Proc. Annual Meeting of the Austrian Society for
Biomedical Engineering 2021

DOI: 10.3217/978-3-85125-826-4-18

DEEP LEARNING BASED IMAGE REGISTRATION IN DYNAMIC CARDIAC CT USING A RECURSIVE CASCADE NETWORK APPROACH

K.A. Lara[1,2], I.A. Juárez[2], M.A. Perez[2], T. Rienmüller[1], C. Baumgartner[1]

[1]Institute of Health Care Engineering with European Testing Center of Medical Devices, Graz University of Technology, Austria
[2]Department of Biomedical Engineering, Galileo University, Guatemala

andrealh@galileo.edu

Abstract— *Registration of dynamic CT image sequences is a necessary preprocessing step for accurate assessment of multiple (patho)physiological determinants in the heart such as myocardial perfusion. In this work we present a recursive-cascade-network approach for deformable image registration using data from myocardial perfusion CT studies. A contrast-agent dependent loss function was introduced which enabled us to further improve the accuracy of sequence registration. In addition, different network configurations were evaluated, showing a good trade-off between spatial registration accuracy and image quality.*

Keywords— *myocardial perfusion, dynamic cardiac computed tomography, deep learning, sequence registration, recursive cascade network.*

Introduction

Deformable image registration (DIR) is essential for clinical applications where spatial alignment of anatomical structures is required. Such applications include image-guided procedures in diagnostics and therapy [1]. In cardiac image analysis, DIR is used in image-guided interventions, perfusion studies and procedures that requires myocardial motion tracking [2]–[4]

Cardiac perfusion studies in dynamic computed tomography (CT) are performed to qualitatively or quantitatively assess myocardial perfusion after contrast agent administration. Such studies evaluate the distribution of contrast agent in the heart and aim to identify and detect ischemic areas in the ventricle characterized by hypo attenuation (reduced CT values) in the image [5]. During a patient examination, an image sequence of the heart is obtained by using an ECG-gated protocol that usually acquires image data at the end-systolic phase. However, due to cardiac stressing, respiratory- and patient motion, spatial misalignment can be present. Hence, the registration of the whole ventricle or a selected ROI over the whole 2D/3D sequence is necessary for an accurate measurement of the time-attenuation curves. Such task has some unique challenges because of the non-rigid dynamic nature of the heart, the motion of the thorax and the lack of anatomical landmarks. Moreover, the dynamic information of the changing contrast agent introduces another degree of complexity to the problem.

Recent advances in image registration have demonstrated the potential of deep learning techniques in applications for multi-modal and inter-patient registration, and motion tracking [1], [6]. Furthermore, approaches using supervised and unsupervised learning have been introduced, however, supervised methods are hardly to be implemented due to the need of ground-truth flow-fields. In contrast, unsupervised methods do not require flow-field labels and perform image registration using a similarity measure between the fixed and the warped moving image [6]. Current state-of-the-art methods on DIR, however, use an unsupervised approach [7], [8], [9].

In this work, we further developed and evaluated the performance of a so-called Recursive-Cascade-Network [9], an established method for DIR using datasets of myocardial perfusion CT sequences. Here, we introduced a contrast-agent dependent loss function to improve the accuracy of sequence registration and evaluated the results using different network configurations. Finally, we evaluated the effects of the number of selected cascades and modified the loss function in terms of optimizing spatial alignment and image quality.

Methods

Recursive Cascade Network. Let $S = \{I_{m_i}\}_{i=0}^{N}$ denote a sequence of images, where $I_{m_i} \in \Omega \subset R^2$ and let $I_f \in \mathrm{S}$ denote a fixed image. We want to predict a flow field $\varphi: \Omega \rightarrow \Omega$ that aligns the sequence S. The Recursive Cascade Network [9] generates a flow prediction function F which takes a fixed image I_f a moving image I_m and predicts φ This field is by the composition of flows (see Eq. 1)

$$\varphi = \varphi_1 \circ \varphi_2 \circ \cdots \circ \varphi_n \qquad (1)$$

where φ_k for $k = 1,2, \ldots, n$ is predicted by the k-th cascade which is a base subnetwork such as [7], [8]. The motivation of using a cascade-based-registration

Published by Verlag der TU Graz
Graz University of Technology

DOI: 10.3217/978-3-85125-826-4-18

concept is to decompose large displacements performed by φ into progressively small displacements generated by the flows $\varphi_1, \varphi_2, \dots, \varphi_n$. The final image is obtained by successively warping of the moving image along all cascades (see Eq. 2).

$$\varphi \circ I_m = (\varphi_1 \circ \varphi_2 \circ \dots \circ \varphi_n) \circ I_m \qquad (2)$$

Loss function. As suggested in [8], the loss function in image registration often consist of a similarity loss L_{sim} for the fixed and warped image, and a regularization loss L_{reg} to smooth the terms of the field as presented in Eq. 3.

$$L_{nc}(I_f, I_m, \varphi) = L_{sim}(I_f, \varphi \circ I_m) + L_{reg}(\varphi) \qquad (3)$$

However, for this application, we found that considering the loss as denoted in Eq. 3, affects negatively to regions of high contrast agent concentrations (right/left atrium or ventricle) in the image sequence. Specifically, for cases in which the contrast regions from the fixed image differ from the moving image. In these cases, contrast regions are introduced or removed from the moving image to make the warped image more similar to the fixed one. Therefore, to address this problem, we added a loss term that penalizes such changes and formulated the loss as presented in Eq. 4:

$$L_c(I_f, I_m, \varphi) = \\ \alpha_1 L_{sim}(I_f, \varphi \circ I_m) + \alpha_2 L_{cont}(I_f, \varphi \circ I_m, C) + L_{reg}(\varphi) \qquad (4)$$

The term L_{sim} is used to reduce the misalignment between I_m and I_f, and the introduced term L_{cont} is used to penalize the changes of the contrast regions in the warped image. Here, we included the parameters α_1 and α_2 to weight the losses. We experimentally found that α_1 has to be lower than 0.5 to reduce the changes in contrast regions. If $C_i \subset I_{mi}$ is a contrast region in the *i*-th moving image of the sequence *S*, we want to preserve as much information about C_i in the warped image, therefore, we want φ to modify this region only the necessary to reduce the misalignment. For this purpose, masks of the contrast agent were used and the regions generated as shown in Eq. 5

$$C_i = I_{m_i} \odot M_{m_i} \qquad (5)$$

where *M* is the mask. Hence, the L_{cont} loss is defined as denoted in Eq. 6

$$L_{cont}(I_f, I_m, \varphi) = \\ d(I_m \odot M_m, (\varphi \circ I_m) \odot M_m) + d(I_f \odot M_f, (\varphi \circ I_m) \odot M_f) \qquad (6)$$

The first term penalizes the loss of contrast from the moving image in the warped image, and the second term penalizes the introduction of contrast regions into the warped image. Both cases are equally relevant for this application, therefore, we used the same weight factor, α_2, for both terms. Moreover the operator $\odot$ is the Hadamard product and *d* is a similarity metric such as correlation coefficient, mutual information or mean squares. For this application we used the *Pearson correlation coefficient*. We ran several experiments with different configurations for L_{cont}, to determine the parameters α_1 and α_2 .

Dataset. We used a dataset comprising 247 CT sequences of 2D myocardial perfusion images. The data was acquired from 19 patients undergoing regular CT examination with a Philips-iCT 256 scanner. All patients gave informed consent. The sequences where obtained from a 13 slices volume stack of slice thickness of 5 mm and matrix size of 512x512, containing 23 – 45 frames over time. However, from each volume stack only 4 to 5 slices representing the ventricular chambers were considered. For each of the frames, masks of the contrast agent (regions with high contrast agent concentrations) were obtained using a CT window of W:450 L:130. For training and validation, we used the default CT window W:750 L:90. Subsequently, we split the data on subject level, 17 for training and 2 for validation. I_f of each sequence was set to the frame with the maximum amount of contrast-agent.

Implementation. Our proposed configuration was implemented in PyTorch using a modified 2D version of the original implementation (for 3D volumes) as published in [9]. We also used the same hyper parameters and selected a VTN [7] for the base subnetworks. Our model was trained using a batch size of 32 on 2 TITAN RTX 25GB.

Experiments. The registration of the sequence was performed using different numbers of cascades and loss functions. The aim was to assess how the network configuration influences the image alignment and quality. Hence, we implemented several networks using 3, 5, 7, 10, 15 and 20 cascades, and trained them based on the loss function according to Eq. 3 and Eq. 4, respectively. The *Pearson correlation coefficient* was used as the similarity metric for L_{sim} and L_{cont}, L_{reg} is the total variation loss as used in [7] and the parameters α_1 and α_2 were set to 0.4 and 0.3, respectively. Finally, we assessed the registration performed by the networks by calculating the evaluation measures.

Evaluation Metrics. The performance of the registration was evaluated based on the accuracy of spatial alignment and the image quality. The spatial alignment was quantified using the Dice score [10] which measures the overlap between two regions and ranges between 0 and 1, where 1 means perfect matching. In particular, we calculate the Dice score between the segmentations obtained from of the whole heart. Furthermore, the image similarity before and after registration was measured to assess the loss of information during the process, i.e. changes in contrast agent concentration over time. Next the Mutual Information (MI)[11] between the moving and the warped image is estimated. In addition, we included the Structural Similarity Index Measure (SSIM)[12] as an additional measure for estimating image quality, this was calculated between the moving and the warped image. The SSIM quantifies the quality between two images and ranges between 0 and 1, where 1 represents the highest quality.

Results

To demonstrate the effectiveness of the proposed model configurations, a qualitative and quantitative evaluation was performed. For qualitative evaluations we registered sequences and visualized them frame by frame by generating cines using the Graphic Interchange Format (GIF). Figure 1 shows the results of a CT perfusion sequence obtained from a 10-cascade model trained with L_{nc} (see Eq.3) and L_c (see Eq.4), respectively. The fixed (reference) image shows a good contrast between the LV cavity and myocardium, while the selected “moving” image represents a frame of the sequence before the contrast agent occurs in the LV (see first line: Fixed image shows perfect contrast between cavities and tissue. Moving image: no contrast agent appeared in the heart. Second line: Fixed image reveals perfect contrast. Moving image: agent already occurred in the right atrium and ventricle).

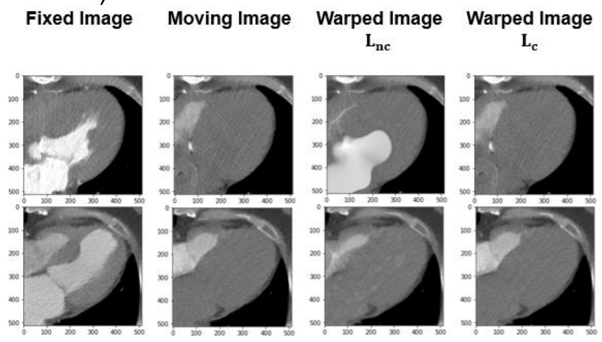

Figure 1: Results of sequence registration using a 10 cascade model with L_{nc} and L_c, respectively.

We also visualized the generated flow fields to assess the deformations performed by the models. Figure 2 shows the warped image and 10th flow field applied to the frames in the first row of Fig. 1. Here the norm of the vectors in the flow field is shown in a color scale, where red and blue represent large and small displacements, respectively. As expected, in the first column, L_{nc}, large displacements can be observed in the contrast regions while in the second column, L_c,, more uniform displacements can be seen in the contours.

The quantitative analysis was carried out using the evaluation metrics described above. We assessed the registration of the sequence by estimating the Dice score at frame level. Moreover, to investigate changes in the image quality, we calculated the MI between the moving and the warped image. The aim was to examine possible quality loss due to deformations. Figure 3 shows the mean Dice score and mean MI of one full CT perfusion sequence using a 10 cascade model trained with L_{nc} and L_c, respectively. The peak values are obtained at the frame selected as fixed.

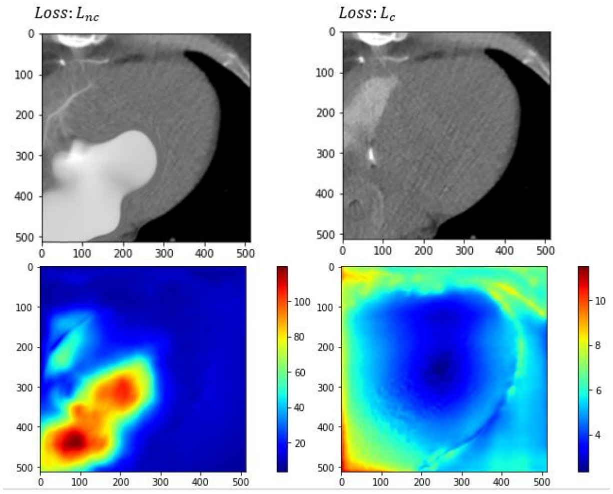

Figure 2: Warped images and flow field of two frames using a 10 cascade model.

Finally, to identify the optimal configuration in terms of the number of selected cascades we performed a quantitative analysis for all model configurations. Table 1 presents the means of the Dice score, MI and SSIM for the 3, 5, 7, 10, 15 and 20 cascade configurations obtained from the sequences in the validation set.

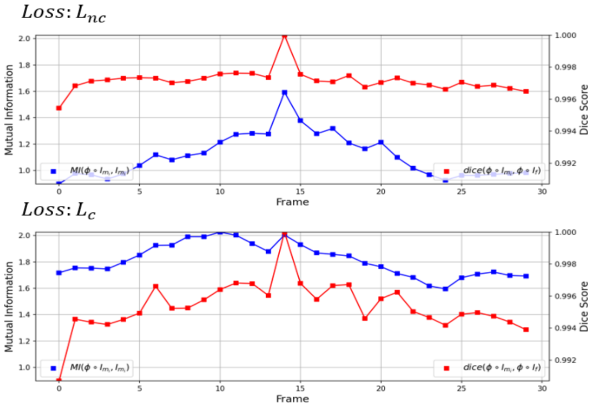

Figure 3: Evaluation metrics for one patient sequence using a 10 cascade-model

DOI: 10.3217/978-3-85125-826-4-18

Table 1: Evaluation metrics for different model configurations for the validation set.

N	Dice		MI		SSIM	
	L_{nc}	L_c	L_{nc}	L_c	L_{nc}	L_c
3	0.998	0.998	1.35	1.95	0.846	0.984
5	0.998	0.996	1.44	1.93	0.874	0.992
7	0.998	0.998	1.21	1.90	0.8	0.982
10	0.998	0.998	1.20	1.87	0.8	0.979
15	0.998	0.998	1.17	1.84	0.79	0.979
20	0.998	0.998	1.16	1.70	0.79	0.965

Discussion

In this work different model configurations of the Recursive-Cascade-Network were implemented and tested with the aim to identify the optimal configuration for the registration of myocardial perfusion CT sequences. Firstly, we evaluated the effect of the loss function at different cascade numbers. The results showed that training the model based on a loss function that does not penalizes the changes in the contrast regions negatively affects the quality of the warped image. Figure 1 shows an example of quality degradation of the warped image when the fixed and moving image have different contrast regions. It can be noted that the model deforms and introduces artifacts in these regions (see Fig 1 and Fig 2) that were observed for all investigated cascade configurations. Moreover, the models' performance was quantitatively assessed as shown in Fig 3 for a 10 cascade model as an example. Here, the quality metrics between the moving and the warped image were considerably lower in the L_{nc} model. Interestingly, the Dice score was higher than for the L_c model. However, considering only this measure as evaluation metric does not fully reflect the overall performance of the model. The latter can also be observed in Table 1, where despite achieving a high Dice score, all remaining quality measures (MI and SSMI) are lower. Finally, according to Table 1 and based on visual inspection, the best trade-off between spatial alignment and image quality can be achieved with a n=7 cascade model trained using the L_c loss function.

In summary, in this work we introduced a powerful loss function for optimizing the registration problem in cardiac CT perfusion imaging. Moreover, the effects of different model configurations were evaluated and they showed that considering the Dice score only as an evaluation metric does not fully represent the model performance of this approach in terms of spatial registration accuracy and image quality. In our future work we will investigate the effect of using other similarity metrics for the loss function and will evaluate the impact of feeding more contrast information into the network.

References

[1] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, “Deep learning in medical image registration: a review,” p. 48, 2020.

[2] T. Mäkelä, T. Katila, and I. E. Magnin, “A Review of Cardiac Image Registration Methods,” *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 21, no. 9, p. 11, 2002.

[3] H. Wiputra, “Cardiac motion estimation from medical images: a regularisation framework applied on pairwise image registration displacement fields,” *Scientific Reports*, p. 14, 2020.

[4] K. A. Lara Hernandez, T. Rienmüller, D. Baumgartner, and C. Baumgartner, “Deep learning in spatiotemporal cardiac imaging: A review of methodologies and clinical usability,” *Computers in Biology and Medicine*, vol. 130, p. 104200, Mar. 2021, doi: 10.1016/j.compbiomed.2020.104200.

[5] D. Caruso *et al.*, “Dynamic CT myocardial perfusion imaging,” *European Journal of Radiology*, vol. 85, no. 10, pp. 1893–1899, Oct. 2016, doi: 10.1016/j.ejrad.2016.07.017.

[6] G. Haskins, U. Kruger, and P. Yan, “Deep Learning in Medical Image Registration: A Survey,” *Machine Vision and Applications*, vol. 31, no. 1–2, p. 8, Feb. 2020, doi: 10.1007/s00138-020-01060-x.

[7] S. Zhao, T. Lau, J. Luo, E. I.-C. Chang, and Y. Xu, “Unsupervised 3D End-to-End Medical Image Registration with Volume Tweening Network,” *IEEE J. Biomed. Health Inform.*, vol. 24, no. 5, pp. 1394–1404, May 2020, doi: 10.1109/JBHI.2019.2951024.

[8] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “VoxelMorph: A Learning Framework for Deformable Medical Image Registration,” *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019, doi: 10.1109/TMI.2019.2897538.

[9] S. Zhao, Y. Dong, E. I.-C. Chang, and Y. Xu, “Recursive Cascaded Networks for Unsupervised Medical Image Registration,” *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10599–10609, Oct. 2019, doi: 10.1109/ICCV.2019.01070.

[10] L. R. Dice, “Measures of the Amount of Ecologic Association Between Species,” *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945, doi: 10.2307/1932409.

[11] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer, “Image Similarity Using Mutual Information of Regions,” in *Computer Vision - ECCV 2004*, vol. 3023, T. Pajdla and J. Matas, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 596–607. doi: 10.1007/978-3-540-24672-5_47.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. on Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.