T1Diff: Cardiac T1 Mapping Synthesis from Cine MRI based on Diffusion Model

Anonymous submission

Abstract

Cardiac T1 mapping has emerged as a highly sensitive diagnostic tool for detecting a range of cardiac pathologies by abnormal T1 signal, particularly myocardial changes stemming from conditions such as myocardial infarction. Cine magnetic resonance imaging(MRI), as a fundamental technique in clinical cardiac MRI, encompass both functional and histological features of the heart. However, these features are often embedded within a high-dimensional feature space, which is difficult to visualize the myocardial lesions. Therefore, the synthesis of T1 mapping from Cine MRI is an effective method for visualizing cardiomyopathy. In this study, we propose a novel method to synthesis T1 mapping based on disentangled conditional denoising diffusion probabilistic model(T1Diff). Leveraging a modality encoder and an anatomical structure encoder to extract both the modality features and morphokinetic feature from Cine MRI. Then guiding the synthesis of high-quality T1 mapping from diffusion model. Which enables the extraction and visualization of myocardial lesion features from Cine MRI, thus transform complex Cine MRI into clinically visible. Through quantitative and qualitative analysis, T1Diff has demonstrated high concordance with native T1 mapping in terms of image quality and the visuospatial distribution of myocardial scarring.

Introduction

Cardiac magnetic resonance imaging (CMR) is crucial for assessing cardiac structure and function, with T1 mapping being particularly notable. It precisely measures myocardial tissue relaxation times, providing quantitative insights into microstructural changes relevant to various cardiac conditions. T1 mapping is essential for evaluating diffuse myocardial fibrosis, showing promise in distinguishing relaxation times between pathological and normal myocardium (Everett et al. 2016; Sparrow et al. 2006). Native T1 values accurately discriminate between normal and diseased myocardium (Puntmann et al. 2013). In myocardial infarction-related diseases, prolonged myocardial T1 and increased extracellular volume suggest diffuse fibrosis, correlating with post-infarction myocardial remodeling (Messroghli et al. 2007). However, T1 scanning can be costly and time-consuming, potentially causing discomfort for patients and leading to underutilization in clinical practice. In MRI scanning, the Cine MRI technique is a mature, readily accessible, and cost-effective method that yields abundant datasets. In the realm of clinical diagnostics, T1 mapping and cine MR complement each other: T1 mapping focuses on the quantitative analysis of myocardial tissue characteristics, while Cine MRI contains basic structural information about the patient's heart and include complete dynamic information about the cardiac cycle. However, the features of the myocardium are located within a high-dimensional space in the cine sequence, which are not easily accessible for direct clinical application. To address these issues, We developed T1Diff to synthesize T1 mapping from Cine MRI, extracting and visualizing complex myocardial features. Chartsias et al. (Chartsias et al. 2019) proposed that Cine MRI naturally disentangled into some spatial factors depicting anatomy and factors that denote the imaging characteristics.Consequently, we have extracted modality features and morphokinetic characteristics from Cine MRI to construct disentangled conditions. Ultimately, we utilize a conditional diffusion model to synthesize these elements into high-fidelity T1 mapping images.

Extensive experiments conducted on actual clinical datasets have demonstrated that T1Diff exhibit significant concordance with native T1 mapping in terms of visuospatial distribution and quantitative analysis of infarct lesions. In cases where native T1 signals are weak, the image quality of T1Diff synthesized images may even surpass that of native T1 imaging. To our knowledge, this represents the first instance of synthesizing T1 mapping from Cine MRI, offering a fresh perspective in the field of cardiac magnetic resonance modality synthesis.

Our contributions are three fold. Firstly, we are the first to synthesize T1 mapping from Cine MRI and generate high-fidelity images based on a conditional diffusion model. Secondly, we extracted modality features and morphokinetic features from Cine MRI to construct disentangled conditions, ensuring that the synthesized T1 mapping maintains a high degree of visual and spatial concordance with native T1 mapping in representing myocardial infarction scars. Thirdly, we validated the proposed method on a clinical dataset, where our approach demonstrated superior performance in both image quality and the accurate mapping of myocardial infarction lesions compared to existing technologies.

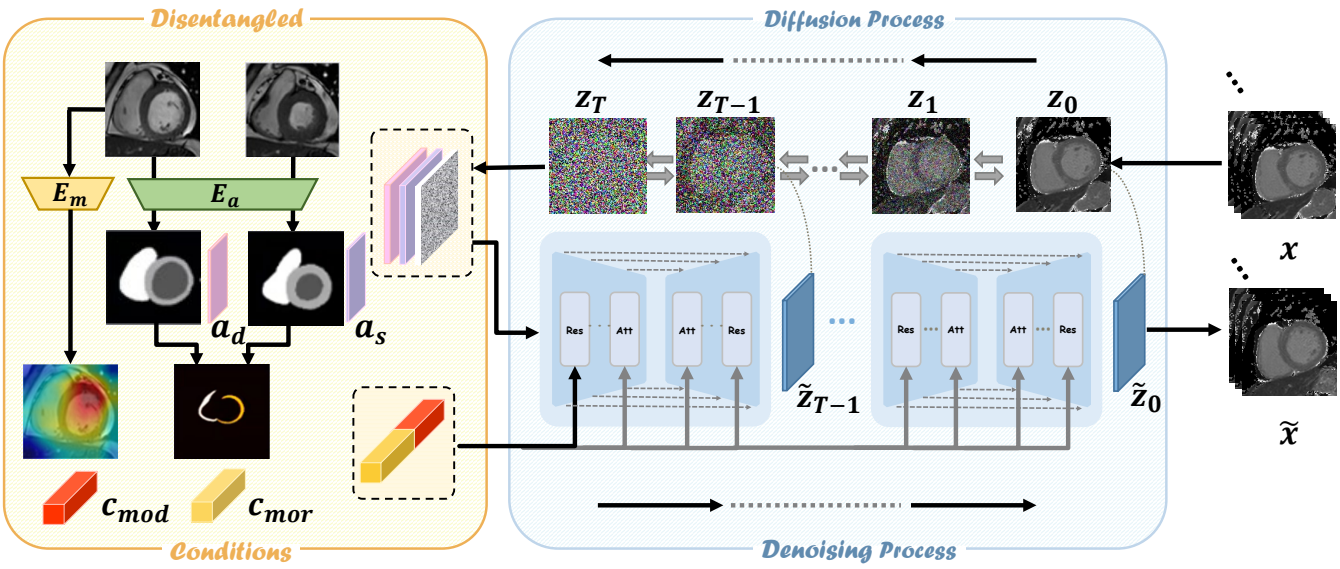

Figure 1: Framework for T1Diff. The conditions mainly consist of a modality encoder and an anatomical encoder, which are respectively responsible for extracting texture features and morphological features from Cine MRI. The diffusion module depicts the cross modal conditional diffusion and reverse generation process.

Related Works

Building an image synthesis model that is capable of image conversion between multi-modality medical images has been a longstanding desire for the artificial intelligence community. We present a brief overview of recent endeavors in this direction.

Diffusion Models (DMs). As a new paradigm for generative models, DMs have brought about a wave of research innovations for their realistic and diverse outputs (Saharia et al. 2022; Austin et al. 2021; Dhariwal and Nichol 2021; Ho and Salimans 2021). The pioneering work DDPM (Ho, Jain, and Abbeel 2020) firstly introduced the principle of DMs, which converts a simple known distribution into a target distribution using a generative Markov chain. Subsequent efforts extend this method to a broader spectrum, such as DDIM (Song, Meng, and Ermon 2020) introduced a method for conditional generation by deterministically sampling noisy hidden variables, and the score-based SDE (Song et al. 2020) demonstrated that conditional generation can be achieved by solving a conditional reverse-time SDE, thereby advancing the conditional generation field. Nichol et al. (Nichol et al. 2022), Bansal et al. (Bansal et al. 2023) and Liu et al. (Liu et al. 2023) then extended category conditions to encompass image, text, and multi-modal conditional generations. Choi et al. (Choi et al. 2021) proposed incorporating reference images during the sampling process to guide the reconstruction of the image, thereby ensuring its similarity to the reference image. In contrast to (Choi et al. 2021), we not only utilize the static shape features found in a single reference image but also integrate the dynamic morion information obtained throughout the entire cine sequence.

Cross-Modality Translation. Image translation enhances medical image analysis by bridging modalities like MRI, CT, and PET, expanding the range of available imaging data (Zhou et al. 2020; Liu et al. 2021; Pan et al. 2018). Most existing attempts to apply diffusion-based models in MRI translation involve using an encoder to acquire the latent representation of the source modality, which is then integrated with the noise estimator to enable conditional generation. Lyu and Wang (Lyu and Wang 2022) investigated both DDPM and SDE methods for generating synthetic CT images from given T2-weighted MRI. Li et al. (Li et al. 2023b) integrated MRI image-guided DM with a regularization term based on Range-Null Space Decomposed CT measurements during sampling to generate high-quality CT images from MRI. Xie et al. (Xie et al. 2022) suggested using the joint probability distribution of the diffusion model to synthesize brain PET images, with ultrahigh field MRI as the guidance. Meng et al. (Meng et al. 2022) proposed the Unified Multi-Modal Conditional Score-based Generative Model, extending SGM to cross-modal conditional synthesis for various missing-modality configurations in a unified framework. Taking inspiration from these studies, we encoded Cine MRI as a conditions acting on the middle layers of the noise estimator in synthesizing T1 mapping.

Method

The proposed framework is illustrated in Fig. 1. T1Diff mainly composed of disentangled conditions and a diffusion model. It is used to extract modality and morphological features from Cine MRI. This framework aims too achieve the synthesis and visualize of T1 mapping through cross modal conditional diffusion and reverse generation processes.

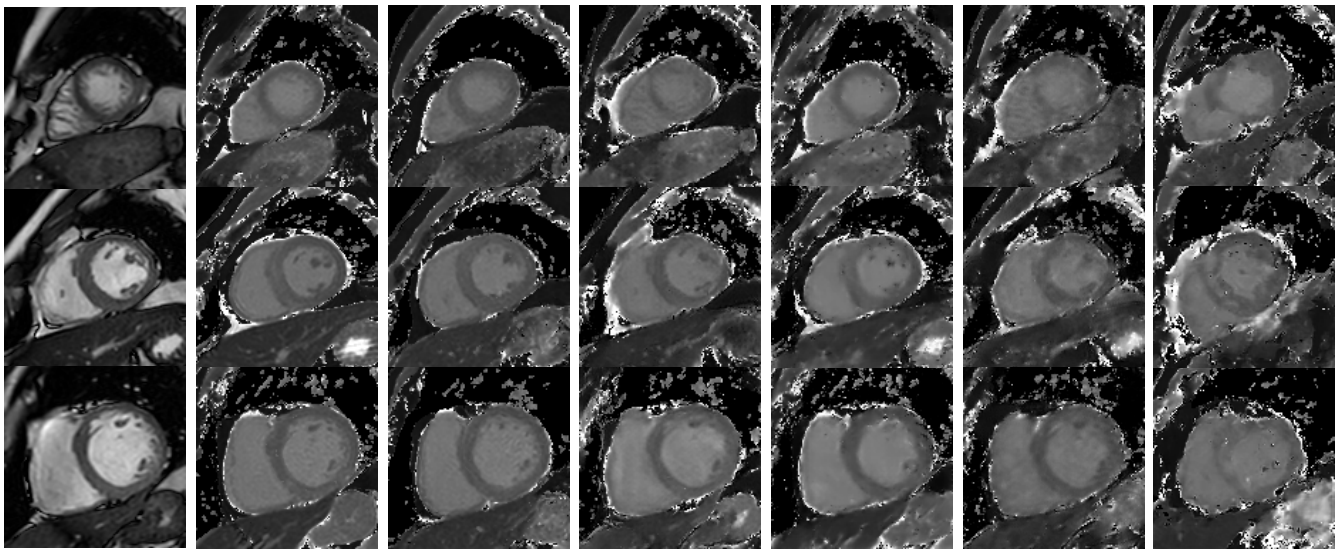

Figure 2: Visual comparative of synthesized T1 Mapping. The leftmost column displays images from cardiac cine MRI at ED. The Reference is the native T1 mapping, whereas the images synthesized by T1Diff are situated in the third column.

Disentangled Conditions

Cine MRI is capable of capturing comprehensive structural information of the heart and the subtle tissue and morphokinetic characteristics of the myocardium. Nevertheless, directly interpreting the health status of the myocardium from these images remains a challenge, as the signs of myocardial health are often concealed within the complex textures and dynamic changes of the images. Current research on cross-modality translation largely focuses on direct transformation between different imaging modalities, with less attention paid to the deep-seated features embedded within these modalities, which are crucial for understanding and visualizing the health status of the myocardium.

To capture the deep-seated features within Cine MRI, we factorize Cine MRI into non-spatial modality factors and spatial anatomical factors. As illustrated in Fig 1, we have introduced a simple U-net as Modality Encode E_m, to extract histological and textural features c_{mod} of the myocardium from Cine MRI. In addition, we have employed an Anatomical Structure Encoder E_a specifically for extracting pivotal anatomical information a_d and a_s during the end-diastolic(ED) and end-systolic(ES) phases of Cine MRI. Our method is not confined to any specific anatomical structure encoder, thus we have opted for a parsimonious, frequency-domain-based network model GF-Unet (Li et al. 2023a) to distill anatomical structures. By precisely capturing anatomical information to delineate cardiac morphokinetic features c_{mor}, we have not only enhanced the representation of cardiac morphology but also described the dynamic behavior of the myocardium throughout the cardiac cycle, effectively circumventing the interference that texture artifacts might introduce. In Fig. 3, Cmod represents the feature map post Cine encoding, while Cmor signifies the differences in anatomical structures from ED to ES stages due to motion. By combining Cmod and Cmor as Disentangled conditions, we can generate T1 that accurately reflect myocardial infarction scars.

Furthermore, we concatenate these modality and morphokinetic features to construct disentangled conditions, which serve as the conditional input for the DDPM. The novelty of this approach lies in its enhancement of the quality of information transfer between modalities, and by leveraging both modality and morphokinetic information comprehensively, it enables the synthesized T1 mapping to depict the myocardial health conditions with greater accuracy and completeness.

Conditional Denoising Diffusion Probability Model

DDPM defines a forward Markovian diffusion process q that gradually adds Gaussian noise to T1 mapping z_0 over T iterations:

$$\begin{aligned} &q\left(z_t \mid z_{t-1}\right)=\mathcal{N}\left(z_t ; \sqrt{1-\beta_t} \cdot z_{t-1}, \beta_t \cdot \mathbf{I}\right), \\ &\forall t \in\{1,2, \ldots, T\} \end{aligned} \tag{1}$$

where the parameters T, $\beta_1, \beta_2, \ldots, \beta_T \in [0, 1)$ represent the number of diffusion steps and the variance schedule across diffusion steps, respectively. $\mathbf{I}$ is the identity matrix and $\mathcal{N} \sim (x; \mu, \sigma)$ represents the normal distribution of mean μ and covariance σ. Considering $\alpha_t = 1 - \beta_t$ and $\bar{\alpha_t} = \prod_{s=0}^{T} \alpha_s$ one can directly sample an arbitrary step of the noised latent conditioned on the input x_0 as follows:

$$z_t = \sqrt{\bar{\alpha_t}} z_0 + \sqrt{1 - \bar{\alpha_t}} \epsilon \tag{2}$$

Leveraging the above definitions, approximating a reverse process to get a sample from $q(z_0)$. To this end, we can

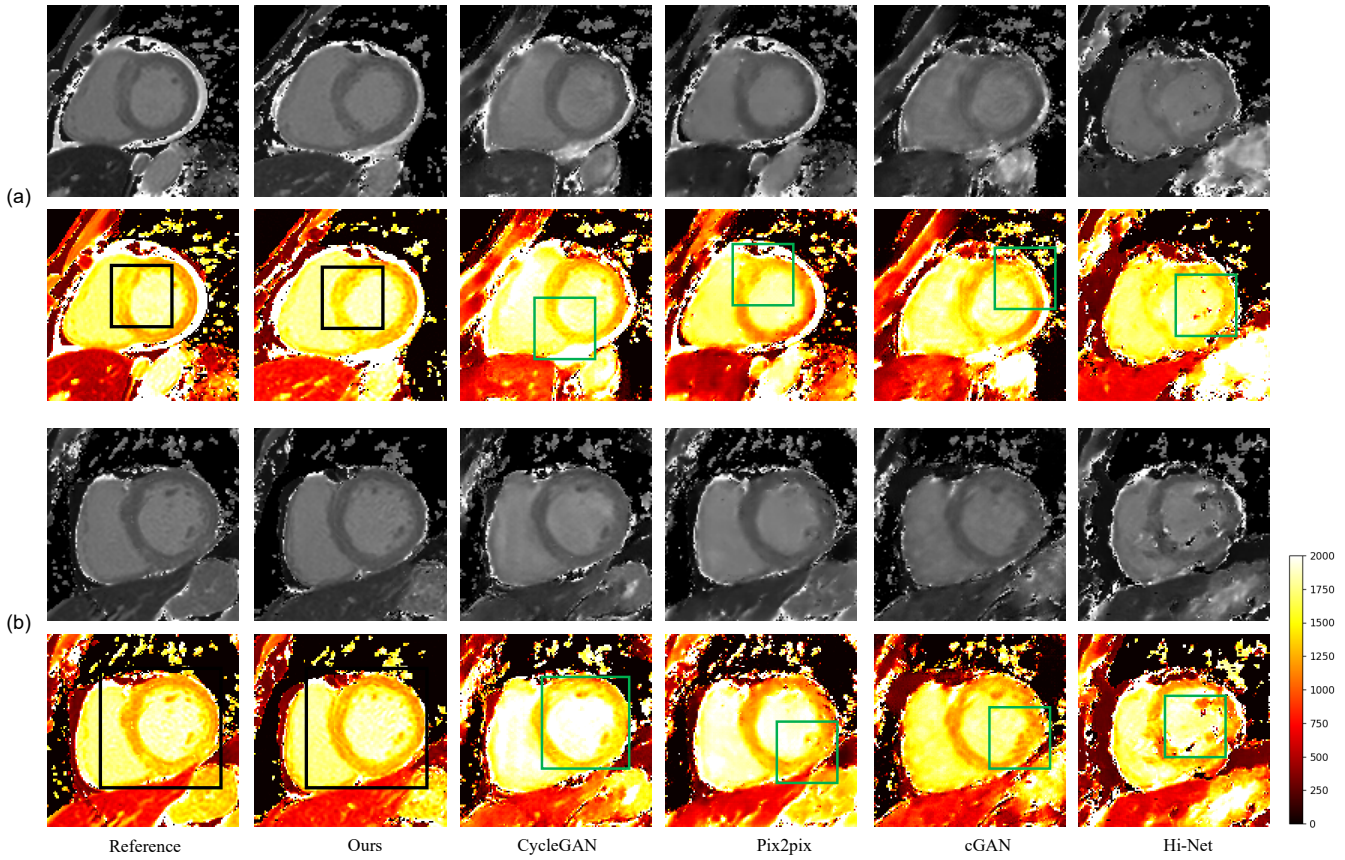

Figure 3: Comparative analysis of T1 Mapping synthesized by different methods using heatmaps: (a) displays images from a patient with myocardial infarction scarring; (b) presents cardiac images from a healthy patient. In the figure, black boxes indicate areas where our proposed method achieves high congruence with native T1 mapping images in visualizing the myocardium, while green boxes highlight the inconsistencies between myocardial visualizations by other methods and the reference images.

parameterize this reverse process by starting at $p_\theta(z_T) = \mathcal{N}(z_T; \mu, \sigma)$ as follows:

$$p_\theta(z_{0:T}) = p(z_T) \prod_{t=1}^{T} p_\theta\left(z_{t-1} \mid z_t, c_{\text{mod}}, c_{\text{mor}}\right) \tag{3}$$

$$\begin{aligned} &p_\theta(z_{t-1} \mid z_t) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t, c_{\text{mod}}, c_{\text{mor}}), \\ &\mathbf{\Sigma}_\theta(z_t, t, c_{\text{mod}}, c_{\text{mor}})\mathbf{I}) \end{aligned} \tag{4}$$

To train this model in a way that allows $p(z_0)$ to learn the true data distribution $q(z_0)$, we can optimize the following variational bound on negative log-likelihood:

$$\begin{aligned} &\mathbb{E}\left[-\log p_\theta\left(z_0\right)\right] \leq \mathbb{E}_q[-\log p\left(z_T\right) - \\ &\sum_{t \geq 1} \log \frac{p_\theta\left(z_{t-1} \mid z_t, c_{\text{mod}}, c_{\text{mor}}\right)}{q\left(z_t, c_{\text{mod}}, c_{\text{mor}} \mid z_{t-1}\right)}] = -L_{\text{VL·B}} \end{aligned} \tag{5}$$

In practical applications, one simply needs to train a model $\epsilon_\theta\left(x_t, t, c_{\text{mod}}, c_{\text{mor}}\right)$ to predict ϵ. Hence, by reparameterizing Eq 6, a simplified objective as follows:

$$\mathcal{L}(\theta) := \mathrm{E}_{\mathrm{t}, z_0, \epsilon}\left[\left\|\epsilon - \epsilon_\theta\left(z_t, t, c_{\text{mod}}, c_{\text{mor}}\right)\right\|^2\right] \tag{6}$$

Once the neural network training is completed, start with random noise $z_T \sim \mathcal{N}(0; \mathbf{I})$, and utilize the neural network to follow the current step T. Predict the noise at step $T-1$ using z_T and compute z_{T-1} based on the predicted mean and variance. Continue this iterative process until $t = 0$ to synthesize the T1 mapping.

Experiments

Dataset and Training Details

To assess the T1Diff model, We have collected a multi-center dataset from multiple hospitals and are working diligently to expand its scale to encompass a variety of diseases. Data preparation is both complex and time-consuming. To validate our idea, we initially selected myocardial infarction patients' MRIs for experimentation, as T1 mapping can effectively display infarct scars. Dataset comprising anonymized cardiac MRI scans from 381 patients diagnosed with acute myocardial infarction. These scans were acquired using a 3.0-T scanner (Ingenia, Philips). The dataset included native T1 mapping (256×256×3) at end-diastole (ED) and Cine MRI sequences (256×256×540) capturing the entire cardiac cycle for each patient. Following a rig-

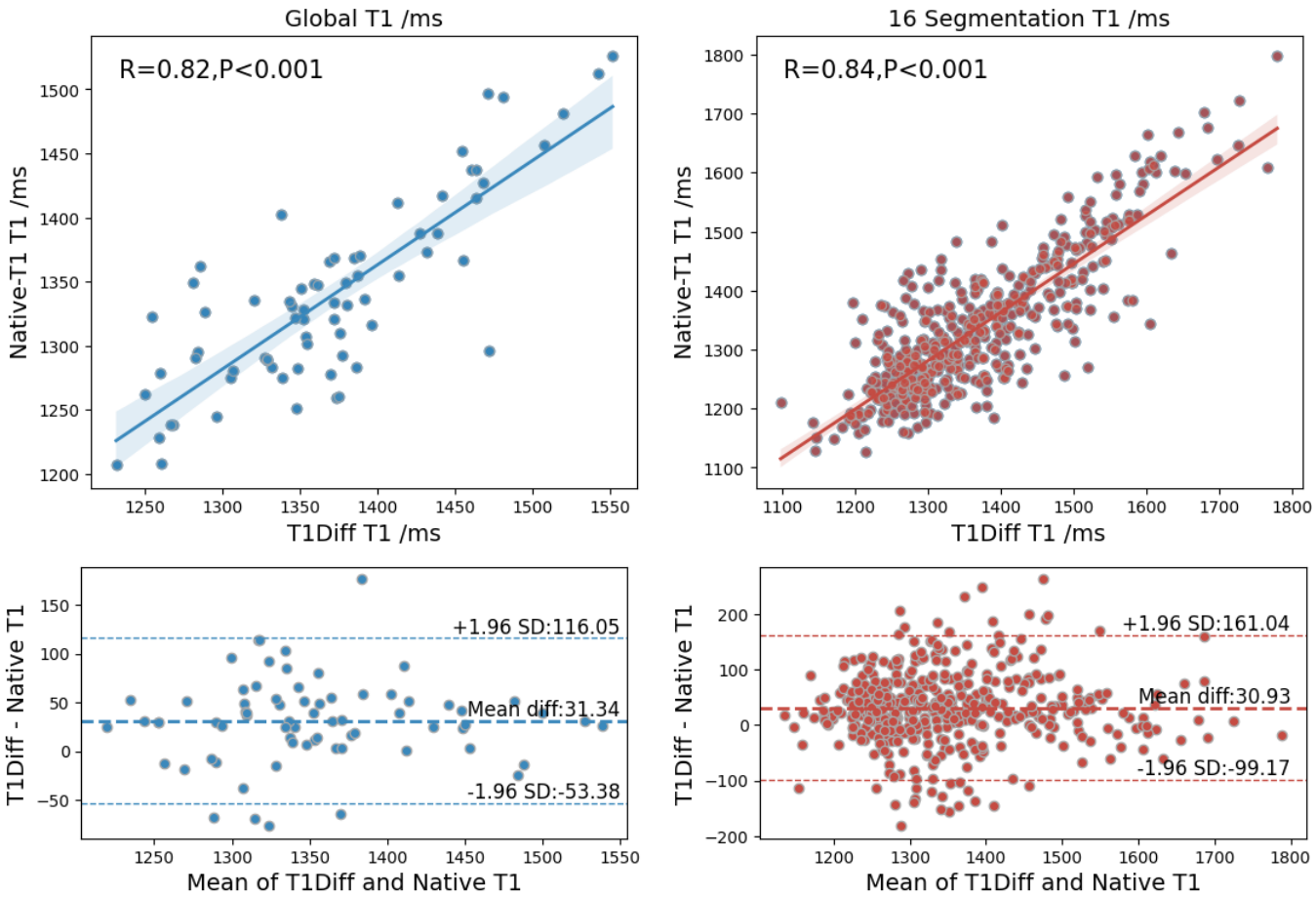

Figure 4: T1Diff correlated strongly with Native T1 in global T1 value (left) and in American Heart Association(AHA) 16 segmentation T1 value (reight) in test patients. Top, Correlation plots with the linear regression equations, Pearson correlation coefficients (R), statistical significance of correlation (P value). Bottom, Bland-Altman plots to analyze any systematic differences between quantification by T1Diff and Native T1.

orous quality control protocol, which involved the exclusion of scans with suboptimal image quality and inconsistent slice positions, a final cohort of 313 patient Cine-T1 mapping pairs was curated for the training and validation of the T1Diff model. Each patient has three pairs of Cine-T1 MRI image sets, which are located at the apex, base, and papillary muscle regions of the heart, respectively. The dataset was partitioned randomly, allocating 90% (846 image pairs from 282 patients) for training and the remaining 10% (93 image pairs from 31 patients) for testing. To register Cine and T1, the clinical doctor first marks the ED and ES in Cine. Then, we use the "slice position" in DICOM information to align Cine and T1. Subsequently, through segmentation, we aligned the left ventricles of Cine and T1 and extracted a 128×128 ROI to accelerate the training process and enhance the fidelity of the heart region in the generated images. The T1Diff model was developed using the PyTorch framework and trained on a workstation equipped with an NVIDIA Tesla V100 GPU, which boasts 32 GB of VRAM. Training commenced from an initial random state, deliberately eschewing the use of any pre-trained models to ensure the model's performance was solely attributable to our methodology. The training regimen spanned a continuous period of 72 hours, utilizing the Adam optimization algorithm with a diffusion step parameter set at 1000.

	PSNR ↑	SSIM ↑	NMSE ↓
Hi-Net	27.76 ± 0.419	0.82 ± 0.079	1.43 ± 0.085
cGAN	28.58 ± 0.481	0.82 ± 0.083	0.99 ± 0.124
CycleGan	29.23 ± 0.326	0.85 ± 0.072	0.88 ± 0.087
pix2pix	30.21 ± 0.522	0.87 ± 0.029	0.78 ± 0.055
Ours	$\mathbf{31.84 \pm 0.452}$	$\mathbf{0.89 \pm 0.061}$	$\mathbf{0.66 \pm 0.101}$

Table 1: In evaluating the performance of various cross-modality methods in comparison with T1Diff in terms of image quality, we adopted three key metrics(mean ± std): PSNR, SSIM, and RMSE. In the comparative results, the metrics that demonstrate optimal performance will be highlighted in bold.

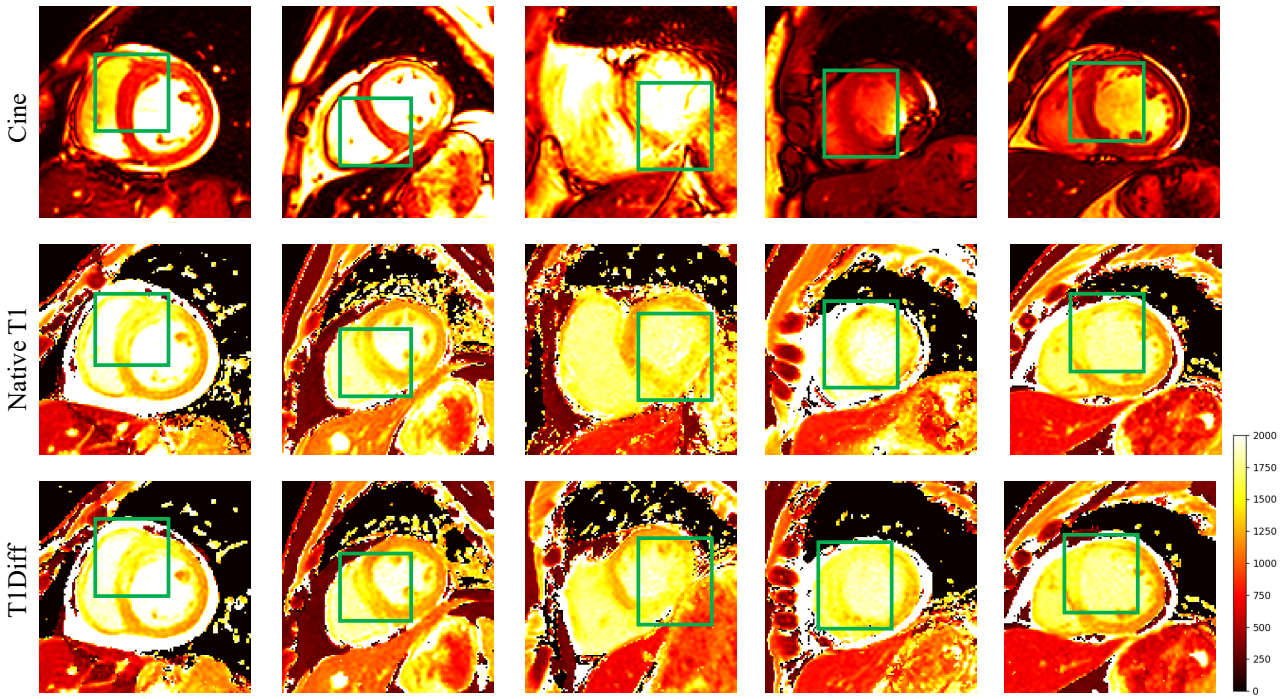

Figure 5: Cine MRI, real T1 mapping, and T1Diff heat maps in the illustrations of some patients with acute myocardial infarction. The green box corresponds to the location of the infarcted scar. It can be observed that it is challenging to identify the location of infarct scars in Cine images; Even if some algorithms can tell that a patient is likely to suffer a myocardial infarction, the exact location of the scar is difficult to clearly show. In contrast, both the native T1 mapping and the T1Diff can clearly reveal the exact location of the infarct scar.

Evaluation Metrics of Synthesis Quality

We conducted a comparison between T1Diff and several typical cross-modal translation methods. Among them are CycleGan (Zhu et al. 2017) and Pix2pix (Isola et al. 2017), which have demonstrated impressive performance on multi-modal datasets. Additionally, we evaluated two specialized cross-modal methods for medical data, cGAN (Dar et al. 2019) and Hi-Net (Zhou et al. 2020). To assess the performance of these methods, we primarily employed three evaluation metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Root Mean Square Error (RMSE). These metrics offer quantitative measures to evaluate image quality and similarity to reference images. By analyzing the results obtained from the evaluation metrics, we were able to gain insights into the comparative performance of T1Diff and the other cross-modal translation methods.

As Table 1 demonstrates, our method surpasses all comparison group methods across three key metrics, highlighting the verisimilitude of our generated images. Given the relative difficulty in acquiring corresponding data in clinical settings, the fact that our method achieves exceptional performance on a smaller dataset further confirms the significant advantages of our proposed approach. As shown in Fig. 2, T1Diff reveals richer myocardial structural details and clearer textural features, with image quality that is comparable to native T1 mapping. This accomplishment is attributable to our Modality Encoder and Anatomical Structure Encoder, which provide more precise contours and histological information for the synthesized T1 mapping. Observing the synthesized T1 mapping images, one can clearly discern that the myocardial edge contours generated by T1Diff are sharper than those produced by other methods, aiding clinicians in more clearly identifying the health status of the myocardium.

Evaluation of Myocardial Infarction Scar

Despite the synthesis of T1 mapping by other methods mentioned in this paper that subjectively resemble those produced by T1Diff, the precise assessment of myocardial segments in the synthesized images is particularly crucial given the high diagnostic accuracy of T1 mapping in differentiating normal from pathological myocardial tissue. As demonstrated in Fig. 3(a), which is a T1 mapping image of a patient with myocardial infarction, the heatmaps reveal that T1Diff effectively visualizes myocardial infarction scars in the anterior wall region, similar to native T1 mapping, while other methods such as CycleGAN and Pix2pix exhibit inaccuracies in scar localization. In Fig. 3(b), which show a healthy individual's native T1 mapping, the myocardial texture quality displayed by T1Diff is on par with the native T1 mapping, whereas other methods display some artifacts or abnormal representations of myocardial tissue.

Although Late Gadolinium Enhancement(LGE) is the gold standard for diagnosing myocardial infarction, its ap-

plication is limited by equipment differences and operator skills, making it difficult to achieve quantitative analysis. In contrast, T1 mapping can effectively diagnose infarcted myocardium through quantitative T1 value analysis, filling the limitations of LGE. In Fig. 4, T1Diff showed strong identity correlation with Native T1 in global T1 value ($P<0.001$) and American Heart Association(AHA) 16 segmentation T1 value ($P<0.001$) of 31 patients. The high agreement in both global T1 value (R=0.82) and AHA 16 segmentation T1 value(R=0.84) supports the promise of T1Diff to replace Native T1 for noninvasive and contrast-free myocardial scar assessment. Bland-Altman plots (Fig. 4) showed excellent accuracy with high precision and no significant bias.

The heatmap shown in Fig. 5 provides us with an important perspective on the localization and recognition of myocardial infarction areas using T1Diff technology. This imaging technique represents the changes in T1 values of myocardial tissue through different color intensities, with the areas with the highest brightness, i.e. the most significant increase in T1 values, accurately depicting the infarcted area. In healthy myocardial tissue, T1 values are typically maintained at around 1200 levels, but in damaged infarcted areas, this value significantly increases, allowing us to clearly distinguish between healthy and damaged tissues. Therefore, whether from a quantitative or qualitative analysis perspective, T1Diff can be effectively used for the diagnosis of myocardial infarction like Native T1.

Ablation Experiments

Table 2 presents the results from a series of ablation experiments designed to validate the effectiveness of incorporating Disentangled Conditions and DDPM in our framework. Each row represents a different configuration of the model, highlighting the impact of disentangled conditions and DDPM on the overall performance metrics.

	PSNR ↑	SSIM ↑	NMSE ↓
GAN w/ Dis C	29.32 ± 0.27	0.85 ± 0.87	1.22 ± 0.41
DDPM w/o Dis C	29.15 ± 0.43	0.84 ± 0.35	1.12 ± 0.13
Ours	$\mathbf{31.84 \pm 0.45}$	$\mathbf{0.89 \pm 0.06}$	$\mathbf{0.66 \pm 0.10}$

Table 2: Ablation studies demonstrating the effectiveness of Disentangled Conditions(Dis C) and DDPM, the metrics that demonstrate optimal performance will be highlighted in bold.

Firstly, the GAN with Disentangled Conditions experiment focused on evaluating the impact of disentangling conditions within the generative adversarial network framework. This setup allowed us to observe how separating various conditioning factors could enhance the model's ability to generate more diverse and accurate representations, by providing a more structured and interpretable latent space. The effectiveness of this approach was measured against a baseline GAN model, highlighting the improvements in generation quality and fidelity when conditions are disentangled.

Secondly, the DDPM without Disentangled Conditions experiment was designed to assess the performance of the denoising diffusion probabilistic model in the absence of disentangled conditioning. This setup helped in understanding the intrinsic capabilities of DDPM in generating high-quality outputs and its robustness across different settings. By comparing this configuration with the complete model, including both DDPM and disentangled conditions, we could evaluate the contribution of DDPM's probabilistic generation process to the overall system.

Conclusion

In this study, we introduce an innovative approach to T1 mapping synthesis, which we have named T1Diff. This method is at the forefront of medical imaging technology, leveraging the power of disentangled conditional diffusion models to transform modality and morphokinetic features extracted from Cine MRI into high-quality cardiac T1 mapping images. The core innovation of T1Diff lies in its ability to accurately simulate T1 mapping images, a critical component in CMR imaging, by analyzing and utilizing the dynamic and structural information present in Cine MRI sequences. This approach addresses several challenges in the current CMR imaging landscape. Firstly, it significantly reduces the need for multiple scanning sessions to capture different modalities, thereby bridging the gap between CMR modalities in a seamless and efficient manner. This synthesis capability is particularly valuable in clinical settings where time and resources are limited.Furthermore, T1Diff can substantially decrease the duration of MRI scans. This reduction in scan time directly translates to less time that patients need to spend in the MRI scanner, thereby alleviating patient discomfort and anxiety. The shorter scan times can also increase the throughput of MRI facilities, allowing more patients to be scanned per day, which is beneficial for both patients and healthcare providers. In addition to these patient-centered benefits, T1Diff streamlines the workflow for clinical practitioners. By synthesizing T1 mapping images from existing Cine MRI data, it enables clinicians to obtain a comprehensive set of CMR modalities without additional scanning, thereby facilitating a more efficient diagnostic process. This streamlined workflow can enhance the overall efficiency of cardiac imaging departments, allowing for quicker diagnosis and treatment planning.

References

Austin, J.; Johnson, D. D.; Ho, J.; Tarlow, D.; and Van Den Berg, R. 2021. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34: 17981–17993.

Bansal, A.; Chu, H.-M.; Schwarzschild, A.; Sengupta, S.; Goldblum, M.; Geiping, J.; and Goldstein, T. 2023. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 843–852.

Chartsias, A.; Joyce, T.; Papanastasiou, G.; Semple, S.; Williams, M.; Newby, D. E.; Dharmakumar, R.; and Tsaftaris, S. A. 2019. Disentangled representation learning in cardiac image analysis. *Medical image analysis*, 58: 101535.

Choi, J.; Kim, S.; Jeong, Y.; Gwon, Y.; and Yoon, S. 2021. ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14367–14376.

Dar, S. U.; Yurt, M.; Karacan, L.; Erdem, A.; Erdem, E.; and Cukur, T. 2019. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE transactions on medical imaging*, 38(10): 2375–2388.

Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34: 8780–8794.

Everett, R.; Stirrat, C.; Semple, S.; Newby, D.; Dweck, M.; and Mirsadraee, S. 2016. Assessment of myocardial fibrosis with T1 mapping MRI. *Clinical radiology*, 71(8): 768–778.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Ho, J.; and Salimans, T. 2021. Classifier-Free Diffusion Guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Li, P.; Zhou, R.; He, J.; Zhao, S.; and Tian, Y. 2023a. A global-frequency-domain network for medical image segmentation. *Computers in Biology and Medicine*, 164: 107290.

Li, X.; Shang, K.; Wang, G.; and Butala, M. D. 2023b. DDMM-Synth: A Denoising Diffusion Model for Cross-modal Medical Image Synthesis with Sparse-view Measurement Embedding. *arXiv preprint arXiv:2303.15770*.

Liu, X.; Park, D. H.; Azadi, S.; Zhang, G.; Chopikyan, A.; Hu, Y.; Shi, H.; Rohrbach, A.; and Darrell, T. 2023. More control for free! image synthesis with semantic diffusion guidance. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 289–299.

Liu, X.; Xing, F.; El Fakhri, G.; and Woo, J. 2021. A unified conditional disentanglement framework for multimodal brain mr image translation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 10–14. IEEE.

Lyu, Q.; and Wang, G. 2022. Conversion between ct and mri images using diffusion and score-matching models. *arXiv preprint arXiv:2209.12104*.

Meng, X.; Gu, Y.; Pan, Y.; Wang, N.; Xue, P.; Lu, M.; He, X.; Zhan, Y.; and Shen, D. 2022. A novel unified conditional score-based generative framework for multi-modal medical image completion. *arXiv preprint arXiv:2207.03430*.

Messroghli, D. R.; Walters, K.; Plein, S.; Sparrow, P.; Friedrich, M. G.; Ridgway, J. P.; and Sivananthan, M. U. 2007. Myocardial T1 mapping: application to patients with acute and chronic myocardial infarction. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(1): 34–40.

Nichol, A. Q.; Dhariwal, P.; Ramesh, A.; Shyam, P.; Mishkin, P.; Mcgrew, B.; Sutskever, I.; and Chen, M. 2022. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In *International Conference on Machine Learning*, 16784–16804. PMLR.

Pan, Y.; Liu, M.; Lian, C.; Zhou, T.; Xia, Y.; and Shen, D. 2018. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III 11*, 455–463. Springer.

Puntmann, V. O.; Voigt, T.; Chen, Z.; Mayr, M.; Karim, R.; Rhode, K.; Pastor, A.; Carr-White, G.; Razavi, R.; Schaeffter, T.; et al. 2013. Native T1 mapping in differentiation of normal myocardium from diffuse disease in hypertrophic and dilated cardiomyopathy. *JACC: Cardiovascular Imaging*, 6(4): 475–484.

Saharia, C.; Chan, W.; Chang, H.; Lee, C.; Ho, J.; Salimans, T.; Fleet, D.; and Norouzi, M. 2022. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, 1–10.

Song, J.; Meng, C.; and Ermon, S. 2020. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2020. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*.

Sparrow, P.; Messroghli, D. R.; Reid, S.; Ridgway, J. P.; Bainbridge, G.; and Sivananthan, M. U. 2006. Myocardial T1 mapping for detection of left ventricular myocardial fibrosis in chronic aortic regurgitation: pilot study. *American Journal of Roentgenology*, 187(6): W630–W635.

Xie, T.; Cao, C.; Cui, Z.; Li, F.; Wei, Z.; Zhu, Y.; Li, Y.; Liang, D.; Jin, Q.; Chen, G.; et al. 2022. Brain pet synthesis from mri using joint probability distribution of diffusion model at ultrahigh fields. *arXiv preprint arXiv:2211.08901*.

Zhou, T.; Fu, H.; Chen, G.; Shen, J.; and Shao, L. 2020. Hi-net: hybrid-fusion network for multi-modal MR image synthesis. *IEEE transactions on medical imaging*, 39(9): 2772–2781.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*.

Reproducibility Checklist

For most authors...

- Includes a conceptual outline and/or pseudocode description of AI methods introduced (yes/partial/no/NA)? yes
- Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (yes/no)yes
- Provides well marked pedagogical references for less-familiare readers to gain background necessary to replicate the paper (yes/no) yes

Does this paper make theoretical contributions? (yes/no) yes

If yes, please complete the list below.

- All assumptions and restrictions are stated clearly and formally. (yes/partial/no)partial
- All novel claims are stated formally (e.g., in theorem statements). (yes/partial/no)no
- Proofs of all novel claims are included. (yes/partial/no)yes
- Proof sketches or intuitions are given for complex and/or novel results. (yes/partial/no)no
- Appropriate citations to theoretical tools used are given. (yes/partial/no)yes
- All theoretical claims are demonstrated empirically to hold. (yes/partial/no/NA)yes
- All experimental code used to eliminate or disprove claims is included. (yes/partial/no/NA)no

Does this paper rely on one or more datasets? (yes/no)yes
If yes, please complete the list below.

- A motivation is given for why the experiments are conducted on the selected datasets (yes/partial/no/NA)yes
- All novel datasets introduced in this paper are included in a data appendix. (yes/partial/no/NA)no
- All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (yes/partial/no/NA)no
- All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations. (yes/no/NA)no
- All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available. (yes/partial/no/NA)no
- All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisficing. (yes/partial/no/NA)yes

Does this paper include computational experiments? (yes/no)yes

If yes, please complete the list below.

- Any code required for pre-processing data is included in the appendix. (yes/partial/no).no
- All source code required for conducting and analyzing the experiments is included in a code appendix. (yes/partial/no)no
- All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (yes/partial/no)yes
- All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (yes/partial/no)partial
- If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results. (yes/partial/no/NA)yes
- This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks. (yes/partial/no)partial
- This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics. (yes/partial/no)yes
- This paper states the number of algorithm runs used to compute each reported result. (yes/no)no
- Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information. (yes/no)no
- The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank). (yes/partial/no)no
- This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments. (yes/partial/no/NA)partial
- This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting. (yes/partial/no/NA)no