Bidirectional Recurrence for Cardiac Motion Tracking with Gaussian Process Latent Coding

Jiewen Yang, Yiqun Lin, Bin Pu, Xiaomeng Li*

The Hong Kong University of Science and Technology

Code: https://github.com/xmed-lab/GPTrack

The Thirty-Eighth Annual Conference on Neural Information Processing Systems

Importance of Cardiac Motion Analysis

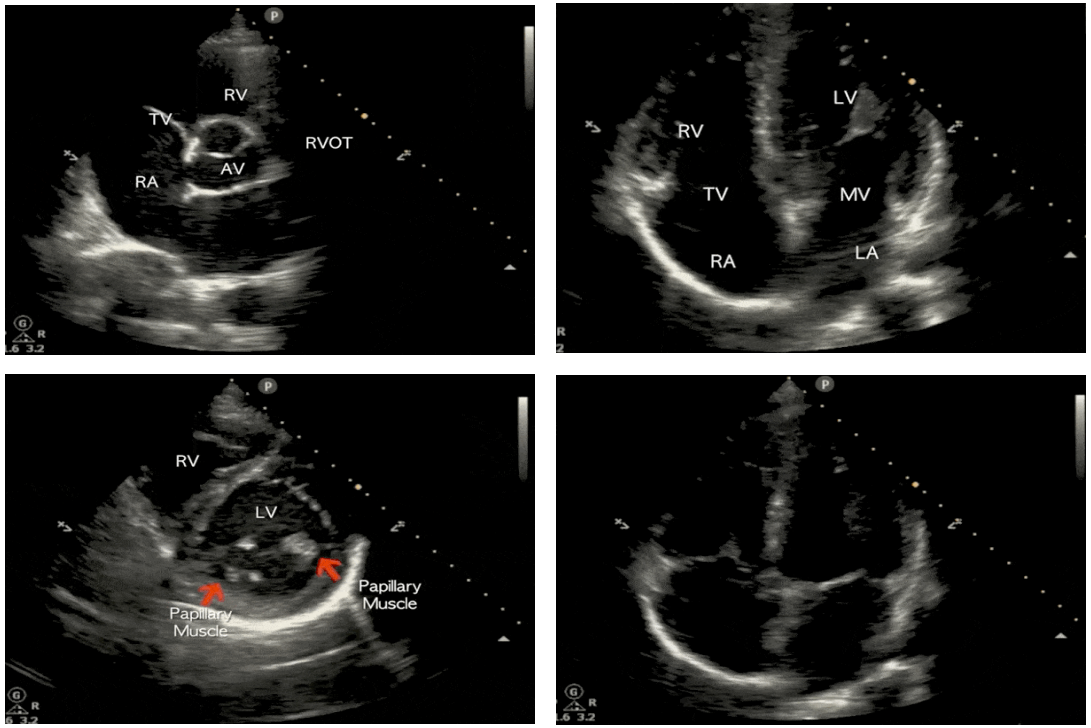


Figure 1 : Examples of Cardiac Motion Scanned via Echocardiogram [1].

Current medical imaging of the human heart will scan cardiac structures and their motion such as valves, vessels, ventricle and atrium.

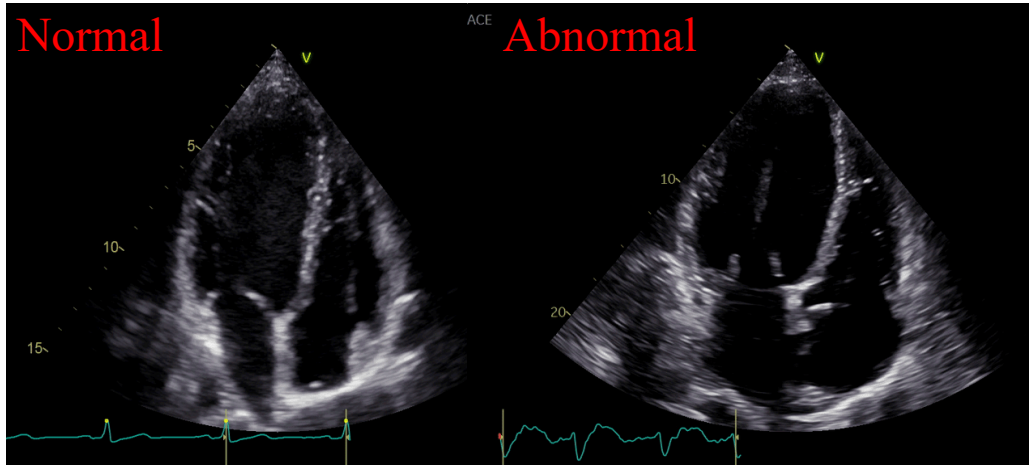


Figure 2 : Examples of Normal (Left) and Heart Failure (Right).

Cardiac motion analysis can help clinicians identify many cardiovascular diseases, such as Heart Failure.

[1] Cardiac Ultrasound (Echocardiography) Made Easy: Step-By-Step Guide, Vi Dinh, https://www.pocus101.com/cardiac-ultrasound-echocardiography-made-easy-step-by-step-guide/
[2] https://en.wikipedia.org/wiki/Ebstein%27s_anomaly [3] https://journals.sagepub.com/doi/10.1177/2150135117692973
[4] https://cardiovascularultrasound.biomedcentral.com/articles/10.1186/1476-7120-6-33

Challenges for Cardiac Motion Analysis

1. Previous methods lack consideration of the motion dynamics and spatial variability;

2. Previous methods often overlook the long-term relationships and regional motion characteristic.

3. It is necessary to leverage temporal and spatial information for cardiac motion tracking;

The Observation of Cardiac Motion

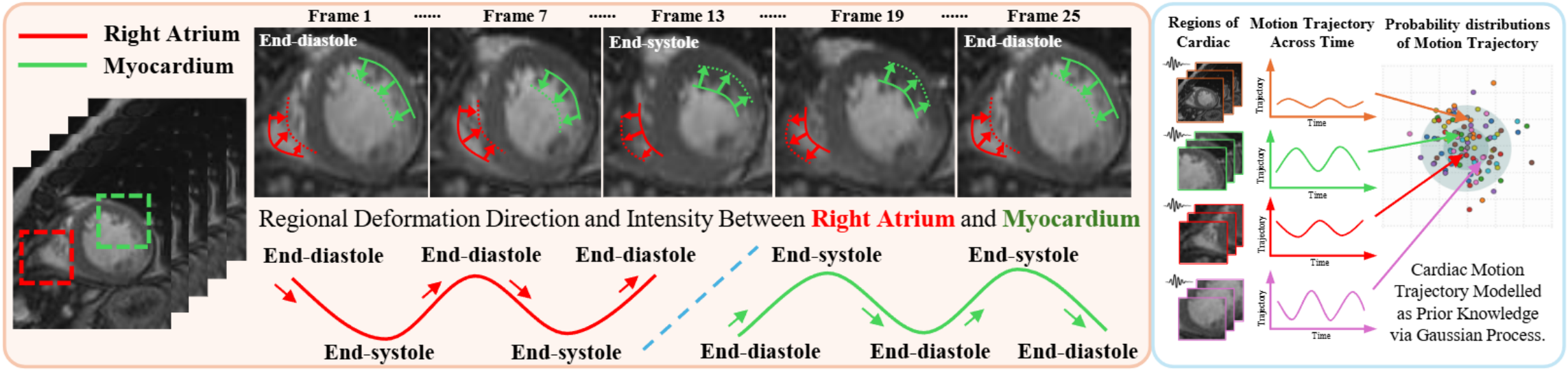


Figure 3 : The motivation of regional motion analysis and formulate cardiac motion as prior knowledge via Gaussian Process

Regional Motions in Cardiac: Regions of the Right Atrium (red) and Myocardium (green) performing the opposite trajectories during the heartbeat cycle.

The deformation is bounded in the space of periodically specific human cardiac motion variation, which means we are able to formulate **Cardiac Motion as a strong Prior Knowledge**.

A Novel Framework Named GPTrack for Cardiac Motion Analysis

Motivations

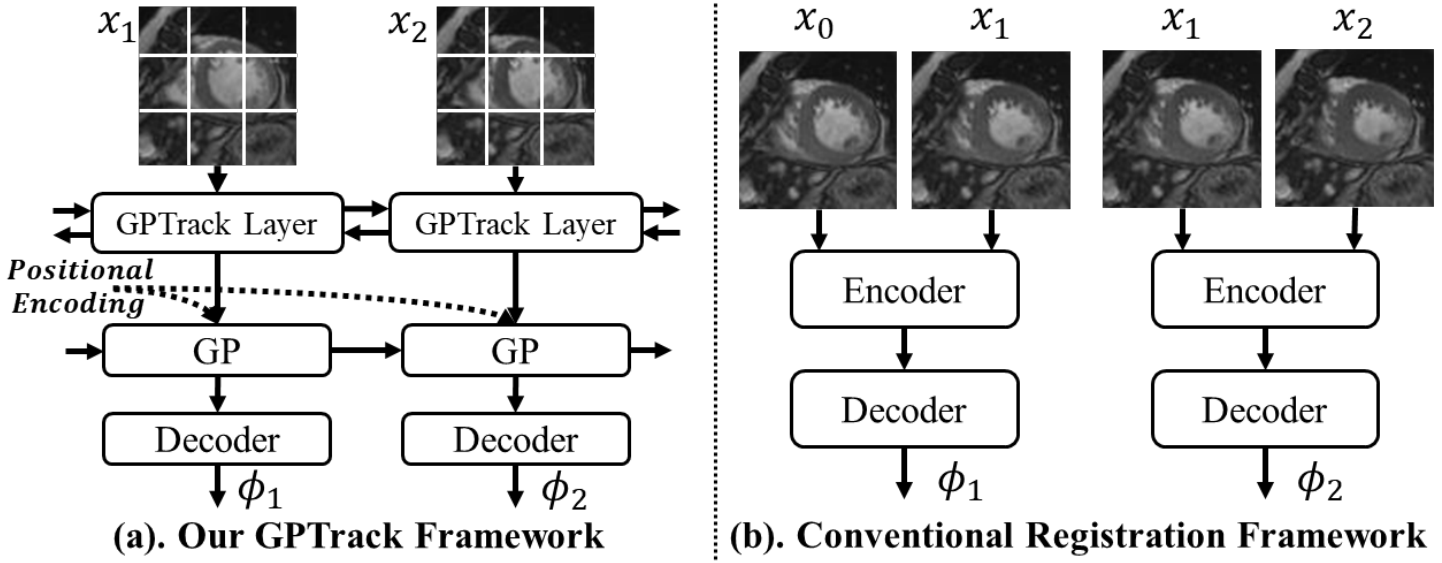
- Pervious studies focus **only the global understanding** of cardiac motion;
- Deformation is bounded in the space of **periodically specific human cardiac motion variation**;
- Enabling both **temporal and spatial understanding** for cardiac motion analysis.

Our Solutions

- We employ the **Gaussian Process (GP) to promote temporal consistency and regional variability incompact latent space**, establishing a robust regularizer to enhance cardiac motion tracking accuracy;
- GPTrack framework is designed to **capture the long-term relationship** of cardiac motion **via a bidirectional recursive manner**;

Difference Between GPTrack and Conventional Methods

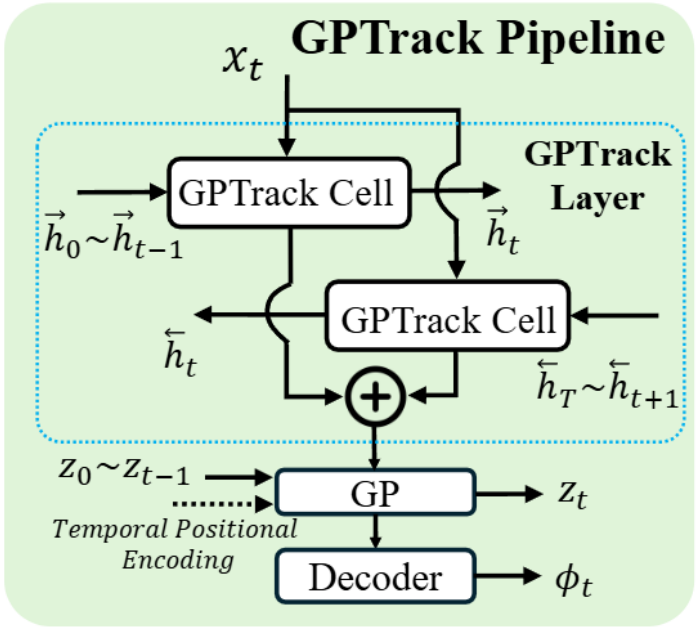
This framework employs the Gaussian Process (GP) to promote temporal consistency and regional variability incompact latent space, establishing a robust regularizer to enhance cardiac motion tracking accuracy.



1. GPTrack allows the registration network to aggregate the spatial information temporally, both forward and backward.
2. GPTrack considering the motion consistency between two adjacent state space

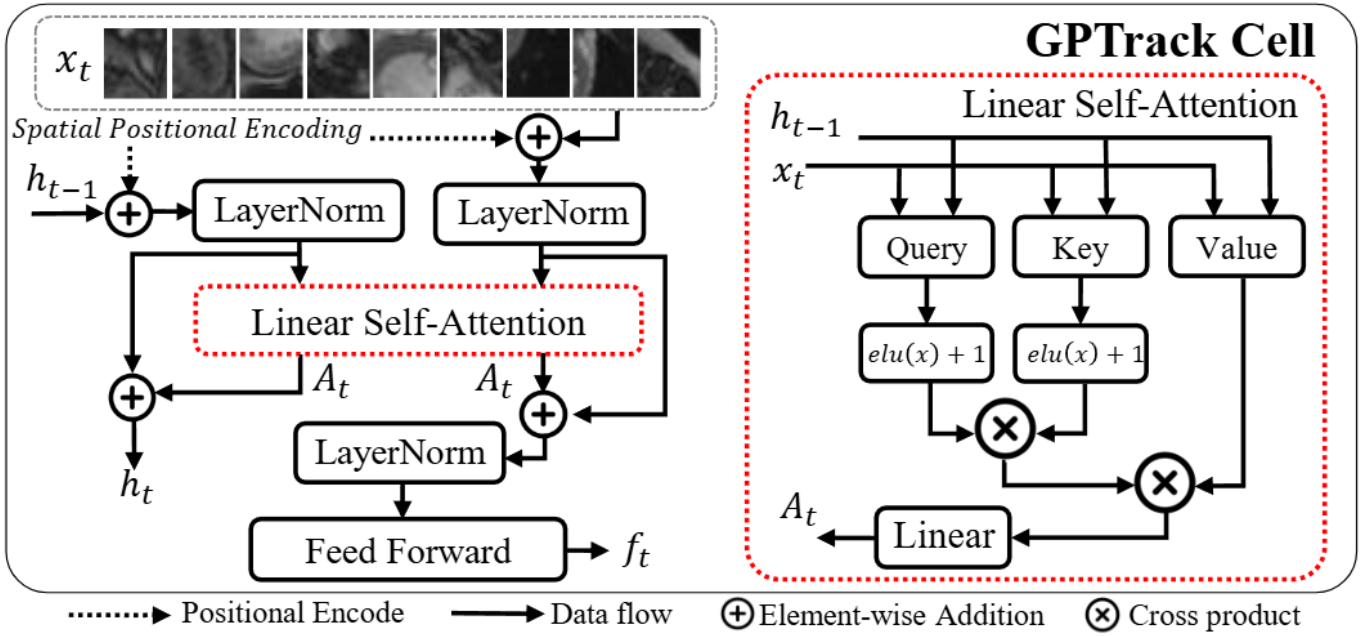
The overview pipeline of GPTrack

The GPTrack layer consists of two independent GPTrack cells that respond to forward and backward computation.



With the recursive manner, our GPTrack is able to formulate the variable temporal information while maintaining the comparable computational cost.

The overview pipeline of GPTrack Cell



The current input x^t and hidden state $\vec{h}_{t-1}$ followed by the addition of learnable position encoding $pos_t \in R^{P\times C}$ are respectively normalized by Layer Normalization LN(·). The linear self-attention then computes the attentive weight $A^t \in R^{P\times C}$ of combined x^t and $\vec{h}_{t-1}$. The above operations can be formulated as follows:

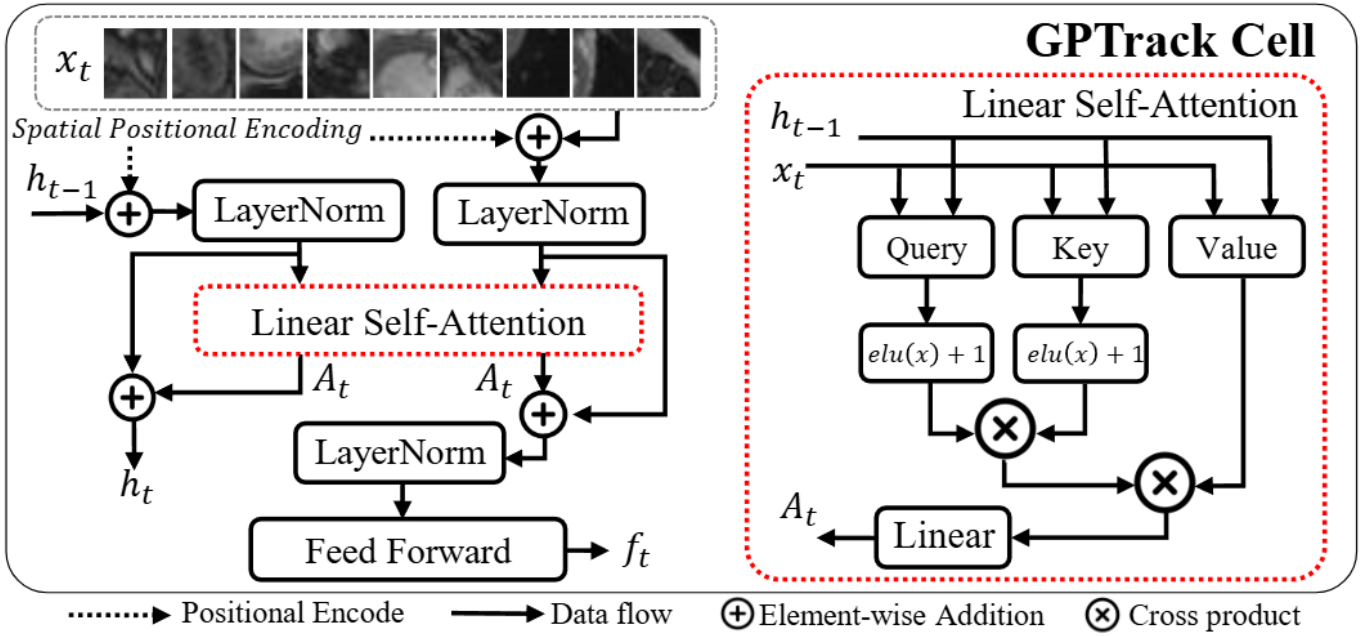
Exponential Linear Units $elu(\cdot)$

$$A_t = (\delta(\mathcal{W}_Q x)+1)(\delta(\mathcal{W}_K x)+1)^{\mathrm{T}}\mathcal{W}_V x;\ x = \mathrm{LN}(x_t+\mathrm{pos}_t)\oplus\mathrm{LN}(\vec{h}_{t-1}+\mathrm{pos}_t)$$

Learnable Weight of Query, Key and Value

Concatenation Operation

The overview pipeline of GPTrack Cell

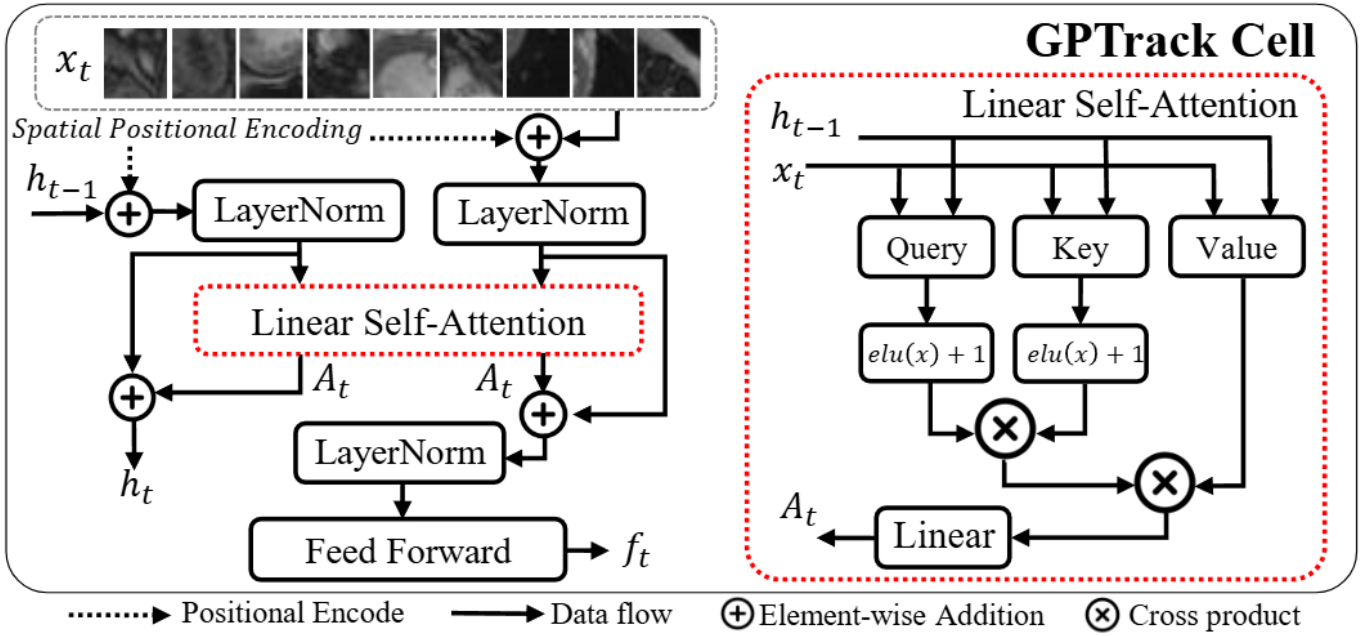


The output feature $\vec{f}_t$ and hidden state $\vec{h}_t$ of the t-th moment is formulated as:

$$\vec{f}_t = \mathrm{FFN}(\mathrm{LN}(A_t + \mathrm{LN}(x_t + \mathrm{pos}_s)))$$

$$\vec{h}_t = A_t + \mathrm{LN}(\vec{h}_{t-1} + \mathrm{pos}_s)$$

Gaussian Process in Cardiac Motion Tracking

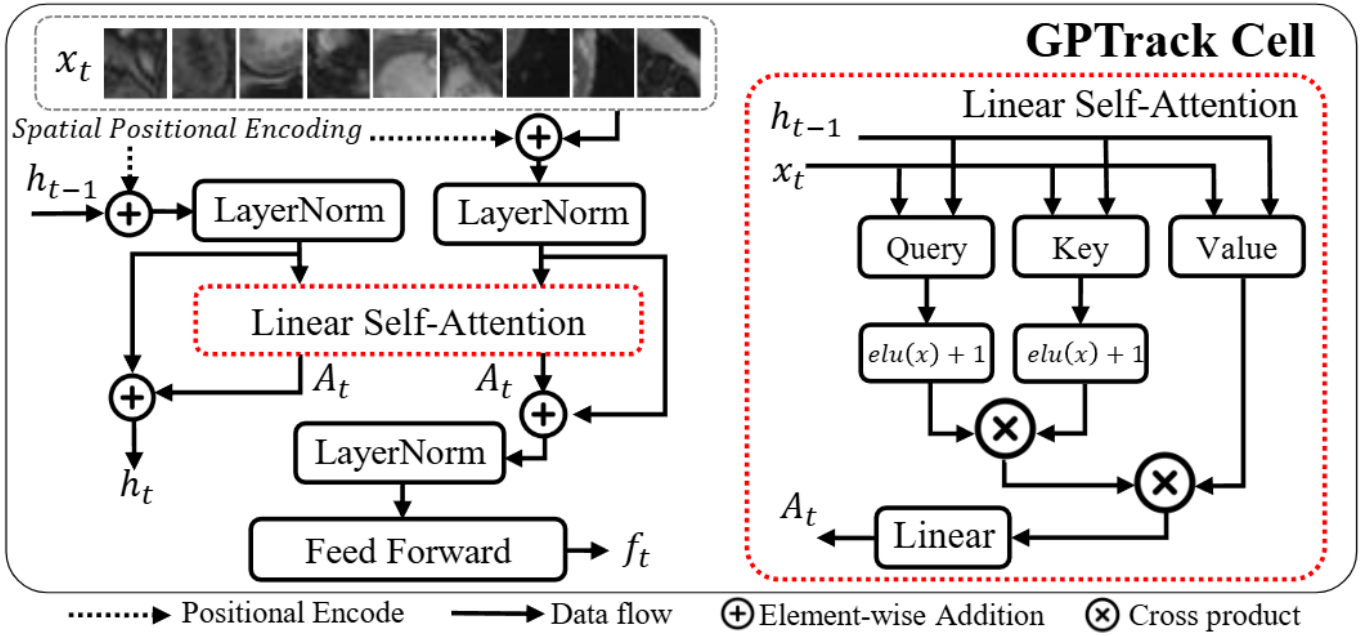


Initially, we define a covariance (kernel) function for the GP layer as depicted in the above figure. We employ the isotropic and stationary Matern kernel to fulfil the required covariance function structure.

$$\kappa(x_t, x_{t-1}) = \sigma \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{D(x_t, x_{t-1})}{l}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\frac{D(x_t, x_{t-1})}{l}\right),$$

where $\nu, \sigma, l > 0$ are the smoothness, magnitude and length scale parameters, K_v is the modified Bessel function, and $D(\cdot,\cdot)$ denotes the distance metric between features of two consecutive motion fields.

Gaussian Process in Cardiac Motion Tracking

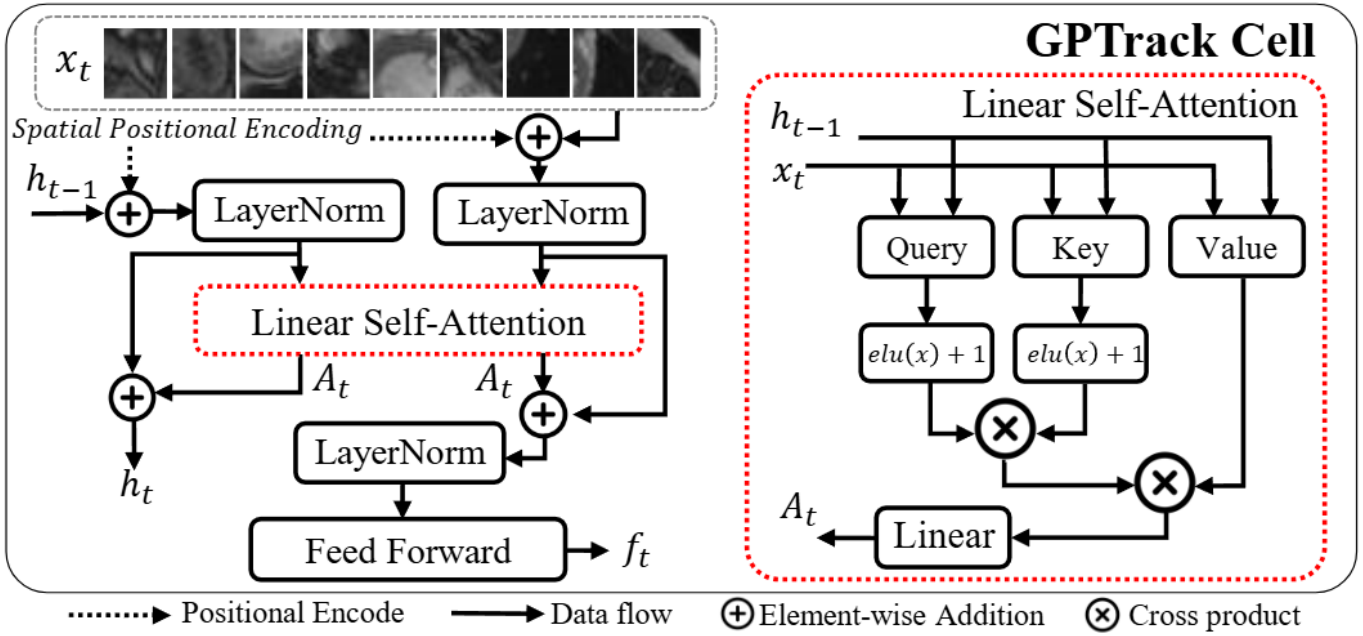


We regard the sequential output $\{f_t\}_t^T = 1$ of GPTrack as noise-corrupted versions of the ideal latent space encodings, formulating the inference as the following GP regression model with noise observations z_t:

$$z_t \sim \mathrm{GP}\left(\mu(\mathrm{pos}_t), \kappa\left(\mathrm{pos}_{t-1}, \mathrm{pos}_t\right)\right), \; f_t = z_t + \epsilon_t, \; \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

where σ^2 is the noise variance of the likelihood model set as the learnable parameter in GPTrack.

Gaussian Process in Cardiac Motion Tracking



Concretely, the Gaussian process corresponds to the following linear stochastic differential equation:

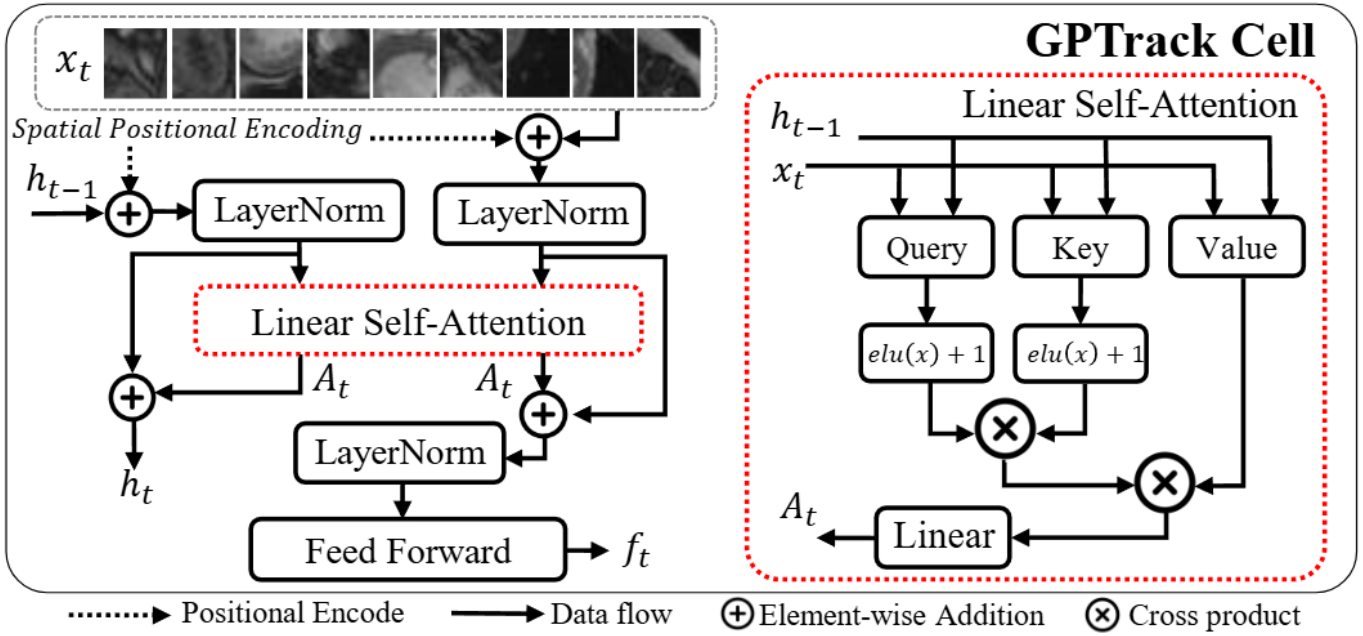
$$\frac{d}{dt}\mathbf{z}(t) = \mathbf{A}\mathbf{z}(t) + \mathbf{b}w(t), \quad f(t) = \mathbf{h}^{\mathsf{T}}\,\mathbf{z}(t) + \epsilon(t),\ \epsilon(t) \sim \mathcal{N}(0, \sigma^2),$$

with the solution as:

$$\mathbf{z}(t) = \exp^{(t-r)\mathbf{A}}\mathbf{z}(r) + \int_r^t \exp^{(t-s)\mathbf{A}}\mathbf{b}w(s)ds,\ \forall r < t,$$

$$f(t) = \mathbf{h}^{\mathsf{T}}\mathbf{z}(t) + \epsilon(t),\ \epsilon(t) \sim \mathcal{N}(0, \sigma^2),$$

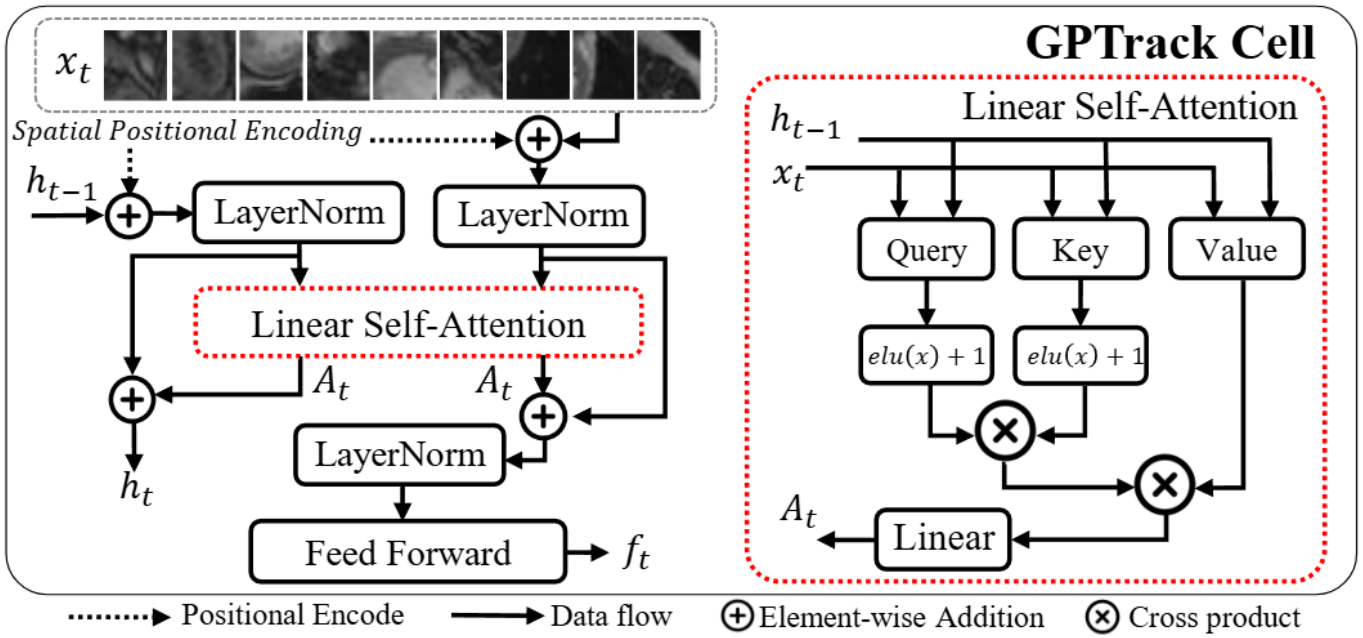
Gaussian Process in Cardiac Motion Tracking



Then we can discretize the pervious solution and get its weakly equivalent state-space model of Equation as:

$$\mathbf{z}_t = \mathbf{\Phi}_t\, \mathbf{z}_{t-1} + \mathbf{n}_t,\ f_t = \mathbf{h}^\top \mathbf{z}_t + \epsilon_t,\ \epsilon(t) \sim \mathcal{N}(0, \sigma^2),\ t = 1, \ldots, T,$$

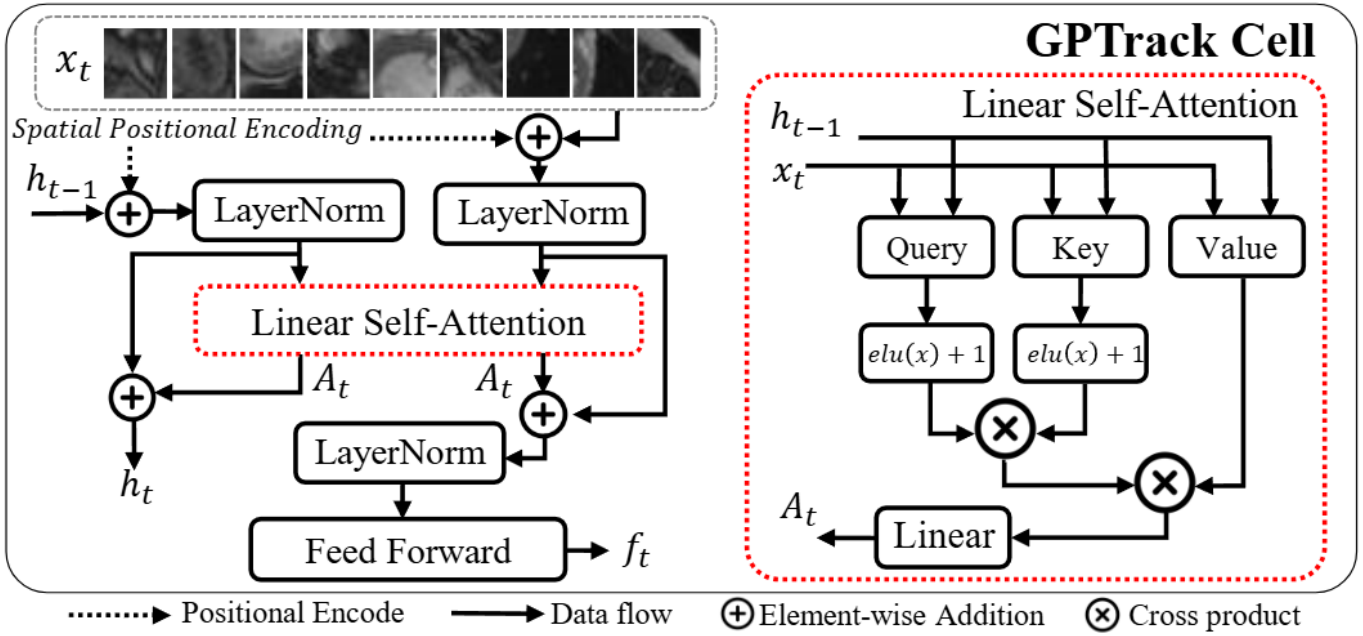
Gaussian Process in Cardiac Motion Tracking



Given the initial value $z_0 \sim N(\mu_0, \Sigma_0)$ with $\mu_0 = 0$ and $\Sigma_0 = diag(\sigma^2/2, 3\sigma^2/l^2)$, we can sequentially calculate the posterior distribution using update criterion of Kalman filter for state space model as:

$$\overline{\boldsymbol{\mu}}_t \leftarrow \boldsymbol{\Phi}_t \overline{\boldsymbol{\mu}}_{t-1}, \qquad \overline{\boldsymbol{\Sigma}}_t \leftarrow \boldsymbol{\Phi}_t \overline{\boldsymbol{\Sigma}}_{t-1} \boldsymbol{\Phi}_t^{\mathsf{T}} + \boldsymbol{\Sigma}_0 - \boldsymbol{\Phi}_t \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}_t^{\mathsf{T}},$$
$$\boldsymbol{\mu}_t \leftarrow \overline{\boldsymbol{\mu}}_t + \mathbf{k}_t (f_t - \mathbf{h}^{\mathsf{T}} \overline{\boldsymbol{\mu}}_t), \quad \boldsymbol{\Sigma}_t \leftarrow \overline{\boldsymbol{\Sigma}}_t - \mathbf{k}_t \mathbf{h}^{\mathsf{T}} \overline{\boldsymbol{\mu}}_t, \; t = 1, \ldots, T,$$

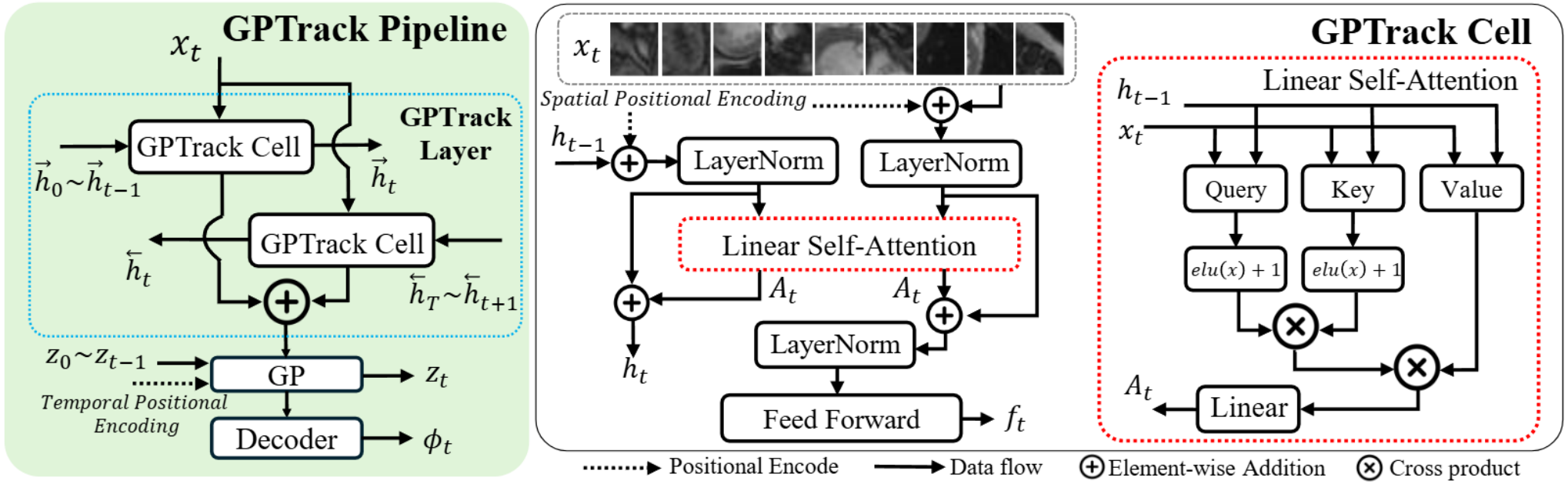
Gaussian Process in Cardiac Motion Tracking



$$\overline{\boldsymbol{\mu}}_t \leftarrow \boldsymbol{\Phi}_t \overline{\boldsymbol{\mu}}_{t-1}, \qquad \overline{\boldsymbol{\Sigma}}_t \leftarrow \boldsymbol{\Phi}_t \overline{\boldsymbol{\Sigma}}_{t-1} \boldsymbol{\Phi}_t^{\mathsf{T}} + \boldsymbol{\Sigma}_0 - \boldsymbol{\Phi}_t \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}_t^{\mathsf{T}},$$
$$\boldsymbol{\mu}_t \leftarrow \overline{\boldsymbol{\mu}}_t + \mathbf{k}_t (f_t - \mathbf{h}^{\mathsf{T}} \overline{\boldsymbol{\mu}}_t), \quad \boldsymbol{\Sigma}_t \leftarrow \overline{\boldsymbol{\Sigma}}_t - \mathbf{k}_t \mathbf{h}^{\mathsf{T}} \overline{\boldsymbol{\mu}}_t, \; t = 1, \ldots, T,$$

The k_t is the optimal Kalman gain at time t. The output of the GP layer in t-th moment thus can be formulated as $z_t^{GP} = \text{ReLU}(k_t z_t)$. In the final, the t-th motion field θ_t is obtained by decoder from the z_t^{GP}.

Gaussian Process in Cardiac Motion Tracking



The overall loss function L is formulated as:

$$\sum_{t=1}^{T-1}[\underbrace{\mathcal{L}_{kl}(x_t, x_{t+1})}_{\text{a)}} + \alpha_1(\underbrace{\mathcal{L}_{sm}(\phi_{t:t+1}) + \mathcal{L}_{sm}(\phi_{t+1:t})}_{\text{b)}}) + \alpha_2 \underbrace{\mathcal{L}_{nc}(x_{t+1}, x_1 \circ \phi_{0:t+1})}_{\text{c)}} + \alpha_3 \underbrace{\mathcal{L}_{sm}(\phi_{1:t+1})}_{\text{d)}}],$$

where α_1, α_2 and α_3 are loss weights, and $\phi_{t_1:t_2}$ is the motion field from state t_1 to t_2. $\mathcal{L}_{kl}(x_t, x_{t+1}) = \mathbb{KL}(q(z_t^{GP}|x_t; x_{t+1})||p(z_t^{GP}|x_t; x_{t+1})) + \mathbb{KL}(q(z_t^{GP}|x_{t+1}; x_t)||p(z_t^{GP}|x_{t+1}; x_t))$ is the summation of forward and backward VAE losses with latent coding z_t^{GT}, posterior distribution q and conditional distribution p, L_{nc} is the negative normalized local cross-correlation metric, and $\mathcal{L}_{sm}(\phi) = ||\nabla\phi||_2^2$ is the ℓ_2-total variation metric.

Experiments and Results

We use the **Peak Signal-to-Noise Ratio (PSNR)** and **Structural Similarity Index (SSIM)** to measure whether the motion field is accurately estimated between the first frame and the following wrapped frames. We also use the **Dice** score to measure the discrepancy between tracked and ground-truth cardiac segmentation

Table 1: The performance[1] of different registration methods in Cardiac-UDA dataset [17]. Results were reported in structures (RV, RA, LV, LA) and the overall averaged Dice score (Avg. %).

2D Methods (256×256)	LV ↑	RV ↑	LA ↑	RA ↑	Avg. ↑	$\|\|J\|-1\|\downarrow$	$det(J_\phi)\leq 0\downarrow$	PSNR ↑	SSIM ↑	Times (s) ↓	Params (M) ↓	TFlops ↓
Non-rigid Registration												
LDDMM [6]	69.44±6.9	70.61±5.3	57.03±12	70.78±5.3	69.22±5.2	13.12±11.05	25.67±23.41	26.45±2.7	76.44±2.4	*177.9±2.3	-	-
RDMM [8]	70.50±7.3	71.12±6.3	57.10±12	72.22±6.0	70.84±6.3	5.102±1.067	8.602±6.350	26.80±2.6	76.92±1.9	*241.0±3.5	-	-
ANTs (SyN) [24]	73.51±6.6	74.12±5.7	60.49±14	74.69±4.6	73.71±5.8	16.09±8.031	40.06±28.56	27.96±2.4	76.52±2.5	*156.4±4.1	-	-
Deep Learning Based Registration												
VM-SSD [10]	74.26±8.3	74.85±5.2	66.78±18	76.24±7.4	75.86±4.2	0.374±0.021	0.262±0.305	29.01±2.5	75.89±1.8	0.011±0.0	0.118	0.010
VM-NCC [10]	74.04±7.2	76.20±5.9	67.54±14	77.36±4.2	76.51±4.2	0.685±0.052	0.905±1.229	28.53±2.5	75.77±2.3	0.011±0.0	0.118	0.010
SYMNet [36]	75.21±7.5	75.33±6.1	69.67±11	77.78±5.5	76.60±4.2	0.454±0.048	0.631±0.108	28.56±2.5	76.87±2.0	0.101±0.0	0.449	0.125
VM-DIF [9]	73.53±7.5	76.37±5.6	68.10±15	78.55±6.1	76.83±5.0	0.387±0.066	0.437±0.508	28.80±2.2	76.87±1.8	0.011±0.0	0.109	0.010
Ahn SS, et al. [31]	75.66±7.6	77.24±6.3	71.41±17	79.20±6.9	77.04±4.3	3.107±1.156	2.664±0.827	29.86±2.5	77.59±2.4	0.017±0.0	7.783	0.851
DiffuseMorph [12]	77.02±6.0	80.45±5.5	72.50±12	80.81±5.3	79.27±5.2	0.319±0.043	0.339±0.478	29.48±2.0	77.02±2.5	0.103±0.0	90.67	0.227
DeepTag [15, 16]	76.83±7.5	80.13±4.8	72.87±14	80.98±4.2	79.41±3.5	0.273±0.056	0.027±0.022	28.53±2.5	76.40±2.3	**0.011**±0.0	**0.107**	**0.010**
GPTrack-M (Ours)	76.94±7.6	81.72±6.4	73.13±16	80.85±6.4	81.64±2.8	0.286±0.069	0.119±0.084	31.28±2.0	78.22±2.4	0.013±0.0	0.467	0.015
GPTrack-L (Ours)	77.07±8.0	82.57±7.1	73.11±15	**81.24**±6.4	82.11±2.7	**0.250**±0.044	**0.019**±0.017	31.57±2.0	78.70±2.1	0.016±0.0	5.161	0.041
GPTrack-XL (Ours)	**78.51**±7.9	**82.48**±6.0	**73.43**±12	81.20±5.9	**82.37**±2.7	0.279±0.085	0.027±0.023	**32.03**±2.4	**80.04**±2.4	0.026±0.0	7.536	0.053

Visualization of Cardiac Motion Tracking in CardiacUDA

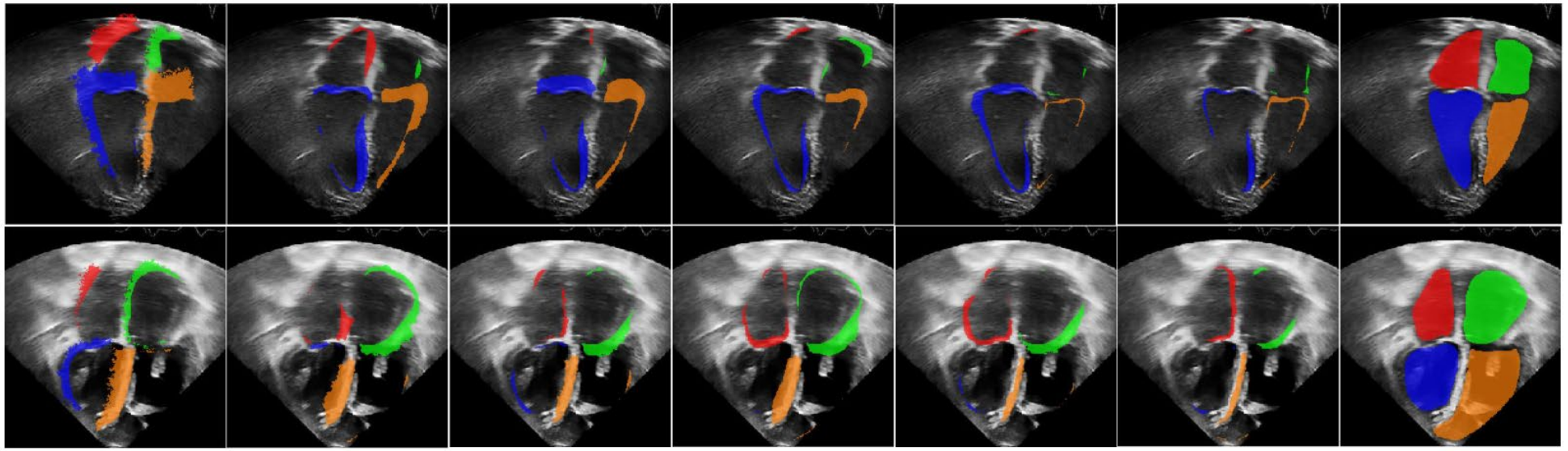


Figure 4: The visualization in 3D Echocardiogram video of motion tracking error. We visualised the last frame of tracking result and ground truth from 32 consecutive frames in CardiacUDA [17]. Colours Red, Blue, Green and Orange denote cardiac structures RA, RV, LV and LA, respectively.

Experiments and Results

We use the **Peak Signal-to-Noise Ratio (PSNR)** and **Structural Similarity Index (SSIM)** to measure whether the motion field is accurately estimated between the first frame and the following wrapped frames. We also use the **Dice** score to measure the discrepancy between tracked and ground-truth cardiac segmentation

Table 2: The performance[1] of different registration methods in ACDC [19] dataset. Results reported in structures (RV, LV, Myo) and overall averaged Dice score (Avg. %).

3D Methods (128×128×32)	RV ↑	LV ↑	Myo ↑	Avg. ↑	$\|\|J\|-1\|\downarrow$	$det(J_\phi)\leq 0\downarrow$	PSNR ↑	SSIM ↑	Times (s) ↓	Params (M) ↓	TFlops ↓
Non-rigid Registration											
LDDMM [6]	73.61±8.5	65.62±8.5	56.44±13	72.39±18	451.8±162.3	653.5±371.2	31.20±3.8	84.59±6.0	*1533±8.4	-	-
RDMM [8]	76.43±7.8	69.50±9.1	62.19±14	75.51±12	144.2±63.67	266.0±165.3	31.66±3.9	84.36±5.4	*1715±26	-	-
ANTs (SyN) [24]	75.30±7.4	66.92±8.6	58.03±11	74.64±13	15.82±22.30	57.26±37.74	30.92±3.6	84.26±5.6	*1166±16	-	-
Deep Learning Based Registration											
VM-SSD [10]]	79.83±7.1	74.27±9.0	64.44±15	77.56±12	3.144±2.242	4.602±3.485	32.61±3.7	83.88±5.2	0.015±0.0	0.327	0.767
VM-NCC [10]	81.60±6.5	77.00±8.6	67.90±13	79.90±11	0.260±0.070	0.079±0.058	34.68±3.3	85.01±5.5	0.015±0.0	0.327	0.767
VM-DIF [9]	81.50±6.6	75.50±9.2	65.90±14	78.90±12	0.286±0.074	0.083±0.063	33.48±3.5	84.22±5.1	0.015±0.0	**0.327**	0.767
SYMNet [36]	80.46±6.4	77.81±9.4	66.22±14	79.47±13	0.341±0.062	0.121±0.054	32.91±3.5	83.55±4.9	0.414±0.0	1.124	0.226
NICE-Trans [52]	79.97±6.0	78.55±8.1	67.02±11	79.66±10	0.278±0.071	0.093±0.044	33.08±3.0	83.88±4.7	0.486±0.0	5.619	0.280
DiffuseMorph [12]	82.10±6.7	78.30±8.6	67.80±15	80.50±11	0.237±0.068	0.061±0.038	34.73±3.6	84.30±5.2	0.458±0.0	0.327	0.642
CorrMLP [53]	80.33±6.5	80.07±7.8	70.51±14	80.44±8.6	0.248±0.055	0.059±0.022	34.90±2.9	84.27±4.5	0.070±0.0	13.36	0.303
DeepTag [15, 16]	81.89±7.0	79.10±7.5	70.37±13	80.83±12	0.185±0.067	0.044±0.025	33.64±3.4	83.09±4.9	**0.015±0.0**	0.362	**0.113**
Transmatch [54]	81.22±7.0	80.34±6.8	71.21±12	81.35±9.8	0.226±0.050	0.077±0.054	33.89±3.3	84.78±4.9	0.325±0.0	70.71	0.603
FSDiffReg [11]	82.70±6.1	80.90±7.7	72.40±12	82.30±9.6	0.214±0.054	0.054±0.026	35.34±3.5	85.85±5.2	1.106±0.0	1.320	0.855
GPTrack-M (Ours)	81.65±7.0	80.77±7.5	71.53±16	81.45±10	0.209±0.081	0.047±0.035	34.82±3.2	85.78±5.3	0.022±0.0	0.418	0.201
GPTrack-L (Ours)	82.78±5.6	81.16±6.8	71.71±14	82.38±11	0.182±0.072	0.035±0.022	34.99±3.0	85.62±4.9	0.023±0.0	0.942	0.204
GPTrack-XL (Ours)	**82.91**±5.8	**81.23**±8.2	**72.86**±9.0	**82.65**±10	**0.178**±0.024	**0.032**±0.021	**35.52**±3.1	**86.19**±5.0	0.034±0.0	1.094	0.205

Visualization of Cardiac Motion Tracking in ACDC

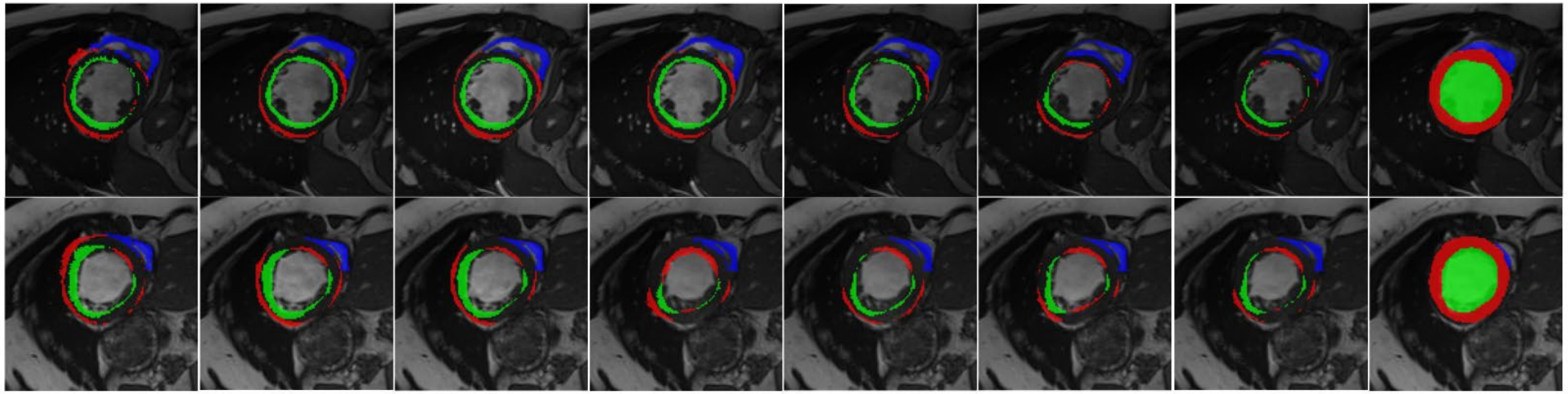


Figure 5: The visualization in 4D Cardiac MRI of motion tracking error. We visualised the result of the last frame tracking from ED to ES and corresponding ground truth in ACDC [17]. Colours Red, Blue, and Green denote cardiac structures MYO, LA, and LV, respectively.

Our Contribution

- We highlights the cardiac motion trajectory that follows a certain pattern can be modelled as prior knowledge via the Gaussian Process.
- We capture the long-term relationship of cardiac motion via a bidirectional recursive manner, mimics the workflows of the classical diffeomorphic registration framework.
- Our framework achieves state-of-the-art performance on both 3D Echocardiogram videos and 4D temporal MRI datasets, maintaining comparable computational efficiency.

NEURAL INFORMATION PROCESSING SYSTEMS

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Thank You!

Code & data: ***https://github.com/xmed-lab/GPTrack***

Contact : ***Jiewen Yang - jyangcu@connect.ust.hk***
Xiaomeng Li - eexmli@ust.hk