Deep Learning-Based Image Registration in Dynamic Myocardial Perfusion CT Imaging

A. Lara-Hernández⍟, *Member, IEEE*, T. Rienmüller⍟, I. Juárez⍟, M. Pérez, F. Reyna⍟, D. Baumgartner⍟, V. N. Makarenko, O. L. Bockeria, M. Maksudov, R. Rienmüller, and C. Baumgartner⍟

Abstract— Registration of dynamic CT image sequences is a crucial preprocessing step for clinical evaluation of multiple physiological determinants in the heart such as global and regional myocardial perfusion. In this work, we present a deformable deep learning-based image registration method for quantitative myocardial perfusion CT examinations, which in contrast to previous approaches, takes into account some unique challenges such as low image quality with less accurate anatomical landmarks, dynamic changes of contrast agent concentration in the heart chambers and tissue, and misalignment caused by cardiac stress, respiration, and patient motion. The introduced method uses a recursive cascade network with a ventricle segmentation module, and a novel loss function that accounts for local contrast changes over time. It was trained and validated on a dataset of n = 118 patients with known or suspected coronary artery disease and/or aortic valve insufficiency. Our results demonstrate that the proposed method is capable of registering dynamic cardiac perfusion sequences by reducing local tissue displacements of the left ventricle (LV), whereas contrast changes do not affect the registration and image quality, in particular the absolute CT (HU) values of the entire CT sequence. In addition, the deep learning-based approach presented reveals a short processing time of a few seconds compared to conventional image registration methods, demonstrating its application potential for quantitative CT myocardial perfusion measurements in daily clinical routine.

Index Terms— Registration, deep learning, dynamic cardiac imaging, computed tomography, myocardial perfusion.

I. INTRODUCTION

DEFORMABLE image registration (DIR) is a preprocessing technique that finds non-linear spatial transformations for the alignment of an image pair. In medical image analysis, this is essential for many clinical applications where spatial alignment of anatomical structures is required. These modalities include image-guided procedures for diagnostics and patient management where images are acquired at different points in time or using different modalities [1], [2]. Especially in cardiac image analysis, DIR is used in image-guided interventions that require myocardial motion tracking or in myocardial perfusion studies [3], [4], [5].

Heart perfusion studies are performed to quantitatively estimate myocardial perfusion and assess different degrees of myocardial ischemia in patients with known or suspected coronary artery disease. Contrast enhanced myocardial perfusion studies allow the evaluation of contrast agent distribution in the heart chambers and tissue using dynamic CT or MRI sequences. During a CT examination, a contrast agent is administered and an image sequence is obtained using an ECG-gated protocol that normally acquires image data at the end-systolic phase. The image sequence is qualitatively assessed and quantitatively evaluated through time attenuation curves that are used to identify and detect ischemic areas in the myocardium characterized by hypo attenuation (reduced CT values) [6], [7], [8]. However, registration of the image sequence is required for accurate generation of such curves, as spatial misalignment may occur due to cardiac stressing, patient and respiratory movement. This task includes some unique challenges besides the motion of the thorax such as the non-rigid dynamic nature of the heart, less accurate anatomical landmarks, and the contrast agent distribution that changes over time in the heart, adding an additional level of complexity to the registration problem.

Manual medical image registration is a difficult, time-consuming and clinician-dependent task. Several methods have

A. Lara-Hernández is with the Institute of Health Care Engineering With European Testing Center of Medical Devices, Graz University, 8010 Graz, Austria, and also with the Department of Biomedical Engineering, Galileo University, Guatemala City 01010, Guatemala (e-mail: andrealh@galileo.edu).

T. Rienmüller and C. Baumgartner are with the Institute of Health Care Engineering With European Testing Center of Medical Devices, Graz University, 8010 Graz, Austria (e-mail: theresa.rienmueller@tugraz.at; christian.baumgartner@tugraz.at).

I. Juárez and M. Pérez are with the Department of Biomedical Engineering, Galileo University, Guatemala City 01010, Guatemala (e-mail: alexanderj@galileo.edu; mperez@galileo.edu).

F. Reyna is with the Faculty of Medicine, Francisco Marroquin University, Guatemala City 01010, Guatemala, and also with the Department of Radiology, La Paz Hospital, Guatemala City 01010, Guatemala (e-mail: freyna@ufm.edu).

D. Baumgartner and R. Rienmüller are with the Clinical Division of Pediatric Cardiology and the Department of General Radiology, Medical University of Graz, 8036 Graz, Austria (e-mail: daniela.baumgartner@medunigraz.at; rainer@rienmueller.net).

V. N. Makarenko and O. L. Bockeria are with the A. N. Bakulev National Medical Research Center of Cardiovascular Surgery, 121552 Moscow, Russia (e-mail: vn7902352@yandex.ru; bockeria@me.com).

M. Maksudov is with the Fedorovich Klinikasi and the Department of Radiology Vakhidov Republican Specialized Center of Surgery, Tashkent 100061, Uzbekistan (e-mail: dr.mfmaksudov@gmail.com).

been proposed to automate this process in response to these issues. Traditional methods such as ANTs [9], demons [10], and ELASTIX [11] use iterative methods to find the optimal transformation. Approaches of this kind have achieved state-of-the-art performance, but they are time-intensive and therefore not sufficiently applicable in the clinical routine, especially when a large number of images needs to be registered.

Motivated by the demand of faster algorithms and the state-of-the-art performance achieved in other medical image analysis applications [5], [12], deep learning-based methods, which represent an emerging area of research, have recently been proposed for image registration [1], [2]. Supervised and unsupervised approaches were introduced for mono- and multi- modal registration, intra- and inter-patient registration, and motion tracking [1], [2], [5]. Supervised methods use ground-truth labels to estimate the deformation field used to register the pair of images. Several supervised approaches have outperformed traditional DIR methods [13], [14], however, they are difficult to implement due to ground-truth deformations. By contrast, unsupervised methods do not require such ground-truth transformations and perform image registration using a deformation field that is estimated by measuring the similarity between pairs of images (fixed and warped moving images). The transformation is applied to the image within the network, for example with a Dense Displacement Vector Field (DVF) and a Spatial Transformer Network (STN) [15]. The latter is a differentiable module that performs the warping of the image and can be introduced into any Convolutional Neural Network (CNN) architecture.

As mentioned above, deep learning-based methods for medical DIR have been proposed recently. Most of these focus on regions of interest (ROIs) in more "static" organs such as the brain, prostate or liver [1]. In the case of the cardiac DIR, however, only a few methods have been suggested. For example, De Vos et al. [16] introduced DIRNet, an unsupervised approach that estimates the parameters for local deformations using a CNN and outputs a grid of 2D control points, which in turn are the input into the STN module to generate a DVF using cubic B-spline. The latter is used by a resampler to warp the moving image. In a follow-up work, the authors presented an extension of [16] and introduced a Deep Learning Image Registration (DLIR) framework [17] that performs unsupervised affine and deformable 3D image registration. Both methods were trained and tested on cardiac cine MRI scans. Sheikhjafari et al. [18] proposed a fully connected neural network (FCN) for unsupervised DIR of 2D cine MRI scans. This method uses a latent vector as input, which is fed into an FCN to generate the two-dimensional DVF. The registration is based on the generated DVF and Bilinear Interpolation. Fechter and Baltas [19] introduced a one-shot learning DIR method for periodic motion tracking in 3D and 4D data. This approach employs a U-Net and a differential spatial transformer module for estimation of the displacement fields; for 3D data, it also generates the inverse DVF. Otherwise, Mahapatra et al. [20] presented a new approach using Generative Adversarial Networks (GANs) for cardiac cine MRI scans (2D data). This model is trained to learn the density function describing all plausible deformation fields to generate realistic registered images. The results obtained with this approach outperformed other methods such as ELASTIX and DIRNet. However, the application of the previous methods is limited to cardiac sequences without changes in the concentration of contrast agent over time. When registering myocardial perfusion sequences, however, local contrast changes over time must be taken into account, which makes registration in perfusion imaging a challenging research area in DIR.

In cardiac MRI perfusion imaging non-deep-learning based methods have been proposed for sequence registration. Milles et al. [21] for example introduced an approach based on independent component analysis (ICA) for translational motion correction. The method consists of an ICA analysis and a two-pass registration framework. ICA is applied to extract relevant perfusion information for computing a time-varying reference image that is used subsequently by the two-pass registration to align the sequence frames. Moreover, Wollny et al. [22] proposed a three stages registration method that exploits the quasi-periodicity of respiratory motion. In the first phase, a global reference and a subset of images corresponding to the same respiratory phase are selected. A semi local B-spline transformation is then applied to register the obtained subset to the global reference. Finally, the remaining images are aligned to synthetic references generated by a weighted linear combination of the frames of the registered subset. Similarly, Scannell et al. [23] introduced a two-stage procedure using robust principal component analysis (RPCA) and synthetic image series. In the first stage, bulk motion is corrected by applying a rigid registration to a low-rank image sequence obtained by RPCA. The images are then registered to a synthetic image series obtained using principal component analysis (PCA) in a refinement phase. Here, the registration is performed using free-form deformations.

Traditional methods have also been introduced for registration in cardiac CT perfusion imaging. Isola et al. [24] introduced a non-rigid registration method using cubic B-splines. In this approach a spatiotemporal diffusion filter is first applied to the sequence for reducing noise and artifacts. Registration is then performed using a multi-resolution approach to create downsampled versions of the original moving and fixed images, and deformation fields are computed using cubic B-splines. The fixed image here is the image frame with the largest contrast agent quantity and the *zero mean normalized cross-correlation* (ZNCC) is used as a similarity metric. Similarly, Tang et al. [25] proposed a registration method based on the estimation of time-dependent motion vector fields (MVF) using cubic B-splines. The method uses a motion compensation image reconstruction to align 4D image data. This reconstruction is based on a motion-tracking algorithm that is guided by the generated MVFs. Moreover, Liu et al. [26] presented a motion compensation algorithm for low dose dynamic CT. Here, RPCA is applied for image decomposition to obtain the low rank components of the sequence. Then, an optical flow-based method is used to estimate the deformations fields and register the low rank frames. In addition, a low rank regularized image reconstruction is performed using the

obtained MVFs to improve the sequence registration. Finally, Janssens et al. [27] introduced a diffeomorphic registration method for dynamic contrast-enhanced images. The method uses morphons to register the images based on the local intensity phase. This follows a multiscale nonparametric approach that computes the diffeomorphic displacements fields applied from a coarse to a fine scale.

In this paper, we introduce for the first time a DIR deep learning-based method for quantitative myocardial perfusion CT studies, which unlike other previous approaches, addresses the following challenges: 1) the algorithm needs to deal with lower image quality, in terms of contrast resolution with less accurate anatomical landmarks and less signal-to-noise ratio in cardiac CT compared to MRI. 2) the dynamic information from the contrast agent must remain unaffected, so deformations should not have an impact on the absolute CT (HU) values in the image, particularly in the LV myocardium and cavity as regions of interests (ROIs). 3) registration must be performed on the entire image sequence, taking into account and correcting for misalignment caused by cardiac stressing, respiration and patient motion, 4) the method must have a significantly improved processing time compared with conventional methods, so that this approach can be used in the clinical routine for quantitative myocardial perfusion measurements.

The presented deep learning-based method has a recursive cascade architecture with a pre-trained segmentation module, and introduces a novel loss function based on contrast and image similarity losses, and auxiliary segmentation data to guide the registration. The method is trained and evaluated on a representative dataset of patients with known or suspected coronary artery disease and/or aortic valve insufficiency to demonstrate its power for spatiotemporal DIR, accounting for local contrast changes over time, while preserving the image quality of the warped images.

II. MATERIAL AND METHODS

A. Dataset

Image sequences were obtained from n = 118 patients with known or suspected coronary artery disease and/or aortic valve insufficiency undergoing dynamic CT myocardial perfusion examinations with a weight-based adjustment of exposure settings. All patients gave informed consent (clinical trial NCT02361996, "Quantitative measurement of myocardial perfusion by cardiac CT in patients (CCT/MPERF)").

Tube voltages of 80/100/120kV were selected depending on patient weight. Scanning parameters were set to axial acquisition mode, 5 mm slice thickness, gantry/detector tilt 0°, and 180 ms exposure time. Image acquisition protocol was prospectively ECG-triggered at 40 % of the R-R interval every heartbeat. A contrast agent with a concentration of 210 mg Iodine/ml (Visipaque R 300 mg) was injected intravenously with a constant flow of 4 ml/s (CT- injector "ohio tandem", ulrich medical®). All examinations were performed using a dose modulation technique with a 256-slice multidetector CT scanner (Brilliance i256; Philips Healthcare). Functional parameters such as heart rate, systolic/diastolic blood pressure, stroke volume (SV), and potential risk factors for decreased perfusion (calcium score, stenotic lesions, high body mass index (BMI)) were recorded for clinical characterization purposes.

In total, image sequences from 13 to 14 slices volume stacks per patient were acquired over 30-40 heartbeats. From the latter, only the 2D sequences that cover the ROI, i.e. LV myocardium, were selected, resulting in 944 2D sequences (image size 512×512 pixels) composed of an average of 35 frames. Image contrast was set to the default CT window W:750 L:90. Masks of the regions with high contrast agent concentration were obtained using a CT window of W:450 L:130.

Finally, the available dataset was split on subject-level, n = 95 (80 %) and n = 23 (20 %) for cross-validation and testing, respectively. Moreover, to further evaluate the performance of the registration, segmentations of LV cavity (including papillary muscles and trabeculae) and LV myocardium from the test dataset (n = 23) were performed manually by a radiologist. Because of the time required for this task, we selected 4 frames per patient for the segmentation (i.e. 92 frames in total). These frames included the fixed image and three moving images at different time points of the contrast agent distribution in the LV.

B. Registration Network

Let $S = \{I_{m_i}\}_{i=0}^{N}$ denote a cardiac CT perfusion sequence of images, where $I_{m_i} \in \Omega \subset R^2$. Let denote I_m and I_f as moving image and fixed image, respectively. We want to generate a flow prediction function, F, that takes I_m and I_f to predict a flow field $\varphi : \Omega \to \Omega$ that aligns the sequence S. Based on the work of Zhao et al. [28] and our previous work [29], we extended the basic recursive cascade network architecture for 2D sequences by (i) including a *ventricle segmentation module* and (ii) introducing a *novel loss function* for optimizing dynamic cardiac image registration.

The method follows an n-cascade architecture which decomposes the registration into progressively small deformations that are recursively applied to the warped image. The final architecture is presented in Fig. 1.

Each cascade represents a base network such as DLIR [17], VTN [30] or Voxelmorph [31], and operates as a flow prediction function f that takes a pair of images $\left(I_m, I_f\right)$ and predicts a flow field φ by aligning the moving image I_m to the fixed image I_f. For the k-th cascade the predicted flow field φ_k can be obtained as shown in (1).

$$\varphi_k = f_k(I_m^{(k-1)}, I_f) \tag{1}$$

where f_k is the flow prediction function of each k-th cascade. The warped image is the composition of the flow field φ_k and the moving image $I_m^{(k-1)}$ as shown in (2).

$$I_m^{(k)} = \varphi_k \circ I_m^{(k-1)} \tag{2}$$

Following the recursive model, the final flow prediction function F is a composition of all generated flow fields as denoted in (3).

$$F\left(I_m, I_f\right) = \varphi_n \circ \varphi_{n-1} \circ \ldots \circ \varphi_1 \tag{3}$$
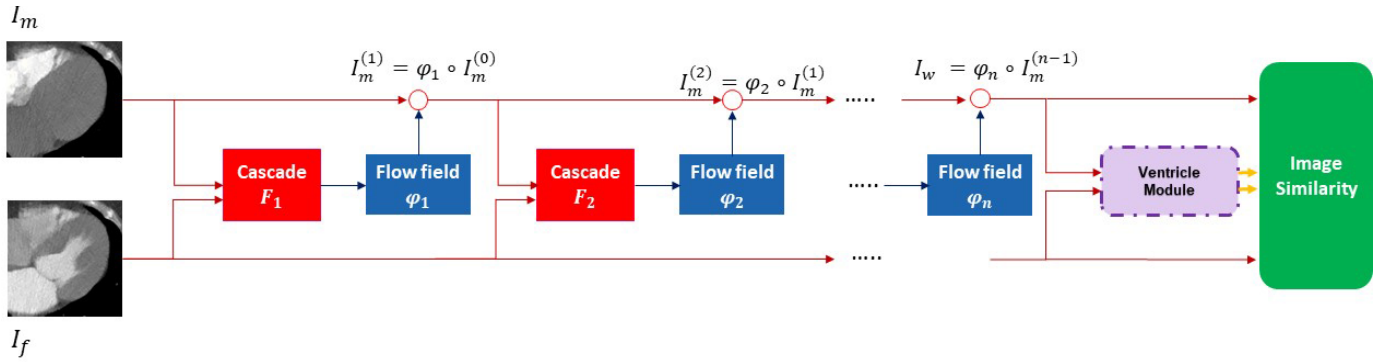
Fig. 1. Recursive cascade network architecture with ventricle segmentation module for dynamic cardiac image registration.

Therefore, the final warped image I_w is obtained by successively warping the moving image along all cascades as presented in (4).

$$I_w = (\varphi_n \circ \varphi_{n-1} \circ \ldots \circ \varphi_1) \circ I_m \qquad (4)$$

1) Ventricle Module: To improve the alignment of the LV, we added a module used to segment the ventricles from the warped image I_w and the fixed image I_f. The objective of this module is to provide auxiliary information that can support the registration of the anatomical structures. It was implemented using the original U-Net architecture presented in [32], and it was trained and validated using masks of the RV and LV chambers. These segmentations were performed manually by a radiologist and a non-clinical expert who was trained by the medical expert on how to recognize and outline cardiac structures using Medviso Segment CT, a commercial software for quantitative cardiac CT analysis (http://segment.heiberg.se). All segmentations were obtained from a randomly selected subset of n = 50 patients from the training dataset and were visually assessed by the expert to identify and remove poor quality segmentations. Finally, a total of 1562 masks of the RV and LV were used to train and validate the model (90 % and 10 %, respectively).

C. Loss Functions

We introduced a novel loss function L_{cv} as to train our model for the registration of dynamic cardiac perfusion sequences in an unsupervised manner, as presented in (5). Analogously to previous unsupervised approaches [30], [31] the loss function is generally based on a similarity loss and a regularization term as described in (6).

$$
\begin{aligned}
L_{cv} &\left(I_f, I_m, \varphi, M_c, M_{RV}, M_{LV} \right) \\
&= \alpha_1 L_{sim} \left(I_f, \varphi I_m \right) + \alpha_2 L_{cont} \left(I_f, \varphi \circ I_m, M_c \right) \\
&\quad + \alpha_3 L_{vent} \left(M_{RV}, M_{LV} \right) + L_{reg}(\varphi) \qquad (5)
\end{aligned}
$$

$$
\begin{aligned}
L_{nc} &\left(I_f, I_m, \varphi \right) \\
&= L_{sim} \left(I_f, \varphi \circ I_m \right) + L_{reg}(\varphi) \qquad (6)
\end{aligned}
$$

where L_{cv} and L_{nc} denote loss functions *with* and *without* contrast information, respectively. In both losses, L_{sim} is a similarity loss used to penalize the difference in appearance between the fixed and warped image, and L_{reg} is the regularization term employed to smooth the flow field and prevent

unrealistic displacements. In contrast to L_{nc}, the loss function L_{cv} includes the contrast concentration loss L_{cont} that uses masks of the regions with high contrast agent concentration, M_C, to account for changes of the contrast agent over time and penalize alterations in the contrast regions, and the ventricle loss L_{vent} that enforces the alignment of the ventricles using the masks generated by the ventricle module i.e. M_{LV} and M_{RV}. The terms $\alpha_1 \alpha_2$ and α_3 are the weighting factors of the losses.

The motivation for adding new terms to the original loss function, L_{nc}, is to perform the registration without affecting the relevant areas needed for the study. We therefore introduced the functions L_{cont} (9) and L_{vent} (11). As demonstrated in this study, L_{nc} negatively affects the regions of high contrast concentration (right/left atrium or ventricles). This applies in particular to cases where the contrast regions of the fixed image differ from those of the moving image. In these cases, the final warped image results with contrast regions that differ from the original moving image.

1) Similarity Loss: We used the *Pearson correlation coefficient PCC* to measure the degree of linear correlation in a pixel manner between the images I_f and I_w. The PCC returns a value between the range $[-1, 1]$ that indicates the degree of correlation between the two images. A PCC value of 1 corresponds to a perfect correlation and a PCC value of 0 indicates two linearly uncorrelated images. Therefore, to penalize the dissimilarity between the pair I_f and I_w, we defined L_{sim} as denoted in (7).

$$L_{sim} \left(I_f, \varphi \circ I_m \right) = 1 - PCC(I_f, \varphi \circ I_m) \qquad (7)$$

2) Regularization Loss: We implemented the *Total Variation Loss* to encourage the continuity of the flow field, as presented in [30]. For a 2D flow field, the loss function is defined as follows:

$$L_{reg}(\varphi) = \frac{1}{2|\Omega|} \sum_x \sum_{i=1}^{2} (\varphi(x + e_i) - \varphi(x))^2 \qquad (8)$$

where $e_{1,2}$ form the natural basis of $\mathbb{R}^2$.

3) Contrast Concentration Loss: If M_c contains the masks of high contrast concentration areas of every image in the sequence S and $C_i \subset I_{m_i}$ denotes the contrast regions in the i-th moving image, we want to preserve as much information

about C_i in the warped image I_{w_i}. as possible. We thus need the flow field φ to minimally affect the regions C_i while correcting for misalignment. We introduce L_{cont}, for this task denoted by (9), as a term that uses the regions C_i to reduce the changes of contrast in such areas. As shown in (10), C_i is generated using the Hadamard product between the image and the mask containing the previously delineated areas of high-contrast concentration.

$$
\begin{aligned}
L_{cont} & \left(I_f, \varphi \circ I_m, M_c \right) \\
& = L_{sim} \left(I_m \odot M_{c_m}, (\varphi \circ I_m) \odot M_{c_m} \right) \\
& \quad + L_{sim} \left(I_m \odot M_{c_f}, (\varphi \circ I_m) \odot M_{c_f} \right) \quad (9)
\end{aligned}
$$

$$
C_i = I_{m_i} \odot M_{c_i} \quad (10)
$$

L_{cont} is defined by two terms that are used to guide the deformation of the warped image. The first term penalizes the alteration of contrast from the moving image in the warped image by estimating the similarity between the original contrast regions and the new contrast regions after deformation. The second term penalizes the introduction of contrast into the warped image. This step is necessary because the fixed image may contain contrast regions that are not present in the moving image. Our aim is thus to preserve the contrast of such regions as in the original moving image. We again estimate the similarity in penalizing the differences in the contrast in such areas. The previous is measured using L_{sim}, and, M_{c_m} and M_{c_f} are the masks of the moving and fixed image, respectively.

4) Ventricle Loss: For this particular task, the registration of the ventricle is very important to accurately estimate myocardial perfusion. For this reason, we included the term L_{vent} that emphasizes the registration of such areas in addition to the overall registration estimated by L_{sim}. M_{RV_i} and M_{LV_i} denote the masks predicted by the *ventricle module* of the right and left ventricle of the *i-th* moving image, respectively. L_{vent} is estimated as shown in (11) to measure the alignment of the right and left ventricle between the fixed and the warped image, respectively.

$$
\begin{aligned}
L_{vent} & (M_{RV}, M_{LV}) \\
& = \frac{1}{2} \left(L_{seg} \left(M_{RV_f}, M_{RV_w} \right) + L_{seg} \left(M_{LV_f}, M_{LV_w} \right) \right) \quad (11)
\end{aligned}
$$

where L_{seg} for each of the ventricle masks is defined as (12).

$$
L_{seg} \left(M_{V_f}, M_{V_i} \right) = 1 - Dice(M_{V_f}, M_{V_i}) \quad (12)
$$

D. Implementation

Our proposed configuration was implemented in PyTorch using a modified 2D version of the original implementation (for 3D volumes) as published in [28]. We chose the VTN [30] for the base subnetwork based on our previous work [29]. The models were trained and validated using 5-fold cross validation on 2 TITAN RTX 25GB. We used the Adam optimizer [33] with two beta values of 0.9 and 0.999 and an initial learning rate of 0.0001. The batch size is 32 pairs per batch and a total of $8*10^4$ iterations were performed for each kth-fold. We used different values of α_i to train the model to find the best weighting factors experimentally, and multiple models were thus trained. For training the *ventricle module*, we used a batch size of 10 and learning rate of 0.001.

E. Evaluation Metrics

The performance of the method was evaluated based on the accuracy of the spatial alignment and the image quality after deformation. The spatial alignment was quantified using the *Dice Score (DSC)* [34] which measures the overlap between two regions. It ranges between 0 and 1, with 1 indicating perfect alignment. In this application, the DCS of segmentations of the whole heart was measured, and additionally the *Hausdorff Distance (HD)* was used as another measure of registration performance. For this evaluation, the distance of the contours of the heart (original versus registered) was determined.

The quality of the warped image with respect to the moving image was evaluated to measure possible alterations during the registration process, i.e. changes of the original CT values (in HU). For this task, the *mutual information (MI)* [35] and the *structural similarity index (SSIM)* were determined. The MI measures the statistical dependence between two images which is used to estimate the amount of information that the warped image contains about the moving image. SSIM measures the similarity between two images and it is used to assess image quality degradation. It ranges from 0 to 1, with 1 indicating the highest quality.

F. Experiments

First, the proposed improved model (LCV) was cross-validated, including the *ventricle module* and using the new *loss function L_{cv}* (5) that processes the contrast concentration loss. The model was 5 fold cross-validated on the training dataset of n = 95 patients with 3, 5, 7, 8, 9 and 10 cascades with the objective to determine empirically the best cascade configuration for this application.

Next, to investigate the strength of the proposed loss function and extension of the network architecture by the ventricle module, two variants of the LCV model were evaluated using the best cascade configuration determined from the previous experiment. The first model variant (LC) was trained and validated *without* the ventricle module, but with a loss function processing the contrast concentration loss (for details see loss function L_c in [29]). The second model variant (LNC) was trained based on the architecture originally presented by Zhao et al. [28] *without* segmentation module and using the loss function L_{nc} (6), *not* processing the contrast concentration loss.

Finally, to demonstrate the effectiveness of the LCV method, we benchmarked the performance of LCV with its two variants, i.e. LC and LNC, and two well established iterative methods presented by Wollny et al. [22], a nonrigid registration method for myocardial perfusion imaging and Janssens et al. [27], a diffeomorphic registration method for dynamic contrast-enhanced images. All five models were tested on 245 2D sequences of the test data set (n = 23 patients) with a total of 5638 image pairs.

III. RESULTS

The proposed method was evaluated quantitatively using the evaluation metrics described above and qualitatively by visual
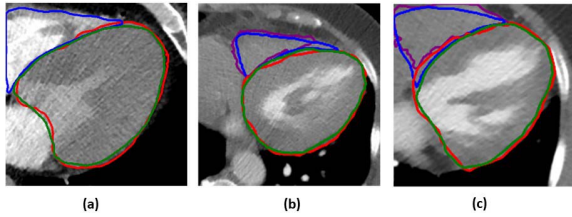
Fig. 2. RV and LV segmentations. Green and blue contours show manual annotations of the LV and RV, respectively. Purple and red contours denote predicted segmentations of the LV and RV by the *ventricle module*. (a)–(c) Image data from three selected patients.

TABLE I
VENTRICLE MODULE RESULTS

ROI	DSC	HD
LV	0.953 (0.02)	4.888 (0.51)
RV	0.857 (0.04)	5.886 (1.10)

*The standard deviations are in parentheses.

TABLE II
EVALUATION METRICS FOR N-CASCADES CONFIGURATIONS

Cascades	3	5	7	8	9	10
DSC	0.9801	0.9901	0.9958	0.9958	0.9960	**0.9962**
HD	5.681	5.072	4.527	4.521	4.515	**4.497**
MI	1.671	1.605	**1.563**	1.549	1.532	1.528
SSIM	0.938	0.905	**0.900**	0.897	0.890	0.887

TABLE III
COMPARISON OF IMAGE REGISTRATION USING DIFFERENT MODEL CONFIGURATIONS

Models	DSC	HD	MI	SSIM
LCV	0.9958	4.527	1.563	0.900
LC	0.9957	4.532	1.495	0.877
LNC	0.9956	4.516	0.877	0.784

inspection of registered 2D cardiac sequences. Videos of the unregistered and registered image sequences from 4 selected patients, comparing the registration performance of the five models, i.e. LCV, LC, LNC [28], Wollny et al. [22] and Janssens et al. [27] can be found in the Supplemental Material.

A. Ventricle Module

To evaluate the performance of the *ventricle module*, the generated masks of the RV and LV were assessed qualitatively and quantitatively. Examples of the qualitative analysis are shown in Fig 2. For the quantitative evaluation, we calculated the DSC and HD using the ground truth masks generated by the radiologist. Table I summarizes the results.

The results shown in Fig. 2 demonstrate that the *ventricle module* is capable of predicting masks of the RV and LV. The best results are obtained when the chambers have different contrast, as shown in Fig 2a. However, in cases where the contrast of the interventricular septum and the RV are similar (Fig 2b-c), the module performs better for the LV than for the RV segmentation (Table I). As mentioned earlier, the goal of this module is to provide auxiliary information that can assist in the alignment of the cardiac structures, particularly the alignment of the LV, as this is the ROI for myocardial perfusion studies. The impact of this module is further evaluated using the ablation studies presented in Section IIIC.

B. Weighting Factors and Number of Cascades

To determine the best weighting factors and number of cascades, our LCV model was 5-fold cross-validated using image sequences of the training patient cohort (n = 95). We evaluated the performance by estimating the average of the evaluation metrics obtained at the validation of each kth-fold. The spatial alignment metrics DSC and HD were computed between the fixed image and the warped images of a sequence and image quality metrics were estimated from the original moving image and the warped images. Based on the results we established experimentally that weighting factors of $\alpha_i = \{0.6, 0.25, 0.15\}$ gave the best results. Table II summarizes the results obtained with α_i and different cascade configurations. From Table II, 10 cascades yielded the best results for the spatial alignment metrics DSC and HD, indicating that increasing the number of cascades improves the registration performance, as suggested in [28]. When comparing the image quality metrics MI and SSIM, we observed a decrease in the values, suggesting that the increase in the number of cascades has a negative impact on the quality of the registered images. To evaluate this observation, we performed a qualitative analysis to assess the registered sequences obtained with each cascade configuration. After visual inspection and the analysis of the quantitative results from Table II, we found that a configuration with 7 cascades provides the best trade-off between spatial alignment and image quality.

C. Ablation Studies

Based on these results, the model variants LC and LNC [28] were implemented in a 7-cascade configuration. To compare the performance with LCV, all models were cross-validated and evaluated as in the previous experiment. Table III summarizes the results obtained for the different model configurations.

Table III shows that all three models based on the DSC and HD measures achieved similar registration performance. However, the results obtained in MI and SSIM indicate that the addition of the contrast concentration loss term L_{cont} (9), as in LCV and the variant LC, helps to minimize the effects of the deformation on the image quality, demonstrating that the segmentation module further improves the registration of the dynamic sequences, especially with respect to the image quality measures MI and SSIM. The original deep learning method LNC of Zhao et al. [28] however showed the worst results.

Based on these findings, all three models were retrained with the complete training dataset of n = 95 subjects (without cross-validation). These models were then used for the performance evaluation of the next section.
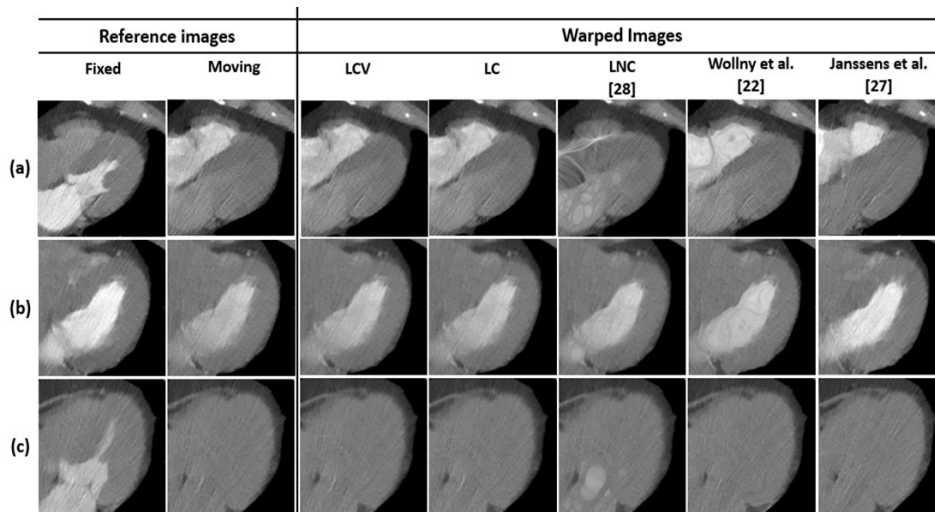
Fig. 3. Comparison of image registration performance of three selected patients using the methods LCV, LC, LNC, Wollny et al., and Jassens et al. Fixed image, moving image to be registered and warped (registered) image. (a)–(c) Image data of three selected patients (see also Fig. 4).
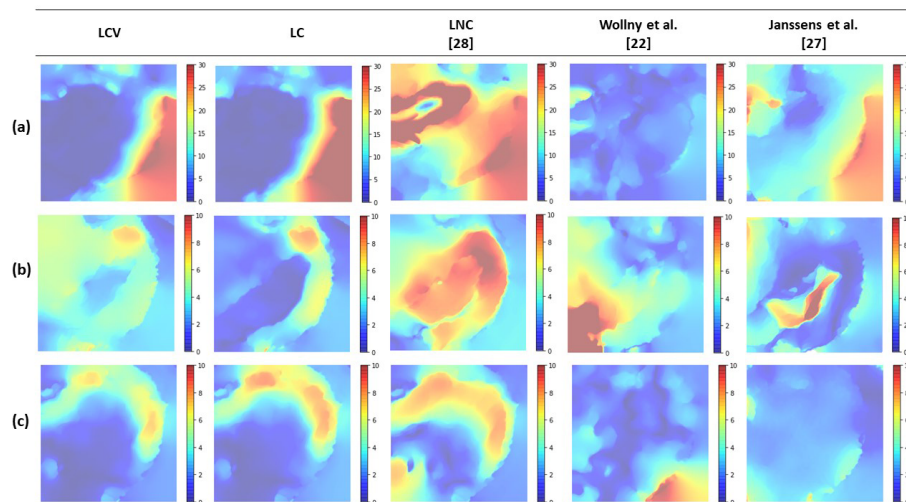

Fig. 4. Generated flow fields of the examples in Fig. 3. Color scale codes the norm of the vectors in the flow field. Color red and blue represent large and small displacements, respectively. (a)–(c) Same patient image data as in Fig. 3.

D. Comparison of Registration Performance With Other Methods

We tested the performance of the three deep learning-based models LCV, LC and LNC on the image dataset of the test patient cohort (n = 23) and compared the results with two established iterative, non-deep learning-based methods of Wollny et al. [22] and Janssens et al. [27]. For these methods, we used the implementation of [22] for motion compensation for myocardial perfusion imaging from [36], and for [27] a MATLAB implementation of the Morphons algorithm from [37]. Fig. 3 shows examples of image sequence registration performed with each method in three selected patients at different time points of contrast agent distribution.

In detail, the first two columns show the fixed and a selected image of the image sequence to be registered (moving image), followed by the registered (warped) images obtained by the five compared methods LCV, LC, LNC [28], Wollny et al.

[22] and Janssens et al. [27]. Fig. 3a shows an example in which the areas of high contrast differ significantly in both the fixed (almost no contrast agent in the right chambers) and the selected image of the sequence (almost no contrast agent in the chambers of the left heart), Fig. 3b is similar to Fig. 3a, but already shows contrast agent in the left chambers of the moving image, and Fig. 3c demonstrates only contrast agent in the left atrium and ventricle of the fixed image, but no contrast agent in the moving image at all. Fig. 4 illustrates the generated deformation flow fields by the five methods presented in Fig. 3. Visual inspection shows that LNC and Wollny et al. [22] produce artifacts in the warped image (patients a and c). Both, LNC and [22] are not able to correctly register the high contrast areas in the right chambers (a), and LNC also introduces artificial gray areas (a, c). In contrast, the models LCV and LC, and the iterative method of Janssens et al. [27] demonstrate good performance, registering the images without introducing artifacts. It should be noted that in [22]
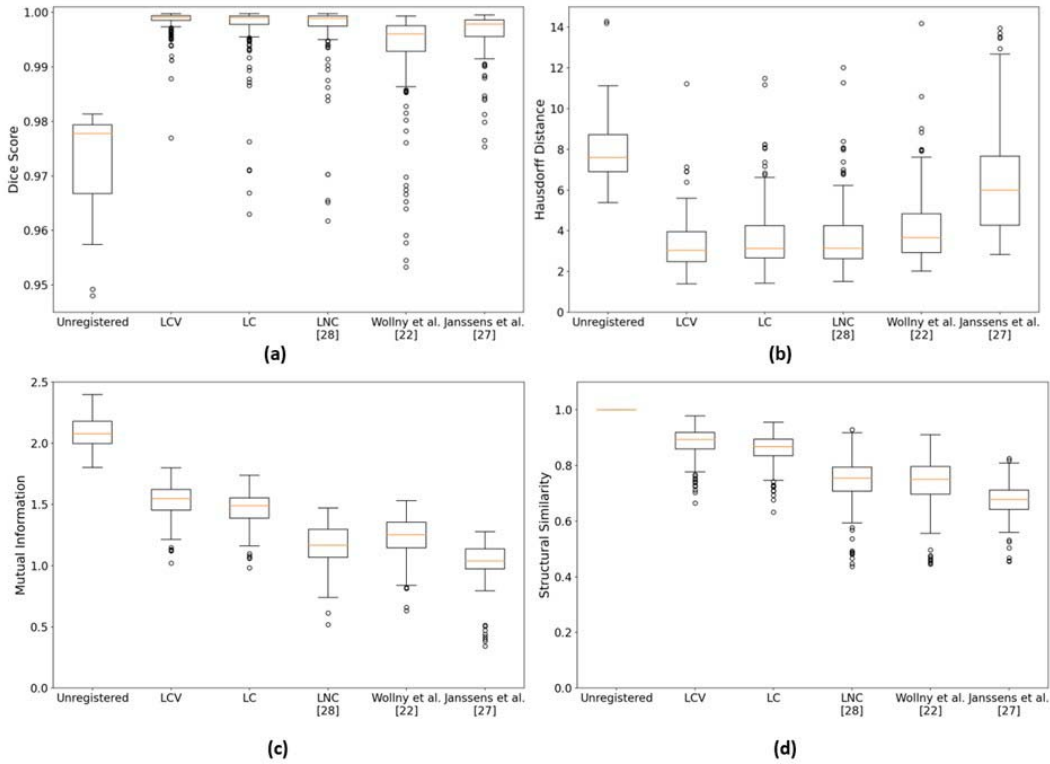
Fig. 5. Box plot analysis of spatial alignment metrics and image quality metrics obtained from 2D cardiac sequence registration of the test dataset. (a) mean DSC, (b) mean HD, (c) mean MI, and (d) mean SSIM of the registered sequences for each method compared with the unregistered images.

TABLE IV
COMPARISON OF THE DIFFERENT REGISTRATION METHODS

Methods	DSC	HD	MI	SSIM	Run Time (s)	
					CPU	GPU
Unregistered	0.973 (0.010)	7.863 (1.37)	2.087 (0.130)	1.000 (0.000)	-	-
LCV	**0.998 (0.002)**	**3.370 (1.63)**	**1.523 (0.139)**	**0.881 (0.056)**	10.79 (2.05)	0.92 (0.18)
LC	0.997 (0.002)	3.617 (1.60)	1.460 (0.138)	0.857 (0.055)	10.78 (2.11)	0.91 (0.18)
LNC [28]	0.997 (0.003)	3.632 (1.61)	1.165 (0.177)	0.745 (0.095)	10.77 (2.09)	0.92 (0.17)
Wollny et al. [22]	0.993 (0.009)	4.071 (1.71)	1.227 (0.167)	0.736 (0.096)	543 (5.89)	-
Janssens et al. [27]	0.995 (0.007)	6.486 (2.71)	1.016 (0.194)	0.673 (0.071)	13800 (1405.0)	-

The standard deviations are in parentheses. Run time is measured and averaged over 245 2D cardiac sequences.

not all the cases were correctly registered, as shown in the Supplemental Material (Example 4), these cases were thus not considered for the following evaluations.

Quantitative evaluation of the methods was performed using the test dataset (n=23 patients). A total of 245 2D sequences were registered and each of them was evaluated based on the spatial alignment (DSC, HD) and image quality measures (MI, SSIM). In addition, the run time for sequence registration for each method was measured. The results are summarized in Table IV. In addition, Fig. 5 visualizes the results (box plots) of each registration method. According to the spatial alignment metrics DSC and HD (Fig 5a-b), all methods showed similar performance in spatial alignment, but the best results for the LCV model (DSC$_{LCV}$ = 0.9980 vs. DSC$_{Unregistered}$ = 0.9734, HD$_{LCV}$ = 3.370 vs. HD$_{Unregistered}$ = 7.863).

To quantitatively evaluate the effect of the deformation on the quality of the warped image, the MI and SSIM between the moving and the warped image were estimated for each of the sequences. Here, the LCV and LC models outperform the comparison methods LNC [22], [28], and [27] (Fig 5c-d). In terms of computational effort, LCV, LC and LNC perform registration with an average runtime of 10.78 seconds for a complete 2D image sequence on the CPU, while the iterative methods [22] and [27] have significantly longer average runtimes of 543 seconds (50 times longer) and 13800 seconds (1280 times longer), respectively.

Finally, to analyze to what extent the absolute CT (HU) values in the image are affected by the registration operations, we determined the histograms of the moving and warped image to identify possible changes in the absolute CT values. We then quantified the similarity between the histograms obtained using measures such as cross-correlation (CCF), Hellinger distance (HeD) and Chi-squared distance (ChiS). Fig 6 shows the results of the histogram analysis performed on the registered images of Fig 3. It can be seen that the absolute CT values registered by LNC [28], Wollny et al. [22],
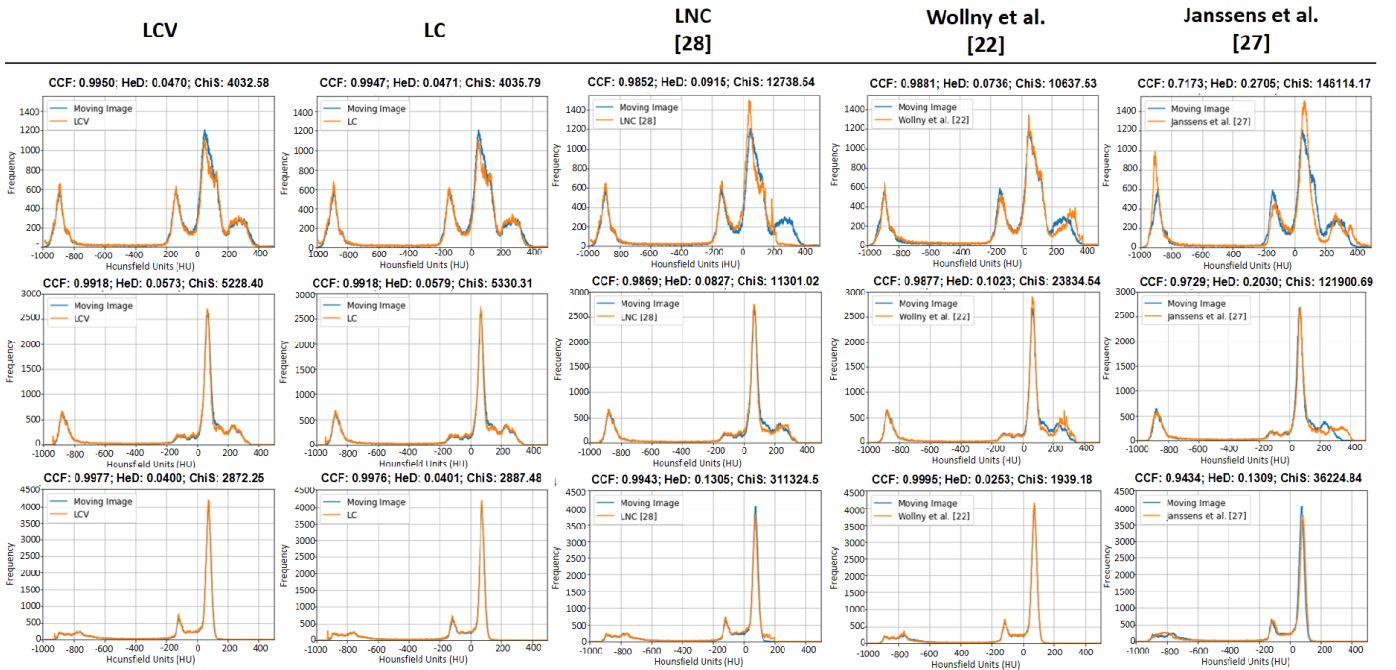
Fig. 6. CT (HU) values analysis in the images. Histogram of CT values of moving and warped image for the examples presented in Fig 3. (CCF) Cross correlation coefficient, (HeD) Hellinger distance and (ChiS) Chi-squared distance between histograms of the moving and warped image.

TABLE V
CT VALUES ANALYSIS

Methods	CCF	HeD	ChiS
LCV	**0.992**	**0.058**	**6983**
LC	0.991	0.060	7734
LNC	0.986	0.101	150476
Wollny et al. [22]	0.987	0.079	12768
Janssens et al. [27]	0.955	0.432	119259

Similarity metrics were obtained from the CT histograms and averaged over 5638 images of the 245 2D cardiac sequences.

TABLE VI
LV REGISTRATION RESULTS

Methods	DSC Myocardium	HD Myocardium	DSC Cavity	HD Cavity
Unregistered	0.85 (0.05)	5.62 (0.13)	0.82 (0.09)	5.00 (0.12)
LCV	**0.90 (0.01)**	**5.17 (0.04)**	**0.87 (0.02)**	4.46 (0.28)
LC	0.89 (0.01)	5.30 (0.05)	0.83 (0.02)	4.99 (0.30)
LNC [28]	0.89 (0.04)	5.41 (0.18)	0.84 (0.02)	5.09 (0.51)
Wollny et al. [22]	0.87 (0.04)	5.47 (0.17)	0.84 (0.04)	4.83 (0.47)
Janssens et al. [27]	0.89 (0.06)	5.19 (0.23)	0.84 (0.03)	**4.31 (0.61)**

Spatial alignment metrics (DSC, HD) averaged over 95 frames from n = 23 subjects. The standard deviations are in parentheses.

and Janssens et al. [27] change, in particular compared to the original images. This is also reflected in the similarity measures CCF, HeD and ChiS. The summary of the results obtained for all the sequences of the test dataset is presented in Table V. Note that the mean CCF was estimated using the Fisher's z-transform. From Table V, we can see that images registered with LCV and LC have the least alterations in the CT values and thus remain almost unchanged, which is essential for quantitative myocardial perfusion measurements.

E. Evaluation of the Alignment of the LV Cavity and Myocardium

Further evaluation of the methods was performed using the manual segmentations of the LV cavity and myocardium carried out by a radiologist with the objective of providing additional information on the accuracy of the alignment of the LV cavity and myocardium. For this purpose, a total of 92 frames from the 23 patients test dataset were used to determine the flow fields between the fixed and the selected warped frames. We then applied the deformations to the masks

of the LV cavity and myocardium. Finally, we calculated the DSC and HD between the warped masks and ground truth segmentations of the fixed frame. The results are shown in Table VI.

It can be seen that in addition to the best image quality (see MI, SSMI in Table IV and CCF, HeD and ChiS in Table V), the LCV model also has the best spatial alignment accuracy for the relevant ROIs of the LV compared to the benchmark methods. These results suggest that the incorporation of auxiliary information of the LV and RV can further promote the alignment of the cardiac structures such as the LV myocardium. Note that all methods achieved similar performance as they obtained better results for LV myocardium than for LV cavity registration. For the LCV and LC models, these results can be caused due to the penalization of changes in high contrast regions, i.e., the LV cavity, and therefore only minimal deformations are expected. However, they still perform well despite the minimal warping.
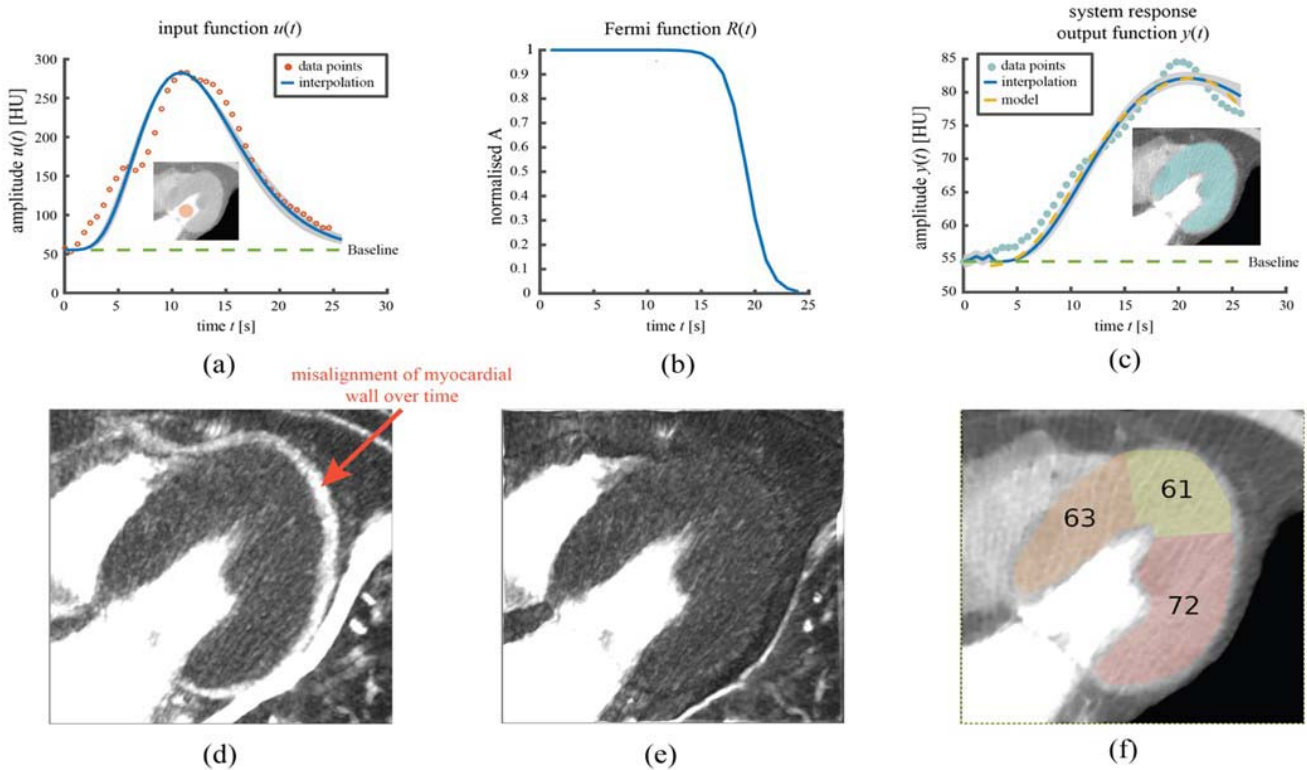
Fig. 7. (a) CT values time curves obtained from a ROI in the LV cavity (input function u(t)) (b) Fermi-function for deconvolution (c) measured and estimated CT values-time curves in the segmented LV myocardial wall (output function y(t)). (d) difference in HU values over time for the unregistered images (misalignment of the myocardium over the sequence). (e) difference in HU values over time for the registered images (aligned LV myocardium after LCV registration). (f) calculated regional myocardial perfusion in ml/100g/min for the apical (yellow), septal (orange) and lateral wall region (red).

F. Clinical Example of Myocardial Perfusion Measurement

To demonstrate the potential clinical applicability of the LCV model, a clinical example of myocardial perfusion measurement is presented in detail. Using a patient of our study cohort with known aortic valve insufficiency, quantitative LV myocardial perfusion was calculated based on the contrast enhancement time curves obtained from a ROI placed in the LV cavity (input function $u(t)$) and the segmented myocardial region (output function $y(t)$) using the well-established Fermi deconvolution method [38]. The acquired CT values-time curves were filtered using a low-pass filter with a filter-length of 2 for the input function and 6 for the output function (dotted values in Fig. 7a and 7c). The data points were fit by standard model functions described by the blue curves in Fig. 7a for the LV cavity and Fig. 7c for the LV myocardial wall, respectively. The confidence interval for the curve-fit is plotted in grey. Fig. 7b shows the obtained Fermi function $R(t)$. The result of the deconvolution process is depicted in Fig. 7c, yellow dashed line, showing the output of the model. In this example, global myocardial blood flow was computed to be 66 ml/100g/min (for the entire myocardium, i.e. septum, apical region and lateral wall). Fig. 7d shows the maximal difference in HU values over the original time series (max-min) for each voxel in the images. The bright band surrounding the lateral wall (marked by an arrow) is caused by misalignments of the unregistered myocardium over different points in time. It can

be seen in Fig. 7e that these misalignments of the LV are removed in the registered image sequence. In order to assess regional differences in myocardial perfusion, the segmented myocardium was split into the septum, an apical region and the lateral wall following the suggestions of the AHA [39]. We can observe small differences with a slightly decreased value of 61 ml/100g/min in the apical region and a slightly higher value in the lateral wall area of 72 ml/100g/min. These values and small deviations correspond well to standard mean blood flow (MBF) in the literature [40].

IV. DISCUSSION

In this work, we present a deformable deep learning-based method for the registration of dynamic myocardial perfusion sequences. The method uses a pre-trained ventricle segmentation module and tailored image similarity loss to guide the registration instead of ground-truth fields used in supervised approaches. The main reason for this approach is that there are no ground-truth deformation fields that can be easily generated manually or synthetically. In the case of manual labeling, a tremendous amount of effort is required to determine such dense flow fields. Synthetic data, on the other hand, is more easily generated by simulations, but must be carefully designed to meet the realistic requirement of the data. In our particular case, such simulations have to consider displacements caused by cardiac stressing, respiration, and patient motion, in addition to the changing contrast agent

which also varies by patient. Due to the complexity of the latter, we have proposed a method using a loss function based on an image similarity loss and auxiliary segmentation data from a pre-trained module.

The proposed method introduces a novel loss function (LCV) that accounts for the changing contrast agent concentration in the image sequence and emphasizes the alignment of the RV and LV. The proposed network has a recursive cascade architecture with a segmentation module, and was trained with a 5-fold cross-validation to evaluate its performance and robustness in highly varying cardiac sequences. The optimal number of cascades was determined experimentally and shows that a configuration with 7 cascades gives the best trade-off between spatial alignment and image quality.

The performance of our extended models LCV and LC was evaluated and compared with the original deep learning method LNC [28] and two established iterative, non-deep learning-based methods by Wollny et al. [22] and Janssens et al. [27]. The experimental results shown in Fig. 3 and Fig. 4 emphasize the influence and necessity of a contrast concentration term in the loss function for the deep learning methods. It is obvious that not considering the contrast term as in LNC leads to a degradation in the quality of the warped image. Fig. 3a shows an example where the contrast regions between the fixed and the moving image differ significantly. Here, the results of LNC, [22] and [27] show changes in the contrast of the right chambers. The latter can also be seen in Fig. 4a where the norm of the flow field vectors is higher in the regions with enhanced contrast concentration. As a result, the DVF applies deformations that lead to artifacts in such regions. Similar results can be seen in Fig. 3b and Fig. 4b, where most of the displacements are in the high contrast regions. Interestingly, the DVF applied by [27] leads to an enhancement of the contrast concentration after the registration. In contrast to the comparison methods, LCV and LC revealed equivalent results with the contrast agent regions not being affected. From the previous examples, LCV applies deformations in the outer contours of the LV and RV as well as within the cavities, whereas LC applies the displacements mainly in the outer contours. The latter is due to the fact that the warping of the contrast regions is penalized inside the heart, which limits the plausible deformations of the contours. The results presented in Fig. 3c show that LCV, LC, and the iterative methods have similar performance without the presence of contrast regions in the moving image, registering the images without artifacts. The latter is also demonstrated in Fig. 4c where most of the displacements are present in the outer contours.

Our quantitative results show that the proposed LCV method achieves excellent accuracy in registering image sequences with spatio-temporal variations (Fig. 5, Table IV). Moreover, it also proves consistent image quality after warping (Fig. 6, Table IV). Similar results are achieved by LC, showing that adding the contrast term L_{cont} helps to preserve the image quality after deformation. It should be noted that some image degradation is expected due to the deformations.

The results of the analysis of absolute CT values demonstrate the effects of the methods on the HU values after registration of the cardiac sequences. The histograms in Fig. 6

show the difference between the CT values of the original and the warped images, and this dissimilarity is quantified by the measures CCF, HeD and ChiS. The histograms of the warped images from LNC, [22] and [27] show deviations in the range of upper HU values representing the high-contrast regions of the images. This result is undesirable as it indicates that information relevant for perfusion studies is altered. As outlined in this work, myocardial perfusion studies require good spatial alignment of dynamic cardiac sequences while retaining the original CT values in the registered images required for quantitative estimation of myocardial perfusion (see example patient in Fig. 7). In particular, as shown in Fig. 6 and Table V, the LCV model performs best in registration in terms of spatial alignment and image quality without significantly affecting the absolute CT values.

Further evaluation of the alignment of the LV myocardium and LV cavity after registration reveals that deformations made by LCV are not only applied to the contours of the heart but also on the relevant ROIs of the LV. Table VI shows that LCV achieves the best results compared to the other methods, especially in the registration of the LV myocardium, which is due to the auxiliary information from the *ventricle module* that enhances the alignment of the outer contour. However, comparing the DSC scores with Table IV, we see that the deformation caused by the methods is lower in these ROIs. We found that in cases where the contrast agent is not present (see Fig. 3c and Fig 4c), most of the displacements are applied only on the outer contours of the heart. In such cases, both the LV myocardium and cavity have similar HU values and lack distinguishable landmarks, making it difficult to establish correspondence of these specific areas and thus to determine the required displacements, which may explain the decrease in the DSC scores. LC also shows good performance despite the lack of the ventricle module, suggesting that LC is also a suitable registration method sharing the advantages of LCV, specifically high processing speed, however with slightly reduced image quality and local alignment accuracy.

Finally, the clinical example presented in Section IIIF demonstrated that our proposed LCV method can be used as an effective image preprocessing step in cardiac perfusion studies. The results show that LCV performs registration without affecting relevant information in the image sequence, i.e. HU values, so that consistent time-attenuation curves in CT can be generated for regional myocardial perfusion calculation.

V. CONCLUSION

We present for the first time here, a deformable deep learning-based image registration method for cardiac CT perfusion imaging that has several advantages and strengths. The architecture of the LCV model is based on a configuration with recursive cascades. This structure was extended and optimized by the introduction of novel loss functions that are especially well-suited to take dynamic changes in the contrast agent concentration into account and a ventricular segmentation module that further reduces local tissue displacement.

The LCV model shows best registration accuracy for the relevant ROIs, compared to the original cascade architecture

and two state-of-the-art iterative, non-deep leaning methods. It processes poor image quality in terms of low signal-to-noise ratio and less accurate anatomical landmarks, and overcomes the problem of dynamic changes of contrast agent concentration over time without affecting the HU values throughout the registered image sequence. This is essential for quantitative global and regional myocardial perfusion estimation, as demonstrated by the clinical example. Due to its high processing speed of a few seconds for the registration of an entire CT perfusion sequence compared to minutes and hours of non-deep learning-based methods, this new method shows high potential to be used in clinical routine, in particular due to its rapidity.

In summary, our experimental results confirm the high accuracy and robustness of this method on a representative dataset of patients with coronary artery disease and/or aortic valve insufficiency. From this we can expect that our approach can easily be extended and applied to other dynamic sequence registration problems in medical imaging, while keeping dynamic contrast changes unaffected across a registered image sequence. In our future work, we will extend our method to other dynamic imaging modalities and further investigate the effects of different image similarity loss functions.

ACKNOWLEDGMENT

REFERENCES

[1] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: A review," *Phys. Med. Biol.*, vol. 65, no. 20, p. 48, 2020, doi: 10.1088/1361-6560/ab843e.

[2] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," *Mach. Vis. Appl.*, vol. 31, pp. 1–18, Feb. 2020, doi: 10.1007/s00138-020-01060-x.

[3] T. Mäkelä, T. Katila, and I. E. Magnin, "A review of cardiac image registration methods," *IEEE Trans. Med. Imag.*, vol. 21, no. 9, p. 11, Sep. 2002.

[4] H. Wiputra, W. X. Chan, Y. Y. Foo, S. Ho, and C. H. Yap, "Cardiac motion estimation from medical images: A regularisation framework applied on pairwise image registration displacement fields," *Sci. Rep.*, vol. 10, no. 1, p. 14, Dec. 2020.

[5] K. A. L. Hernandez, T. Rienmüller, D. Baumgartner, and C. Baumgartner, "Deep learning in spatiotemporal cardiac imaging: A review of methodologies and clinical usability," *Comput. Biol. Med.*, vol. 130, Mar. 2021, Art. no. 104200, doi: 10.1016/j.compbiomed.2020.104200.

[6] D. Caruso et al., "Dynamic CT myocardial perfusion imaging," *Eur. J. Radiol.*, vol. 85, no. 10, pp. 1893–1899, Oct. 2016, doi: 10.1016/j.ejrad.2016.07.017.

[7] M. Weininger et al., "Adenosine-stress dynamic real-time myocardial perfusion CT and adenosine-stress first-pass dual-energy myocardial perfusion CT for the assessment of acute chest pain: Initial results," *Eur. J. Radiol.*, vol. 81, no. 12, pp. 3703–3710, Dec. 2012, doi: 10.1016/j.ejrad.2010.11.022.

[8] Y. Tanabe et al., "Peak enhancement ratio of myocardium to aorta for identification of myocardial ischemia using dynamic myocardial computed tomography perfusion imaging," *J. Cardiol.*, vol. 70, no. 6, pp. 565–570, Dec. 2017, doi: 10.1016/j.jjcc.2017.04.002.

[9] B. Avants, N. J. Tustison, and G. Song, "Advanced normalization tools: V1.0," *Insight J.*, p. 29, Jul. 2009. [Online]. Available: http://hdl.handle.net/10380/3113, doi: 10.54294/uvnhin.

[10] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, Mar. 2009, doi: 10.1016/j.neuroimage.2008.10.040.

[11] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 196–205, Jan. 2010, doi: 10.1109/TMI.2009.2035616.

[12] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[13] X. Cao et al., "Deformable image registration based on similarity-steered CNN regression," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 10433, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, Eds. Cham, Switzerland: Springer, 2017, pp. 300–308, doi: 10.1007/978-3-319-66182-7_35.

[14] J. Lv, M. Yang, J. Zhang, and X. Wang, "Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: A feasibility study," *Brit. J. Radiol.*, vol. 91, Jan. 2018, Art. no. 20170788, doi: 10.1259/bjr.20170788.

[15] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," 2015, *arXiv:1506.02025*.

[16] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, vol. 10553, M. J. Cardoso et al., Eds. Cham, Switzerland: Springer, 2017, pp. 204–212, doi: 10.1007/978-3-319-67558-9_24.

[17] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Med. Image. Anal.*, vol. 52, pp. 128–143, Feb. 2019, doi: 10.1016/j.media.2018.11.010.

[18] A. Sheikhjafari, M. Noga, K. Punithakumar, and N. Ray, "Unsupervised deformable image registration with fully connected generative neural network," in *Proc. Int. Conf. Med. Imag. With Deep Learn.*, 2018. [Online]. Available: https://openreview.net/pdf?id=HkmkmW2jM

[19] T. Fechter and D. Baltas, "One shot learning for deformable medical image registration and periodic motion tracking," 2019, *arXiv:1907.04641*.

[20] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Washington, DC, USA, Apr. 2018, pp. 1449–1453, doi: 10.1109/ISBI.2018.8363845.

[21] J. Milles, R. J. V. D. Geest, M. Jerosch-Herold, J. H. C. Reiber, and B. P. F. Lelieveldt, "Fully automated motion correction in first-pass myocardial perfusion MR image sequences," *IEEE Trans. Med. Imag.*, vol. 27, no. 11, pp. 1611–1621, Nov. 2008, doi: 10.1109/TMI.2008.928918.

[22] G. Wollny, M. J. Ledesma-Carbayo, P. Kellman, and A. Santos, "Exploiting quasiperiodicity in motion correction of free-breathing myocardial perfusion MRI," *IEEE Trans. Med. Imag.*, vol. 29, no. 8, pp. 1516–1527, Aug. 2010, doi: 10.1109/TMI.2010.2049270.

[23] C. M. Scannell, A. D. M. Villa, J. Lee, M. Breeuwer, and A. Chiribiri, "Robust non-rigid motion compensation of free-breathing myocardial perfusion MRI data," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1812–1820, Aug. 2019, doi: 10.1109/TMI.2019.2897044.

[24] A. A. Isola et al., "Image registration and analysis for quantitative myocardial perfusion: Application to dynamic circular cardiac CT," *Med. Phys.*, vol. 56, no. 18, pp. 5925–5947, Sep. 2011, doi: 10.1088/0031-9155/56/18/010.

[25] Q. Tang, J. Cammin, S. Srivastava, and K. Taguchi, "A fully four-dimensional, iterative motion estimation and compensation method for cardiac CT: Motion estimation and compensation method for cardiac CT," *Med. Phys.*, vol. 39, no. 7, pp. 4291–4305, Jun. 2012, doi: 10.1118/1.4725754.

[26] J. Liu et al., "Motion compensation combining with local low rank regularization for low dose dynamic CT myocardial perfusion reconstruction," *Phys. Med. Biol.*, vol. 66, no. 14, Jul. 2021, Art. no. 145007, doi: 10.1088/1361-6560/ac0f2f.

[27] G. Janssens, L. Jacques, J. O. de Xivry, X. Geets, and B. Macq, "Diffeomorphic registration of images with variable contrast enhancement," *Int. J. Biomed. Imag.*, vol. 2011, pp. 1–16, Apr. 2011, doi: 10.1155/2011/891585.

[28] S. Zhao, Y. Dong, E. Chang, and Y. Xu, "Recursive cascaded networks for unsupervised medical image registration," in *Proc. IEEE CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10599–10609, doi: 10.1109/ICCV.2019.01070.

[29] K. A. Lara, I. A. Juárez, M. A. Perez, T. Rienmüller, and C. Baumgartner, "Deep learning based image registration in dynamic cardiac CT using a recursive cascade network approach," in *Proc. Annu. Conf. Austrian Soc. Biomed. Eng.*, 2021, pp. 67–70, doi: 10.3217/978-3-85125-826-4-18.

[30] 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 Informat.*, vol. 24, no. 5, pp. 1394–1404, May 2020, doi: 10.1109/JBHI.2019.2951024.

[31] 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. Imag.*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019, doi: 10.1109/TMI.2019.2897538.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[34] 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.

[35] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer, "Image similarity using mutual information of regions," in *Proc. Eur. Conf. Comput. Vis.*, vol. 3023, T. Pajdla and J. Matas, Eds. Berlin, Germany: Springer, 2004, pp. 596–607, doi: 10.1007/978-3-540-24672-5_47.

[36] G. Wollny, P. Kellman, M.-J. Ledesma-Carbayo, M. M. Skinner, J.-J. Hublin, and T. Hierl, "MIA—A free and open source software for gray scale medical image analysis," *Source Code Biol. Med.*, vol. 8, no. 1, p. 20, Dec. 2013, doi: 10.1186/1751-0473-8-20.

[37] J. Plumat, M. Andersson, G. Janssens, J. O. de Xivry, H. Knutsson, and B. Macq, "Image registration using the Morphon algorithm: An ITK implementation," *Insight J.*, p. 13, Mar. 2009. [Online]. Available: http://hdl.handle.net/1926/1527, doi: 10.54294/kmysv2.

[38] G. Morton et al., "Quantification of absolute myocardial perfusion in patients with coronary artery disease: Comparison between cardiovascular magnetic resonance and positron emission tomography," *J. Amer. College Cardiol.*, vol. 60, no. 16, pp. 1546–1555, Oct. 2012, doi: 10.1016/j.jacc.2012.05.052.

[39] M. D. Cerqueira et al., "Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. A statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the American heart association," *Circulation*, vol. 105, no. 4, pp. 539–542, Jan. 2002, doi: 10.1161/hc0402.102975.

[40] G. Degtiarova et al., "Low septal to lateral wall ^{18}F-FDG ratio is highly associated with mechanical dyssynchrony in non-ischemic CRT candidates," *EJNMMI Res.*, vol. 9, no. 1, p. 105, Dec. 2019, doi: 10.1186/s13550-019-0575-9.

[41] S. A. Kajander et al., "Clinical value of absolute quantification of myocardial perfusion with ^{15}O-water in coronary artery disease," *Circulat., Cardiovascular Imag. Logo*, vol. 4, no. 6, pp. 678–684, Nov. 2011, doi: 10.1161/CIRCIMAGING.110.960732.