Multi-step, Learning-Based, Semi-supervised Image Registration Algorithm

Marek Wodzinski

Department of Measurement and Electronics, AGH University of Science and Technology, Krakow, Poland

wodzinski@agh.edu.pl

Abstract. This paper presents a contribution to the Learn2Reg challenge organized jointly with the MICCAI 2020, more specifically, to the task related to inter-patient hippocampus registration in magnetic resonance images. The proposed algorithm is a multi-step, learning-based, and semi-supervised procedure. The method consists of a sequentially stacked U-Net-like architecture, trained in alternation. The method was ranked as the second-best (for the hippocampus registration task) in terms of the combined challenge evaluation criteria.

Keywords: Image registration · Deep learning · Medical imaging · L2R · Learn2Reg

1 Introduction

This paper presents a contribution to the Learn2Reg challenge organized jointly with the MICCAI 2020 conference [1]. The method aims to perform inter-patient registration of the hippocampus in mono-modal magnetic resonance images (MRI). The proposed method is a multi-step, learning-based algorithm. It combines the self-supervision, based on the MIND-loss [2], with the weak supervision using available segmentation masks.

2 Methods

2.1 Method

The proposed method is a learning-based, multi-step registration procedure. It consists of a sequentially stacked U-Net-like architecture [3]. The algorithm starts with initializing an identity transformation and concatenating the source and target images. The images are passed through the network and the calculated displacement field is composed with the current transformation. Then, the source image is warped with the displacement field, again concatenated with the target image, and passed to the following level (Fig. 1). The process is repeated...
for a predefined number of levels. The models do not share weights and are trained simultaneously and independently with different regularization parameters. As a result, the later levels get slightly different inputs during each epoch. It was experimentally verified that this approach serves as a self-augmentation, it decreases the results for the training set but improves them for the validation set, compared to training the networks sequentially or simultaneously. The leaky ReLU was used as the activation function. The networks were trained with a small batch size, thus the batch normalization was replaced by the group normalization [4].

The objective function is a weighted sum of the modality independent neighbourhood descriptor self-similarity context (MIND-SSC) [2,5], the diffusion regularizer, and the mean squared error between the segmentation masks:

\[
C(S, T, S_m, T_m, u) = MINDSSC(S \circ u, T) + \alpha R(u) + \beta MSE(S_m \circ u, T_m),
\]

where \(S, T, S_m, T_m\) denotes the source, target, source mask, and target mask respectively, \(u\) is the dense displacement field, \(R\) denotes the diffusive regularization, and \(\alpha, \beta\) are the parameters controlling the transformation smoothness and the influence of segmentation masks respectively. The segmentation masks were warped with the linear interpolation to make the cost function differentiable with respect to the transformation grid.

### 2.2 Dataset and Experimental Setup

The dataset consists of MRI scans acquired in 195 adult subjects, both healthy and with non-affective psychotic disorders. The images show the hippocampus head and body with a small surrounding neighborhood. There are 394 volumes, 263 in the training set and 131 in the test set. The training set contains segmentation masks of the hippocampus head and body. The volumes were resampled to the same voxel resolution. A more detailed dataset description can be found in [1,6].

The training was performed for a predefined number of epochs (30), using Adam, with an exponentially decaying learning rate (initial learning rate: 0.002, decaying rate: 0.92), batch size equal to 4, and 3 network levels. A single network consists of 2 660 739 trainable parameters. The total number of trainable parameters is equal to the number of levels times the value above (7 982 217). The value for \(\beta\) was constant and set to 0.8, while the values for \(\alpha\) were equal to 2.0, 2.6 and 3.4 for each level respectively. The MIND-SSC radius and dilation were set to 2. The networks were trained in alternation. This means that the objective function was evaluated, and the weights were updated independently between levels, while the other networks were in forward mode only. The motivation behind this approach was first to reduce the memory required during training, and second to provide a slight self-augmentation mechanism. Since the outputs of the first level were changing during each epoch, the inputs to the later levels were different. It was observed that this approach slightly improved the results on the validation set (Table 1). The training time was roughly 30 h using
RTX 2080 Ti. The method was implemented using PyTorch [7] and the source code is available at [8].

Fig. 1. Visualization of the proposed multi-step procedure and the network architecture. The segmentations masks are used during training only. Each level of the deep network consists of 5 convolutional layers (3-D) and 4 transposed convolutional layers (3-D).
3 Results

The method was evaluated using the following metrics: (i) Dice similarity coefficient of segmentation masks (DSC), (ii) robustness score (DSC30 - 30% cases with the lowest DSC), (iii) 95% percentile of Hausdorff distance of segmentations (HD95), and (iv) standard deviation of log Jacobian determinant (SDlogJ).

The quantitative results are presented in the Table 1. The table presents also the results for the validation set with $\beta$ set to 0, and the results for training the 3 levels simultaneously without alternation, and a comparison to state-of-the-art VoxelMorph [9]. An exemplary visualization of the registered images and the warped segmentation masks is shown in Fig. 2. The method was ranked as the second-best in terms of the combined challenge evaluation criteria for the hippocampus registration task [1]. A comparison to other challenge participants is presented in Table 1, is available on the challenge website [1], and will be summarized in the challenge overview article.

Fig. 2. Exemplary visual results of the registration for a validation pair. The columns (from left) show the middle slice of source, target, registered source, source mask, target mask, and warped source mask respectively.
Table 1. Quantitative results on the validation and test sets for the proposed method, as well as the other participants’ methods. The high difference between the validation and test set arises from an unobserved overfitting due to a slightly incorrect experimental setup. The ranks and times are unavailable for the methods evaluated on the validation set. The “w/o alter” denotes results acquired training the 3 levels simultaneously, while the “w/o labels” shows results with $\beta$ set to 0. Details about the rank calculation are available at [1].

| Dataset                      | Avg. DSC | Avg. DSC30 | Avg. HD95 | Avg. SDlogJ | Time  | Rank |
|------------------------------|----------|------------|-----------|-------------|-------|------|
| Initial                      | 0.55     | 0.36       | 3.91      | --          | --    | --   |
| VoxelMorph (val. w/o labels)| 0.74     | 0.72       | 2.65      | 0.07        | --    | --   |
| Validation (w/o alter)       | 0.86     | 0.85       | 1.49      | 0.05        | --    | --   |
| Validation (w/o labels)      | 0.78     | 0.77       | 2.36      | 0.05        | --    | --   |
| Validation (proposed)        | 0.87     | 0.86       | 1.36      | 0.07        | --    | --   |
| Test (proposed)              | 0.79     | 0.76       | 2.20      | 0.08        | 0.76  | 2    |
| LapIRN                       | 0.88     | 0.86       | 1.30      | 0.05        | 1.00  | 1    |
| CentraleSupelec              | 0.85     | 0.84       | 1.51      | 0.09        | 1.43  | 3    |
| PDD-Net                      | 0.78     | 0.76       | 2.23      | 0.07        | 0.31  | 4    |
| LibReg                       | 0.85     | 0.84       | 1.55      | 0.05        | --    | 5    |
| Deeds                        | 0.76     | 0.71       | 2.49      | 0.11        | 3.14  | 6    |
| Uppsala                      | 0.74     | 0.67       | 2.82      | 0.16        | 21.96 | 7    |
| Nifty (unoptimized)          | 0.73     | 0.65       | 3.37      | 1.00        | 4.72  | 8    |

4 Conclusion

The proposed method provides slightly better results than state-of-the-art unsupervised methods (Table 1). The method could be further improved by a proper augmentation of the dataset to prevent overfitting. The results for the validation set are considerably better compared to the test set because the validation pairs contained images that were used in other training pairs (a small mistake during the experiment preparation that was then repeated for other experiments to maintain consistency). It is also uncertain how the semi-supervised method will behave on a dataset acquired with different scanners or using another acquisition protocol. Noteworthy, based on the low average SDlogJ, the calculated deformations are relatively smooth and regular.

To conclude, the presented method is a semi-supervised, multi-step, learning-based registration procedure trained in alternation that outperforms state-of-the-art unsupervised methods, however, it still requires substantial improvements in terms of generalizability to new data.

Acknowledgments. This work was funded by NCN Preludium project no. UMO-2018/29/N/ST6/00143.
References

1. Dalca, A., Hering, A., Hansen, L., Heinrich, M.: The Challenge. https://learn2reg.grand-challenge.org/
2. Heinrich, M., et al.: MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. 16(7), 1423–1435 (2012)
3. 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
4. Wu, Y., He, K.: Group normalization. arXiv:1803.08494 (2018)
5. Heinrich, M.P., Hansen, L.: Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5d displacement search. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12263, pp. 190–200. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_19
6. Simpson, A., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv:1902.09063 (2019)
7. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32, pp. 8024–8035. Curran Associates, Inc. (2019)
8. Wodzinski, M.: The Source Code. https://github.com/lNefarin/L2R-Task4
9. Balakrishnan, G., Zhao, A., Sabuncu, M., Guttag, J., Dalca, A.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging 38(8), 1788–1800 (2019)