



Unsupervised mitral valve segmentation in  
echocardiography with neural network matrix  
factorization.

---

Luca Corinzia, Jesse Provost, Alessandro Candreva,  
Maurizio Tamarasso, Francesco Maisano and  
Joachim M. Buhmann

EasyChair preprints are intended for rapid  
dissemination of research results and are  
integrated with the rest of EasyChair.

February 5, 2019

# Unsupervised mitral valve segmentation in echocardiography with neural network matrix factorization

Luca Corinzia<sup>1\*</sup>, Jesse Provost<sup>1\*</sup>, Alessandro Candreva<sup>2</sup>, Maurizio Tamarasso<sup>2</sup>, Francesco Maisano<sup>2</sup>, and Joachim M. Buhmann<sup>1</sup>

<sup>1</sup> ETH Zurich, Institute for Machine Learning, Zurich, Switzerland

{luca.corinzia,jbuhmann}@inf.ethz.ch jprovost@student.ethz.ch

<sup>2</sup> University Hospital Zurich, Department of Cardiology, 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 the 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) is a valve of the heart that 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 the healthy blood flow between left atrium and left ventricle. Briefly, two possible scenarios are possible. i) In case of a reduction of the valve surface or leaflets mobility, a decline of the blood flow to the left ventricle occurs, with an increase of the pressure on the MV. This condition is better known as *mitral stenosis* and it is most often secondary to inflammatory conditions, such rheumatic cardiac disease. ii) Conversely, if the integrity of the coaptation line between the two leaflets is affected, blood can flow back to the

---

\* These authors contributed equally to the work

left atrium during systole in what is called *mitral regurgitation*. This condition defines the second most common cardiac valvular defect amenable of surgical intervention and might be caused by inborn abnormalities of the valvular tissue, geometric disparities between the left ventricle and the MV, various connective tissue disorders and valve infections [6].

Echocardiography (echo), a minimally invasive medical imaging technique, 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 [14,1,6] since echo 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 [8,2] require practitioners to initialize the segmentation algorithm, placing manually a contour close to the desired position in a given frame or on multiple frames [9,12]. 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 [10] and/or of a dynamical model of the MV [11]. 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 [13] 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 [15,4]. 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. The low-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)<sup>3</sup>, a method that improves on [4] with a two-fold contribution: (i) we use a neural network matrix factorization [5] (also known as neural collaborative filtering in [7]) 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 concatenate each column to get a matrix  $\mathbf{X} \in \mathbb{R}_+^{N \times T}$  where  $N = h \cdot w$ . Given the matrix  $\mathbf{X}$  we build a low dimensional embedding of it 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  $\mathbf{U}_n, \mathbf{V}_t \in \mathbb{R}_+^{D \times K}$ . Let  $f_{\theta_{LD}}$  denote the *low dimensional* network with weights  $\theta_{LD}$  and  $f_{\theta_T}$  denote the *threshold* network with weights  $\theta_T$ . The low dimensional network reconstructs the inputs as

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

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  $\text{diag}(\mathbf{U}_n \mathbf{V}_t^T)$ , 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 [7]). The non-negativity of the latent features is imposed using a non-negative activation function. 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 a single 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 we train the *low-dimensional* network to provide an accurate approximation  $\hat{\mathbf{X}}$ . Subsequently, we train both the *low-dimensional* network and the *threshold* network

<sup>3</sup> code is available at <https://github.com/jprovost14/NN-MitralSeg>

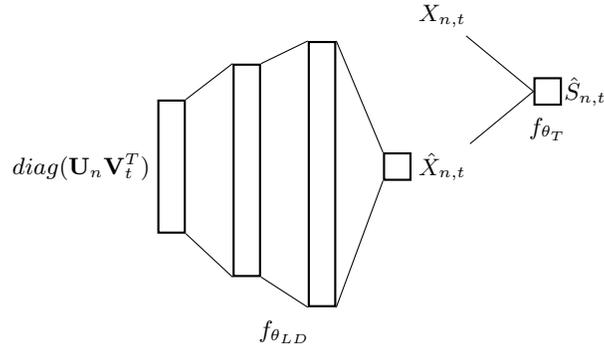


Fig. 1: Diagram of the model used in NN-MitralSeg. The network  $f_{\theta_{LD}}$  maps the pixel and frame latent features  $\mathbf{U}_n, \mathbf{V}_t$  to the reconstruction  $\hat{\mathbf{X}}_{n,t}$ . Then the threshold operator  $f_{\theta_T}$  is applied to the difference  $\hat{\mathbf{X}}_{n,t} - X_{n,t}$  to give the sparse signal  $\hat{\mathbf{S}}_{n,t}$ .

iteratively such that the threshold network fully reconstructs  $\mathbf{X} - \hat{\mathbf{X}}$ . The final stage of the training consists of imposing the sparse structure on the  $\hat{\mathbf{S}}$  using a  $\ell_1$ -norm regularizer.

**Pre-training the Low-Dimensional and Threshold Networks.** Pre-training the parameters  $\theta_{LD}$  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 done in the same way as in [5]; freezing the latent (pixel and frame) features  $\{(\mathbf{U}_n, \mathbf{V}_t)\}_{n,t}$  while updating  $\theta_{LD}$ , and then freezing the low dimensional network  $\theta_{LD}$  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  $\beta$  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: update  $\theta_T$  while freezing  $\theta_{LD}$  and  $\{(\mathbf{U}_n, \mathbf{V}_t)\}_{n,t}$ ; then updating  $\theta_{LD}$  and  $\{(\mathbf{U}_n, \mathbf{V}_t)\}_{n,t}$  as described above while freezing  $\theta_T$  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  $\ell_1$ -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  $\lambda$  is the sparsity coefficient and  $\|\cdot\|_1$  denotes the  $\ell_1$ -norm. Network specifications and other details of the learning procedures are given in appendix B.

### 3.3 Mitral Valve Window Detection and Segmentation

The sparse matrix  $\hat{\mathbf{S}}$  captures the motion of the mitral valve. In [4] the authors compute the Frobenius norm on all possible 3D window 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 matrix 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 matrix. The norm of the dense optical flow can measure the motion in a video and a large norm is indicative of fast motion. First the sparse signal  $\hat{\mathbf{S}}$  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 [4,15], 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 window as a binary mask, the ROI selection can be summarized as

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

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 [17]. 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 box around the MV and three selected frames were densely 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 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 perfor-

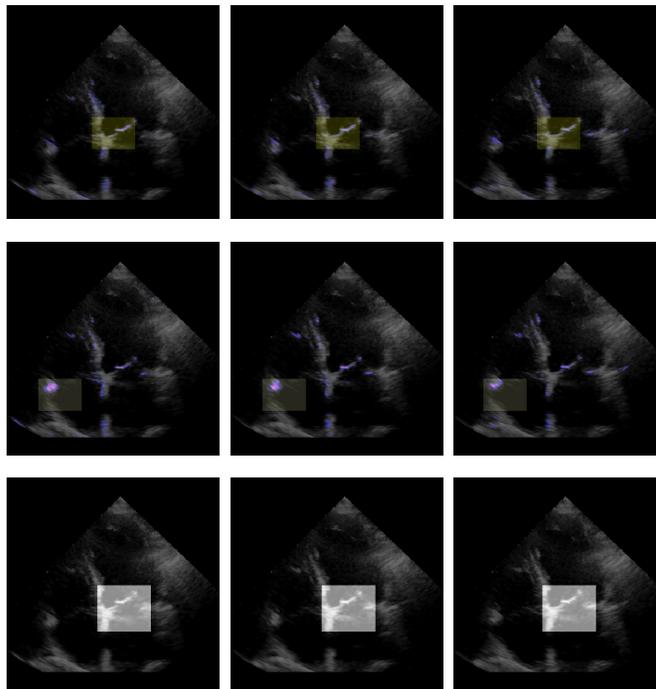


Fig. 2: A failure case for the window detection method of R-NNMF [4]. The sparse signal (in blue) is given for both our method (NN-MitralSeg, top row) and R-NNMF (middle row) with reference to the original frames (bottom row) for three consecutive frames. The mitral valve region is always highlighted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

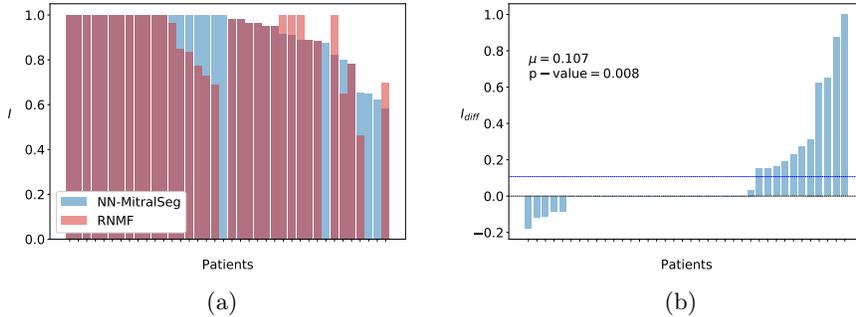


Fig. 3: A comparison of window detection accuracies  $I$  between our method and the R-NNMF [4]. (a) shows the accuracies sorted in descending order by our method. (b) shows the difference of accuracies  $I_{diff}$  between our method and R-NNMF sorted in ascending order.

mance of the mitral valve window detection according to the percentage of pixels in the computed ROI that intersect with the ground truth window. The window detection accuracy ( $I$ ) is sorted in descending order according to our method in fig. 3a. In fig. 3b the difference between the window detection accuracies of our method and of R-NNMF is sorted in ascending order, alongside the average accuracy difference over all echos  $\mu$  and the p-value of the one-sided t-test. In table 1 we also report the number of success cases where the accuracy reached by the window detection algorithm is higher than a given threshold, and the Intersection over Union score (IoU) averaged over all echos.

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

	Accuracy > 0.65	Accuracy > 0.85	IoU
<b>NN-MitralSeg</b>	<b>35</b>	<b>31</b>	<b>0.35132</b>
<b>R-NNMF [4]</b>	32	25	0.30883

### 4.3 Mitral Valve segmentation

The output of the segmentation algorithms are compared with the ground truth in fig. 4 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. 4b 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 appendix A. In particular figs. 5a and 5b show the masks compared to the ground truth respectively for the highest and lowest five scoring echos (according to our method) and fig. 6 shows the time coherence of the masks.

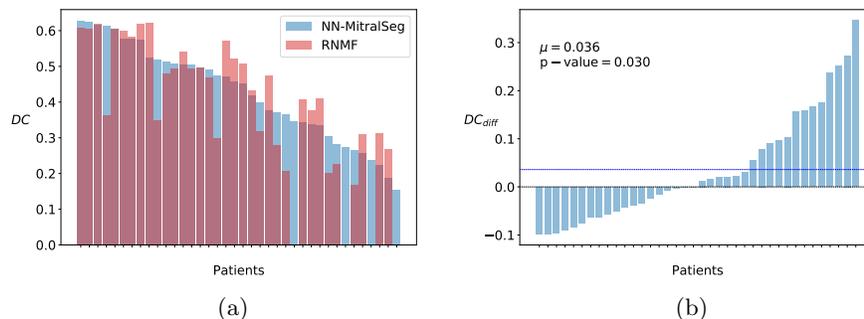


Fig. 4: A comparison of Dice coefficient measures between our NN-MitralSeg method and the R-NNMF method in [4]. (a) shows the Dice coefficients sorted in descending order according to our method. (b) shows the difference of the Dice coefficients between NN-MitralSeg and R-NNMF sorted in ascending order.

## 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 outperform 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.

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 [16]. This concept would also provide practitioners with an online segmentation algorithm that could be deployed in the real-time echocardiography during mitral valve intraoperative procedures.

## References

1. Baumgartner, H., Hung, J., Bermejo, J., Chambers, J.B., Evangelista, A., Griffin, B.P., Iung, B., Otto, C.M., Pellikka, P.A., Quiñones, M.: Echocardiographic assessment of valve stenosis: Eae/ase recommendations for clinical practice. *Journal of the American Society of Echocardiography* **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 Science & Business Media (2012)
3. Burlina, P., Sprouse, C., DeMenthon, D., Jorstad, A., Juang, R., Contijoch, F., Abraham, T., Yuh, D., McVeigh, E.: Patient-specific modeling and analysis of the mitral valve using 3d-tee. In: *International Conference on Information Processing in Computer-Assisted Interventions*. pp. 135–146. Springer (2010)
4. Dukler, Y., Ge, Y., Qian, Y., Yamamoto, S., Yuan, B., Zhao, L., Bertozzi, A.L., Hunter, B., Llerena, R., Yen, J.T.: 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. Hayek, E., Gring, C.N., Griffin, B.P.: Mitral valve prolapse. *The Lancet* **365**(9458), 507–518 (2005)
7. 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)
8. Isard, M., Blake, A.: Contour tracking by stochastic propagation of conditional density. In: *European conference on computer vision*. pp. 343–356. Springer (1996)
9. 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)
10. Mikic, I., Krucinski, S., Thomas, J.D.: Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates. *IEEE transactions on medical imaging* **17**(2), 274–284 (1998)
11. Schneider, R.J., Tenenholtz, N.A., Perrin, D.P., Marx, G.R., Pedro, J., Howe, R.D.: Patient-specific mitral leaflet segmentation from 4d ultrasound. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 520–527. Springer (2011)
12. Shang, Y., Yang, X., Zhu, L., Deklerck, R., Nyssen, E.: Region competition based active contour for medical object extraction. *Computerized Medical Imaging and Graphics* **32**(2), 109–117 (2008)
13. Siefert, A.W., Icenogle, D.A., Rabbah, J.P.M., Saikrishnan, N., Rossignac, J., Leraakis, S., Yoganathan, A.P.: Accuracy of a mitral valve segmentation method using j-splines for real-time 3d echocardiography data. *Annals of biomedical engineering* **41**(6), 1258–1268 (2013)
14. Zamorano, J., Cordeiro, P., Sugeng, L., de Isla, L.P., Weinert, L., Macaya, C., Rodríguez, E., Lang, R.M.: Real-time three-dimensional echocardiography for rheumatic mitral valve stenosis evaluation: an accurate and novel approach. *Journal of the American College of Cardiology* **43**(11), 2091–2096 (2004)

15. Zhou, X., Yang, C., Yu, W.: Automatic mitral leaflet tracking in echocardiography by outlier detection in the low-rank representation. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 972–979. IEEE (2012)
16. Zhou, Z.H.: A brief introduction to weakly supervised learning. National Science Review **5**(1), 44–53 (2017)
17. Zoghbi, W.A., Enriquez-Sarano, M., Foster, E., Grayburn, P.A., Kraft, C.D., Levine, R.A., Nihoyannopoulos, P., Otto, C.M., Quinones, M.A., Rakowski, H., et al.: Recommendations for evaluation of the severity of native valvular regurgitation with two-dimensional and doppler echocardiography. Journal of the American Society of Echocardiography **16**(7), 777–802 (2003)

## A Mitral valve masks comparison

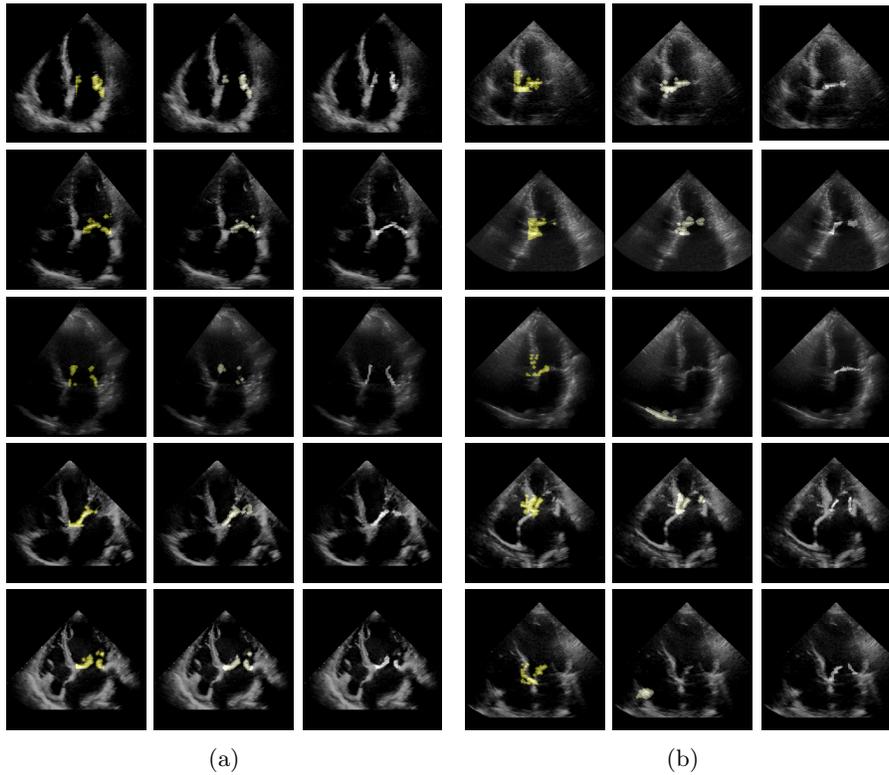


Fig. 5: The mitral valve segmentation masks for the echos with the (a) five highest and (b) lowest Dice coefficients according to NN-MitralSeg are given. From left to right: NN-MitralSeg (yellow), R-NNMF (green) and ground truth (white).

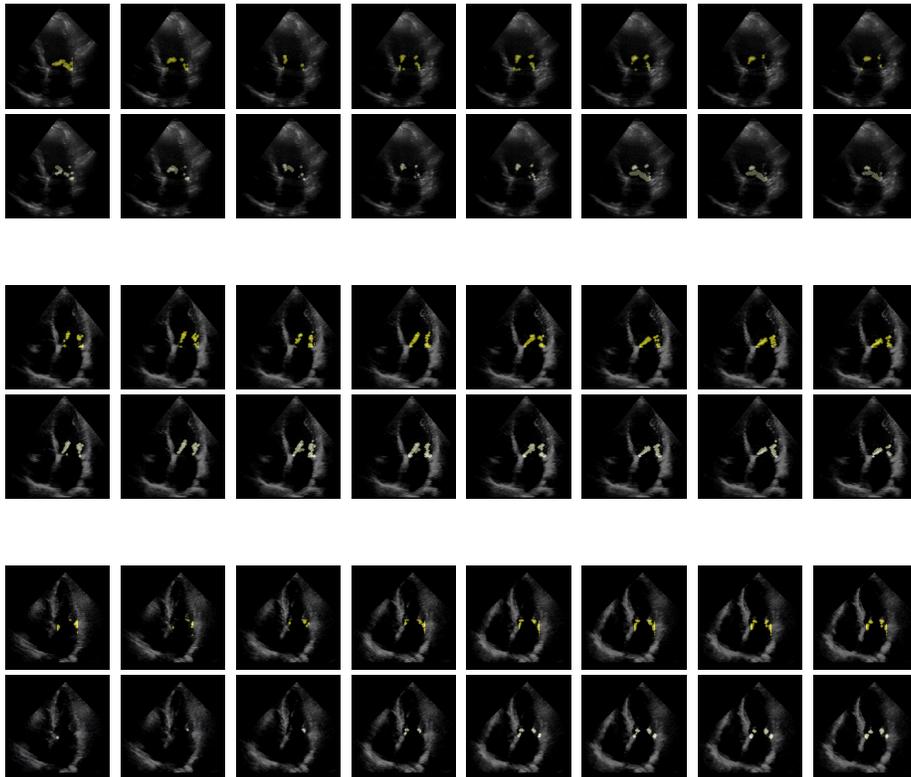


Fig. 6: The mitral valve segmentation masks for 8 consecutive frames and for 3 different echos. NN-MitralSeg (top row, yellow) and R-NNMF (bottom row, green).

## B Hyperparameters and network specifications

The RMSProp optimizer is used with a learning rate of 0.005 and a batch size of 200,000. The dimensions of the latent features is kept constant across all echos with  $D = 10$ , and  $K = 2$ . The sparsity coefficient and regularization parameter are also kept constant across all videos at  $\lambda = 0.2$  and  $\beta = 0.1$ . The window size used in optical flow is chosen to be roughly the same size as the width of the mitral valve, and it varies across echos since the echos are not height and width standardized. The standard deviation of the Gaussian to smooth the temporal derivatives in the optical flow algorithm is set to 3.5. The low dimensional network consists of three fully connected (FC) layers with 10 units each. The latent features and the first two FC layers have ReLU activation and the last FC layer has sigmoid activation to give  $\hat{X}_{n,t} \in [0, 1]$ . The number of training epochs is held constant for each stage of the training. The first pre-training stage is extended for 10 epochs, the second pre-training stage for 5 epochs, and the final training on the full objective for 5 epochs.