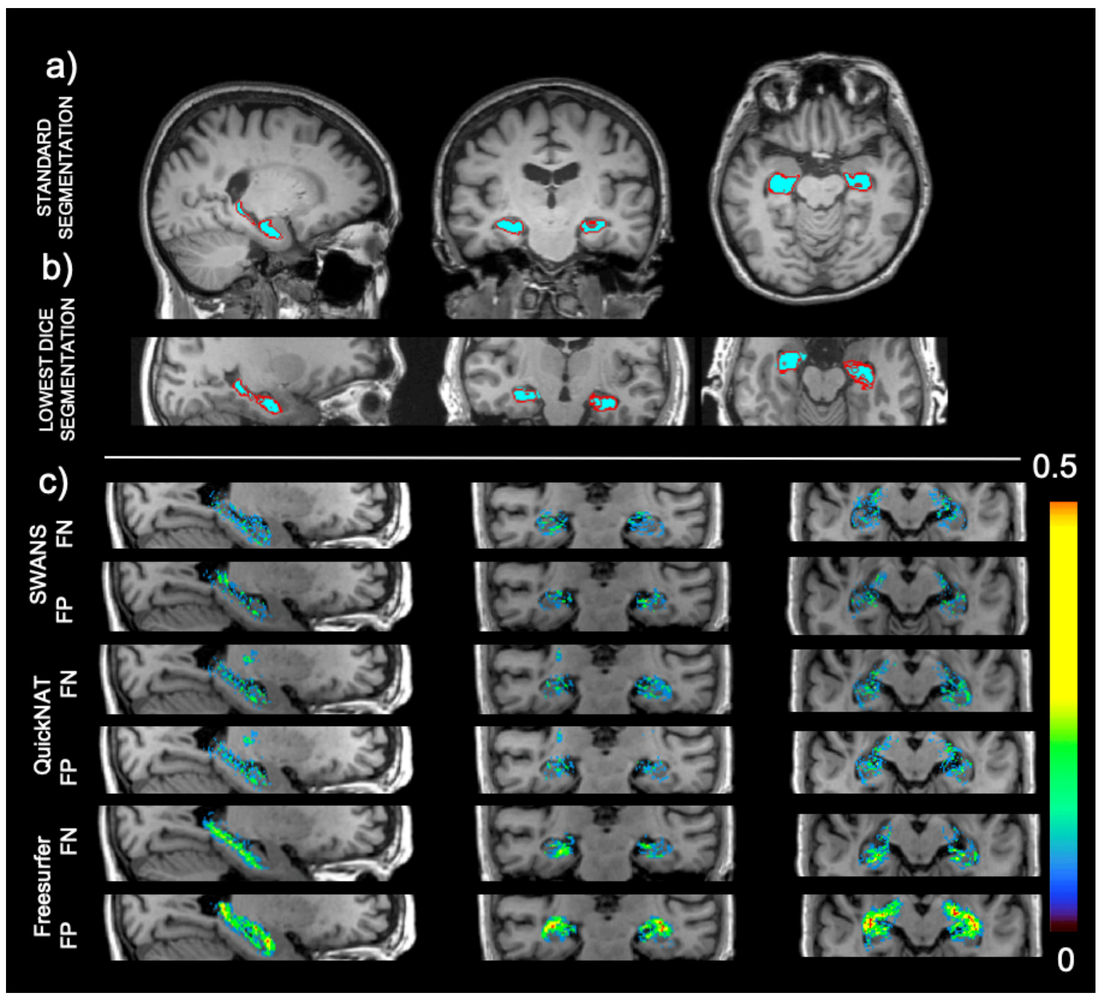
Network Design



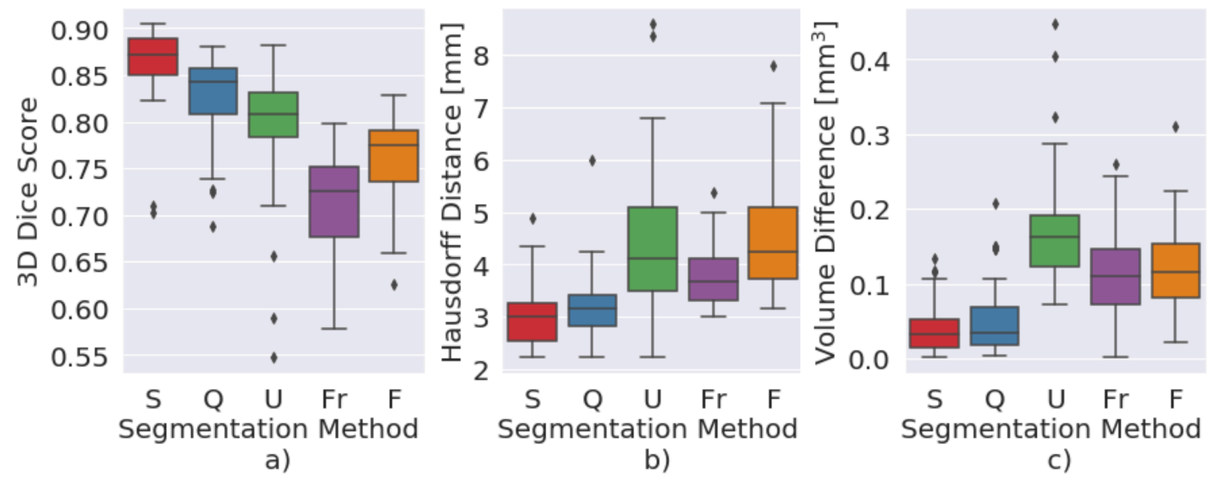
Segmentation Results

- The network performance for hippocampal segmentation was compared to four existing methods:
  - Two model based methods: FSL FIRST [4] and Freesurfer [5]
  - Two CNNs: U-Net [1] and QuickNAT [6]

Method	CN		MCI		AD	
	L	R	L	R	L	R
FIRST [4]	0.782 ± 0.03	0.779 ± 0.03	0.747 ± 0.03	0.834 ± 0.02	0.778 ± 0.04	0.780 ± 0.04
Freesurfer [5]	0.716 ± 0.04	0.709 ± 0.04	0.713 ± 0.06	0.699 ± 0.04	0.723 ± 0.05	0.720 ± 0.03
U-Net [1]	0.787 ± 0.03	0.799 ± 0.04	0.775 ± 0.08	0.825 ± 0.04	0.778 ± 0.06	0.826 ± 0.04
QuickNAT [6]	0.840 ± 0.02	0.831 ± 0.04	0.816 ± 0.06	0.834 ± 0.02	0.843 ± 0.03	0.853 ± 0.02
SWANS	**0.845 ± 0.06**	**0.868 ± 0.03***	**0.872 ± 0.02***	**0.873 ± 0.02***	**0.862 ± 0.03***	**0.880 ± 0.02***

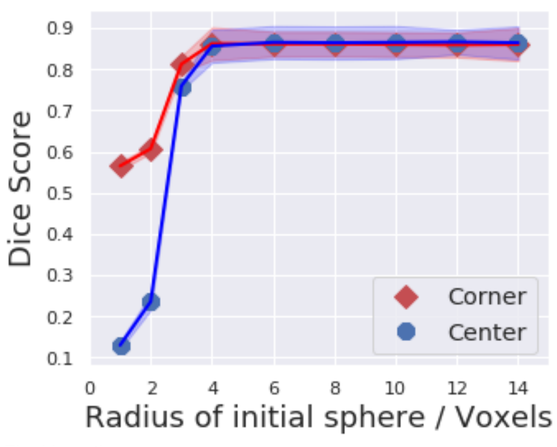


a) Representative output segmentation: blue shows the manual segmentation and the red outline is the segmentation produced by SWANS. b) Lowest Dice score segmentation. c) Error maps comparing SWANS, QuickNAT and Freesurfer. FP = False Positive, FN = False negative. **It can be seen that SWANS has the smallest error in all categories and is not systematically over-or-underestimating the segmentation.**

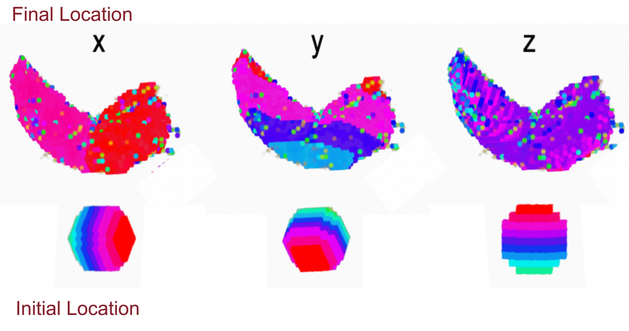


The methods were compared for three metrics to give an overall view of the segmentation quality (above). a) The Dice score gives an understanding of the degree of overlap between the segmentation and the true mask, whereas the b) Hausdorff Distance and c) Volume Difference give an indication of the extent of the outliers. It can be seen that SWANS performed consistently well across the metrics and **SWANS is consistently outperforming the other methods across the metrics.** S = SWANS, Q=QuickNAT, U = U-Net, Fr = Freesurfer, F = FSL FIRST.

Effect of Initial Binary Mask



- Finally, it was important to understand the effect of the initial binary mask on the output segmentation.
- From a) it can be seen that once the initial binary mask reached a radius of 4 voxels (approx. 1% of the volume of the hippocampus) **the size and location of the initial binary mask had no effect on the output segmentation**.
- b) shows how the initial binary mask, bottom row, was resampled in each axis – a colour region in the initial mask was resampled to produce the hippocampus in the corresponding colour, top row. It can be seen that the initial binary mask is resampled in a methodical manner and that there is significant redundancy with regions of the sphere not contributing to the final segmentation.



References
[1] Ronneberger et al (2015)
[2] Jaderberg et al (2015)
[3] Boccardi et al (2015)
[4] Patenaude et al (2011)
[5] Fischl et al (2002)
[6] Roy et al (2019)

This work was supported by funding from the Engineering and Physical Sciences Research Council (EPSRC) and Medical Research Council (MRC) [grant number EP/L016052/1]. AN is grateful for support from the UK Royal Academy of Engineering under the Engineering for Development Research Fellow- ships scheme. MJ is supported by the National Institute for Health Research (NIHR) and the Oxford Biomedical Research Centre (BRC).

EPSRC-MRC Centre for Doctoral Training in Biomedical Imaging