Temporal Interpolation of Abdominal MRIs Acquired During Free-Breathing

Neerav Karani[1(✉)], Christine Tanner[1], Sebastian Kozerke[2], and Ender Konukoglu[1]

[1] Computer Vision Laboratory, ETH Zurich, Zurich, Switzerland
nkarani@student.ethz.ch
[2] Institute for Biomedical Engineering, University & ETH Zurich, Zurich, Switzerland

Abstract. We propose a convolutional neural network (CNN) based solution for temporal image interpolation in navigated 2D multi-slice dynamic MRI acquisitions. Such acquisitions can achieve high contrast time-resolved volumetric images without the need for breath-holding, which makes them essential for quantifying breathing induced motion for MR guided therapies. Reducing the number of navigator slices needed in these acquisitions would allow increasing through-plane resolution and reducing overall acquisition time. The proposed CNN achieves this by interpolating between successive navigator slices. The method is an end-to-end learning based approach and avoids the determination of the motion field between the input images. We evaluate the method on a dataset of abdominal MRI sequences acquired from 14 subjects during free-breathing, which exhibit pseudo-periodic motion and sliding motion interfaces. Compared to an interpolation-by-registration approach, the method achieves higher interpolation accuracy on average, quantified in terms of intensity RMSE and residual motion errors. Further, we analyze the differences between the two methods, showing the CNN's advantages in peak inhale and exhale positions.

1 Introduction

Dynamic volumetric magnetic resonance imaging (4D-MRI) is an essential technology for non-invasive quantification of breathing induced motion of anatomical structures [1]. It is of particular importance for learning motion models, which are used for planning and guiding radiotherapy [2] and high intensity focused ultrasound therapy [3]. One particular approach to 4D-MRI is navigated 2D multi-slice acquisition, which is performed by continuously switching between acquiring a navigator slice $\mathbf{N}_t$ (at same anatomical location) and a data slice $\mathbf{D}^p$ (at different locations p), e.g. for 3 locations the acquisition sequence would be $\{\mathbf{N}_1, \mathbf{D}^1, \mathbf{N}_2, \mathbf{D}^2, \mathbf{N}_3, \mathbf{D}^3, \mathbf{N}_4, \mathbf{D}^1, \dots\}$. 3D MRI for different time points are retrospectively created by stacking the data slices enclosed by navigators that show the same organ position. The main advantages of 4D-MRI are that it allows imaging without breath-holding, which facilitates quantifying irregular motion patterns over long periods and does not impose additional discomfort to

M. Descoteaux et al. (Eds.): MICCAI 2017, Part II, LNCS 10434, pp. 359–367, 2017.
DOI: 10.1007/978-3-319-66185-8_41

the patient. Compared to other temporal MRI techniques [4], the chosen image protocol yields higher inflow contrast which provides stronger image contrast between vessels and soft tissue, an important advantage for radiotherapy applications.

Reducing the number of navigator acquisitions without sacrificing temporal resolution is very attractive. For example, changing to a scheme where 3 data slices are acquired between navigators would reduce the required acquisition time by 2/3, which could be used for improving through plane resolution while keeping the same total acquisition time (same FOV covered by 6 slices $\{\mathbf{N}_1, \mathbf{D}^1, \mathbf{D}^2, \mathbf{D}^3, \mathbf{N}_2, \mathbf{D}^4, \mathbf{D}^5, \mathbf{D}^6, \mathbf{N}_3, \mathbf{D}^1 \dots\}$) or for reducing overall acquisition time while keeping the same plane thickness ($\{\mathbf{N}_1, \mathbf{D}^1, \mathbf{D}^2, \mathbf{D}^3, \mathbf{N}_2, \mathbf{D}^4, \mathbf{D}^1 \dots\}$). Accurate temporal interpolation of the navigators can achieve such a reduction.

In this work we propose a convolutional neural network (CNN) for temporal interpolation of 2D MRI slices. The network takes as input the images of the same slice acquired at different time points, e.g. $\mathbf{N}_1, \mathbf{N}_3, \mathbf{N}_5$ and $\mathbf{N}_7$, and interpolates the image in between, e.g. $\mathbf{N}_4$. The proposed network is a basic fully convolutional architecture that takes multiple images and produces an image of the same size. We evaluate the proposed method with a dataset composed of navigator images from 4D-MRI acquisitions in 14 subjects with a mean temporal resolution of 372 ms. We compare our algorithm with a state-of-the-art registration based approach, which interpolates between successive time points using the displacement field estimated by a non-rigid registration algorithm. The results suggest that the proposed CNN-based method outperforms the registration based method. Analyzing the differences, we observed that the network produces more accurate results when interpolating at peak inhalation and exhalation points, where the motion between time points is highly non-linear. Registration-based interpolation that considers multiple past and future images might be able to account for some of this non-linear motion, but will require a more sophisticated approach including inversion of non-rigid transformation fields (potentially introducing errors) and thus much higher computation times.

Related Work: Temporal interpolation in MRI has been studied in the literature for the problem of dynamic MRI reconstruction. Majority of these works interpolate k-space data [4] or use temporal coherency to help reconstruction [5]. Sampling patterns in the k-space is an important part of these methods while the proposed method here, directly works on the image space. On the other hand, 4D-MRI reconstruction methods without 2D navigators have also been proposed, relying, for example, on an external breathing signal [6] or the consistency between neighbouring data slices after manifold embedding [7]. However, continuously observing organ motion through navigators potentially provides superior reconstructions.

Temporal interpolation in the image space has been mostly studied for ultrasound imaging. Several works tackled this problem by explicitly tracking pixel-wise correspondences in the input images. These include approaches based on optical flow estimation [8], non-rigid registration [9,10] and motion

compensation [11]. Authors in [12] interpolate the temporal intensity variation of each pixel with sparse reconstruction using over-complete dictionaries.

Following the success of CNNs several computer vision studies proposed temporal interpolation in non-medical applications. Authors in [13] use CNN-based frame interpolation as an intermediate step for estimating dense correspondences between two images. Their CNN architecture is inspired by [14], where the goal is dense optical flow estimation. Variants of deep neural networks that have been proposed for the closely related task of future frame prediction in videos include recurrent neural networks [15] and an encoder-decoder style network with a locally linear latent space [16]. Authors in [17] and [18] use generative adversarial networks [19] and variational autoencoders [20] to predict future video frames and for facial expression interpolation respectively.

2 Method

CNN-Based Temporal Interpolation: The general architecture of the proposed temporal interpolation CNN is shown in Fig. 1. The network is trained to increase the temporal resolution of an input image sequence ($\mathbf{N}_1, \mathbf{N}_3, \mathbf{N}_5, \dots$) by generating the intermediate images ($\mathbf{N}_2, \mathbf{N}_4, \dots$). For generating the intermediate image at any time instance, 2T input images, T from the past and T from the future, are concatenated in the order of their time-stamps, and passed through multiple convolutional blocks in order to generate the target image. Each convolutional block consists of a spatial dimension preserving convolutional layer, followed by a rectified linear unit (ReLU) activation function. As the network is fully convolutional, it can be used to temporally interpolate image sequences of any spatial resolution without retraining. During training we optimize a loss function L between the ground truth images $\mathbf{N}_t$ and the interpolated ones $\hat{\mathbf{N}}_t$, i.e. $L(\mathbf{N}_t, \hat{\mathbf{N}}_t)$. We experimented with different loss functions that we detail in Sect. 3.

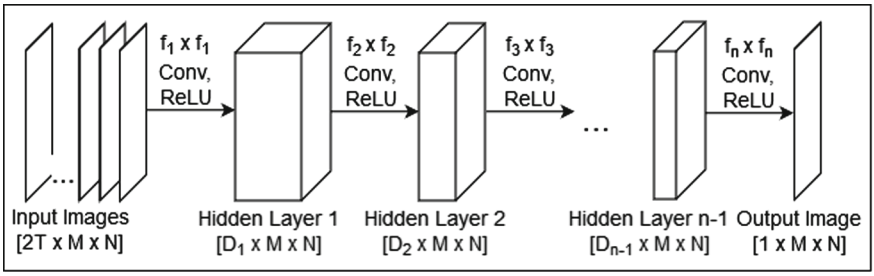

Fig. 1. Architecture of the temporal interpolation CNN.

Long range spatial dependencies are captured by increasing the convolution kernel sizes or the depth of the network. Other ways to do this, such as pooling

or higher stride convolutions, may reduce the spatial dimensionality in the hidden layers, which might lead to losing high-frequency details in the generated images. These alternatives often require skip connections [13] or multi-resolution approaches [17] to preserve details.

Some of the previously proposed CNN-based methods for frame interpolation in computer vision, such as [13], use only the immediate neighbours for interpolation, i.e. $T = 1$. Due to lack of additional temporal context, these approaches may be unable to resolve certain motion ambiguities and capture non-linearities. In the proposed algorithm, we consider larger temporal context similar to [17], to deal with such challenges. Indeed, our experiment analysis demonstrates the benefits of using $T > 1$.

Registration-Based Interpolation: We employ the widely used interpolation-by-registration approach to compare with the proposed CNN. The method is based on the principles proposed in [9], however, we employ a recently devised image registration method that can cope with sliding boundaries and has a state-of-the-art performance for 4D-CT lung and 4D-MRI liver image registration [21]. It uses local normalized cross correlation as image similarity measure and isotropic total variation for spatial regularization besides a linearly interpolated grid of control points $\mathbf{G}$ with displacements $\mathbf{U}$.

For $T = 1$, intermediate slices $\mathbf{N}_t$ are created by registering the enclosing slices ($\mathbf{N}_{t-1}$, $\mathbf{N}_{t+1}$) and then applying half of the transformation to the moving image. To improve SNR and avoid possible bias, we make use of both transformations ($\mathbf{N}_{t+1} \rightarrow \mathbf{N}_{t-1}$, $\mathbf{N}_{t-1} \rightarrow \mathbf{N}_{t+1}$) and average the resulting two interpolated slices. For T=2, 3 moving images ($\mathbf{N}_{t-2}$, $\mathbf{N}_{t+1}$, $\mathbf{N}_{t+2}$) are registered to fixed image $\mathbf{N}_{t-1}$, providing grid displacements $\mathbf{U}_{t-2}$, $\mathbf{U}_{t+1}$, $\mathbf{U}_{t+2}$. Per grid point and displacement component, a third order polynomial is fitted to the displacement values to deduce $\mathbf{U}_t$. Finally the inverse transformation $\mathbf{U}_t^{-1}$ is approximated and applied to $\tilde{\mathbf{N}}_{t-1}$ (mean of the fixed and warped moving images) to provide the interpolated image.

3 Experiments and Results

Dataset: The dataset consists of temporal sequences of sagittal abdominal MR navigator slices from 14 subjects. Images were acquired on a 1.5T Philips Achieva scanner using a 4-channel cardiac array coil, a balanced steady-state free precession sequence, SENSE factor 1.7, 70° flip angle, 3.1 ms TR, and 1.5 ms TE. Spatial resolution is $1.33 \times 1.33 \times 5\,\text{mm}^3$ and temporal resolution is 2.4–3.1 Hz. For each subject the acquisition was done over 3 to 6 blocks with each block taking 7 to 9 min and with 5 min resting periods in between. Each block consists of between 1100 and 1500 navigator images. We divide the 14 subjects into two groups of 7 subjects each, which are used for two-fold cross-validation experiments.

Training Details: The network is implemented in Tensorflow [22]. The architecture parameters (see Fig. 1) are empirically set to a depth n = 9, kernel sizes

(f_1, f_2, ... f_9) = (9,7,5,3,3,3,3,3,3), and (D_1, D_2,... D_8) = (32,16,8,8,8,8,8,8). The weights are initialized as recommended in [23] for networks with ReLUs as activation functions. We use the Adam optimizer [24] with a learning rate of 1e−4 and set the batch size to 64. Per block, the image intensities are linearly normalized to their 2 to 98%tile range. The CNN trains in about 48 h. No overfitting is observed, with training and testing errors being similar (mean RMSE +2.1%).

Evaluation: The interpolation performance was quantified by (i) the RMSE between the intensities of the interpolated and the ground truth image, and (ii) the residual mean motion when registering the interpolated image to the ground truth image. We summarize the performance by the mean, median and 95%tile after pooling all test results.

We evaluated the benefit of providing additional temporal context for interpolation by comparing the proposed CNN's performance using $T = 1$ and $T = 2$. Setting $T = 2$, we then studied the effect of training the network on 3 different loss functions, namely $L2$ ($\|\mathbf{N}_t - \hat{\mathbf{N}}_t\|_2$), $L1$ ($\|\mathbf{N}_t - \hat{\mathbf{N}}_t\|_1$), and L1-GDL ($L1 + \|\partial\mathbf{N}_t/\partial_x - \partial\hat{\mathbf{N}}_t/\partial_x\|_1 + \|\partial\mathbf{N}_t/\partial_y - \partial\hat{\mathbf{N}}_t/\partial_y\|_1$), where GDL stands for the Gradient Difference Loss [17] that is shown to improve sharpness and correct edge placement. In GDL computation, the target image gradients are computed after denoising with a median filter of size 5×5 and the gradient operators are implemented with first order finite differences. The GDL is equally weighted with the reconstruction cost, as in [17].

Results: We evaluated the performance of the registration algorithm in aligning 2D liver MR sequences based on manually annotated landmarks inside the liver (20 landmarks from sequences of 10 subjects, 300 frames each). Its mean registration accuracy was 0.75 mm and average runtime per slice registration was 1.19 s on a 2 processor machine with Intel i7-3770K CPUs @ 3.50 GHz.

Table 1. Intensity RMSE and residual mean motion comparison.

Method		RMSE			ResMotion [mm]			Runtime [s]
		Mean	Median	95%	Mean	Median	95%	
(a) Interpolation by registration versus CNN								
Registration	T=1	8.45	7.97	12.86	0.45	0.32	1.14	2.377
Registration	T=2	8.34	7.84	12.30	0.36	0.29	0.80	94.691
CNN	T=1, L2	8.46	7.98	12.68	0.42	0.30	1.08	**0.006**
CNN	T=2, L2	**7.92**	7.63	**11.62**	**0.30**	**0.24**	**0.66**	0.007
CNN	T=3, L2	7.99	7.66	11.67	0.31	0.25	0.67	0.007
(b) CNN trained on different loss functions								
CNN	T=2, L2	**7.92**	7.63	**11.62**	**0.30**	**0.24**	**0.66**	0.007
CNN	T=2, L1	7.93	**7.61**	11.64	0.31	**0.24**	0.70	0.007
CNN	T=2, L1-GDL	9.44	8.95	12.58	0.31	**0.24**	0.71	0.007

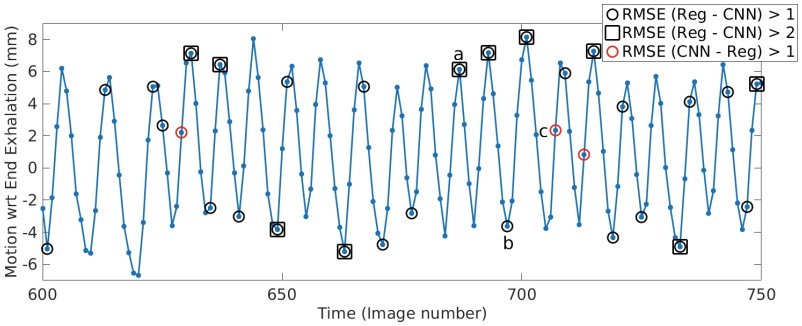

Fig. 2. Relative performance of the two methods along several breathing cycles. Labels a-c indicate rows in Fig. 3 showing the corresponding images.

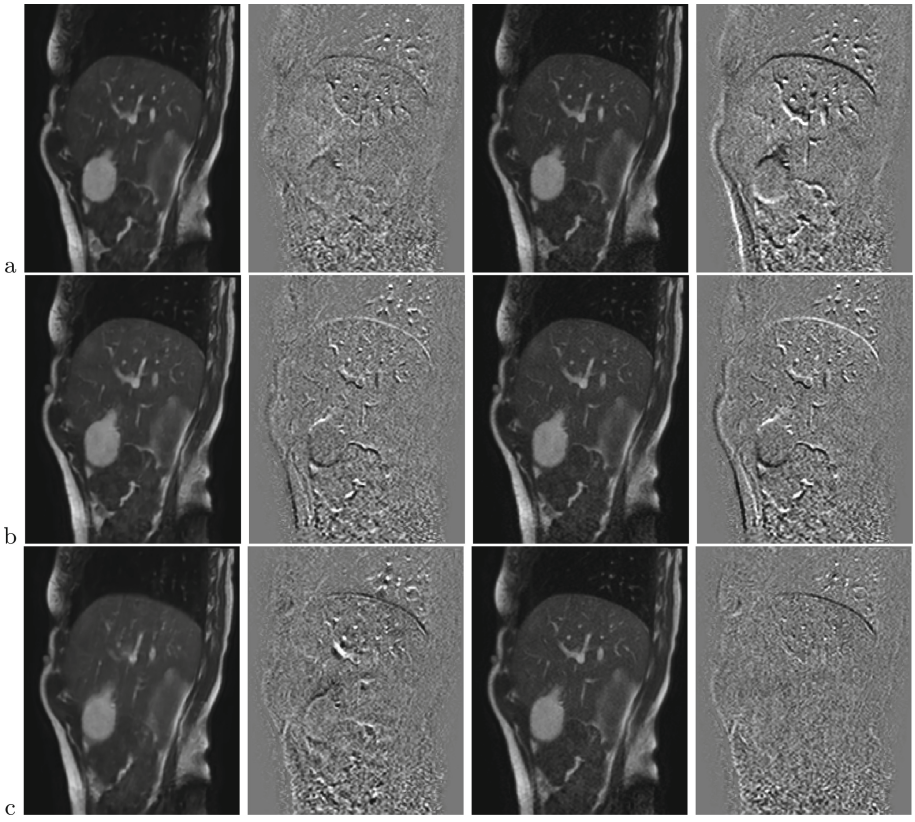

Fig. 3. Visualization of cases marked in Fig. 2. Each row (from left to right): CNN ($T = 2$, L2) result and error image, registration result and error image. Rows (a, b) show examples of the CNN performing better at an (a) end-inhale and (b) end-exhale position, while row (c) shows the registration performing better when the motion is high and linear.

Table 1 summarizes the two-fold cross-validation interpolation results. The performances of the registration and the CNN ($T = 1$) are similar, with the latter needing much less time for interpolation. Using CNN ($T = 2$) leads to an improvement in mean RMSE and mean residual motion by 6.27% and 33.33% respectively. More temporal context (CNN, $T = 3$) does not improve results further. L1 and L2 losses lead to similar results, while the introduction of the GDL worsens the RMSE. The relevant evaluation measure for 4D reconstruction, the residual mean motion, seems insensitive to the choice of training loss function.

To gain insight about the method's performance, we extracted the superior-inferior mean motion within the liver by registering all images to a reference end-exhale image, see Fig. 2. Then we marked cases were the RMSE values of CNN and registration differed substantially. It can be observed that CNN had substantially lower RMSE values for most end-inhale extrema (positive SI displacements) while a registration was better for a few frames during the high motion phase. Example interpolated images and their differences to the ground truth image are shown in Fig. 3 for the selected cases with large difference in RMSE. The difference is also visually apparent.

4 Conclusion

In this article, we proposed a convolutional neural network for temporal interpolation of 2D MR images. Experimental results suggest that the CNN based method reaches a higher accuracy than interpolation by non-rigid registration. The difference is especially pronounced at the peak inhalation and exhalation points. We believe the proposed method can be useful for 4D MRI acquisition. For the same acquisition time, it can improve the through-plane resolution or SNR, and for the same through-plane resolution and SNR, it can reduce the acquisition time. The proposed method is evaluated using retrospective data in this work. In our future work, we will extend this to prospective evaluation with new data acquisitions to quantify improvements on through-plane resolution and acquisition time reduction.

The results also suggest that there is room for improvement. Better network architectures [14–18] and objective functions [17] might preserve high-frequency details better, which will be examined in the continuation of this work. Lastly, we demonstrated the temporal interpolation for the problem of interpolating navigator slices in 4D MRI. The same methodology can also be used for temporal interpolation of segmentation labels for more accurate object tracking and longitudinal studies with irregular temporal sampling.

Acknowledgments. This work was supported by a K40 GPU grant from Nvidia.

References

1. Von Siebenthal, M., Székely, G., Gamper, U., Boesiger, P., Lomax, A., Cattin, P.: 4D MR imaging of respiratory organ motion and its variability. Phys. Med. Biol. **52**, 1547 (2007)

2. Bert, C., Durante, M.: Motion in radiotherapy: particle therapy. Phys. Med. Biol. **56**(16), R113 (2011)
3. Arnold, P., Preiswerk, F., Fasel, B., Salomir, R., Scheffler, K., Cattin, P.C.: 3D organ motion prediction for MR-guided high intensity focused ultrasound. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011 Part II. LNCS, vol. 6892, pp. 623–630. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23629-7_76
4. Tsao, J., Kozerke, S.: MRI temporal acceleration techniques. J. Magn. Reson. Imaging **36**(3), 543 (2012)
5. Uecker, M., Zhang, S., Voit, D., Karaus, A., Merboldt, K.-D., Frahm, J.: Real-time MRI at a resolution of 20 ms. NMR Biomed. **23**(8), 986 (2010)
6. Tryggestad, E., Flammang, A., Han-Oh, S., Hales, R., Herman, J., McNutt, T., Roland, T., Shea, S.M., Wong, J.: Respiration-based sorting of dynamic MRI to derive representative 4D-MRI for radiotherapy planning. Med. Phys. **40**(5), 051909 (2013)
7. Baumgartner, C.F., Kolbitsch, C., McClelland, J.R., Rueckert, D., King, A.P.: Groupwise simultaneous manifold alignment for high-resolution dynamic MR imaging of respiratory motion. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) IPMI 2013. LNCS, vol. 7917, pp. 232–243. Springer, Heidelberg (2013). doi:10.1007/978-3-642-38868-2_20
8. Nam, T.-J., Park, R.-H., Yun, J.-H.: Optical flow based frame interpolation of ultrasound images. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2006 Part I. LNCS, vol. 4141, pp. 792–803. Springer, Heidelberg (2006). doi:10.1007/11867586_72
9. Penney, G.P., Schnabel, J.A., Rueckert, D., Viergever, M.A., Niessen, W.J.: Registration-based interpolation. IEEE Trans. Med. Imaging **23**(7), 922 (2004)
10. Zhang, W., Brady, J.M., Becher, H., Noble, J.A.: Spatio-temporal (2D+T) non-rigid registration of real-time 3D echocardiography and cardiovascular MR image sequences. Phys. Med. Biol. **56**(5), 1341 (2011)
11. Lee, G.-I., Park, R.-H., Song, Y.-S., Kim, C.-A., Hwang, J.-S.: Real-time 3D ultrasound fetal image enhancment techniques using motion-compensated frame rate up-conversion. In: Medical Imaging, p. 375 (2003)
12. Gifani, P., Behnam, H., Haddadi, F., Sani, Z.A., Shojaeifard, M.: Temporal super resolution enhancement of echocardiographic images based on sparse representation. IEEE Trans. Ultrason. Ferroelectr. **63**(1), 6 (2016)
13. Long, G., Kneip, L., Alvarez, J.M., Li, H., Zhang, X., Yu, Q.: Learning image matching by simply watching video. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 434–450. Springer, Cham (2016). doi:10.1007/978-3-319-46466-4_26
14. Fischer, P., Dosovitskiy, A., Ilg, E., Häusser, P., Hazırbaş, C., Golkov, V., van der Smagt, P., Cremers, D., Brox, T.: Flownet: learning optical flow with convolutional networks. arXiv:1504.06852 (2015)
15. Srivastava, N., Mansimov, E., Salakhutdinov, R.: Unsupervised learning of video representations using LSTMs. In: ICML, p. 843 (2015)
16. Goroshin, R., Mathieu, M.F., LeCun, Y.: Learning to linearize under uncertainty. In: Advances in Neural Information Processing Systems, p. 1234 (2015)
17. Mathieu, M., Couprie, C., LeCun, Y.: Deep multi-scale video prediction beyond mean square error. arXiv:1511.05440 (2015)
18. Yeh, R., Liu, Z., Goldman, D.B., Agarwala, A.: Semantic facial expression editing using autoencoded flow. arXiv:1611.09961 (2016)
19. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, p. 2672 (2014)

20. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. arXiv:1312.6114 (2013)
21. Vishnevskiy, V., Gass, T., Szekely, G., Tanner, C., Goksel, O.: Isotropic total variation regularization of displacements in parametric image registration. IEEE Trans. Med. Imaging **36**(2), 385 (2016)
22. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al.: Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467 (2016)
23. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: ICCV, p. 1026 (2015)
24. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv:1412.6980 (2014)