

RadGS-Reg: Registering Spine CT with Biplanar X-rays via Joint 3D Radiative Gaussians Reconstruction and 3D/3D Registration

Ao Shen¹, Xueming Fu², Junfeng Jiang³ ✉, Qiang Zeng³, Ye Tang¹, Zhengming Chen¹, Luming Nong⁴, Feng Wang⁵, and S. Kevin Zhou² ✉

¹ College of Information Science and Engineering, Hohai University (HHU), Changzhou Jiangsu, 213200, China

² School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China (USTC), Hefei Anhui, 230026, China
skevinzhou@ustc.edu.cn

³ College of Artificial Intelligence and Automation, HHU, Changzhou Jiangsu, 213200, China
jiangjf.hhu@gmail.com

⁴ The Third Affiliated Hospital of Nanjing Medical University, Changzhou Jiangsu, 213164, China

⁵ Tuodao Medical Technology Co., Ltd., Nanjing Jiangsu, 210012, China

Abstract. Computed Tomography (CT)/X-ray registration in image-guided navigation remains challenging because of its stringent requirements for high accuracy and real-time performance. Traditional "render and compare" methods, relying on iterative projection and comparison, suffer from spatial information loss and domain gap. 3D reconstruction from biplanar X-rays supplements spatial and shape information for 2D/3D registration, but current methods are limited by dense-view requirements and struggles with noisy X-rays. To address these limitations, we introduce RadGS-Reg, a novel framework for vertebral-level CT/X-ray registration through joint 3D Radiative Gaussians (RadGS) reconstruction and 3D/3D registration. Specifically, our biplanar X-rays vertebral RadGS reconstruction module explores learning-based RadGS reconstruction method with a Counterfactual Attention Learning (CAL) mechanism, focusing on vertebral regions in noisy X-rays. Additionally, a patient-specific pre-training strategy progressively adapts the RadGS-Reg from simulated to real data while simultaneously learning vertebral shape prior knowledge. Experiments on in-house datasets demonstrate the state-of-the-art performance for both tasks, surpassing existing methods. The code is available at: github.com/shenao1995/RadGS_Reg.

Keywords: 2D/3D Registration · 3D Reconstruction · Synergistic Training · 3D Gaussian Splatting · Counterfactual Attention.

✉ Junfeng Jiang and S. Kevin Zhou are joint corresponding authors.

1 Introduction

Computed Tomography (CT)/X-ray registration has been extensively studied as an unresolved issue in the realm of image-guided navigation, characterized by its stringent requirements for accuracy and real-time performance [10]. Over the past two decades, numerous methods have been devised to address these critical clinical demands; however, persistent challenges remain. Primarily, akin to the "render-and-compare" paradigm in computer vision [16,21,4], the majority of traditional approaches have been predominantly projection-based. These methods iteratively refine the CT volume's pose by repeatedly projecting the 3D CT volume onto the detector plane, followed by a comparison of the resultant projected image, known as Digitally Reconstructed Radiographs (DRR), with the target X-ray. Nonetheless, the DRR image inherently loses information from its 3D CT counterpart. Furthermore, contingent upon the nature of the projection process, various strategies have been proposed to mitigate the domain gap between DRR and X-ray [27,7], as well as to enhance the efficiency of DRR generation [18,9]. The advent of the 3D Gaussian Splatting (3DGS) offers the potential to significantly reduce the domain gap and improve efficiency substantially [15,2]. Additionally, existing methods [17,13] for CT/X-ray registration concerning spinal registration often fail to account for variations in the patient's spinal poses as captured in CT and X-rays.

Inspired by recent progress in 3D reconstruction in the field of computer vision, we present a novel approach termed **RadGS-Reg** to joint learning of **R**adiative **G**aussians [26] **R**econstruction and 3D/3D **R**egistration for CT/X-ray Registration. This methodology entails the transformation of biplanar X-rays inputs into Radiative Gaussians (RadGS), which is subsequently registered with the provided CT volume. The task of accurately reconstructing RadGS using only biplanar X-rays presents a significant challenge. To address this, we unveil the synergistic interaction between 3D reconstruction and CT/3D Gaussians registration. Specifically, the shape of the preoperative CT volume constitutes the ultimate target within the 3D reconstruction process. Concurrently, the pose of the RadGS derived from biplanar X-rays is registered with the target pose of the preoperative CT volume. To mitigate the issue of spinal pose variation in CT and X-ray, we utilize the visually-grounded, region-specific vertebral-level identified in X-rays, along with the vertebral-level in the CT volume segmented using existing methods [19] for vertebrae segmentation. To address the complication posed by the overlap of adjacent vertebral joints, which may impede the reconstruction of individual vertebrae, we integrate Counterfactual Attention Learning (CAL) [22], concentrating on vertebra regions to enhance reconstruction accuracy.

Our primary contributions are delineated as follows. (1) A comprehensive deep learning framework for CT/X-ray registration through sequential 3D reconstruction and 3D/3D registration, leveraging the synergy between reconstruction and registration to enhance accuracy and efficiency. (2) A vertebral-level approach using integrated CAL to reduce interference from adjacent vertebrae, focusing on vertebral regions during 3D reconstruction from X-rays. (3) Improved

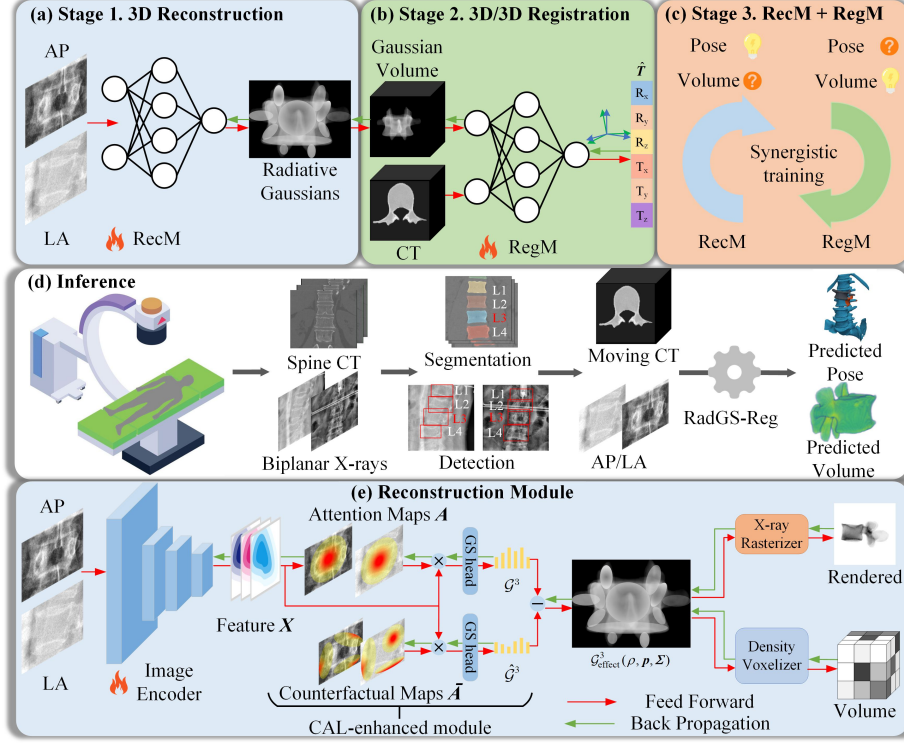


Fig. 1. The overall framework of RadGS-Reg (a-d) and the biplanar X-rays RadGS reconstruction module (e). The framework consists of four parts: (a) 3D reconstruction, (b) 3D/3D registration, (c) synergistic training, and (d) inference phase.

training efficiency through module-aware and data-aware strategies, with results surpassing state-of-the-art methods in 3D reconstruction and registration.

2 Method

The proposed end-to-end deep learning framework is comprehensively depicted in Fig. 1, consisting of two fundamental components: the 3D reconstruction module and the 3D/3D registration module, denoted RecM and RegM, respectively. Specifically, as shown in Fig. 1(a), RecM utilizes biplanar X-rays as inputs to yield predictions of RadGS $\mathcal{G}^3 = \{\mathcal{G}_i^3(\rho_i, \mathbf{p}_i, \Sigma_i)\}_{i=1, \dots, N}$, where $\rho_i, \mathbf{p}_i \in \mathbb{R}^3$ and $\Sigma_i \in \mathbb{R}^{3 \times 3}$ are learnable parameters representing central density, position, and covariance, respectively. The predicted RadGS are voxelized into a volumetric form and fed into the registration module (Fig. 1(b)) along with the input CT volume, with RegM outputting the intraoperative reconstructed volume 6-Degrees Of Freedom (DOF) pose relative to the preoperative CT.

During inference (as illustrated in Fig. 1(d)), biplanar X-rays with camera parameters and CT volume are preprocessed using X-ray detection [14] and CT segmentation [19] to extract same-level vertebrae before input to RadGS-Reg, enabling vertebral-level reconstruction and registration. The output includes both the reconstructed volume from biplanar X-rays and the rigidly registered CT volume. The training procedures for reconstruction, registration, and their synergistic training are detailed separately.

3D Reconstruction Module Our RecM consists of four sub-modules: an image encoder, a CAL-enhanced module, an X-ray rasterizer [26], and a density voxelizer [26], as illustrated in Fig. 1(e). The structure of RecM is described below.

To address the challenges of blurred vertebral features in real X-rays, we propose adding a CAL module [22] before the Gaussians Prediction Head (GSHead), consisting of a sequence of a linear layer, a ReLU layer, and another linear layer. The CAL module guides the network to focus on factual regions by leveraging counterfactual causality to quantify and optimize attention quality. Specifically, given input features $\mathbf{X}$ and factual attention maps $\mathbf{A}$, we leverage an intervention operation $do(\cdot)$ to generate counterfactual attention maps $\bar{\mathbf{A}}$ by replacing the factual attention maps $\mathbf{A}$ with random weights. The final RadGS prediction $\mathcal{G}_{\text{effect}}^3$ can be expressed as the difference between the factual prediction $\mathcal{G}^3$ and its counterfactual alternative $\hat{\mathcal{G}}^3$, as shown in Eq. (1).

$$\mathcal{G}_{\text{effect}}^3 = \mathbb{E}_{A \sim \gamma} [Y(A = \mathbf{A}, X = \mathbf{X}) - Y(do(A = \bar{\mathbf{A}}), X = \mathbf{X})], \quad (1)$$

where γ is the distribution of counterfactual attentions, Y is the GSHead.

Then the predicted RadGS are rendered by X-ray rasterizer and concurrently voxelized by a density voxelizer. Finally, the overall reconstruction loss Eq. (2) is defined as follows.

$$\mathcal{L}_{CAL-rec} = \mathcal{L}_1(I_{\text{effect}}, I_m) + \lambda_1 \mathcal{L}_{SSIM}(I_{\text{effect}}, I_m) + \lambda_2 \mathcal{L}_{tv}(V_{\text{effect}}) + \mathcal{L}_{rec}, \quad (2)$$

where I_{effect} and V_{effect} represent the projected image and volume, respectively, obtained by $\mathcal{G}_{\text{effect}}^3$ through the rasterizer and voxelizer, while I_m denotes the measured projection. In addition to utilizing photometric L1 loss $\mathcal{L}_1$ and Structural Similarity Index Measure (SSIM) [25] loss $\mathcal{L}_{SSIM}$, we further incorporate a 3D total variation (TV) [24] regularization term $\mathcal{L}_{TV}$ as a homogeneity prior for tomography. The loss $\mathcal{L}_{rec}$ also comprises $\mathcal{L}_1$, $\mathcal{L}_{SSIM}$, and $\mathcal{L}_{TV}$, but its input is derived from the output generated from the original RadGS $\mathcal{G}^3$.

3D/3D Registration Module The RegM processes the concatenated Gaussian volume channel-wise and the preoperative CT volume V_{rec} as input, and regresses the reconstructed volume rigid 6-DOF pose relative to the CT volume. We simply employ an image encoder and a pose head for pose regression. The registration loss is defined as follows.

$$\mathcal{L}_{reg} = \mathcal{L}_{NCC}(\mathbf{V}_{CT} \cdot \hat{\mathbf{T}}, \mathbf{V}_{rec}) + \mathcal{L}_{SSIM}(\mathbf{V}_{CT} \cdot \hat{\mathbf{T}}, \mathbf{V}_{rec}) + \lambda \mathcal{L}_{geo}(\mathbf{T}_{gt}, \hat{\mathbf{T}}), \quad (3)$$

where $\mathcal{L}_{NCC}$ represents a 3D-extended of the normalized cross-correlation [20] loss, and $\mathcal{L}_{geo}$ is utilized to compute the geodesic distance [23] between the estimated $\hat{\mathbf{T}}$ and Ground Truth (GT) poses $\mathbf{T}_{gt}$.

RadGS-Reg Training Strategy We implement module-aware and data-aware strategies to train RadGS-Reg model. For module-aware training, we first train the 3D reconstruction (Fig. 1(a)) and registration (Fig. 1(b)) modules independently, followed by synergistic training (Fig. 1(c)) with the total loss, as shown in Eq. (4):

$$\mathcal{L}_{total} = \mathcal{L}_{reg} + \mathcal{L}_{CAL-rec}, \quad (4)$$

making the RadGS-Reg model-agnostic. For data-aware training, we adopt a three stage pre-training strategy, as detailed in Sec. 3.2.

3 Experiments

3.1 Data Preparation

Our datasets are sourced from the publicly available dataset VERSE '20 [1] and an in-house dataset collected from a medical institution.

The VERSE '20 dataset includes 253 lumbar spine CT scans extracted from 300 cases, containing a total of 1,280 vertebrae. DRRs were synthesized using a C-arm simulator (976×976 resolution, 1,124 mm focal length) with $\pm 15^\circ$ perturbations in Anterior Posterior (AP) and Lateral (LA) views, yielding 32,000 biplanar DRR pairs for RecM first stage pre-training. For RegM, 1,280 CT pairs were generated through pose sampling with parameters randomly sampled: rotations in $[-20^\circ, 20^\circ]$ and translations in $[-50 \text{ mm}, 50 \text{ mm}]$.

The in-house dataset includes 10 real-scent intraoperative cases, each containing preoperative CT volumes (512×512×401 voxels, spacing: 0.3×0.3×0.625 mm) and paired AP/LA X-rays (976×976 resolution, 1,124 mm focal length). Thirty vertebral-level samples were extracted and utilized for the second and third stage of pre-training with five-fold cross-validation. GT poses (CT-to-C-arm coordinate transformation) were recorded during C-arm circular acquisitions to ensure geometric consistency.

3.2 Experimental Setup and Evaluation Metrics

To implement the data-aware strategy, we proposed a three stage pre-training strategy for the RadGS-Reg, training on three distinct datasets: publicly available CT volumes with DRRs, our proprietary real dataset, and the target CT volume with DRRs. First, the model was pre-trained on the VERSE '20 dataset

using simulated data, followed by pre-training on the in-house dataset with five-fold cross-validation. For each fold, the validation set of preoperative CT volumes generated 1,800 biplanar DRR pairs and 600 CT pairs for the third stage. Results were analyzed using five-fold cross-validation on real X-ray data from the in-house dataset.

The evaluation results are presented in two main aspects: first, the SSIM and Peak Signal-to-Noise Ratio (PSNR) were used to gauge the quality of the reconstructed vertebra volumes. Second, the mean Target Registration Error (mTRE), Capture Range (CR), time efficiency, and Success Rate (SR) were used to measure the performance of 2D/3D registration [27]. The success of SR was defined as the mTRE of less than 2 mm, and the acceptable CR interval was set to 5 mm.

Table 1. Our method was compared with the existing approaches in both reconstruction and registration. Statistical significance was evaluated against each method, with an asterisk (*) indicating $p < 0.05$ (mean $\pm$ std, best results in bold).

Method	*	Reconstruction		Registration			
		SSIM(%) $\uparrow$	PSNR(dB) $\uparrow$	mTRE(mm) $\downarrow$	CR(mm) $\uparrow$	SR(%) $\uparrow$	RT(s) $\downarrow$
DiffVox	✓	15.48 $\pm$ 11.42	10.22 $\pm$ 0.76	-	-	-	-
SAX-NeRF	✓	57.23 $\pm$ 8.96	16.17 $\pm$ 0.55	-	-	-	-
3DGR	✓	39.93 $\pm$ 12.35	15.06 $\pm$ 2.35	-	-	-	-
R ² -GS	✓	52.36 $\pm$ 2.21	15.79 $\pm$ 0.54	-	-	-	-
DiffPose	✓	-	-	10.49 $\pm$ 9.90	10-15	20.00	139.24
DDGS-CT	✓	-	-	13.90 $\pm$ 11.84	15-20	16.67	72.99
TS-SAR	✓	-	-	4.89 $\pm$ 1.20	0-5	6.67	1.72
Ours	-	94.51$\pm$0.39	28.80$\pm$0.33	1.14$\pm$1.01	20-25	93.33	0.82

All experiments were conducted on an AMD Core EYPC 7R32 48-core processor with 128 GB RAM and an NVIDIA GeForce RTX 4090 GPU (24 GB memory). The Adam optimizer was used for training, which lasted 300 epochs with an initial learning rate of 0.1, decaying by 0.5 every 50 epochs. The batch size was set to 8 to fit within GPU memory. For hyperparameters, we used $\lambda_1 = 0.2$ and $\lambda_2 = 0.05$ in Eq. (2) as per [26], and $\lambda = 0.02$ in Eq. (3) as suggested by [9].

3.3 Results

Comparison with Existing Methods We conducted experiments on 3D reconstruction and registration, as detailed in Table 1, with results marked by the dotted line. For reconstruction, we compared classical (DiffVox), NeRF-based (SAX-NeRF [3]), and 3DGS-based methods (3DGR [6], R²-GS [26]), using $\mathcal{L}_{rec}$ from Eq. (2). RadGS-Reg achieves efficient, optimization-free reconstruction by encoding implicit vertebral CT shape priors into an encoder, enabling inference based only on biplanar X-rays. For registration, we included learning-based

(DiffPose [9]), 3DGS-based (DDGS-CT [8]), and Lift3D-based (TS-SAR [28]) methods. RadGS-Reg outperformed the others in both reconstruction and registration. As shown in Fig. 2, the shape and voxel intensity distribution of the vertebrae reconstructed by our method are more similar to the GT volume. In Fig. 3, our approach yielded the most precise registration, with the preoperative vertebral CT registering accurately to the intraoperative position. The intraoperative volume reconstructed from two X-rays elevated the 2D/3D registration to 3D space, and the registration facilitated by RecM effectively bridged the domain gap inherent in the "render and compare" approaches [9,8].

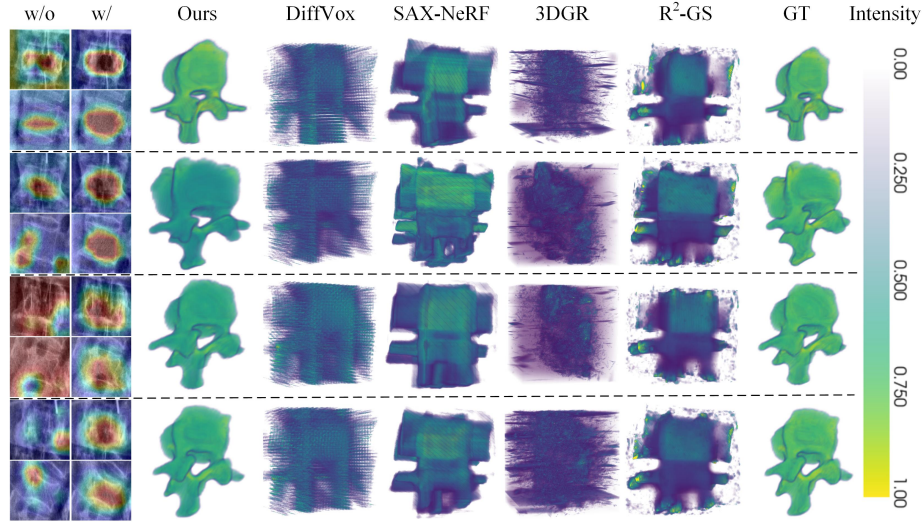


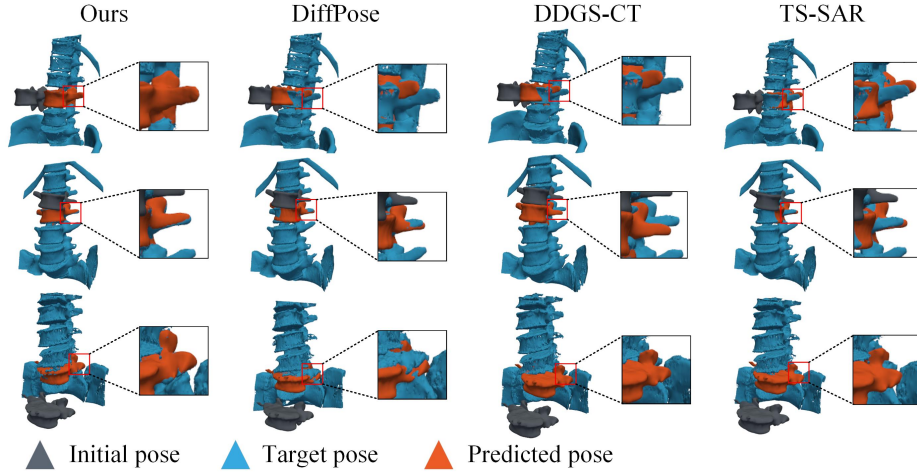
Fig. 2. Qualitative comparison with the different reconstruction methods. For each case separated by a dashed line, column 1 displays attention maps from RecM without CAL, while column 2 shows those with CAL, with the color bar indicating the voxel intensity distribution.

Ablation Study Ablation studies were conducted to assess the efficacy of the modules within RadGS-Reg, encompassing the CAL module, synergistic training, three stage pre-training, and model-agnostic characteristics, as delineated in Table 2. Results from experiments (1) and (2) validated that the CAL module enhanced the accuracy of vertebral reconstruction. Analysis of X-ray attention maps in Fig. 2 indicated that the CAL module directed the network’s focus towards the vertebral regions within the X-rays. A comparative analysis of experiment results (3) and (4) revealed that synergistic training substantially improved both registration and reconstruction performance. In this context, RadGS-Reg epitomized the synergistic training between RecM and RegM. Find-

Table 2. Ablation study of our method. P_1 , P_2 , and P_{Full} represent the first stage, first and second stage, and entire three stage pre-training, respectively.

No. Method	Reconstruction		Registration		
	SSIM(%) $\uparrow$	PSNR(dB) $\uparrow$	mTRE(mm) $\downarrow$	CR(mm) $\uparrow$	SR(%) $\uparrow$
(1) RecM+ P_1	20.56 $\pm$ 1.31	14.87 $\pm$ 0.31	-	-	-
(2) RecM+CAL+ P_1	22.37 $\pm$ 1.64	15.45 $\pm$ 0.47	-	-	-
(3) RegM+ P_1	-	-	4.70 $\pm$ 8.90	0-5	6.67
(4) RadGS-Reg+ P_1	32.78 $\pm$ 0.45	18.99 $\pm$ 0.17	4.03 $\pm$ 3.42	5-10	30.00
(5) RadGS-Reg+ P_2	86.42 $\pm$ 0.97	27.29 $\pm$ 0.30	1.36 $\pm$ 1.19	15-20	80.00
(6) RadGS-Reg+ P_{Full}	94.51$\pm$0.39	28.80$\pm$0.33	1.14$\pm$1.01	20-25	93.33
(7) Dense-backbone	93.77 $\pm$ 0.32	28.27 $\pm$ 0.26	1.28 $\pm$ 0.97	20-25	90.00
(8) ViT-backbone	94.08 $\pm$ 1.09	28.20 $\pm$ 0.79	1.27 $\pm$ 0.95	20-25	93.33

ings from experiments (4)-(6) demonstrated that exclusive training of RadGS-Reg on simulated data yielded suboptimal performance on the in-house dataset. Incrementally incorporating real X-rays and patient-specific DRRs, which provided shape priors, enabled RadGS-Reg to realize optimal performance across both tasks. Furthermore, a series of experiments employing various backbones (e.g., ResNet-50 [11], DenseNet-121 [12], and ViT-12 [5]) in the reconstruction and registration modules of RadGS-Reg were conducted. The backbone network was uniformly configured as ResNet-50 for experiments (1)-(6). Findings from experiments (6)-(8) demonstrated that the proposed synergistic training possessed model-agnostic properties, consistently achieving favorable outcomes across both tasks, irrespective of the backbone employed.

**Fig. 3.** Qualitative comparison with different registration methods. Each row represents an individual case.

4 Conclusion

We present RadGS-Reg, a unified framework for accurate and efficient vertebral CT/X-ray registration by synergizing 3D RadGS reconstruction and 3D/3D registration. Key innovations include the integration of CAL to mitigate vertebral overlap interference, a module-aware synergistic training strategy, and a three stage data-aware pre-training strategy. RadGS-Reg significantly outperforms existing methods in reconstruction quality (94.51% SSIM) and registration accuracy (1.14 mm mTRE). Its model-agnostic design ensures flexibility, while its 0.82s runtime meets real-time clinical demands. Future work will extend the applicability of RadGS-Reg by: 1) incorporating deformation models for complex non-rigid regions involving soft tissues and vasculature, and 2) implementing post-reconstruction Gaussian segmentation to streamline the current preprocessing procedures.

Acknowledgments. This work was supported partly by Jiangsu Province Key Research Program-Social Development-Clinical Frontier Technologies (BE2022718).

Disclosure of Interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

1. Verse: A vertebrae labelling and segmentation benchmark for multi-detector ct images. *Medical Image Analysis* **73**, 102166 (2021)
2. Bao, Y., Ding, T., Huo, J., Liu, Y., Li, Y., Li, W., Gao, Y., Luo, J.: 3d gaussian splatting: Survey, technologies, challenges, and opportunities. *IEEE Transactions on Circuits and Systems for Video Technology* (2025)
3. Cai, Y., Wang, J., Yuille, A., Zhou, Z., Wang, A.: Structure-aware sparse-view x-ray 3d reconstruction. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 11174–11183 (2024)
4. De Roovere, P., Daems, R., Croenen, J., Bourgana, T., de Hoog, J., Wyffels, F.: Cendernet: Center and curvature representations for render-and-compare 6d pose estimation. In: *European Conference on Computer Vision*. pp. 97–111. Springer (2022)
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020)
6. Fu, X., Li, Y., Tang, F., Li, J., Zhao, M., Teng, G.J., Zhou, S.K.: 3dgr-car: Coronary artery reconstruction from ultra-sparse 2d x-ray views with a 3d gaussians representation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 14–24. Springer (2024)
7. Gao, C., Feng, A., Liu, X., Taylor, R.H., Armand, M., Unberath, M.: A fully differentiable framework for 2d/3d registration and the projective spatial transformers. *IEEE transactions on medical imaging* (2023)

8. Gao, Z., Planche, B., Zheng, M., Chen, X., Chen, T., Wu, Z.: Ddgs-ct: Direction-disentangled gaussian splatting for realistic volume rendering. *arXiv preprint arXiv:2406.02518* (2024)
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. Grupp, R.B., Unberath, M., Gao, C., Hegeman, R.A., Murphy, R.J., Alexander, C.P., Otake, Y., McArthur, B.A., Armand, M., Taylor, R.H.: Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2d/3d registration. *International journal of computer assisted radiology and surgery* **15**, 759–769 (2020)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
12. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4700–4708 (2017)
13. Jaganathan, S., Kukla, M., Wang, J., Shetty, K., Maier, A.: Self-supervised 2d/3d registration for x-ray to ct image fusion. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 2788–2798 (2023)
14. Jiang, J., Shuai, L., Zeng, Q.: Ablspinelevelcheck: Localization of vertebral levels on fluoroscopy via semi-supervised abductive learning. In: *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. pp. 3347–3350. IEEE (2024)
15. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139–1 (2023)
16. Labbé, Y., Manuelli, L., Mousavian, A., Tyree, S., Birchfield, S., Tremblay, J., Carpentier, J., Aubry, M., Fox, D., Sivic, J.: Megapose: 6d pose estimation of novel objects via render & compare. *arXiv preprint arXiv:2212.06870* (2022)
17. Naik, R.R., Anitha, H., Bhat, S.N., Ampar, N., Kundangar, R.: Realistic c-arm to pct registration for vertebral localization in spine surgery: A hybrid 3d-2d registration framework for intraoperative vertebral pose estimation. *Medical & Biological Engineering & Computing* **60**(8), 2271–2289 (2022)
18. Otake, Y., Armand, M., Armiger, R.S., Kutzer, M.D., Basafa, E., Kazanzides, P., Taylor, R.H.: Intraoperative image-based multiview 2d/3d registration for image-guided orthopaedic surgery: incorporation of fiducial-based c-arm tracking and gpu-acceleration. *IEEE transactions on medical imaging* **31**(4), 948–962 (2011)
19. Payer, C., Stern, D., Bischof, H., Urschler, M.: Coarse to fine vertebrae localization and segmentation with spatialconfiguration-net and u-net. In: *VISIGRAPP (5: VISAPP)*. pp. 124–133 (2020)
20. Penney, G.P., Weese, J., Little, J.A., Desmedt, P., Hill, D.L., et al.: A comparison of similarity measures for use in 2-d-3-d medical image registration. *IEEE transactions on medical imaging* **17**(4), 586–595 (1998)
21. Ponimatin, G., Labbé, Y., Russell, B., Aubry, M., Sivic, J.: Focal length and object pose estimation via render and compare. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 3825–3834 (2022)
22. Rao, Y., Chen, G., Lu, J., Zhou, J.: Counterfactual attention learning for fine-grained visual categorization and re-identification. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 1025–1034 (2021)
23. Salehi, S.S.M., Khan, S., Erdogmus, D., Gholipour, A.: Real-time deep pose estimation with geodesic loss for image-to-template rigid registration. *IEEE transactions on medical imaging* **38**(2), 470–481 (2018)

24. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences* **1**(3), 248–272 (2008)
25. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
26. Zha, R., Lin, T.J., Cai, Y., Cao, J., Zhang, Y., Li, H.: R^2 -gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction. *arXiv preprint arXiv:2405.20693* (2024)
27. Zhang, B., Faghihroohi, S., Azampour, M.F., Liu, S., Ghotbi, R., Schunkert, H., Navab, N.: A patient-specific self-supervised model for automatic x-ray/ct registration. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 515–524. Springer (2023)
28. Zhao, H., Zhu, W., Deng, X., Zhang, G., Zou, W., et al.: Automatic 2d/3d spine registration based on two-step transformer with semantic attention and adaptive multi-dimensional loss function. *Biomedical Signal Processing and Control* **95**, 106384 (2024)