World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:14, No:8, 2020

NANCY: Combining Adversarial Networks with Cycle-Consistency for Robust Multi-Modal Image Registration

Mirjana Ruppel, Rajendra Persad, Amit Bahl, Sanja Dogramadzi, Chris Melhuish, Lyndon Smith

Abstract—Multimodal image registration is a profoundly complex task which is why deep learning has been used widely to address it in recent years. However, two main challenges remain: Firstly, the lack of ground truth data calls for an unsupervised learning approach, which leads to the second challenge of defining a feasible loss function that can compare two images of different modalities to judge their level of alignment. To avoid this issue altogether we implement a generative adversarial network consisting of two registration networks G_{AB}, G_{BA} and two discrimination networks D_A, D_B connected by spatial transformation layers. G_{AB} learns to generate a deformation field which registers an image of the modality B to an image of the modality A. To do that, it uses the feedback of the discriminator D_B which is learning to judge the quality of alignment of the registered image B. G_{BA} and D_A learn a mapping from modality A to modality B. Additionally, a cycle-consistency loss is implemented. For this, both registration networks are employed twice, therefore resulting in images $\hat{A}$, $\hat{B}$ which were registered to $\tilde{B}$, $\tilde{A}$ which were registered to the initial image pair A, B. Thus the resulting and initial images of the same modality can be easily compared. A dataset of liver CT and MRI was used to evaluate the quality of our approach and to compare it against learning and non-learning based registration algorithms. Our approach leads to dice scores of up to 0.80 ± 0.01 and is therefore comparable to and slightly more successful than algorithms like SimpleElastix and VoxelMorph.

Keywords—Multimodal image registration, GAN, cycle consistency, deep learning.

I. Introduction

In a medical context automatic image registration can drastically simplify the process of combining images that were obtained using different modalities such as CT, MRI or ultrasound (US). The goal of this project is to align pre-procedure MRI to intra-procedure US images via a multi-modal real-time image registration algorithm so surgeons get access to higher quality imaging during prostate brachytherapy and biopsy. However, the dataset of prostate MRI and US is still under construction which is why for the present contribution liver MRI and CT images from the CHAOS dataset [1] are used to test the application of the developed algorithm to multi-modal image registration in general. Since this task is very complex, Ferrante et al. [2] suggest that a promising and a still not fully explored approach to image registration is the use of machine learning. But as the complexity of multi-modal image registration lies in the difficulty of comparing two images that stem from modalities whose depiction of intensities cannot easily be compared analytically, it remains challenging to define a suitable similarity measure as a loss function for a machine learning approach that aims to register multi-modal images.

In [3] the focus lies on finding optimal similarity measures for comparing medical images that were acquired by different modalities. The method outperforms the mutual information similarity measure which is usually applied for multi-modal image registration [4], [5]. Nonetheless, with this approach it remains necessary to apply a computationally expensive optimizer for the calculation of a registration map. Several approaches exist to train deep CNNs to learn how to perform the registration process itself. In [6]- [9] unsupervised learning-based approaches consisting of CNNs were implemented for deformable 3D/3D image registration. They used normalized cross correlation (NCC) and sum of squared differences (SSD) as similarity measures which is sufficient for the registration of single modality images but the resulting errors after multi-modal registration remain considerably higher.

To combine the advantages of having an end-to-end approach and a learned similarity function and to overcome the difficulty of comparing multi-modal images, the ideas in [10] inspired the system depicted in Fig. 1. Two discriminators D_A and D_B learn a similarity measure between both modalities to rate the quality of alignment between an image pair. Two generators G_{AB} and G_{BA} learn to generate deformation fields that register image pairs. Each generator is employed twice to exploit the cycle-consistent character of the generated mappings. The image pair is registered twice to each other which should ideally result in images close to the original image pair. The image pairs can be compared by calculating the difference between images of the same modality.

The trained networks are tested and compared with other state of the art algorithms such as SimpleElastix [11] and VoxelMorph [6], [7].

II. Related Work

There are a great number of already existing image registration methodologies that are used in a medical context. Most non-learning based image registration algorithms that are used for multi-modal registration use an iconic matching criterion in combination with another matching criterion such

M. Ruppel, C. Melhuish, and L. Smith are with the Bristol Robotics Laboratory, UWE Bristol, BS16 1QY, UK (e-mail: mirjana.ruppel@uwe.ac.uk).
R. Persad is with the North Bristol NHS Trust, Bristol, UK.
A. Bahl is with the University Hospitals Bristol NHS Trust, Bristol, UK.
S. Dogramadzi is with the Department of Automation Control and Systems Engineering, The University of Sheffield, UK.

as geometric in [12] and [13] or a sensor based matching criterion in [14]. All three use a continuous optimizer to find a rigid [14], [12] or non-rigid [13] transformation model.

Learning-based approaches seem more promising in this area. In [15] CNNs learn a general similarity function to compare image patches. In [16] a CNN regression approach was successfully implemented to achieve 2D/3D rigid image registration in real-time by directly estimating the rigid transformation parameters between a given image pair.

Since human tissue is mostly soft, the here applied image registration algorithm needs to not only yield rigid but also non-rigid transformation parameters. In [6] and [8] unsupervised learning-based approaches based on CNNs were implemented for deformable 3D/3D image registration. The algorithms were both tested successfully on 3D MR brain scans but are applicable to image registration tasks in general.

Another very recent trend in machine learning applications for image registration is the implementation of Recurrent NNs instead of CNNs, especially RNNs that are Long Short-Termn Memory (LSTM), since they lead to very successful results in medical image segmentation tasks as seen in [17]. In [18] LSTMs were used to rigidly register US and MR images of the fetal brain via a dual-modality atlas image.

Another promising approach is the use of Generative Adversarial Networks (GANs) became a recent focus of research within image registration and segmentation, as seen in [19] and [20], respectively.

III. Formulation

We want the model to learn functions which map a multi-modal image pair $\{\mathrm{A}, \mathrm{B}\}$ to deformation fields $\{\Phi_{\mathrm{AB}}, \Phi_{\mathrm{BA}}\}$ which align each image to the respective other one. In our case A represents a CT image volume and B represents an MR image volume. As depicted in Fig. 1, the model consists of two generators $\mathrm{G_{AB}}$ and $\mathrm{G_{BA}}$ which produce deformation fields which align image B to image A and vice versa. Additionally, two discriminators $\mathrm{D_A}$ and $\mathrm{D_B}$ are implemented. $\mathrm{D_B}$ aims to classify whether the registered image $\hat{\mathrm{B}}$ is well aligned to image A, while $\mathrm{D_A}$ learns to classify whether image $\hat{\mathrm{A}}$ is well registered to image B. To learn this it is necessary to define positive cases which determine a 'perfect' registration. In this work we use images aligned via SimpleElastix deformable [11] as positive cases. If the positive cases are well defined, the discriminator learns how to compare a multi-modal image pair. A task that often is too complex for analytical methods.

The full objective includes four different kinds of terms: *adversarial losses* for classifying the quality of the generated alignment; a *cycle consistency loss* to overcome the issue of having to compare multi-modal images mathematically and to also rate the quality of the alignment; an *identity loss* to make sure the generated deformation field does not change the moving image if the image pair is already well aligned; and a *continuity loss* to assure that the generated deformation fields are smooth.

A. Adversarial Loss

Both mapping functions are directed by an individual adversarial loss. The generator $\mathrm{G_{AB}}$ works against the discriminator $\mathrm{D_B}$. Their adversarial loss is defined as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{GAN}}(\mathrm{G_{AB}}, \mathrm{D_B}) = (\mathrm{D_B}(\mathrm{A_T}) - 1)^2 \\ + \mathrm{D_B}(\tau(\mathrm{B}, \mathrm{G_{AB}}(\mathrm{A}, \mathrm{B})))^2\end{aligned} \tag{1}$$

$\mathrm{G_{AB}}$ generates deformation fields that are supposed to register the moving image B to fixed image A well enough so the discriminator rates them as well aligned. During the same time the discriminator $\mathrm{D_B}$ learns to rate the quality of the resulting alignments. We define a similar loss for the second mapping: $\mathcal{L}_{\mathrm{GAN}}(\mathrm{G_{BA}}, \mathrm{D_A})$

B. Cycle Consistency Loss

To further increase the quality of the registration we use the fact that the learned mappings should be cycle-consistent. The images of image pair $\{\mathrm{A}, \mathrm{B}\}$ are registered to each other via the deformation fields that were generated by $\mathrm{G_{AB}}$ and $\mathrm{G_{BA}}$. This results in the image pair $\{\hat{\mathrm{A}}, \hat{\mathrm{B}}\}$ where $\hat{\mathrm{A}}$ is image A registered to image B and $\hat{\mathrm{B}}$ is image B registered to image A. We then employ both generators a second time to register the image pair $\{\hat{\mathrm{A}}, \hat{\mathrm{B}}\}$ to each other. This results in the image pair $\{\tilde{\mathrm{A}}, \tilde{\mathrm{B}}\}$ which should be ideally identical to image pair $\{\mathrm{A}, \mathrm{B}\}$. Comparing the image pairs $\{\tilde{\mathrm{A}}, \tilde{\mathrm{B}}\}$ and $\{\mathrm{A}, \mathrm{B}\}$ with each other can be done via the L1 norm since we only compare the images of the same modality with each other.
We write this characteristic as the loss function:

$$\begin{aligned}\mathcal{L}_{\mathrm{cyc}}(\mathrm{G_{AB}}, \mathrm{G_{BA}}) = \left\|\tau(\hat{\mathrm{A}}, \mathrm{G_{BA}}(\hat{\mathrm{B}}, \hat{\mathrm{A}})) - \mathrm{A}\right\|_1 \\ + \left\|\tau(\hat{\mathrm{B}}, \mathrm{G_{AB}}(\hat{\mathrm{A}}, \hat{\mathrm{B}})) - \mathrm{B}\right\|_1\end{aligned} \tag{2}$$

C. Identity Loss

To make sure that well aligned images are not further changed, the following loss function is implemented:

$$\begin{aligned}\mathcal{L}_{\mathrm{ident}}(\mathrm{G_{AB}}, \mathrm{G_{BA}}) = \left\|\tau(\mathrm{A_T}, \mathrm{G_{AB}}(\mathrm{A}, \mathrm{A_T})) - \mathrm{A_T}\right\|_1 \\ + \left\|\tau(\mathrm{B_T}, \mathrm{G_{BA}}(\mathrm{B}, \mathrm{B_T})) - \mathrm{B_T}\right\|_1\end{aligned} \tag{3}$$

The positive case as defined for the adversarial loss is used as moving image. The generated deformation field should not alter the moving image as it is already well aligned.

D. Continuity Loss

As naturally occurring deformations in soft tissue are considered to be smooth, we also expect our deformation field to be smooth. The following loss function therefore penalizes non-smoothness:

$$\mathcal{L}_{\mathrm{reg}}(\mathrm{G_{AB}}, \mathrm{G_{BA}}) = \left\|\mathrm{G_{AB}}(\mathrm{A}, \mathrm{B})\right\|_2 + \left\|\mathrm{G_{BA}}(\mathrm{B}, \mathrm{A})\right\|_2 \tag{4}$$

E. Full Objective

The full objective is:

$$\begin{aligned}\mathcal{L}(\mathrm{G_{AB}}, \mathrm{G_{BA}}, \mathrm{D_A}, \mathrm{D_B}) = \mathcal{L}_{\mathrm{GAN}}(\mathrm{G_{AB}}, \mathrm{D_B}) + \mathcal{L}_{\mathrm{GAN}}(\mathrm{G_{BA}}, \mathrm{D_A}) \\ + \mu\mathcal{L}_{\mathrm{cyc}}(\mathrm{G_{AB}}, \mathrm{G_{BA}}) + \nu\mathcal{L}_{\mathrm{ident}}(\mathrm{G_{AB}}, \mathrm{G_{BA}}) \\ + \xi\mathcal{L}_{\mathrm{reg}}(\mathrm{G_{AB}}, \mathrm{G_{BA}})\end{aligned} \tag{5}$$

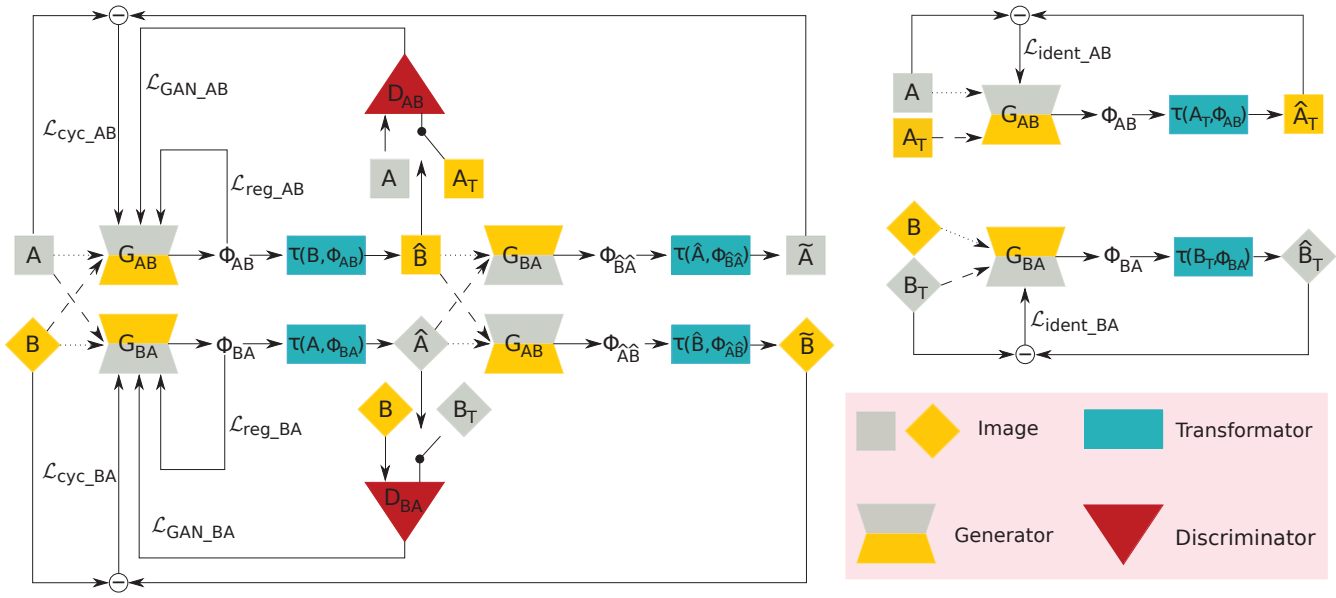

Fig. 1 Overview of the developed network. The differently coloured squares represent images A and B of different modalities. Dotted and dashed lines indicate the flow of fixed and moving images, respectively

where μ, ν, ξ control the weight of each part of the full objective.
The network is trained to solve:

$$G^{\star}_{AB}, G^{\star}_{BA} = \arg \min_{G_{AB}, G_{BA}} \max_{D_A, D_B} \mathcal{L}(G_{AB}, G_{BA}, D_A, D_B) \quad (6)$$

The loss functions rating the output of the generator should be minimized while the output of the discriminator should be maximized to favour correct judgment of the alignment of the image pair.

IV. Implementation

A. Network

As depicted in Fig. 1, two generators G_{AB} and G_{BA} - each consisting of a UNet similar to the one used in [6], [7] and introduced in [21] - learn how to generate smooth deformation fields Φ_{AB} and Φ_{BA} which align both images A and B to the other one, respectively. A spatial transformation layer τ applies the deformation fields to the moving images to generate aligned images $\hat{A}$ and $\hat{B}$. The discriminators D_A and D_B are consisting of Patch-GANs [22], [23] and try to classify whether the registered images $\hat{A}$ and $\hat{B}$ are well aligned to images B and A, respectively. The generators G_{AB} and G_{BA} are applied again. This time to align $\hat{A}$ and $\hat{B}$. The newly registered images $\tilde{A}$ and $\tilde{B}$ are then compared to the initial images A and B, respectively.

B. Training

For training we employ the Adam optimizer with a $\lambda = 0.0002$ learning rate. We choose $\mu = 1000$, $\nu = 10000$, and $\xi = 100$ to ensure that all parts of the objective are in the same range so they have equal impact on the network. We use 30 MRI and CT images of the open-source CHAOS dataset [1] for training, respectively, and combine them to 900 inter-patient image pairs for training. We hold out 10 images of each modality to generate 100 image pairs for testing. The model is trained for 100 epochs with a constant learning rate. For the next 100 epochs the learning rate is consistently decreased until it reaches 0.

V. Evaluation

For the evaluation of the resulting registration the liver segmentations of both the fixed and warped image are compared by calculating two different values. The Dice coefficient counts the amount of overlapping voxels. If Dice=1 the volumes overlap exactly [24]. The Hausdorff distance measures the distance between two subsets of a metric space, it is 0 if both volumes have the same boundary [25]. We calculate both values for the case of registration from CT to MRI and from MRI to CT. We then compare the results of our algorithm with the coefficients from three baseline algorithms: SimpleElastix affine, SimpleElastix deformable, and VoxelMorph [11], [7]. To improve readability, we refer to our algorithm as NANCY.

VI. Results

The registration results of all employed algorithms are depicted in Tables I and II. The average of Dice coefficient and Hausdorff distance are calculated for the registration of 100 inter-patient test image pairs that were not part of the training data.

Our approach is comparable to or slightly more successful than SimpleELastix and VoxelMorph, two state of the art algorithms. However, dice coefficients of maximum 0.80 leave room for improvement. Also notable is the fact that registration from CT to MRI leads to smaller dice coefficients and higher Hausdorff distances for most algorithms and is therefore less successful than registration from MRI to CT.

Exemplary registration results are shown in Fig. 2. Here NANCY results in a slightly smaller dice coefficient, yet the

TABLE I
EVALUATION OF THE REGISTRATION RESULTS FOR THE TEST DATASET. MRI ARE REGISTERED TO CT IMAGES

	Dice Coefficient	Hausdorff Distance
unaligned	0.49 ± 0.01	19.64 ± 28.23
SimpleElastix Affine	0.60 ± 0.01	17.23 ± 32.40
SimpleElastix Deformable	0.77 ± 0.01	16.84 ± 37.28
VoxelMorph	0.79 ± 0.02	15.31 ± 30.13
NANCY	0.80 ± 0.01	14.82 ± 35.17

TABLE II
EVALUATION OF THE REGISTRATION RESULTS FOR THE TEST DATASET. CT IMAGES ARE REGISTERED TO MRI

	Dice Coefficient	Hausdorff Distance
unaligned	0.49 ± 0.01	19.64 ± 28.23
SimpleElastix Affine	0.62 ± 0.01	17.10 ± 35.24
SimpleElastix Deformable	0.65 ± 0.02	16.48 ± 22.32
VoxelMorph	0.66 ± 0.02	16.02 ± 31.56
NANCY	0.66 ± 0.01	15.81 ± 30.77

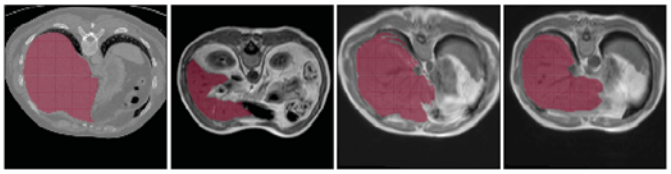

Fig. 2 From left to right: CT of liver, unaligned MRI, aligned MRI via SimpleElastix Deformable, aligned MRI via NANCY. Dice coefficient for unaligned pair: 0.30, for aligned pair using SimpleElastix: 0.75, for aligned pair using NANCY: 0.73. As depicted, both registration approaches result in more overlap of the liver. However, using NANCY the registered liver remains in a more physiological shape than the result obtained with SimpleElastix

resulting liver segmentation keeps a more physiological shape than the deformable SimpleElastix approach.

VII. DISCUSSION AND OUTLOOK

A novel approach of multi-modal image registration was implemented using GANs and cycle-consistency. The algorithm was tested on an open-source dataset consisting of 3D CT and MRI of the liver [1]. The results show that the newly developed algorithm is at least comparable if not more successful than the chosen baseline algorithms. However, the best results are a dice coefficient of 0.80 and a Hausdorff distance of 14.82 for the registration of MRI to CT via NANCY which leaves space for improvement.

One flaw of the algorithm is the definition of a 'positive case' for the learning process of the discriminator. Using images that were aligned via the SimpleElastix deformable registration algorithm can obviously not lead to perfect results since SimpleElastix does not generate perfectly aligned images. Our definition of a 'positive case' therefore needs to change. One idea to do this would be by using translated images instead of aligned images as positive cases. E.g use an MRI that was translated to look like a CT as the positive case for how a perfectly aligned CT to said MRI would look like. Algorithms that could be employed for this strategy are either cycleGAN as introduced in [10] and [20].

Another idea to improve NANCY is to change the implemented model from a UNet-type model to a long short-term memory network (LSTM) or to a kervolutional neural network as introduced by [26]. Both networks seem to be better equipped to handle the entirety of the spatio-temporal context of a 3D image volume as opposed to CNNs that only perceive small local voxel contexts [18], [26]. Both networks have not yet been applied to multi-modal deformable image registration.

REFERENCES

[1] A. E. Kavur, M. A. Selver, O. Dicle, M. Bar, and N. S. Gezer, "Chaos - combined (ct-mr) healthy abdominal organ segmentation challenge data," Apr. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3362844

[2] E. Ferrante and N. Paragios, "Slice-to-volume medical image registration: A survey," *Medical image analysis*, vol. 39, pp. 101–123, 2017.

[3] M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, and N. Komodakis, "A deep metric for multimodal registration," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 10–18.

[4] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Nonrigid multimodality image registration," in *Medical Imaging 2001: Image Processing*, vol. 4322. International Society for Optics and Photonics, 2001, pp. 1609–1620.

[5] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1449–1453.

[6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An unsupervised learning model for deformable medical image registration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9252–9260.

[7] G. Balakrishnan, A. Zhao, M. R. Sabuncu, and J. Guttag, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE transactions on medical imaging*, 2019.

[8] X. Yang, R. Kwitt, and M. Niethammer, "Fast predictive image registration," in *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 48–57.

[9] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration–a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[11] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The design of simpleitk," *Frontiers in neuroinformatics*, vol. 7, p. 45, 2013.

[12] J. Mitra, S. Ghose, D. Sidibé, A. Oliver, R. Marti, X. Llado, J. C. Vilanova, J. Comet, and F. Mériaudeau, "Weighted likelihood function of multiple statistical parameters to retrieve 2d trus-mr slice correspondence for prostate biopsy," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 2949–2952.

[13] C. Reynier, J. Troccaz, P. Fourneret, A. Dusserre, C. Gay-Jeune, J.-L. Descotes, M. Bolla, and J.-Y. Giraud, "Mri/trus data fusion for prostate brachytherapy. preliminary results," *Medical physics*, vol. 31, no. 6, pp. 1568–1575, 2004.

[14] S. Xu, J. Kruecker, B. Turkbey, N. Glossop, A. K. Singh, P. Choyke, P. Pinto, and B. J. Wood, "Real-time mri-trus fusion for guidance of targeted prostate biopsies," *Computer Aided Surgery*, vol. 13, no. 5, pp. 255–264, 2008.

[15] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4353–4361.

[16] S. Miao, Z. J. Wang, and R. Liao, "A cnn regression approach for real-time 2d/3d registration," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.

[17] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation," in *Advances in neural information processing systems*, 2015, pp. 2998–3006.

[18] R. Wright, B. Khanal, A. Gomez, E. Skelton, J. Matthew, J. V. Hajnal, D. Rueckert, and J. A. Schnabel, "Lstm spatial co-transformer networks for registration of 3d fetal us and mr brain images," in *Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis*. Springer, 2018, pp. 149–159.

[19] J. Fan, X. Cao, Q. Wang, P.-T. Yap, and D. Shen, "Adversarial learning for mono- or multi-modal registration," *Medical Image Analysis*, vol. 58, p. 101545, 2019.

[20] Z. Zhang, L. Yang, and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network," *CoRR*, vol. 1802.09655, 2018.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[22] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[23] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *European conference on computer vision*. Springer, 2016, pp. 702–716.

[24] T. Sørenson, *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons*, 1948.

[25] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: measuring error on simplified surfaces," in *Computer Graphics Forum*, vol. 17, no. 2. Wiley Online Library, 1998, pp. 167–174.

[26] C. Wang, J. Yang, L. Xie, and J. Yuan, "Kervolutional neural networks," *CoRR*, vol. 1904.03955, 2019.