

Supplementary Material for DeepTag: An Unsupervised Deep Learning Method for Motion Tracking on Cardiac Tagging Magnetic Resonance Images

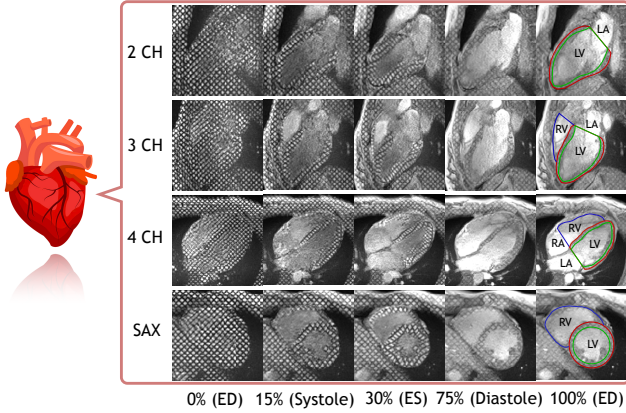


Figure 1. Typical scan views (2-, 3-, 4-chamber views and short-axis view) of cardiac tagging MRI. Number under the figure means percentage of one cardiac cycle. Red and green contours depict the epi- and endo-cardial borders of left ventricle (LV) myocardium (MYO) wall. Bright area within the green contour is the LV blood pool. Blue contour depicts the right ventricle (RV). LA: left atrium. RA: right atrium.

1. Challenges for T-MRI Motion Tracking

Tags are physical properties of the tissue which will deform with the heart, as it contracts and relaxes during a cardiac cycle. Tracking the deformation of tags can help us retrieve a 2D displacement field in the imaging plane and reconstruct local motion of the myocardium. Challenges for motion tracking on cardiac tagging magnetic resonance imaging (t-MRI) images can include the following.

(1) Image appearance changes a lot even within a 2D sequence. One can observe in Fig. 1 that, at the beginning of the heart cycle, which is the end diastole (ED) phase, tag grids cover the imaging plane except for the background lung cavity. As the heart contracts to the end systole (ES) phase, approximately 30% of a cycle, untagged blood replace tagged blood, leaving a brighter untagged blood pool. After ES phase, as the heart relaxes towards the ED phase (100% of a cycle), tag signal in the myocardium fades because of T1 relaxation of the perturbed magnetization vectors. So the dark tags get brighter and brighter in the later frames. This can pose a big challenge for the tag deformation

estimation.

(2) Different frames can have a very different image appearance, related to the changing myocardium shape.

(3) t-MRI images have relatively low temporal resolution, reaching at the best 20 ~ 30 frames in a cycle. Motion in between consecutive two frames could be large, especially during the rapid early filling phase.

(4) t-MRI is 2D imaging, and through-plane motion of the heart through the fixed imaging plane could make tags disappear at one frame and reappear at some other frame within a sequence. Non identification of such tags will incur motion-tracking errors.

(5) Due to imaging condition changes in the imaging process, such as magnetic field drift and patients' respiratory motion by unsuccessful breath holding, other artifacts and noise can degrade image quality.

2. Derivation of KL Loss

Detailed derivation of the KL loss is as following:

$$\begin{aligned}
 \mathcal{KL}[q_\psi(z|\mathbf{y}; \mathbf{x}) || p(z|\mathbf{y}; \mathbf{x})] &= \mathbb{E}_q \left[\log \frac{q_\psi(z|\mathbf{y}; \mathbf{x})}{p(z|\mathbf{y}; \mathbf{x})} \right] \\
 &= \mathbb{E}_q \left[\log \frac{q_\psi(z|\mathbf{y}; \mathbf{x}) p(\mathbf{y}; \mathbf{x})}{p(z|\mathbf{y}; \mathbf{x}) p(\mathbf{y}; \mathbf{x})} \right] \\
 &= \mathbb{E}_q \left[\log \frac{q_\psi(z|\mathbf{y}; \mathbf{x}) p(\mathbf{y}; \mathbf{x})}{p(z, \mathbf{y}; \mathbf{x})} \right] \\
 &= \mathbb{E}_q \left[\log \frac{q_\psi(z|\mathbf{y}; \mathbf{x}) p(\mathbf{y}; \mathbf{x})}{p(\mathbf{y}|\mathbf{z}; \mathbf{x}) p(\mathbf{x}|\mathbf{z}) p(\mathbf{z})} \right] \\
 &= \mathbb{E}_q \left[\log \frac{q_\psi(z|\mathbf{y}; \mathbf{x}) p(\mathbf{y}; \mathbf{x})}{p(\mathbf{y}|\mathbf{z}; \mathbf{x}) p(\mathbf{x}) p(\mathbf{z})} \right] \\
 &= \mathbb{E}_q \left[\log \frac{q_\psi(z|\mathbf{y}; \mathbf{x})}{p(z)} \right] - \mathbb{E}_q [\log p(\mathbf{y}|\mathbf{z}; \mathbf{x})] + \log p(\mathbf{y}|\mathbf{x}) \\
 &= \mathcal{KL}[q_\psi(z|\mathbf{y}; \mathbf{x}) || p(z)] - \mathbb{E}_q [\log p(\mathbf{y}|\mathbf{z}; \mathbf{x})] + \text{const.}
 \end{aligned} \tag{1}$$

For the first term, we have

$$\begin{aligned}
& \mathcal{KL}[q_\psi(\mathbf{z}|\mathbf{y}; \mathbf{x})||p(\mathbf{z})] \\
&= \mathcal{KL}[\mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}, \boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}})||\mathcal{N}(\mathbf{z}; \mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{z}})] \\
&= \frac{1}{2} \left[\log \frac{|\boldsymbol{\Sigma}_{\mathbf{z}}|}{|\boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}|} - n + \text{tr}(\boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}) \right. \\
&\quad \left. + \boldsymbol{\mu}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}^T \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\mu}_{\mathbf{z}|\mathbf{x}, \mathbf{y}} \right], \tag{2}
\end{aligned}$$

where n is the total number of the variables in $p(\mathbf{z})$.

According to our setting, $\boldsymbol{\Lambda}_{\mathbf{z}} = \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} = \lambda \mathbf{L}$ [2], where $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the Laplacian of a neighborhood graph defined on the pixel grid, \mathbf{D} is the graph degree matrix, \mathbf{A} is a pixel neighborhood adjacency matrix. Therefore, $\log|\boldsymbol{\Sigma}_{\mathbf{z}}|$ is constant. Since $\boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}$ is set to be diagonal, $\log|\boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}| = \text{tr} \log \boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}$. And $\text{tr}(\boldsymbol{\Sigma}_{\mathbf{z}}^{-1} \boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}) = \text{tr}(\lambda(\mathbf{D} - \mathbf{A})\boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}) = \text{tr}(\lambda \mathbf{D} \boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}})$. So we can get

$$\begin{aligned}
& \mathcal{KL}[q_\psi(\mathbf{z}|\mathbf{y}; \mathbf{x})||p(\mathbf{z})] \\
&= \frac{1}{2} [\text{tr}(\lambda \mathbf{D} \boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}} - \log \boldsymbol{\Sigma}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}) + \boldsymbol{\mu}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}^T \boldsymbol{\Lambda}_{\mathbf{z}} \boldsymbol{\mu}_{\mathbf{z}|\mathbf{x}, \mathbf{y}}] \tag{3} \\
&+ \text{const.}
\end{aligned}$$

For the second term, if we model $p(\mathbf{y}|\mathbf{z}; \mathbf{x})$ as a Gaussian, we can get

$$\begin{aligned}
& -\mathbb{E}_q[\log p(\mathbf{y}|\mathbf{z}; \mathbf{x})] \\
&= -\mathbb{E}_q[\log \mathcal{N}(\mathbf{y}; \mathbf{x} \circ \boldsymbol{\phi}, \sigma^2 \mathbb{I})] \\
&= \frac{1}{2} \mathbb{E}_q[\log (2\pi\sigma^2)^n + \frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{x} \circ \boldsymbol{\phi}\|_2^2] \tag{4} \\
&= \frac{1}{2\sigma^2} \mathbb{E}_q[\|\mathbf{y} - \mathbf{x} \circ \boldsymbol{\phi}\|_2^2] + \text{const.},
\end{aligned}$$

where the term $\|\mathbf{y} - \mathbf{x} \circ \boldsymbol{\phi}\|_2^2$ corresponds to the sum-of-squared difference (SSD) metric.

If we model $p(\mathbf{y}|\mathbf{z}; \mathbf{x})$ as a Boltzmann distribution, we can get

$$\begin{aligned}
& -\mathbb{E}_q[\log p(\mathbf{y}|\mathbf{z}; \mathbf{x})] \\
&\propto -\mathbb{E}_q[\log \exp(-\gamma \text{NCC}(\mathbf{y}, \mathbf{x} \circ \boldsymbol{\phi}))] \tag{5} \\
&= \gamma \mathbb{E}_q[\text{NCC}(\mathbf{y}, \mathbf{x} \circ \boldsymbol{\phi})],
\end{aligned}$$

where γ is a negative scalar hyperparameter, NCC is the normalized local cross correlation metric.

We can approximate the expectation \mathbb{E}_q with K samples $\mathbf{z}_k \sim q_{\mathbf{z}}$, so we get

$$-\mathbb{E}_q[\log p(\mathbf{y}|\mathbf{z}; \mathbf{x})] = \frac{\gamma}{K} \sum_k [\text{NCC}(\mathbf{y}, \mathbf{x} \circ \boldsymbol{\phi}_k)]. \tag{6}$$

Note that by Eq. (1), we can get:

$$\begin{aligned}
& \log p(\mathbf{y}|\mathbf{x}) = \mathcal{KL}[q_\psi(\mathbf{z}|\mathbf{y}; \mathbf{x})||p(\mathbf{z}|\mathbf{y}; \mathbf{x})] \\
&\quad - (\mathcal{KL}[q_\psi(\mathbf{z}|\mathbf{y}; \mathbf{x})||p(\mathbf{z})] - \mathbb{E}_q[\log p(\mathbf{y}|\mathbf{z}; \mathbf{x})]) \\
&= \mathcal{KL}[q_\psi(\mathbf{z}|\mathbf{y}; \mathbf{x})||p(\mathbf{z}|\mathbf{y}; \mathbf{x})] + \text{ELBO} \\
&\geq \text{ELBO}, \tag{7}
\end{aligned}$$

where

$$\text{ELBO} = -(\mathcal{KL}[q_\psi(\mathbf{z}|\mathbf{y}; \mathbf{x})||p(\mathbf{z})] - \mathbb{E}_q[\log p(\mathbf{y}|\mathbf{z}; \mathbf{x})]). \tag{8}$$

Thus, maximizing the ELBO of the log marginalized likelihood $\log p(\mathbf{y}|\mathbf{x})$ in Eq. (7) is equivalent to minimizing $\mathcal{KL}[q_\psi(\mathbf{z}|\mathbf{y}; \mathbf{x})||p(\mathbf{z}|\mathbf{y}; \mathbf{x})]$ in Eq. (1).

3. Backward Registration

With the SVF representation, we can also compute an inverse motion field $\boldsymbol{\phi}^{-1}$ by inputting $-\mathbf{z}$ into the SS layer: $\boldsymbol{\phi}^{-1} = \text{SS}(-\mathbf{z})$. Thus we can warp image \mathbf{y} backward towards image \mathbf{x} and get the observation distribution of warped image, $\mathbf{y} \circ \boldsymbol{\phi}^{-1}$, which is also modeled by a Boltzmann distribution:

$$p(\mathbf{x}|\mathbf{z}; \mathbf{y}) \sim \exp(-\gamma \text{NCC}(\mathbf{x}, \mathbf{y} \circ \boldsymbol{\phi}^{-1})), \tag{9}$$

where \mathbf{x} denotes the observation of warped image \mathbf{y} . We call the process of warping image \mathbf{y} towards \mathbf{x} the backward registration. We minimize the KL divergence between $q_\psi(\mathbf{z}|\mathbf{x}; \mathbf{y})$ and $p(\mathbf{z}|\mathbf{x}; \mathbf{y})$, which leads to maximizing the ELBO of the log marginalized likelihood $\log p(\mathbf{x}|\mathbf{y})$ as follows:

$$\begin{aligned}
& \min_{\psi} \mathcal{KL}[q_\psi(\mathbf{z}|\mathbf{x}; \mathbf{y})||p(\mathbf{z}|\mathbf{x}; \mathbf{y})] \\
&= \min_{\psi} \mathcal{KL}[q_\psi(\mathbf{z}|\mathbf{x}; \mathbf{y})||p(\mathbf{z})] - \mathbb{E}_q[\log p(\mathbf{x}|\mathbf{z}; \mathbf{y})] \tag{10} \\
&\quad + \log p(\mathbf{x}|\mathbf{y}).
\end{aligned}$$

4. Algorithm 1

We use Algorithm 1 to compute Lagrangian motion field Φ from interframe (INF) motion field $\boldsymbol{\phi}$.

5. Evaluation Metric Definitions

Following the metric used in [1], we use the root mean squared (RMS) error of distance between the centers of predicted landmark X' and ground truth landmark X for evaluation of motion tracking accuracy:

$$\text{RMS} = \sqrt{\frac{1}{M} \sum_{i=0}^{M-1} \|X'_i - X_i\|_2^2}, \tag{11}$$

where M is the total number of predefined ground truth landmarks.

In addition, we evaluate the diffeomorphic property of the predicted INF motion field $\boldsymbol{\phi}$, using the following Jacobian determinant:

$$\det(J_{\boldsymbol{\phi}}(\mathbf{p})) = \det \left(\begin{bmatrix} \frac{\partial \phi_x(\mathbf{p})}{\partial x} & \frac{\partial \phi_x(\mathbf{p})}{\partial y} \\ \frac{\partial \phi_y(\mathbf{p})}{\partial x} & \frac{\partial \phi_y(\mathbf{p})}{\partial y} \end{bmatrix} \right), \tag{12}$$

Algorithm 1: INF motion field recombination as Lagrangian motion field

Input: INF motion field: ϕ
Output: Lagrangian motion field: Φ

```

1 // shape: frames, channels, size x, size y
2  $shape \leftarrow \text{shape of } \phi$ ;
3  $N \leftarrow shape[0]$ ;
4  $C \leftarrow shape[1]$ ;
5  $\Phi \leftarrow \phi$ ;
6 for  $n$  from 1 to  $N - 1$  do
7    $src \leftarrow \phi[n, ::]$ ;
8    $flow \leftarrow \Phi[n - 1, ::]$ ;
9   for  $d$  from 0 to  $C - 1$  do
10    //  $\circ$  is linear interpolation
11     $\Phi[n, d, ::] \leftarrow flow[0, d, ::]$ 
12     $\quad + src[0, d, ::] \circ flow$ ;
13  end
14 end

```

where p is a certain position. Such a Jacobian determinant could be used to analyze the local behavior of the motion field. A positive Jacobian determinant $\det(J_\phi(p))$ indicates the motion field at position p preserves the orientation in the neighborhood of p . However, a negative Jacobian determinant $\det(J_\phi(p))$ indicates the motion field at position p reverses the orientation in the neighborhood of p , which will lose the one-to-one mapping.

6. More Detailed Results

6.1. T-MRI Image Sequence Registration Results

In the supplementary folder `Registration`, we show representative t-MRI image sequence registration results: (upper-left) tagging image sequence; (upper-right) forward registration results; (bottom-left) backward registration results; (bottom-right) Lagrangian registration results.

6.2. Landmarks Tracking Results

In the supplementary folder `LM_Tracking`, we show representative landmarks tracking results on basal, middle and apical slice: red is ground truth, green is prediction. Note that in the basal slice, on the septum wall, which is between RV and LV, tags may apparently disappear for some frames, due to through-plane motion, as do the ground truth landmarks, but we still showed the predicted landmarks on the closest position. Our method can even track the motion on the last several frames accurately, in spite of the significant image quality degradation.

6.3. Motion Field Results

In supplementary folder `Motion_Field_Quiver`, we show representative INF motion fields and Lagrangian motion fields, represented as a “quiver” form. Note that our method accurately captures the back-and-forth motion in the left ventricle myocardium wall during systole. Also note that our method can even track the right ventricle’s motion accurately.

In the folder `Motion_Field_Map`, we show corresponding Lagrangian motion fields: (left) x component; (right) y component.

6.4. Virtual Tag Grid Tracking Results

In the supplementary folder `Virtual_Tag_Grid`, we show representative virtual tag grid tracking results on the short-axis view: (left) tagging image sequence; (middle) warped virtual tag grid by the Lagrangian motion field; (right) virtual tag grid superimposed on tagging images. Note that the virtual tag grid has been aligned with the tag pattern at time $t = 0$. As time goes on, the virtual tag grid is deformed by the predicted Lagrangian motion field and follows the underlying tag pattern in the images very well.

In the folder `Virtual_Tag_Grid_LAX`, we show representative virtual tag grid tracking results on the long axis (2-, 3-, 4-chamber) views: (upper) tagging image sequence; (bottom) virtual tag grid superimposed on tagging images.

6.5. Full Sequence Motion Tracking Results

In Fig. 2 and Fig. 3, we show the motion tracking results on a full t-MRI image sequence.

References

- [1] Raghavendra Chandrasekara, Raad H Mohiaddin, and Daniel Rueckert. Analysis of 3-d myocardial motion in tagged mr images using nonrigid image registration. *IEEE Transactions on Medical Imaging*, 23(10):1245–1250, 2004. 2
- [2] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 729–738. Springer, 2018. 2

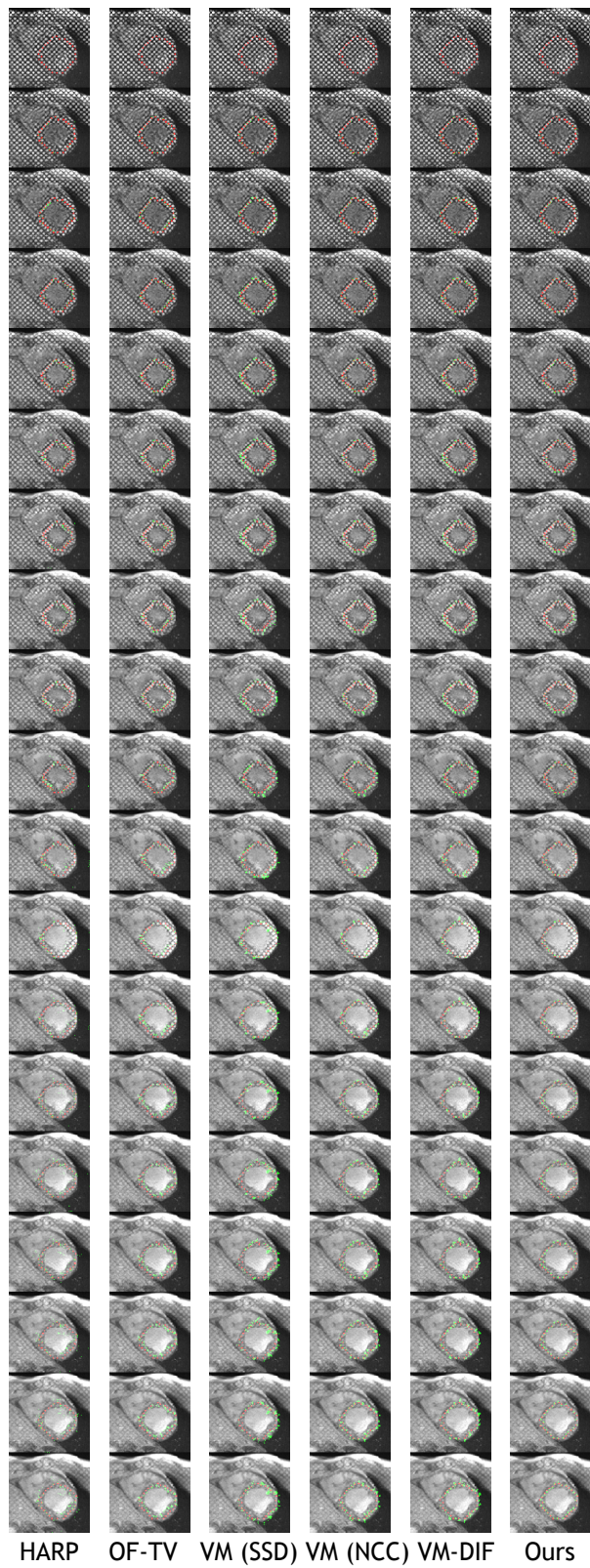


Figure 2. Motion tracking results shown on a full t-MRI image sequence (best viewed zoomed in). Red is ground truth, green is prediction.



Figure 3. Ablation study motion tracking results shown on a full t-MRI image sequence (best viewed zoomed in). Red is ground truth, green is prediction.