



LSTM Spatial Co-transformer Networks for Registration of 3D Fetal US and MR Brain Images

Robert Wright¹(✉), Bishesh Khanal¹, Alberto Gomez¹, Emily Skelton¹,
Jacqueline Matthew¹, Jo V. Hajnal¹, Daniel Rueckert², and Julia A. Schnabel¹

¹ School of Biomedical Engineering and Imaging Sciences,
King's College London, London, UK
robert.wright@kcl.ac.uk

² Department of Computing, Imperial College London, London, UK

Abstract. In this work, we propose a deep learning-based method for iterative registration of fetal brain images acquired by ultrasound and magnetic resonance, inspired by “Spatial Transformer Networks”. Images are co-aligned to a dual modality spatio-temporal atlas, where computational image analysis may be performed in the future. Our results show better alignment accuracy compared to “Self-Similarity Context descriptors”, a state-of-the-art method developed for multi-modal image registration. Furthermore, our method is robust and able to register highly misaligned images, with any initial orientation, where similarity-based methods typically fail.

1 Introduction

Registration, the process of aligning images, is an important technique which allows visual inspection and computational analysis of images in a common coordinate system. For fetal abnormality screening, registered Magnetic Resonance (MR)/Ultrasound (US) images may assist diagnosis as the two modalities capture complementary anatomical information. For example, in the fetal brain, MR images have better contrast between important structures such as cortical Grey Matter (GM) and White Matter (WM), whereas the higher spatial resolutions of US gives better discrimination between fine structures such as the septum pellucidum and the choroid plexus [7].

A voxel-wise image similarity measure or cost function is commonly used in medical imaging to register images. This function quantifies the alignment of images, where an extremum gives the optimum alignment between images. Unfortunately, image similarity-based methods are ill-suited to the challenging task of US/MR image registration as there is no global intensity relationship between the two modalities. Primarily this is due to the imaging artefacts present in US images, such as view-dependent shadows, speckle noise, anisotropy, attenuation, reverberation and refraction. Popular similarity measures developed specifically for other multi-modal registration problems in the past, such

as Normalised Mutual Information (NMI), often fail, even with a good initialisation [12].

Consequently, an alternative approach has arisen for registration of images with non-global intensity relationships whereby image intensities are first transformed to a modality independent representation. These are typically derived from hand-crafted descriptors which capture structural information from images such as edges and corners. Representations used by previous authors include local gradient orientation [4], local phase [9] and local entropy [16]. Notably, [5] use the concept of self-similarity, computing the similarity of small image patches in a local neighbourhood within an image, which achieved state-of-the-art performance on a challenging US/MRI registration dataset. Another approach to this problem is modality synthesis, which aims to transform image intensities from one modality to another allowing the registration task to be treated as a mono-modal problem. [7] made use of this approach to register the fetal brain imaged by US and MR for the first time.

More recently, deep neural networks have been applied to the problem of registration. Two common strategies for registration with deep learning include estimating a similarity measure [2, 15] and predicting transformations directly [1, 13]. An advantage of the first approach is that it allows established transformation models and optimizers to be used, however, this could be a hindrance if the learnt similarity function is not smooth or convex. The second approach, predicting the parameters of a transformation model directly, is receiving more research focus recently as it allows more robust transformation estimates.

1.1 Proposed Method

In this work, we adopt a deep learning approach to tackle the challenging task of paired 3D MR/US fetal brain registration. Our Long Short-Term Memory (LSTM) network simultaneously predicts a joint isotropic rescaling plus independent rigid transformations for both MR/US images, aligning them to a dual-modality spatio-temporal atlas (Sect. 2.6). Transformation estimates are refined iteratively over time, allowing for higher accuracy. For this, we extend the iterative spatial transformer [8] for co-transformation of multiple images (see Fig. 1). The main contributions of this work are as follows:

- A network architecture inspired by spatial transformer networks [6] for group-wise registration of images to a common pose.
- A loss function which encourages convergence and fine alignment of images.

2 Methods

2.1 Overview

The spatial transformer module [6] allows geometric transformation of network inputs or feature maps within a network, conditioned on the input or feature

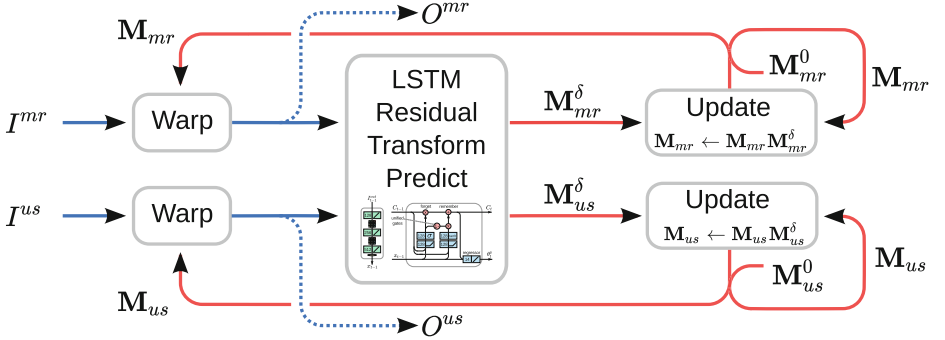


Fig. 1. Proposed LSTM spatial co-transformer for coalignment of 3D MR/US images. Flow of image intensities is shown in blue while flow of transformation parameters is shown in red. An LSTM network predicts residual transformations \mathbf{M}_{mr}^{δ} , \mathbf{M}_{us}^{δ} conditioned on the current warped images O^{us} , O^{mr} , iteratively refining their alignment. (Color figure online)

map itself. Importantly, the spatial transformer module is differentiable, allowing end-to-end training of any network it is inserted into. This allows reorientation of an image into a canonical pose, simplifying the task of subsequent layers. [8] proposed an elegant iterative version of the spatial transformer that passes composed transformation parameters through the network instead of warped images, preserving image intensities until the final transformation. The same geometric predictor with a much simpler network architecture can be used in a recurrent manner, for more accurate alignment.

In this work, we propose a novel extension the “recurrent/LSTM spatial co-transformer”, which allows simultaneous transformation of multiple images to a common pose. Commonly, registration algorithms estimate a warp from one image (the source) towards another (the target). However, we found that fine alignment is more easily learnt between images in a common pose. Thus, we simultaneously co-align pairs of MR/US images to a common atlas-space (Sect. 2.6), which will also facilitate future computational image analysis.

Additionally, we propose an LSTM-based parameter prediction network (Fig. 2) and a temporally varying loss function (Sect. 2.5) for more accurate alignments.

2.2 Recurrent Spatial Co-transformer

The recurrent spatial co-transformer consists of three main components: (1) the warper, (2) the residual parameter prediction network and (3) the composer. The first component, the warper, is the computational machinery needed to transform an image and does not contain any learnable parameters. For simplicity of discourse, we treat this as a single function f_{warp} and refer the reader to [6] for a detailed description of grid transformation and differentiable interpolation. The second component, the parameter prediction network, $f_{predict}$, predicts

residual transformations conditioned on the current warped output images. Finally, the third component, the composer, updates the transformation estimates. The recurrent spatial co-transformer iterates between three steps, which will now be described in more detail.

Step 1 - Image Warping. For iteration t , Let $\mathcal{I} = (I^0, I^1, \dots, I^N)$ denote an N -tuple of input images, $\Theta_t = (\theta_t^0, \theta_t^1, \dots, \theta_t^N)$ denote an N -tuple of corresponding transformation estimates and $\mathcal{O}_t = (O_t^0, O_t^1, \dots, O_t^N)$ denote an N -tuple of corresponding warped output images. Then each input image I^i is first warped independently given its last transformation estimate θ_{t-1}^i

$$O_{t-1}^i = f_{warp}(I^i, \mathbf{G}, \theta_{t-1}^i) \quad \forall i \in [1, \dots, N]. \quad (1)$$

Here, $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_g] \in \mathbb{R}^{4 \times g}$ is a matrix of homogeneous grid coordinates.

Step 2 - Residual Parameter Prediction. Warped images \mathcal{O}_{t-1} are concatenated along the channel axis and passed as a single tensor to $f_{predict}$ which simultaneously predicts an N -tuple of corresponding residual transformations $\Delta_t = (\delta_t^0, \delta_t^1, \dots, \delta_t^N)$

$$\Delta_t = f_{predict}(O_{t-1}^0 \wedge O_{t-1}^1 \wedge \dots \wedge O_{t-1}^N). \quad (2)$$

$f_{predict}$ can take any form but typically consists of a feed-forward network with several interleaved convolutional and max pooling layers followed by a fully connected layer and a final fully connected regression layer with the number of units equalling the number of model parameters.

Step 3 - Parameter Composition. Finally, each transformation estimate θ_{t-1}^i , is composed with its residual transformation estimate δ_t^i , yielding a new transformation estimate θ_t^i

$$\theta_t^i = f_{update}(\theta_{t-1}^i, \delta_t^i) \quad \forall i \in [1, \dots, N]. \quad (3)$$

The composition function f_{update} will vary depending on the transformation model. For example, if θ parametrises a homogeneous transformation matrix, f_{update} would be matrix multiplication.

2.3 LSTM Spatial Co-transformer

For more accurate parameter prediction, we propose an LSTM network architecture for $f_{predict}$. LSTMs are an extremely powerful network architecture capable of storing information in a cell state allowing them to learn long term dependencies in sequential data much better than recurrent neural networks. For this we modify the prediction function $f_{predict}$ (Eq. 2) so that it now takes a feature vector \mathbf{x}_t , and a cell state vector \mathbf{c}_t

$$\Delta_t = f_{predict}(\mathbf{x}_t, \mathbf{c}_t), \text{ where } \mathbf{x}_t = f_{extract}(O_{t-1}^0 \wedge O_{t-1}^1 \wedge \dots \wedge O_{t-1}^N). \quad (4)$$

Here $f_{extract}$ is a function that extracts the feature vector \mathbf{x}_t from the concatenation of the output images, $O_{t-1}^0 \wedge O_{t-1}^1 \wedge \dots \wedge O_{t-1}^N$. For this, we chose a

neural network with a series of convolutions and max pooling operations followed by a flattening procedure (see Fig. 2 for a schematic, however any network architecture may be used that produces a vector). At each iteration t , the cell state \mathbf{c}_t is updated by a linear blend of the previous cell state \mathbf{c}_{t-1} and a vector of candidate values $\tilde{\mathbf{c}}_t$ [3]

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + (1 - \mathbf{f}_t) \odot \tilde{\mathbf{c}}_t. \quad (5)$$

Here, \odot is the Hadamard or element-wise product and \mathbf{f}_t is the forget mask, a real valued vector that determines which information is forgotten from the cell state and which candidate values are added. We define \mathbf{f}_t as the result of a single function f_{forget} that takes the extracted feature vector \mathbf{x}_t and also the previous cell state \mathbf{c}_{t-1} . We implement both the forget and candidate functions as a sequence of two dense layers with weight matrices \mathbf{W}_{f1} , \mathbf{W}_{f2} and \mathbf{W}_{c1} , \mathbf{W}_{c2} , respectively

$$\mathbf{f}_t = f_{forget}(\mathbf{c}_{t-1}, \mathbf{x}_t) = \sigma(\mathbf{W}_{f2} \cdot \max(\mathbf{W}_{f1} \cdot [\mathbf{c}_{t-1}, \mathbf{x}_t], 0)), \quad (6)$$

$$\tilde{\mathbf{c}}_t = f_{candidate}(\mathbf{c}_{t-1}, \mathbf{x}_t) = \tanh(\mathbf{W}_{c2} \cdot \max(\mathbf{W}_{c1} \cdot [\mathbf{c}_{t-1}, \mathbf{x}_t], 0)). \quad (7)$$

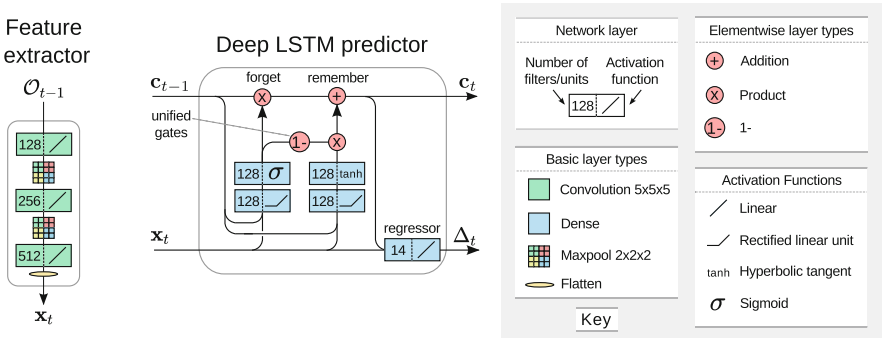


Fig. 2. LSTM parameter prediction architecture for rigid alignment of MR/US images. The image feature extractor encodes a dual-channel image as a vector that is passed into an LSTM network which predicts a residual transformation. Fourteen parameters are predicted: three for rotation, three for translation and one for isotropic scale, per modality (note, weights for scaling are shared between modalities).

2.4 Rigid Parameter Prediction

For rigid coalignment, our network predicts seven residual update parameters per image: an isotropic log scaling s , three rotation parameters r_x , r_y , r_z and three translation parameters t_x , t_y , t_z . Here, $[r_x, r_y, r_z]$ gives an axis of rotation, while $\phi = \|[r_x, r_y, r_z]\|_2$, gives the angle of rotation. Note, weights are shared between images for scaling parameters. Our transformation parameters

now become rigid transformation matrices $\delta_t = \mathbf{M}_t^\delta$, $\theta_t = \mathbf{M}_t$. Note, for simplicity, transformations \mathbf{M} are applied to the target grid \mathbf{G} before resampling, i.e. the inverse transformation. For consistency, we define \mathbf{M}^δ as the inverse update and $(\mathbf{M}^\delta)^{-1}$ as the forward update. Learning a series of forward update transformations is inherently easier for the network, thus we post-multiply the current transformation matrix by the residual matrix, $\mathbf{M} \leftarrow \mathbf{M}\mathbf{M}^\delta$. This is equivalent to updating the forward transformation as follows $\mathbf{M}^{-1} \leftarrow (\mathbf{M}^\delta)^{-1}\mathbf{M}^{-1}$. The forward update transformation is composed as a translation, followed by a rotation, followed by an isotropic rescaling, $(\mathbf{M}^\delta)^{-1} = \mathbf{SRT}$. In practice, we predict the inverse of the update directly by reversing the composition and inverting the operations $\mathbf{M}^\delta = \mathbf{T}^{-1}\mathbf{R}^{-1}\mathbf{S}^{-1}$.

2.5 Training and Loss Function

Let $\mathbf{X} = \{\mathcal{I}_0, \mathcal{I}_1, \dots, \mathcal{I}_n\}$ denote a training set of n aligned image tuples. Images in the training set are initially aligned to a common pose (in our case we affinely align our MR and US images to a dual-modality atlas, see Sect. 2.6). For each training iteration, an image tuple is selected $\mathcal{I} = (I^0, I^1, \dots, I^N)$ and each image I^i is transformed by a randomly generated matrix \mathbf{D}^i , before being fed into the network. \mathbf{D}^i incorporates an affine augmentation (shared across the input tuple) and an initial rigid disorientation. For augmentation, we randomly sample and compose a shearing, an anisotropic scaling and an isotropic scaling. For disorientation, we compose a random rotation and translation. Crucially, the use of a recurrent network allows us to back-propagate errors through time. We took advantage of this by designing a temporally varying loss function comprising of a relative and an absolute term, which allows our network to learn a long term strategy for alignment. For k alignment iterations of N images, we define our loss

$$\mathcal{L} = \sum_{i=1}^N \sum_{t=1}^k d(\mathbf{M}_t^i \mathbf{D}^i) / d(\mathbf{M}_{t-1}^i \mathbf{D}^i) + \lambda \frac{t}{k} d(\mathbf{M}_t^i \mathbf{D}^i). \quad (8)$$

Here, d is a distance function of a transformation matrix from the identity and λ is a weighting between the loss terms. The first loss term rescales distance errors $d(\mathbf{M}_t^i \mathbf{D}^i)$, relative to the previous distance error, $d(\mathbf{M}_{t-1}^i \mathbf{D}^i)$. This encourages the network to learn fine alignments and convergence. Note, $d(\mathbf{M}_{t-1}^i \mathbf{D}^i)$ is treated as a constant here. The second term penalises the absolute error with increasing weight, encouraging initial exploration but still penalising poor final alignments. The distance function $d(\mathbf{M})$ is computed by first decomposing matrix \mathbf{M} into a isotropic scale s , a translation vector \mathbf{t} and a rotation matrix \mathbf{R} . We then compute $d(\mathbf{M})$ as a sum of separate distance measures for each of these components

$$d(\mathbf{M}) = d_{scale}(s) + d_{rotate}(\mathbf{R}) + d_{translate}(\mathbf{t}), \quad \text{where} \quad d_{translate}(\mathbf{t}) = \|\mathbf{t}\|_2, \\ d_{scale}(s) = \mu |\log(s)| \quad \text{and} \quad d_{rotate}(\mathbf{R}) = \frac{1}{g} \sum_{i=1}^g \|\mathbf{g}_i - \mathbf{R}\mathbf{g}_i\|_2. \quad (9)$$

Here, μ weights d_{scale} relative to the other two distance measures. Rotation distance, d_{rotate} , is given by the mean distance between transformed grid points $\mathbf{R}\mathbf{g}_i$ and their initial locations \mathbf{g}_i . This gives a natural weighting between translation and rotation components.

2.6 Joint Affine MR/US Spatio-Temporal Atlas (Ground Truth)

We followed the approach of [14], by constructing average image intensity templates for each week of gestation (20–31 weeks), from 166 3D reconstructed MR/3D US image pairs. A set of templates was constructed for each modality separately with a final registration step between templates to establish correspondences across modalities. This process comprised of three parts: (1) manual reorientation (2) age-dependant template bootstrapping and (3) unified template bootstrapping. All images were carefully manually reoriented to a standard pose with the yz plane aligned with the brain midline and the top of the brain stem centred at the origin. Averaging reoriented image intensities yielded an initial template estimate which was refined using a bootstrapping procedure. This involved alternating between two steps: (1) affinely registering images to the current template and (2) averaging registered image intensities. The bootstrapping procedure was then repeated between templates to establish correspondences across time. MR templates were constructed first, allowing us to fix the shearing and scaling parameters for US template construction. For US registration, we restrict the optimisation to three degrees of freedom, rotation around x, and translation along y and z, thus respecting the manual definition of the midline. With additional masking, this allowed robust registration of US images for template construction using [10].

3 Results and Discussion

3.1 Alignment Error

To demonstrate the accuracy of our method (LSTM ST) we compute registration errors with respect to two ground truth alignments: the first, derived from our spatio-temporal atlas and the second, derived from anatomical landmarks picked by clinical experts (fourteen per image), which offers an unbiased alternative. For comparison, two image similarity-based registration methods were chosen, NMI with block-matching (NMI+MI) [10] and self-similarity context descriptors with discrete optimisation (SSC+DO) [5]. Both of these methods were developed for robust registration and have been used for multi-modality registration tasks previously. To compare the accuracy of the methods and also their ability to register highly misaligned images, we created three test sets with different ranges of disorientation: $[3-5^\circ, 3-5\text{ mm}]$, $[30-60^\circ, 10-20\text{ mm}]$ and $[90-180^\circ, 30-50\text{ mm}]$.

As we can see from Table 1 our method outperforms both similarity-based methods for all disorientation levels and both ground truth datasets. Furthermore, our method converges to the same alignment for each image pair, irrespective of initial orientation and positioning, which explains the very similar

mean errors seen for the three disorientation levels. Conversely, similarity-based methods failed to register images for higher levels of disorientation. All pairs of images registered by our method were visually inspected and a reasonable alignment was found in all cases (see Fig. 3 for example alignments). The worst

Table 1. Mean alignment error. Mean rotation and translation errors over our test set are shown for three automated registration methods, relative to two ground truth alignments.

(a) Atlas-based ground truth alignment							
Disorientation		NMI+BM		SSC+DO		LSTM ST	
3–5°	3–5mm	23.00°	3.49 mm	4.08°	1.01 mm	2.97°	0.63 mm
30–60°	10–20 mm	42.11°	5.26 mm	36.62°	3.97 mm	2.94°	0.63 mm
90–180°	30–50 mm	131.25°	9.18 mm	129.74°	13.01 mm	2.91°	0.62 mm
(b) Landmark-based ground truth alignment							
Disorientation		NMI+BM		SSC+DO		LSTM ST	
3–5°	3–5 mm	23.84°	3.57 mm	5.49°	1.73 mm	4.06°	1.60 mm
30–60°	10–20 mm	42.58°	5.22 mm	35.16°	4.11 mm	4.03°	1.60 mm
90–180°	30–50 mm	131.70°	8.98 mm	131.28°	11.68 mm	4.03°	1.60 mm

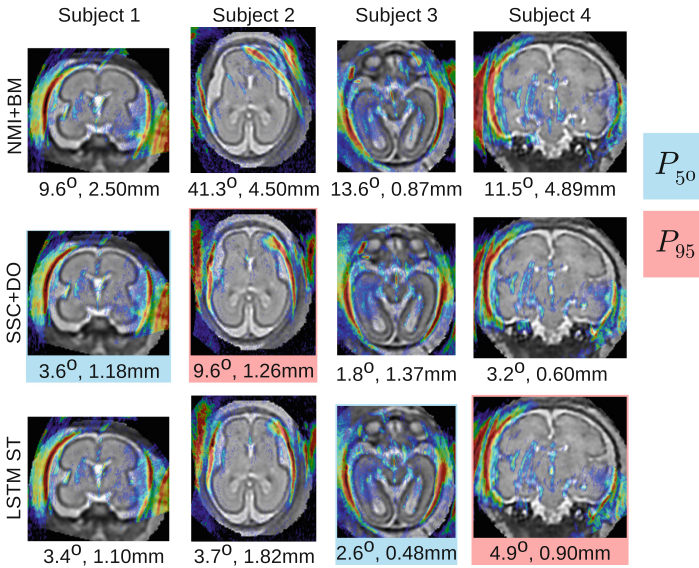


Fig. 3. Median (blue) and 95th percentile (red) alignments by rotation error for SSC+DO and LSTM ST. Alignments for other methods are shown for comparison. Each column shows the same MR image for a subject from our test set with its corresponding US image thresholded, colour-mapped, overlaid and aligned, by each of the automated methods. (Color figure online)

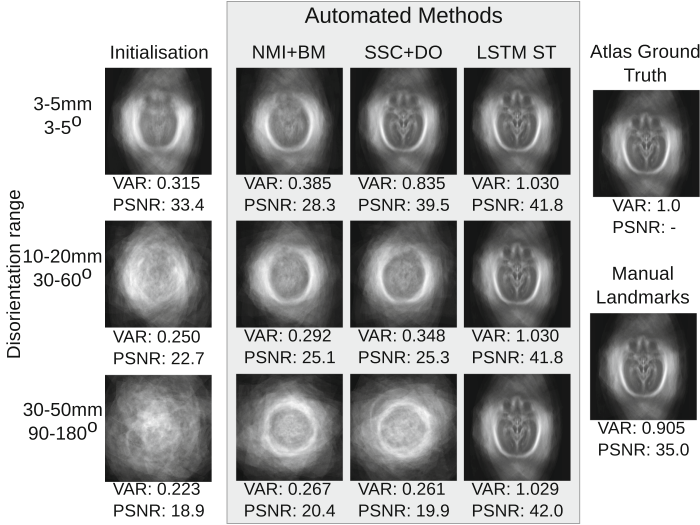


Fig. 4. Template sharpness. Templates are constructed by averaging image intensities for US images registered to an MR template via their corresponding MR images (see Sect. 3.2). Higher Variance of the Laplacian (VAR) indicates sharper templates and better registration accuracy, while higher Peak Signal-to-Noise Ratio (PSNR) indicates greater similarity with the atlas ground truth template.

rotation and translation errors seen were 7.9° and 1.8 mm respectively, showing our method is relatively robust.

3.2 Mean Templates

We construct US mean templates by first registering each US image to its corresponding MR image, rigidly, then affinely transforming the image pair to the MR atlas space and finally averaging the intensities for all transformed US images. If registration between modalities is accurate, then the constructed US template should be crisp. To evaluate the constructed templates, we compute two measures, Peak Signal-to-Noise Ratio (PSNR) with respect to our ground truth US template (Sect. 2.6) and the Variance of the image Laplacian (VAR), which provides an unbiased measure of sharpness [11]. Figure 4 shows that our method produces the sharpest template as measured by VAR and also has the highest PSNR. Furthermore, templates for our method have the same sharpness for any level of initial disorientation.

4 Conclusion

In this work, we proposed the LSTM spatial co-transformer, a deep learning-based method for group-wise registration of images to a standard pose.

We applied this to the challenging task of fetal MR/US brain image registration. Our method automatically coaligns brain images with a dual-modality spatio-temporal atlas, where future computational image analysis may be performed. Our results show that our method registers images more accurately than state-of-the-art similarity-based registration method “self-similarity context descriptors” [5]. Furthermore, it is able to robustly register highly misaligned images, where similarity-based will fail.

Acknowledgements. This work was supported by the Wellcome/EPSRC Centre for Medical Engineering [WT 203148/Z/16/Z], Wellcome Trust IEH Award [102431] and NVIDIA with the donation of a Titan Xp GPU.

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: CVPR (2018)
2. Cheng, X., Zhang, L., Zheng, Y.: Deep similarity learning for multimodal medical images. *Comput. Meth. Biomech. Biomed. Eng. Imaging Vis.* **6**(3), 248–252 (2018)
3. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: EMNLP, pp. 1724–1734 (2014)
4. Haber, E., Modersitzki, J.: Intensity gradient based registration and fusion of multi-modal images. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 726–733. Springer, Heidelberg (2006). https://doi.org/10.1007/11866763_89
5. Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, S.M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8149, pp. 187–194. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40811-3_24
6. Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K.: Spatial transformer networks. In: NIPS, pp. 2017–2025 (2015)
7. Kuklisova-Murgasova, M., et al.: Registration of 3D fetal brain US and MRI. In: MICCAI, pp. 667–674 (2012)
8. Lin, C.H., Lucey, S.: Inverse compositional spatial transformer networks. In: CVPR, pp. 2252–2260 (2017)
9. Mellor, M., Brady, M.: Phase mutual information as a similarity measure for registration. *Med. Image Anal.* **9**(4), 330–343 (2005). *Functional Imaging and Modeling of the Heart - FIMH 2003*
10. Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3D structure from serial histological sections. *Image Vis. Comput.* **19**(1), 25–31 (2001)
11. Pech-Pacheco, J.L., Cristobal, G., Chamorro-Martinez, J., Fernandez-Valdivia, J.: Diatom autofocusing in brightfield microscopy: a comparative study. *ICPR* **3**, 314–317 (2000)
12. Rivaz, H., Karimaghloo, Z., Collins, D.L.: Self-similarity weighted mutual information: a new nonrigid image registration metric. *Med. Image Anal.* **18**(2), 343–358 (2014)

13. Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-Net: learning deformable image registration using shape matching. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 266–274. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_31
14. Serag, A., Aljabar, P., Ball, G., Counsell, S.J., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive Kernel regression. *NeuroImage* **59**(3), 2255–2265 (2012)
15. Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., Komodakis, N.: A deep metric for multimodal registration. In: MICCAI, pp. 10–18 (2016)
16. Wachinger, C., Navab, N.: Entropy and Laplacian images: structural representations for multi-modal registration. *Med. Image Anal.* **16**(1), 1–17 (2012)