Learn to fuse input features for large-deformation registration with differentiable convex-discrete optimisation

Hanna Siebert⬤ and Mattias P. Heinrich⬤

Institute of Medical Informatics, Universität zu Lübeck, Germany
{siebert,heinrich}@imi.uni-luebeck.de

Abstract. Hybrid methods that combine learning-based features with conventional optimisation have become popular for medical image registration. The ConvexAdam algorithm that ranked first in the comprehensive Learn2Reg registration challenges completely decouples semantic and/or hand-crafted feature extraction from the estimation of the transformation due to the difficulty of differentiating the discrete optimisation step. In this work, we propose a simple extension that enables backpropagation through discrete optimisation and learns to fuse the semantic and hand-crafted features in a supervised setting. We demonstrate state-of-the-art performance on abdominal CT registration.

Keywords: large deformation registration · convex optimisation · end-to-end learning.

1 Introduction and Related Work

While end-to-end learning of fully-convolutional networks is the method of choice for semantic segmentation, image registration continues to benefit from integrating conventional optimisation steps, e.g. pairwise instance optimisation [7], a discretised search of displacements [1] or iterative recurrent updates [9]. Discrete optimisation has been shown to yield excellent registration quality for numerous tasks [2,7,5] but does rely on non-differentiable steps which would prevent its use in end-to-end learning. We aim for a method that offers the possibility to use discrete optimisation in an end-to-end learning setting. Therefore, we introduce a differentiable convex discrete optimisation approach that is able to align images with large deformations. This differentiable optimisation is used to learn the fusion of semantic and hand-crafted image features.

2 Method

Figure 1 gives an overview of our method: First, hand-crafted and semantic features are extracted from the input images, concatenated and passed to a small network comprising layers for feature fusion. The fixed and moving features output from this network are then used for our differentiable discretised convex optimisation method to align images with large deformations.
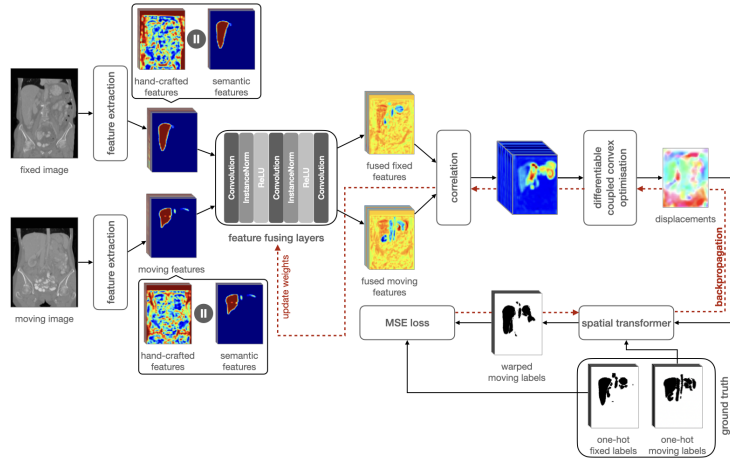
Fig. 1: Overview of our method: Hand-crafted and semantic features are concatenated and fused with feature fusing network layers. The fused features are used for our differentiable optimisation method to compute displacements. For backpropagation, warped moving and fixed labels are passed to a MSE loss function to update the feature fusing network's weights whereas the feature extraction part of the framework remains frozen.

2.1   Differentiable convex-discrete optimisation

For pairwise deformable image registration, a deformation field $\mathbf{u}$ is sought that minimises the cost function $E(I_F, I_M, \mathbf{u})$ to align a fixed image I_F and a moving image I_M. In [3], a non-differentiable convex-discrete method has been proposed to find a deformation field $\mathbf{u}$ by solving a combined cost function

$$E(\mathbf{v}, \mathbf{u}) = DSV(\mathbf{v}) + \frac{1}{2\theta}(\mathbf{v} - \mathbf{u})^2 + \alpha|\nabla \mathbf{u}|^2 \tag{1}$$

that ensures similarity and smoothness optimisation. In this function, $\mathbf{v}$ is an auxiliary second deformation field used to compute the displacement space volume DSV. The regularisation parameter α controls the smoothness of the deformation field and the parameter θ models the coupling between similarity and regularisation penalty and is decreased during iterative solving of the equation. The optimal selection of $\mathbf{v}$ with respect to the similarity term can be performed globally optimal using local cost aggregation [3].

In this work, we introduce a differentiable discretised convex optimisation by replacing argmin operators with their corresponding softmin counterparts and make suitable adjustments to hyper-parameters that reduce memory requirements for end-to-end learning. Coupled-convex discrete optimisation [3] approximates more complex MRF-solutions by the following steps:

(0) initialisation of the current displacement field to zeros
(1) computation of a correlation volume based on sum of squared differences of feature tensors (the volume comprises 6 dimensions, 3 spatial dimension and 3 displacement dimensions)

(2a) a regularising coupling term that adds 3D parabolas in displacement dimensions that are rooted at the current displacement solution

(2b) the argmin operator (across all possible displacements) that defines a new regularised displacement field

(2c) a spatial smoothing step (e.g. a box-filter)

The correlation volume (step (1)) directly depends on the feature maps obtained from fixed and moving scans. By defining a large enough capture range and correspondingly a discrete mesh grid of relative displacements the method can robustly find a near global optimum without multiple warping steps or cascaded architectures. Steps (2a) - (2c) are iteratively repeated with a continuously increasing weight for the coupling term that helps to ensure convergence of the optimisation. Step (2b), which takes the argmin is not differentiable and will be replaced with a softmin operator along the displacement dimension followed by a point-wise multiplication with the relative displacements of the predefined discrete mesh grid and subsequent reduction.
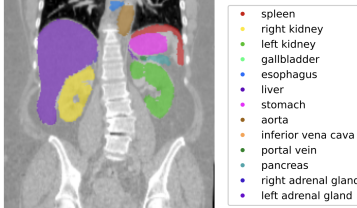
2.2   Learning of input feature fusion

Previous work [3,7] has shown that hand-crafted MIND features [4] or automatic nnU-Net segmentations [6] can be used as input for a coupled convex optimisation method for image registration. In this work, we combine hand-crafted and semantic features by fusing them with help of trainable feature fusing network layers comprising two $1 \times 1 \times 1$-convolutions followed by instance normalisations and ReLU activations. The first convolution increases the number of feature channels to 32 and a third $1 \times 1 \times 1$-convolution reduces the number of feature channels to 15. The resulting feature maps are then used to solve the differentiable convex-discrete optimisation problem described in section 2.1 in order to compute the displacement fields that are then used to warp the moving label maps. One-hot representations of warped and fixed label maps weighted inversely proportional to the square root of the class frequency are passed to a MSE loss function that is used to train the feature fusing network's parameters whereas the feature extraction part of the framework stays frozen.

3   Experiments and Results

For our experiments we use the Learn2Reg-2020 challenge's (task 3) dataset containing 30 abdominal inter-patient CT scans with 13 manually labeled abdominal organs and a resolution of $192 \times 160 \times 256$ [5,10]. The scans have been linearly pre-registered and split into 20 training cases and 10 test cases. For evaluation we consider all possible pairwise combinations of the test cases. From the image data, we extract MIND features (leading to 12 feature channels) and compute one-hot encoded label features by applying a nnU-Net trained on the 20 training cases (leading to 14 feature channels). We downsample the features to a resolution of $48 \times 40 \times 64$, concatenate them and pass the 26-channel input

Table 1: Left: Quantitative results: Accuracy is measured by the Dice similarity of segmentations and the 95% Hausdorff distance for segmentations. Plausibility of the deformations is measured by the standard deviation of the logarithmic Jacobian determinant. Right: example visualisation of fixed image and warped moving labels.

	Dice [%]	HD [mm]	SDlogJ
initial	25.14	40.21	—
MIND features	37.79	37.22	0.050
nnU-Net features	50.56	24.71	0.021
concatenated features	49.71	28.33	0.050
fused features	56.37	24.13	0.049

to our feature fusing network. The network's 15-channel output is then used for displacement computation with the differentiable convex optimisation method. Therefore, we use a displacement range that covers ~ 32 mm within the scanned abdominal region and scale the softmin operation's output (step (2b)) by half of the downsampled feature dimensions. The feature fusing network is trained for 50 epochs using Adam and a learning rate of 0.005.

For evaluation, we upsample the obtained displacement fields to the original image resolution. We compare our fused features with the direct use of MIND features, nnU-Net label features, and concatenation of MIND and nnU-Net features. The results given in Table 1 show that the fusion of MIND and nnU-Net features clearly outperforms the other investigated feature variants with an average Dice score of 56.37% compared to 50.56% when using only nnU-Net features. As using nnU-Net features yields to a deformation field that is optimised to warp the foreground structures, the SDlogJ value is lower than when MIND features are involved. We evaluated the potential problem of label bias with an experiment on additional structures (lumbar and thoracic vertebrae[1] [8]) unseen for the nnU-Net segmentation training and our fusion learning. While using only MIND features yields the highest accuracy we see great potential for the proposed feature fusion that only reduced the Dice score of the spine by 5% while the nnU-Net-based registration results in a drop of 42%. Hence the influence of label bias is substantially reduced.

4   Discussion and Conclusion

This work introduced a differentiable version of coupled convex discrete optimisation for image registration with large deformation. It has opened up the possibility of end-to-end feature learning and has well-performed for our feature fusing network. We show that the fusion of semantic label features and hand-crafted features based on image self-similarities leads to an improved registration performance compared to either using only semantic or only hand-crafted features or the simple concatenation of both.

---

[1] https://github.com/MIRACLE-Center/CTSpine1K

References

1. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2758–2766 (2015)
2. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: Mrf-based deformable registration and ventilation estimation of lung ct. IEEE Transactions on Medical Imaging **32**(7), 1239–1248 (2013)
3. Heinrich, M.P., Papież, B.W., Schnabel, J.A., Handels, H.: Non-parametric discrete registration with convex optimisation. In: International Workshop on Biomedical Image Registration. pp. 51–61. Springer (2014)
4. 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: International conference on medical image computing and computer-assisted intervention. pp. 187–194. Springer (2013)
5. Hering, A., Hansen, L., Mok, T.C., Chung, A., Siebert, H., Häger, S., Lange, A., Kuckertz, S., Heldmann, S., Shao, W., et al.: Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. arXiv preprint arXiv:2112.04489 (2021)
6. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
7. Siebert, H., Hansen, L., Heinrich, M.P.: Fast 3d registration with accurate optimisation and little learning for learn2reg 2021. pp. 174–178. Springer International Publishing (2021)
8. Smith, K., Clark, K., Bennett, W., Nolan, T., Kirby, J., Wolfsberger, M., Moulton, J., Vendt, B., Freymann, J.: Data from ct_colonography. The Cancer Imaging Archive (2015). https://doi.org/10.7937/K9/TCIA.2015.NWTESAY1
9. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: European conference on computer vision. pp. 402–419. Springer (2020)
10. Xu, Z., Lee, C.P., Heinrich, M.P., Modat, M., Rueckert, D., Ourselin, S., Abramson, R.G., Landman, B.A.: Evaluation of six registration methods for the human abdomen on clinically acquired ct. IEEE Transactions on Biomedical Engineering **63**(8), 1563–1572 (2016)