CT Scan Registration with 3D Dense Motion Field Estimation Using LSGAN

Essa R. Anas(✉), Ahmed Onsy, and Bogdan J. Matuszewski

Computer Vision and Machine Learning (CVML) Group, School of Engineering, University of Central Lancashire, Preston PR1 2HE, UK
{EAnas, AOnsy, BMatuszewskil}@Uclan.ac.uk

Abstract. This paper reports on a new CT volume registration method, using 3D Convolutional Neural Networks (CNN). The proposed method uses the Least Square Generative Adversarial Network (LSGAN) model consisting of the Contraction-Expansion registration network as the LSGAN's generator and a deep 3D CNN classification network as the LSGAN's discriminator. The training of the generator is performed first on its own, using Charbonnier and smoothness loss functions, with progressive weights update moving from lower to higher resolution layers of the Expander. Subsequently, the complete network (Contraction-Expansion with the Discriminator) is trained as a LSGAN network. For the training, CREATIS and COPDgene datasets have been used in a self-supervised paradigm, using 3D warping of the moving volume to estimate the error with respect to the reference volume. The input to the network has $256 \times 256 \times 128 \times 2$ voxels and the output is displacement field of $128 \times 128 \times 64 \times 3$ voxels. The Contraction-Expansion registration network, on its own, achieves mean error of 1.30 mm with 1.70 standard deviation (SD) on the DIR-LAB dataset. When the whole proposed LSGAN network is used, the mean error is further reduced to 1.13 mm with 0.67 (SD). Therefore, the use of the GAN paradigm reduces the mean error by approximately 15%, providing the state-of-the-art performance.

Keywords: Image registration · Convolutional neural network · Generative adversarial network

1 Introduction

Image registration is an essential step in the Radiotherapy workflows. For many patients, multiple CT scans are performed during and after treatment. Common deformable image registration algorithms require cost function optimization between any two volumes [1]. This means that for any new registration, the algorithm needs to go through computationally demanding and time-consuming optimization process. These problems could cause a bottleneck for some radiation therapy (RT) workflows including treatment planning. Clinical applications of the 4D CT data registration including semi-automated target volume and organ at risk contour propagation; assessment of motion effects on dose distributions (4D RT quality assurance, dose warping) [2] and 4D CT-based lung ventilation estimation and its incorporation into

B. W. Papież et al. (Eds.): MIUA 2020, CCIS 1248, pp. 195–207, 2020.
https://doi.org/10.1007/978-3-030-52791-4_16

RT treatment planning [3]. Image registration plays also vital role in computer aided diagnosis pipelines, radiation treatment planning, and image-guided interventions. For all these applications, it is valuable to quantify the registration error locally. Image registration is often validated using 'overlap' measures (as Dice index of organs overlap). The most popular way to determine registration accuracy, however, is the Target Registration Error (TRE) on corresponding points in the registered images [4]. These points are commonly relevant anatomical landmarks annotated by experts. For deformable registration problems, these landmarks should cover the entire region of interested to be accurate descriptors of the local registration.

More recently, many researchers are tackling the medical image registration utilizing convolutional neural network. The authors in [4] presented a supervised approach using sliding window to directly estimate the registration error using synthetically deformed CT images as well as publicly available dataset containing landmarks annotations. Yang et al. [5] also used sliding window technique to estimate the deformation field for 3D brain DIR. The main issue that could be mentioned about the patch-based learning is the lack of the global information about the transformation that may not be adequately represented in small motion displacement but could appear in large displacement cases as in the CT scans. In the context of Generative Adversarial Network (GAN), researchers perform multimodal registration using GAN [6], Wasserstein GAN (WGAN) [7], InfoGAN [8], and Cycle GAN [9, 10].

In [6] vanilla GAN is utilized to register Magnetic Resonance (MR) and 3D intra-procedural transrectal ultrasound (TRUS). They used supervised learning to train a discriminator part of the network by providing the output of the generator and a simulated motion field. Authors in [11] developed a probabilistic generative model and demonstrate their VoxelMorph network on a 3D brain registration task.

In this paper, the deformable registration problem in performed using 3D CNN. The required deformation is achieved using motion vectors to estimate the per-pixel displacement estimation in 3D using both local and global representations by training with the full size of the CT volume of $256 \times 256 \times 128$ voxel. Then a Bayesian version of the model is introduced by converting it to probabilistic architecture [12]. This conversion achieved by adding a discriminator to the network to be train it in a GAN paradigm. The output of the model represents the 3D motion vectors of $128 \times 128 \times 64 \times 3$ dimension. Which consists of three channels that required to compensate for the deformation of the moving volume with respect to a reference volume. Unlike [6] the discriminator trained with the CT volumes to during the network regularization. Introducing the LSGAN to the network improved the training and increased the performance of the network on the test set.

2 Method

In this paper, the deformable image registration is performed by designing the Transformation function to perform the mapping from the moving image to the reference image. Let $I_F(\boldsymbol{x}) : \Omega_F \rightarrow \mathbb{R}$ and $I_M(\boldsymbol{x}) : \Omega_M \rightarrow \mathbb{R}$ be two real value images defined on their corresponding spatial domains $\Omega_F \subset \mathbb{R}^3$ and $\Omega_M \subset \mathbb{R}^3$ respectively.

The task is to find the function $\widehat{\boldsymbol{T}} : \Omega_M \rightarrow \Omega_F$, $\widehat{\boldsymbol{T}}(\boldsymbol{x}) = \boldsymbol{x} + \widehat{\boldsymbol{U}}(\boldsymbol{x})$, mapping pixels in the moving domain Ω_M to their corresponding pixels in the reference domain Ω_F. In the non-rigid registration framework this function is usually estimated through solving following optimization problem:

$$\widehat{\boldsymbol{U}} = \arg \min_{\widetilde{\boldsymbol{U}}} \left(sim(I_F, I_M(\widetilde{\boldsymbol{U}})) + \beta reg(\widetilde{\boldsymbol{U}}) \right) \tag{1}$$

where $sim()$ is a so called fidelity term depending on the observed data and $reg()$ is so called regularization term which reflects known or assumed properties of the displacement filed $\widehat{\boldsymbol{U}}$, typically encoding information about some form of $\widehat{\boldsymbol{U}}$ smoothness.

The quality of the registration, i.e. estimated function $\widehat{\boldsymbol{T}}$, can be assessed in a number of ways, in this paper the Target Registration Error (TRE) [13] is used. The TRE measure the displacement (here using the Euclidean distance) of $\widehat{\boldsymbol{T}}(\boldsymbol{x})$ from the true positions $\boldsymbol{T}(\boldsymbol{x})$ of the registered points [4]:

$$TRE : \Omega_F \rightarrow \mathbb{R}^+ : \boldsymbol{x} \rightarrow \left\| \boldsymbol{T}(\boldsymbol{x}) - \widehat{\boldsymbol{T}}(\boldsymbol{x}) \right\| \tag{2}$$

In practice the true positions are not know and manually annotated set of target points (landmarks) is used as a surrogate of these true positions. Therefore, manually annotated corresponding landmarks $\boldsymbol{p_F} \in \Omega_F$ and $\boldsymbol{p_M} \in \Omega_M$ are used and the TRE measure Euclidean distance between $\boldsymbol{p_F}$ and the relocated, by the $\widehat{\boldsymbol{T}}$, $\boldsymbol{p_M}$ points.

2.1 Network

The implementation of the convolutional network for this deformable image registration consists of five 3D convolutional layers with down-sampling in the contractive part with the expander part consisting of up-sampling followed by corresponding 3D convolutional layer, see Table 1. At each layer of the contractive part, the features are down-sampled across the three dimensions, x, y, and z. Similarly, at the expander part the features are up-sampled at each layer across the three dimensions. The input to the network consists of the two volumes the reference and the moving volumes with volume size of $128 \times 256 \times 256$ as input size and the output size of $64 \times 128 \times 128$ pixels, which is up-sampled later for the TRE computations. The size of the output has been chosen due to the limited computational resources. The activation function utilized in this implementation is leaky ReLU. At each layer of the expander a reconstruction sampler is attached to map the moving volume to the reference volume using the transformation or the displacement function that the convolutional network is learning. This displacement consists of three maps corresponding to the x, y, and z components of the transformation function.

The discriminator consists of five blocks of VGG network layers [14] with input of $64 \times 128 \times 128 \times 2$. The layer block consists of 3D convolution, down-sampling, batch normalization, and then leaky ReLU activation function. At the end of the discriminator there are two Fully Connected Neural Network (FCNN) layers. In addition to the activation between these FCNN layers, dropout layers with 50% dropout

Table 1. The details network configuration. (a) Contractive part. (b) Expanding part. (c) Discriminative part.

Contractive Part				
Layer	Input	Output	No. of Features	Downscale
Conv1	256x256x128x2	128x128x64x32	32	2
Conv2	128x128x64x32	64x64x32x64	64	2
Conv3	64x64x32x64	32x32x16x128	128	2
Conv4	32x32x16x128	16x16x8x256	256	2
Conv5	16x16x8x256	8x8x4x512	512	2

(a)

Expanding Part					
Layer	Input	Skip-Conn	Output	No. of Feature maps	Upscale
Up-Conv1	8x8x4x512	16x16x8x256	16x16x8x256	256	2
Up-Conv2	16x16x8x256	32x32x16x128	32x32x16x128	128	2
Up-Conv3	32x32x16x128	64x64x32x64	64x64x32x64	64	2
Up-Conv4	64x64x32x64	128x128x64x32	128x128x64x32	32	2
Up-Conv5	128x128x64x32		128x128x64x3	3	1

(b)

Discriminator				
Layer	Input	Output	No. of Feature maps	Downscale
Conv1	128x128x64x2	64x64x32x8	8	2
Conv2	64x64x32x8	32x32x16x16	16	2
Conv3	32x32x16x16	16x16x8x32	32	2
Conv4	16x16x8x32	8x8x4x64	64	2
Conv5	8x8x4x64	4x4x2x128	128	2
FCNN	4x4x2x128	64	64	
Drop-out				
FCNN	64	16	16	
Dropout				
FCNN	16	1	1	

(c)

rate are attached. Then the discriminator is terminated with a regression layer to classify the input. The input to the network can be either: the reference volume concatenated with the same volume after adding uniform noise in a range between $[-0.01, 0.01]$ or a reference volume concatenated with the warped moving volume. In the context of GAN model, the real class corresponds to the reference volume concatenated with its noisy version, while the fake class is represented by the reference volume concatenated with the moving volume. As the discriminator is starting to classify the fake input as a legitimate input, the generator performance improved. A simplified diagram representing the proposed architecture is shown in Fig. 1.

2.2 Training Procedure

The training of the network consists of two phases; the Contractive-Expander phase and the GAN phase. The Contractive-Expander training phase includes the training of each convolutional layer of the Expander using 5000 iterations. It starts at the lowest resolution and gradually involving higher resolution levels of the expander, in each case trained using 5000 iterations. To estimate the error between the reconstructed or warped moving volume and the reference volume the Generalized Charbonnier (GC) penalty function $\rho(x) = \left(||x||^2 + \epsilon^2\right)^{\alpha/2}$ [15] is considered as the fidelity term, see Eq. (3). This loss function is often used for optical flow and depth estimation. The function behaves as an L2 loss function close to the zero, to encourage smoothness, and L1 otherwise, to encourage robustness against outliers. It is also sometimes called L1-L2 loss function.

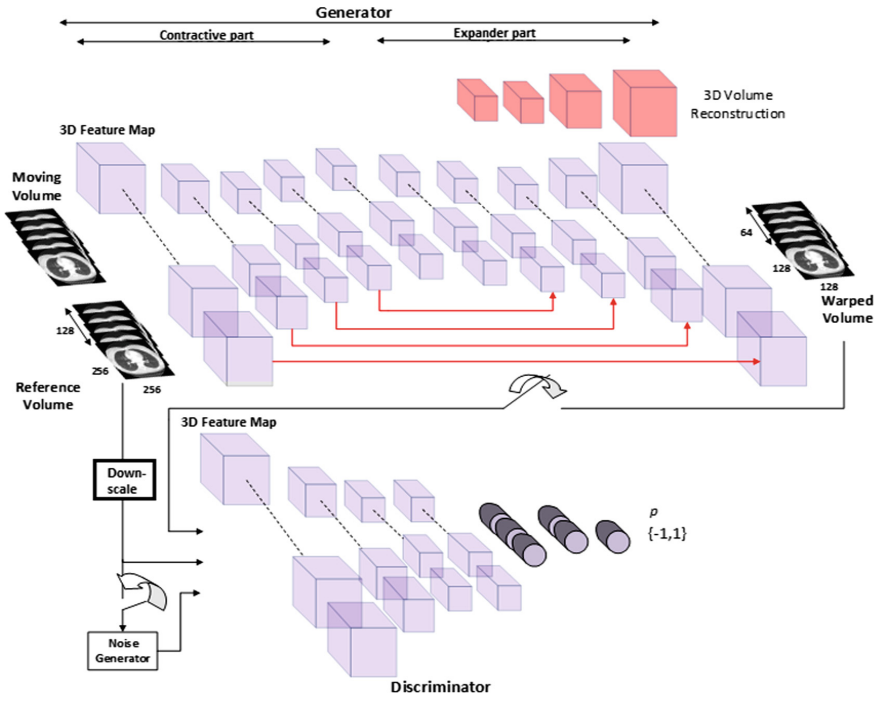


Fig. 1. A simplified diagram representing the proposed network architecture. It consists of two parts: the contractive-expander as generator and the discriminator.

The function is defined as:

$$sim = \ell_{GC} = \sum_{i,j,k} \left(\left(I_F(i,j,k) - I_{Mwarped}(i+u, j+v, k+l) \right)^2 + \epsilon^2 \right)^{\alpha/2} \quad (3)$$

where α and ϵ are chosen experimentally to be 0.7 and 0.0001 respectively, and u, v, and l are elements of the displacement field $\widetilde{\boldsymbol{U}}$. The adopted fidelity term implies assumption that the true corresponding points in both volumes have the same intensity. However, this assumption is not exactly true in case of lung CT data, which can change intensity values of the corresponding points due to changing air volume in the lungs in the inhale and exhale phases. In the follow on work a suitable intensity correction will be considered, which can be learned independently or as a function of the local Jacobians. In the current registration model, it is hypothesized that the adopted GAN implementation may implicitly compensate for that fidelity term modeling inaccuracy.

In addition to this loss function, while acting globally on the volume, the effect on regions with less image structure could create unambiguity in the displacement vector estimation (aperture problem). To address this ambiguity, smoothing function that can minimize the multiple velocity scoring area is utilized, Eq. (4).

$$reg = \ell_S(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{l}) = \sum\nolimits_{W,H,D} \rho(\nabla \widehat{\boldsymbol{U}}) \tag{4}$$

where $\widehat{\boldsymbol{U}}$ is the estimated deformable displacement and the loss function estimated across the volume (W, H, D). ρ is the realization of the GC function that is applied on these smooth surfaces like the organs with insufficient image structure. This function will also enhance the smoothness of the moving voxels and close any gap within the representation of the displacement across the organ's boundaries. The total loss estimation is performed using the following weighting:

$$\ell_{Total} = \ell_{GC} + \beta \times \ell_S \tag{5}$$

The warping of the moving volume is performed using the method reported in [16] as in Fig. 2. To reconstruct the volume, a bilinear interpolation.

The next training phase (GAN training) starts while training the last layer of the generator part of the GAN network. In this phase the discriminator starts to be involved in the training. The dropout layers in the discriminator were quite effective in stabilizing the discriminator during the training of the model and mitigate against overfitting. The last layer is left without activation and the results value is compared to (1) in the case of the real image and (−1) in the case of fake image using LSGAN.

With the Least Square GAN, the discriminator is trained following loss function of Eq. (6) [17]:

$$\min_D V_{LSGAN}(D) = \frac{1}{2}\left(E_{x\sim P_{data}(x)}\left[(D(x)-1)^2\right] + E_{z\sim P_z(z)}\left[(D(G(z))+1)^2\right]\right) \tag{6}$$

here, $D(x)$ is the prediction of the discriminator when the input is the true reference volume, while $D(G(z))$ is the prediction of the discriminator when the input is the fake (which is the generator output or the warped moving volume $I_{Mwarped}$). Simultaneously the target of the generator is to learn the distribution $P_z(z)$ by sampling the input variable z from the dataset distribution and map it though differentiable network. The loss function to train the generator to perform the above-mentioned function for this particular GAN is defined as in Eq. (7):

$$\min_{G} V_{LSGAN}(G) = \frac{1}{2} E_{z \sim P_z(z)} \left[(D(G(z)))^2 \right] \tag{7}$$

Using this arrangement, the model continued in its training using the optimization provided by the LSGAN network. It is worth mentioning that during the inferencing stage the discriminator network is stripped and only the generator network is involved in the operation.

During the training, an Adam optimizer is adopted with a starting learning rate of 0.0001 that reduced after each 5000 iteration by a factor of 1.4. Each 5000 iterations, a layer weights of the expander part are updated starting from the lowest layer and going to the highest layer. The loss function is calculated using Eqs. (5) during this phase of training. When the training reaches the last layer of the expander part, the discriminator joins the training and Eqs. (5) and (7) are utilized during the training of the generator in the following fashion:

$$L_{i > 2000} = \ell_{Total} + \lambda \times \min_{G} V_{LSGAN}(G) \tag{8}$$

where $\lambda = 0.4$.

In typical implementation, the recommendation is to have a reasonably batch size that is more than one sample. However, both the model and the volume sizes allowed only one sample per batch during the training, meaning that the model is updating the weight after each sample (volume of $128 \times 256 \times 256 \times 2$). On the other hand, the volume of 128 images per volume can still represents the statistical properties of the dataset to some extent.

2.3 Dataset and Data Acquisition

For the training dataset both the CREATIS [18] and COPDgene [19] dataset have been used. The CREATIS dataset contains 4D CT scans for 6 patients, each patient having 10 volumes of 141 images that represent different stages of the inhale and exhale cycle. When the sample acquired from this dataset, a random start such that 128 images is available for date input. COPDgene consists of inhale and exhale set of images for 10 patients with around 100 images or slab per volume, at each acquisition the sample has been resized to fit the network input using bilinear interpolation. Furthermore, during the acquisition of the data, the volumes randomly swapped to be sometime as fixed volume and other time as a moving volume which increased the variability of the dataset. Both datasets include landmarks, but they were not considered during the model testing.

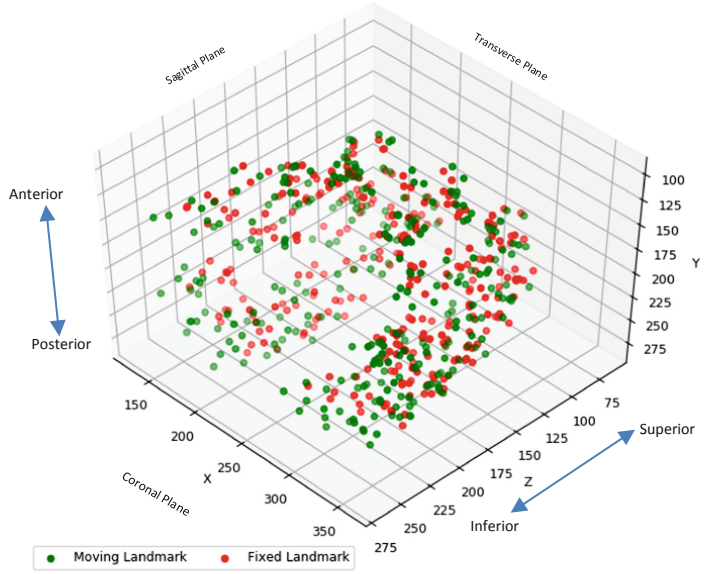
For the test set the 4D DIR-LAB dataset [20] has been used. This dataset consists of 10 patients with each patient entry includes 10 volumes. The other important feature of this dataset is that it includes 300 landmark per volume for volume 1 and 5. By conventions, the researchers map volume 5 to volume 1 of each patient.

3 Results

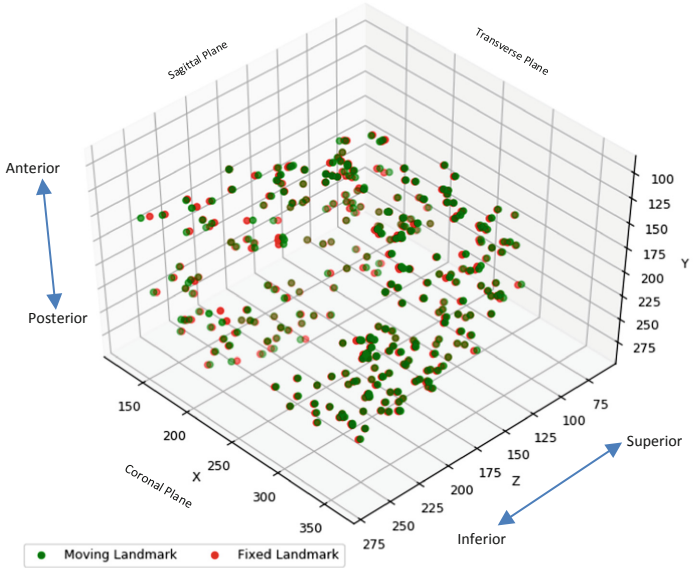
In this work the results were obtained for two cases. The first case is to train the model without the GAN implementation. The training is performed using the reconstruction error obtained from the difference between the reconstructed (warped) moving volume and the reference volume, Eqs. (5). The training for this case of training included 60K iteration. In the second case, the LSGAN network is implemented as in Fig. 1 and the training included: a) Contractive-Expanding network trained for 20000 iteration using the warping loss, Eqs. (5). b) The discriminator joined the training for the rest of 40K iterations. The error used to train the GAN during the 40K iterations is obtained from the classification of the combination of the reconstructed moving volume with the reference volume in one instance and the reference volume only in the second instance, Eqs. (6–7).

Figure 2 shows landmarks for two volumes (Patient 8) of DIR-LAB before the registration and after the registration. It is worth mentioning at this point that the volume that has been used for this experiment is quite big (128 slides) and covers a large space of the human body, it is expected that the model provides similar quality of registration for the other parts of the body like the liver, stomach and other organs under the pulmonary system and the diaphragm.

Table 2 lists on first column the landmark errors estimation before registration, followed by registration obtained from [21], following to that the result obtained from training only the Contractive-Expander (The generator part only) network in one case and training the LSGAN in the second case. In the first case, when the model consists of the generator only, the regularizations techniques involved in Eqs. (3–4) are utilized. The model performed with good performance; however, it could be due to the relatively limited dataset size the model didn't generalized will with new dataset. This can be noticed by the amount of the maximum error produced. The first case of training achieved 1.30, 1.70, and 16.34 for average, SD and maximum error respectively. The average TRE in the case of the LSGAN training shows an improvement of about 15% on the mean error. For the LSGAN the error was 1.13, 0.67, 5.70 for average, SD and maximum error respectively. This improvement can be related to the fact that the model was less prone to the overfitting. When the discriminator incorporated in the network to be training in GAN paradigm, the regularization effect due to the Dropout layers provided better conditions for the network training and improved the generalization of the model.



(a)



(b)

Fig. 2. Annotated points of DIR-LAB, volume 8. (a): Before registration, (b): After registration, all axes in millimeters. The more color saturated points represent points closer to the viewer.

Furthermore, it has been noticed that part of the better performance in the LSGAN is related to the lower maximum error obtained. It can be seen in Fig. 2 that the max error appears when a large displacement is required. Having lower maximum error in the estimation of the LSGAN tells us that the model's responses to large displacement improved which is usually challenging in the deformable registration problem. Figure 3, shows the error distributions (using the configuration with the discriminator network) for the averaged results reported in Table 2.

Table 2. Target Registration Error (TRE) for DIR-LAB dataset. Each column consists of three values, Error Mean, (Error SD), and (Maximum Error) except for results obtained from reference [21] which includes Error Mean, (Error SD). The first column (Before Registration) is the error of the landmarks between the reference and the moving volumes. The send column (Registration) is the convolutional network of Fig. 1 trained without GAN. The third column (Registration with GAN) is the complete implementation of Fig. 1 which include both the generator and the discriminator. The measurement shows improvements of the registration results while training the model with Mean Square Generative Adversarial Network (LSGAN). About 15% improvement in the registration obtained after training with GAN paradigm.

DIR-LAB	Before	After from [21]	After with Contractive-Expander (mm)	After with LSGAN (mm)
P1	3.89(2.77)(10.90)	1.05(0.5)	2.55(1.39)(16.17)	1.29(0.85)(6.28)
P2	4.33(3.89)(17.69)	1.08(0.6)	1.04(1.44)(15.89)	1.30(0.64)(3.29)
P3	6.94(4.04)(16.55)	1.46(0.9)	0.82(0.81)(4.87)	0.99(0.66)(4.61)
P4	9.72(4.89)(20.25)	2.05(1.5)	1.51(2.57)(21.67)	1.27(0.57)(4.44)
P5	7.34(5.52)(24.77)	2.02(1.7)	1.11(2.35)(27.33)	1.153(0.65)(6.43)
P6	10.89(6.9)(27.59)	2.48(1.8)	0.89(1.16)(10.77)	0.95(0.49)(4.19)
P7	11.02(7.4)(30.63)	2.78(2.3)	1.12(1.36)(17.08)	1.03(0.59)(6.05)
P8	14.99(8.9)(30.57)	3.96(3.8)	1.74(2.66)(22.65)	1.50(1.04)(9.38)
P9	7.91(3.97)(15.76)	1.89(1.2)	0.82(1.18)(8.30)	0.75(0.29)(2.10)
P10	7.30(6.3)(27.79)	2.35(2.5)	1.42(2.05)(18.58)	1.07(0.92)(10.25)
Mean	8.43(5.48)(22.19)	2.11(0.9)	**1.30(1.70)(16.34)**	**1.13(0.67)(5.70)**

The model can perform the dense motion field prediction in about 19.3 ms using INVIDIA TITAN X Pascal graphical interface with 16 GB computer memory and Intel i7 CPU.

It can be seen in the table that the summation error generated regarding the volumes (P1-P5) is higher than the error summation of (P6-P10). This difference in the error is related to the fact that number of slides of (P1-P5) is smaller than 128. During the test, to complete the volume to 128 slides, the slide padding is applied by replicating the last slide to complete the volume to 128 slides. For instance, for volume P1, the number of slides is 94, which means that the padding needed to complete the volume is 34 slides to complete the input volume to be $256 \times 256 \times 128$. Figure 4 shows a cross section for the output x-direction displacement field of P1. It is hypothesized that such padding could introduce error which is bigger than expected for what seems to be simpler displacement field.

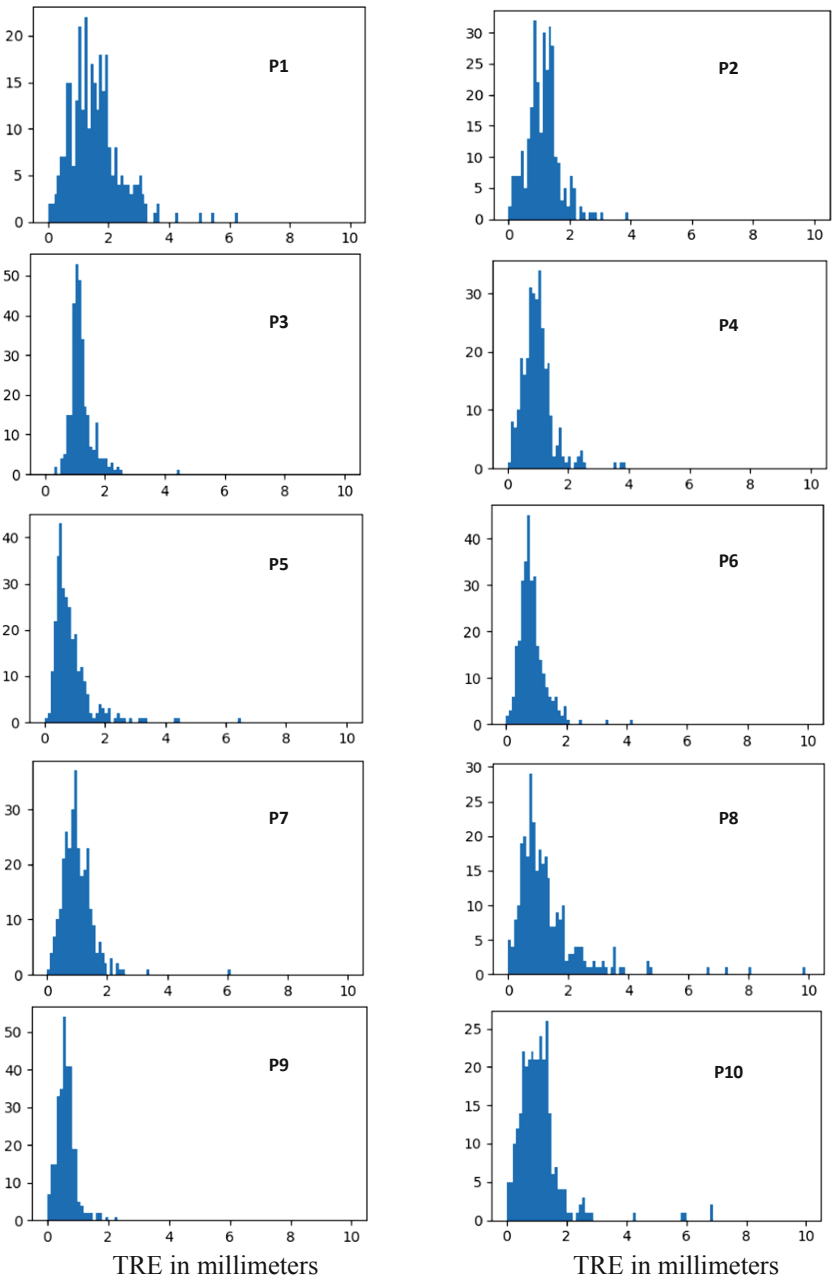


Fig. 3. The TRE histogram shows for all the volumes of DIR-LAB dataset. The horizontal axis shows the Error values.

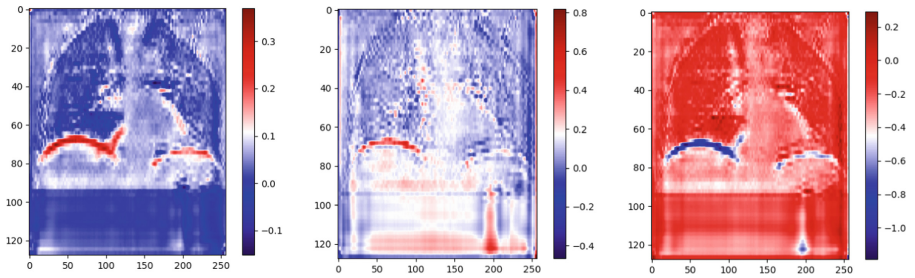


Fig. 4. A cross-second of the displacement field. From the left, the x, y, and z respectively of volume P1 shows the padding performed to complete the required input of 256 × 256 × 128 to the model.

4 Conclusion

This paper describes a novel deformable volume registration method using contraction-expansion CNN, configured without and with a discriminator sub-network within a GAN training framework. The proposed architecture is evaluated on the DIR-LAB dataset with registration performed on exhale – inhale sequences of lung CT scans. The results show that the proposed method achieves better performance when trained with the discriminator sub-network in the GAN training regime. The use of the discriminator in the GAN like-training improves the performance of the network by approximately 15%, with the state-of-the-art TRE mean error of 1.13 mm and 0.67 mm SD. These results are competitive when compared to previously reported method evaluated on the same dataset. Although the inference time has not been always clearly reported in previously published work, it is worth mentioning that the estimation of the dense motion field when using the proposed method seems also to be competitive as it enables to estimate the entire 3D registration motion field within 19.3 ms. The future work will be focused on better understanding of the role of the discriminator and changes to configuration of the described model to directly reflect intensity differences between corresponding points in the exhale – inhale sequences of lung CT scans.

References

1. Sentker, T., Madesta, F., Werner, R.: GDL-FIRE^{4D}: deep learning-based fast 4D CT image registration. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 765–773. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_86
2. Rosu, M., Hugo, G.D.: Advances in 4D radiation therapy for managing respiration: part II–4D treatment planning. Zeitschrift für Medizinische Physik. **22**(4), 272–280 (2012)
3. Yamamoto, T., Kabus, S., Bal, M., Keall, P., Benedict, S., Daly, M.: The first patient treatment of computed tomography ventilation functional image-guided radiotherapy for lung cancer. Radiother. Oncol. **118**(2), 227–231 (2016)
4. Eppenhof, K.A., Pluim, J.P.: Error estimation of deformable image registration of pulmonary CT scans using convolutional neural networks. J. Med. Imaging **5**(2), 024003 (2018)

5. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: fast predictive image registration–a deep learning approach. NeuroImage **158**, 378–396 (2017)
6. Hu, Y., et al.: Adversarial deformation regularization for training image registration neural networks. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 774–782. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_87
7. Yan, P., Xu, S., Rastinehad, A.R., Wood, B.J.: Adversarial image registration with application for MR and TRUS image fusion. In: Shi, Y., Suk, H.-I., Liu, M. (eds.) MLMI 2018. LNCS, vol. 11046, pp. 197–204. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00919-9_23
8. Qin, C., Shi, B., Liao, R., Mansi, T., Rueckert, D., Kamen, A.: Unsupervised deformable registration for multi-modal images via disentangled representations. In: Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S. (eds.) IPMI 2019. LNCS, vol. 11492, pp. 249–261. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20351-1_19
9. Tanner, C., Ozdemir, F., Profanter, R., Vishnevsky, V., Konukoglu, E., Goksel, O.: Generative adversarial networks for MR-CT deformable image registration. arXiv preprint arXiv:1807.07349, 19 July 2018
10. Mahapatra, D., Sedai, S., Garnavi, R.: Elastic registration of medical images with GANs. arXiv preprint arXiv:1805.02369, 7 May 2018
11. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Med. Image Anal. **57**, 226–236 (2019)
12. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: International Conference on Machine Learning, pp. 1050–1059, 11 June 2016
13. Alansary, A., et al.: Evaluating reinforcement learning agents for anatomical landmark detection. Med. Image Anal. **53**, 156–164 (2019)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 4 September 2014
15. Barron, J.T.: A general and adaptive robust loss function. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4331–4339 (2019)
16. Jaderberg, M., Simonyan, K., Zisserman, A.: Spatial transformer networks. In: Advances in Neural Information Processing Systems, pp. 2017–2025 (2015)
17. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794–2802 (2017)
18. Vandemeulebroucke, J., Bernard, O., Rit, S., Kybic, J., Clarysse, P., Sarrut, D.: Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT. Med. Phys. **39**(2), 1006–1015 (2012)
19. Castillo, R., et al.: A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. Phys. Med. Biol. **58**(9), 2861 (2013)
20. Castillo, R., et al.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys. Med. Biol. **54**(7), 1849 (2009)
21. Papież, B.W., Heinrich, M.P., Fehrenbach, J., Risser, L., Schnabel, J.A.: An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. Med. Image Anal. **18**(8), 1299–1311 (2014)