GSMorph: Gradient Surgery for Cine-MRI Cardiac Deformable Registration

Haoran Dou[1], Ning Bi[1], Luyi Han[2,3], Yuhao Huang[4,5,6], Ritse Mann[2,3],
Xin Yang[4,5,6,7], Dong Ni[4,5,6], Nishant Ravikumar[1,8],
Alejandro F. Frangi[1,8,9,10], and Yunzhi Huang[11(✉)]

[1] Centre for Computational Imaging and Simulation Technologies in Biomedicine
(CISTIB), University of Leeds, Leeds, UK
[2] Department of Radiology and Nuclear Medicine, Radboud University Medical
Centre, Nijmegen, The Netherlands
[3] Department of Radiology, Netherlands Cancer Institute, Amsterdam, The
Netherlands
[4] National-Regional Key Technology Engineering Laboratory for Medical
Ultrasound, School of Biomedical Engineering, Health Science Center, Shenzhen
University, Shenzhen, China
[5] Medical Ultrasound Image Computing (MUSIC) Lab, Shenzhen University,
Shenzhen, China
[6] Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen,
China
[7] Shenzhen RayShape Medical Technology Co., Ltd., Shenzhen, China
[8] Division of Informatics, Imaging and Data Science, Schools of Computer Science
and Health Sciences, University of Manchester, Manchester, UK
[9] Medical Imaging Research Center (MIRC), Electrical Engineering and
Cardiovascular Sciences Departments, KU Leuven, Leuven, Belgium
[10] Alan Turing Institute, London, UK
[11] Institute for AI in Medicine, School of Artificial Intelligence, Nanjing University of
Information Science and Technology, Nanjing, China
yunzhi.huang.scu@gmail.com

Abstract. Deep learning-based deformable registration methods have
been widely investigated in diverse medical applications. Learning-based
deformable registration relies on weighted objective functions trading off
registration accuracy and smoothness of the deformation field. Therefore,
they inevitably require tuning the hyperparameter for optimal registra-
tion performance. Tuning the hyperparameters is highly computationally
expensive and introduces undesired dependencies on domain knowledge.
In this study, we construct a registration model based on the gradient
surgery mechanism, named GSMorph, to achieve a hyperparameter-free
balance on multiple losses. In GSMorph, we reformulate the optimiza-
tion procedure by projecting the gradient of similarity loss orthogonally
to the plane associated with the smoothness constraint, rather than addi-
tionally introducing a hyperparameter to balance these two competing

---

H. Dou and N. Bi—Contributed equally to this work.

terms. Furthermore, our method is model-agnostic and can be merged into any deep registration network without introducing extra parameters or slowing down inference. In this study, We compared our method with state-of-the-art (SOTA) deformable registration approaches over two publicly available cardiac MRI datasets. GSMorph proves superior to five SOTA learning-based registration models and two conventional registration techniques, SyN and Demons, on both registration accuracy and smoothness.

Keywords: Medical image registration · Gradient surgery · Regularization

1   Introduction

Image registration is fundamental to many medical image analysis applications, e.g., motion tracking, atlas construction, and disease diagnosis [5]. Conventional registration methods usually require computationally expensive iterative optimization, making it inefficient in clinical practice [1,19]. Deep learning has recently been widely exploited in the registration domain due to its superior representation extraction capability and fast inference speed [2,7]. Deep-learning-based registration (DLR) formulates registration as a network learning process minimizing a composite objective function comprising one similarity loss to penalize the difference in the appearance of the image pair, and a regularization term to ensure the smoothness of deformation field. Typically, to balance the registration accuracy and smoothness of the deformation field, a hyperparameter is introduced in the objective function. However, performing hyperparameter tuning is labor-intensive, time-consuming, and *ad-hoc*; searching for the optimal parameter setting requires extensive ablation studies and hence training tens of models and establishing a reasonable parameter search space. Therefore, alleviating, even circumventing, hyperparameter search to accelerate development and deployment of DLR models remains challenging.

Recent advances [6,11,13] in DLR have primarily focused on network architecture design to boost registration performance. Few studies [9,16] investigated the potential in preventing hyperparameter searching by hypernetwork [8] and conditional learning [10]. Hoopes *et al.* [9] leveraged a hyper-network that takes the hyperparameter as input to generate the weight of the DLR network. Although effective, it introduces a large number of additional parameters to the basic DLR network, making the framework computationally expensive. In parallel, Mok *et al.* [16] proposed to learn the effect of the hyperparameter and condition it on the feature statistics (usually illustrated as *style* in computer vision [10]) to manipulate the smoothness of the deformation field in the inference phase. Both methods can avoid hyperparameter tuning while training the DLR model. However, they still require a reasonable sampling space and strategy of the hyperparameter, which can be empirically dependent.

Gradient surgery (GS) projects conflicting gradients of different losses during the optimization process of the model to mitigate gradient interference. This has proven useful in multi-task learning [20] and domain generalization [15]. Motivated by these studies, we propose utilizing the GS to moderate the discordance between the similarity loss and regularization loss. The proposed method can further avert searching the weight for balancing losses in training the DLR.

– We propose GSMorph, a gradient-surgery-based DLR model. Our method can circumvent tuning the hyperparameter in composite loss function with a gradient-level reformulation to reach the trade-off between registration accuracy and smoothness of the deformation field.
– Existing GS approaches have operated the parameters' gradients independently or integrally. We propose a layer-wise GS to group by the parameters for optimization to ensure the flexibility and robustness of the optimization process.
– Our method is model-agnostic and can be integrated into any DLR network without extra parameters or losing inference speed.

2 Methodology

Deformable image registration estimates the non-linear correspondence field ϕ between the moving, M, and fixed, F, images (Fig. 1). Such procedure is mathematically formulated as $\phi = f_\theta(F, M)$. For learning-based registration methods, f_θ (usually adopted by a neural network) takes the fixed and moving image pair as input and outputs the deformation field via the optimal parameters θ. Typically, θ can be updated using standard mini-batch gradient descent as follows:

$$\theta := \theta - \alpha \nabla_\theta \left(\mathcal{L}_{sim}(\theta; F, M \circ \phi) + \lambda \mathcal{L}_{Reg}(\theta; \phi) \right) \qquad (1)$$

where α is the learning rate; $\mathcal{L}_{sim}$ is the similarity loss to penalize differences in the appearance of the moving and fixed images (e.g., mean square error, mutual information or local negative cross-correlation); $\mathcal{L}_{reg}$ is the regularization loss to encourage the smoothness of the deformation field (this can be computed by the gradient of the deformation field); λ is the hyperparameter balancing the trade-off between $\mathcal{L}_{sim}$ and $\mathcal{L}_{reg}$ to achieve desired registration accuracy while preserving the smoothness of the deformation field in the meantime. However, hyperparameter tuning is time-consuming and highly experience-dependent, making it tough to reach the optimal solution.

Insight into the optimization procedure in Eq. 1, as registration accuracy and spatial smoothness are potentially controversial in model optimization, the two constraints might have different directions and strengths while going through the gradient descent. Based on this, we provide a geometric view to depict the gradient changes for θ based on the *gradient surgery* technique. The conflicting relationship between two controversial constraints can be geometrically projected as orthogonal vectors. Depending on the orthogonal relationship, merely updating the gradients of the similarity loss would automatically associate with the
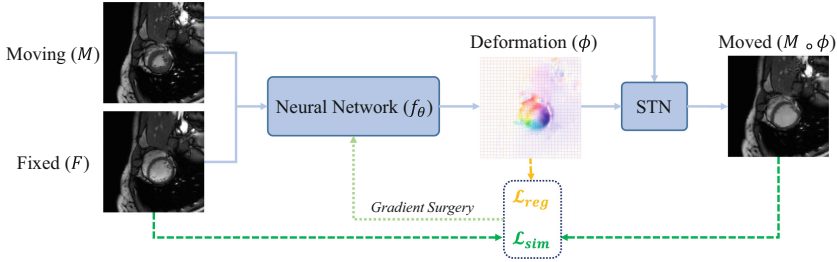
Fig. 1. Schematic illustration of our proposed GSMorph. GS modifies the gradients computed by similarity loss $\mathcal{L}_{sim}$ and regularization loss $\mathcal{L}_{reg}$, then updates the model's parameters θ.

updates of the regularization term. In this way, we avoid tuning the hyperparameter λ to optimize θ. The Eq. 1 can then be rewritten into a non-hyperparameter pattern:

$$\theta := \theta - \alpha \Phi(\nabla_\theta \mathcal{L}_{sim}(\theta; F, M \circ \phi)) \tag{2}$$

where $\Phi(\cdot)$ is the operation of proposed GS method.

2.1 Layer-Wise Gradient Surgery

Figure 2 illustrates the two scenarios of gradients while optimizing the DLR network via vanilla gradient descent or gradient surgery. We first define that the gradient of similarity loss, g_{sim}, and that of regularization loss, g_{reg}, are conflicting when the angle between g_{sim} and g_{reg} is the obtuse angle, viz. $\langle g_{sim}, g_{reg} \rangle < 0$. In this study, we propose updating the parameters of neural networks by the original g_{sim} independently, when g_{sim} and g_{reg} are non-conflicting, representing g_{sim} has no incompatible component of the gradient along the direction of g_{reg}. Consequently, optimization with sole g_{sim} within a non-conflicting scenario can inherently facilitate the spatial smoothness of deformations.

Conversely, as shown in Fig. 2, conflicting gradients are the dominant reason associated with non-smooth deformations. Hence, deconflicting gradients in the optimization of the DLR network to ensure high registration accuracy, as well as smooth deformation, is the primary goal of our study. Following a simple and intuitive procedure, we project the g_{sim} onto the normal plane of the g_{reg}, where the projected similarity gradient g and g_{reg} are non-conflicting along each gradient's direction.

Existing studies [15,20] performed the GS in terms of independent parameters or the entire network. Despite the effectiveness, these can be either unstable or inflexible. Considering that a neural network usually extracts features through the collaboration of each parameter group in the convolution layers, we introduce a layer-wise GS to ensure its stability and flexibility. The parameter updating rule is detailed in the Algorithm 1. Specifically, in each gradient updating iteration, we first compute the gradients of two losses for the parameter group in
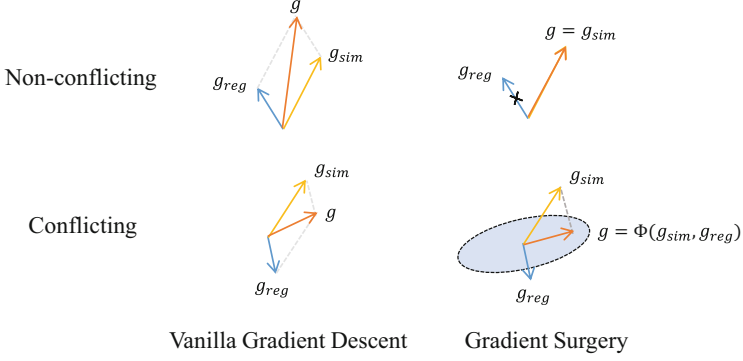
Fig. 2. Visualization of vanilla gradient descent and gradient surgery for non-conflicting and conflicting gradients. Regarding vanilla gradient descent, the gradient, g, is computed based on the average of g_{sim} and g_{reg}. Our GS-based approach projects the g_{sim} onto the normal vector of g_{reg} to prevent disagreements between the similarity loss and regularization loss. On the other hand, we only update the g_{sim} in non-conflicting scenarios.

each layer separately. Then, the conflicting relationship between the two gradients is calculated based on their inner production. Once the two gradients are non-conflicting, the gradients used to update its corresponding parameter group will be only the original gradients of similarity loss; on the contrary, the gradients will be the projected similarity gradients orthogonal to the gradients of regularization, which can be calculated as $g^i_{sim} - \frac{\langle g^i_{sim}, g^i_{reg} \rangle}{\|g^i_{reg}\|^2} g^i_{reg}$. After performing GS on all layer-wise parameter groups in the network, the final gradients will be used to update the model's parameters.

Algorithm 1. Gradient surgery

Require: Parameters θ_i in ith layer of the network; Number of
    layers in the network N.
1: $g_{sim} \leftarrow \nabla_\theta \mathcal{L}_{sim}$
2: $g_{reg} \leftarrow \nabla_\theta \mathcal{L}_{reg}$
3: **for** $i = 1 \rightarrow N$ **do**
4:     **if** $\langle g^i_{sim}, g^i_{reg} \rangle > 0$ **then**
5:         $g_i = g^i_{sim}$
6:     **else**
7:         $g_i = g^i_{sim} - \frac{\langle g^i_{sim}, g^i_{reg} \rangle}{\|g^i_{reg}\|^2} g^i_{reg}$
8:     **end if**
9:     $\Delta\theta_i = g_i$
10: **end for**
11: Update θ

2.2   Network Architecture

Our network architecture (seen in Fig. 1) is similar to VoxelMorph [7] that comprises naive U-Net [18] and spatial transform network (STN) [12]. The U-Net takes the moving and fixed image pair as input and outputs the deformation field, which is used to warp the moving image via STN. The U-Net consists of an encoder and a decoder with skip connections, which forward the features from each layer in the encoder to the corresponding layer in the decoder by concatenation to enhance the feature aggregation and prevent gradient vanishing. The number of feature maps in the encoder part of the network is 16, 32, 64, 128, and 256, increasing the number of features as their size shrinks, and vice versa in the decoder part. Each convolutional block in the encoder and decoder has two sequential convolutions with a kernel size of 3, followed by a batch normalization and a leaky rectified linear unit.

3   Experiments and Results

3.1   Datasets and Implementations

Datasets.   In this study, we used two public cardiac cine-MRI datasets for investigation and comparison: ACDC [3] and M&M [4]. ACDC and M&M contain 100 and 249 subjects, respectively. We followed a proportion of 75%, 5%, and 20% to split each dataset for training, validation, and testing. We selected the image from the cine-MRI cardiac sequence at the End Systole (ES) time point of the cardiac cycle as the moving image, and that at the End Diastole (ED) as the fixed one. All images were cropped into the size of 128×128 centralized to the heart. We normalized the intensity of images into the range from 0 to 1 before inputting them into the model.

Implementation Details.   We implemented our model in PyTorch [17], using a standard PC with an NVIDIA GTX 2080ti GPU. We trained the network through Adam optimizer [14] with a learning rate of 5e-3 and a batch size of 32 for 500 epochs. We also implemented and trained alternative methods for comparison with the same data and similar hyper-parameters for optimization. Our source code is available at https://github.com/wulalago/GSMorph.

3.2   Alternative Methods and Evaluation Criteria

To demonstrate the advantages of our proposed method in medical image registration, we compared it with two conventional deformable registration methods, i.e., **Demons** [19] and **SyN** [1], and a widely-used DLR model, **Voxel-Morph** [7]. These methods usually need laborious effort in hyperparameter tuning. Additionally, we reported the results of VoxelMorph trained with different λ (i.e., 0.1, 0.01, and 0.001, denoted as **VoxelMorph-l**, **VoxelMorph-m**, **VoxelMorph-s**). Meanwhile, we compared our approach to one alternative

DLR model based on the hyperparameter learning, i.e., **HyperMorph** [9]. This method only require additional validations in searching the optimal hyperparameter without necessarily tuning it from scratch. Finally, we reformulated two variations of GS based on our concept for further comparison. Specifically, **GS-Agr** [15] treats the gradient of each parameter independently. It updates the parameter with the gradient of similarity loss in the non-conflicting scenario, and a random gradient sampled from the Gaussian distribution when conflicting. While **GS-PCGrad** [20] uses the same GS strategy as ours, but with respect to the whole parameters of the entire network. The **Initial** represents the results without any deformation.

In this study, we used six criteria to evaluate the efficacy and efficiency of the investigated methods, including Dice score (Dice) and 95% Hausdorff distance (HD95) to validate the registration accuracy of the regions of interest, Mean square error (MSE) to evaluate the pixel-level appearance difference between the moved and fixed image-pairs, the percentage of pixels with negative Jacobian determinant (NJD) values to compare the smoothness and diffeomorphism of the deformation field, the number of parameters (Param) of the neural network and inference speed (Speed) to investigate the efficiency.

3.3   Results

As summarized in Table 1, our method could obtain the best MSE in the ACDC dataset and Dice in the M&M dataset while achieving comparable performance with the tuned VoxelMorph over other metrics. The Dice and HD95 reported in Table 1 were averaged over three anatomical regions of interest in the heart, i.e., Left ventricle, Myocardium, and Right Ventricle (LV, Myo, and RV). Consequently, the proposed model achieved superior registration accuracy and spatial regularization with faster inference speed than the two conventional registration methods. We also observed that our approach gained higher registration performance than HyperMorph in both datasets. Regarding the GS-based methods, GS-Agr totally collapsed, as the conflicting gradients accounted for most have

Table 1. Quantitative comparison of investigated methods on the testing datasets over ACDC and M&M.

Methods	Dataset							
	ACDC				**M&M**			
	Dice(%)	HD95(mm)	MSE(10^{-2})	NJD(%)	Dice(%)	HD95(mm)	MSE(10^{-2})	NJD(%)
Initial	61.81±8.68	4.40±1.33	1.58±0.52	–	61.03±10.16	4.79±1.82	1.90±1.08	–
Demons	85.38±3.52	1.67±0.75	0.46±0.21	1.31±0.59	75.66±10.30	17.79±6.23	0.71±0.61	1.84±1.19
SyN	79.28±8.23	2.24±1.28	0.65±0.21	0.30±0.27	81.97±9.36	**2.45±2.04**	0.84±0.12	0.49±0.47
VoxelMorph-s	86.69±2.17	1.30±0.24	0.39±0.14	2.01±0.96	77.12±9.36	3.43±2.18	**0.42±0.29**	3.45±2.33
VoxelMorph-m	**87.47±2.21**	**1.29±0.30**	0.42±0.15	0.67±0.48	79.93±8.57	2.91±1.98	0.48±0.32	1.31±1.10
VoxelMorph-l	82.12±4.30	1.87±0.64	0.59±0.18	**0.10±0.14**	77.18±8.69	2.81±1.60	0.74±0.43	**0.16±0.22**
HyperMorph	83.44±3.55	1.75±0.64	0.47±0.20	1.60±0.86	77.21±8.45	3.28±1.99	0.59±0.37	2.50±1.22
GS-Agr	63.40±9.15	4.20±1.35	1.33±0.43	0	63.41±9.85	4.50±1.77	1.55±0.86	<0.001
GS-PCGrad	84.59±3.53	1.62±0.53	0.51±0.16	0.11±0.17	80.67±8.18	2.48±1.67	0.59±0.36	0.41±0.44
GSMorph	87.45±2.27	1.34±0.40	**0.31±0.11**	0.87±0.52	**82.26±6.59**	2.66±1.93	0.49±0.27	0.98±0.84
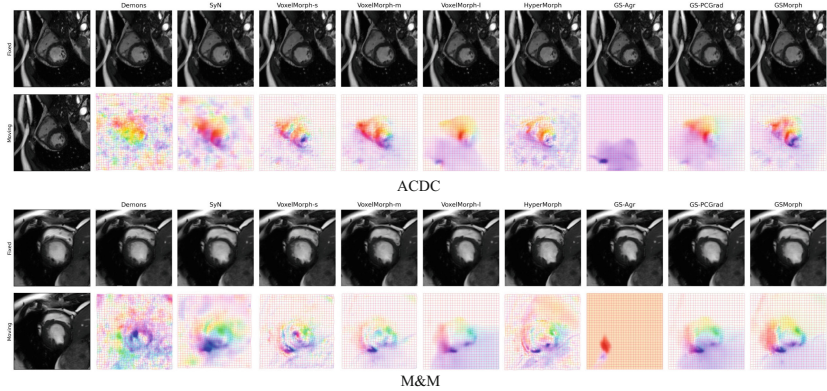
Fig. 3. Visual comparison of the registration results of the investigated methods for two representative test cases in ACDC and M&M datasets. The top rows are the fixed images and moved images from different methods; the bottom rows are the moving images and deformation fields. (We encourage you to zoom in for better visualization)

been replaced by random noise. On the other hand, GS-PCGrad only yielded an inadequate registration performance with an inclination of over-regularization. The comparison in the GS-based method shows the flexibility and robustness of our approach.

Figure 3 illustrates the sample cases of the warped images and the corresponding deformation fields from the compared methods. It can be observed our methods could obtain the moved images most similar to the fixed ones. Voxelmorph could achieve comparable results to us but still require a time-consuming hyperparameter tuning. Overall, the results of the comparisons in Table 1 and Fig. 3 indicate that our method performed the best among all the techniques that we implemented and examined, showing the effectiveness of our model in balancing the trade-off between registration accuracy and smoothness of deformations.

In Table 2, we have also reported the number of parameters and inference speed. We observed that DLR methods could obtain faster speed compared with conventional ones in general. As our proposed approach only modified the optimization procedure of the backbone network, it could maintain the original inference speed and the number of parameters. Conversely, HyperMorph intro-

Table 2. Number of parameters and inference speed of investigated methods on the testing datasets over ACDC and M&M.

Methods	Demons	SyN	VoxelMorph	HyperMorph	GSMorph
Params	–	–	1.96M	126M	1.96M
Speed	7.55±1.79	16.59±5.48	2.29±0.83	2.96±1.09	2.29±0.83

duced tremendous extra parameters and loss of inference speed as they adopted the secondary network to generate the conditions or weights of the main network architecture.

4   Conclusion

This work presents a gradient-surgery-based registration framework for medical images. To the best of our knowledge, this is the first study to employ gradient surgery to refine the optimization procedure in learning the deformation fields. In our GSMorph, the gradients from the similarity constraint were projected onto the plane orthogonal to those from the regularization term. In this way, merely updating the gradients in optimizing the registration accuracy would result in a joint updating of the gradients from the similarity and regularity constraints. Then, no additional regularization loss is required in the network optimization and no hyperparameter is further required to explicitly trade off between registration accuracy and spatial smoothness. Our model outperformed the conventional registration methods and the alternative DLR models. Finally, the proposed method is model-agnostic and can be integrated into any DLR network without introducing extra parameters or compromising the inference speed. We believe GSMorph will facilitate the development and deployment of DLR models and alleviate the influence of hyperparameters on performance.

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. **12**(1), 26–41 (2008)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9252–9260 (2018)
3. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging **37**(11), 2514–2525 (2018)
4. Campello, V.M., et al.: Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&Ms challenge. IEEE Trans. Med. Imaging **40**(12), 3543–3554 (2021)

5. Chen, X., Diaz-Pinto, A., Ravikumar, N., Frangi, A.F.: Deep learning in medical image registration. Prog. Biomed. Eng. **3**(1), 012003 (2021)

6. Chen, X., Xia, Y., Ravikumar, N., Frangi, A.F.: A deep discontinuity-preserving image registration network. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12904, pp. 46–55. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87202-1_5

7. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 729–738. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_82

8. Ha, D., Dai, A., Le, Q.V.: Hypernetworks. arXiv preprint arXiv:1609.09106 (2016)

9. Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., Dalca, A.V.: HyperMorph: amortized hyperparameter learning for image registration. In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds.) IPMI 2021. LNCS, vol. 12729, pp. 3–17. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78191-0_1

10. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510 (2017)

11. Huang, Y., Ahmad, S., Fan, J., Shen, D., Yap, P.T.: Difficulty-aware hierarchical convolutional neural networks for deformable registration of brain MR images. Med. Image Anal. **67**, 101817 (2021)

12. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in Neural Information Processing Systems, vol. 28 (2015)

13. Jia, X., Thorley, A., Chen, W., Qiu, H., Shen, L., Styles, I.B., Chang, H.J., Leonardis, A., De Marvao, A., O'Regan, D.P., et al.: Learning a model-driven variational network for deformable image registration. IEEE Trans. Med. Imaging **41**(1), 199–212 (2021)

14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

15. Mansilla, L., Echeveste, R., Milone, D.H., Ferrante, E.: Domain generalization via gradient surgery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6630–6638 (2021)

16. Mok, T.C.W., Chung, A.C.S.: Conditional deformable image registration with convolutional neural network. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12904, pp. 35–45. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87202-1_4

17. Paszke, A., et al.: PyTorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, vol. 32 (2019)

18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

19. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. Neuroimage **45**(1), S61–S72 (2009)

20. Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., Finn, C.: Gradient surgery for multi-task learning. In: Advances in Neural Information Processing Systems, vol. 33, pp. 5824–5836 (2020)