Weak Bounding Box Supervision for Image Registration Networks

Mona Schumacher[1,2], Hanna Siebert[1], Ragnar Bade[2], Andreas Genz[2], and Mattias Heinrich[1]

[1] University of Luebeck, Institute of Medical Informatics, Luebeck, Germany
[2] MeVis Medical Solutions AG, Bremen, Germany
mona.schumacher@mevis.de

Abstract. Image registration is a fundamental task in medical image analysis. Many deep learning based methods use multi-label image segmentations during training to reach the performance of conventional algorithms. But the creation of detailed annotations is very time-consuming and expert knowledge is essential. To avoid this, we propose a weakly supervised learning scheme for deformable image registration that uses bounding boxes during training. By calculating the loss function based on these bounding box labels, we are able to perform an image registration with large deformations without using densely labeled annotations. The performance of the registration of inter-patient 3D Abdominal CT images can be enhanced by approximately 10% only with little annotation effort in comparison to unsupervised learning methods. Taken into account this annotation effort, the performance also exceeds the performance of the label supervised training.

Keywords: deformable image registration · weak supervision · bounding box supervision.

1 Introduction

Medical image registration is the process of the alignment of the anatomical structures of two or more images in order to be able to do follow up studies, image-guidance or to plan a treatment. Deep learning methods have become increasingly important. They have demonstrated low computation times and are promising to enable real time registration approaches. For the case of brain image registration [1], which only require small deformation, already satisfactory results could be achieved. The registration of images of highly deformable body regions, such as the abdominal region or thorax are, due to the respiration or digestion, more complex and still often solved with conventional algorithms [2, 3]. Deep learning methods have started to address the challenge of handling large deformations (for example in the Learn2Reg Challenge, cf. learn2reg.grandchallenge.org) [4, 5]. Mok et al. [6] use Laplacian pyramids to solve the registration in a coarse-to-fine scheme inspired by classical algorithms. They show that label supervision substantially increases the registration accuracy, which is also

shown by Siebert et al. [7]. In image segmentation, weak label supervision has already gained interest. Rajchl et al. [8], for example, use an extension of the GrabCut algorithm and learn segmentation from bounding box annotations. In this paper, our aim is to close the gap between supervised and unsupervised registration methods and propose a weakly supervised learning scheme for deformable image registration including large deformations and introduce a loss function based on 3D bounding boxes to decrease the effort of the labeling process. We use inter-patient 3D Abdominal CT images and are able to increase the overlap of organs by approximately 10% in comparison to unsupervised image registration methods. If the time of the labeling process is taken into account, the performance of supervised algorithms can also be exceeded.

2 Methods

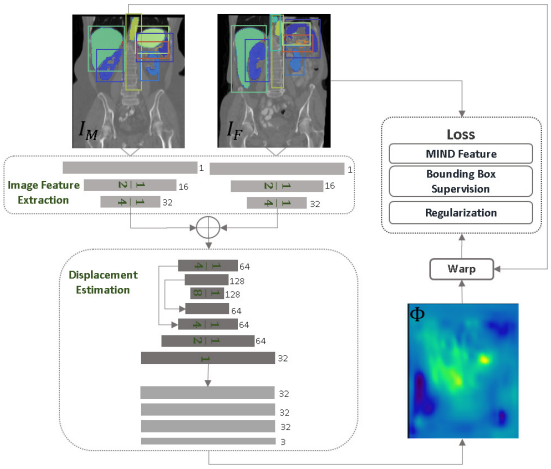

Fig. 1. Architecture of proposed method: Image features are extracted for I_F and I_M separately in two decoders (shared weights). The concatenated features are passed through a U-Net-like architecture and are finally used to estimate a displacement Φ to warp I_M. The loss consists of three parts: MIND features, regularization and the proposed bounding box supervision. The resolution in relation to the input resolution of the different steps are displayed in the layers.

The network consists of two parts: an image feature extraction part and a displacement estimation part. An overview of the architecture is shown in Fig. 1. The image feature extraction part extracts the low level features of the input images in two streams (with shared weights for monomodal registration). The displacement estimation part uses the concatenated low level features and estimates the displacement field. The 32 concatenated feature maps of I_F and I_M are used as input to extract 32 joint feature maps with a U-Net-like network with

three encoder and four decoder blocks. Three additional sequences are added to estimate the displacement field. The final displacement field is generated by reducing the 32 feature maps to the three displacement dimensions with a $1 \times 1 \times 1$ convolution and transformed to normalized sampling voxel locations (value range from -1 to 1) with the *tanh* activation function to match the PyTorch grid definition. The deformation has the same size as the input images.

To train the network, weak label supervision is used. Instead of using detailed labels for the calculation of the loss function, bounding boxes are used. The advantage of this method is that a significant reduction in time can be achieved and the variance between raters is also lower. A combination of three loss functions is used: the modality independent neighbourhood descriptor (MIND) with self-similar context (SSC) [10], a diffusion regularization and the mean squared error for the bounding boxes. The bounding box loss is multiplied by a factor of two.

To generate the final registration result including large deformations, we apply the network twice. The first input images are I_F and I_M. Then, I_M is warped with the first displacement field. The resulting warped moving image is used as second input.

3 Experiments

To train and evaluate our method, we use the publicly available Learn2Reg challenge dataset (Task3, 2020). This dataset contains 30 abdominal CT scans with thirteen manually labeled abdominal organs [4, 5]. For training and testing, we use the split and validation pairs as in the official challenge. The data is already preprocessed to same voxel sizes and spatial dimensions. We downsample the images for the experiments to a size of $144 \times 112 \times 144$ due to GPU memory requirements. For all labels, tight bounding boxes as well as a bounding box with a random error of $\pm 5\%$ are generated. The network is trained using Adam optimizer with a learning rate of 0.001 for 7500 iterations.

We train our network three times: unsupervised (not using the label loss), with the proposed bounding box loss, and with the voxelwise manually labeled organ segmentations. To establish comparability between training with label and weak label loss, we perform additional runs of supervised training with less training data. In this way, we simulate manual generation of labels or bounding boxes that takes the same amount of time. In total, we have five experiments: unsupervised, tight-weakly-supervised, weakly-supervised, supervised and supervised_50%. Tight-weakly refers to perfect bounding boxes, weakly refers to bounding boxes with an additional error of $\pm 5\%$ and supervised_50% refers to the experiment with less labeled data.

4 Results

In Table 1 the average Dice scores for all organs are listed for the different trainings. In comparison to the initial overlap of the organs, the overlap can

Table 1. Dice scores [%] for spleen, right kidney, left kidney, gall bladder, esophagus, liver, stomach, aorta, inferior vena cava, portal and splenic vein, pancreas, left adrenal gland, and right adrenal gland.

														avg ± std
initial	42	34	35	2	23	62	24	33	36	5	15	8	9	25 ± 13
unsupervised	67	57	61	5	33	81	35	54	50	15	21	18	14	39 ± 14
tight-weakly-supervised	70	67	69	7	33	86	41	53	56	20	27	25	17	44 ± 13
weakly-supervised	67	64	64	6	32	83	40	54	56	18	28	24	16	43 ± 13
supervised	81	73	78	8	43	86	50	67	61	17	25	21	16	48 ± 11
supervised-_50%	67	55	59	6	38	81	39	51	42	10	18	23	9	38 ± 13

be increased by approximately 14%. For the tight bounding box training, the overlap can be increased by approximately 19% and 18% for the bounding box training with random error. The label supervised trained network increased the overlap by approximately 22%. The standard deviation of the Jacobian determinant as well as the proportion of negative values are comparable for all trainings. It can be shown that a higher Dice score can be obtained for larger organs or for organs that initially already have a high overlap. The largest organ, the liver, for example, has the highest initial Dice overlap of 62%, and also the highest Dice overlap after registration for all variants (in a range of $81 - 85\%$). Organs with a small initial overlap, e.g. left adrenal gland (initial overlap 8%), also have a relatively low overlap after registration for all methods (in a range of $18 - 25\%$). For these organs, however, the Dice of weakly-supervised is higher than for supervised (e.g. left adrenal gland: 25% for weakly-supervised and 21% for supervised).

5 Discussion and Conclusion

We presented a deep-learning-based method for deformable image registration with weak bounding box supervision. We compared our method with an unsupervised and a label supervised training. The resulting registration of our method shows an improvement of about 5% for the Dice overlap in comparison to the unsupervised training. To simulate a realistic annotation of bounding boxes, we added an inter-observer-error of 5% per bounding box side, and showed that the quality of the result does not change significantly (approximately 1%) compared to tight bounding boxes. Organs with small initial overlap show the highest Dice score after the registration with the weak bounding box supervised network.

If the time for the labeling process was taken into account, so that less labels are available than bounding boxes, the accuracy of the label supervised training is less than for our bounding box supervision. Hence, for the purpose of medical image registration the proposed weak supervision strategy (labeling more images with lower effort) is beneficial.

References

1. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Isgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis **52**, 128–143 (2019)
2. Eppenhof, K.A., Pluim, J.P.: Pulmonary CT registration through supervised learning with convolutional neural networks. IEEE Trans Med Imag **38**(5), 1097—1105 (2018)
3. Sentker, T., Madesta, F., Werner, R.: Gdl-fire 4D: Deep learning-based fast 4D CT image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 765—773. Springer (2018)
4. Hansen L, Hering A, Heinrich MP, et al.. Learn2Reg: 2020 MICCAI registration challenge; 2020. https://learn2reg.grand-challenge.org.
5. Xu Z, Lee CP, Heinrich MP, et al. Evaluation of six registration methods for the human abdomen on clinically acquired CT. IEEE Trans Biomed Eng. **63**(8) 1563–1572 (2016).
6. Mok, T. C., Chung, A. C.: Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 211–221. Springer (2020)
7. Siebert, H., Hansen, L., Heinrich, M.: Architecture matters: evaluating design choices for deep learning registration networks. In Bildverarbeitung für die Medizin 2021. Springer Vieweg, Wiesbaden (2021)
8. Rajchl, M., Lee, M. C., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Rueckert, D.: Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. IEEE transactions on medical imaging, **36**(2), 674–683 (2016)
9. Hering, A., Kuckertz, S., Heldmann, S., Heinrich, M. P.: Memory-efficient 2.5 D convolutional transformer networks for multi-modal deformable registration with weak label supervision applied to whole-heart CT and MRI scans. International journal of computer assisted radiology and surgery, **14**(11), 1901–1912. (2019)
10. Heinrich, M.P., Jenkinson, M., Papiez˙, B.W., Brady, 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)