# MULTIMODAL LEARNING TO IMPROVE CARDIAC LATE MECHANICAL ACTIVATION DETECTION FROM CINE MR IMAGES

*Jiarui Xing*[a] *Nian Wu*[a]
*Kenneth C. Bilchick*[d] *Frederick H. Epstein*[c] *Miaomiao Zhang*[a,b]

[a] Department of Electrical and Computer Engineering, University of Virginia, USA
[b] Department of Computer Science, University of Virginia, USA
[c] Department of Biomedical Engineering, University of Virginia Health System, USA
[d] School of Medicine, University of Virginia Health System, USA

## ABSTRACT

This paper presents a multimodal deep learning framework that utilizes advanced image techniques to improve the performance of clinical analysis heavily dependent on routinely acquired standard images. More specifically, we develop a joint learning network that for the first time leverages the accuracy and reproducibility of myocardial strains obtained from Displacement Encoding with Stimulated Echo (DENSE) to guide the analysis of cine cardiac magnetic resonance (CMR) imaging in late mechanical activation (LMA) detection. An image registration network is utilized to acquire the knowledge of cardiac motions, an important feature estimator of strain values, from standard cine CMRs. Our framework consists of two major components: (i) a DENSE-supervised strain network leveraging latent motion features learned from a registration network to predict myocardial strains; and (ii) a LMA network taking advantage of the predicted strain for effective LMA detection. Experimental results show that our proposed work substantially improves the performance of strain analysis and LMA detection from cine CMR images, aligning more closely with the achievements of DENSE.

## 1. INTRODUCTION

Myocardial strain has demonstrated its significance in identifying LMA regions for an optimized pacing site for cardiac resynchronization therapy (CRT) [1, 2]. The quantification of myocardial strains can be achieved through various specialized imaging techniques that offer information of ventricular deformation patterns and cardiac motion abnormalities from MR images. Commonly used methods include MR tagging [3], cine SSFP with feature tracking (FT) [4, 5, 6, 7], and cine DENSE [8], with DENSE standing out for its high accuracy in capturing myocardial deformations [9]. Despite the advantages of DENSE, its widespread clinical use is hindered by limited accessibility, primarily due to the high-cost facilities and specialized expertise required for image collection and analysis. Many clinical centers often opt for cine FT. However, the accuracy of FT is compromised by inherent limitations in image quality, including low spatial and temporal resolution. Additionally, these registration-based tracking algorithms focus solely on motions along contours [10].

Recent research has explored the application of deep learning to enhance the accuracy of predicting myocardial motion from cine images, guided by the supervision of DENSE [11]. In this study, the authors employed a neural network to capture the intricate relationship between a time sequence of left ventricular (LV) myocardium segmented from DENSE, and the corresponding encoded displacement fields. By assuming a minimal domain gap between cine and DENSE segmentations in predicting displacement fields, the researchers directly evaluated the trained model on cine input.

Inspired by [11], this paper introduces a multimodal deep learning method that for the first time leverages DENSE to guide the analysis of cine CMRs for an improved LMA detection. Using DENSE strain as ground truth data, we develop an end-to-end joint learning framework that predicts LMA regions (measured by the onset of circumferential shortening (TOS) of segmental myocardial regions [12]) from cine images. Our framework includes two main components: (i) a registration-based strain network to predict the myocardium strain using the learned latent motion/deformation features from cine images, and (ii) a LMA network to predict TOS based on the learned strains. These subnetworks are simultaneously trained to mutually benefit each other, resulting in improved overall performance.

To the best of our knowledge, our method is the first to leverage machine learning to improve LMA detection from cine images, guided by DENSE. This opens promising research venues for transferring knowledge from advanced strain imaging to routinely acquired CMR data. Additionally, our method increases the accessibility to DENSE, particularly in under-resourced regions and populations. Our experimental results demonstrate a substantial improvement in LMA detection accuracy compared to exsiting approaches. Future work will involve meticulous validation of model generalizability as additional patient data becomes available.

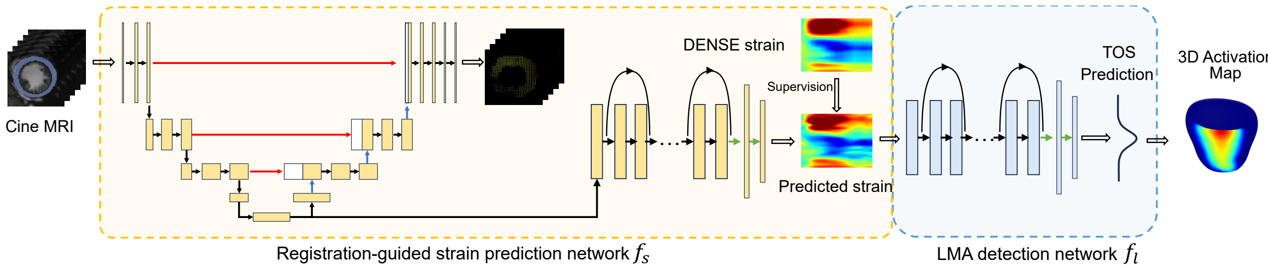


**Fig. 1**: The multimodal joint learning framework with registration-guided strain prediction and LMA detection networks.

## 2. METHODOLOGY

This section presents our joint learning framework of two submodules, including a registration-guided strain prediction network guided by DENSE and a LMA detection network (as illustrated in Fig. 1). Before introducing our model, we will first briefly review CMR myocardial strain analysis.

**CMR strain analysis.** Consider a time-sequence of CMR images with $T$ time frames (see Fig. 2(a)). For each time frame, we compute the circumferential strain along the myocardium based on the displacement fields, and sample the strain values into $N$-dimensional strain vector from a number of $N$ evenly divided myocardial sectors, beginning from the the middle of the intersection points and following counter-clockwise order (see Fig. 2(b)). A $N \times T$ strain matrix containing information from all time frames is built by concatenating the strain vectors across time. A TOS curve labeled from the 2D strain matrix is shown in Fig. 2(c). Here, each TOS value represents the start time of contraction in the corresponding sector, with higher values indicating more severe LMA due to delayed contraction [1].

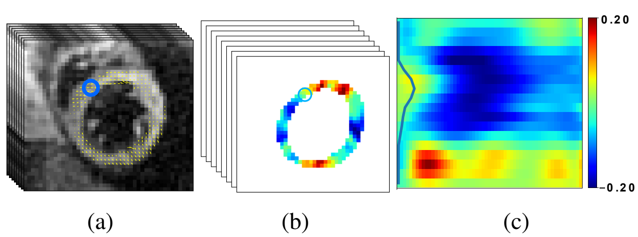


**Fig. 2**: Example of (a) temporal CMRs overlaid with displacement fields; (b) LV strain (contraction/stretching in blue/red; the blue circle shows the sampling starting location); and (c) 2D strain matrix and its corresponding TOS curve.

### 2.1. Our Multimodal Learning Network

**Registration-based strain network.** Given a time-sequence of cine CMRs, $\{I_t\}$, where $t \in [1, \cdots, T]$, we employ a registration network to first learn the latent features of cardiac motions, represented by Lagrangian displacement fields $\{\phi_t\}$, from images. Such latent features, denoted as $z$, are directly utilized to predict strains with the supervision of DENSE strain data. We employ a UNet architecture backbone [13] for our registration network, and a ResNet network for the strain prediction [14, 15]. Analogues to [16], we apply a low-rank singular value decomposition to the predicted strain matrix for smoothness constraints.

Defining DENSE strain as $S$, and $\theta_r, \theta_s$ as the registration network, and strain network parameters respectively, we can now formulate the loss function of our registration-based strain network as

$$l_{\text{strain}} = \sum_{t=1}^{T} \left[ \frac{1}{2\sigma^2} \| I_1 \circ \phi_t(\theta_r) - I_t \|_2^2 + \text{Reg}(\phi_t(\theta_r)) \right] + \alpha \| f_s(z; \theta_s) - S \|_2^2 + \lambda \| \theta_r \|_2^2 + \mu \| \theta_s \|_2^2, \quad (1)$$

where $\circ$ represents interpolation, and $(\sigma, \alpha, \lambda, \mu)$ are positive weighting parameters. The $\text{Reg}(\cdot)$ is a regularization term that encourage the smoothness of the predicted displacement field, $\phi_t$. We adopt the regularization term used in large deformation diffeomorphic metric mapping [17].

**LMA regression network to predict TOS.** Analogous to [15], we develop a LMA regression network to predict the TOS (a $N$-dimensional vector). Given the predicted strain matrix from Eq. (1), we utilize a mean-squared-error of predicted TOS and manually labeled ground truth TOS, denoted as $\boldsymbol{y}$, for network loss, i.e.,

$$l_{\text{TOS}} = \beta \| f_l(f_s(z; \theta_s); \theta_l) - \boldsymbol{y} \|_2^2 + \gamma \| \theta_l \|_2^2, \quad (2)$$

where $\theta_l$ represents network parameters, with $\beta$ and $\gamma$ being the weighting parameters.

**Joint loss optimization.** We jointly optimize the registration-based strain network and the LMA regression network in the training process. The total loss function is the sum of strain loss (in Eq. (1)) and TOS loss (in Eq. (2)), i.e., $l_{\text{strain}} + l_{\text{TOS}}$.

## 3. EXPERIMENTS

We validate our method on cine CMR images paired with DENSE. A comparison of our multimodal joint learning model with existing deep learning methods, including cine FT based on deformable image registration [18] and DENSE for LMA detection [19], is performed.

**Data acquisition.** All short-axis cine bSSFP images were acquired during repeated breath holds covering the LV (temporal resolution, 30-55 ms). Cine DENSE was performed in 4 short-axis planes at basal, two mid-ventricular, and apical levels (with temporal resolution of 17 ms, pixel size of $2.65^2$ mm$^2$, and slice thickness=8mm). Other parameters included displacement encoding frequency $k_e = 0.1$ cycles/mm, flip angle $15°$, and echo time $= 1.08$ ms. All cine and DENSE images are cropped to the size of $128^2$, with $T = 40$ time frames for cine and $T = 20$ for DENSE. All LV myocardium segmentation and ground-truth TOS curves were manually labeled by experts.

**Experimental settings**. In our experiments, we utilized 118 left ventricle MRI scan slices from 24 subjects, divided into 66 slices for training, 26 for validation, and 26 for testing from different subsets of subjects. We first compare our multimodal joint learning model with the baseline algorithms, including cine FT and DENSE-strain for LMA detection and using TOS ($N = 128$) mean square error as the evaluation metric. The TOS error (the MSE between predicted TOS from all methods vs. ground truth) is used a evaluation metric. We also employ a second evaluation metric, which is LMA sector classification accuracy. More specifically, we classify sector as LMA region if its TOS value is greater than a specified variable. While any region where the LV myocardium does not start contraction at the first frame (i.e., TOS=17ms) is considered as LMA, we take the LMA threshold as 18ms to avoid small numerical perturbations in all experiments.

We visualize 3D activation maps reconstructed from the TOS prediction. Using myocardium segmentation from sparsely scanned CMR slices, we first construct coordinates for densely sampled points on the myocardium surface through spatial interpolation. A similar interpolation strategy is then used to estimate TOS at those sampled points.

All experiments were trained on an Nvidia 2080Ti GPU using an Adam optimizer. The hyper-parameters are tuned with grid search strategy, and the optimal values are $\sigma = 0.03$, $\lambda = \mu = \gamma = 0.0001$, $\alpha = 1000$ and $\beta = 0.005$.

**Experimental results.** The top panel of Fig. 3 shows examples of estimated TOS by our method and all baselines. It shows that DENSE-strain predicted TOS fits the ground truth better than cine FT, especially in the peak region of TOS. Our method is able to bridge the gap between DENSE and cine FT, reaching closer TOS prediction to DENSE. The bottom panel of Fig. 3 displays quantitative results of TOS error and LMA region classification error of all methods. Similarly, our method achieves closer accuracy to DENSE with substantially improved performance over cine FT.

Fig. 4 shows a comparison of reconstructed 3D activation maps using all methods vs. the ground-truth TOS data. Note that regions with TOS values much larger than 18ms (shown in red) indicate severe late activation, and normal regions (shown in blue) are typically with small TOS values. Our approaches provide the more accurate LMA region estimation than cine FT.

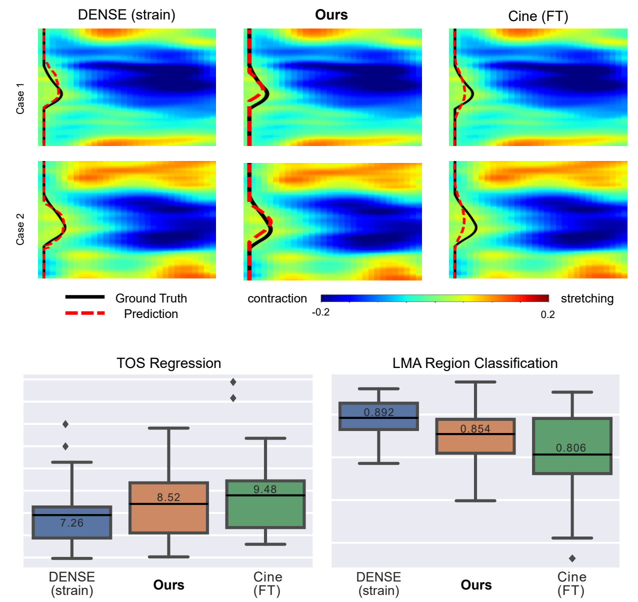


**Fig. 3**: Top to bottom panel: a comparison of TOS prediction from all methods vs. manually labeled TOS (marked in solid black) overlaid on strain matrix; TOS regression mean square error vs. LMA classification accuracy from all methods.

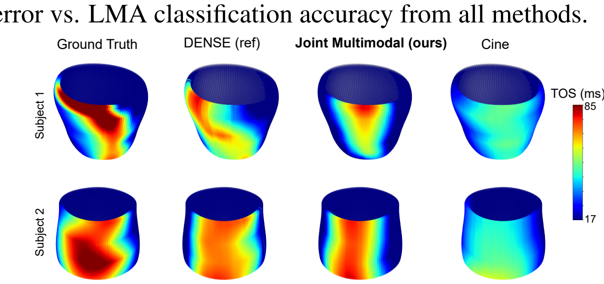


**Fig. 4**: Left to right: a comparison of 3D Activation Maps from ground truth vs. reconstructed from all methods.

## 4. CONCLUSION

This paper presents a multimodal deep learning framework that provides improved cardiac LMA detection accuracy from routinely acquired standard cine CMR images. Experimental results on LMA detection tasks and 3D activation map visualization show that our work substantially outperforms current approaches based on cine FT, and offers performance that aligns more closely with the achievements with DENSE. Experimental findings in this paper indicate a promising convergence of accessibility and accuracy in the analysis of CMR strain imaging. Our future work will focus on (i) further improve the model accuracy to match the DENSE performance; and (ii) thoroughly validate the model's generalizability as more patient data becomes available.

**Compliance with ethical standards.** This work was supported by NIH 1R21EB032597. All studies involving human subjects and waiver of consent were approved by our institutional review board.

## 5. REFERENCES


[1] K. C. Bilchick, D. A. Auger, M. Abdishektaei, R. Mathew, M.-W. Sohn, X. Cai, C. Sun, A. Narayan, R. Malhotra, A. Darby, et al., "Cmr dense and the seattle heart failure model inform survival and arrhythmia risk after crt," *Cardiovascular Imaging*, vol. 13, no. 4, pp. 924–936, 2020.

[2] L. P. Budge, A. S. Helms, M. Salerno, C. M. Kramer, F. H. Epstein, and K. C. Bilchick, "Mr cine dense dyssynchrony parameters for the evaluation of heart failure: comparison with myocardial tissue tagging," *JACC: Cardiovascular Imaging*, vol. 5, no. 8, pp. 789–797, 2012.

[3] L. Axel and L. Dougherty, "Heart wall motion: improved method of spatial modulation of magnetization for mr imaging.," *Radiology*, vol. 172, no. 2, pp. 349–350, 1989.

[4] M. Tee, J. A. Noble, and D. A. Bluemke, "Imaging techniques for cardiac strain and deformation: comparison of echocardiography, cardiac magnetic resonance and cardiac computed tomography," *Expert review of cardiovascular therapy*, vol. 11, no. 2, pp. 221–231, 2013.

[5] M. A. Morales, M. Van den Boomen, C. Nguyen, J. Kalpathy-Cramer, B. R. Rosen, C. M. Stultz, D. Izquierdo-Garcia, and C. Catana, "Deepstrain: a deep learning workflow for the automated characterization of cardiac mechanics," *Frontiers in Cardiovascular Medicine*, vol. 8, pp. 730316, 2021.

[6] M. Qiao, Y. Wang, Y. Guo, L. Huang, L. Xia, and Q. Tao, "Temporally coherent cardiac motion tracking from cine mri: Traditional registration method and modern cnn method," *Medical Physics*, vol. 47, no. 9, pp. 4189–4198, 2020.

[7] C. Qin, W. Bai, J. Schlemper, S. E. Petersen, S. K. Piechnik, S. Neubauer, and D. Rueckert, "Joint motion estimation and segmentation from undersampled cardiac mr image," in *Machine Learning for Medical Image Reconstruction - MICCAI Workshop*. Springer, 2018, pp. 55–63.

[8] D. Kim, W. D. Gilson, C. M. Kramer, and F. H. Epstein, "Myocardial tissue tracking with two-dimensional cine displacement-encoded mr imaging: development and initial evaluation," *Radiology*, vol. 230, no. 3, pp. 862–871, 2004.

[9] J. Kar, A. K. Knutsen, B. P. Cupps, and M. K. Pasque, "A validation of two-dimensional in vivo regional strain computed from displacement encoding with stimulated echoes (dense), in reference to tagged magnetic resonance imaging and studies in repeatability," *Annals of biomedical engineering*, vol. 42, pp. 541–554, 2014.

[10] A. A. Young, B. Li, R. S. Kirton, and B. R. Cowan, "Generalized spatiotemporal myocardial strain analysis for dense and spamm imaging," *Magnetic resonance in medicine*, vol. 67, no. 6, pp. 1590–1599, 2012.

[11] Y. Wang, C. Sun, S. Ghadimi, D. C. Auger, P. Croisille, M. Viallon, K. Mangion, C. Berry, C. M. Haggerty, L. Jing, et al., "Strainnet: Improved myocardial strain analysis of cine mri by deep learning from dense," *Radiology: Cardiothoracic Imaging*, vol. 5, no. 3, pp. e220196, 2023.

[12] B. T. Wyman, W. C. Hunter, F. W. Prinzen, and E. R. McVeigh, "Mapping propagation of mechanical activation in the paced heart with mri tagging," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 276, no. 3, pp. H881–H891, 1999.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[15] J. Xing, S. Wang, K. C. Bilchick, F. H. Epstein, A. R. Patel, and M. Zhang, "Multitask learning for improved late mechanical activation detection of heart from cine dense mri," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023, pp. 1–5.

[16] R. Ramachandran, X. Chen, C. M. Kramer, F. H. Epstein, and K. C. Bilchick, "Singular value decomposition applied to cardiac strain from mr imaging for selection of optimal cardiac resynchronization therapy candidates," *Radiology*, vol. 275, no. 2, pp. 413–420, 2015.

[17] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *International journal of computer vision*, vol. 61, pp. 139–157, 2005.

[18] M. Zhang and P. T. Fletcher, "Finite-dimensional lie algebras for fast diffeomorphic image registration," in *Information Processing in Medical Imaging: 24th International Conference*. Springer, 2015, pp. 249–260.

[19] J. Xing, S. Ghadimi, M. Abdi, K. C. Bilchick, F. H. Epstein, and M. Zhang, "Deep networks to automatically detect late-activating regions of the heart," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 1902–1906.