Intraoperative X-ray-Guided Tumor Tracking via Dynamic 3D/2D Registration with Task-Driven 3D Gaussian Priors

Haixiao Geng[1], Jingfan Fan[2✉], Danni Ai[2], Hong Song[3], Jinying Zhang[2], Feng Duan[4], and Jian Yang[2]

[1] School of Data Science, Qingdao University of Science and Technology, Qingdao, China
[2] School of Optics and Photonics, Bejing Institute of Technology, Bejing, China
fjf@bit.edu.cn
[3] School of Computer Science and Technology, Bejing Institute of Technology, Bejing, China
[4] Department of Interventional Radiology, Chinese PLA General Hospital First Medical Center, Beijing, China

Abstract. Intraoperative tumor tracking is critical for accurate targeting and safe margin control in X-ray-guided liver interventions. However, accurate tracking is difficult because fluoroscopy often provides weak or missing tumor appearance, and single-view X-ray images are insufficient to reliably constrain dynamic 3D deformation under severe structural overlap. This motivates a two-stage dynamic 3D/2D registration framework that first constructs a static Gaussian organ-target prior from preoperative CT and then deforms it under intraoperative X-ray sequence guidance. In Stage I, we build a label-aware static 3D Gaussian model with task-driven non-uniform sampling and constrained initialization, which preserves global liver morphology while improving the stability of internal tumor representation. In Stage II, we extract explicit inter-frame 2D motion cues from X-ray images and map them to 3D Gaussian deformation using a decomposed motion formulation with rigid-like sliding and residual deformation, parameterized by a sparse 3D control-point lattice. The deformation model is trained in a closed registration loop with joint 2D and 3D supervision for accurate and stable dynamic alignment. Experiments on simulated and retrospective clinical data demonstrate state-of-the-art performance in intraoperative 3D tumor tracking and deformation estimation.

Keywords: Tumor tracking· 3D/2D registration· Respiratory motion.

1 Introduction

In liver tumor intervention, intraoperative X-ray fluoroscopy is widely used for guidance because it is real-time and workflow-compatible. However, X-ray provides only 2D projections with weak soft-tissue contrast, severe anatomical overlap, and missing depth cues, making stable tumor localization and continuous

tracking difficult. Dynamic 3D tracking therefore aims to recover anatomy synchronized with X-ray images for 3D fusion guidance.

Existing dynamic 3D tracking methods can be grouped into reconstruction- and registration-based approaches. The former estimate 3D structures directly from single-view or multi-view X-ray images. Representative studies include GAN-based CT reconstruction from orthogonal X-ray views [20], neural-field/self-supervised CBCT reconstruction [22], and single-view dynamic reconstruction via differentiable DRR synthesis or patient-specific 4D modeling [16,17,20,21,22]. While these methods show the feasibility of X-ray-based 3D recovery, dynamic single-view reconstruction remains highly ill-posed and often relies on large-scale data or patient-specific optimization [17,21,18]. Registration-based methods are more consistent with interventional workflows because preoperative 3D priors are usually available and the target anatomy is known. Existing methods include mesh-based and voxel-based formulations [5,7,10,13,15]. Mesh-based methods can track organ surface deformation using DRR/X-ray image cues [15,18], but they mainly model surface geometry and are less suitable for liver tumor intervention, where the tumor is inside the liver and localization often depends on internal intensity/projection cues. Voxel-based methods can model both surface and internal deformation [10,5], but they are primarily applied to skeletal structures [10] or rely on patient-specific optimization [5], leaving cross-patient soft tissue and tumor motion tracking underexplored.

Current methods still face a key bottleneck in single-view X-ray-guided dynamic 3D tracking, where severe anatomical overlap, weak soft-tissue contrast, and missing depth cues make 3D-2D matching highly ill-posed. To address this issue, we introduce a registration-based framework that jointly exploits structural priors from segmentation labels and volumetric intensity information for tumor-relevant dynamic tracking, and adopt a differentiable Gaussian representation as a compact bridge between preoperative 3D anatomy and intraoperative X-ray observations. Thus, we consider dynamic 3D/2D registration as static Gaussian prior modeling followed by intraoperative dynamic guidance. The main contributions are threefold: (1) A clinical-task-driven two-stage framework is proposed for X-ray-guided liver tumor tracking, in which intraoperative 3D tumor tracking is formulated as a tumor-relevant dynamic 3D/2D registration task. (2) We construct a label-aware 3D Gaussian prior from preoperative CT and liver-tumor masks with tumor-focused non-uniform sampling, improving the representation of small tumors and subtle local deformation. (3) We introduce a 2D-motion-guided sparse-control deformation model to decompose liver-tumor motion into rigid-like sliding and residual deformation, enabling accurate 3D tumor tracking with motion consistency to inter-frame X-ray observations.

2 Method

2.1 Overview

We propose a two-stage framework for dynamic liver and intrahepatic tumor registration from intraoperative X-ray images, and formulate dynamic 3D/2D

registration as static 3D Gaussian prior modeling followed by intraoperative motion-driven deformation. In Stage I, we build a label-aware static 3D Gaussian model from preoperative CT and liver-tumor masks with non-uniform sampling and constrained initialization. In Stage II, we extract explicit inter-frame 2D motion cues from X-ray images and map them to 3D Gaussian deformation using a decomposed motion formulation with rigid-like sliding and residual deformation. Our method is designed for registration-oriented intraoperative dynamic tracking rather than full 4D anatomical reconstruction.

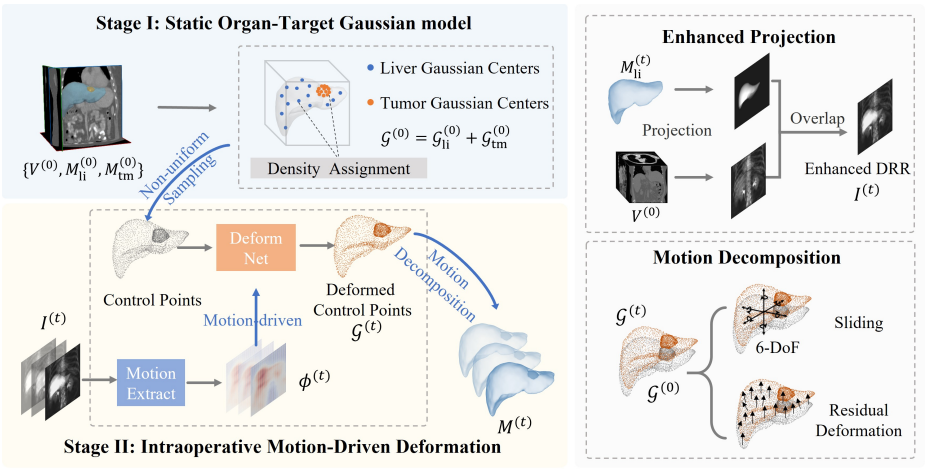

Fig. 1. An overview of the proposed two-stage framework.

2.2 Stage I: Static Organ-Target Gaussian Model

Given a preoperative CT volume $V^{(0)}$, a liver mask $M_{\text{li}}^{(0)}$ and a tumor mask $M_{\text{tm}}^{(0)}$ consist of $M^{(0)}$, i.e., $M^{(0)} = \{M_{\text{li}}^{(0)}, M_{\text{tm}}^{(0)}\}$ we construct a static Gaussian model as the union of liver Gaussians and tumor Gaussians:

$$\mathcal{G}^{(0)} = \mathcal{G}_{\text{li}}^{(0)} \cup \mathcal{G}_{\text{tm}}^{(0)}. \tag{1}$$

Each Gaussian is parameterized as $g_i = (\mu_i, \Sigma_i, a_i, \ell_i)$. $\mu_i \in \mathbb{R}^3$ is the Gaussian center, a_i denotes an X-ray absorption-related coefficient, and $\ell_i \in \{\text{li}, \text{tm}\}$ is the discrete label of liver and tumor. $\Sigma_i \in \mathbb{R}^{3\times3}$ is the covariance matrix, which is parameterized by a rotation matrix $R_i \in \mathbb{R}^{3\times3}$ and a scaling matrix $S_i \in \mathbb{R}^{3\times3}$, i.e., $\Sigma_i = R_i S_i S_i^\top R_i^\top$. For an arbitrary spatial location $\mathbf{x} \in \mathbb{R}^3$, the attenuation-related field induced by the static Gaussian model is written as $\rho^{(0)}(\mathbf{x}) = \rho_{\text{li}}^{(0)}(\mathbf{x}) + \rho_{\text{tm}}^{(0)}(\mathbf{x})$, where

$$\rho_{\text{li}}^{(0)}(\mathbf{x}) = \textstyle\sum_{i=1}^{N_{\text{li}}} a_i^{\text{li}} \mathcal{N}_i(\mathbf{x}), \qquad \rho_{\text{tm}}^{(0)}(\mathbf{x}) = \textstyle\sum_{i=1}^{N_{\text{tm}}} a_i^{\text{tm}} \mathcal{N}_i(\mathbf{x}), \tag{2}$$

where $\mathcal{N}_i(\mathbf{x}) = \exp\big(-\frac{1}{2}(\mathbf{x}-\mu_i)^\top \Sigma_i^{-1}(\mathbf{x}-\mu_i)\big)$ defines the i-th 3D Gaussian kernel. The quadratic form characterizes the Mahalanobis distance between $\mathbf{x}$ and the Gaussian center μ_i.

Instead of using a uniformly sampled generic Gaussian representation, we construct the static model with a label-aware sampling and initialization scheme for liver-tumor dynamic registration from X-ray images. Specifically, we sample $\mathcal{G}_{\text{li}}^{(0)}$ and $\mathcal{G}_{\text{tm}}^{(0)}$ separately to preserve an explicit internal tumor representation. $\mathcal{G}_{\text{tm}}^{(0)}$ is assigned a higher sampling density than $\mathcal{G}_{\text{li}}^{(0)}$, improving target geometry and centroid stability under respiratory motion and weak tumor visibility in fluoroscopy. μ_i, Σ_i and a_i are initialized from mask-based volumetric sampling, local sampling spacing, and CT intensities, respectively. This strategy provides a stable patient-specific organ-target Gaussian prior by assigning denser sampling to tumor-relevant regions, without requiring unconstrained Gaussian optimization.

2.3 Stage II: Intraoperative Motion-Driven Gaussian Deformation

Stage II is formulated as a registration-by-rendering loop that deforms the static organ-target Gaussian prior over time and aligns it to intraoperative X-ray images in the projection domain. Inspired by the sparse-control deformation idea in [14], we use a simpler control-point-driven deformation design tailored to dynamic 3D/2D registration, which avoids the more complex generative dynamic modeling pipeline and reduces training difficulty.

DRR-based registration loop. Given a static Gaussian prior $\mathcal{G}^{(0)}$ and corresponding label mask $M^{(0)}$, we estimate frame-wise deformation parameters to obtain a deformed Gaussian model $\mathcal{G}^{(t)}$ and deformed mask $M^{(t)}$, and render $\mathcal{G}^{(t)}$ to the X-ray plane via a DRR step [9,3]: $\hat{I}^{(t)} = \text{DRR}\big(\mathcal{G}^{(t)}\big)$. This forms a 3D/2D registration loop between 3D Gaussian deformation and 2D X-ray observations.

2D motion extraction from X-ray images. To improve stability under weak soft-tissue visibility and structural overlap, we first generate an enhanced DRR by combining the projection of the deformed liver mask with the CT-derived DRR, providing a clearer liver-region cue for subsequent motion estimation. We then extract an explicit inter-frame 2D motion cue from adjacent X-ray frames:

$$\phi^{(t)} = \mathcal{F}_{\text{2D\_mot}}(I^{(t)}, I^{(t+1)}), \tag{3}$$

where $\mathcal{F}_{\text{2D\_mot}}(\cdot)$ is a VoxelMorph-style U-Net encoder-decoder [2] for 2D motion extraction, and $\phi^{(t)}$ serves as a temporal motion cue.

Motion-guided sparse-control Gaussian deformation. We parameterize dynamic deformation using a sparse 3D control-point lattice in the liver region. A motion-guided deformation network adapted from a Transformer backbone [4] predicts rigid-like sliding motion and residual control-point displacements from the 2D motion cue $\phi^{(t)}$ and 3D control points $\mathcal{P}$:

$$(\boldsymbol{\xi}^{(t)}, \Delta\mathcal{P}_{\text{def}}^{(t)}) = \mathcal{F}_{\text{def}}(\phi^{(t)}, \mathcal{P}). \tag{4}$$

$\boldsymbol{\xi}^{(t)} \in \mathbb{R}^6$ denotes sliding parameters and $\Delta\mathcal{P}_{\text{def}}^{(t)}$ denotes residual control-point displacements. We then update each Gaussian center by combining the sliding transform and the interpolated residual deformation:

$$\mu_i^{(t)} = T(\boldsymbol{\xi}^{(t)})\,\mu_i^{(0)} + \Delta\mu_i^{(t)}(\Delta\mathcal{P}_{\text{def}}^{(t)}), \tag{5}$$

where $\Delta\mu_i^{(t)}(\cdot)$ denotes the residual displacement interpolated from the control-point lattice at the sliding-transformed Gaussian center $T(\boldsymbol{\xi}^{(t)})\mu_i^{(0)}$. In Stage II, only the Gaussian centers are updated by default, while the remaining attributes are kept fixed for stable registration-oriented optimization.

Training optimization. Stage II is optimized in the closed registration loop using the joint 2D/3D consistency and smoothness objectives defined in Sec. 2.4.

2.4 Training Objectives

We optimize the Stage II model using a joint objective with 2D projection consistency, 3D label-volume supervision, and deformation smoothness regularization:

$$\mathcal{L} = \mathcal{L}_{\text{2D}} + \lambda_{\text{3D}}\mathcal{L}_{\text{3D}} + \lambda_{\text{sm}}\mathcal{L}_{\text{sm}}, \tag{6}$$

where λ_{3D} and λ_{sm} are the corresponding weights.

2D and 3D losses. $\mathcal{L}_{\text{2D}}$ enforces projection alignment, while $\mathcal{L}_{\text{3D}}$ reduces depth ambiguity and preserve organ-target structural consistency. After deforming the static Gaussian model and projecting it to generate the DRR image $\hat{I}^{(t)}$, we compare $\hat{I}^{(t)}$ with the input X-ray image $I^{(t)}$ using a 2D image loss that combines LNCC and gradient consistency terms. Also, we use an MSE term to enforce voxel-wise label consistency and a DSC term to improve overlap between the deformed organ-target volumes [1,2].

Smoothness regularization. To stabilize optimization and suppress implausible high-frequency deformation, we impose smoothness regularization on the residual deformation control-point displacements in the sparse 3D lattice. We use spatial smoothness terms: $\mathcal{L}_{\text{sm}} = \|\nabla_{\text{sp}}\Delta\mathcal{P}_{\text{def}}^{(t)}\|_2^2$, where $\Delta\mathcal{P}_{\text{def}}^{(t)}$ denotes the residual deformation control-point displacements at frame t, and ∇_{sp} denotes spatial finite differences on the 3D control lattice.

3 Experiments and results

3.1 Experiment Settings

Datasets. We use 16 cases of 4D liver-tumor CT with 10 respiratory phases (training) and 7 cases of CT/real X-ray image pairs (testing) for retrospective study. All CT volumes are resampled to isotropic 1 mm spacing and normalized to 128 axial slices. To improve the structure and deformation variation of our dataset, we expand each 4D-CT data into multiple training samples using adjacent, interval, and bidirectional phase pairs, yielding substantially more training samples. Further, we apply initial perturbation during DRR generation.

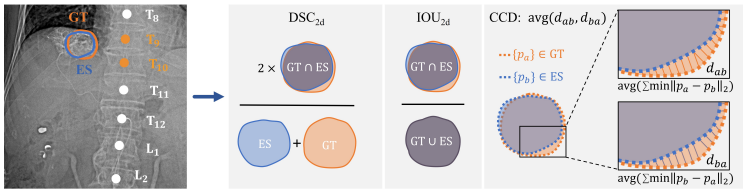

Fig. 2. Examples of tumor region annotations and clinical evaluation metrics.

Implementation Details. A two-stage construction-and-optimization strategy is adopted in our framework, where Stage I constructs and initializes the static Gaussian prior and Stage II optimizes the motion module while updating only Gaussian centers by default. We use the Adam optimizer for Stage II with an initial learning rate of 1×10^{-5} and a decay rate of 0.95. λ_{3D} and λ_{sm} are set to 2 and 0.1, respectively. The coarse control lattice resolution is $4 \times 4 \times 4$. The training adopts a 3-fold cross-validation strategy. All experiments are conducted on a single NVIDIA GeForce RTX 4090 GPU.

Evaluation metrics. We adopt evaluation metrics from prior registration studies [12,2,6], including mean distance (MD), Hausdorff distance (HD), and Dice similarity coefficient (DSC) and volume difference rate (VDR), to quantitatively evaluate registration performance on the simulated data. A clinical ground truth for quantitative evaluation is established by three senior physicians, with tumor contours jointly delineated on from retrospective X-ray image sequences using the lipiodol region and vertebral position (Fig. 2). Three 2D metrics are then adopted to assess tumor localization and deformation estimation accuracy: tumor area Dice score (DSC_{2d}), tumor boundary intersection over union (IoU_{2d}), and closest contour distance error (CCD).

Table 1. Quantitative results (mean ± std) of compared methods on simulated data.

Anatomy	Method	MD (mm) ↓	HD (mm) ↓	DSC (%) ↑	VDR (%) ↓
Liver	Affine [2]	3.37 ± 1.92	11.86 ± 4.93	90.81 ± 5.78	3.89 ± 1.40
	VoxelMorph [2]	3.12 ± 1.37	11.45 ± 3.79	91.84 ± 4.17	3.54 ± 1.86
	TransMorph [4]	2.82 ± 1.03	9.23 ± 3.14	93.94 ± 2.56	2.85 ± 1.39
	MambaMorph [19]	2.98 ± 1.21	10.48 ± 3.62	92.32 ± 1.88	2.96 ± 1.55
	DiffuseMorph [11]	2.75 ± 0.62	8.74 ± 2.35	94.53 ± 1.95	3.17 ± 1.65
	Ours	**2.13 ± 0.63**	**7.52 ± 1.94**	**95.03 ± 1.74**	**2.14 ± 1.05**
Tumor	Affine [2]	2.84 ± 1.77	7.93 ± 3.41	78.25 ± 14.68	7.22 ± 3.36
	VoxelMorph [2]	2.46 ± 1.45	6.35 ± 1.94	83.39 ± 12.17	6.32 ± 2.72
	TransMorph [4]	1.95 ± 0.92	6.12 ± 1.96	85.72 ± 10.32	5.04 ± 2.23
	MambaMorph [19]	1.87 ± 0.78	4.98 ± 1.85	88.15 ± 6.30	3.41 ± 1.82
	DiffuseMorph [11]	1.74 ± 0.81	5.14 ± 1.80	87.89 ± 8.21	4.38 ± 1.97
	Ours	**1.21 ± 0.40**	**4.19 ± 1.75**	**91.34 ± 5.68**	**2.58 ± 1.51**

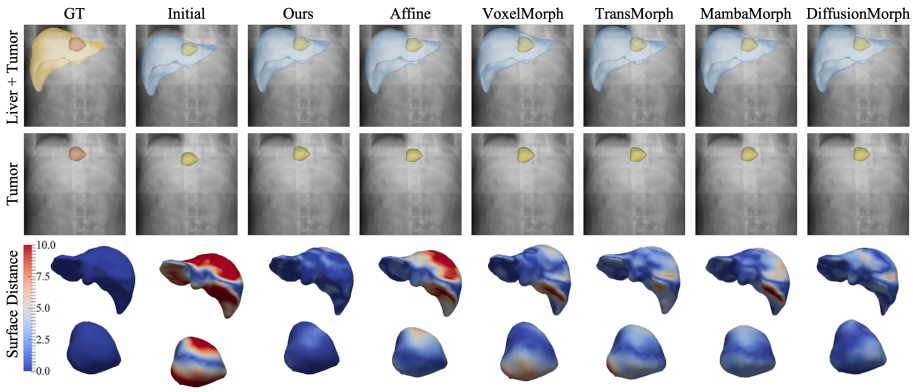

Fig. 3. Visual comparison of liver and tumor registration. "GT" denotes the ground truth at end-exhalation, "Initial" indicates before registration at end-inhalation.

3.2 Comparison with Related Methods

We compared our method with five voxel-based registration baselines, including VoxelMorph [2], TransMorph [4], MambaMorph [19], DiffuseMorph [11] and Affine derives from Voxelmorph to achive affine matrix. For a fair comparison, all baselines were adapted to the same 3D/2D registration protocol, with CT liver-tumor labels as the source and DRR or X-ray images as the target. The quantitative results of simulated data in Table 1 demonstrate that our method outperforms the competing methods (p-value < 0.01 by Wilcoxon Test). The qualitative results in Fig. 3 further show that the liver and tumor predicted by our method are the closest to the ground truth in both shape and volume.

To evaluate the generalization capability of the proposed method to real X-ray image guided liver tumor intervention, we directly evaluated it on the clinical cases. Before tumor tracking, the patient pose between CT and X-ray is first initialized by rigid registration [8]. Subsequently, following [8], we use a CycleGAN-based domain translation model [23] to translate X-ray images into simulated DRR images with enhanced liver-region contrast, which are then fed into the subsequent pipeline. Fig. 4 presents the qualitative (left) and quantitative (right) results of our method over a continuous respiratory cycle. The liver and tumor exhibit motion and deformation consistent with the X-ray images, and show high overlap with the green ground-truth contours. The average inference runtime is 62 ms per frame, corresponding to 16.1 fps.

3.3 Ablation Study

We design a ablation study and summarize all results in Table 2. The ablations focus on one key design from each stage, including **w/o Task-Driven Sampling:** replacing the label-aware non-uniform Gaussian sampling in Stage I with uniform sampling, **w/o 2D Motion Guidance:** removing the explicit

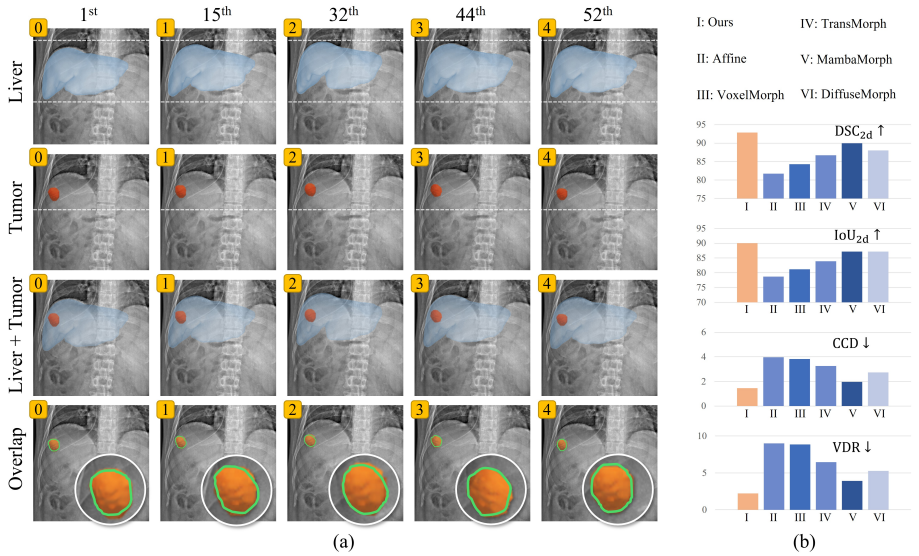

Fig. 4. (a) Clinical registration results across a continuous respiratory cycle. Yellow tags 0–4 denote end-inhalation, mid-exhalation, end-exhalation, mid-inhalation, and approximately 90% of the cycle. The green contour indicates the physician-annotated ground truth. (b) Quantitative results of tumor registration on four metrics.

Table 2. Tumor registration results of ablation studies on simulated data.

Methods	MD (mm) ↓	HD (mm)(%) ↓	DSC (%) ↑	VDR (%) ↓
Full (Ours)	**1.21 ± 0.40**	**4.19 ± 1.75**	**91.34 ± 5.68**	**2.58 ± 1.51**
W/o Task-Driven Sampling	1.83 ± 0.99	4.48 ± 1.83	88.95 ± 7.62	3.01 ± 1.94
W/o 2D Motion Guidance	2.35 ± 1.63	5.26 ± 2.21	85.24 ± 10.96	4.93 ± 2.38
W/o Motion Decomposition	2.07 ± 1.67	4.82 ± 1.89	87.73 ± 8.32	3.78 ± 1.27

inter-frame 2D motion cue in Stage II, and **w/o 3D Motion Decomposition:** removing the sliding and deformation decomposition by replacing it with a single-branch deformation model. The results show that removing explicit 2D motion guidance causes the largest performance drop, mainly due to reduced dynamic alignment stability under weak X-ray visibility and structural overlap. Replacing the motion decomposition with a single-branch deformation model also degrades tumor localization and deformation estimation, indicating the benefit of explicitly separating global rigid-like motion from residual deformation. Removing task-driven sampling further reduces tumor registration accuracy, confirming the value of label-aware non-uniform Gaussian construction for preserving weakly visible internal targets.

4 Conclusion

We present a task-driven framework for dynamic 3D liver-tumor localization and deformation estimation under continuous X-ray guidance. The framework decomposes dynamic registration into static 3D Gaussian modeling and motion-guided refinement. The label-aware Gaussian representation improves tumor modeling stability while preserving global liver morphology, and the motion-guided refinement enhances registration accuracy and temporal stability under weak X-ray visibility. Experiments on simulated and clinical data show improved tumor localization, deformation estimation, and continuous-frame tracking over competing methods, with good temporal consistency and structural fidelity. These results support the potential of our framework for reliable intraoperative navigation in X-ray-guided liver tumor intervention and other physiological-motion-affected scenarios, such as lung and cardiac interventions. Future work will expand the clinical dataset to assess generalizability and robustness.

Acknowledgments. This study was supported in part by the National Key R&D Program of China (2023YFC2415300) and the National Science Foundation Program of China (62422102).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis **12**(1), 26–41 (2008)
2. 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)
3. Cai, Y., Liang, Y., Wang, J., Wang, A., Zhang, Y., Yang, X., Zhou, Z., Yuille, A.: Radiative gaussian splatting for efficient x-ray novel view synthesis. In: European Conference on Computer Vision. pp. 283–299. Springer (2024)
4. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. Med. Image Anal. **82**, 102615 (2022)
5. Dai, J., Dong, G., Zhang, C., He, W., Liu, L., Wang, T., Jiang, Y., Zhao, W., Zhao, X., Xie, Y.: Volumetric tumor tracking from a single cone-beam x-ray projection image enabled by deep learning. Medical Image Analysis **91**, 102998 (2024)
6. Fan, J., Cao, X., Wang, Q., Yap, P.T., Shen, D.: Adversarial learning for mono-or multi-modal registration. Med. Image Anal. **58**, 101545 (2019)
7. Geng, H., Fan, J., Yang, S., Chen, S., Xiao, D., Ai, D., Fu, T., Song, H., Yuan, K.: DSC-Recon: Dual-Stage complementary 4-D organ reconstruction from X-Ray image sequence for intraoperative fusion. IEEE Transactions on Medical Imaging **43**(11), 3909–3923 (2024)
8. Geng, H., Xiao, D., Yang, S., Fan, J., Fu, T., Lin, Y., et al.: CT2X-IRA: CT to X-ray image registration agent using domain-cross multi-scale-stride deep reinforcement learning. Phys. Med. Biol. **68**(17), 175024 (2023)

9. Gopalakrishnan, V., Dey, N., Golland, P.: Intraoperative 2d/3d image registration via differentiable x-ray rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11662–11672 (2024)
10. Jiang, Y., Pei, Y., Xu, T., Yuan, X., Zha, H.: Towards semantically-consistent deformable 2d-3d registration for 3d craniofacial structure estimation from a single-view lateral cephalometric radiograph. IEEE Transactions on Medical Imaging (2024)
11. Kim, B., Han, I., Ye, J.C.: Diffusemorph: Unsupervised deformable image registration using diffusion model. In: European conference on computer vision. pp. 347–364. Springer (2022)
12. Kim, J., Valdes-Hernandez, M.d.C., Royle, N.A., Park, J.: Hippocampal shape modeling based on a progressive template surface deformation and its verification. IEEE transactions on medical imaging **34**(6), 1242–1261 (2014)
13. Laumer, F., Amrani, M., Manduchi, L., Beuret, A., Rubi, L., Dubatovka, A., Matter, C.M., Buhmann, J.M.: Weakly supervised inference of personalized heart meshes based on echocardiography videos. Medical image analysis **83**, 102653 (2023)
14. Li, Z., Chen, Y., Liu, P.: Dreammesh4d: Video-to-4d generation with sparse-controlled gaussian-mesh hybrid representation. Advances in Neural Information Processing Systems **37**, 21377–21400 (2024)
15. Nakao, M., Nakamura, M., Matsuda, T.: Image-to-graph convolutional network for 2D/3D deformable model registration of low-contrast organs. IEEE Transactions on Medical Imaging **41**(12), 3747–3761 (2022)
16. Peng, C., Liao, H., Wong, G., Luo, J., Zhou, S.K., Chellappa, R.: Xraysyn: realistic view synthesis from a single radiograph through ct priors. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 436–444 (2021)
17. Shen, L., Zhao, W., Xing, L.: Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. Nature Biomedical Engineering **3**(11), 880–888 (2019)
18. Wang, Y., Zhong, Z., Hua, J.: Deeporgannet: on-the-fly reconstruction and visualization of 3d/4d lung models from single-view projections by deep deformation network. IEEE transactions on visualization and computer graphics **26**(1), 960–970 (2019)
19. Wang, Y., Guo, T., Yuan, W., Shu, S., Meng, C., Bai, X.: Mamba-based deformable medical image registration with an annotated brain mr-ct dataset. Computerized Medical Imaging and Graphics **123**, 102566 (2025)
20. Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., Zheng, Y.: X2CT-GAN: Reconstructing CT from biplanar X-rays with generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10619–10628 (2019)
21. Zakeri, A., Hokmabadi, A., Nix, M.G., Gooya, A., Wijesinghe, I., Taylor, Z.A.: 4d-precise: Learning-based 3d motion estimation and high temporal resolution 4dct reconstruction from treatment 2d+ t x-ray projections. Computer Methods and Programs in Biomedicine **250**, 108158 (2024)
22. Zha, R., Zhang, Y., Li, H.: Naf: neural attenuation fields for sparse-view cbct reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 442–452. Springer (2022)
23. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. IEEE Int. Conf. Comput. Vis. pp. 2223–2232 (2017)