# TVnet: Automated Time-Resolved Tracking of the Tricuspid Valve Plane in MRI Long-Axis Cine Images with a Dual-Stage Deep Learning Pipeline

Ricardo A. Gonzales[1,2,3], Jérôme Lamy[1], Felicia Seemann[1,3,4], Einar Heiberg[3,4,5], John A. Onofrey[1], and Dana C. Peters[1]

[1] Department of Radiology and Biomedical Imaging, Yale School of Medicine, Yale University, New Haven, Connecticut, United States of America
[2] Department of Electrical Engineering, Universidad de Ingeniería y Tecnología, Lima, Peru
ricardo.gonzales@utec.edu.pe
[3] Department of Clinical Physiology, Lund University, Skåne University Hospital, Lund, Sweden
[4] Department of Biomedical Engineering, Lund University, Lund, Sweden
[5] Wallenberg Center for Molecular Medicine, Lund University, Lund, Sweden

**Abstract.** Tracking the tricuspid valve (TV) in magnetic resonance imaging (MRI) long-axis cine images has the potential to aid in the evaluation of right ventricular dysfunction, which is common in congenital heart disease and pulmonary hypertension. However, this annotation task remains difficult and time-demanding as the TV moves rapidly and is barely distinguishable from the myocardium. This study presents TVnet, a novel dual-stage deep learning pipeline based on ResNet-50 and automated image linear transformation, able to automatically derive tricuspid annular plane systolic excursion. Stage 1 uses a trained network for a coarse detection of the TV points, which are used by stage 2 to reorient the cine into a standardized size, cropping, resolution, and heart orientation and to accurately locate the TV points with another trained network. The model was trained and evaluated on 4170 images from 140 patients with diverse cardiovascular pathologies. A baseline model without standardization achieved a Euclidean distance error of $4.0 \pm 3.1$ mm and a clinical-metric agreement of ICC = 0.87, whereas a standardized model improved the agreement to $2.4 \pm 1.7$ mm and an ICC = 0.94, on par with an evaluated inter-observer variability of $2.9 \pm 2.9$ mm and an ICC = 0.92, respectively. This novel dual-stage deep learning pipeline substantially improved the annotation accuracy compared to a baseline model, paving the way towards reliable right ventricular dysfunction assessment with MRI.

**Keywords:** Right ventricular dysfunction · Cine MRI · Annotation · Residual neural networks.

## 1 Introduction

The analysis of the tricuspid valve (TV) motion provides anatomic and functional evaluation of the right ventricle (RV) [1]. Compared to the left ventricle, whose functional evaluation is paramount in the diagnosis of cardiovascular disease, the RV involvement in conditions such as pulmonary hypertension and congenital heart disease, and the high prognostic value of RV function evaluation are often overlooked [2]. The maximal longitudinal displacement, known as tricuspid annular plane systolic excursion (TAPSE), and the annular early diastolic velocity (RV e') are commonly evaluated in echocardiography to assess right ventricular dysfunction [3]. However, this imaging modality can be limited by the dependency of image quality on beam angle [4] and influenced by translational motion [5].

Magnetic resonance imaging (MRI) is a more reproducible imaging modality and is considered the reference standard for cardiac volume assessment. Its accuracy and reliability serve as a promising tool for serial examinations of RV function [6]. Although MRI would never be performed just to measure TAPSE or RV e', developing its capability to do so is important as these parameters can be assessed retrospectively from standard cardiovascular cine images acquired systematically in clinical practice. Furthermore, this analysis of both mitral and tricuspid valve motion by MRI has been shown to be reliable and to provide an evaluation of both systolic and diastolic function [7, 8]. However, their measurement requires a precise annotation in the valve insertion points in every temporal frame of a long-axis cine, typically consisting of 30 temporal frames, which is time-demanding and prone to errors [9]. Although some progress has been made to semi-automatically track the valve insertion points in these images [10, 11], a fully automated, fast, and accurate method for tracking the TV points using standard clinical MRI images is lacking.

The variability of long-axis views of the heart in orientation and location within the cine image is one obstacle towards precise automated identification of the TV points. We addressed this using a novel two-stage framework with an intermediate automated linear transformation task. A first trained neural network produces coarse annotation, which in turn is used to fully standardize the image with respect to size, centering, and rotation. A second network is then used to precisely annotate in these preprocessed images, to automatically track the tricuspid valve insertion points and derive TAPSE and RV e' on par with expert human-level performance. This method, trained and validated in a diverse population of 140 patients, showed an excellent agreement with the inter-observer variability and with a smoother trajectory annotation.

## 2 Methods

### 2.1 Imaging Data and Manual Annotation

One hundred forty patients were retrospectively included in this study (59 females, 53±14 years old) as part of an IRB-approved chart-review study. The

reported pathologies included pulmonary hypertension (n=25), atrial fibrillation (n=23), ventricular tachycardia or paroxysmal fibrillation (n=16), hypertrophic cardiomyopathy (n=17), heart failure (n=9), right or left bundle branch block (n=6), coronary artery disease (n=6), sarcoid (n=6), obstructive sleep apnea (n=5), other cardiac diseases (n=5), Lyme disease (n=4), and healthy volunteers (n=18). All subjects were scanned on a 1.5T (n=115) or 3T (n=25) conventional clinical MRI scanner (Siemens Healthcare, Erlangen) including a standard four-chamber long-axis cine exam, systematically acquired in a clinical setting. Such images, mostly retrospectively ECG-gated, were acquired during breath-hold with the following parameters: repetition time of 3 ms, echo time of 1.5 ms, and flip angle of 60°. Four chamber data had an average spatial resolution of $2 \times 2$ mm$^2$, slice thickness of 8 mm, and mostly 30 temporal frames per cardiac cycle were reconstructed. The final dataset comprised a total of 4170 images (140 sets of time-resolved images) analyzed in previous studies [12–14].

RV lateral and septal points were manually placed at the intersection between the tricuspid valve and the right ventricular myocardium by observer 1, a researcher with 4 years of cardiac MRI experience. The landmark annotations were performed in every image, using the software Segment [15]. The landmarks, annotated by observer 1, were considered as the ground truth.

### 2.2 Dual-Stage Residual Neural Network

The proposed framework, illustrated in Fig. 1, involves two trained networks, using an established convolutional neural network with a residual framework of 50 layers, ResNet-50 [16]. The first stage, with a network, provides valve points sufficiently accurate to define the TV plane. The second stage uses these points to perform a linear transformation to the image and, with another network, predicts highly accurate points. For both networks, the input and output layers were adapted to input a grayscale image with a fixed size, and to output four numbers (i.e. two pairs of coordinates) with a regression layer. For both networks, the training and testing sets were partitioned into 3330 images (112 subjects) and 840 images (28 subjects), respectively. This distribution was performed homogeneously along the reported cardiovascular diseases in a random manner. To further mitigate selection bias, a 5-fold cross-validation with homogeneous distribution was additionally performed.

**Initial Coarse Annotation** The first ResNet-50 network was trained with the ground-truth manual annotations and images "as is" with only an image resizing to 160×160 with cubic interpolation. Every input image was processed independently. The resulting coordinates were sufficiently accurate to localize both TV points and to roughly follow the valve motion.

**Automated Linear Transformation** The output coordinates of the first network were used to standardize the input of the second network: (i) the image was interpolated to a standard spatial resolution of 0.75 mm, (ii) the TV was

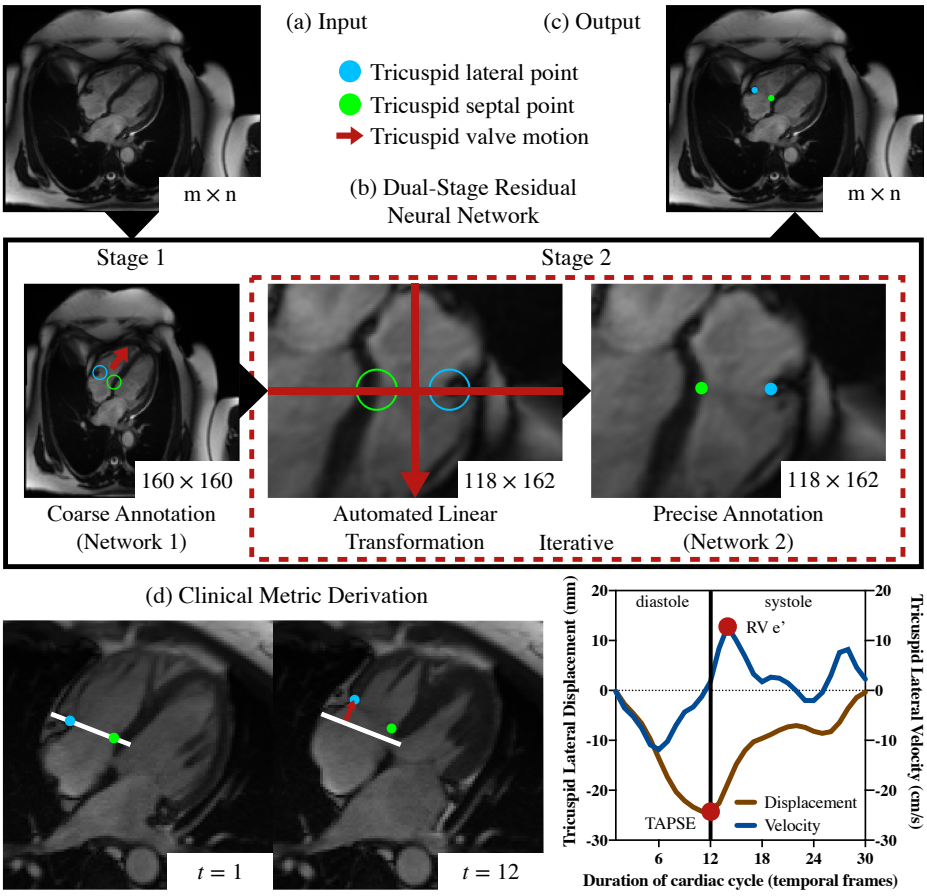


**Fig. 1.** TVnet pipeline. (a) The input cine images with an inherent clinical variability (size $(m \times n)$, resolution, orientation, and cropping) are fed to (b) the proposed dual-stage residual neural network. Stage 1 produces coarse annotations, marked in circumferences representing uncertainty, in every cine image in a fixed image size of $160 \times 160$. Stage 2 uses these points to apply a linear transformation to a standard spatial resolution of 0.75 mm, orientation and cropping around the TV center for a size of $118 \times 162$, and predicts precise annotations, marked in circles representing higher accuracy, which are adjusted again to the original input image. Stage 2, dashed red box, can be done iteratively as indicated. (c) The output time-resolved coordinates are used to (d) derive the tricuspid lateral displacement curve by measuring the perpendicular distance from the initial TV plane, marked in white at end-diastole $(t = 1)$, to the lateral point at every temporal frame, and the tricuspid lateral velocity, as its time-derivative. TAPSE is calculated as the maximum displacement, and RV e' is calculated as the second global velocity peak.

oriented horizontally (rotated) with the apex pointing down and with the left ventricle placed to the left, and (iii) the image was finally cropped around the TV center for a size of 118×162. The orientation and centering tasks were based on the first temporal frame and applied to the rest of the cine, accordingly. The first network processed the cine images independently and its output trajectory allowed to determine the valve plane for each image. The direction of the RV apex was determined by the valve movement direction along the cardiac cycle.

**Final Precise Annotation** The second ResNet-50 network was trained with the ground-truth manual annotations and the linearly transformed images. These standardized images were used as input to the second network which generated the TV points with high accuracy, processing every image independently. These new predicted points were readjusted, with an inverse standardization transformation, to match the original input cine images.

**Network Training** Both networks were trained in the same manner. A pre-trained version of the ResNet-50, from the ImageNet database [17], was used for weights initialization as part of a transfer learning process to reduce convergence time. Image pixel distribution was centered and reduced using the median and the interquartile range. Training data augmentation was performed 10-fold by scaling ±10%, rotating ±10° and translating ±3 pixels, to add more inherent variability in stage 1, and to compensate for any error from stage 1 in stage 2, i.e., a slight misalignment of the valve plane center from the ground truth. The Adam method [18] was used for optimizing the mean square error loss function with a learning rate of $1 \times 10^{-4}$. Training consisted of 20 epochs with a mini-batch size of 8. The networks were trained and tested on MATLAB R2019b (Mathworks, Natick, MA) with a NVIDIA Titan RTX GPU.

### 2.3 Evaluation and Statistical Analysis

**Clinical-Metric Derivation** Clinically relevant RV parameters extracted were the TAPSE and the RV e', which were calculated from the lateral displacement curve, as illustrated in Fig. 1(d). The perpendicular distance from the lateral point to the initial TV plane, defined at the beginning of the cardiac cycle (end-diastole, usually $t = 1$) with both lateral and septal points, was measured in every temporal frame. TAPSE was measured as the maximum value of this displacement curve. RV e' was calculated as the second global peak of the time-derivative of this curve.

The test set, comprising 840 annotations for each TV insertion point, from 28 patients, was evaluated against manual annotation. The spatial annotation error was measured with: (i) the Euclidean distance, which measured in millimeters the distance error between the ground-truth and predicted annotations, (ii) the angular distance, which measured in degrees the inner intersection angle of the ground-truth and automated planes defined by both TV points, and (iii) the trajectory jitter, which evaluated the difference ratio between the actual point

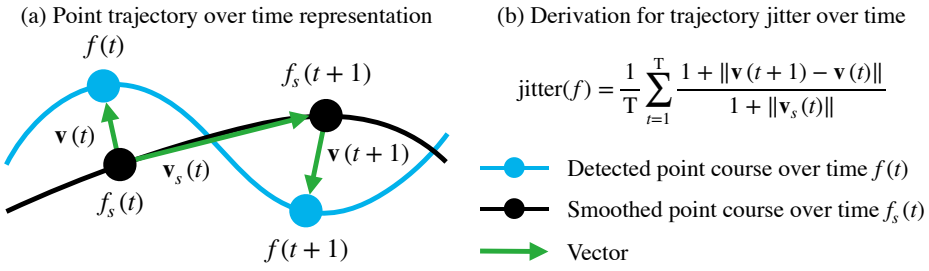


**Fig. 2.** Point trajectory jitter illustration (a) and derivation (b). The jitter for two successive points over time $t$ of the trajectory with T temporal frames was measured as the difference between their vectors $\mathbf{v}(t)$ and $\mathbf{v}(t+1)$ from $f(t)$ to its smoothed contour $f_s(t)$, after a low-pass filter and normalized by the distance, magnitude of $\mathbf{v}_s(t)$, between the two corresponding points on $f_s(t)$. The trajectory jitter was evaluated by averaging every two points jitter along the trajectory $f_s(t)$ and was performed for every annotated point course.

course over time and its smoothed version, representing a more physiological pattern, as illustrated in Fig. 2. Clinical metric (TAPSE and RV e') comparisons were performed using linear regression analysis, Bland-Altman plots, the intraclass correlation coefficient (ICC), and the coefficient of variance (CoV) between the automated and manual measures. For inter-observer variability analysis, observer 2, a researcher of 7 years of cardiac MRI experience, performed manual annotations on the same test set and the same evaluation was assessed. The threshold for statistical significance was considered for $p<0.05$.

**Additional Stage Influence** This evaluation was performed for the results of using only the first stage (stage 1), both stages (stage 1+2), and three successive iterations (stage 1+2+2, stage 1+2+2+2, and stage 1+2+2+2+2), to show the influence of the additional stage (see red box in Fig. 1(b)).

## 3 Experiments and Results

### 3.1 Implementation

TVnet was implemented as a plug-in in the medical image analysis software Segment v3.1 R8109 [15] (http://segment.heiberg.se) and uploaded to GitHub (https://github.com/ra-gonzales/TVnet), which is freely available for research purposes. Training time took 22 h for each network. For the dual-stage pipeline, on the GPU, testing time took 1.8 s per patient with an additional 0.8 s for every iteration, whereas on a CPU, it took 3.8 s per patient with an additional 1.6 s for every iteration, compared to an average manual annotation time from 4 to 20 min. Additional movie files, in Supplementary Material, demonstrate automated TV tracking across a cardiac cycle.

**Table 1.** Automated spatial annotation accuracy (n=840) of tricuspid valve tracking.

| Stage | Lateral point error (mm) | Septal point error (mm) | Angular error (°) | TAPSE error (mm) |
|---|---|---|---|---|
| 1 | $4.03 \pm 3.13$ | $3.98 \pm 3.16$ | $4.40 \pm 3.91$ | $0.32 \pm 2.86$ |
| 1+2 | $2.61 \pm 2.01$ | $2.61 \pm 1.59$ | $3.82 \pm 3.36$ | $0.29 \pm 1.86$ |
| 1+2+2 | $2.43 \pm 1.92$ | $2.45 \pm 1.48$ | $3.78 \pm 3.34$ | $0.13 \pm 1.90$ |
| 1+2+2+2 | $2.41 \pm 1.97$ | $2.43 \pm 1.50$ | $3.87 \pm 3.38$ | $-0.02 \pm 1.80$ |
| 1+2+2+2+2 | $2.38 \pm 1.93$ | $2.48 \pm 1.51$ | $3.95 \pm 3.37$ | $0.07 \pm 2.03$ |
| Observer 2 | $2.18 \pm 1.78$ | $3.67 \pm 3.49$ | $4.83 \pm 6.01$ | $-0.20 \pm 2.30$ |

The mean ± standard deviation are reported for Euclidean (for lateral and septal points) and angular distances and tricuspid annular plane systolic excursion (TAPSE) errors between detected and ground-truth manual annotations by observer 1.

### 3.2 Annotation Accuracy

The reproducibility analysis of the TVnet annotations and the clinical metrics is reported in Table 1. The proposed framework improved the accuracy after every iteration converging on stage 1+2+2, with a marginal improvement afterward. With this chosen model, the method achieved a mean Euclidean distance error of $2.44 \pm 1.71$ mm, an angular distance error of $3.78 \pm 3.34°$, and a TAPSE error of $0.13 \pm 1.90$ mm with an ICC = 0.94. The mean Euclidean distance error with standard deviation across folds was $2.79 \pm 0.39$ mm. The inter-observer variability for manual annotation achieved $2.92 \pm 2.87$ mm, $4.83 \pm 6.01°$, $-0.20 \pm 2.30$ mm, and ICC = 0.92, respectively, with a marked discordance in tracing the septal point. The trajectory jitter with the framework was $0.70 \pm 0.10$ whereas the trajectory jitter of observer 2 was $0.81 \pm 0.13$. Compared to the difference between automated and ground truth annotations, the second observer achieved a stronger agreement with the lateral point annotation but weaker agreement with the septal point annotation, whereas the automated annotations exhibited less jitter and achieved a better agreement with the clinical metrics.

The automatically derived annotations yielded average values of $19.9 \pm 5.5$ mm for TAPSE and $7.8 \pm 3.3$ cm/s for RV e'. These were comparable with manually derived average values of these clinical metrics from both observers (observer 1: $19.8 \pm 5.7$ mm and $6.8 \pm 2.8$ cm/s and observer 2: $19.6 \pm 5.8$ mm and $9.2 \pm 4.1$ cm/s respectively for TAPSE and RV e'). The regression and Bland-Altman plots for TAPSE and RV e' between the automated (stage 1+2+2) and manual measurements are presented in Fig. 3(a), where an excellent correlation and agreement were observed for each parameter. Figure 3(b) shows the corresponding inter-observer variability of these metrics, demonstrating similar excellent agreement for TAPSE, but less agreement for RV e'. This is due to the sensitivity of RV e' to jitter, as RV e' requires a time derivative. All reported correlation values were significant ($p<0.0001$).

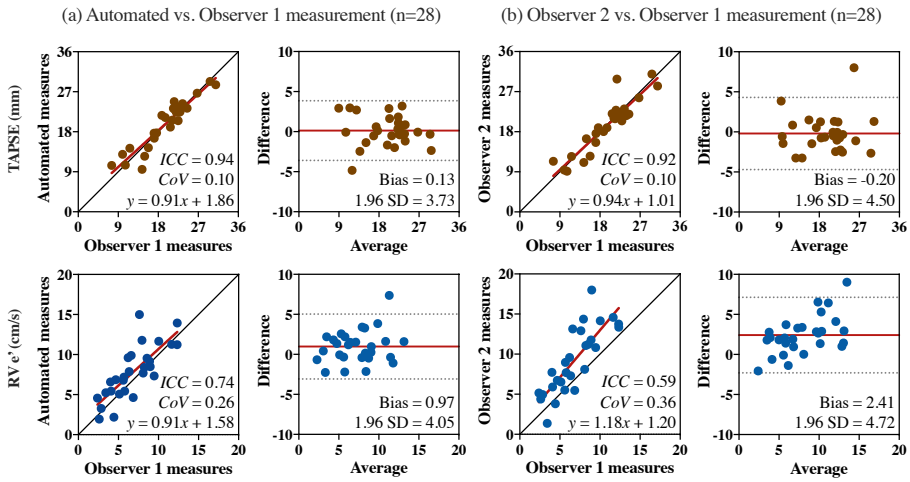


**Fig. 3.** Correlation and bias of tricuspid valve MRI-derived clinical measures between an expert manual annotation (observer 1) and (a) the automated method, and (b) annotation by observer 2, on the testing set of 28 patients. The first column of each analysis shows the regression plots whereas the second shows the Bland-Altman plots of the tricuspid annular plane systolic excursion (TAPSE), in brown, and the right ventricular early diastolic velocity (RV e'), in blue. In each scatter plot the red line denotes the regression line and the black line denotes the identity line, whereas in each Bland-Altman plot, the red line denotes the mean difference (bias) and the two light dotted lines denote ± 1.96 standard deviations from the mean.

## 4 Discussion and Conclusion

In this work, we proposed a dual-stage residual learning framework, TVnet, for time-resolved annotation of the TV in four-chamber views from standard long-axis cine MRI images. The proposed method was fast, fully automated, and showed excellent agreement with manual annotation by expert readers in terms of valve points positioning as well as with the subsequently extracted RV function parameters. Tedious manual labor is not needed, reducing the processing time from 4-20 min to 2 s. This enables fully automated, accurate, fast, and reproducible assessment of RV function in clinical routine. Moreover, this method can be applied retrospectively to any four-chamber image acquisition, which is routinely acquired in a standard cardiovascular MRI exam.

The initial step, used to standardize the images with a linear transformation, substantially reduced the spatial annotation and clinical-metric errors. It further exploited the utility of artificial neural networks in readjusting the result from a baseline model using the TV anatomical position and reducing the image variability. The low Euclidean distance and angular orientation errors displayed consistency in the tracking. The former assured a point placement around the ground-truth annotation, whereas the latter assured a concordant TV plane dis-

placement, which for the RV is known to vary greatly throughout the cardiac cycle. The high parity with human-level performance may be explained with the low trajectory jitter of the model, as the automated point prediction was smoother compared to a second observer, which is vital in calculating TV velocity. Future work includes validation on a multi-vendor, multi-center population and expansion to other cardiac landmark annotations.

**Acknowledgements** The authors gratefully acknowledge funding from NHLBI R01HL144706, RAG acknowledges Magnus Caspersen, MSc for his guidance in deep learning, and DCP acknowledges James W. Goldfarb, PhD for his ideas on the utilization of deep learning for cine valve-tracking, many years ago.

## References

1. Dimopoulos, K., Condliffe, R., Tulloh, R.M., Clift, P., Alonso-Gonzalez, R., Bedair, R., Chung, N.A., Coghlan, G., Fitzsimmons, S., Frigiola, A., Howard, L.S., Jenkins, P., Kenny, D., Li, W., MacDonald, S.T., McCabe, C., Oliver, J.J., Spence, M.S., Szantho, G.V., von Klemperer, K., Wilson, D.G., Wort, S.J.: Echocardiographic screening for pulmonary hypertension in congenital heart disease: JACC review topic of the week. Journal of the American College of Cardiology 72(22), 2778–2788 (2018)
2. Amsallem, M., Mercier, O., Kobayashi, Y., Moneghetti, K., Haddad, F.: Forgotten no more: a focused update on the right ventricle in cardiovascular disease. JACC: Heart Failure 6(11), 891–903 (2018)
3. D'Andrea, A., Vriz, O., Carbone, A., Ferrara, F., Di Maio, M., Cocchia, R., Tagliamonte, G., Acri, E., Driussi, C., Pezzullo, E., Citro, R., Cittadini, A., Calabrò, R., Giovanna Russo, M., Bossone, E.: The impact of age and gender on right ventricular diastolic function among healthy adults. Journal of Cardiology 70(4), 387–395 (2017)
4. Ho, C.Y., Solomon, S.D.: A clinician's guide to tissue doppler imaging. Circulation 113(10), e396–e398 (2006)
5. Abraham, T.P., Dimaano, V.L., Liang, H.Y.: Role of tissue doppler and strain echocardiography in current clinical practice. Circulation 116(22), 2597–2609 (2007)
6. Valsangiacomo Buechel, E.R., Mertens, L.L.: Imaging the right heart: the use of integrated multimodality imaging. European Heart Journal 33(8), 949–960 (2012)
7. Carlsson, M., Ugander, M., Mosén, H., Buhre, T., Arheden, H.: Atrioventricular plane displacement is the major contributor to left ventricular pumping in healthy adults, athletes, and patients with dilated cardiomyopathy. American Journal of Physiology-Heart and Circulatory Physiology 292(3), H1452–H1459 (2007)
8. Seemann, F., Heiberg, E., Carlsson, M., Gonzales, R.A., Baldassarre, L.A., Qiu, M., Peters, D.C.: Valvular imaging in the era of feature-tracking: A slice-following cardiac MR sequence to measure mitral flow. Journal of Magnetic Resonance Imaging 51(5), 1412–1421 (2020)
9. Caudron, J., Fares, J., Vivier, P.H., Lefebvre, V., Petitjean, C., Dacher, J.N.: Diagnostic accuracy and variability of three semi-quantitative methods for assessing right ventricular systolic function from cardiac MRI in patients with acquired heart disease. European Radiology 21(10), 2111–2120 (2011)

10. Leng, S., Jiang, M., Zhao, X.D., Allen, J.C., Kassab, G.S., Ouyang, R.Z., Tan, J.L., He, B., Tan, R.S., Zhong, L.: Three-dimensional tricuspid annular motion analysis from cardiac magnetic resonance feature-tracking. Annals of Biomedical Engineering 44(12), 3522–3538 (2016)
11. Seemann, F., Pahlm, U., Steding-Ehrenborg, K., Ostenfeld, E., Erlinge, D., Dubois-Rande, J.L., Jensen, S.E., Atar, D., Arheden, H., Carlsson, M., et al.: Time-resolved tracking of the atrioventricular plane displacement in cardiovascular magnetic resonance (CMR) images. BMC Medical Imaging 17(1), 19 (2017)
12. Hu, C., Sinusas, A.J., Huber, S., Thorn, S., Stacy, M.R., Mojibian, H., Peters, D.C.: T1-refBlochi: high resolution 3D post-contrast T1 myocardial mapping based on a single 3D late gadolinium enhancement volume, Bloch equations, and a reference T1. Journal of Cardiovascular Magnetic Resonance 19(63), 1–17 (2017)
13. Seemann, F., Baldassarre, L.A., Llanos-Chea, F., Gonzales, R.A., Grunseich, K., Hu, C., Sugeng, L., Meadows, J., Heiberg, E., Peters, D.C.: Assessment of diastolic function and atrial remodeling by MRI-validation and correlation with echocardiography and filling pressure. Physiological Reports 6(17), e13828 (2018)
14. Gonzales, R.A., Seemann, F., Lamy, J., Arvidsson, P.M., Heiberg, E., Murray, V., Peters, D.C.: Automated left atrial time-resolved segmentation in MRI long-axis cine images using active contours. BMC Medical Imaging 21(1), 1–12 (2021)
15. Heiberg, E., Sjögren, J., Ugander, M., Carlsson, M., Engblom, H., Arheden, H.: Design and validation of segment-freely available software for cardiovascular image analysis. BMC Medical Imaging 10(1), 1 (2010)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
17. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255. IEEE (2009)
18. Kingma, D.P., Ba, J.L.: Adam: A method for stochastic gradient descent. In: ICLR: International Conference on Learning Representations (2015)