Intensity-Based Wasserstein Distance As A Loss Measure For Unsupervised Deformable Deep Registration

Conference Paper · April 2021

DOI: 10.1109/ISBI48211.2021.9433818

4 authors, including:

Roozbeh Shams
Polytechnique Montréal
11 PUBLICATIONS **55** CITATIONS
SEE PROFILE

William Trung Le
University of Montreal Hospital Research Centre
13 PUBLICATIONS **38** CITATIONS
SEE PROFILE

Samuel Kadoury
Polytechnique Montréal
204 PUBLICATIONS **4,586** CITATIONS
SEE PROFILE

2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)
April 13-16, 2021, Nice, France

INTENSITY-BASED WASSERSTEIN DISTANCE AS A LOSS MEASURE FOR UNSUPERVISED DEFORMABLE DEEP REGISTRATION

Roozbeh Shams[⋆†]*, William Le*[⋆†]*, Adrien Weihs*[⋆]*, and Samuel Kadoury*[⋆†]

[⋆] MedICAL Laboratory, Polytechnique Montréal, Montréal, Canada
[†] Centre de recherche du CHUM (CRCHUM), Montréal, Canada

2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-1246-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISBI48211.2021.9433818

ABSTRACT

Traditional pairwise medical image registration techniques are based on computationally intensive frameworks due to numerical optimization procedures. While there is increasing adoption of deep neural networks to improve deformable image registration, achieving a clinically suitable solution remains scarce. One of the primary difficulties lies in the choice of tractable distance functions to assess image similarity. Recent works have explored the Wasserstein distance as a loss function in generative deep neural networks. In this work, we evaluate a fast approximation variant — the sliced Wasserstein distance — for deep image registration of brain MRI datasets. Based on a VoxelMorph backbone architecture, which includes a combination of UNet and spatial transformer networks (STN) for deformable registration, we propose three implementation variants to compare the model's performance: the standard sliced Wasserstein, the Radon transform performing a low dimensional embedding, and a novel patch-based method that allows fine-grained deformation comparison. Experiments performed on public datasets of brain images from the Learn2Reg open challenge demonstrate the Wasserstein methods converge faster than the baseline mean square error method, with the proposed patch-based method yielding similar performance to baseline methods, and improved overall accuracy compared with other implementations. This makes the sliced Wasserstein a valuable metric for deep mono-modal and multi-modal deformable medical image registration problems with our proposed implementation.

Index Terms— Medical image registration, Deformable registration, Optimal transport, Wasserstein distance, Brain MRI, Deep neural networks.

1. INTRODUCTION

The demand for faster registration techniques has motivated the development of one-step deep learning registration approaches, estimating the non-linear transformations between pairs of images. However, quantifying dense deformation fields due to changes in morphology or structural features between pairs of images remains a difficult and open problem in medical image analysis. One of the challenges associated with this framework is the well established problem of the image similarity metric.

Traditional registration approaches require a computationally intensive optimization step to align each image pair. This is slow at prediction time, but also due to the large feature space that must be determined. Recently, several works with convolutional neural networks (CNNs) have attempted to estimate the optimal registration parameters [1, 2, 3]. CNNs are especially well-suited for this task as they can explore a large parameter space and automatically learn the set of hierarchical registration features. Furthermore, as opposed to optimization-based methods, CNNs take advantage from large datasets and perform alignment prediction very quickly; indeed, once training is completed, the optimization does not need to be recomputed. This is particularly relevant for interventional use such as in radiation oncology or surgical guidance where time constraints are elevated and can lead to improved efficiency.

Deep image registration often necessitates unsupervised models, since medical images often do not possess the required labels — deformation parameters or fields — which are required for supervised training. Moreover, clinical datasets with paired and ground-truth registration of multi-modal images are scarce. In most cases, the choice of an appropriate loss function, based on an image similarity metric, is paramount to the performance of the model. Image comparison is non-trivial, more-so as a metric. Mean-squared error (MSE), mutual information (MI) or normalized cross-correlation (NCC) are popular metrics, which are easy to implement. Being non-specific to image registration, it is based on pixel-wise differences. However, it is highly susceptible to intensity-based perturbations between image pairs and does not work across modalities.

Recently, the Wasserstein distance has seen a surge in popularity in training deep neural networks. Also referred to as the earth mover's distance, it has been traditionally difficult to implement, having no closed-form solution in high dimensional cases. Recent work with Wasserstein GAN [4] provided a method to estimate the measure, using the image pairs directly as distributions, showing the method was suitable as a loss function for training a deep generative model. In this work, we implement the Wasserstein loss as part of a deep image registration model and compare the performance of three implementations to the standard registration loss of MSE on brain MRI datasets.

1.1. Related work

Early methods for end-to-end unsupervised deformable registration used a combination of a regression CNN that outputs transformation parameters [5], using the Spatial Transformer Network (STN) [6] and an image resampler. This allowed to train a CNN for deformable image registration directly. Following this work, the DLIR framework was introduced [7] which stacked multiple CNN-STN-resampler modules to perform affine and deformable, coarse to fine-grained registration. In recent years, VoxelMorph [8] has emerged as one of the leading frameworks, replacing the CNN module with a standard UNet [9], used for image segmentation. With this method, it allows the model to learn an output of deformation fields directly, allowing a greater flexibility in possible transformations. Kuang and Schmah [10] noted however that such flexibility comes at the cost of allowing non-invertible "folding" deformations. To address this, they introduced a novel penalty loss based on negative Jacobian determinants. Hu et al. improved on the VoxelMorph model by intro-

ducing a dual-stream architecture [11], allowing training registration at multiple scales. Finally, Shen et al. combined both affine and deformable registration networks, the later using a vector momentum-parametrized stationary velocity field model instead of the usual deformation vector field [12].

2. MATERIALS AND METHODS

We propose an end-to-end framework for deep 3D MRI registration integrating a modular Wasserstein distance function, allowing to evaluate the effects of different Wasserstein distance variants. The architecture, as shown in Figure 1, is based on VoxelMorph [9], in which a U-Net is trained to generate a deformation vector field (DVF), used then by the spatial transformer network (STN) [6] to deform the input volume $\mathcal{M}$ into $\mathcal{D}$ = STN(DVF, $\mathcal{M}$), while minimizing the difference between the moving and the target image $\mathcal{F}$ using one of the proposed Wasserstein distances as a loss function.

2.1. Optimal Transport and Wasserstein Distance

While one natural strategy would be to use the transport map to find an ideal registration map, this approach is very computationally intensive. The idea behind trying to circumvent the difficulty of computing the optimal transport map involves minimizing the Wasserstein distance. This type of approach has been widely used in several works, first in [13] and later on in several GAN implementations as it greatly increases the stability of training [14]. The typical implementation of the Wasserstein loss follows [4], which relies on the following duality result, called the Kantorovitch-Rubinstein formula:

$$\min_{\pi} \int c(x,y)\, d\pi(x,y) = \sup \left\{ \int_X f(x)\, d(\mu-\nu) \mid f \in L^1(|\mu-\nu|),\ \ \|f\|_{\text{Lip}} \leq 1 \right\}. \tag{1}$$

with $c(x,y)$ as the cost function measuring distributions x and y in domain π, $f\|_{\text{Lip}}$ as the differentiable Lipshitz function and L the Lebesgue measure. The above implies that given two point clouds as described above by μ and ν, we can simply sort the points and directly obtain their Wasserstein distance. One dimensional optimal transport exhibits a unique feature since it allows for closed form computations of the optimal transport for n number of samples:

$$d_{W^p}(\mu,\nu)^p = \frac{1}{n}\sum_{i=1}^{n} |x_i - y_i|^p. \tag{2}$$

where x_i and y_i are the sorted points in clouds μ and ν, respectively.

The sliced Wasserstein distance then aims to take advantage of the computational advantage of the 1D case: as seen previously, a closed-form formula exists to facilitate the computation. Following this intuition, for $\mu = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ and $\nu = \frac{1}{n}\sum_{i=j}^{n} \delta_{y_j}$, we define the distance as:

$$d_{SW^p}(\mu,\nu)^p = \int_{\{|\theta|=1,\ \theta\in\mathbb{R}^m\}} d_{W^p}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i\cdot\theta}, \frac{1}{n}\sum_{j=1}^{n}\delta_{y_j\cdot\theta}\right)^p d\theta. \tag{3}$$

which is differentiable given the properties from Radamacher's theorem, where if the ensemble of functions is an open subset and f is Lipschitz continuous, then f is differentiable almost everywhere, meaning points in the ensemble at which f is not differentiable form a set of Lebesgue measure zero. The integral defined in domain θ can be computed through the closed-form formulas, with δ as Dirac functions. For the more general formula, one can describe the projection through Radon transforms [15].

2.2. Wasserstein Distance Approximation

As the Wasserstein distance is computationally expensive to calculate, we used an approximation of the real distance to reduce the inference timings. The three approaches described here are based on the sliced Wasserstein distance, projecting the N-dimensional measure to one dimension and use the closed form equation for calculating the 1D Wasserstein distance to approximate the distance.

In sliced Wasserstein (SWD), the Wasserstein distance is approximated by projecting the N-dimensional distribution of random vectors on the unit sphere. Then, the closed form 1D Wasserstein solution is used to calculate the distance of the projected samples. The approximation is obtained by taking the mean over the projected distances. In the Radon implementation [16], the N-Dimensional distribution is projected into 1D distributions. However as the name suggests, the distribution is instead summed over the vectors with different angles to obtain the distance approximation. In the patch-based sliced Wasserstein (PSWD), the volume is divided into patches and SWD is calculated between every patch pairs. The metric is obtained as the average SWD of the patches. The idea behind this approach is to consider finer scale similarities. This also helps alleviate the computational complexity due to smaller input pairs as:

$$d_{PSW}(\mathcal{F},\mathcal{M}) = \frac{1}{n}\sum_{i=1}^{n} d_{SW^p}(P_i^{\mathcal{F}}, P_i^{\mathcal{M}}) \tag{4}$$

where $\mathcal{F}$ and $\mathcal{M}$ are the fixed and moving volumes, $P_i^{\mathcal{F}}$ and $P_i^{\mathcal{M}}$ are corresponding patches of $\mathcal{F}$ and $\mathcal{M}$, and n is the total number of patches.

2.3. Datasets

For this work, we evaluated the models on brain MRI datasets, obtained from the publicly available Learn2Reg image registration challenge [17]. The first dataset consisted 90 healthy patients and 105 with non-affective psychotic disorder. The non-healthy portion of the dataset was provided by the Psychiatric Genotype/Phenotype Project data repository at Vanderbilt University Medical Center (Nashville, TN, USA). All patients were adults with MR imaging and expert segmentations of the hippocampus shape. MRI were acquired using the Philips Achieva scanner (Philips Healthcare, Inc., Best, The Netherlands) using a 3D T1-weighted MP-RAGE sequence (TI/TR/TE, 860/8.0/3.7 ms; in-plane resolution of 256 x 256; 170 sagittal slices; 1.0 mm^3 voxel size). The hippocampus segmentation mask was available on the MRI, with the input volumes normalized to zero mean and unit standard deviation.

The second dataset was a multimodal brain image dataset of 22 adult patients with low-grade gliomas (Grade II), with intra-operative 3D ultrasound (US) using a 6-12 MHz linear probe, and pre-operative T1-w MRI. The MRI was acquired with both 1.5T and 3T Siemens MR scanners, with an in-plance resolution of 256 x 256, and pixel size of 1.0mm^3. Each image had sets of corresponding expert landmark annotations to evaluate registration performance. Images were initially rigidly aligned using external fiducials.

The first dataset was split into training, validation and testing split using 65%, 15% and 20% cases respectively, with a 5-fold

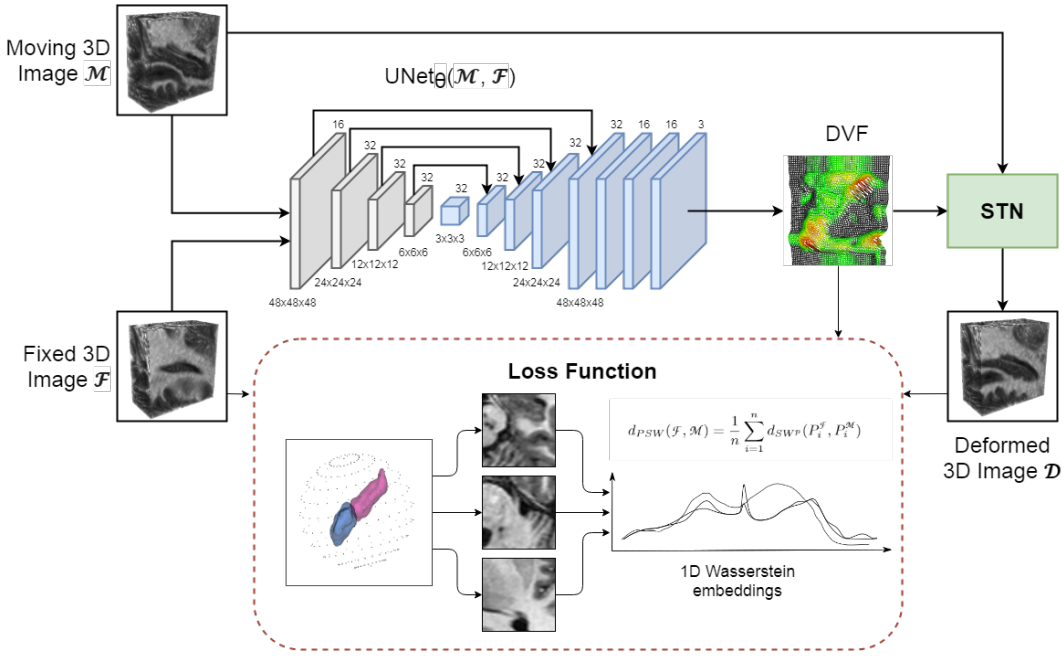


Fig. 1: Overview of the deep registration framework with the proposed novel patch-based sliced Wasserstein loss.

cross-validation scheme was used for the second. Within each split, registration pairs were obtained via random selection between cases. In terms of data preprocessing for Voxelmorph, the volumes were padded to a uniform shape $64 \times 64 \times 64$, before being cropped to the contoured ROI and downscaled to $48 \times 48 \times 48$ for computational purposes.

3. RESULTS AND DISCUSSION

The VoxelMorph model was trained and evaluated with each of the described Wasserstein approximations (SWD, Radon, PSWD) as well as MSE, used as a baseline comparison of the image registration performance. Model selection was performed using a hyperparameter grid search on a GeForce RTX 2080 Ti GPU. The models were trained using the Adam optimizer at a learning rate of 1×10^{-3} for 100 epochs regularized by early stopping with a patience of 5 and delta of 1×10^{-4}, while minimizing the proposed loss functions. The sliced Wasserstein based methods used 400 projections and an embedding size of 50. For the patch-based method, the patches were generated with size $11 \times 11 \times 11$ with a stride of 5. In the case of the first dataset, to establish the registration performance on the deformed hippocampus segmentation, the Dice score, representing the organ overlap, and the 95% Hausdorff distance based on the mean edge distance, were used as auxiliary metrics. For the MRI-US dataset, evaluations were based on mean target registration errors (mTRE) on manual landmarks.

Figure 2 shows qualitative results of the registration results. Overall, the deformed hippocampus matches the fixed image, with sub-region volumes being well preserved. However, edge structures indicated by the dark orange fiducial markers show that the MSE and PSWD methods perform the correct deformation, whereas the Radon method fails to do so. The light yellow marker shows that the PSWD performs the correct flattening deformation, with a less pronounced effect for both MSE and Radon.

Table 1 shows the performance results of the model for the baseline MSE and each of the proposed Wasserstein implementations, as well as the number of epochs before convergence was reached. Convergence time, as triggered by the early stopping mechanism, was achieved earlier for SWD ($33 < 42$), Radon ($36 < 42$ epochs) and PSWD ($32 < 42$ epochs) which converged the fastest compared with the baseline MSE. This shows the benefit of the Wasserstein distance as a metric, with a faster optimization as compared to other methods.

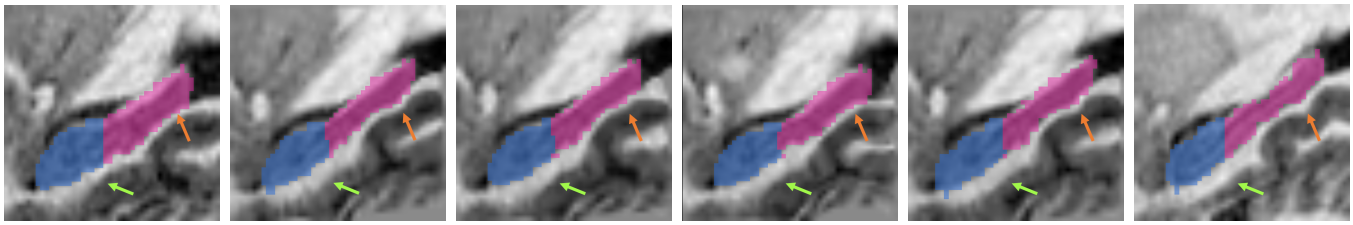
Table 1: Auxiliary registration metrics on the validation set for the mono-modal dataset.

Loss Function	End epoch	HD95	Dice score
MSE [8]	42	3.20 ± 1.16	$\mathbf{72.33 \pm 0.06}$
SWD	33	3.37 ± 1.12	71.32 ± 0.06
Radon	36	3.52 ± 1.22	68.45 ± 0.06
PSWD	**32**	$\mathbf{3.19 \pm 1.08}$	71.98 ± 0.05

Table 2: Auxiliary registration metrics on the test set for the mono-modal dataset.

Loss Function	HD95	Dice score
MSE [8]	3.14 ± 1.47	72.83 ± 0.11
SWD	3.43 ± 1.45	69.72 ± 0.12
Radon	3.46 ± 1.36	68.48 ± 0.10
PSWD	$\mathbf{3.08 \pm 1.25}$	$\mathbf{73.11 \pm 0.10}$

Table 2 shows the performance on the test set. Both previous implementations of sliced Wasserstein showed inferior hippocampus registration performance compared with MSE baseline, with SWD achieving a HD95 of 3.43 ± 1.45mm and Dice score of 69.72 ± 0.12, while the Radon approach yielded a HD95 of 3.46 ± 1.36mm



(a) Moving image (b) MSE (c) SWD (d) Radon (e) PSWD (f) Fixed image

Fig. 2: Sample registration results with the moving and fixed images, as well as sample deformed images for each of the compared distance metrics. Images shown are axial center slice of the hippocampus volume. Blue mask: anterior hippocampus. Pink mask: posterior hippocampus.

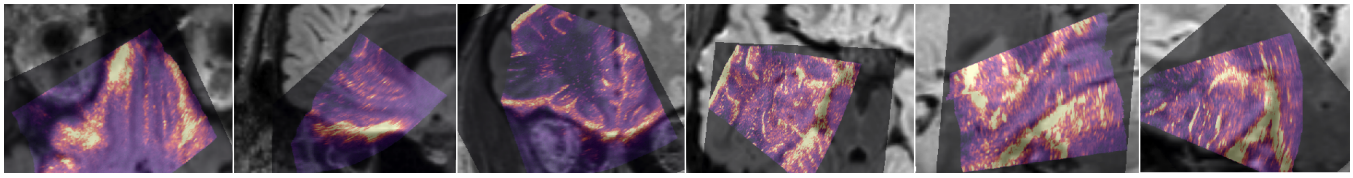


Fig. 3: Sample multi-modal registration with MRI and 3D US images obtained with the PSWD metric used in the framework.

and Dice score of 69.48 ± 0.10. On the other hand, the PSWD version showed the best overall performance compared with baseline MSE, with a HD95 of 3.08 ± 1.25mm and a Dice score of 73.11 ± 0.10. Overall, previous state of the art implementations of the sliced Wasserstein distance underperforms compared with the standard MSE method. However, our proposed metric performs better on both auxiliary metrics, effectively capturing the approximate latent distance between unpaired images. The fine-grained detail comparisons offered by the patch-based approach benefits the parameter search capability of the CNN over the standard global image as a distributions embedding.

For the multi-modal dataset, the proposed Wasserstein method demonstrates an improvement in internal landmark localization errors, yielding a mTRE for PSWD of 3.5 ± 1.9mm, compared to an initial mTRE of 5.3 ± 4.2mm. This compares favourably to the mTRE of the MSE (4.4 ± 2.7mm), Radon (4.5 ± 2.9mm) and SWD (4.3 ± 2.8mm) approaches. Figure 3 shows qualitative results of the multimodal dataset, with the MRI non-rigidly aligned with the US.

4. CONCLUSIONS

In this work, we propose a deep deformable registration network integrating different implementation variants of the sliced Wasserstein distance used as an image similarity metric. The VoxelMorph framework was trained for brain MRI registration, with evaluations based on structural Dice coefficients and registration errors. Results showed that compared with the tradition MSE loss, the Wasserstein metrics helped the model to converge faster. The proposed patch-based method also yielded the best registration performance from all the Wasserstein variants, whereas the Radon implementation showed the worst performance. This demonstrates that the sliced Wasserstein distance could be a powerful and efficient metric for deep multi-modal image registration, and that specific implementation variants may affect performance.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by the Grand Challenge [17]. Ethical approval was not required as confirmed by the license attached with the open access data.

6. ACKNOWLEDGMENTS

This research has been funded in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors have no relevant financial or non-financial interests to disclose. There were no conflicts of interests.

7. REFERENCES

[1] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 729–738.

[2] Hongming Li and Yong Fan, "Non-rigid image registration using self-supervised fully convolutional networks without training data," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1075–1078.

[3] Xiaohuan Cao, Jianhuan Yang, Li Wang, Zhong Xue, Qian Wang, and Dinggang Shen, "Deep learning based inter-modality image registration supervised by intra-modality similarity," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2018, pp. 55–63.

[4] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville, "Improved training of wasserstein gans," *arXiv preprint arXiv:1704.00028*, 2017.

[5] Bob D de Vos, Floris F Berendsen, Max A Viergever, Marius Staring, and Ivana Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural net-

work," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 204–212. Springer, 2017.

[6] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al., "Spatial transformer networks," in *Advances in neural information processing systems*, 2015, pp. 2017–2025.

[7] 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*, vol. 52, pp. 128–143, 2019.

[8] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[10] Dongyang Kuang and Tanya Schmah, "Faim–a convnet method for unsupervised 3d medical image registration," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2019, pp. 646–654.

[11] Xiaojun Hu, Miao Kang, Weilin Huang, Matthew R Scott, Roland Wiest, and Mauricio Reyes, "Dual-stream pyramid registration network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 382–390.

[12] Zhengyang Shen, Xu Han, Zhenlin Xu, and Marc Niethammer, "Networks for joint affine and non-parametric image registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4224–4233.

[13] Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya, and Tomaso A Poggio, "Learning with a wasserstein loss," in *Advances in neural information processing systems*, 2015, pp. 2053–2061.

[14] Martin Arjovsky, Soumith Chintala, and Léon Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[15] Soheil Kolouri, Gustavo K Rohde, and Heiko Hoffmann, "Sliced wasserstein distance for learning gaussian mixture models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3427–3436.

[16] Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo Rohde, "Generalized sliced wasserstein distances," in *Advances in Neural Information Processing Systems*, 2019, pp. 261–272.

[17] 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," Mar. 2020.