

Generative Brownian Bridge Diffusion In Motion Space For Enhanced Myocardial Strain Analysis

Rishov Paul¹, Frederick H. Epstein³, and Miaomiao Zhang^{1,2}

¹ Department of Computer Science, University of Virginia, USA

² Department of Electrical and Computer Engineering, University of Virginia, USA

³ Department of Biomedical Engineering, University of Virginia, USA

Abstract. Myocardial strain analysis of cardiac magnetic resonance (CMR) images provides an important tool for evaluating cardiac function. However, current techniques require either human-adjusted post-processing with suboptimal regional accuracy, or specialized acquisitions with limited availability. In this paper, we propose to leverage the power of generative models to synthesize high-quality motion-derived strain values from routinely acquired CMR sequences. Specifically, we develop a novel Brownian bridge diffusion model in motion space to learn the probabilistic mapping between standard CMR motion estimated from widely adopted registration methods and highly accurate motion provided by advanced strain imaging techniques. To promote the fidelity of anatomical structure in the generation process, our model is conditioned on the corresponding CMR images. We validate our method on large-scale multi-center CMR datasets including subjects of paired standard cine CMR and advanced strain imaging acquisitions. Experimental results demonstrate that our framework significantly improves the accuracy of motion prediction and strain analysis from standard CMRs compared to existing learning-based approaches. Our research represents a new paradigm for potentially developing cost-effective, clinically deployable AI tools for cardiac function assessment with enhanced strain accuracy in busy clinical workflows. Our code is publicly available at <https://github.com/Rishov-MIA/Brownian-Bridge-strain-analysis>.

Keywords: Myocardial Strain Imaging · Brownian Bridge Diffusion · Spatiotemporal Motion.

1 Introduction

Quantification of myocardial strain from cine CMR image sequences plays an important role in assessing cardiac ventricular function and diagnosing heart diseases [1, 24]. Current approaches often rely on post-processing image registration [29, 26] to track myocardial motion that provides a direct measure for strain values throughout the cardiac cycle. However, these methods are limited by heavy dependence on image texture and susceptibility to noise with compromised regional registration accuracy [26]. Advanced imaging techniques, such as displacement encoding with stimulated echoes (DENSE), have gained

increasing attention and are emerging as a gold standard for myocardial strain analysis due to their ability to directly encode highly accurate tissue displacement/motion [36, 12]. Despite these advantages, there has not been widespread adoption of DENSE in routine clinical workflows due to limited availability of the methods and the need for additional acquisitions.

To bridge this gap, earlier studies have made substantial progress in developing deep learning-based image registration to improve the accuracy of myocardial motion and strain estimation from standard cine CMR images. These approaches incorporate spatial and temporal regularization to enforce motion continuity and stabilize the subsequent strain quantification [5, 20, 18]. In parallel, architectural improvements, including multi-scale feature learning [22] and transformer-based self-attention [28], have been introduced to better capture motion patterns across spatial or temporal resolutions. More recent efforts further integrate biomechanical priors and physics-based constraints to improve anatomical plausibility and physiological consistency of predicted motion fields [11, 21]. Additionally, supervised paradigms have also emerged by leveraging advanced imaging modalities (i.e., DENSE) to provide high-fidelity motion supervision during training [31, 34]. While these methods have achieved state-of-the-art performance in motion and strain estimation, they remain limited in modeling the inherently multimodal nature of cardiac motion, particularly under pathological variability. This limitation comes from either deterministic modeling assumptions [13, 31] or diffusion-based frameworks that primarily operate through iterative latent feature refinement rather than explicitly modeling the multimodal motion distribution [34]. Furthermore, many supervised approaches require spatiotemporal pre-alignment between paired acquisitions [13, 31], which can introduce artifacts and reduce model robustness.

Inspired by recent advances in Brownian bridge diffusion models [16], we introduce a new generative diffusion framework in motion space to learn a stochastic mapping between paired motion representations. Specifically, our model is trained to map motion predicted from registration networks into high-quality advanced imaging modality; thereby serving as a refinement module of myocardial motion estimated from CMRs. In addition, we incorporate CMR image features as conditional guidance during the reverse diffusion process to ensure the generated motion trajectories remain consistent with underlying cardiac anatomy and temporal dynamics. Our main contributions are summarized as

- (i) Propose the first Brownian bridge diffusion framework operating directly in spatiotemporal displacement space to learn probabilistic mappings between readily obtainable registration-derived motion and difficult-to-acquire high-fidelity motion distributions from advanced imaging.
- (ii) Introduce structure-conditioned diffusion denoising that integrates CMR image features during the reverse diffusion process to enforce biomechanically and anatomically plausible spatiotemporal motion evolution.
- (iii) Demonstrate consistent and significant improvements over state-of-the-art methods in both motion reconstruction fidelity and subsequent myocardial strain quantification across a large multi-center dataset.

2 Background: Registration-based Motion Estimation

In this section, we will review a general mathematical framework for motion estimation based on deformable image registration [33, 8, 19]. The goal is to estimate displacement fields that describe myocardial motion over time by aligning a temporal image sequence to a reference frame.

Let $\Omega \subset \mathbb{R}^d$ denote a d -dimensional image domain. Given a reference image $I_0 : \Omega \rightarrow \mathbb{R}$ and a sequence of moving images $\{I_t\}_{t=1}^T, t \in \{1, \dots, T\}$, the objective is to estimate a sequence of spatial transformations, $\{\phi_t\}_{t=1}^T : \Omega \rightarrow \Omega$, that map spatial locations from the reference frame to the target frame at time t . The corresponding displacement field is defined as $u_t = \phi_t - e$, where e denotes an identity transformation. A classical deformable registration [26, 7, 3] formulates motion estimation as an energy minimization problem over a sequence of transformations $\{\phi_t\}$, parameterized by velocity fields v_t [3, 2, 7], as

$$\min \mathcal{L}(\phi_t) = \sum_{t=1}^T \text{Dist}(I_0 \circ \phi_t^{-1}, I_t) + \lambda \text{Reg}(v_t), \quad (1)$$

where $\circ$ denotes an interpolation operator, $\text{Dist}(\cdot, \cdot)$ measures image dissimilarity, and $\text{Reg}(\cdot)$ is a regularization term that enforces the spatiotemporal smoothness of the transformation fields. The parameter $\lambda > 0$ controls the trade-off between image matching fidelity and deformation regularity. Various choices of spatial regularizations, such as stationary velocity field (SVF) [3] and large deformation diffeomorphic metric mapping (LDDMM) [2, 7], enforce diffeomorphic constraints (a.k.a. one-to-one smooth and inverse smooth mapping) to ensure anatomically plausible transformations. In particular, under the SVF parameterization, the transformation is obtained from the velocity field through the exponential map, i.e., $\phi_t = \exp(v_t)$, and its inverse is directly given by $\phi_t^{-1} = \exp(-v_t)$. In this paper, we adopt commonly used sum-of-squared intensity differences (SSD) as the similarity metric [33, 23], and SVF for the regularity.

Once displacement fields are estimated, myocardial strain can be computed from the local deformation gradient tensor [34, 6], i.e., $F_t = \nabla \phi_t = e + \nabla u_t$, where ∇u_t is the Jacobian of the displacement field. For finite deformation analysis, myocardial strain is commonly quantified using the Green-Lagrange strain tensor [25, 17], i.e.,

$$S_t = \frac{1}{2}(F_t^T F_t - e). \quad (2)$$

3 Our Method

In this section, we develop a conditional Brownian bridge diffusion model in a vectorized spatiotemporal displacement space to explicitly learn the probabilistic mapping between two distinct motion domains. Once trained, the proposed model takes motion fields estimated from standard cine CMR sequences and translates them through the learned diffusion bridge to generate high-fidelity

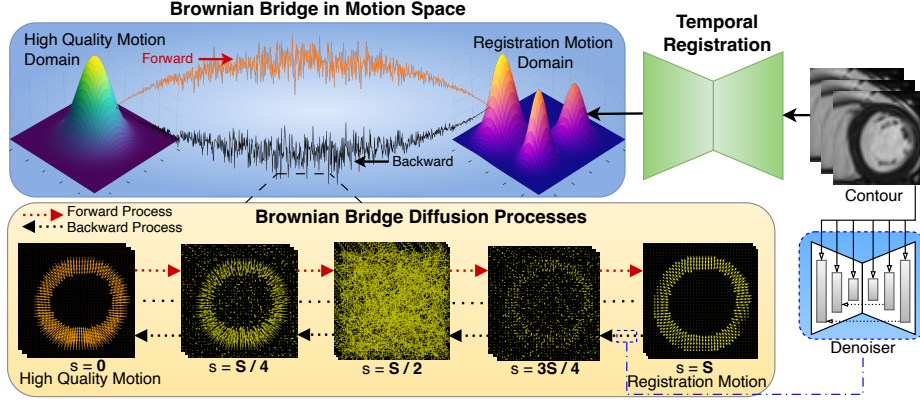


Fig. 1: An overview of our Brownian bridge diffusion model in motion space.

motion estimates without specialized acquisition. Our framework includes two key components: (i) a temporal registration network that predicts spatiotemporal motion fields from input CMR videos, and (ii) a conditional Brownian bridge video diffusion model that synthesizes high-quality motion trajectories given previously learned registration-derived motion. An overall architecture of our proposed method is shown in Fig. 1.

Motion learning from spatiotemporal registration network. Given a training dataset of N cardiac image sequences, $\{\mathcal{I}^i\}_{i=1}^N, i \in \{1, \dots, N\}$, where each sequence $\mathcal{I}^i$ contains $T + 1$ temporal frames, i.e., $\mathcal{I}^i = \{\mathcal{I}_t^i\}_{t=0}^T$. Our objective is to predict a sequence of velocity fields, $\{v_t^i\}$, from which the deformation fields $\{\phi_t^i\}$ associated with the reference image and following images $\{\mathcal{I}_0^i, \mathcal{I}_t^i\}_{t=1}^T$ are obtained via the SVF regularization [3]. Analogous to [33], we employ a spatiotemporal registration network parameterized by θ , and formulate the corresponding loss function as

$$\mathcal{L}(\theta) = \sum_{i=1}^N \left[\sum_{t=1}^T \|\mathcal{I}_0^i \circ \phi_t^i(\theta) - \mathcal{I}_t^i\|_2^2 + \sum_{t=1}^T \lambda \|\nabla v_t^i(\theta)\|^2 + \sum_{t=2}^T \beta \|\Delta v_t^i(\theta)\|^2 \right], \quad (3)$$

where $\Delta v_t^i(\theta) \triangleq v_t^i(\theta) - v_{t-1}^i(\theta)$, representing a temporal regularization, and β is a controlling weighting parameter. Given the predicted transformation fields $\{\phi_t^i\}$, the displacement fields can be computed as $u_t^i = \phi_t^i - e$. Note that, in this paper, we adopt a U-Net backbone with a latent residual network to explicitly model temporal correlations in the registration framework [33]. Our proposed methodology is architecture-agnostic, and alternative backbone designs [10, 4] can be readily incorporated without modification to the overall formulation.

Conditional Brownian Bridge Diffusion in Motion Space. Let $(\mathbf{x}, \mathbf{u}) \in \mathbb{R}^{d \times T \times H \times W}$ denote paired high quality motion acquired by advanced imaging

and registration-derived motion fields in the same spatial coordinate system using the same reference image. Here, H and W denote the height and width of each image. We model motion refinement as a stochastic Brownian bridge defined over the motion space, with boundary conditions $\mathbf{x}_0 = \mathbf{x}$ and $\mathbf{x}_S = \mathbf{u}$, where S represents the ending point.

Following a similar principle of [16], the *forward diffusion process* evolves intermediate motion states by progressively transporting $\mathbf{x}$ toward $\mathbf{u}$ via a variance-preserving linear interpolation. Note that since the paired motions are defined with respect to the same reference image, we assume the interpolation is performed without introducing coordinate misalignment or reparameterization errors. Let $m_s \in [0, 1]$ denote a monotonically increasing schedule for $s \in \{0, \dots, S\}$. The forward process in the motion space can be formulated as

$$q_{\text{BB}}(\mathbf{x}_s | \mathbf{x}, \mathbf{u}) = \mathcal{N}((1 - m_s)\mathbf{x} + m_s\mathbf{u}, \delta_s \mathbf{I}); \quad m_s = s/S. \quad (4)$$

where $\delta_s = 2\tau(m_s - m_s^2)$ with τ being the variance scaling factor [16]. Sampling from this process gives $\mathbf{x}_s = (1 - m_s)\mathbf{x} + m_s\mathbf{u} + \sqrt{\delta_s}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

The *reverse process* is initialized by enforcing the ending boundary condition $\mathbf{x}_S = \mathbf{u}$. Starting from this fixed endpoint, the model progressively refines the registration-motion trajectory toward the high-quality motion domain. To improve data fidelity and enforce anatomical consistency, we condition the reverse dynamics on image features, denoted as $\mathbf{c}$, derived from the input data (e.g., myocardial contours or latent CMR image representations). Formally, our reverse transition can be parameterized as a conditional Gaussian

$$p_\theta(\mathbf{x}_{s-1} | \mathbf{x}_s, \mathbf{u}, \mathbf{c}) = \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{x}_s, s, \mathbf{c}), \tilde{\delta}_s \mathbf{I}), \quad (5)$$

where $\boldsymbol{\mu}_\theta$ is predicted by a neural network (e.g., with a commonly used 3D UNet backbone) and variance $\tilde{\delta}_s$ follows a fixed variance schedule [16].

Similar to [16], instead of predicting the noise directly, we parameterize the denoising objective using a bridge gradient formulation defined as $\mathbf{b}_s = m_s(\mathbf{u} - \mathbf{x}) + \sqrt{\delta_s}\boldsymbol{\epsilon}$. At each timestep, the denoising network predicts the Brownian bridge update along the motion trajectory, i.e., $\hat{\mathbf{b}}_s = f_\theta(x_s, s, \mathbf{c})$, from which the displacement field is recovered as $\hat{\mathbf{x}}_0 = \mathbf{x}_s - \hat{\mathbf{b}}_s$. To explicitly enforce physically plausible and smooth motions, we incorporate a spatial regularization into the loss function to constrain non-smooth or anatomically implausible updates. The network parameters θ are then learned by minimizing loss function as

$$\min_\theta \mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}_0, \mathbf{u}, \boldsymbol{\epsilon}, \mathbf{c}, s} \left[\left\| \mathbf{b}_s - \hat{\mathbf{b}}_s \right\|^2 + \lambda_{\text{reg}} \sum_{t=1}^T \left\| \nabla \hat{\mathbf{x}}_0^t \right\|_2^2 \right], \quad (6)$$

where λ_{reg} is a regularization parameter. In this paper, we also incorporate segmented CMR myocardium contours as conditioning information into each motion state $\mathbf{x}_s$ through a multi-head cross-attention mechanism [28].

During inference, we initialize the reverse process directly with the registration motion $\mathbf{x}_S = \mathbf{u}$ and iteratively sample $\mathbf{x}_{s'} = (1 - m_{s'})\hat{\mathbf{x}}_0 + m_{s'}\mathbf{u} + \sqrt{(\delta_{s'} - \tilde{\delta}_s)/\delta_s}(\mathbf{x}_s - (1 - m_s)\hat{\mathbf{x}}_0 - m_s\mathbf{u}) + \sqrt{\tilde{\delta}_s}\boldsymbol{\epsilon}$, where $s' < s$ is the next retained timestep in a subsampled non-Markovian schedule and $\tilde{\delta}_s = \delta_{s'}/\delta_s(\delta_s -$

$\delta_{s'}(1 - m_s)^2 / (1 - m_{s'})^2$). At the final timestep $s = 0$, the model directly returns the reconstructed motion $\hat{\mathbf{x}}_0$ without noise injection.

4 Experimental Evaluation

We compare our model against state-of-the-art methods for myocardial strain estimation, covering both deterministic (TransUNet [9], and StrainNet [31]) and generative approaches (Flow-matching [27], vid2vid [30], ControlNet [35], and LaMoD [34]). For a fair comparison, all baselines were retrained from scratch, and every model was trained with Adam [15] on a single NVIDIA A100 80GB GPU. The hyperparameters of our model were selected via grid search, resulting in a batch size of 16, learning rate and weight decay of 1×10^{-4} , spatial regularization weight $\lambda_{\text{reg}} = 0.01$, and $\tau = 1$ for all experiments.

Dataset. We include two CMR modalities: routinely acquired left-ventricular (LV) short-axis cine CMR sequences and advanced DENSE-CMR acquisitions, along with manually annotated LV masks [34, 31]. The DENSE-CMR provides both magnitude images and high-quality displacement/motion measurements that serve as a reference standard for circumferential strain evaluation. The DENSE dataset was collected from eight clinical centers [31] and consists of 1,200 sequences from 284 subjects (124 healthy volunteers and 160 patients with heart disease). Sequences are divided subject-wise into training, validation, and testing sets at an 80:10:10 ratio to prevent data leakage across splits. Each of the 116 test sequences is paired with a cine acquisition from the same subject at approximately the same slice location (within ± 4 mm) which lets the DENSE-derived displacements serve as ground-truth motion for the cine sequences. All CMR videos are cropped to a $48^2 \times 20$ LV region of interest, with average in-plane spatial resolution being 2.41 mm/pixel for DENSE and 1.68 mm/pixel for cine. Our model is trained exclusively on DENSE-CMR and evaluated on both DENSE and the paired cine CMR sequences.

Evaluation. We first evaluate the accuracy of predicted displacement vs. DENSE ground truth using a normalized pixel-wise end-point error (nEPE) [31, 34], defined as the EPE normalized by the ground-truth displacement magnitude. Our model takes DENSE magnitude images as input, and its predicted displacement fields are compared with ground-truth fields computed from the corresponding DENSE phase images. We report nEPE on DENSE-CMR, where magnitude images and displacement fields share the same spatial coordinate system.

We then compute myocardial circumferential strain from the motions predicted from cine CMR images using Eq. (2), and compare our method against all baselines using mean absolute error (MAE) under four evaluation settings: whole-slice and segmental strain, each across all frames and at end-systole (ES), defined as the frame of minimum strain value. Strain MAE is reported in percentage points (%), representing absolute error rather than relative percent difference. Note that myocardial strain values typically range up to approximately 25% rather than 100% [31, 18]. Following the standard clinical convention [1], the myocardium is partitioned into six segments for basal and mid-ventricular

slices and four segments for apical slices, beginning at the right ventricular insertion point and proceeding counterclockwise. Whole-slice strain is obtained by averaging the segmental strain values prior to error computation. All predicted strain measurements are compared against reference strain values derived from the paired high-quality DENSE. Due to space limit, we report significance on segmental ES MAE using two-sided Wilcoxon signed-rank tests [32] against each baseline, with Holm correction [14] for multiple comparisons ($p < 0.05$).

Experiment Results. Tab. 1 reports both DENSE displacement errors and cine-derived strain errors on paired Cine-DENSE test data. The results demonstrate that our method achieves top-tier motion tracking fidelity on par with the strongest baseline (LaMoD [34]) in displacement nEPE. Similarly, our method consistently offers the best performance for cine-derived strain estimation across all evaluation settings, with particularly notable improvements in ES strain analysis. Since myocardial strain spans a narrow range of roughly 25%, a strain MAE of 1% already corresponds to about 4-5% of the clinically relevant range. In Tab. 1, the most substantial gains are observed in ES strain estimation, which underscore our method’s ability to accurately capture peak myocardial deformation. The Holm-corrected Wilcoxon signed-rank tests confirm that the improvement in segmental ES strain is significant against every baseline ($p < 0.05$). Furthermore, the lower standard deviations of errors suggest the enhanced robustness and stability of our approach over all baselines. Fig. 2 reports the performance of our method across stratified cohort (healthy vs. diseases). It shows that errors (both on displacements and strain values) are generally higher on diseased subjects where more heterogeneous motion exhibits.

Table 1: Comparison of displacement/strain error on paired Cine-DENSE data.

Method	DENSE Displacement nEPE ↓		Cine Strain MAE (%) ↓				Segmental ES p -value
	All-frame	ES	Whole- slice	Segmental	Whole- slice ES	Segmental ES	
TransUNet	0.60 ± 0.14	0.52 ± 0.16	5.37 ± 2.24	5.88 ± 1.98	7.84 ± 3.93	8.46 ± 3.42	<0.001
StrainNet	0.56 ± 0.14	0.49 ± 0.13	3.87 ± 1.76	4.57 ± 1.40	5.55 ± 3.05	6.37 ± 2.39	<0.001
vid2vid	0.65 ± 0.20	0.54 ± 0.21	7.76 ± 3.12	8.21 ± 2.82	11.52 ± 4.33	11.82 ± 3.83	<0.001
Flow-matching	0.53 ± 0.21	0.44 ± 0.18	3.96 ± 2.43	5.22 ± 2.18	5.26 ± 4.65	6.79 ± 4.03	0.013
ControlNet	0.54 ± 0.19	0.46 ± 0.19	4.42 ± 2.29	4.87 ± 2.06	5.87 ± 3.77	6.53 ± 3.35	<0.001
LaMoD	0.53 ± 0.19	0.44 ± 0.16	3.69 ± 1.71	4.22 ± 1.45	5.03 ± 3.16	5.80 ± 2.68	0.013
Ours	0.52 ± 0.16	0.43 ± 0.15	3.18 ± 1.75	4.14 ± 1.49	4.06 ± 2.83	5.17 ± 2.35	—

Ablation Study. Tab. 2 reports the ablation study on evaluating the individual contribution of Brownian bridge, as well as CMR feature conditioning in our proposed framework. Results show that incorporating the Brownian bridge module substantially improves performance, yielding significant reductions in both displacement and strain errors compared with the registration-only model. Further

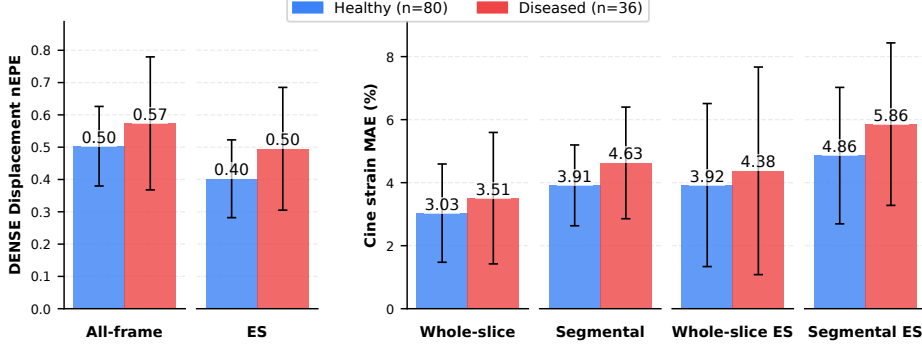


Fig. 2: Left to right: errors of displacements and strain on stratified cohorts.

integrating anatomical contour guidance leads to additional performance gains and consistently achieves the best results across all evaluation metrics.

Qualitative Analysis. Fig. 3 visualizes four representative cases spanning varying pathologies. For an acute myocardial infarction (MI) patient and a healthy subject (top), we compare the DENSE ground truth (GT), our method, and the four best-performing baseline models reported from Tab. 1. The results are demonstrated using segmental strain curves and ES displacement fields overlaid with error maps. Our method achieves the highest agreement with the DENSE-strain across all segments throughout the cardiac cycle. Similarly, the ES displacement shows the highest similarity to DENSE (colored in yellow). In contrast, all baseline methods present noticeably larger deviations (colored in red). Due to space limitations, the additional LBBB and ischemic cases (bottom) show only our method against the DENSE GT, where the baselines follow the same trends reported above. Our model preserves the inter-segmental dyssynchrony characteristic of LBBB and the reduced deformation of the affected inferior segment in the ischemic case.

Discussion. While our method demonstrates improved agreement with DENSE-derived myocardial strain measurements, several limitations also point to promising directions for future work. First, our validation focuses on agreement with DENSE-derived displacement and strain, and does not yet establish whether the synthesized motion captures clinically discriminative patterns. Evaluating the generated motion on downstream tasks, such as disease classification, will

Table 2: Ablation study to assess each proposed module in our proposed method.

Registration	Components		DENSE Displacement	Cine Strain MAE (%) ↓			
	Brownian Bridge	Contour Guidance	nEPE ↓	Whole-slice	Segmental	Whole-slice ES	Segmental ES
✓	✗	✗	0.73 ± 0.18	5.35 ± 2.10	6.01 ± 1.87	6.05 ± 3.36	6.89 ± 2.86
✓	✓	✗	0.62 ± 0.18	3.78 ± 2.03	4.78 ± 2.01	5.01 ± 3.58	5.94 ± 3.15
✓	✓	✓	0.52 ± 0.16	3.18 ± 1.75	4.14 ± 1.49	4.06 ± 2.83	5.17 ± 2.35

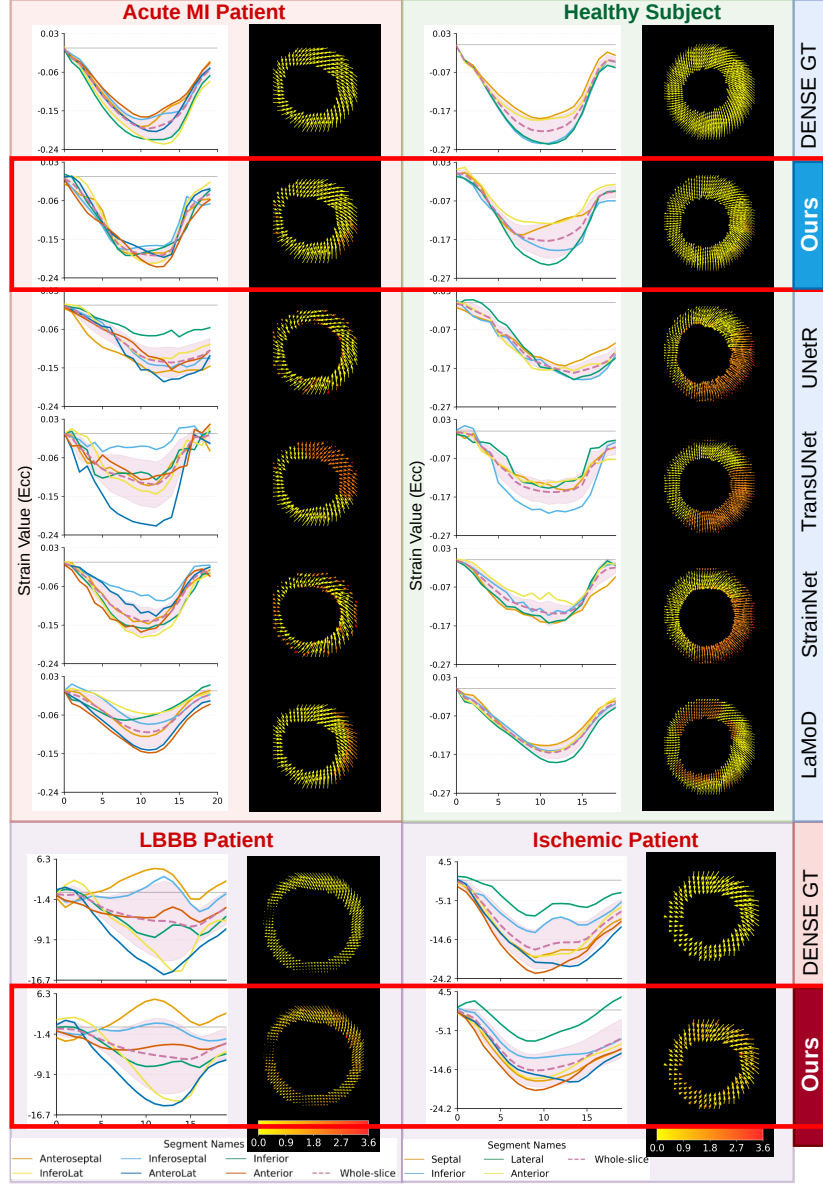


Fig. 3: Top: a comparison of strain curves and ES displacement fields from DENSE ground truth vs. those predicted by all methods. Bottom: additional examples from LBBB and ischemic patients, comparing DENSE vs. predictions from our model.

be an important next step. Second, our model errors relatively higher in the diseased cohort indicate underrepresented heterogeneous motion in the training data. Improving training balance and representation of these challenging cases may further improve performance. Third, our method may inherit errors from registration networks when used as initial conditions for the Brownian bridge learning. This represents another promising direction for future work.

5 Conclusion

In this paper, we introduce the first Brownian bridge framework in the cardiac motion space to synthesize advanced yet clinically less accessible DENSE-quality motion from spatiotemporal registration-predicted displacement trajectories. In order to enhance the fidelity and anatomical consistency of the synthesized motion, we incorporate complementary CMR image features as an effective conditioning during the reverse process. Experimental results on a large multi-site cohort demonstrate that our approach consistently outperforms state-of-the-art myocardial strain analysis methods without requiring specialized acquisitions at inference. By bridging the gap between routine cine imaging and advanced myocardial motion acquisition, our framework provides a promising approach for improving myocardial motion and strain estimation from standard cine CMR, with the potential of future translation into clinical workflows.

Disclosure of Interests. The authors have no conflicts of interest to disclose.

Acknowledgment. This work was supported by NIH 1R21EB032597 and funding from the UVA Wallace H. Coulter Center for Translational Research.

References

1. Amzulescu, M.S., De Craene, M., Langet, et al.: Myocardial strain imaging: review of general principles, validation, and sources of discrepancies. *European Heart Journal-Cardiovascular Imaging* **20**(6), 605–619 (2019)
2. Arnold, V.: Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l’hydrodynamique des fluides parfaits. In: *Annales de l’institut Fourier*. vol. 16, pp. 319–361 (1966)
3. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 924–931. Springer (2006)
4. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging* **38**(8), 1788–1800 (2019)
5. Balakrishnan, G., Zhao, A., Sabuncu, M.R., et al.: An unsupervised learning model for deformable medical image registration. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 9252–9260 (2018)
6. Barnafi, N., Regazzoni, F., et al.: Reconstructing relaxed configurations in elastic bodies: Mathematical formulations and numerical methods for cardiac modeling. *Computer Methods in Applied Mechanics and Engineering* **423**, 116845 (2024)

7. Beg, M.F., Miller, M.I., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision* **61**, 139–157 (2005)
8. Chandrashekar, R., Rao, A., Sanchez-Ortiz, G.I., Mohiaddin, R.H., Rueckert, D.: Construction of a statistical model for cardiac motion analysis using nonrigid image registration. In: *Biennial International Conference on Information Processing in Medical Imaging*. pp. 599–610. Springer (2003)
9. Chen, J., Lu, Y., Yu, Q., et al.: Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306* (2021)
10. Chen, J., Frey, E.C., He, Y., et al.: Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis* **82**, 102615 (2022)
11. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical image analysis* **57**, 226–236 (2019)
12. Ghadimi, S., Abdi, M., Epstein, F.H.: Improved computation of lagrangian tissue displacement and strain for cine dense mri using a regularized spatiotemporal least squares method. *Frontiers in Cardiovascular Medicine* **10**, 1095159 (2023)
13. Hatamizadeh, A., Tang, Y., Nath, V., et al.: Unetr: Transformers for 3d medical image segmentation. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. pp. 574–584 (2022)
14. Holm, S.: A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* pp. 65–70 (1979)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
16. Li, B., Xue, K., Liu, B., Lai, Y.K.: Bbmd: Image-to-image translation with brownian bridge diffusion models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern Recognition*. pp. 1952–1961 (2023)
17. Marwick, T.H., Shah, S.J., Thomas, J.D.: Myocardial strain in the assessment of patients with heart failure: a review. *JAMA cardiology* **4**(3), 287–294 (2019)
18. Morales, M.A., Van den Boomen, M., Nguyen, C., et al.: Deepstrain: a deep learning workflow for the automated characterization of cardiac mechanics. *Frontiers in Cardiovascular Medicine* **8**, 730316 (2021)
19. Pan, J., Rueckert, D., Küstner, T., Hammernik, K.: Efficient image registration network for non-rigid cardiac motion estimation. In: *International Workshop on Machine Learning for Medical Image Reconstruction*. pp. 14–24. Springer (2021)
20. Qiao, M., Wang, Y., Guo, Y., Huang, L., Xia, L., Tao, Q.: Temporally coherent cardiac motion tracking from cine mri: Traditional registration method and modern cnn method. *Medical Physics* **47**(9), 4189–4198 (2020)
21. Qin, C., Wang, S., Chen, C., Bai, W., Rueckert, D.: Generative myocardial motion tracking via latent space exploration with biomechanics-informed prior. *Medical Image Analysis* **83**, 102682 (2023)
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. pp. 234–241. Springer (2015)
23. Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A.: Diffeomorphic registration using b-splines. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 702–709. Springer (2006)
24. Seo, D., Ho, J., Traverse, J.H., Forder, J., Vemuri, B.: Computing diffeomorphic paths with applications to cardiac motion analysis. In: *4th MICCAI Workshop on Mathematical Foundations of Computational Anatomy*. pp. 83–94. Citeseer (2013)

25. Smiseth, O.A., Torp, H., Opdahl, A., et al.: Myocardial strain imaging: how useful is it in clinical decision making? *European heart journal* **37**(15), 1196–1207 (2016)
26. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. *IEEE transactions on medical imaging* **32**(7), 1153–1190 (2013)
27. Tong, A., Fatras, K., et al.: Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482* (2023)
28. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
29. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric log-domain diffeomorphic registration: A demons-based approach. *IEEE transactions on medical imaging* **29**(1), 1–18 (2010)
30. Wang, T.C., Liu, M.Y., Zhu, J.Y., et al.: Video-to-video synthesis. In: *Advances in Neural Information Processing Systems (NeurIPS)* (2018)
31. Wang, Y., Sun, C., Ghadimi, S., et al.: Strainnet: Improved myocardial strain analysis of cine mri by deep learning from dense. *Radiology: Cardiothoracic Imaging* **5**(3), e220196 (2023)
32. Wilcoxon, F.: Individual comparisons by ranking methods. In: *Breakthroughs in statistics: Methodology and distribution*, pp. 196–202. Springer (1992)
33. Wu, N., Xing, J., Zhang, M.: Tltn: Temporal latent residual networks for large deformation image registration. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 728–738. Springer (2024)
34. Xing, J., Jayakumar, N., Wu, N., Wang, Y., Epstein, F.H., Zhang, M.: Lamod: Latent motion diffusion model for myocardial strain generation. In: *International Workshop on Shape in Medical Imaging*. pp. 164–177. Springer (2024)
35. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models (2023)
36. Zhong, X., Spottiswoode, B.S., Meyer, C.H., Kramer, C.M., Epstein, F.H.: Imaging three-dimensional myocardial mechanics using navigator-gated volumetric spiral cine dense mri. *Magnetic resonance in medicine* **64**(4), 1089–1097 (2010)