

Automating in vivo cardiac diffusion tensor postprocessing with deep learning-based segmentation

Pedro F. Ferreira^{1,2}   | Raquel R. Martin³ | Andrew D. Scott^{1,2}  | Zohya Khalique^{1,2} | Guang Yang^{1,2}  | Sonia Nelles-Vallespin^{1,2}  | Dudley J. Pennell^{1,2}  | David N. Firmin^{1,2} 

¹Cardiovascular Magnetic Resonance Unit, Royal Brompton Hospital, London, United Kingdom

²National Heart and Lung Institute, Imperial College, London, United Kingdom

³Department of Bioengineering, Imperial College, London, United Kingdom

Correspondence

Pedro F. Ferreira, Cardiovascular Magnetic Resonance Unit, Royal Brompton Hospital, London, SW3 6NP, United Kingdom.
Email: p.ferreira@rbht.nhs.uk

Funding information

British Heart Foundation, Grant/Award Number: RG/19/1/34160

Purpose: In this work we develop and validate a fully automated postprocessing framework for in vivo diffusion tensor cardiac magnetic resonance (DT-CMR) data powered by deep learning.

Methods: A U-Net based convolutional neural network was developed and trained to segment the heart in short-axis DT-CMR images. This was used as the basis to automate and enhance several stages of the DT-CMR tensor calculation workflow, including image registration and removal of data corrupted with artifacts, and to segment the left ventricle. Previously collected and analyzed scans (348 healthy scans and 144 cardiomyopathy patient scans) were used to train and validate the U-Net. All data were acquired at 3 T with a STEAM-EPI sequence. The DT-CMR postprocessing and U-Net training/testing were performed with *MATLAB* and *Python TensorFlow*, respectively.

Results: The U-Net achieved a median Dice coefficient of 0.93 [0.92, 0.94] for the segmentation of the left-ventricular myocardial region. The image registration of diffusion images improved with the U-Net segmentation ($P < .0001$), and the identification of corrupted images achieved an F1 score of 0.70 when compared with an experienced user. Finally, the resulting tensor measures showed good agreement between an experienced user and the fully automated method.

Conclusion: The trained U-Net successfully automated the DT-CMR postprocessing, supporting real-time results and reducing human workload. The automatic segmentation of the heart improved image registration, resulting in improvements of the calculated DT parameters.

KEYWORDS

cardiac, deep learning, diffusion tensor imaging, image processing, machine learning

1 | INTRODUCTION

Diffusion tensor cardiac magnetic resonance (DT-CMR) is an emerging technique that provides information on myocardial microstructure. Many diffusion parameters can be extracted from the tensors, including tensor orientation and rotational invariant measures. For tensor orientation, it is common to report the primary and either secondary or tertiary eigenvector orientations that have been shown to relate to the orientation of the local cardiomyocytes and their sheetlet structure, respectively.¹ These angles are commonly known as helix angle (HA) and sheetlet angle or secondary eigenvector angle (E2A) respectively. In terms of rotational invariant measures, it is common to report fractional anisotropy (FA), a measure of diffusion organization, and mean diffusivity (MD), which measures the magnitude of diffusion. Both tensor orientation and rotational invariant measures have been shown to change in cardiac disease.²⁻⁶

The computation of diffusion tensors from in vivo data is particularly challenging in the heart. Diffusion imaging yields low SNR images, and it is common to compensate with long acquisitions spanning multiple breath-holds. For most data sets, image registration is needed to correct for residual image shifts due to inconsistency with intrabreath and interbreath holding. This registration is challenging due to the different diffusion weightings and encoding directions between images, which translate into differences in signal intensities and contrast, not only in the myocardium, but also the neighboring structures of the chest wall, liver, and stomach (Figure 1A). Additionally, these peripheral structures do not move rigidly with the heart during the respiratory cycle, further complicating registration.

In addition to the challenges of registration, a typical DT-CMR study of one midventricular slice scanned at two stages of the cardiac cycle will typically contain more than 160 images. Approximately 14% of the images are corrupted by signal loss due to cardiac and respiratory motion and are usually manually excluded based on visual assessment (Figure 1B).⁷

In vivo DT-CMR postprocessing is typically done retrospectively. Multiple steps require manual input, including removing frames corrupted by motion artifacts, and thresholding and segmentation of the left ventricular (LV) myocardial borders. Thresholding removes unwanted structures and background voxels, whereas segmentation not only allows quantification of diffusion parameters in the LV myocardium, but it is also used to define the local cardiac coordinate system consisting of the longitudinal, circumferential, and radial orthogonal directions (Figure 1C).^{8,9} Visualization of tensor orientation measures in relation to the local cardiac coordinates aids in the interpretation of the tensor results.

In this work we develop and validate a fully automated postprocessing framework of in vivo DT-CMR data powered by a convolutional neural network (CNN) trained to perform

semantic segmentation. The aim is to improve robustness, reduce human postprocessing workload, and ultimately to produce on-the-fly DT-CMR results at the scanner, providing real-time visual feedback on scan quality.

2 | METHODS

In vivo DT-CMR postprocessing requires several steps before tensor calculation. In summary, the data need to go through the following steps: (1) removal of data corrupted by motion-induced signal loss, (2) image registration of the remaining diffusion images, (3) thresholding and segmentation of the LV myocardial borders, and (4) voxel-wise calculation of the diffusion tensors (Figure 2A).

A CNN for semantic segmentation was used as the basis to automate and enhance some of the stages of the DT-CMR tensor calculation workflow (Figure 2B). A U-Net-based CNN was designed with five encoder/decoder levels. A diagram of its architecture is depicted in Figure 3A. A U-Net was chosen, as it is known to produce good results for biomedical data without requiring large training data sets.¹⁰ The specific network design was configured ad hoc to yield the highest Dice scores.

The U-Net was trained to segment multiple classes, including LV myocardium; right ventricular myocardium and papillary muscle; the two interventricular insertion points; and the remaining pixels as the background (Figure 3B). For the training, the position of the ventricular insertion points, previously designated by the user, were converted to a circular region with a 6-pixel radius but excluding any pixels inside the LV myocardium. The predicted insertion point regions by the U-Net are therefore similar shapes. The centroid of these regions was considered as the position of the predicted insertion points.

To train the U-Net, we used previously collected DT-CMR scans acquired with a STEAM-EPI sequence. The data included a total of 492 scans (348 healthy scans and 144 scans from patients with either hypertrophic cardiomyopathy [HCM], dilated cardiomyopathy, or cardiac amyloidosis). This study was approved by the National Research Ethics Service. Data were acquired during either the diastolic or the systolic pause. All data had been previously examined for artifacts, and the LV myocardium segmented by 1 of 2 experienced clinicians and subsequently checked/edited by an experienced physicist. All available data were primarily focused on the LV myocardium, although, when thresholding the heart during the analysis, the right-ventricular (RV) muscle was also segmented. The imaging protocol has been described previously,^{11,12} and in summary all imaging was performed using a Skyra 3T MRI scanner (Siemens AG, Erlangen, Germany) with a STEAM-EPI sequence with zonal excitation and fat saturation, TR = 2RR intervals,

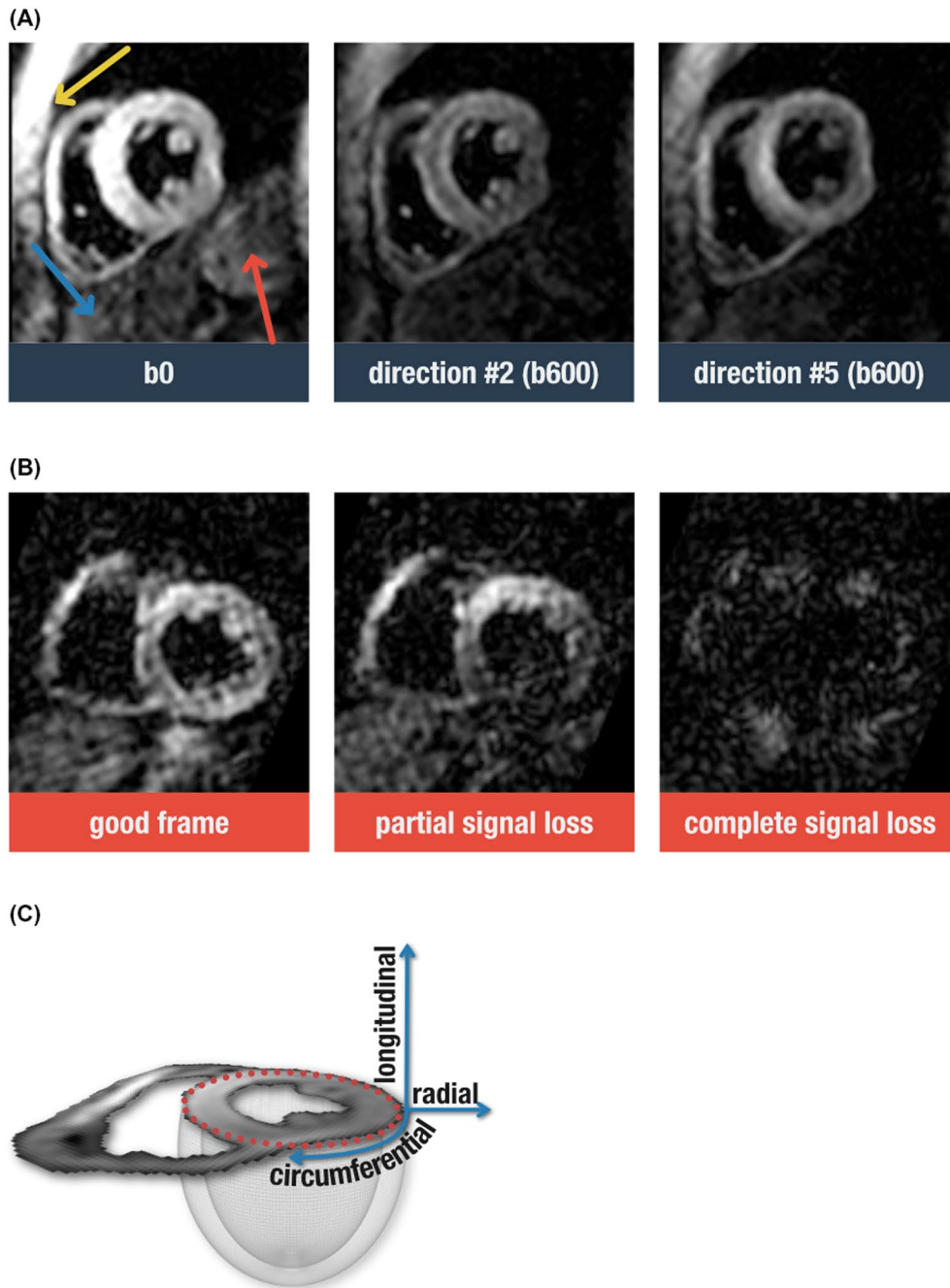


FIGURE 1 A, Diffusion images showing the heart and neighboring structures: the chest wall (yellow arrow), the liver (blue arrow), and the stomach (red arrow) at different diffusion weightings (b-value units, $s\ mm^{-2}$) and encoding directions. B, Three diffusion frames with varying degrees of motion-induced signal-loss artifacts. C, Local cardiac orthogonal coordinate system diagram

TE = 23 ms, SENSE or GRAPPA $\times 2$, echo train duration = 13 ms, at a spatial resolution of $2.8 \times 2.8 \times 8.0\ mm^3$ with diffusion weightings of 150 and $600\ s/mm^2$ for six diffusion-encoding directions.

Each DT-CMR scan consists of multiple images with different diffusion weightings acquired in a short-axis slice at one cardiac phase. A typical clinical scan will contain approximately 80 images per slice and cardiac phase. The network input images were the mean image of each scan after spatially registering the entire diffusion series. This produces

magnitude input images with a much higher SNR than the individual diffusion images and are named here as *scan-mean images*.

Once successfully trained, the U-Net was integrated within the in vivo DT-CMR postprocessing at different stages of the workflow. It was used primarily to enhance the image registration, to identify frames corrupted with motion-induced signal loss, and to segment the LV myocardial region for quantification of diffusion tensor parameters (Figure 2B). These steps were compared with the current established

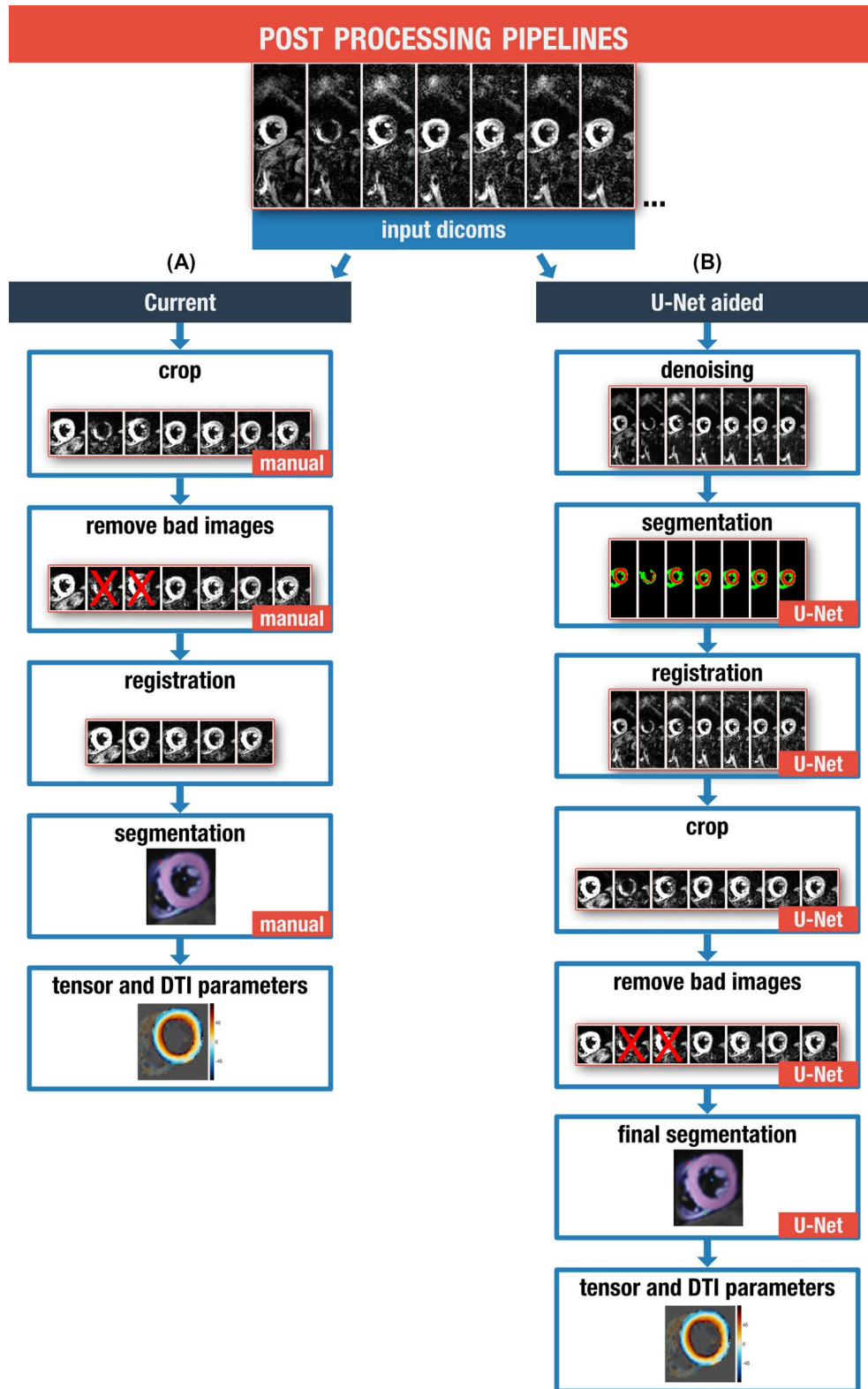


FIGURE 2 Workflow diagrams of the two postprocessing techniques. A, Current workflow in which some steps require manual input (manual label) from the user. B, New workflow in which the manual steps are automated and others enhanced by using the segmentation information given by the trained U-Net (U-Net label)

method of postprocessing the DT-CMR data with manual input from an experienced user, which was considered as the ground truth (Figure 2A).

Finally, this work concluded by comparing the DT-CMR results of the fully automated workflow against the manual user workflow for the independent test data set as described

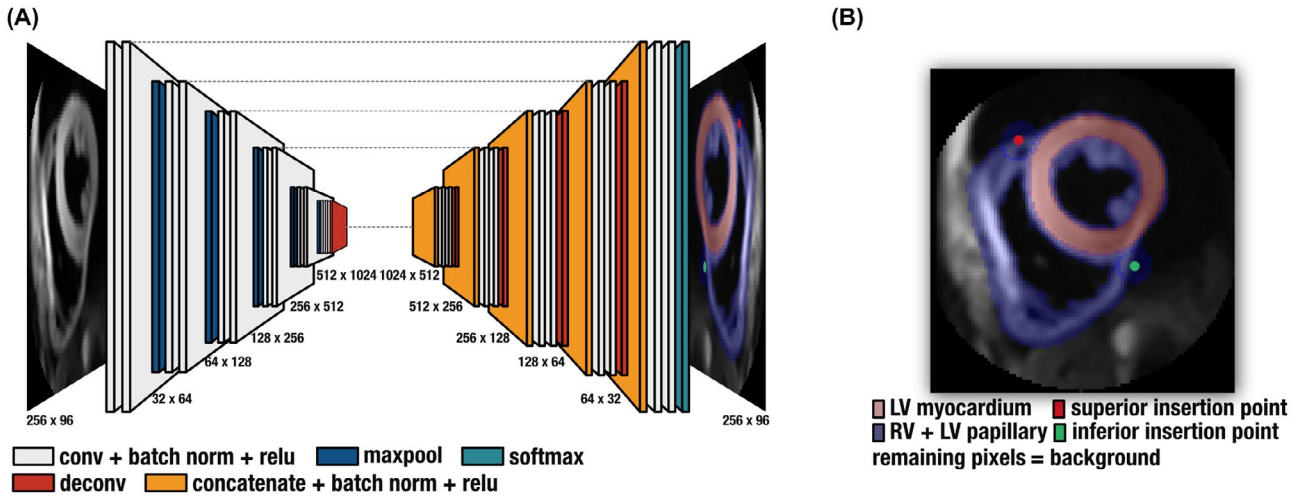


FIGURE 3 A, U-Net architecture used. B, U-Net classes for one example. Abbreviations: LV, left ventricular; RV, right ventricular

subsequently. This includes global median values and a voxel-wise comparison of FA, MD, E2A, and HA in the left ventricle.

The trained U-Net can be downloaded from https://github.com/Pedro-Filipe/DT_CMV_short_axis_conv_net.

2.1 | Left ventricular myocardial segmentation

The first step of this work was to train the U-Net to segment/identify the five different regions/locations shown in Figure 3B. The scans were divided into three groups: training data with 393 (80%) DT-CMR scans, validation data of 29 (6%) scans used to assess the network weights at the end of each training epoch, and finally the test data set with 70 (14%) scans to test the CNN final weights. The data distribution discretized by disease is shown in Supporting Information Figure S1. Data augmentation was used to improve training with random uniform distributions: rigid translation (range: 10% of FOV) and rotation (range: ± 0.6 radians). These ranges were chosen to constrain these operations to realistic levels. A generalized Dice loss function¹³ was used to mitigate class imbalance and assess the proportion of specific agreement between the experienced user and the U-Net segmented masks. Batch normalization layers were added to the U-Net, as they are known to reduce overfitting, and when tested in our data, they were found to improve the segmentation Dice scores of the test data. For training we used the Adam optimizer¹⁴ (learning rate: 0.001, epsilon $1E^{-07}$), a batch size of four images, and 200 epochs. The hyperparameters were tuned until we could not improve the Dice scores on the validation data and with sufficient epochs for the loss function to reach a plateau.

The DT-CMR postprocessing was carried out using *MATLAB* (MathWorks, Natick, MA), and the U-Net training/testing was performed using *Python TensorFlow* (Google,

Mountain View, CA). A Workstation with Ubuntu 18.04, two Intel Xeon 8 core CPUs, 12 GB of RAM, and an NVIDIA Quadro P6000 GPU was used for the GPU-based training.

2.2 | Registration of diffusion-weighted images

The U-Net was originally trained with the scan-mean images. To enhance registration, we attempted to use the same U-Net to segment the heart in each individual diffusion image. The segmentation masks derived from the U-Net were used to remove all peripheral structures outside the heart (ie, everything outside the U-Net-defined LV + RV + papillary region). The segmentation information was also used to increase the image intensity toward the LV myocardial region with a Gaussian filter (SD = 40% of myocardial diameter), to weight favorably the registration to the LV at a cost to the right ventricle and to avoid sharp edges when masking. For the segmentation to work reliably, the diffusion images were strongly denoised with a nonlocal means algorithm and magnitude scaled before applying the U-Net.¹⁵ A degree of filtering of 8 times the estimated noise was used in the denoising. This level of denoising was empirically found to smooth the image without adverse blurring. The degree of filtering controls the rate of decay of the weights in the nonlocal means algorithm,¹⁵ while the noise variance estimation is described by Immerkaer et al.¹⁶ The denoised diffusion frames were used only to aid segmentation and classification as described in the next section, but were not used for any diffusion tensor calculations. A multiresolution subpixel rigid image registration (translation only) by cross-correlation of the magnitude signal was subsequently performed.¹⁷ This registration workflow, named here as *Reg U-Net*, was compared to the image registration workflow currently used, in which a user-defined square region around the heart is cropped before using the

same rigid registration algorithm. This workflow was defined in this work as *Reg*. A flowchart of these methods is shown in Supporting Information Figure S2 for clarity. The image in the diffusion series with the highest mean intensity was used as the reference frame for registration.

Differences in the two image registration algorithms (*Reg* vs *Reg U-Net*) were quantified using the registered LV myocardial masks and by calculating the median Dice coefficient of all the diffusion images in the scan in comparison to the reference frame.

2.3 | Identification of data corrupted by motion artifacts

The next step in using the U-Net mask information in the processing workflow was to identify the images corrupted with motion artifacts. After image registration we recalculated the scan-mean image, which was used here as the reference image for identifying motion-corrupted images. Two different metrics were calculated for each diffusion image: The Pearson coefficient of the heart region between the scan-mean image and each registered diffusion image measures the image similarity, and the Dice coefficient of the LV myocardium masks measures the image similarity of the masks. If an image contains significant signal loss in the heart, then these two metrics are likely to yield values considerably lower than 1. An outlier algorithm was then applied to the mean of the two metrics. The outlier algorithm consisted of finding images in which the mean score is more than three scaled median absolute deviations below the median. These images are excluded and not used for tensor calculation. An example of this workflow is shown in Supporting Information Figure S3. The Pearson correlation coefficient could be used to identify the corrupted frames without the need to mask the images, but this method would be sensitive to contrast changes within the heart due to the different diffusion weightings and would be affected by peripheral organs. By applying the mask obtained from the U-Net to all diffusion images, we can remove the other organs and combine the Pearson coefficient with the Dice score of the LV masks to increase the robustness of identifying images with artifacts automatically.

The efficiency of the classification was measured with an F1 score. The F1 score is a measure of accuracy for a binary classification. It accounts for both the precision P and the recall R and is defined as the harmonic mean of these two measures:

$$P = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$R = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$F1 = 2 \times \frac{P \times R}{P + R}$$

2.4 | Diffusion tensor results

Finally, the entire fully automated artificial intelligence (AI) that aided in vivo DT-CMR postprocessing workflow was compared with the currently manual processing by an experienced user on the test data set. After registration, removal of corrupted frames and tensor calculation, the U-Net was finally used for the task it was originally trained: to automatically segment the LV in order to quantify the DT-CMR parameters in the left ventricle. The LV epicardial border was also used to calculate the local heart orthogonal coordinate system for each voxel in the LV, and to obtain maps of tensor orientation in relation to the local cardiac coordinates, providing the HA and E2A maps.

The agreement in tensor parameters between the experienced user and the fully automated U-Net method was analyzed with Bland-Altman plots and by calculating voxel-wise mean absolute errors for the overlapping region of the myocardial masks between the two methods.

Data were tested for normality with a Kolmogorov-Smirnov test, and it was found throughout to be nonnormal; therefore, nonparametric statistics were used with a Wilcoxon signed rank test to quantify statistical difference between the user and the U-Net. Measures are quoted as median [interquartile range].

3 | RESULTS

3.1 | Left ventricular myocardial segmentation

Training of the U-net plateaued on the validation data well within 200 epochs, and took less than an hour to train (Figure 4A shows the training progress). The network weights were updated for the last time at epoch 84; after that the segmentation of the validation data did not improve, and it is likely that the remaining training was overfitting.

The U-Net achieved an overall median Dice coefficient of 0.93 [0.92, 0.94] in the test data set for the segmentation of the LV myocardial region when compared with an experienced user (Figure 4B). The Dice scores have a similar range when divided by cohort: healthy 0.93 [0.92, 0.95], dilated cardiomyopathy 0.93 [0.92, 0.94], HCM 0.92 [0.91, 0.93], and Amyloid 0.94 [0.93, 0.94]. When visually assessing the individual segmentations, the U-Net performance was deemed appropriate with the exception of two cases as shown in Figure 5B,F. The HCM case (Figure 5F) was removed from the remaining analysis, as the segmentation was inadequate. All remaining 69 test cases were used for further analysis.

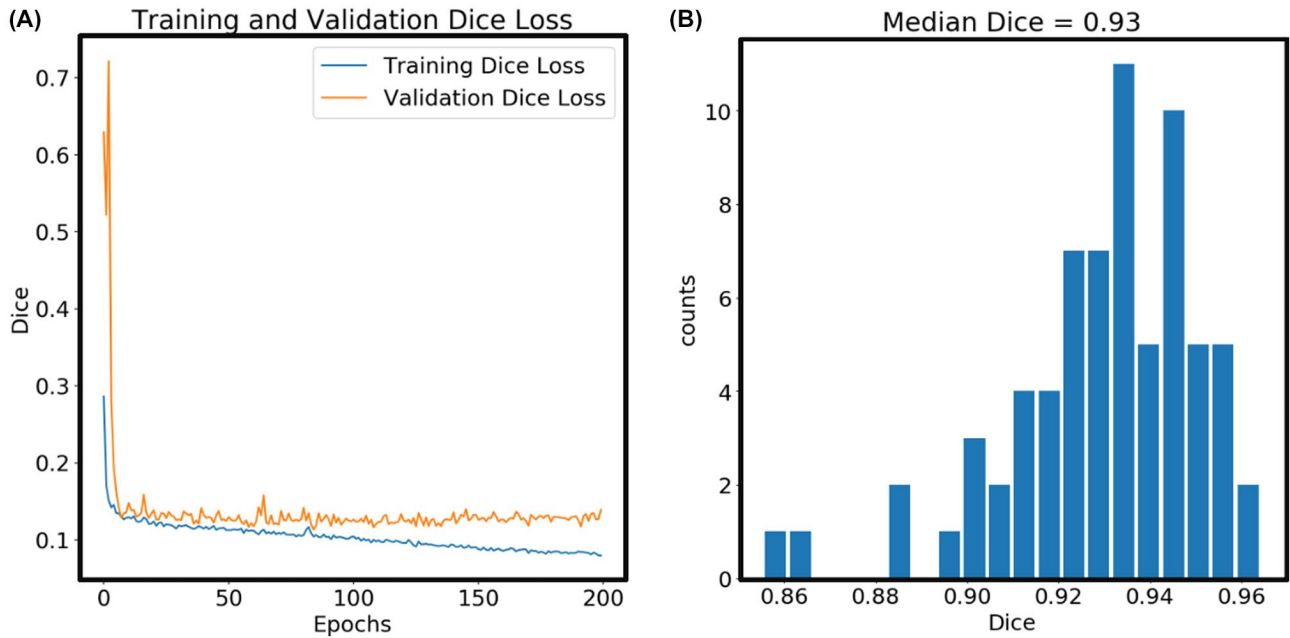


FIGURE 4 A, Loss function evolution during training. B, Histogram of the Dice coefficients between the trained network and an experienced user for the LV myocardial segmentation in the test data

Measuring the axis from the LV center to the insertion points, a median angular difference of 10° [6, 14] was measured for the insertion point positions between the experienced user and the U-Net. Like the Dice scores, we measured a similar range of angular differences among the cohorts: healthy 10 [6, 14]; dilated cardiomyopathy 10 [5, 16]; HCM 11 [5, 16]; and Amyloid 8 [6, 11].

Please see Supporting Information Video S1 for a comparison of the segmentation masks between the user and the U-Net of all 70 test cases.

3.2 | Registration of diffusion-weighted images

Dice scores were used to test whether U-Net was capable of segmenting the denoised individual diffusion images as effectively as the scan-mean images. Overall, a median Dice score of 0.94 [0.92, 0.95] was measured for the segmentation of the denoised individual diffusion images when compared with the corresponding scan-mean image segmentations (after registration). The trained U-Net was therefore capable of segmenting the denoised individual diffusion images.

We also compared the Dice scores after the registration algorithm between the denoised individual diffusion images and the registration of the reference frame. The median Dice coefficient for Reg and Reg U-Net were 0.91 [0.89, 0.92] and 0.93 [0.90, 0.94], respectively. A better registration was therefore measured for the Reg U-Net method, as demonstrated by a significant improvement in the Dice scores when compared with the existing Reg method ($P < .0001$) (Supporting

Information Figure S4). For most cases, the improvement in the Dice score was small (median improvement 0.012 [0.005, 0.027]), but there were a few cases in which the Reg U-Net method visibly improved registration. This was due to the removal of bright bordering structures around the heart, before the registration algorithm, using the U-Net-derived masks. One of these cases is shown in Figure 6 with a Dice improvement of 0.06.

3.3 | Identification of data corrupted by motion artifacts

The next processing step was to use the U-Net to identify images corrupted with significant signal loss artifacts. The confusion matrix comparing the experienced user (ground truth) versus the U-Net method is shown in Figure 7. As expected, the data are imbalanced with 94% of the data considered to be good by both user and U-Net. An F1 score of 0.70 was measured for the U-Net. As anticipated, the mismatches between the user and U-Net were the borderline cases, with partial signal loss only.

3.4 | Diffusion tensor results comparison: user versus machine

Finally, we compared the resulting LV DTI parameters between the two postprocessing workflows. The median DTI parameters are given in Table 1, separated by cardiac phase and cohort. When comparing the overall results between user and U-Net, there was a small but strongly significant drop

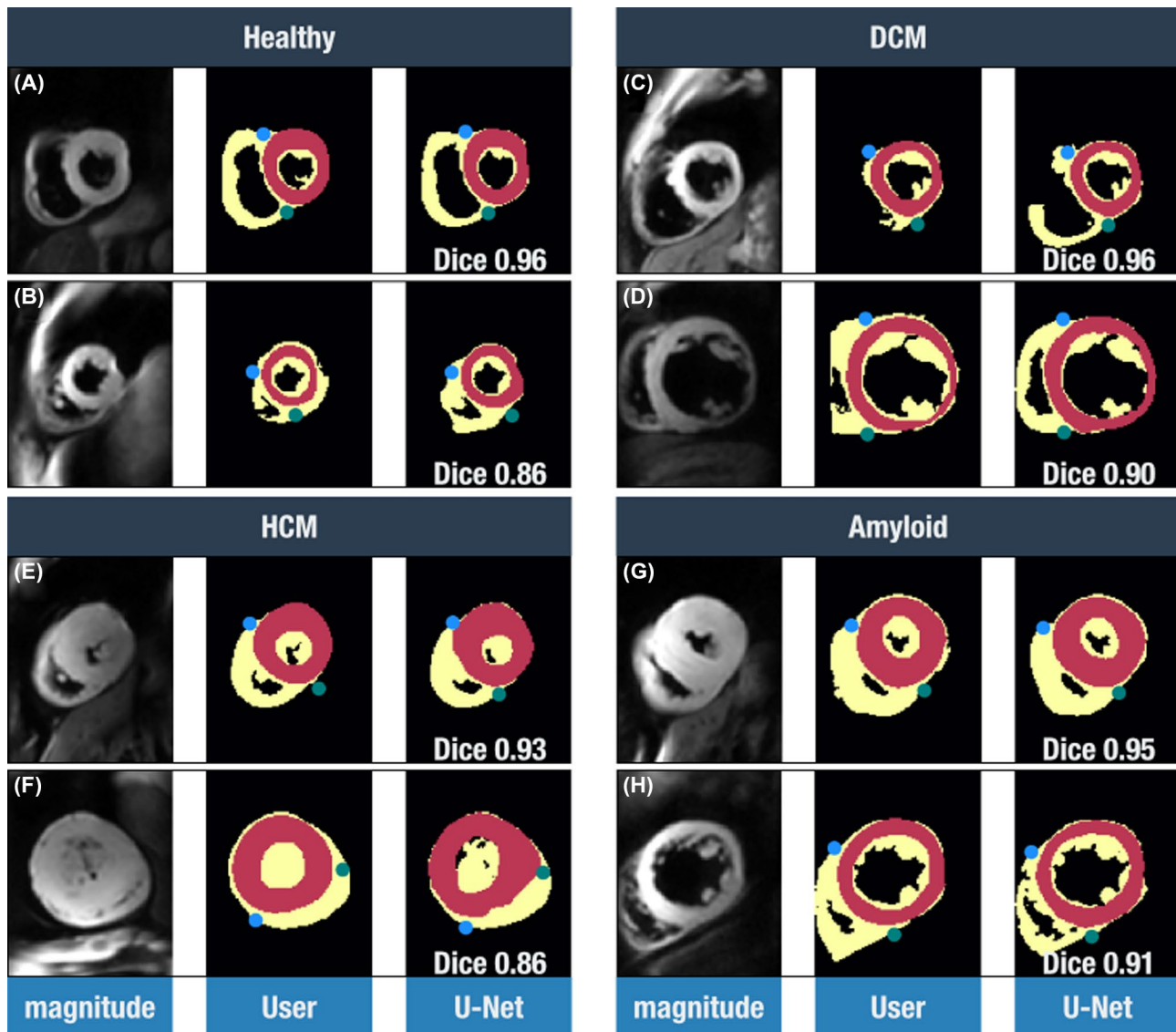


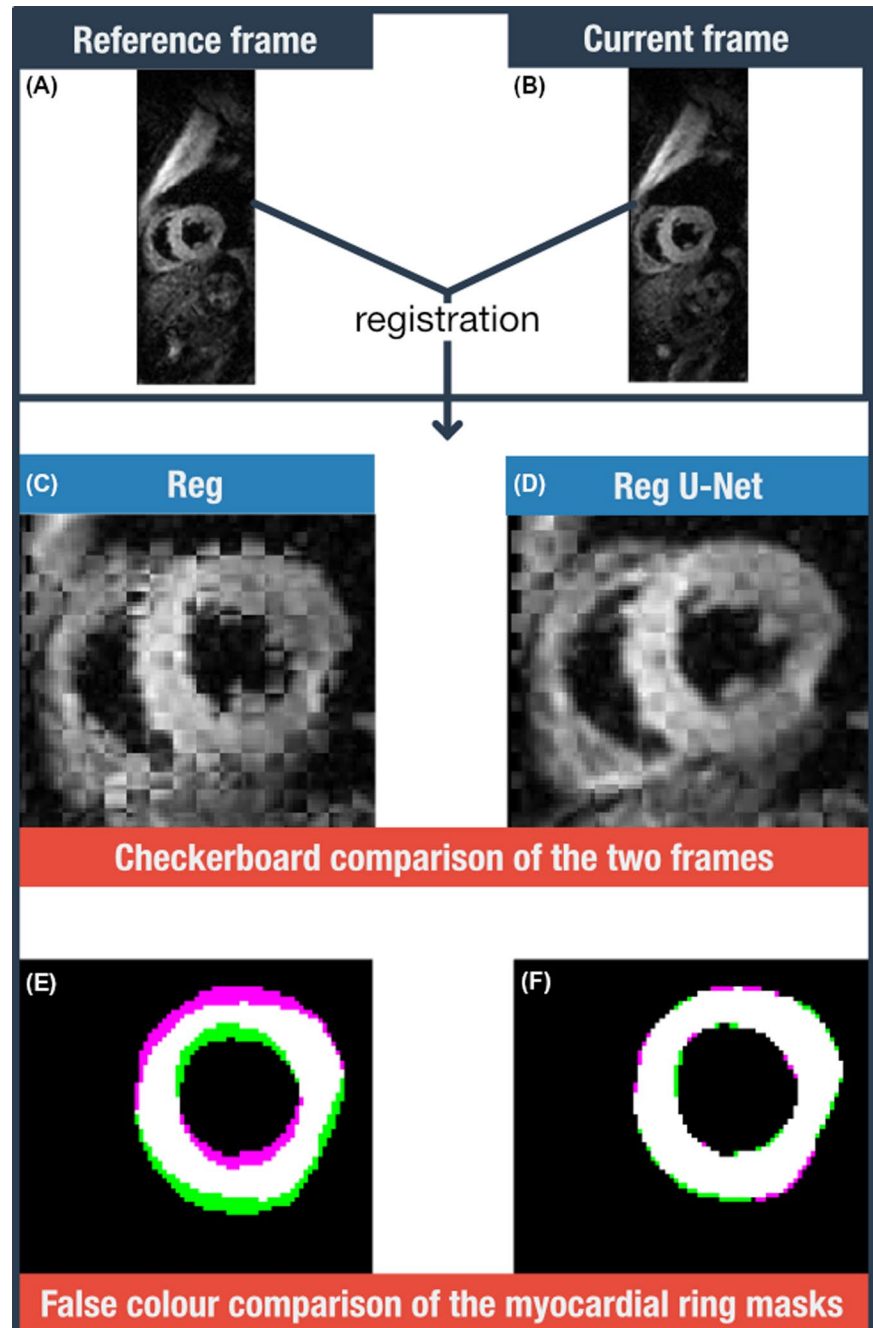
FIGURE 5 Best and worst Dice scores for each cohort: healthy (A,B), dilated cardiomyopathy (DCM; C,D), hypertrophic cardiomyopathy (HCM; E,F), and amyloid (G,H). The Dice score is shown on the bottom right of the U-Net segmentations

in FA ($P < .0001$), a small significant difference in absolute E2A ($P = .011$), and no statistically significant difference was found for MD ($P = .12$). Figure 8A shows Bland-Altman plots that visually compare these parameters and indicate the following bias [interquartile range]: healthy FA 0.019 $[-0.005, 0.045]$, MD $0.001 \cdot 10^{-3} \text{ mm}^2 \text{ s}^{-1} [-0.073, 0.051]$, and E2A $1.2^\circ [-2.7, 5.8]$; and clinical FA 0.020 $[0.002, 0.051]$, MD $-0.014 \cdot 10^{-3} \text{ mm}^2 \text{ s}^{-1} [-0.080, 0.027]$, and E2A $-0.1^\circ [-8.6, 5.7]$. A global median value for HA was not calculated due to its distinct transmural gradient.

A voxel-wise comparison was also performed in the overlapping myocardial masks between the two methods. We found the following intersubject mean absolute errors: FA 0.04 $[0.03, 0.06]$, MD $0.05 \cdot 10^{-3} \text{ mm}^2 \text{ s}^{-1} [0.03, 0.09]$, E2A $7.7^\circ [6, 10]$, and HA $7.9^\circ [5.5, 12]$. Figure 8B shows this same analysis but divided by cohort.

For some cases there was a visible difference in the DTI parameter maps when using the U-Net. Figure 9A,B shows a healthy control case in which conspicuous artifacts can be seen at the edges of the DTI maps for the manual processing, produced by residual misregistration of some of the images and resulting in notably higher FAs and lower MDs at the myocardial edges, and a less circularly symmetric HA than expected from previous healthy studies. Figure 9C,D shows a clinical amyloidosis case in which the manual processing, due to human error, failed to identify and remove diffusion-weighted images with a motion-induced signal loss artifact close to the anterior insertion point, resulting in higher FAs and MDs and abnormal HA values in this region. The fully automated method detected and rejected these corrupted images before tensor calculation, resulting in more homogeneous DTI values in the left ventricle.

FIGURE 6 Example of Reg U-Net performing noticeably better than the current Reg method for a particular diffusion-weighted image. A, Reference frame. B, Frame to be registered to the reference frame. C,D, Checkerboard composite of the two registration techniques. E,F, False color comparison of the myocardial masks for the two registration techniques



4 | DISCUSSION

The fully automated DT-CMR analysis, aided with a CNN, performed effectively, yielding similar results to an experienced user. The myocardial segmentation learned to correctly exclude papillary and RV trabeculation from the LV myocardium, even in patient data with a more varied morphology, including thin dilated cardiomyopathy and asymmetric HCM hearts. The only exception in which the U-Net differed considerably to the user was in two examples shown in Figure 5B,F. This is likely due to the very conspicuous EPI geometric distortion in the lateral wall and the very narrow LV cavity for the HCM scan. The U-Net was also capable of localizing the LV-RV insertion points.

This can be used to automatically divide the LV myocardium into different regions, such as the American Heart Association standardized segmentation model.¹⁸

The use of CNNs for cardiac segmentation is currently a hot topic in cardiac MRI. Most of the published work focuses on segmenting the left and right ventricle for cine imaging.¹⁹⁻²² To the best of our knowledge, this is the first published CNN trained specifically for in vivo STEAM DT-CMR data. The final aim is not only for LV segmentation, but to use this knowledge to automate and improve the entire postprocessing workflow of DT-CMR data.

When an experienced user processes the DT-CMR data manually, the RV region is thresholded but not carefully

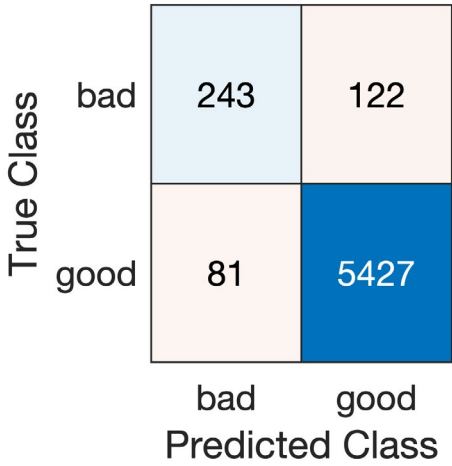


FIGURE 7 Confusion matrix of the identification of images corrupted with significant signal loss, labeled here as “bad.” The ground truth is considered to be the human visual analysis

delineated. This is due to our focus primarily on the LV myocardium in previous in vivo studies. Therefore, training the U-Net to segment the RV region was not as effective, and unsurprisingly results in a lower accuracy between the user and the U-Net with a Dice score of 0.85 [0.80, 0.89]. All of the DT-CMR results quoted in this work are for the LV region only; therefore, the lower RV accuracy is noteworthy but not as important as the LV.

Using a trained U-Net capable of segmenting the heart, and in particular the LV myocardial ring, before performing a rigid image registration algorithm improves the results of in vivo DT-CMR scans. Improvements are greatest when there are bright structures adjacent to the heart, commonly the chest wall, the liver, and the stomach. These organs do not move rigidly with the heart during the respiratory cycle; thus, image registration requires nonsmooth displacement field algorithms.²³ These methods have been shown to handle respiratory motion in the chest but are typically computationally expensive and thus may not be practical to implement with the large number of images in a typical DTI scan. The AI segmentation therefore provides a valid alternative to refine image registration toward a robust pixelwise diffusion tensor calculation.

The detection and removal of diffusion images corrupted with motion-induced signal-loss artifacts was also possible to automate with the aid of the U-Net. While scanning, we prescribe diffusion weightings that will keep the myocardial signal above the noise floor. If the heart contains significant fibrotic tissue or any other microstructural change that will lead to higher diffusion-induced signal loss, then ideally the b-value should be reduced to keep a reasonable amount of myocardial signal. This is important in order to be able to automatically separate diffusion encoding from more severe artifactual motion-induced signal loss. The lower magnitude signals caused by high diffusivity will lower primarily the Pearson coefficients. Our method of considering also the LV

TABLE 1 Global median DTI parameter comparison

		Controls		DCM		HCM		Amyloid	
		Diastole	Systole	Diastole	Systole	Diastole	Systole	Diastole	Systole
FA	User	0.54 [0.52, 0.59]	0.44 [0.42, 0.49]	0.50 [0.48, 0.51]	0.46 [0.43, 0.50]	–	0.43 [0.38, 0.46]	0.45 [0.44, 0.49]	0.35 [0.35, 0.35]
	U-Net	0.52 [0.49, 0.55]	0.43 [0.40, 0.47]	0.48 [0.45, 0.50]	0.44 [0.42, 0.47]	–	0.42 [0.37, 0.44]	0.41 [0.39, 0.46]	0.34 [0.34, 0.34]
MD	User	1.13 [1.08, 1.21]	1.01 [0.95, 1.05]	1.23 [1.18, 1.28]	1.08 [1.07, 1.12]	–	1.16 [1.02, 1.33]	1.26 [1.21, 1.50]	1.33 [1.33, 1.33]
	U-Net	1.12 [1.07, 1.19]	1.02 [0.95, 1.06]	1.23 [1.17, 1.28]	1.11 [1.06, 1.18]	–	1.16 [1.07, 1.31]	1.27 [1.25, 1.48]	1.33 [1.33, 1.33]
E2A	User	16.8 [13.7, 22.9]	55.2 [47.4, 66.6]	25.1 [23.9, 26.2]	50.5 [35.1, 57.0]	–	56.4 [51.4, 73.3]	50.7 [29.1, 53.0]	75.8 [75.8, 75.8]
	U-Net	14.4 [11.8, 18.8]	56.3 [46.9, 65.8]	23.0 [22.7, 23.4]	48.9 [36.2, 57.9]	–	56.2 [46.1, 73.1]	50.1 [29.6, 52.5]	76.6 [76.6, 76.6]

Note: Comparison is between user and U-Net: fractional anisotropy, mean diffusivity ($10^{-3} \text{ mm}^2 \text{ s}^{-1}$), and secondary eigenvector angle (degrees) separated by cardiac phase and cohort. The test data did not include any diastolic HCM examples.

Abbreviations: E2A, secondary eigenvector angle; FA, fractional anisotropy; MD, mean diffusivity.

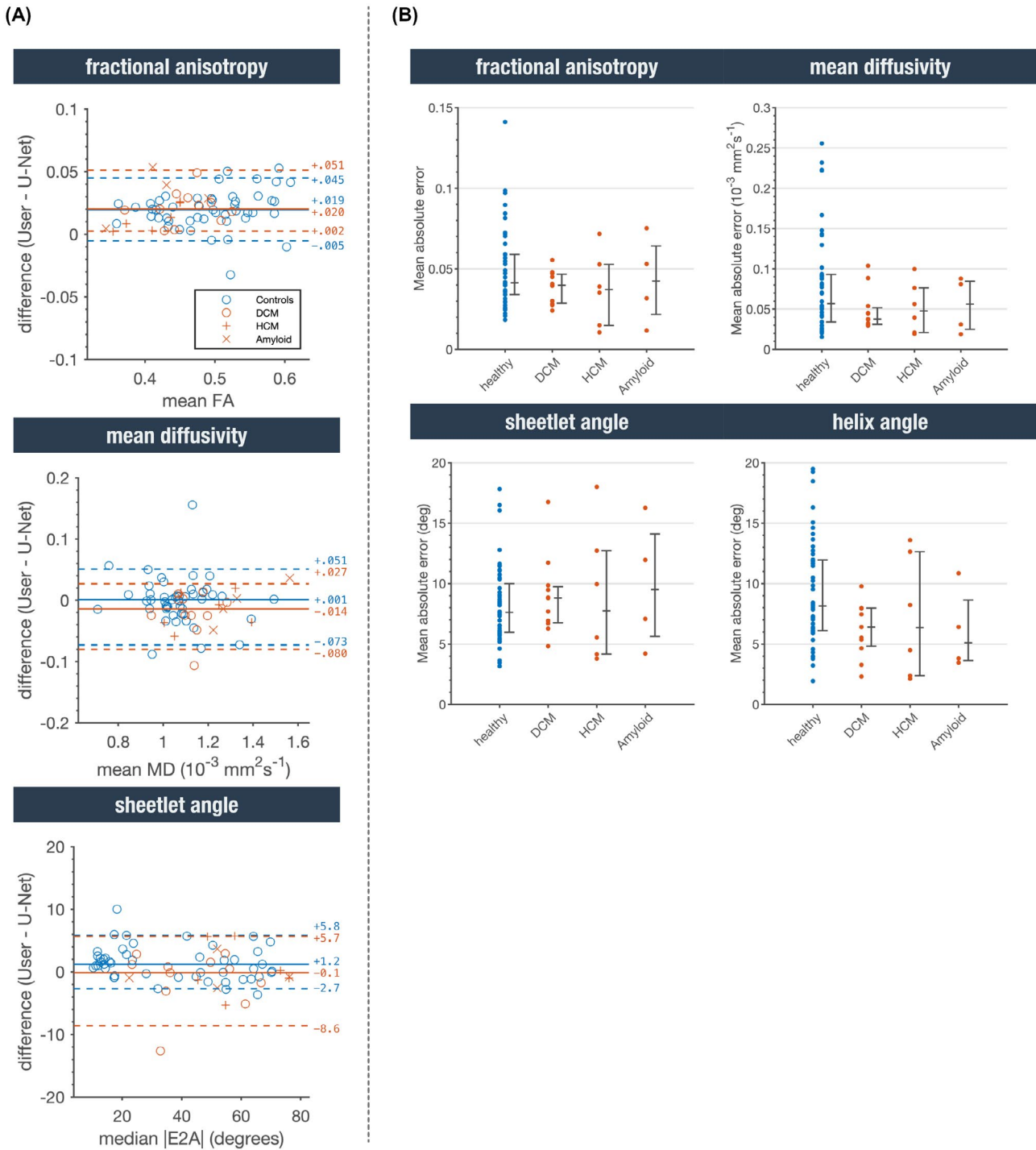


FIGURE 8 A, Bland-Altman plots for median fractional anisotropy, mean diffusivity, and sheetlet angle. The vertical lines and respective numbers represent the median and 5% and 95% quantile of the cohort. B, Voxel-wise difference in the overlapping myocardial masks. The data points indicate the mean absolute error for each subject discretized by cohort, and the bars indicate the respective interquartile ranges

myocardial mask and three scaled median absolute deviations below the median score is robust to the level of diffusion signal loss encountered routinely in our clinical scans. Overall, 6% of the diffusion images were deemed to be corrupted with signal loss by both the user and the U-Net. This was less than a previous study from our group, in which we found a 14% ratio.⁷ This is likely due to the increase in number and image quality

of our scan database. In this previous study we used a support vector machine approach to detect these corrupted frames. We have also experimented with a simple classification neural network.²⁴ Both previous methods were tested in less data and yielded higher F1 scores of 0.8 and 0.9, respectively, when compared with the U-Net approach used here with an F1 score of 0.7. The reason we selected the U-Net approach was

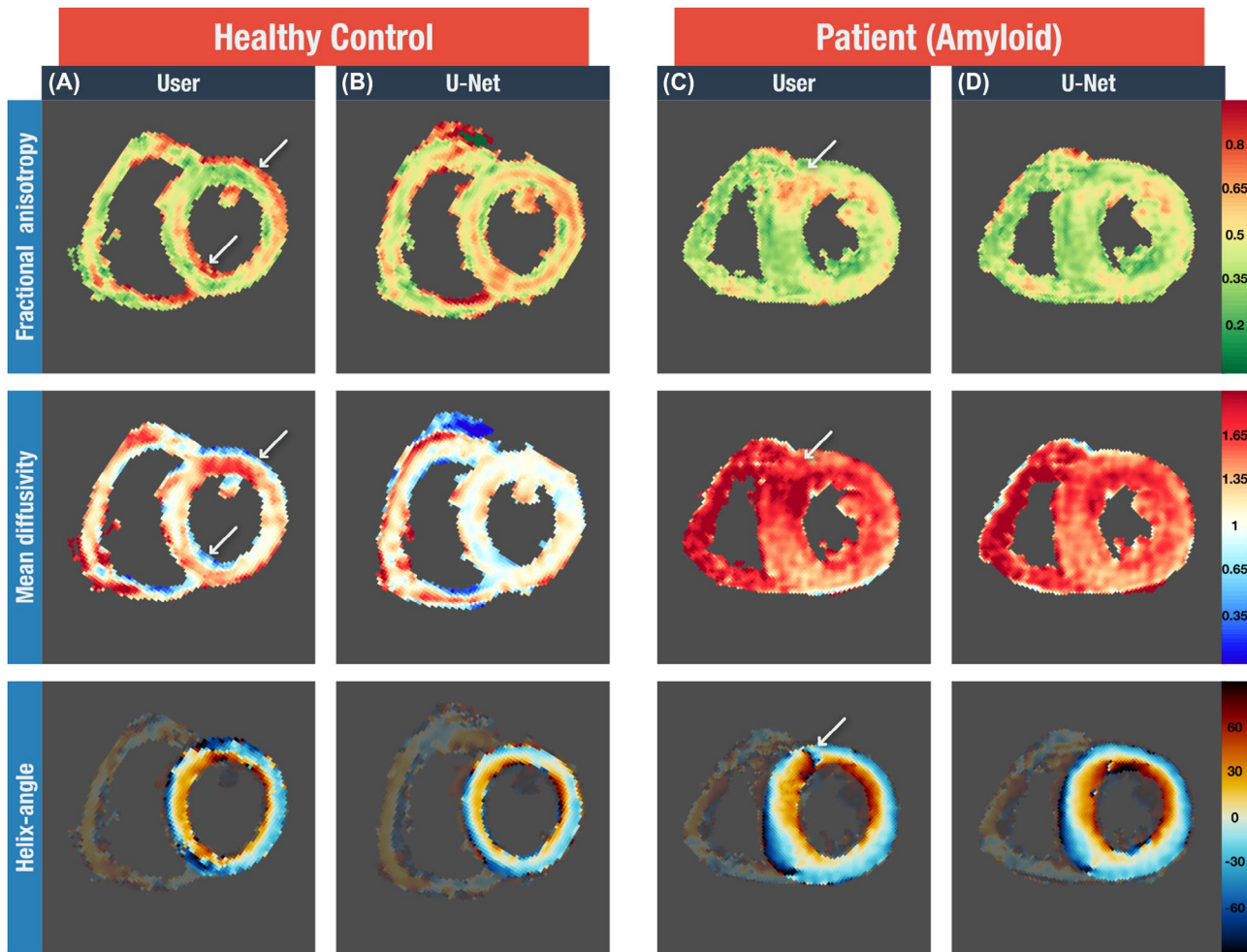


FIGURE 9 Diffusion tensor cardiac magnetic resonance parameter maps from the user (A,C) and U-Net (B,D) postprocessing workflow for 2 subjects (left: healthy control [A,B]; right: clinical amyloid [C,D]). In the healthy case (A,B), conspicuous artifacts can be seen at the edges of the DTI maps for the manual processing, produced by residual misregistration of some of the images and resulting in notably higher fractional anisotropy values and lower mean diffusivity values ($\times 10^{-3} \text{mm}^2/\text{s}$) at the myocardial edges (arrows), and a less circularly symmetric helix angle ($^\circ$). For the clinical amyloid case (C,D), the manual processing, due to human error, failed to identify and remove some diffusion-weighted images with a motion-induced signal loss artifact close to the anterior insertion point, resulting in higher fractional anisotropy values and mean diffusivity values, and unexpected helix angles in this region (arrows)

because of the SNR sensitivity of the previous methods when analyzing image features. They were found to be sensitive to SNR changes and not just myocardial signal loss. For a small number of low SNR scans, both previous methods rejected most of the images. Even though overall in the total pool of test images the match between the human observer and these previous methods was better, this SNR bias made us feel less confident in their clinical robustness. A similar human SNR bias, and the subjective nature of identifying the most borderline cases, as for example the partial signal loss image shown in Figure 1B, can perhaps be an explanation for the higher F1 scores of previous methods trained on the “ground truth” provided by previous human visual assessment.

In summary, using a DT-CMR trained CNN not only enabled the automation of postprocessing but it also increased the quality of the diffusion tensor calculation for some scans.

The U-Net improved the registration, resulting in less edge artifacts, and this is potentially a reason for the small but significant reduction in FA when using the U-Net, as generally these edge artifacts contain high FA values, as shown in Figure 9.

An important point when using AI is the issue of training networks to meet an expectation that may differ from the truth. This is especially important when using AI to reconstruct MR images, but also important for semantic segmentation networks, as the final results are influenced by the segmentation masks. The training data segmentation masks were drawn by 1 of 2 experienced clinicians and subsequently checked by an experienced physicist. We also have a mixture of cardiac phases and heart morphologies. This heterogeneity should reduce some of the learning bias. However, the ground truth for the data is unknown, and therefore it is likely the network assimilated human error and bias.

A potential area to explore in future work is to use the uncertainty of the U-Net segmentation to detect other image artifacts. For example, the EPI readout of the STEAM sequence occasionally leads to geometric distortions in regions with field inhomogeneities. This is typically seen in the lateral wall as in Figure 5B. The U-Net is expected to be less certain in artifact areas (ie, closer probabilities among the different classes). An uncertainty map could therefore be a valuable tool in visualizing the image quality and detecting regions affected by artifacts.

We are currently porting this postprocessing pipeline to the Gadgetron²⁵ framework, to enable real-time online reconstruction of DT-CMR results for on-the-fly feedback and to be packaged as DICOM files with the rest of the study. In terms of processing time, the added denoising and U-Net segmentation introduces a significant processing overhead of approximately 200%, although it is likely to be quicker overall when the user manual input time needed to remove corrupted frames, to threshold the heart and segment the left ventricle, is also considered.

The main limitation of this work is the sparse availability of in vivo DT-CMR data, currently acquired only in a few specialized centers around the world. To be efficient, deep learning requires a large number of data sets for training and validation. This limitation can be mitigated to a certain extent by using data augmentation techniques as used in this work. We have used data acquired by our group only; it contains healthy controls and cardiomyopathy scans. Other cardiac diseases, such as myocardial infarction, were not included at this time due to the low number of scans available. The automatic postprocessing is therefore untested for this condition, although data availability will increase as we continue our DT-CMR research. Additionally, the trained U-Net is currently untested for different scanners, different sequences, and different protocols. Motion-compensated spin-echo methods are often used in other centers as an alternative to the STEAM sequence.²⁶⁻²⁸ This will result in different contrasts, particularly in the blood pool; however, widely used transfer learning-based methods can be applied for multiscanner and multicenter studies. Other conditions like congenital heart diseases present a large spectrum of more exotic heart morphologies and are therefore more likely to be more challenging to automate. Because of these limitations, we have implemented a second AI aid mode in which the postprocessing presents the initial AI guess and waits for user acceptance or correction.

5 | CONCLUSIONS

The postprocessing of DT-CMR was successfully automated with a trained CNN, supporting real-time results and reducing human workload. The automatic segmentation of the

heart improved image registration, resulting in improvements of the calculated diffusion tensor parameters.

ACKNOWLEDGMENTS

This work was supported by the British Heart Foundation (RG/19/1/34160). The NVIDIA Quadro P6000 GPU used for this research was kindly donated by NVIDIA (Santa Clara, California).

CONFLICT OF INTEREST

Professor Dudley Pennell receives research support from Siemens and is a stockholder and director of Cardiovascular Imaging Solutions. Professor David Firmin receives research support from Siemens.

ORCID

Pedro F. Ferreira  <https://orcid.org/0000-0002-0436-3496>

Andrew D. Scott  <https://orcid.org/0000-0001-7656-3123>

Guang Yang  <https://orcid.org/0000-0001-7344-7733>

Sonia Nelles-Vallespin  <https://orcid.org/0000-0003-0412-2796>

Dudley J. Pennell  <https://orcid.org/0000-0001-5523-1314>

David N. Firmin  <https://orcid.org/0000-0003-3894-7489>

TWITTER

Pedro F. Ferreira  @pedro_filipe_

REFERENCES

1. Nelles-Vallespin S, Khalique Z, Ferreira PF, et al. Assessment of myocardial microstructural dynamics by in vivo diffusion tensor cardiac magnetic resonance. *J Am Coll Cardiol*. 2017;69:661-676.
2. Khalique Z, Ferreira PF, Scott AD, et al. Diffusion tensor cardiovascular magnetic resonance of microstructural recovery in dilated cardiomyopathy. *JACC Cardiovasc Imaging*. 2018;11:1548-1550.
3. Nguyen CT, Dawkins J, Bi X, Marbán E, Li D. Diffusion tensor cardiac magnetic resonance reveals exosomes from cardiomyocyte-derived cells preserve myocardial fiber architecture after myocardial infarction. *JACC Basic Transl Sci*. 2018;3:97-109.
4. von Deuster C, Sammut E, Asner L, et al. Studying dynamic myofiber aggregate reorientation in dilated cardiomyopathy using in vivo magnetic resonance diffusion tensor imaging. *Circ Cardiovasc Imaging*. 2016;9:e005018.
5. Ariga R, Tunnicliffe EM, Manohar SG, et al. Identification of myocardial disarray in patients with hypertrophic cardiomyopathy and ventricular arrhythmias. *J Am Coll Cardiol*. 2019;73:2493-2502.
6. Gotschy A, von Deuster C, van Gorkum RJH, et al. Characterizing cardiac involvement in amyloidosis using cardiovascular magnetic resonance diffusion tensor imaging. *J Cardiovasc Magn Reson*. 2019;21:56.
7. Cavallo F, Ferreira PF, Khalique Z, Scott AD, Nelles-Vallespin S, Firmin DN. Automatic detection of corrupted frames in cardiac DTI with machine learning. In Proceedings of the 25th Annual Meeting of ISMRM, Honolulu, HI, 2017. Abstract #4763.

8. Streeter DD Jr, Spotnitz HM, Patel DP, Ross J Jr, Sonnenblick EH. Fiber orientation in the canine left ventricle during diastole and systole. *Circ Res*. 1969;24:339-347.
9. Ferreira PF, Kilner PJ, McGill L-A, et al. In vivo cardiovascular magnetic resonance diffusion tensor imaging shows evidence of abnormal myocardial laminar orientations and mobility in hypertrophic cardiomyopathy. *J Cardiovasc Magn Reson*. 2014;16:87.
10. Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Cham, Switzerland: Springer; 2015.
11. Nielles-Vallespin S, Mekkaoui C, Gatehouse P, et al. In vivo diffusion tensor MRI of the human heart: reproducibility of breath-hold and navigator-based approaches. *Magn Reson Med*. 2013;70:454-465.
12. Scott AD, Ferreira PF, ADC, Nielles-Vallespin S, et al. Optimal diffusion weighting for in vivo cardiac diffusion tensor imaging. *Magn Reson Med*. 2015;74:420-430.
13. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Lecture Notes in Computer Science*. Springer International Publishing; 2017:240-248.
14. Kingma DP, Ba J. "Adam: a method for stochastic optimization. In: *Proceedings from the 3rd International Conference for Learning Representation*, San Diego, California, 2015.
15. Buades A, Coll B, Morel J-M. A non-local algorithm for image denoising. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, 2005.
16. Immerkær J. Fast noise variance estimation. *Comput Vis Image Underst*. 1996;64:300-302.
17. Guizar-Sicairos M, Thurman ST, Fienup JR. Efficient subpixel image registration algorithms. *Opt Lett*. 2008;33:156-158.
18. Cerqueira MD, Weissman NJ, Dilsizian V, 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*. 2002;105:539-542.
19. Bernard O, Lalande A, Zotti C, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging*. 2018;37:2514-2525.
20. Duan J, Bello G, Schlemper J, et al. Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. *IEEE Trans Med Imaging*. 2019;38:2151-2164.
21. Oktay O, Ferrante E, Kamnitsas K, et al. Anatomically Constrained Neural Networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Trans Med Imaging*. 2018;37:384-395.
22. Zotti C, Luo Z, Lalande A, Jodoin P-M. Convolutional neural network with shape prior applied to cardiac MRI segmentation. *IEEE J Biomed Health Inform*. 2019;23:1119-1128.
23. Vishnevskiy V, Gass T, Székely G, Tanner C, Goksel O. Isotropic Total variation regularization of displacements in parametric image registration. *IEEE Trans Med Imaging*. 2017;36:385-395.
24. Ferreira PF, Scott AD, Khalique Z, et al. Fully automated DT-CMR analysis with deep learning. In *Proceedings of the 27th Annual Meeting of ISMRM*, Montreal, Canada, 2019. Abstract #2036.
25. Hansen MS, Sørensen TS. Gadgetron: an open source framework for medical image reconstruction. *Magn Reson Med*. 2013;69:1768-1776.
26. Scott AD, Nielles-Vallespin S, Ferreira PF, et al. An in-vivo comparison of stimulated-echo and motion compensated spin-echo sequences for 3 T diffusion tensor cardiovascular magnetic resonance at multiple cardiac phases. *J Cardiovasc Magn Reson*. 2018;20:1.
27. Stoeck CT, von Deuster C, Genet M, Atkinson D, Kozerke S. Second-order motion-compensated spin echo diffusion tensor imaging of the human heart. *Magn Reson Med*. 2016;75:1669-1676.
28. von Deuster C, Stoeck CT, Genet M, Atkinson D, Kozerke S. Spin echo versus stimulated echo diffusion tensor imaging of the in vivo human heart. *Magn Reson Med*. 2016;76:862-872.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

FIGURE S1 Diagram showing the population distribution of healthy and clinical data discretised by cohort

FIGURE S2 Flowchart of the two image registration algorithms: old motion registration method Reg (top) and the new AI-aided method Reg U-Net (bottom)

FIGURE S3 Image classification workflow. A, examples of denoised diffusion images from the entire scan (the top right image has partial signal loss, while the bottom right image has total signal loss in the heart due to residual cardiac motion). B, the corresponding heart segmentation masks (yellow LV myocardium, green RV and papillary muscle). C, the mean(Pearson, Dice) scores of all the scan images with two flagged outliers corresponding to the images with signal loss artefacts shown above

FIGURE S4 Dice coefficients comparison between Reg and Reg U-Net image registration workflows

VIDEO S1 Segmentation comparison between the user and the trained U-Net for all 70 test cases

How to cite this article: Ferreira PF, Martin RR, Scott AD, et al. Automating in vivo cardiac diffusion tensor postprocessing with deep learning-based segmentation. *Magn Reson Med*. 2020;00:1-14. <https://doi.org/10.1002/mrm.28294>