

# Deep learning based registration using spatial gradients and noisy segmentation labels

Théo Estienne<sup>1,2</sup>, Maria Vakalopoulou<sup>1</sup>, Enzo Battistella<sup>1,2</sup>, Alexandre Carré<sup>2</sup>, Théophraste Henry<sup>2</sup>, Marvin Lerousseau, Charlotte Robert<sup>2</sup>, Nikos Paragios<sup>3</sup>, and Eric Deutsch<sup>2</sup>

<sup>1</sup> Université Paris-Saclay, CentraleSupélec, Mathématiques et Informatique pour la Complexité et les Systèmes, Inria Saclay, 91190, Gif-sur-Yvette, France.

{theo.estienne}@centralesupelec.fr

<sup>2</sup> Université Paris-Saclay, Institut Gustave Roussy, Inserm, Radiothérapie Moléculaire et Innovation Thérapeutique, 94800, Villejuif, France.

<sup>3</sup> Therapanacea, Paris, France

**Abstract.** Image registration is one of the most challenging problems in medical image analysis. In the recent years, deep learning based approaches became quite popular, providing fast and performing registration strategies. In this short paper, we summarise our work presented on Learn2Reg challenge 2020. The main contributions of our work rely on *(i)* a symmetric formulation, predicting the transformations from source to target and from target to source simultaneously, enforcing the trained representations to be similar and *(ii)* integration of variety of publicly available datasets used both for pretraining and for augmenting segmentation labels. Our method reports a mean dice of 0.64 for task 3 and 0.85 for task 4 on the test sets, taking third place on the challenge. Our code and models are publicly available at [https://github.com/TheoEst/abdominal\\_registration](https://github.com/TheoEst/abdominal_registration) and [https://github.com/TheoEst/hippocampus\\_registration](https://github.com/TheoEst/hippocampus_registration).

## 1 Introduction

In the medical field, the problem of deformable image registration has been heavily studied for many years. The problem relies on establishing the best dense voxel-wise transformation ( $\Phi$ ) to wrap one volume (source or moving,  $M$ ) to match another volume (reference or fixed,  $F$ ) in the best way. Traditionally, different types of formulations and approaches had been proposed in the last years [18] to address the problem. However, with the recent advances of deep learning, a lot of learning based methods became very popular currently, providing very efficient and state-of-the art performances [10]. Even if there is a lot of work in the field of image registration there are still a lot of challenges to be addressed. In order to address these challenges and provide common datasets for the benchmarking of learning based [5,6] and traditional methods [11,1], the Learn2Reg challenge is organised [4]. Four tasks were proposed by the organisers with different organs and modalities. In this work, we focused on two tasks: the CT abdominal (task 3) and the MRI hippocampus registration (task 4).

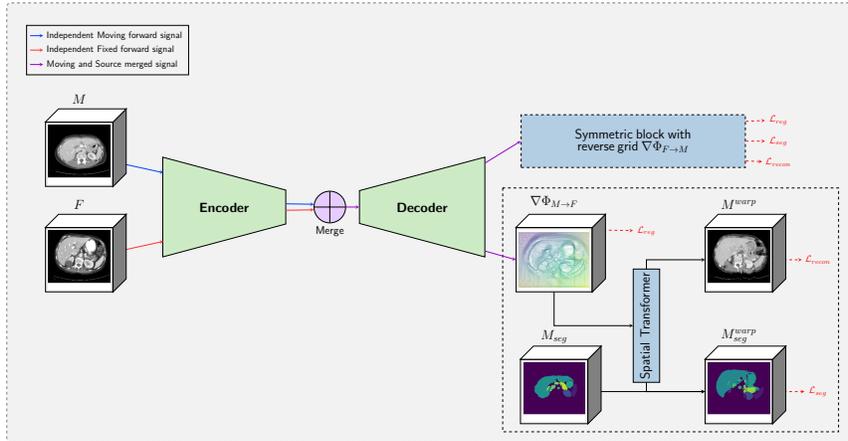


Fig. 1: Schematic representation of the proposed methodology.

In this work, we propose a learning based method that learns how to obtain spatial gradients in a similar way to [19,7]. The main contributions of this work rely on (i) enforcing the same network to predict both  $\Phi_{M \rightarrow F}$  and  $\Phi_{F \rightarrow M}$  deformations using the same encoding and implicitly enforcing it to be symmetric and (ii) integrating noisy labels from different organs during the training, to fully exploit publicly available datasets. In the following sections, we will briefly summarise these two contributions and present our results that gave to our method the third position in the Learn2Reg challenge 2020 (second for task 3 and third for task 4).

## 2 Methodology

An overview of our proposed framework is presented in the Figure 1. Our method uses as backbone a 3D UNet [3] based architecture, which consists of 4 blocks with 64, 128, 256 and 512 channels for the encoder part (**E**). Each block consists of a normalisation layer, Leaky ReLU activation, 3D convolutions with a kernel size of  $3 \times 3 \times 3$  and convolution with kernel size and stride 2 to reduce spatial resolution. Each of the  $F, M$  volumes passes independently through the encoder part of the network. Their encoding is then merged using the subtraction operation before passing through the decoder (**D**) part for the prediction of the optimal spatial gradients of the deformation field  $\nabla\Phi$ . We obtained the deformation field  $\Phi$  from its gradient using integration which we approximated with the cumulative summation operation.  $\Phi$  is then used to obtain the deformed volume together with its segmentation mask using warping  $M^{warp} = \mathcal{W}(M, \Phi_{M \rightarrow F})$ . Finally, we apply deep supervision to train our network in a way similar to [14].

**Symmetric training** Even if our grid formulation has constraints for the spatial gradients to avoid self-crossings on the vertical and horizontal directions

for each of the  $x,y,z$ -axis, our formulation is not diffeomorphic. This actually indicates that we can not calculate the inverse transformation of  $\Phi_{M \rightarrow F}$ . To deal with this problem, we predict both  $\Phi_{M \rightarrow F}$  and  $\Phi_{F \rightarrow M}$  and we use both for the optimization of our network. Different methods such as [13,9] explore similar concepts using however different networks for each deformation. Due to our fusion strategy on the encoding part, our approach is able to learn both transformations with less parameters. In particular, our spatial gradients are obtained by:  $\nabla \Phi_{M \rightarrow F} = \mathbf{D}(\mathbf{E}(M) - \mathbf{E}(F))$  and  $\nabla \Phi_{F \rightarrow M} = \mathbf{D}(\mathbf{E}(F) - \mathbf{E}(M))$ .

**Pretraining and Noisy Labels** Supervision has been proved to boost the performance of the learning based registration methods integrating implicit anatomical knowledge during the training procedure. For this reason, in this study, we investigate ways to use publicly available datasets to boost performance. We exploit available information from publicly available datasets namely KITS 19 [12], Medical Segmentation Decathlon (sub-cohort Liver, Spleen, Pancreas, Hepatic Lesion and Colon) [17] and TCIA Pancreas[16,8]. In particular, we trained a 3D UNet segmentation network on 11 different organs (spleen, right and left kidney, liver, stomach, pancreas, gallbladder, aorta, inferior vena cava, portal vein and oesophagus). To harmonise the information that we had at disposal for each dataset, we optimised the dice loss only on the organs that were available per dataset. The network was then used to provide labels for the 11 organs for approximately 600 abdominal scans. These segmentation masks were further used for the pretraining of our registration network for the task 3. After the training the performance of our segmentation network on the validation set in terms of dice is summarised to: 0.92 (Spl), 0.90 (RKid), 0.91 (LKid), 0.94 (Liv) 0.83 (Sto), 0.74 (Pan), 0.72 (GBla), 0.89 (Aor), 0.76 (InfV), 0.62 (PorV) and 0.61 (Oes). The validation set was composed of 21 patients of Learn2Reg and TCIA Pancreas dataset.

Furthermore, we explored the use of pretraining of registration networks on domain-specific large datasets. In particular, for task 3 the ensemble of the publicly available datasets together with their noisy segmentation masks were used to pretrain our registration network, after a small preprocessing including an affine registration step using Advanced Normalization Tools (ANTs)[2] and isotropic resampling to 2mm voxel spacing. Moreover, for task 4, we performed an unsupervised pretraining using approximately 750 T1 MRI from OASIS 3 dataset [15] without segmentations. For both tasks, the pretraining had been performed for 300 epochs.

## 2.1 Training Strategy and Implementation Details

To train our network, we used a combination of multiple loss functions. The first one was the reconstruction loss optimising a similarity function over the intensity values of the medical volume  $\mathcal{L}_{sim}$ . For our experiments, we used the mean square error function and normalized cross correlation, depending on the experiment, between the warped image  $M^{warp}$  and the fixed image  $F$ . The second loss integrated anatomical knowledge by optimising the dice coefficient

between the warped segmentation and the segmentation of the fixed volume:  $\mathcal{L}_{sup} = Dice(M_{seg}^{warp}, F_{seg})$ . Finally, a regularisation loss was also integrated to enforce smoothness of the displacement field by keeping it close to zero deformation:  $\mathcal{L}_{smo} = \|\nabla\Phi_{M \rightarrow F}\|$ . These losses composed our final optimization strategy calculated for both  $\nabla\Phi_{M \rightarrow F}$  and  $\nabla\Phi_{F \rightarrow M}$

$$\mathcal{L} = (\alpha\mathcal{L}_{sim} + \beta\mathcal{L}_{sup} + \gamma\mathcal{L}_{smo})_{M \rightarrow F} + (\alpha\mathcal{L}_{sim} + \beta\mathcal{L}_{sup} + \gamma\mathcal{L}_{smo})_{F \rightarrow M}$$

where  $\alpha$ ,  $\beta$  and  $\gamma$  were weights that were manually defined. The network was optimized using Adam optimiser with a learning rate set to  $1e^{-4}$ .

Regarding the implementation details, for task 3, we used batch size 2 with patch size equal to  $144 \times 144 \times 144$  due to memory limitations. Our normalisation strategy included the extraction of three CT windows, which all of them are used as additional channels and min-max normalisation to be in the range  $(0, 1)$ . For our experiments we did not use any data augmentation and we set  $\alpha = 1$ ,  $\beta = 1$  and  $\gamma = 0.01$ . The network was trained on 2 Nvidia Tesla V100 with 16 GB memory, for 300 epochs for  $\approx 12$  hours. For task 4, the batch size was set to 6 with patches of size  $64 \times 64 \times 64$  while data augmentation was performed by random flip, random rotation and translation. Our normalisation strategy in this case included:  $\mathcal{N}(0, 1)$  normalisation, clipping values outside of the range  $[-5, 5]$  and min-max normalisation to stay to the range  $(0, 1)$ . The weights were set to  $\alpha = 1$ ,  $\beta = 1$  and  $\gamma = 0.1$  and the network was trained on 2 Nvidia GeForce GTX 1080 GPUs with 12 GB memory for 600 epochs for  $\approx 20$  hours.

The segmentation network, used to produce noisy segmentations, was a 3D UNet trained with batch size 6, learning rate  $1e^{-4}$ , leaky ReLU activation functions, instance normalisation layers and random crop of patch of size  $144 \times 144 \times 144$ . During inference, we kept the ground truth segmentations of the organs available, we applied a normalisation with connected components and we checked each segmentations manually to remove outlier results.

### 3 Experimental Results

For each task, we performed an ablation study to evaluate the contribution of each component and task 3, we performed a supplementary experiment integrating the noisy labels during the pretraining. The evaluation was performed in terms of Dice score, 30% of lowest Dice score, Hausdorff distance and standard deviation of the log Jacobian. These metrics evaluated the accuracy and robustness of the method as well as the smoothness of the deformation. Our results are summarised in Table 1, while some qualitative results are represented in Figure 2. For the inference on the test set, we used our model trained on both training and validation datasets. Concerning the computational time, our approach needs 6.21 and 1.43 seconds for the inference respectively for task 3 and 4. This is slower than other participants to the challenge, probably due to the size of our deep network which have around 20 millions parameters.

Concerning task 3, one can observe a significant boost on the performance when the pretraining with the noisy labels was integrated. Due to the challenging

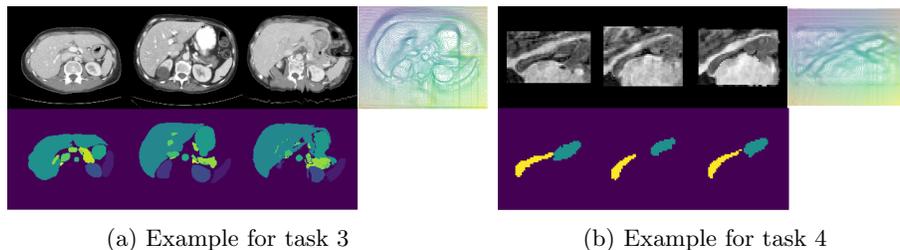


Fig. 2: Results obtained on the validation set. From left to right : moving, fixed, deformed images and the deformation grid. For the task 3, we displayed an axial view with the different organs (second row). For the task 4, we displayed a sagittal view with the head and tail masks (second row)

nature of this registration problem, the impact of the symmetric training was not so high in any of the metrics. On the other hand, for task 4, the symmetric component with the pretraining boosted the robustness of the method while the pretraining had a lower impact than on task 3. One possible explanation is that for this task, the number of provided volumes in combination with the nature of the problem was enough for training a learning based registration method.

Dataset		Task 3				Task 4			
		Dice	Dice30	Hd95	StdJ	Dice	Dice30	Hd95	StdJ
Val	Unregistered	0.23	0.01	46.1		0.55	0.36	3.91	
Val	Baseline	0.38	0.35	45.2	1.70	0.80	0.78	2.12	<b>0.067</b>
Val	Baseline + sym.	0.40	0.36	45.7	1.80	0.83	0.82	1.68	0.071
Val	Baseline + sym. + pretrain	0.52	0.50	42.3	<b>0.32</b>	<b>0.84</b>	<b>0.83</b>	<b>1.63</b>	0.093
Val	Baseline + sym. + pretrain + noisy labels	<b>0.62</b>	<b>0.58</b>	<b>39.3</b>	1.77				
Test	Baseline + sym. + pretrain + noisy labels	0.64	0.40	37.1	1.53	0.85	0.84	1.51	0.09

Table 1: Evaluation of our method for the Tasks 3 & 4 of Learn2Reg Challenge on the validation set (val) and on the test set (test).

## 4 Conclusions

In this work, we summarise our method that took the 3rd place in the Learn2Reg challenge, participating on the tasks 3 & 4. Our formulation is based on spatial gradients and explores the impact of symmetry, pretraining and integration of public available datasets. In the future, we aim to further explore symmetry in our method and investigate ways that our formulation could hold diffeomorphic properties. Finally, adversarial training is also something that we want to explore in order to be deal with multimodal registration.

## References

1. Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008.
2. Brian B Avants, Nick Tustison, and Gang Song. Advanced normalization tools (ants). *Insight j*, 2(365):1–35, 2009.
3. Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
4. Adrian Dalca, Yipeng Hu, Tom Vercauteren, Mattias Heinrich, Lasse Hansen, Marc Modat, Bob de Vos, Yiming Xiao, Hassan Rivaz, Matthieu Chabanas, Ingerid Reinertsen, Bennett Landman, Jorge Cardoso, Bram van Ginneken, Alessa Hering, and Keelin Murphy. Learn2reg - the challenge, March 2020.
5. Adrian V. Dalca, Guha Balakrishnan, John V. Guttag, and Mert R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *MICCAI*, 2018.
6. Bob D de Vos, Floris F Berendsen, Max A Viergever, Hessam Sokooti, Marius Staring, and Ivana Išgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, 52, 2019.
7. Théo Estienne, Marvin Lrousseau, Maria Vakalopoulou, Emilie Alvarez Andres, Enzo Battistella, et al. Deep learning-based concurrent brain registration and tumor segmentation. *Frontiers in Computational Neuroscience*, 14:17, 2020.
8. Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, et al. Automatic multi-organ segmentation on abdominal ct with dense v-networks. *IEEE transactions on medical imaging*, 37(8), 2018.
9. Yi Guo, Xiangyi Wu, Zhi Wang, Xi Pei, and X George Xu. End-to-end unsupervised cycle-consistent fully convolutional network for 3d pelvic ct-mr deformable registration. *Journal of Applied Clinical Medical Physics*, 2020.
10. Grant Haskins, Uwe Kruger, and Pingkun Yan. Deep learning in medical image registration: a survey. *Machine Vision and Applications*, 31(1):8, 2020.
11. Mattias P Heinrich, Mark Jenkinson, Michael Brady, and Julia A Schnabel. Mrf-based deformable registration and ventilation estimation of lung ct. *IEEE transactions on medical imaging*, 32(7):1239–1248, 2013.
12. Nicholas Heller, Niranjana Sathianathan, Arveen Kalapara, Edward Walczak, et al. The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. *arXiv preprint arXiv:1904.00445*, 2019.
13. Boah Kim, Jieun Kim, June-Goo Lee, Dong Hwan Kim, Seong Ho Park, and Jong Chul Ye. Unsupervised deformable image registration using cycle-consistent cnn. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019.
14. Julian Krebs, Hervé Delingette, Boris Mailhé, Nicholas Ayache, and Tommaso Mansi. Learning a probabilistic model for diffeomorphic registration. *IEEE Transactions on Medical Imaging*, 38(9), 2019.
15. Daniel S Marcus, Anthony F Fotenos, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults. *Journal of cognitive neuroscience*, 22(12):2677–2684, 2010.

16. Holger R Roth, Amal Farag, E Turkbey, Le Lu, Jiamin Liu, and Ronald M Summers. Data from pancreas-ct. the cancer imaging archive, 2016.
17. Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019.
18. Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7), 2013.
19. Christodoulidis Stergios, Sahasrabudhe Mihir, Vakalopoulou Maria, Chassagnon Guillaume, et al. Linear and deformable image registration with 3d convolutional neural networks. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Springer, 2018.