GDL-FIRE^{4D}: Deep Learning-Based Fast 4D CT Image Registration

Thilo Sentker[1,2]($\boxtimes$), Frederic Madesta[1,2], and René Werner[1]

[1] Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
[2] Department of Radiotherapy and Radiation Oncology, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
{t.sentker,f.madesta,r.werner}@uke.de

Abstract. Deformable image registration (DIR) in thoracic 4D CT image data is integral for, e.g., radiotherapy treatment planning, but time consuming. Deep learning (DL)-based DIR promises speed-up, but present solutions are limited to small image sizes. In this paper, we propose a General Deep Learning-based Fast Image Registration framework suitable for application to clinical 4D CT data (GDL-FIRE^{4D}). Open source DIR frameworks are selected to build GDL-FIRE^{4D} variants. In-house-acquired 4D CT images serve as training and open 4D CT data repositories as external evaluation cohorts. Taking up current attempts to DIR uncertainty estimation, dropout-based uncertainty maps for GDL-FIRE^{4D} variants are analyzed. We show that (1) registration accuracy of GDL-FIRE^{4D} and standard DIR are in the same order; (2) computation time is reduced to a few seconds (here: 60-fold speed-up); and (3) dropout-based uncertainty maps do not correlate to across-DIR vector field differences, raising doubts about applicability in the given context.

Keywords: Non-linear image registration · Registration uncertainty
4D CT · Deep learning

1 Introduction

Acquisition of 4D image data (3D+t images, respiration-correlated data) is an integral part of current radiation therapy (RT) workflows for RT planning and treatment of thoracic and abdominal tumors. Especially 4D CT imaging is meanwhile widespread and currently estimated to be routinely applied in approximately 70% of the RT facilities in the United States [1]. Typical clinical use cases of 4D CT data are (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 RT treatment planning [1].

T. Sentker and F. Madesta—Equal contribution.

At this, a key step is the application of deformable image registration (DIR) to the phase images of the 4D CT data. Traditional DIR approaches tackle the underlying task of finding an optimal transformation mapping two phase images by minimization of a dissimilarity measure that controls local correspondences of voxel intensities [3]. Yet, the algorithms are time consuming and there exists the risk of getting stuck in local minima during optimization.

Motivated by the exceptional success of deep learning (DL) and especially convolutional neural networks (CNNs) for image segmentation and classification tasks, meanwhile a number of approaches has been proposed to also solve image registration tasks by CNNs – first in the context of optical flow estimation in computer vision [4], and later similarly for medical image registration [3,5–7]. Yang *et al.* further extended a CNN-based DIR architecture to a probabilistic framework using dropouts [5], resulting in DIR uncertainty maps that could be of great value for RT treatment planning [8].

However, Uzunova *et al.* noted that "dense 3D registration with CNNs is currently computationally infeasible" [6], and focused on 2D (brain and cardiac) DIR only. To overcome this issue, patch-based approaches have been proposed for, e.g., 3D brain DIR [5], with the side effect that global information about the transformation to learn might be missing [3]. In turn, Rohé *et al.* indeed proposed using a fully convolutional architecture; with a size of $64 \times 64 \times 16$ voxel, their cardiac MR images were, however, not even close to typical sizes of 4D CT images (in the order of $512 \times 512 \times 150$ voxel per phase image).

This paper is therefore dedicated to CNN-based registration suitable for application to fast DIR in clinical thoracic 4D CT data. Taking up the aforementioned challenges and trends in current DL-based DIR,

C1 we propose a general and efficient CNN-based framework for deep learning of dense motion fields in clinical thoracic 4D CT, called GDL-FIRE^{4D},
C2 build variants of GDL-FIRE^{4D} using common open source DIR frameworks,
C3 perform a first comprehensive evaluation thereof using publicly available 4D CT data repositories (thereby presenting first respective benchmark baseline results for DL-based DIR in 4D CT data), and
C4 compare and discuss dropout-generated registration uncertainty maps for the different GDL-FIRE^{4D} variants.

To the best of our knowledge, all aspects C1-C4 are novel contributions in the given application context.

The remainder of the paper is structured as follows: In Sect. 2, the problem formulation and the concept of GDL-FIRE^{4D} are detailed. Applied data sets and performed experiments are described in Sect. 3 and respective results given and discussed in Sect. 4. The paper closes with concluding remarks in Sect. 5.

2   Methods: DL-Based Deformable Image Registration

A 4D CT image is a series $(I_i)_{i \in \{1,\dots,n_{\mathrm{ph}}\}}$ of 3D CT images $I_i : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^3$, representing the patient geometry at different breathing phases i with n_{ph} as

number of available images and breathing phases, respectively. The phases i sample the patient's breathing cycle in time and are usually denoted by cycle fractions, i.e. $\{1, \ldots, n_{\mathrm{ph}}\} \equiv \{0\%, \ldots, 50\%, \ldots\}$ with 0% as end inspiration and 50% as end expiration phase. Deformable registration in 4D CT data then aims to estimate a corresponding series of transformations $(\varphi_i)_{i \in \{1, \ldots, n_{\mathrm{ph}}\}}$ between the I_i and a reference image I_{ref}, with $\varphi_i : \Omega \to \Omega$. For the applications outlined in Sect. 1, I_{ref} usually represents one of the phase images I_i and the transformation φ_i and vector fields $u_i : \Omega \to \mathbb{R}^3$, $u_i = \varphi_i - \mathrm{id}$ (id: identity map) the respiration-induced motion of the image structures between phase i and the reference phase.

2.1 Traditional Deformable Image Registration (DIR) Formulation

In a traditional 4D CT DIR setting, the reference image is considered the fixed image, $I_{\mathrm{ref}} \equiv I_{\mathrm{F}}$, and the phase images as moving images, $I_i \equiv I_{\mathrm{M}}$, which are sequentially registered to I_{F} by $\varphi_i = \arg\min_{\varphi_i^* \in \mathcal{C}^2[\Omega]} \mathcal{J}\left[I_{\mathrm{F}}, I_{\mathrm{M}}; \varphi_i^*\right]$ to compute the sought transformations $(\varphi_i)_{i \in \{1, \ldots, n_{\mathrm{ph}}\}}$. The exact functional $\mathcal{J}$, i.e. dissimilarity measure, applied regularization approach and considered transformation model, and the optimization strategy vary in the community; see [9] for details.

2.2 Convolutional Neural Networks (CNNs) for DIR

Different to traditional DIR, we now assume a database of n_{pat} training tuples $\left(I_i^p, I_j^p, \varphi_{ij}^p\right)$, $i, j \in \{1, \ldots, n_{\mathrm{ph}}\}$, $p \in \{1, \ldots, n_{\mathrm{pat}}\}$ to be given; $\varphi_{ij}^p = id + u_{ij}^p$ represents a DIR result of the phase images $I_i \equiv I_{\mathrm{F}}$ and I_j of patient p. The goal is to learn the relationship between the input data $\left(I_i^p, I_j^p\right)$ and u_{ij}^p by a convolutional neural network.

As noted by Uzunova *et al.* [6], it is currently computationally not feasible to directly feed the entire images and vector fields into a CNN or GPU memory. Instead, we propose a slab-based approach: Let $I|_{\hat{x}} := I|_{\Omega_{\hat{x}}}$ be the restriction of image I to $\Omega_{\hat{x}} = \{(x, y, z) \in \Omega \mid x = \hat{x}\}$, i.e. the sagittal slice of I at x-position $\hat{x}$. Similarly, let $I|_{[\hat{x}_1, \hat{x}_2]}$ be the restriction of I to $\Omega_{[\hat{x}_1, \hat{x}_2]} = \{(x, y, z) \in \Omega \mid \hat{x}_1 \leq x \leq \hat{x}_2\}$, i.e. an image slab comprising the sagittal slices $\hat{x}_1, \ldots, \hat{x}_2$ of I. Using this notation, the aforementioned training tuples were converted to slab-based training samples $\left(I_i^p|_{[x-2, x+2]}, I_j^p|_{[x-2, x+2]}, u_{ij}^p|_x\right)$ with $x \in \{x_{\min}, \ldots, x_{\max}\}$ covering all sagittal slices of I. The rationale was to represent maximum information along main motion directions *inferior-superior* and *anterior-posterior* for each training sample, but also to provide some anatomical context in lateral direction.

Furthermore, the image dynamics were rescaled to $[0, 1]$, the slabs resampled to isotropic resolution of $2\,\mathrm{mm}$ and cropped/zero-padded to identical size, and the non-patient background intensity set to zero. Similar pre-processing was applied to the displacement fields (resampling and -sizing of sagittal slices, background set to zero). In addition, x-, y- and z-displacement components were z-transformed on a voxel-level to avoid unintended suppression of small displacements during CNN training. Thus, the CNN aimed to learn normalized 3D-vectors for the individual voxels of sagittal slices, which are back-transformed
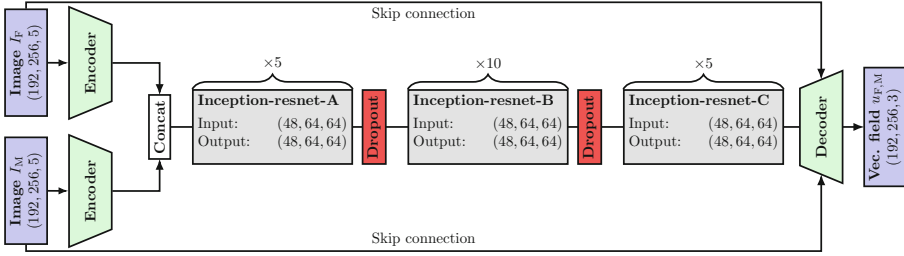
Fig. 1. CNN architecture implemented for DL-based DIR.

to actual motion fields during final reconstruction of the fields. The pre-processed slab-based samples $(\tilde{I}_i^p|_{[x-2,x+2]}, \tilde{I}_j^p|_{[x-2,x+2]}, \tilde{u}_{ij}^p|_x)$ with $x \in \{x_{\min}, \ldots, x_{\max}\}$ of the n_{pat} patients were finally shuffled and used for CNN training.

We tested different CNN architectures, including the classical U-Net [10]. Due to an observed increased robustness for DL-based DIR compared to the U-Net, we finally used an iterative CNN architecture with an Inception-ResNet-v2 [11] embedded in the encoder part of a pre-trained CT autoencoder, see Fig. 1, with MSE (mean squared error) loss function and NADAM optimizer (implemented in Tensorflow). *Iterative* means that we cascaded copies of the trained networks for improved coverage of large motion patterns.

2.3   Probabilistic CNN-Based DIR

As detailed by Yang *et al.* [5] and references therein, deterministic CNN architectures can be extended to probabilistic using dropouts [12]. Briefly speaking, the dropout layers incorporated into the CNN architecture to prevent overfitting during model training remain enabled during motion prediction. Repeated motion prediction with respectively sampled connections to be dropped eventually enable computing the sought motion field as the mean of the sampled predicted fields; further, corresponding voxel-wise variances can be interpreted as local registration uncertainty estimates [5].

3   Materials and Study Design

All experiments were run on a desktop computer with Intel Xeon CPU E5-1620 and Nvidia Titan Xp GPU. Models and scripts required can be found at https://github.com/IPMI-ICNS-UKE/gdl-fire-4d.

3.1   Training and Testing 4D CT Data Cohorts

For CNN training and model optimization, a cohort of 69 in-house acquired RT treatment planning ten-phase 4D CT data sets of patients with small lung and liver tumors was used (image size: $512 \times 512 \times 159$ voxel) and a 85%/15% split
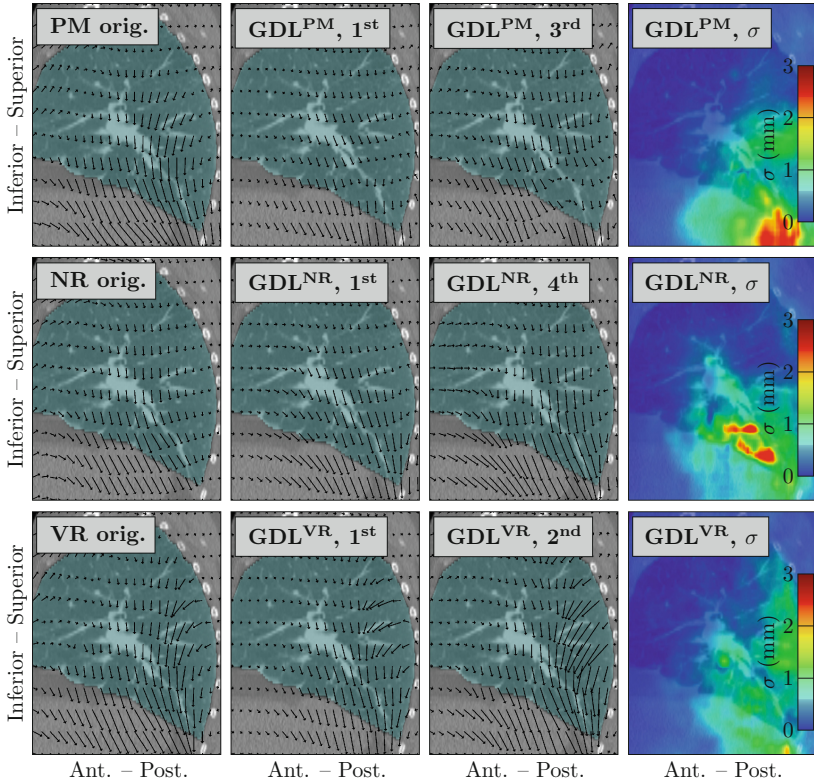
Fig. 2. Motion fields estimated by the original DIR algorithms (left column); GDL-FIRE[4D] with only a single iteration (2nd column); GDL-FIRE[4D] n iterations (3rd column); and GDL-FIRE[4D] variant-specific registration uncertainty maps (right column). Data set: DIRLAB case 08, DIR of 0% and 50% phase images.

into training and testing data performed. The 4D CT images of the open data repositories DIRLAB [13] and CREATIS [14] (see also www.creatis.insa-lyon.fr/rio/popi-model) served as *external* evaluation cohort of the trained CNNs (i.e. no model optimization performed by means of the external 4D CT cohorts).

3.2   Applied DIR Frameworks and Algorithms

To provide motion field training data, the in-house 4D CT data were registered using three common open source DIR frameworks: PlastiMatch [15], NiftyReg [16], and VarReg [17]. All approaches have been proven suitable for 4D CT registration [9]; the applied parameters were similar to respective EMPIRE10 parameters [9]. However, the algorithms are applied in a plug-and-play manner (no data pre-processing or pre-registration, no masks used). For each DIR algorithm, motion fields were provided between the 20% phase image (served as I_F) and all other phase images.

3.3   Experiments and Evaluation Measures

For each DIR algorithm, a respective probabilistic GDL-FIRE^{4D} variant was built (up to 4 cascaded CNNs, 20% dropouts). DIR accuracy was evaluated by the target registration error (TRE), computed by means of the landmarks publicly available for the DIRLAB and CREATIS data. In addition, the smoothness of transformations of the different DIR approaches and GDL-FIRE^{4D} variants was analyzed in terms of the standard deviation of transformation Jacobian determinant values of the lung voxels of the evaluation data.

4   Results and Discussion

Motion fields estimated by the original DIR algorithms and respective GDL-FIRE^{4D} variants as well as corresponding registration uncertainty maps are shown in Fig. 2 for DIRLAB case 08 (DIRLAB case with maximum motion amplitude) and phase 50% to phase 0% DIR. The similarity of the original and the GDL-FIRE^{4D} predicted fields is striking, i.e. the CNN obviously learned the DIR-specific transformation properties. This includes that the NiftyReg GDL-FIRE^{4D} variant has (similar to the original DIR) problems to directly cover larger motion amplitudes – and thereby motivates cascading several trained models for *iterative* CNN-based DIR. The success can be seen in Table 1, where the NiftyReg GDL-FIRE^{4D} outperforms the original NiftyReg DIR in terms of accuracy especially for cases with larger motion.

Still, GDL-FIRE^{4D} DIR accuracy as well as transformation properties for the other DIR approaches also resemble respective values of the traditional registration algorithm – but GDL-FIRE^{4D} offers a reduction of the runtime from approx. 15 min to a few seconds (speedup of approx. 60-fold).

Finally, it can be seen that the computed DIR uncertainty maps differ greatly between the GDL-FIRE^{4D} variants. In Fig. 3, a dataset of our internal testing cohort is shown that exhibits an artifact in the liver. This artifact led to
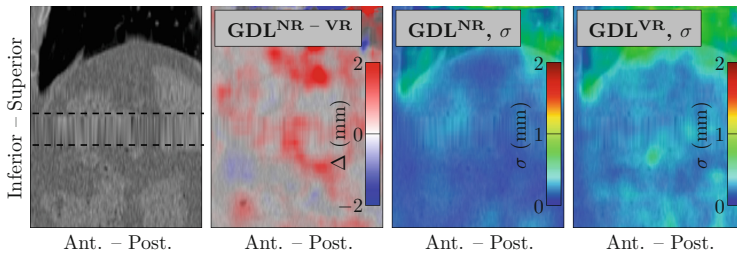

Fig. 3. From left to right: CT image serving as reference image with artifact in liver; difference of motion amplitudes estimated by the NiftyReg and the VarReg GDL-FIRE^{4D} variants, illustrating large across-DIR approach differences; NiftyReg and VarReg GDL-FIRE^{4D} uncertainty maps, showing negligible uncertainties for both variants.

Table 1. TRE values (in mm) and transformation smoothness (measured by standard deviation of lung voxel Jacobian determinant values), listed for the DIRLAB and CREATIS data, the individual DIR algorithms, and respective GDL-FIRE[4D] variants (PM: PlastiMatch; NR: NiftyReg; VR: VarReg). Landmark distance before registration: (8.46 ± 6.58) mm for the DIRLAB and (8.11 ± 4.76) mm for the CREATIS data.

		Original DIR algorithms			GDL-FIRE[4D]		
		PM [16]	NR [15]	VR [17]	PM	NR	VR
DIRLAB 4DCT	01	1.54 ± 0.98	1.46 ± 0.92	1.13 ± 0.54	1.69 ± 0.92	1.58 ± 0.80	1.20 ± 0.60
	02	1.74 ± 1.76	1.55 ± 1.06	1.17 ± 0.83	1.58 ± 1.07	1.65 ± 1.15	1.19 ± 0.63
	03	2.78 ± 2.20	2.53 ± 2.41	1.33 ± 0.69	2.39 ± 1.76	2.68 ± 1.78	1.67 ± 0.90
	04	2.70 ± 2.27	3.01 ± 2.45	3.08 ± 3.83	2.72 ± 1.97	2.48 ± 1.68	2.53 ± 2.01
	05	3.30 ± 3.06	3.21 ± 2.77	1.57 ± 1.33	2.83 ± 2.21	3.09 ± 2.50	2.06 ± 1.56
	06	3.80 ± 3.03	5.40 ± 3.94	5.23 ± 4.67	3.01 ± 1.97	2.73 ± 1.63	2.90 ± 1.70
	07	5.62 ± 5.32	8.36 ± 6.59	4.64 ± 3.91	4.48 ± 4.83	4.12 ± 4.21	3.60 ± 2.99
	08	7.65 ± 7.45	11.45 ± 9.08	4.58 ± 5.95	7.44 ± 6.87	8.26 ± 6.47	5.29 ± 5.52
	09	3.74 ± 2.60	5.66 ± 3.24	2.66 ± 2.46	3.56 ± 2.35	3.26 ± 1.90	2.38 ± 1.46
	10	3.15 ± 2.99	4.39 ± 4.21	2.14 ± 2.42	2.48 ± 1.99	2.55 ± 2.01	2.13 ± 1.88
∅ TRE		3.60 ± 1.83	4.70 ± 3.17	2.75 ± 1.57	3.22 ± 1.71	3.24 ± 1.81	2.50 ± 1.16
∅ $\sigma_{|\nabla\varphi|}$		0.10 ± 0.02	0.11 ± 0.03	0.39 ± 0.08	0.30 ± 0.13	0.24 ± 0.09	0.39 ± 0.14
CREATIS	01	1.13 ± 0.78	1.79 ± 1.26	0.90 ± 0.39	1.49 ± 0.83	1.73 ± 0.97	1.34 ± 0.74
	02	3.29 ± 3.10	4.29 ± 4.33	1.95 ± 2.87	3.59 ± 2.92	4.25 ± 3.47	2.98 ± 2.38
	03	1.95 ± 2.14	2.39 ± 2.60	1.14 ± 1.37	1.83 ± 1.42	2.05 ± 1.26	1.57 ± 1.01
	04	2.32 ± 2.95	2.51 ± 2.87	1.28 ± 2.13	1.79 ± 1.79	1.92 ± 1.73	1.64 ± 1.62
	05	1.88 ± 1.84	2.51 ± 2.73	1.17 ± 1.17	2.10 ± 1.78	2.18 ± 1.67	1.62 ± 1.09
	06	1.13 ± 0.78	1.52 ± 1.38	0.97 ± 0.72	1.60 ± 1.07	1.63 ± 1.11	1.26 ± 0.73
∅ TRE		2.01 ± 0.68	2.50 ± 0.88	1.24 ± 0.34	2.07 ± 0.78	2.29 ± 0.89	1.74 ± 0.57
∅ $\sigma_{|\nabla\varphi|}$		0.09 ± 0.02	0.11 ± 0.05	0.28 ± 0.05	0.31 ± 0.12	0.26 ± 0.08	0.30 ± 0.10

very different motion patterns estimated by the NiftyReg and the VarReg GDL-FIRE[4D] variant, but almost no measurable uncertainty for both DIR approaches. Being a direct consequence of the concept of probabilistic CNN-based DIR, this does, however, not match our understanding of DIR uncertainty and raises doubts regarding its applicability for RT planning and estimation of uncertainties therein.

5   Conclusions

The presented GDL-FIRE[4D] framework illustrates feasibility and potential of deep learning of dense vector fields for motion estimation in clinical thoracic 4D CT image data (TRE values of CNN-based DIR were in the same order than for the underlying DIR algorithms, accompanied by a speed-up factor of approximately 60), and thereby motivates continuing optimization of the framework.

References

1. Yamamoto, T., et al.: The first patient treatment of computed tomography ventilation functional image-guided radiotherapy for lung cancer. Radiother Oncol. **118**, 227–31 (2016)
2. Rosu, M., Hugo, G.D.: Advances in 4D radiation therapy for managing respiration: part II - 4D treatment planning. Z Med. Phys. **22**, 272–80 (2012)
3. Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-Net: learning deformable image registration using shape matching. In: Descoteaux, M., Descoteaux, M., et al. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 266–274. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_31
4. Dosovitskiy, A., et al.: FlowNet: learning optical flow with convolutional networks. In: 2015 IEEE International Conference on Computer Vision (ICCV), vol. 2015 Inter., pp. 2758–2766. IEEE (2015)
5. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: fast predictive image registration - a deep learning approach. Neuroimage **158**, 378–396 (2017)
6. Uzunova, H., Wilms, M., Handels, H., Ehrhardt, J.: Training CNNs for image registration from few samples with model-based data augmentation. In: Descoteaux, M., et al. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 223–231. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_26
7. Krebs, J., et al.: Robust non-rigid registration through agent-based action learning. In: Descoteaux, M., et al. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 344–352. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_40
8. Amir-Khalili, A., Hamarneh, G., Zakariaee, R., Spadinger, I., Abugharbieh, R.: Propagation of registration uncertainty during multi-fraction cervical cancer brachytherapy. Phys. Med. Biol. **62**, 8116–8135 (2017)
9. Murphy, K., van Ginneken, B., Reinhardt, J.M.: Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. IEEE Trans. Med. Imaging **30**, 1901–20 (2011)
10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
11. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, Inception-Resnet and the impact of residual connections on learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), pp. 4278–4284 (2017)
12. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**, 1929–58 (2014)
13. Castillo, R., Castillo, E., Guerra, R.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys. Med. Biol. **54**, 1849–70 (2009)
14. Vandemeulebroucke, J., Rit, S., Kybic, J., Clarysse, P., Sarrut, D.: Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. Med. Phys. **38**, 166–78 (2011)
15. Modat, M., et al.: Fast free-form deformation using graphics processing units. Comput. Methods Programs Biomed. **98**, 278–84 (2010)

16. Shackleford, J.A., Kandasamy, N., Sharp, G.C.: On developing B-spline registration algorithms for multi-core processors. Phys. Med. Biol. **55**, 6329–51 (2010)
17. Werner, R., Schmidt-Richberg, A., Handels, H., et al.: Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: a comparison and evaluation study. Phys. Med. Biol. **59**, 4247–4260 (2014)