

# Deep Lesion Tracker: Monitoring Lesions in 4D Longitudinal Imaging Studies

Jinzheng Cai<sup>1</sup> Youbao Tang<sup>1</sup> Ke Yan<sup>1</sup> Adam P. Harrison<sup>1</sup> Jing Xiao<sup>2</sup> Gigin Lin<sup>3</sup> Le Lu<sup>1</sup>

<sup>1</sup>PAII Inc. <sup>2</sup>Ping An Technology <sup>3</sup>Chang Gung Memorial Hospital

## Supplementary Materials

**Robustness analysis with more details.** Due to constraints of space, we have omit “CPM@Radius” in the Table 2 of the main manuscript and kept “CPM@10mm” as it is a more tighter evaluation. In STable 1, measurements of performance of different trackers delivered by “CPM@Radius” follow the same trend as the results measured by “CPM@10mm”. With “CPM@Radius”, DLT-Mix remains to be the best approach. DEEDS is the most vulnerable method with over 10% drop in “CPM@Radius”. In comparison, DLT-Mix only drops 1.87%.

**Parameter analysis with more details.** Due to constraints of space, we have omit some details of experiments that reported in Table. 3 in the main manuscript. As promised, we show the complete version here in STable 2.

**More visualization examples for method comparison.** In Fig. 4 of the main manuscript, we compared our methods with three state-of-the-art trackers. Here, we show more examples in SFig. 1, SFig. 2, and SFig. 3. All case are shown with representative axial, coronal, and sagittal slices to accurately illustrate 3D locations.

For 2D visualization, we orthographically projected the lesion center from 3D. These centers were projected from any axial slices within 10mm of the ground truth axial slices (most CTs have 5mm slice thickness). Thus, in the second example of Fig. 4 in the main manuscript, DEEDS is actually located further away in the  $z$  direction, despite the visual appearance. Some centers in samples 5 and 7 are invisible because they overlap and/or they are located outside of the  $\pm 10mm$  limit.

**Visualization examples for lesion tracking with multiple follow ups.** We show lesion tracking using deep lesion tracker (DLT) with three follow-ups in SFig. 4 and SFig. 5. In SFig. 6, we show DLT tracks lesions up to six follow-ups. For lesion tracking with multiple follow-ups, DLT is only provided with the location of the target lesion in the initial template image.

## References

- [1] M. P. Heinrich, M. Jenkinson, M. Brady, and J. A. Schnabel. Mrf-based deformable registration and ventilation estimation

of lung ct. *IEEE Trans. Med. Imaging*, 32(7):1239–1248, 2013. 2

- [2] Bo Li, Wei Wu, Qiang Wang, Fangyi Zhang, Junliang Xing, and Junjie Yan. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 4282–4291. 2019. 2

| Method        | CPM@Radius            | CPM@10mm              | MED (mm)                |
|---------------|-----------------------|-----------------------|-------------------------|
| SiamRPN++ [2] | 71.52 (↓ 8.79)        | 51.27 (↓ 17.6)        | 10.6±10.3 (↑ 2.3)       |
| DEEDS [1]     | 74.82 (↓ 10.7)        | 53.85 (↓ 18.0)        | 9.8±8.9 (↑ 2.4)         |
| DLT-SSL       | 78.38 (↓ 3.14)        | 64.24 (↓ 6.80)        | 10.0±11.4 (↑ 1.2)       |
| DLT           | 83.18 (↓ 3.70)        | 70.36 (↓ 8.49)        | 8.1±8.7 (↑ 1.2)         |
| DLT-Mix       | <b>86.88 (↓ 1.87)</b> | <b>75.03 (↓ 3.62)</b> | <b>8.0±10.5 (↑ 0.9)</b> |

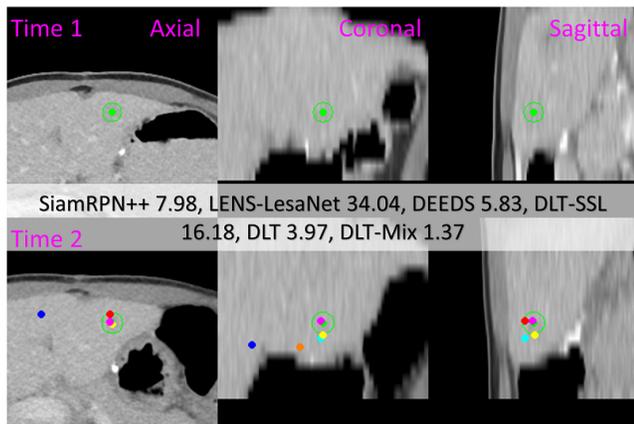
STable 1: Robustness evaluation. ↓ and ↑ demonstrate decrease and increase of measurements, respectively, compared with the values reported in Table 1 in the main script.

| Model id | Ablation study             | Eq. 6: $K_g$ |       | $\psi, \phi$ | Eq. 3    | Eq. 2: $G$ |               | Valid             | Test      | Speed |
|----------|----------------------------|--------------|-------|--------------|----------|------------|---------------|-------------------|-----------|-------|
|          |                            | size         | learn | dim.         | fusion   | size       | $\Delta\mu_t$ | MED (mm)          | MED (mm)  | spv   |
| <i>a</i> | w/o $K_g$                  | NA           | NA    | 64           | multiply | 4r         | ✗             | 8.77±9.88 (↑1.69) | 9.29±10.2 | 1.44  |
| <i>b</i> | smaller $K_g$              | 7,7,7        | ✓     | 64           | multiply | 4r         | ✗             | 8.26±9.40 (↑1.18) | 9.41±10.2 | 2.38  |
| <i>c</i> | greater $K_g$              | 7,15,15      | ✓     | 64           | multiply | 4r         | ✗             | 7.24±5.64 (↑0.16) | 7.67±8.78 | 24.1  |
| <i>d</i> | smaller $G_t, G_s$         | 7,11,11      | ✓     | 64           | multiply | 2r         | ✗             | 7.56±8.95 (↑0.48) | 7.51±8.39 | 3.51  |
| <i>e</i> | greater $G_t, G_s$         | 7,11,11      | ✓     | 64           | multiply | 8r         | ✗             | 8.40±9.23 (↑1.32) | 8.81±9.80 | 3.51  |
| <i>f</i> | smaller feat. dim.         | 7,11,11      | ✓     | 32           | multiply | 4r         | ✗             | 7.23±6.17 (↑0.15) | 8.72±16.6 | 2.25  |
| <i>g</i> | greater feat. dim.         | 7,11,11      | ✓     | 128          | multiply | 4r         | ✗             | 7.15±6.99 (↑0.07) | 7.91±9.29 | 5.83  |
| <i>h</i> | w/o ASE                    | 7,11,11      | ✓     | 64           | NA       | NA         | NA            | 8.23±9.44 (↑1.15) | 9.34±10.0 | 3.51  |
| <i>i</i> | w/o learn $K_g$            | 7,11,11      | ✗     | 64           | multiply | 4r         | ✗             | 7.61±9.02 (↑0.53) | 7.98±9.26 | 3.51  |
| <i>j</i> | <b>comparison baseline</b> | 7,11,11      | ✓     | 64           | multiply | 4r         | ✗             | 7.08±5.25 (↑0.00) | 7.95±8.96 | 3.51  |
|          | Eq. 3 with concat.         | 7,11,11      | ✓     | 64           | concat.  | 4r         | ✓             | 6.85±9.47 (↓0.23) | 7.94±9.22 | 5.91  |
|          | <b>final configuration</b> | 7,11,11      | ✓     | 64           | multiply | 4r         | ✓             | 6.69±5.62 (↓0.39) | 6.98±8.95 | 3.51  |

STable 2: Parameter analysis and ablation study of the proposed components.

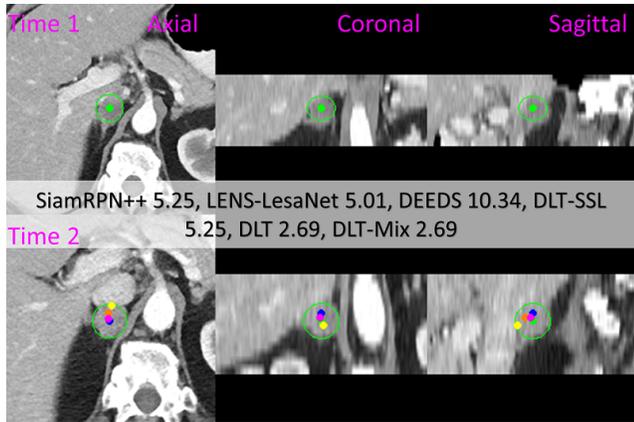


SFigure 1: Comparison of our methods, *i.e.*, DLT, DLT-SSL, DLT-Mix, with three state-of-the-art trackers including a Siamese networks based tracker – SiamRPN++, a leading registration algorithm – DEEDS, and a detector based tracker – LENS-LesaNet. Offsets from the predicted lesion centers to the manually labeled center are reported in *mm*.



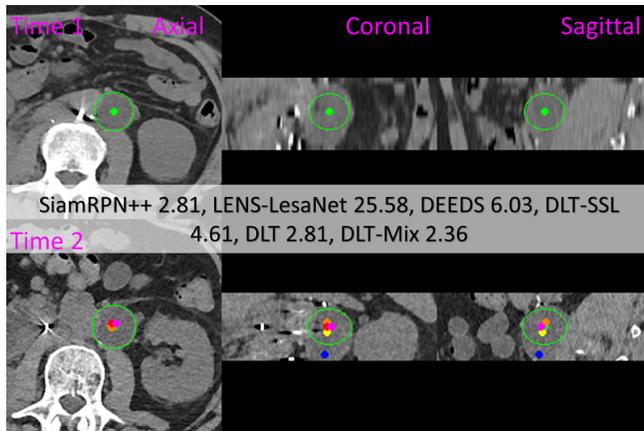
SiamRPN++ 7.98, LENS-LesaNet 34.04, DEEDS 5.83, DLT-SSL 16.18, DLT 3.97, DLT-Mix 1.37

● Ground truth ● SiamRPN++ ● LENS-LesaNet  
● DEEDS ● DLT-SSL ● DLT ● DLT-Mix



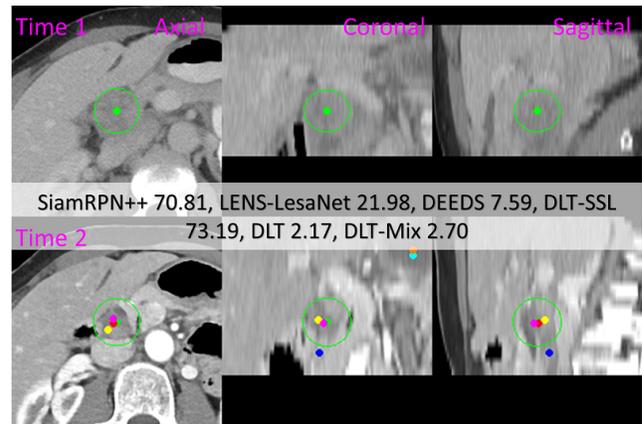
SiamRPN++ 5.25, LENS-LesaNet 5.01, DEEDS 10.34, DLT-SSL 5.25, DLT 2.69, DLT-Mix 2.69

● Ground truth ● SiamRPN++ ● LENS-LesaNet  
● DEEDS ● DLT-SSL ● DLT ● DLT-Mix



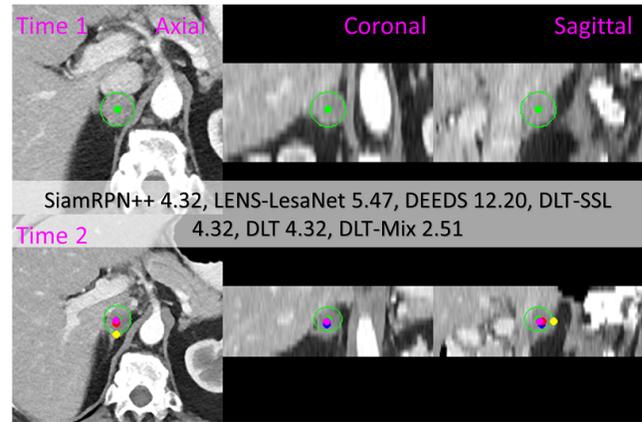
SiamRPN++ 2.81, LENS-LesaNet 25.58, DEEDS 6.03, DLT-SSL 4.61, DLT 2.81, DLT-Mix 2.36

● Ground truth ● SiamRPN++ ● LENS-LesaNet  
● DEEDS ● DLT-SSL ● DLT ● DLT-Mix



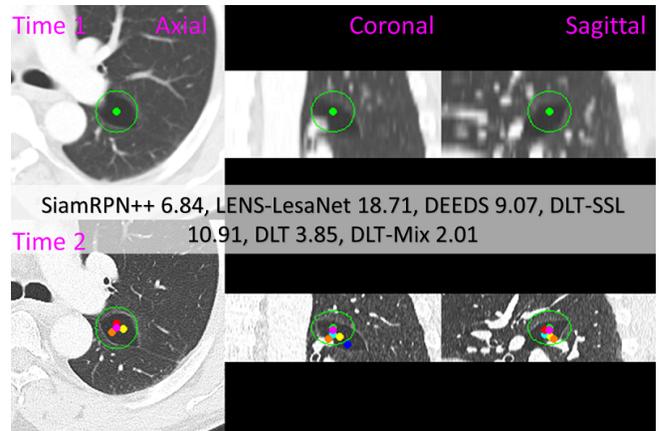
SiamRPN++ 70.81, LENS-LesaNet 21.98, DEEDS 7.59, DLT-SSL 73.19, DLT 2.17, DLT-Mix 2.70

● Ground truth ● SiamRPN++ ● LENS-LesaNet  
● DEEDS ● DLT-SSL ● DLT ● DLT-Mix



SiamRPN++ 4.32, LENS-LesaNet 5.47, DEEDS 12.20, DLT-SSL 4.32, DLT 4.32, DLT-Mix 2.51

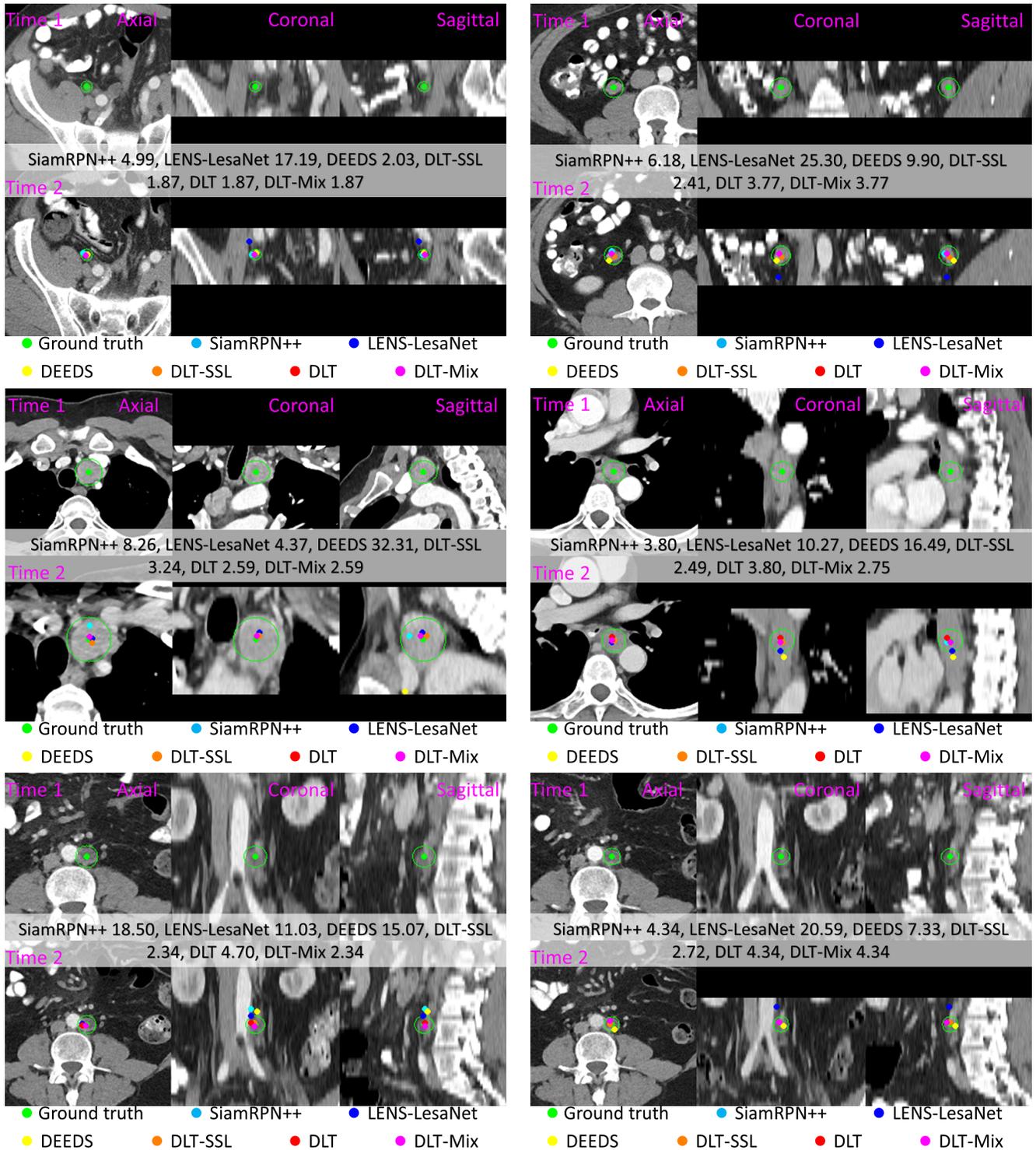
● Ground truth ● SiamRPN++ ● LENS-LesaNet  
● DEEDS ● DLT-SSL ● DLT ● DLT-Mix



SiamRPN++ 6.84, LENS-LesaNet 18.71, DEEDS 9.07, DLT-SSL 10.91, DLT 3.85, DLT-Mix 2.01

● Ground truth ● SiamRPN++ ● LENS-LesaNet  
● DEEDS ● DLT-SSL ● DLT ● DLT-Mix

SFigure 2: comparison of our methods, *i.e.*, DLT, DLT-SSL, DLT-Mix, with three state-of-the-art trackers including a Siamese networks based tracker – SiamRPN++, a leading registration algorithm – DEEDS, and a detector based tracker – LENS-LesaNet. Offsets from the predicted lesion centers to the manually labeled center are reported in *mm*.



SFigure 3: comparison of our methods, *i.e.*, DLT, DLT-SSL, DLT-Mix, with three state-of-the-art trackers including a Siamese networks based tracker – SiamRPN++, a leading registration algorithm – DEEDS, and a detector based tracker – LENS-LesaNet. Offsets from the predicted lesion centers to the manually labeled center are reported in *mm*.

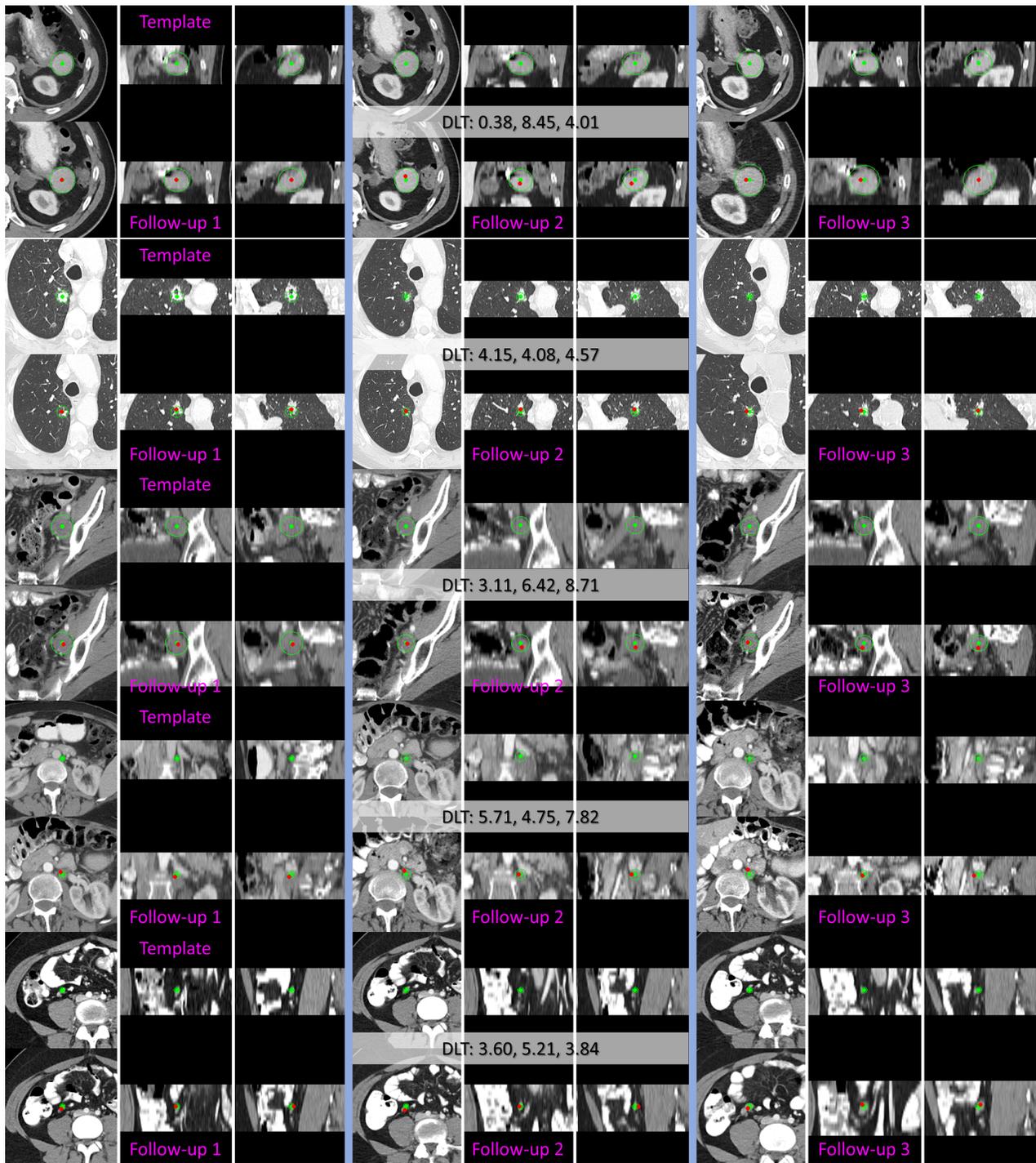
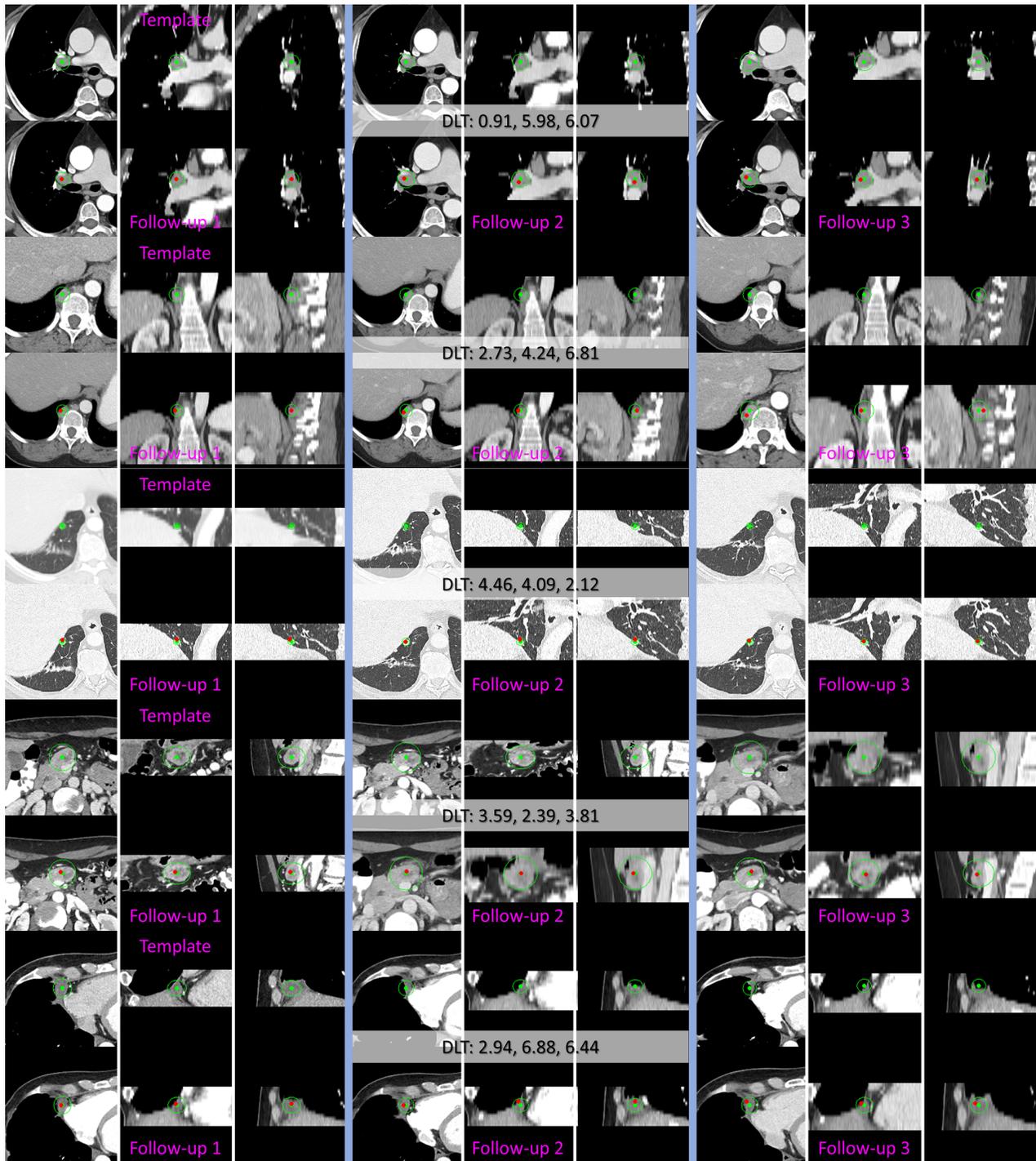
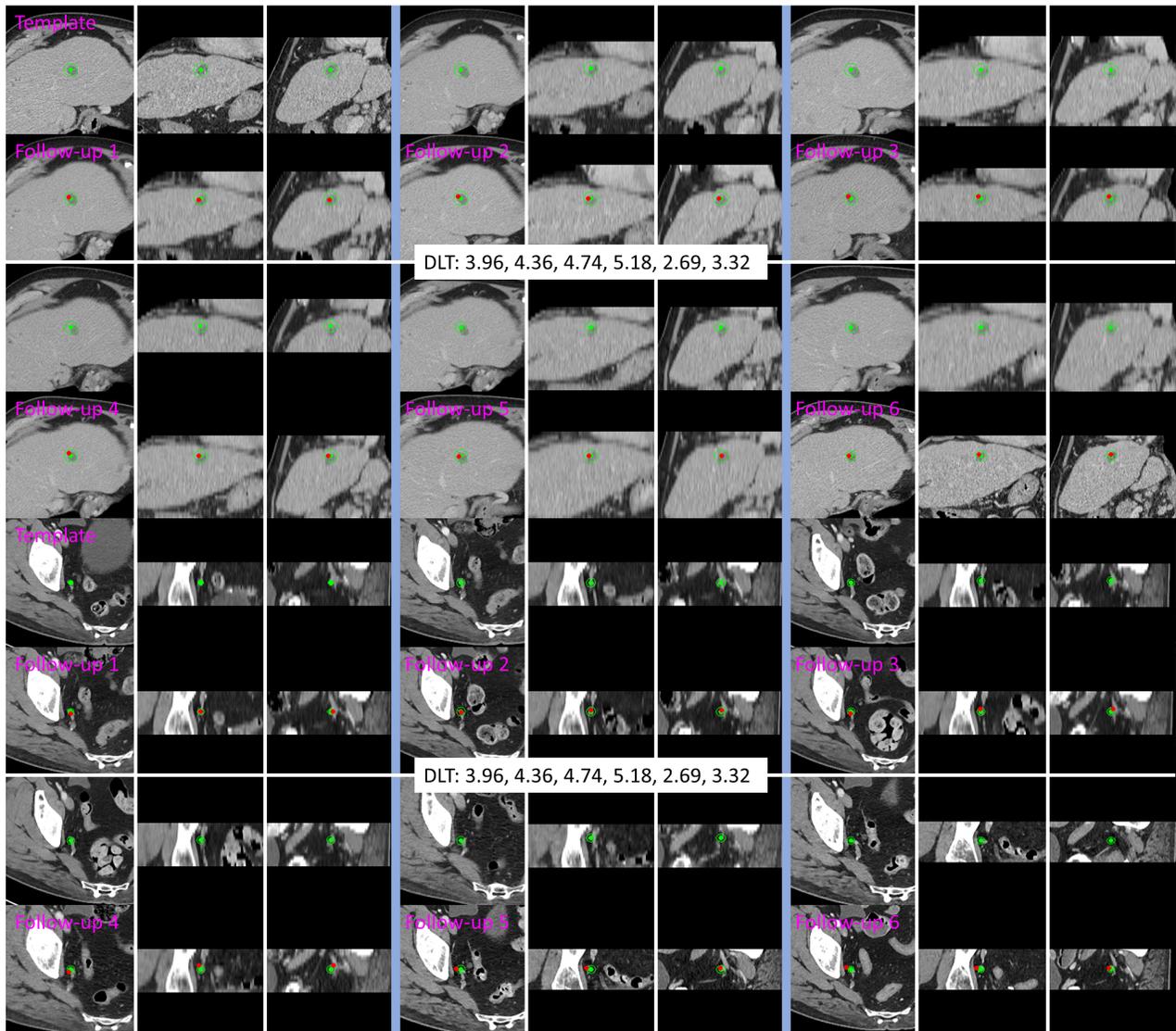


Figure 4: Lesion tracking through three follow ups using the proposed DLT. The template image is sampled from the first exam, and then follow-up 1, 2, and 3 are sampled from times of the second, third, and fourth exams, respectively. Green and red points represent the manually labeled and DLT predicted centers, respectively. Only the lesion center and radius at the first time point is given. Offsets from the DLT predicted lesion center to the manually labeled center are reported in *mm*.



SFigure 5: Lesion tracking through three follow ups using the proposed DLT. Green and red points present the manual labeled and DLT predicted centers, respectively. Only the lesion center and radius at the first time point is given. Offsets from the DLT predicted lesion center to the manual labeled center are reported in *mm*.



SFigure 6: Lesion tracking through six follow ups using the proposed DLT. Green and red points represent the manual labeled and DLT predicted centers, respectively. Only the lesion center and radius at the first time point is given. Offsets from the DLT predicted lesion center to the manual labeled center are reported in *mm*.