Unsupervised Mitral Valve Segmentation in Echocardiography with Neural Network Matrix Factorization

Luca Corinzia[1(✉)], Jesse Provost[1], Alessandro Candreva[2], Maurizio Tamarasso[2], Francesco Maisano[2], and Joachim M. Buhmann[1]

[1] Institute for Machine Learning, ETH Zurich, Zurich, Switzerland
{luca.corinzia,jbuhmann}@inf.ethz.ch, jprovost@student.ethz.ch
[2] Department of Cardiology, University Hospital Zurich, Zurich, Switzerland
{alessandro.candreva,maurizio.taramasso,francesco.maisano}@usz.ch

Abstract. Mitral valve segmentation specifies a crucial first step to establish a machine learning pipeline that can support practitioners into performing diagnosis of mitral valve diseases, surgical planning, and intraoperative procedures. To this end, we propose a totally automated and unsupervised mitral valve segmentation algorithm, based on a low-dimensional neural network matrix factorization of echocardiography videos. The method is evaluated in a collection of echocardiography videos of patients with a variety of mitral valve diseases and exceeds the state-of-the-art method in all the metrics considered.

Keywords: Mitral valve segmentation · Echocardiography · Neural network matrix factorization

1 Introduction

The mitral valve (MV) of the heart regulates the blood flow between two heart chambers, namely the left atrium and the left ventricle. It is formed by two leaflets, the anterior and the posterior leaflet, that are attached to a fibrous ring known as the mitral annulus. In healthy patients, the left atrium contracts during diastole and the blood flows through the open MV into the left ventricle that is dilating. During systole the left ventricle contracts and pushes the blood into the aorta through the aortic valve, and the MV closes so that the blood does not flow back into the atrium. Various diseases concern the MV causing an alteration of healthy blood flow between left atrium and left ventricle. Briefly, two possible scenarios are possible: (i) *mitral stenosis* that is characterized by a narrowing of the mitral annulus and hence a decline of the blood flow to the left ventricle and (ii) *mitral regurgitation* that causes blood to flow back to the left atrium during systole. This last condition defines the second most common cardiac valvular defect amenable of surgical intervention [7].

L. Corinzia and J. Provost contributed equally to the work.

D. Riaño et al. (Eds.): AIME 2019, LNAI 11526, pp. 410–419, 2019.
https://doi.org/10.1007/978-3-030-21642-9_51

Echocardiography (echo) is a medical imaging technique that produces 2D and 3D pictures and videos using ultrasound waves generated by a transducer, scattered/reflected by biological tissue and read by a detector. Echo is the standard imaging tool in the clinical routine to perform the diagnosis of most of heart diseases and dysfunctions, including MV diseases [1,7,15] since it is inexpensive, non-invasive and it enables both qualitative and quantitative assessment of the myocardium and of the MV functions. The current clinical protocol requires practitioners to manually measure a plethora of diagnostic parameters of the cardiac valves as well as of the cardiac chambers. In this paper we propose NN-MitralSeg, an unsupervised MV segmentation algorithm that supports a systematic and fast evaluation of MV health status for the medical practitioners. Our method improves on the Robust Non Negative Matrix Factorization method (R-NNMF) proposed in [4] and it outperforms R-NNMF on a dataset of 38 patients affected with MV dysfunction and mitral regurgitation.

2 Related Work

MV segmentation in 2D and 3D echo enables automatic diagnosis and personalized prognosis and, therefore, it has received a lot of attention recently. Many early methods are based on active contour algorithms or on other methods that depend heavily on human-in-the-loop contributions. Active contour algorithms [2,9] require practitioners to initialize the segmentation algorithm, placing manually a contour close to the desired position in a given frame or on multiple frames [10,13]. Then the MV is segmented on the given frames optimizing a fixed energy function, and the mask is propagated over time with the support of the optical flow [11] and/or of a dynamical model of the MV [12]. In [3] the authors proposed a method that leverages both an active contour algorithm that segments the myocardial walls and a thin tissue detector that finds the valve leaflets. Also in [14] medical practitioners initialize the segmentation denoting multiple points that are then connected using J-splines.

The first attempts to design a fully automated MV segmentation algorithm are proposed in [4,16]. The 2D echo video is factorized using (robust) non-negative 2-rank matrix factorization. Every frame of the video is decomposed as a non-negative linear mixture of two frames and a sparse signal, where the frames, the mixture coefficients and the sparse signal are obtained optimizing an l_2 loss. The 2-rank factorization captures most of the myocardium wall motion, while the high dimensional sparse signal represents the echo noise and the MV movement. Then the MV is segmented using simple diffusion and thresholding of the sparse signal. Despite producing satisfactory results on high quality echos, these methods performs poorly on noisy low quality videos, due mostly to the misplacement of the region of interest (ROI).

We propose Neural Network Mitral Segmentation (NN-MitralSeg)[1], a method that improves on [4] with a two-fold contribution: (i) we use a neural network

[1] Code, hyperparameters and network specifications are available at https://github.com/jprovost14/NN-MitralSeg.

matrix factorization [5] (also know as neural collaborative filtering in [8]) to account for both linear and non-linear contributions of the myocardial wall motion, in combination with a parametrized threshold operator to learn the high dimensional sparse signal that captures the MV, and (ii) we leverage the information of both the sparse signal and of the dense optical flow to delineate the ROI.

3 Method

3.1 Model

Each echo is initially represented as a tensor $\mathbf{T} \in \mathbb{R}_+^{h \times w \times T}$, where h and w are respectively the height and the width of a single frame and T is the number of frames in the video. We reshape each frame of the echo into a column vector and then concatenate all the columns to get a matrix $\mathbf{X} \in \mathbb{R}_+^{N \times T}$ where $N = h \cdot w$. The matrix $\mathbf{X}$ is then embedded in a low dimensional space as follows. For each row (pixel) $n \in N$ and each column (frame) $t \in T$, we associate the latent feature matrices with non-negative entries $\mathbf{U}_n$, $\mathbf{V}_t, \in \mathbb{R}_+^{D \times K}$, where D and K are two hyperparameters of the model. Let $f_{\theta_{LD}}$ denote the *low dimensional* network with weights θ_{LD} and f_{θ_T} denote the *threshold* network with weights θ_T. The low dimensional network reconstructs the inputs as

$$\hat{X}_{n,t} = f_{\theta_{LD}}(\mathbf{u}_{n,1} \cdot \mathbf{v}_{t,1}, \ldots, \mathbf{u}_{n,D} \cdot \mathbf{v}_{t,D}).$$

where $\mathbf{u}_{n,j} \cdot \mathbf{v}_{t,j}$ is the inner product between the j-th row vectors of $\mathbf{U}_n$ and $\mathbf{V}_t$. It is easy to see that the input of the $f_{\theta_{LD}}$ is equivalent to the diagonal of the product matrix $\mathbf{U}_n \mathbf{V}_t^T$, hence it is a D-dimensional latent feature vector. Notice that K-rank non-negative matrix factorization is obtained enforcing $f_{\theta_{LD}} = \mathbf{1}$ and $D = 1$, where $\mathbf{1}$ is the identity function (see [8]), hence K can be interpreted as the generalized rank of the factorization. The non-negativity of the latent features is imposed using non-negative activation functions. Given the reconstruction $\hat{X}_{n,t}$, the difference between $X_{n,t}$ and $\hat{X}_{n,t}$ serves as the scalar input to the threshold network and is transformed to get the scalar output:

$$\hat{S}_{n,t} = f_{\theta_T}(X_{n,t} - \hat{X}_{n,t}).$$

The threshold network is composed by just one node and a ReLu activation function and acts as a parametrized threshold operator. A diagram for the general architecture is given in Fig. 1.

3.2 Learning

The training of this model occurs in three stages. In the first stage the *low-dimensional* network is trained to provide an accurate approximation $\hat{\mathbf{X}}$. Subsequently, both the *low-dimensional* network and the *threshold* network are trained iteratively such that the threshold network fully reconstructs the error $\mathbf{X} - \hat{\mathbf{X}}$. The final stage consists of imposing the sparseness on $\hat{\mathbf{S}}$ using a ℓ_1 regularizer.

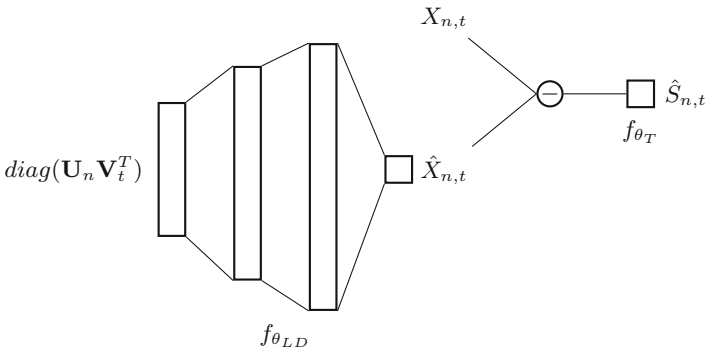


Fig. 1. Diagram of the model used in NN-MitralSeg. The network $f_{\theta_{LD}}$ maps the pixel and frame latent features $\mathbf{U}_n$, $\mathbf{V}_t$ to the reconstruction $\hat{X}_{n,t}$. The input of the network is the diagonal of $\mathbf{U}_n\mathbf{V}_t^T$ denoted as $diag(\mathbf{U}_n\mathbf{V}_t^T)$. The threshold operator f_{θ_T} is then applied to $\hat{X}_{n,t} - X_{n,t}$ to give the sparse signal $\hat{S}_{n,t}$.

Pre-training the Low-Dimensional and Threshold Networks. Pre-training the parameters θ_{LD} and $\{(\mathbf{U}_n, \mathbf{V}_t)\}_{n,t}$ of the low-dimensional network ensures that the network can produce an accurate approximation of $\hat{\mathbf{X}}$, which is used as input into the threshold operator. The pre-training of the low-dimensional network is performed as in [5]; freezing the latent (pixel and frame) features $\{(\mathbf{U}_n, \mathbf{V}_t)\}_{n,t}$ while updating θ_{LD}, and then freezing the low dimensional network θ_{LD} while updating $\{(\mathbf{U}_n, \mathbf{V}_t)\}_{n,t}$. The objective that is optimized during this stage is given by:

$$\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 + \beta \left[\sum_n \|\mathbf{U}_n\|_F^2 + \sum_t \|\mathbf{V}_t\|_F^2 \right],$$

where β is a regularization parameter and $\|\cdot\|_F$ is the Frobenius norm. In the second stage also the threshold network is trained in an iterative fashion: updating θ_T while freezing θ_{LD} and $\{(\mathbf{U}_n, \mathbf{V}_t)\}_{n,t}$; then updating θ_{LD} and $\{(\mathbf{U}_n, \mathbf{V}_t)\}_{n,t}$ as described above while freezing θ_T according to the loss function given by:

$$\|\mathbf{X} - \hat{\mathbf{X}} - \hat{\mathbf{S}}\|_F^2 + \beta \left[\sum_n \|\mathbf{U}_n\|_F^2 + \sum_t \|\mathbf{V}_t\|_F^2 \right].$$

Training on the Full Objective. The goal of pre-training is to obtain two networks that can fully reconstruct the echo. The low-dimensional network captures the myocardium movement and the threshold operator captures the echo noise and the mitral valve movement. Sparsity is enforced by regularizing the loss function with the ℓ_1-norm:

$$\|\mathbf{X} - \hat{\mathbf{X}} - \hat{\mathbf{S}}\|_F^2 + \beta \left[\sum_n \|\mathbf{U}_n\|_F^2 + \sum_t \|\mathbf{V}_t\|_F^2 \right] + \lambda \|\hat{\mathbf{S}}\|_1,$$

where λ is the sparsity coefficient and $\|\cdot\|_1$ denotes the ℓ_1-norm.

3.3 Mitral Valve Window Detection and Segmentation

The sparse signal $\hat{\mathbf{S}}$ captures the motion of the mitral valve. In [4] the authors compute the Frobenius norm on all possible 3D windows of the sparse matrix $\hat{\mathbf{S}}$ and define the MV ROI as the window with the maximum Frobenius norm. However, it often occurs that part of the myocardium movement is also captured in the sparse signal due to low quality of the echos and then the ROI does not contain the mitral valve or it captures it only partially.

We propose an alternative method for MV window detection that leverages also movement information. The motion of the MV is much faster compared to the myocardium, even when the myocardium appears in the sparse signal. The norm of the dense optical flow [6] can measure the motion in a video and a large norm is indicative of fast motion. First the sparse signal $\hat{\mathbf{S}} \in \mathbb{R}_+^{n\times\times T}$ is reshaped into a 3D array of the same shape of the original video $\mathbb{R}_+^{h\times w\times T}$ and then thresholding is applied in order to retain only the p percent high intensity pixels. The dense optical flow is then computed for every frame of $\hat{\mathbf{S}}$ and is denoted as $optical\_flow(\hat{\mathbf{S}})_t$. Similar to the window detection method in [4,16], the ROI of the MV is then identified as the window with largest sum among the frames of the optical flow norms. The selection is made between windows spanning the whole 2D frame, with a fixed stride. Denoting by $\mathbf{W}_l \in \{0,1\}^{w\times h}$ the windows as binary masks, the ROI selection can be summarized as

$$l^* = \arg\max_l \sum_{t=1}^{T} \|optical\_flow(\hat{\mathbf{S}})_t \cdot \mathbf{W}_l\|_2^2$$

The segmentation is consequently performed on the sparse signal enclosed in the ROI similarly to [4] using simple isotropic 2D diffusion on each frame.

4 Experiments and Results

4.1 Dataset Description

A total of 38 transthoracic echos were obtained from the MitraSwiss Registry, a Swiss-wide prospective registry which includes patients undergoing percutaneous mitral valve repair using the MitraClip system. All patients had moderate-to-severe (3+) or severe (4+) mitral regurgitation of functional or degenerative origin as graded according to current recommendations of the American Society of Echocardiography [18]. Imaging data were processed in an anonymized way and all patients provided written informed consent to be entered into the database. Only 4-chamber echo views are used, and for every echo, a rectangular window around the MV and three selected frames were annotated by an expert medical doctor.

4.2 Window Detection

A comparison of the sparse signal according to R-NNMF [4] and our method NN-MitralSeg is showed in Fig. 2 for a R-NNMF failure case. As it can be seen

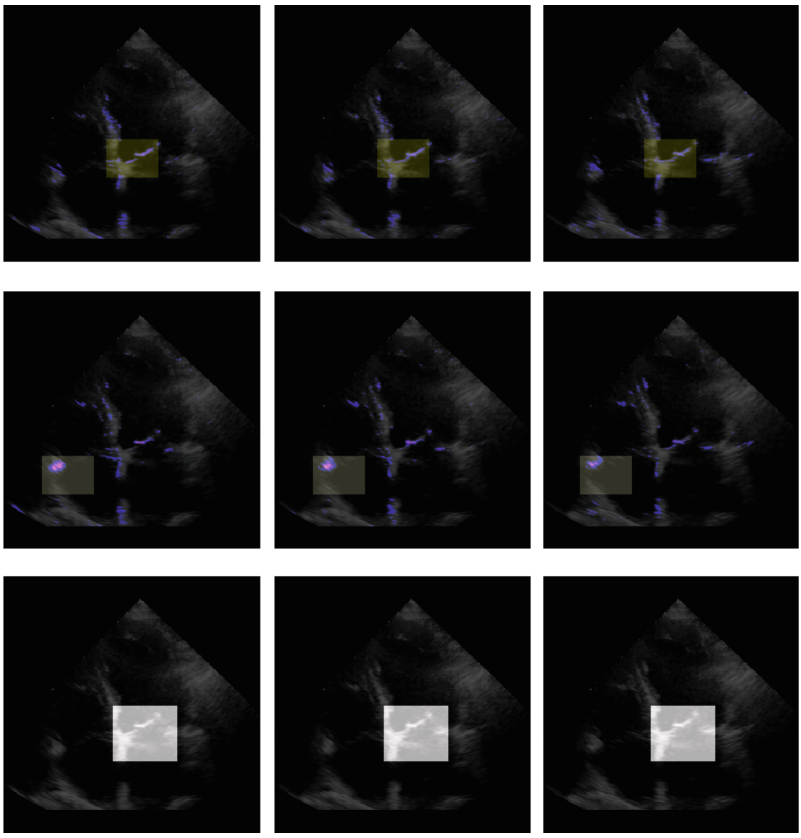


Fig. 2. A failure case for the window detection method of R-NNMF [4]. The sparse signal (in blue) is given for both our method (NN-MitralSeg, top row) and R-NNMF (middle row) with reference to the original frames (bottom row) for three consecutive frames. The mitral valve region is always highlighted as the shaded area. The region is misplaced by R-NNMF due to a strong myocardium movement contribution in the sparse signal. (Color figure online)

the failures of the R-NNMF window detection method are due to a strong presence of the myocardium movement in the sparse signal, as a consequence of the low expressiveness of the linear model used in R-NNMF. We compare the performance of the mitral valve window detection according to accuracy (I), defined as the percentage of pixels in the computed ROI that intersect the gold standard window. Note that in this specific task, the window sizes are fixed and not inferred by the model, hence the accuracy is a reliable measure of performance. I is sorted in descending order according to our method in Fig. 3a. In Fig. 3b the difference between the accuracies of our method and of R-NNMF I_{diff} is sorted in ascending order, alongside the average I_{diff} over all echos μ and the p-value of the one-sided t-test. In Table 1 we also report the number of success

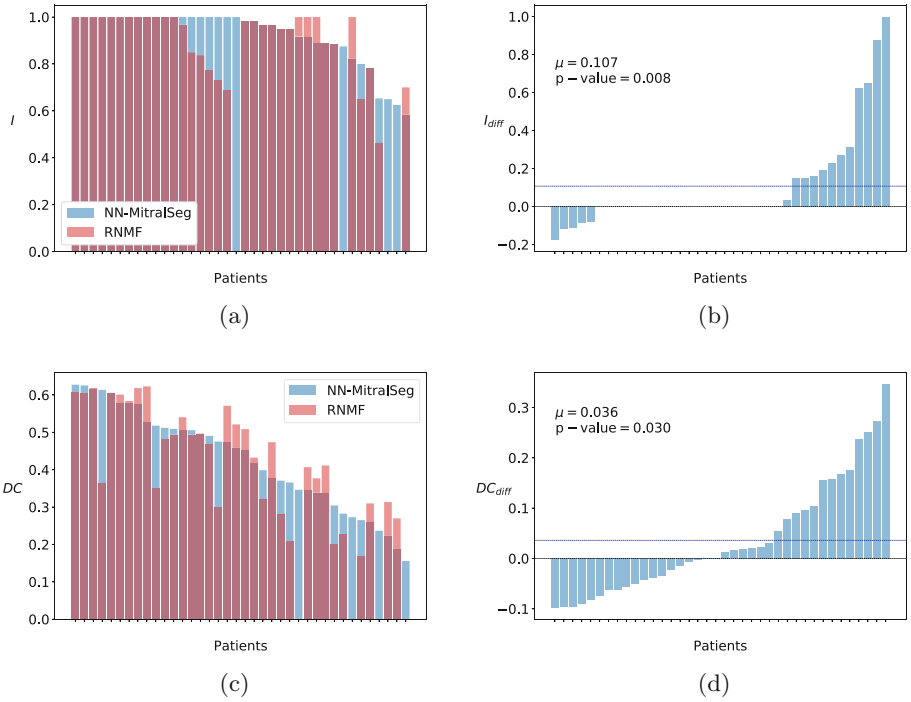
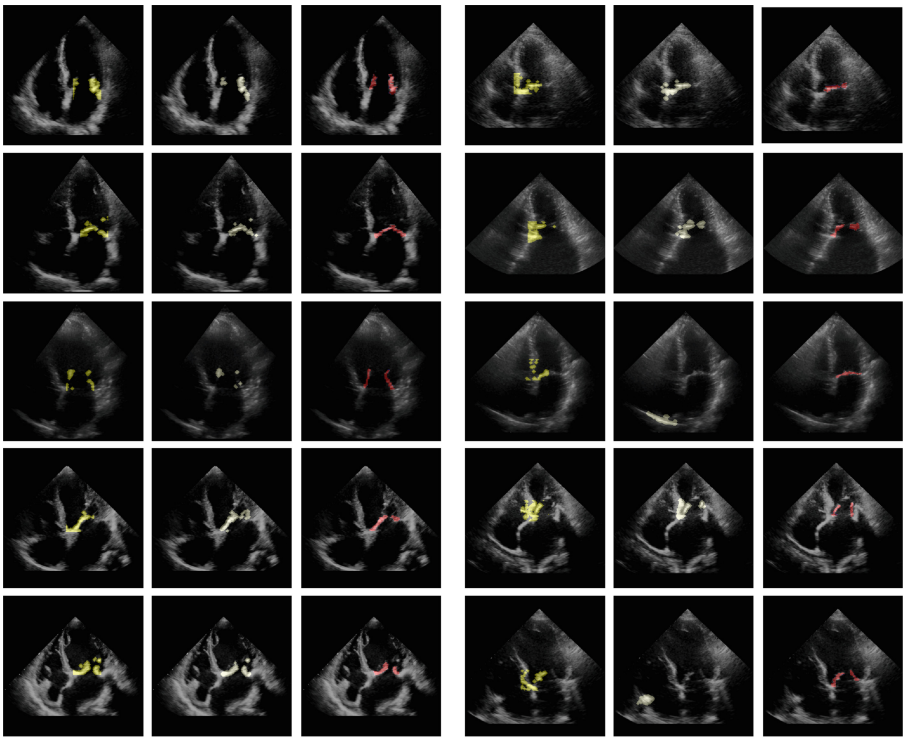


Fig. 3. A comparison of our method NN-MitralSeg and the R-NNMF [4] according to window detection accuracies I (a-b) and Dice coefficient (c-d). (a) and (c) show respectively the accuracy and the Dice coefficient sorted in descending order by our method. (b) and (d) show respectively the difference of accuracies I_{diff} and the difference of dice coefficients DC_{diff} between NN-MitralSeg and R-NNMF, sorted in ascending order.

cases where the accuracy I is higher then a given threshold, and the average Intersection over Union score (IoU).

4.3 Mitral Valve Segmentation

The output of the segmentation algorithms are compared with the ground truth in Figs. 3a and d according to the Dice coefficient (DC). The DC is reported for every echo and it is sorted in descending order according to the score of our method. The DC difference DC_{diff} between the two methods is also reported in Fig. 3d sorted in ascending order. We observe that NN-MitralSeg outperforms the state-of-the-art in both window detection and in the dense MV annotation by a statistically significant margin. A detailed comparison of the MV segmentations produced by the two algorithms is documented in Figs. 4a and b where we show the masks and the ground truth respectively for the highest and lowest five scoring echos (according to our method).



(a) (b)

Fig. 4. The mitral valve segmentation masks for the echos with the (a) five highest and (b) lowest Dice coefficients according to NN-MitralSeg. From left to right: NN-MitralSeg (yellow), R-NNMF (green) and ground truth (red). (Color figure online)

Table 1. Number of success cases and average Intersection over Union score for the window detection algorithm. The total number of echo is 38.

	$I > 0.65$	$I > 0.85$	IoU
NN-MitralSeg (Ours)	**35**	**31**	**0.35132**
R-NNMF [4]	32	25	0.30883

5 Conclusion and Future Work

We proposed NN-MitralSeg, a fully automated and unsupervised mitral valve segmentation algorithm based on non-linear matrix factorization. An echocardiography video is decomposed into a low dimensional signal that captures the linear and non-linear myocardial wall motion, and a high dimensional sparse signal that accounts for the echocardiography noise and mitral valve movement.

The mitral valve is then segmented from the sparse signal using thresholding and diffusion algorithms. This method outperforms the state-of-the-art fully automated algorithm in a data-set of 38 videos with patients suffering various mitral valve dysfunctions, in both the task of positioning the rectangular region of interest, and in the accuracy of the dense mitral valve mask. Despite being a small dataset, due to the health status of the patients it contains a larger variability then an healthy control dataset of the same size. A possible future development includes the use of both the sparse ground truth segmentation masks and the dense (inaccurate) annotation generated by unsupervised algorithms (like NN-MitralSeg) to train segmentation deep networks in a weakly-supervised-learning scenario [17] in an online fashion. This would provide practitioners with segmentation algorithms that could be deployed in the real-time echocardiography during mitral valve intraoperative procedures.

References

1. Baumgartner, H., et al.: Echocardiographic assessment of valve stenosis: EAE/ASE recommendations for clinical practice. J. Am. Soc. Echocardiogr. **22**(1), 1–23 (2009)
2. Blake, A., Isard, M.: Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion. Springer, London (2012). https://doi.org/10.1007/978-1-4471-1555-7
3. Burlina, P., et al.: Patient-specific modeling and analysis of the mitral valve using 3D-TEE. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 135–146. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13711-2_13
4. Dukler, Y., et al.: Automatic valve segmentation in cardiac ultrasound time series data. In: Medical Imaging 2018: Image Processing, vol. 10574, p. 105741Y. International Society for Optics and Photonics (2018)
5. Dziugaite, G.K., Roy, D.M.: Neural network matrix factorization. arXiv preprint arXiv:1511.06443 (2015)
6. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Bigun, J., Gustavsson, T. (eds.) SCIA 2003. LNCS, vol. 2749, pp. 363–370. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45103-X_50. http://dl.acm.org/citation.cfm?id=1763974.1764031
7. Hayek, E., Gring, C.N., Griffin, B.P.: Mitral valve prolapse. Lancet **365**(9458), 507–518 (2005)
8. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.S.: Neural collaborative filtering. In: Proceedings of the 26th International Conference on World Wide Web, pp. 173–182. International World Wide Web Conferences Steering Committee (2017)
9. Isard, M., Blake, A.: Contour tracking by stochastic propagation of conditional density. In: Buxton, B., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1064, pp. 343–356. Springer, Heidelberg (1996). https://doi.org/10.1007/BFb0015549
10. Mikic, I., Krucinski, S., Thomas, J.D.: Segmentation and tracking of mitral valve leaflets in echocardiographic sequences: active contours guided by optical flow estimates. In: Medical Imaging 1996: Image Processing, vol. 2710, pp. 311–321. International Society for Optics and Photonics (1996)

11. Mikic, I., Krucinski, S., Thomas, J.D.: Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow estimates. IEEE Trans. Med. Imaging **17**(2), 274–284 (1998)
12. Schneider, R.J., Tenenholtz, N.A., Perrin, D.P., Marx, G.R., del Nido, P.J., Howe, R.D.: Patient-specific mitral leaflet segmentation from 4D ultrasound. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6893, pp. 520–527. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23626-6_64
13. Shang, Y., Yang, X., Zhu, L., Deklerck, R., Nyssen, E.: Region competition based active contour for medical object extraction. Comput. Med. Imaging Graph. **32**(2), 109–117 (2008)
14. Siefert, A.W., et al.: Accuracy of a mitral valve segmentation method using J-splines for real-time 3D echocardiography data. Ann. Biomed. Eng. **41**(6), 1258–1268 (2013)
15. Zamorano, J., et al.: Real-time three-dimensional echocardiography for rheumatic mitral valve stenosis evaluation: an accurate and novel approach. J. Am. Coll. Cardiol. **43**(11), 2091–2096 (2004)
16. Zhou, X., Yang, C., Yu, W.: Automatic mitral leaflet tracking in echocardiography by outlier detection in the low-rank representation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 972–979. IEEE (2012)
17. Zhou, Z.H.: A brief introduction to weakly supervised learning. Natl. Sci. Rev. **5**(1), 44–53 (2017)
18. Zoghbi, W.A., et al.: Recommendations for evaluation of the severity of native valvular regurgitation with two-dimensional and doppler echocardiography. J. Am. Soc. Echocardiogr. **16**(7), 777–802 (2003)