Transformed Grid Distance Loss for Supervised Image Registration

Xinrui Song[1], Hanqing Chao[1], Sheng Xu[2], Baris Turkbey[3], Bradford J. Wood[2], Ge Wang[1], and Pingkun Yan✉[1]

[1] Department of Biomedical Engineering and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
yanp2@rpi.edu
[2] Center for Interventional Oncology, Radiology & Imaging Sciences, National Institutes of Health, Bethesda, MD 20892, USA
[3] Molecular Imaging Program, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Abstract. Many deep learning image registration tasks, such as volume-to-volume registration, frame-to-volume registration, and frame-to-volume reconstruction, rely on six transformation parameters or quaternions to supervise the learning-based methods. However, these parameters can be very abstract for neural networks to comprehend. During the optimization process, ill-considered representations of rotation may even trap the objective function at local minima. This paper aims to expose these issues and propose the Transformed Grid Distance loss as a solution. The proposed method not only solves the problem of rotation representation but unites the gap between translation and rotation. We test our methods both with synthetic and clinically relevant medical image datasets. We demonstrate superior performance in comparison with conventional losses while requiring no alteration to the network input, output, or network structure at all.

1 Introduction

Existing deep learning-based image registration methods have explored many types of supervision. Unsupervised methods such as [1, 4, 11] relies on image intensity-based similarity metrics to supervise the network. These methods, however, are limited to single-modality registration tasks, or multi-modal images with very similar content and texture. Weakly supervised registration [2, 7] incorporated weak labels such as organ segmentation to guide the training process.

In contrast, supervised methods require the ground truth annotations of registration for training [3, 5]. For deformable image registration, providing such annotations can be unrealistically difficult. However, for tasks in which no significant differences were found between rigid and deformable registrations [10], using rigid registration reduces the annotation cost significantly. For example, in image-fusion guided prostate cancer biopsies, the manual registration between the MR and ultrasound images has been a routine for the clinical procedure. Requesting these manual registration labels for training come at no additional cost

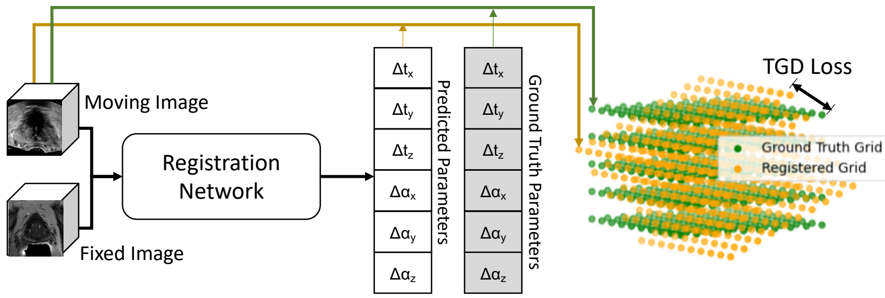

Fig. 1. Illustration of the transformed grid distance (TGD) loss.

to the clinicians. In other scenarios where deformable registration is preferred, a rigid transformation is also often required to pre-align the images before any deformable registration can be performed. For the above reasons, supervised deep learning based rigid image registration has been intensively studied, and will be the focus of this work.

Labels and loss function are critical components of supervised image registration. Since 3D rigid transformation is commonly represented by six transformation parameters, including three rotation angles and a 3D translation vector, a straightforward option is to use the distance between the ground truth and estimated transformation parameters as the loss to train the image registration network. However, numerous works [6, 8, 12] pointed out that the Euler angle representation is problematic for loss computation. In some cases, quaternion angles are used instead. In this paper, we argue that neither of them is the optimal choice for being used directly in a loss function. Instead, these abstract mathematical expressions should be first converted into more physically intuitive values. We propose a new loss – the Transformed Grid Distance (TGD) loss for network training.

2 Transformed Grid Distance

In supervised rigid registration, transformation parameters are often used as the label for network supervision. Compact transformation parameters, either in Euler or quaternion representation, can be difficult for neural networks to learn through conventional loss functions (e.g. L1 and MSE loss).

Instead of directly supervising the transformation parameters themselves, we apply the estimated transformation on the moving image grid, and supervise the distance between the transformed points and their corresponding points in the ground truth grid as illustrated in Fig. 1. Let $\boldsymbol{G} \in \mathbb{R}^{m\times n\times l}$ denote a 3D moving image grid. TGD loss is computed as

$$L_{TGD} = \left\| T_\theta(\boldsymbol{G}) - T_{gt}(\boldsymbol{G}) \right\|_2 , \tag{1}$$

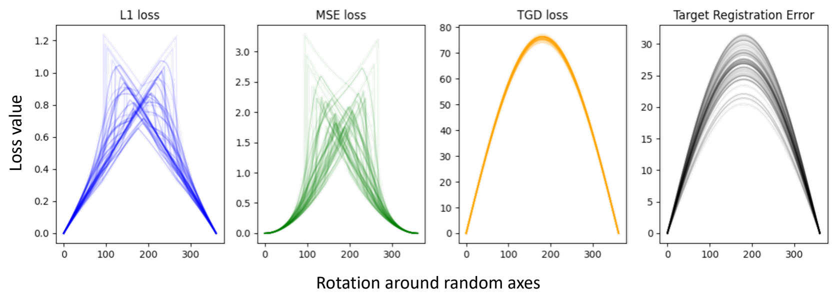

Fig. 2. We generated the rotations in this figure by randomly selecting 27 unit vectors and ranging the rotation amplitude from 0 to 360 degrees. The Target Registration Error serves as the evaluation metric, as in many registration tasks.

where T denotes a 3D transformation matrix converted from θ. The key difference here is to convert the abstract representation θ into a dense and intuitive representation, which guides the network optimization process through circumventing any non-linear transformation conversions that the network would otherwise have to figure out.

The proposed TGD loss elegantly unifies both rotation and transformation into point-wise distance, which results in a smooth loss landscape that guides the network learning process. Had more meaningful points been acquired (*i.e.* anatomical landmarks), the loss can be simply adapted into Target Registration Error (TRE) by replacing the grid with those clinically relevant points. One major weakness of the Euler angles is that they must be applied in a fixed order, which is not reflected at all with L1 or MSE loss. During training, each line from Fig. 2 can be regarded as a training sample. The loss curves for either L1 or MSE loss on Euler angles vary wildly from sample to sample, while the proposed TGD loss stays consistent with the Target Registration Error (TRE).

The quaternion system seems to be a better solution than the Euler angles. However, due to the fact that the quaternion expression is divided into two intertwined parts, it is hard to guarantee that the direction of optimization is at all correct. For example, slight error in the rotation axis would result in a large TRE regardless of the rotation angle.

3 Experiments

In this section, we present both a synthetic and a clinically relevant experiment. Our dataset consists of 528 manually labeled cases of MR-transrectal ultrasound (TRUS) volume pair for training, 66 cases for validation, and 68 cases for testing.

In the first experiment, we use an MR volume as the fixed image, and its own perturbed result as the moving image. We have also included the result of TRE-TGD loss, which is another version of the proposed method that replaces

Table 1. Performance of different loss functions in MR-MR registration.

Method	Mean TRE (mm)	Percentiles [25th, 50th, 75th, 95th]
Initial	12.66±7.30	[6.39, 12.83, 18.79, 23.88]
Quaternion loss	12.95±7.36	[6.64, 12.93, 19.00, 24.67]
MSE Euler angle loss	2.68±2.31	[1.19, 2.04, 3.43, 6.82]
L1 Euler angle loss	2.80±2.68	[1.04, 2.09, 3.72, 7.48]
TGD loss	**1.51±1.45**	[0.62, 1.15, 1.93, 3.85]
TRE-TGD loss	**1.50±1.53**	[0.65, 1.14, 1.87, 3.83]

Table 2. Performance of different loss functions in MR-TRUS registration.

Method	Mean TRE (mm)	Percentiles [25th, 50th, 75th, 90th]
Initial	9.93±5.87	[4.89, 9.82, 14.89, 19.10]
MSE Euler angle loss	5.57±2.86	[3.47, 5.06, 7.07, 10.98]
SRE-TGD loss	**4.40±2.49**	[2.57, 3.97, 5.77, 8.88]

the regular grid points in TGD loss with the target prostate surface points. The quaternion loss, on the other hand, failed to converge in this experiment where large rotation errors are concerned. Results in Table 1 show that simply through 'rephrasing' the transformation parameters into physical distance between grid points, the network was guided to converge at a lower minimum.

The second experiment treats the TRUS volume as the moving image, and the corresponding MR volume as the fixed image. This is a use case, where an accurate alignment between the transrectal ultrasound (TRUS) and MR volume greatly benefits the ultrasound-guided prostate cancer biopsy [9]. For each pair of MR and TRUS volume, we are provided with the manual label for rigid registration from TRUS to MR, as well as the prostate surface points in MR. Similar to the TRE-TGD loss in the first experiment, the SRE-TRD loss also calculates the distance between corresponding points, thereby a subset to the proposed TGD loss. Table 2 compares the result of multi-modal registration between the conventional MSE loss and SRE-TGD loss. With the same network architecture and other settings, the proposed loss function results in a significant (p <0.001 under paired t-test) improvement over the conventional MSE loss.

4 Discussions and Conclusion

In this paper, we revealed the limitation of directly using abstract transformation parameters for loss computation in supervised training of image registration networks. With such insight, we introduced a simple yet effective tool to boost the performance of supervised rigid volume registration. Although the analysis and experiments are mainly conducted in a rigid setting, this idea can be easily adapted for a non-rigid affine registration task.

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging **38**(8), 1788–1800 (2019)
2. Baum, Z., Hu, Y., Barratt, D.C.: Multimodality biomedical image registration using free point transformer networks. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis, pp. 116–125. Springer (2020)
3. Guo, H., Kruger, M., Xu, S., Wood, B.J., Yan, P.: Deep adaptive registration of multi-modal prostate images. Computerized Medical Imaging and Graphics **84**, 101769 (2020)
4. Hansen, L., Heinrich, M.P.: Graphregnet: Deep graph regularisation networks on sparse keypoints for dense registration of 3d lung cts. IEEE Transactions on Medical Imaging **40**(9), 2246–2257 (2021)
5. Haskins, G., Kruecker, J., Kruger, U., Xu, S., Pinto, P.A., Wood, B.J., Yan, P.: Learning deep similarity metric for 3d mr–trus image registration. International journal of computer assisted radiology and surgery **14**(3), 417–425 (2019)
6. Hou, B., Khanal, B., Alansary, A., McDonagh, S., Davidson, A., Rutherford, M., Hajnal, J.V., Rueckert, D., Glocker, B., Kainz, B.: 3-d reconstruction in canonical co-ordinate space from arbitrarily oriented 2-d images. IEEE transactions on medical imaging **37**(8), 1737–1750 (2018)
7. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., et al.: Weakly-supervised convolutional neural networks for multimodal image registration. Medical image analysis **49**, 1–13 (2018)
8. Kendall, A., Grimes, M., Cipolla, R.: Posenet: A convolutional network for real-time 6-dof camera relocalization. In: Proceedings of the IEEE international conference on computer vision. pp. 2938–2946 (2015)
9. Song, X., Guo, H., Xu, X., Chao, H., Xu, S., Turkbey, B., Wood, B.J., Wang, G., Yan, P.: Cross-modal attention for mri and ultrasound volume registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 66–75. Springer (2021)
10. Venderink, W., de Rooij, M., Sedelaar, J.M., Huisman, H.J., Fütterer, J.J.: Elastic versus rigid image registration in magnetic resonance imaging–transrectal ultrasound fusion prostate biopsy: a systematic review and meta-analysis. European Urology Focus **4**(2), 219–227 (2018)
11. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 204–212. Springer (2017)
12. Wei, W., Haishan, X., Alpers, J., Rak, M., Hansen, C.: A deep learning approach for 2d ultrasound and 3d ct/mr image registration in liver tumor ablation. Computer Methods and Programs in Biomedicine **206**, 106117 (2021)