

Robotic Ultrasound Makes CBCT Alive

Feng Li^{1,2*}, Ziyuan Li^{1*}, Zhongliang Jiang³, Nassir Navab^{1,2}, and Yuan Bi^{1,2}

¹ Chair for Computer-Aided Medical Procedures and Augmented Reality,
Technical University of Munich, Munich, Germany

² Munich Center for Machine Learning, Munich, Germany

³ The University of Hong Kong, Hong Kong, China
`feng.li@tum.de`

Abstract. Intraoperative Cone Beam Computed Tomography (CBCT) provides a reliable 3D anatomical context for interventional planning. However, its static nature fails to provide continuous monitoring of soft-tissue deformations induced by respiration, probe pressure, and surgical manipulation, leading to navigation discrepancies. We propose a deformation-aware CBCT updating framework that leverages robotic ultrasound as a dynamic proxy to infer tissue motion and update static CBCT slices in real time. Starting from calibration-initialized alignment with linear correlation of linear combination (LC2)-based rigid refinement, our method establishes multimodal correspondence. To capture intraoperative dynamics, we introduce the ultrasound correlation UNet (USCorUNet), a lightweight network trained with optical flow-guided supervision to learn deformation-aware correlation representations, enabling accurate, real-time dense deformation field estimation from ultrasound streams. The inferred deformation is spatially regularized and transferred to the CBCT reference to produce deformation-consistent visualizations without repeated radiation exposure. We validate the approach through deformation estimation and ultrasound-guided CBCT updating experiments. Results demonstrate real-time end-to-end CBCT slice updating and physically plausible deformation estimation, enabling dynamic refinement of static CBCT guidance during robotic ultrasound-assisted interventions. The source code is publicly available at <https://github.com/ziyuan-li/robotic-us-cbct-updating>.

Keywords: Robotic ultrasound · CBCT · Deformation.

1 Introduction

In recent years, robotic ultrasound has emerged as a promising paradigm for autonomous and reproducible intraoperative imaging, offering real-time visualization with high soft-tissue contrast and vascular sensitivity [2,3]. Robotic control enables precise probe positioning, regulation of contact force, and automated scanning [4,10,21], supporting applications such as 3D compounding, motion-aware imaging, deformation recovery, and elastography [11]. However, its limited acoustic window restricts imaging to local soft tissues, which remain prone

* These authors contributed equally to this work.

to artifacts and occlusions. Ultrasound captures only deformation-dependent observations without a consistent global reference, providing dynamic tissue information but not comprehensive anatomical context.

Cone beam computed tomography (CBCT) provides high-resolution volumetric imaging and has become an important intraoperative modality due to its compact design, lower radiation dose, and flexible integration into the operating room. Commercial systems such as C-arm-based platforms [23] and Loop-X devices [15] offer on-demand 3D imaging with a large field of view, supporting surgical navigation and serving as detailed volumetric priors. However, CBCT captures only a static snapshot [22], while anatomy evolves during procedures due to respiration, patient movement, or probe interaction. Repeated CBCT is limited by radiation and workflow constraints, motivating integration with real-time ultrasound. Prior work using electromagnetic [19] and optical tracking [17] has enabled rigid CBCT-ultrasound alignment, but deformation-aware multimodal updating remains a key challenge for accurate intraoperative guidance.

In CBCT-ultrasound integration, deformation poses a fundamental challenge [12]. Ultrasound sequences encode rich temporal information, with adjacent frames reflecting tissue motion from respiration and probe interaction. Classical methods such as speckle tracking [6] and optical flow [18] have been explored for motion estimation, but modern speckle suppression reduces tracking reliability, and large probe-induced deformations remain difficult to capture. Deep learning-based optical flow models, like recurrent all-pairs field transforms (RAFT) [24], improve robustness to large displacements, yet ultrasound artifacts and depth-dependent distortions violate standard assumptions. Purely data-driven approaches also lack physical constraints, potentially producing biomechanically implausible deformation fields. These limitations motivate dedicated deformation modeling for CBCT-ultrasound integration.

To address these challenges, we propose a robotic-ultrasound-driven framework for deformation-aware CBCT updating in image-guided intervention. Our main contributions are: (1) a workflow-compatible robotic US-CBCT pipeline that integrates calibration, LC2-based refinement, ultrasound deformation estimation, and ultrasound-informed CBCT slice updating; (2) USCorUNet, a lightweight bidirectional correlation-enhanced network for real-time ultrasound-to-ultrasound deformation estimation with optical-flow-guided supervision; (3) real-time CBCT slice updating for probe- and externally induced deformations; and (4) multi-dataset in vivo and phantom validation, showing a favorable accuracy-efficiency trade-off against RAFT-based and classical baselines.

2 Methods

Our system has four modules (Fig. 1). Calibration-based rigid initialization establishes the CBCT-ultrasound spatial relationship (Fig. 1(a)), refined by LC2 to correct residual alignment errors (Fig. 1(b)). Non-rigid deformation between consecutive ultrasound frames is estimated with USCorUNet using pedagogical supervision (Fig. 1(c)). The resulting deformation is applied to update the corre-

sponding CBCT slice, producing a dynamically deformed CBCT for visualization and guidance (Fig. 1(d)).

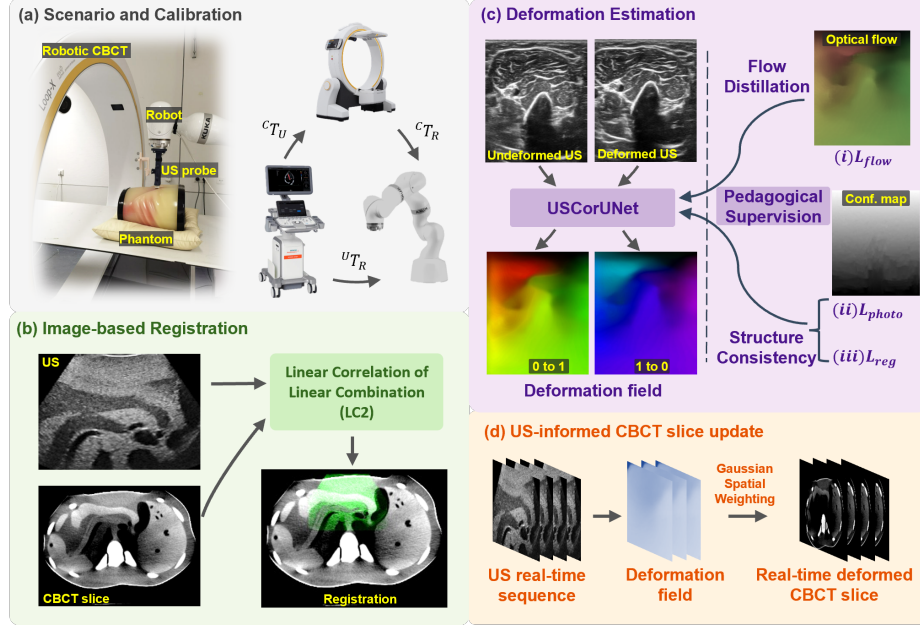


Fig. 1: System overview. (a) Rigid calibration between robotic ultrasound and CBCT. (b) Image-based registration refinement. (c) Deformation estimation with USCorUNet. (d) Deformation transfer for CBCT slice updating. Ultrasound examples from the in vivo arm dataset (c) and the CT-mapped phantom dataset (b,d) illustrate cross-domain applicability. Conf. map is confidence map.

2.1 Calibration and Registration

For spatial calibration, we adopted the hand-eye calibration framework proposed in [17]. Ultrasound (US) images can be projected into the CBCT volume via the spatial relationships among the CBCT, robot, and probe. The rigid transformation between a US image and CBCT slice, ${}^C\mathbf{T}_U$, is computed as ${}^C\mathbf{T}_U = {}^C\mathbf{T}_R({}^U\mathbf{T}_R)^{-1}$, where ${}^C\mathbf{T}_R$ and ${}^U\mathbf{T}_R$ map the CBCT and US image to the robot base frame.

Although calibration achieves a mean alignment error of approximately 1–2 mm, residual misalignment persists without image-based refinement, which can be clinically significant in precision-sensitive procedures such as needle insertion. To improve accuracy, we incorporate multimodal rigid registration using the LC2 similarity metric [26], which models the relationship between CBCT intensities

and ultrasound appearance via a local linear approximation: $f(x_i) = \alpha \mathbf{p}_i + \beta \mathbf{g}_i + \gamma$, where p_i and g_i denote CBCT intensity and gradient magnitude, and α, β, γ estimate the local cross-modal relationship. Initialized by calibration, LC2 searches only in-plane translations within ± 10 CBCT pixels around the projected US position, while the slice and orientation are fixed by the calibrated geometry. This constrained search reduces computational cost and runtime.

2.2 Deformation Field Acquisition

Building on the rigid alignment, USCorUNet estimates dense bidirectional deformation fields between ultrasound frames $I_0, I_1 \in \mathbb{R}^{H \times W}$, yielding $F_{01}, F_{10} \in \mathbb{R}^{H \times W \times 2}$. We distill pseudo-labels from RAFT [24]. For each pair, we generate a direct RAFT candidate on (I_0, I_1) and a bisect candidate via an intermediate frame, composed as $(F \oplus G)(\mathbf{x}) = F(\mathbf{x}) + G(\mathbf{x} + F(\mathbf{x}))$, and select the candidate with lower post-warp misalignment under differentiable warping $\mathcal{W}(I, F)(\mathbf{x}) = I(\mathbf{x} + F(\mathbf{x}))$.

Fig. 2 shows the architecture of USCorUNet. A preprocessing module forms a five-channel input $[I_0, I_1, (I_1 - I_0), |\nabla I_0|, |\nabla I_1|]$. The network combines a ResUNet-style context encoder-decoder with a shared-weight correlation encoder g_ϕ . From features $f_i = g_\phi([I_i, |\nabla I_i|])$, we build local correlation volumes, e.g., $CV_{01}(\mathbf{x}, \Delta) = \frac{1}{\sqrt{C}} \langle f_0(\mathbf{x}), f_1(\mathbf{x} + \Delta) \rangle$, and define CV_{10} analogously, where C is the feature dimension and Δ indexes a local displacement neighborhood. The correlation volumes are fused with same-scale context features and decoded into dense fields at 1/8 resolution, balancing efficiency and reliable matching.

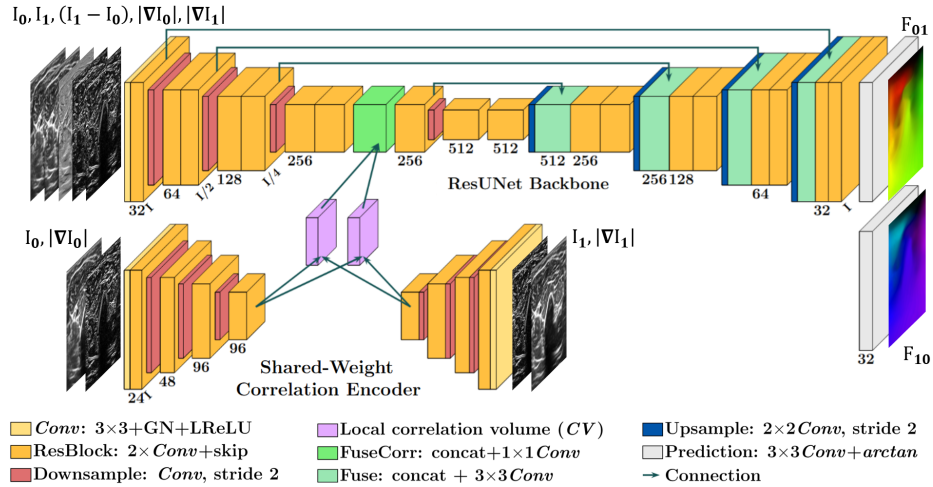


Fig. 2: Architecture of USCorUNet.

Since appearance-driven pseudo-labels do not explicitly enforce physical plausibility [8], we train USCorUNet with a bidirectional objective that combines optical flow distillation $\mathcal{L}_{\text{flow}}$, confidence-weighted photometric consistency $\mathcal{L}_{\text{photo}}$, and regularization $\mathcal{L}_{\text{reg}}$. Specifically, $\mathcal{L}_{\text{flow}}$ is an ℓ_1 loss to the optical flows, $\mathcal{L}_{\text{photo}}$ is a confidence-weighted Charbonnier penalty [5] on post-warp intensity residuals using random-walk confidence maps [14], and $\mathcal{L}_{\text{reg}}$ combines edge-aware smoothness [9] and a Jacobian-based folding penalty [20]. The final objective is $\mathcal{L} = \lambda_{\text{flow}}\mathcal{L}_{\text{flow}} + \lambda_{\text{photo}}\mathcal{L}_{\text{photo}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}$, with $\lambda_{\text{flow}} = 1$, $\lambda_{\text{photo}} = 0.2$, and $\lambda_{\text{reg}} = 0.05$, chosen to prioritize flow distillation while using photometric consistency as a complementary cue and regularization as a mild physical prior.

2.3 Ultrasound Guided CBCT Updating

The estimated deformation field updates the CBCT slice in real time, accounting for probe and external motion. Specifically for probe-induced motion, we correct non-uniform convex compression using a Gaussian profile $P(x) = d_{\text{robot}} \cdot \exp(-(x - c_x)^2 / 2\sigma_{\text{probe}}^2)$, where d_{robot} denotes the probe displacement magnitude, c_x the lateral midpoint, and σ_{probe} the curvature. This profile corrects the vertical deformation component: $\mathbf{D}_{\text{geo}}^y = \mathbf{D}_{\text{raw}}^y - P(x)$.

To align the local ultrasound field with the larger CBCT ROI, we use Euclidean Distance Transform (EDT)-based spatial weighting. The corrected field is padded to $\mathbf{D}_{\text{pad}}$ and scaled by $W(x, y) = \exp(-\mathcal{D}(x, y) / \sigma_{\text{smooth}})$, where $\mathcal{D} = \text{EDT}(1 - M)$ denotes the distance to the ultrasound boundary. The final field $\mathbf{D}_{\text{final}} = \mathbf{D}_{\text{pad}} \odot W$ enforces smooth decay of deformation from the probe contact region across the CBCT slice.

3 Experiments and Results

3.1 Experimental Protocol

The experimental setup integrates a KUKA LBR iiwa 14 R820 robot and Siemens ACUSON Juniper ultrasound (5C1 probe) via a 3D-printed holder, with images acquired through an Epiphan frame grabber. A Loop-X Imaging Ring provided CBCT data.

Experiments were conducted on four datasets: (A) in vivo forearm/upper-arm ultrasound from DefCor-Net [13], collected under IRB approval No. 2022-87-S-KK; (B) a pork-tissue gel phantom; (C) a chicken/pork gel phantom; and (D) an abdominal phantom (Kyoto Kagaku US-22). Datasets B–D are phantom datasets used for controlled probe-induced deformation experiments, with Dataset B additionally including finger-press compression to simulate externally induced deformation.

USCorUNet was trained on Dataset A (8:1:1 split) as a base model using (I_0, I_1) pairs and AdamW for 50 epochs (batch size 4, learning rate 2×10^{-4} , weight decay 10^{-4} , mixed precision). From the base model, we fine-tuned two variants: one for probe-induced motion on Datasets B–D and one for externally

induced motion on Dataset B. Each variant was fine-tuned for 20 epochs with a learning rate of 10^{-4} , with the split and other settings unchanged. Across datasets, validation and test splits were held out and excluded from fine-tuning.

We evaluate bidirectional deformation estimates using (i) post-warp alignment (MAE, NCC), (ii) forward-backward (FB) consistency via the mean ℓ_2 norm of $r_{01} = F_{01} \oplus F_{10}$ and $r_{10} = F_{10} \oplus F_{01}$ (mean FB residual) [25], and (iii) physical plausibility via the folding ratio, i.e., the fraction of pixels with $\det(I + \nabla F) < 0$. On Dataset A, we additionally report Dice on SAM-segmented [16] bone masks after nearest-neighbor warping. For deformation-warped CT volumes, we report bone-mask Dice and SSIM.

As the primary baseline, we evaluate the selected optical flow under the same metrics as a reference for USCorUNet. We further conduct Dataset A ablations by (i) removing the correlation branch (w/o Corr.), (ii) disabling confidence-map weighting (w/o Conf.), and (iii) simplifying the training objective to $\mathcal{L}_{\text{flow}} + \mathcal{L}_{\text{photo}}$ or $\mathcal{L}_{\text{flow}} + \mathcal{L}_{\text{reg}}$.

3.2 Results of Deformation Field Acquisition

Base model. Table 1 shows that USCorUNet matches the optical flow in photometric alignment, slightly improves bone-mask Dice, and markedly improves deformation quality, reducing the mean FB residual by around 53% and lowering the folding ratio. Ablations further show that removing the correlation branch causes the largest drop in photometric alignment with only a mild Dice decrease, suggesting its primary benefit is improved correspondence in non-bone texture regions. Disabling confidence weighting or removing $\mathcal{L}_{\text{reg}}$ mainly worsens FB consistency and increases foldings, whereas removing $\mathcal{L}_{\text{photo}}$ yields the largest Dice drop, indicating that photometric supervision is important for anatomically faithful deformations.

Table 1: Test-set performance of the base model on Dataset A, averaged over both directions ($I_0 \rightarrow I_1$, $I_1 \rightarrow I_0$) and reported as mean $\pm$ standard deviation.

	Method	MAE $\downarrow$	NCC $\uparrow$	FB $\downarrow$	Fold(%) $\downarrow$	Dice(%) $\uparrow$
Baseline	RAFT	0.05 $\pm$ 0.02	0.85 $\pm$ 0.10	1.81 $\pm$ 1.22	0.24 $\pm$ 0.14	90.31 $\pm$ 4.14
Ours	USCorUNet	0.05 $\pm$ 0.02	0.85 $\pm$ 0.09	0.85 $\pm$ 0.57	0.13 $\pm$ 0.10	90.62 $\pm$ 3.76
Ablation Study	w/o Corr.	0.08 $\pm$ 0.03	0.61 $\pm$ 0.18	1.32 $\pm$ 3.14	0.56 $\pm$ 0.51	89.10 $\pm$ 3.33
	w/o Conf.	0.06 $\pm$ 0.06	0.74 $\pm$ 0.16	1.55 $\pm$ 1.20	0.17 $\pm$ 0.11	87.03 $\pm$ 6.17
	w/o $\mathcal{L}_{\text{reg}}$	0.06 $\pm$ 0.02	0.79 $\pm$ 0.12	1.21 $\pm$ 0.92	0.41 $\pm$ 0.22	89.30 $\pm$ 4.23
	w/o $\mathcal{L}_{\text{photo}}$	0.07 $\pm$ 0.03	0.72 $\pm$ 0.17	1.38 $\pm$ 1.03	0.19 $\pm$ 0.19	85.45 $\pm$ 6.28

Table 2 compares USCorUNet with DefCor-Net [13] on the same in-vivo ultrasound dataset (Dataset A) under their force-stratified evaluation protocol. In this protocol, the contact force between the robot and the probe is synchronized with B-mode frames and grouped into 1–6 N force levels. Improvements are more pronounced at higher forces, which correspond to larger and more challenging compressions.

Table 2: Bone-mask Dice under the force-stratified protocol of DefCor-Net [13] on Dataset A. Dice is reported for the $I_1 \rightarrow I_0$ direction.

Method	1 N	2 N	3 N	4 N	5 N	6 N
DefCor-Net	95.9 ± 3.3	92.4 ± 4.8	91.1 ± 7.5	87.8 ± 7.1	87.8 ± 12.2	82.6 ± 12.1
USCorUNet	95.4 ± 2.6	94.1 ± 2.4	94.0 ± 3.6	93.2 ± 1.8	89.8 ± 1.5	87.5 ± 1.2

Fine-tuned models. Table 3 shows that regime-specific fine-tuning improves performance, especially FB consistency and physical plausibility. For probe-induced motion, fine-tuning reduces mean FB residuals and folding ratios while preserving alignment. For externally induced motion, it mitigates the domain shift of the base model and restores alignment. Overall, the base checkpoint provides a robust, transferable initialization across interaction patterns. Visual results are shown in Fig. 3.

Table 3: Test-set performance of fine-tuned models on Datasets B–D (probe-induced) and Dataset B (externally induced), averaged over both directions.

Motion	Method	MAE $\downarrow$	NCC $\uparrow$	FB $\downarrow$	Fold(%) $\downarrow$
Probe-induced	RAFT	0.03 ± 0.01	0.94 ± 0.03	1.03 ± 2.27	0.23 ± 0.14
	Base model	0.04 ± 0.01	0.84 ± 0.08	0.84 ± 0.65	0.15 ± 0.11
	Probe-adapted	0.03 ± 0.01	0.93 ± 0.03	0.33 ± 0.28	0.09 ± 0.10
Externally induced	RAFT	0.03 ± 0.01	0.93 ± 0.05	0.91 ± 1.35	0.14 ± 0.13
	Base model	0.05 ± 0.02	0.71 ± 0.17	1.28 ± 1.95	0.45 ± 0.55
	External-adapted	0.03 ± 0.01	0.92 ± 0.05	0.23 ± 0.24	0.07 ± 0.07

3.3 Results of Ultrasound Guided CBCT Updating

Table 4 compares our method with RAFT-based and classical LC2-FFD [7] baselines. Our approach achieves the best quality–efficiency trade-off, slightly outperforming RAFT while reducing runtime by $5\times$. In addition, RAFT can exhibit tearing artifacts. Although LC2-FFD shows comparable MAE and SSIM, it introduces geometric distortions and is $512\times$ slower. As absolute metrics are affected by robotic artifacts, this serves as a controlled relative comparison. Visual results are shown in Fig. 4.

Table 4: CBCT updating results on Dataset D using ultrasound-based deformation. Methods share identical conditions; runtime is end-to-end.

Model	MAE $\downarrow$	SSIM $\uparrow$	Dice (%) $\uparrow$	Timing (ms) $\downarrow$
USCorUNet	0.33 ± 0.04	0.22 ± 0.02	82.22 ± 10.54	11.25 ± 0.44
RAFT	0.34 ± 0.04	0.21 ± 0.03	79.86 ± 12.31	56.24 ± 1.78
LC2-FFD	0.33 ± 0.05	0.22 ± 0.03	58.91 ± 8.77	5764.26 ± 510.63

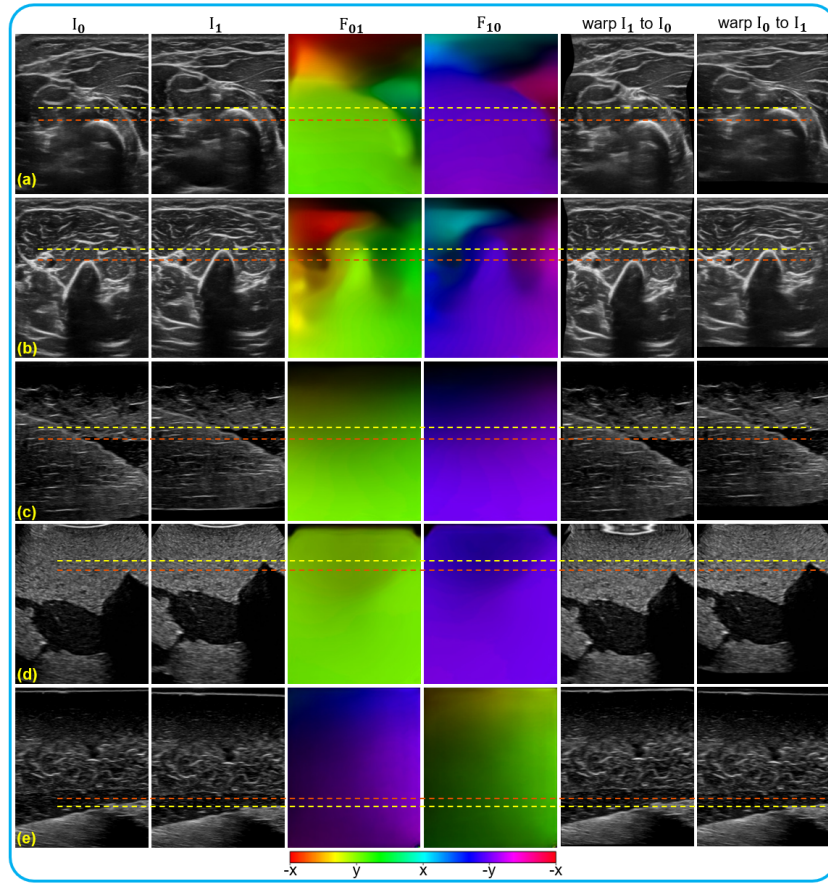


Fig. 3: Bidirectional deformation estimation results. (a,b) In vivo arm examples (Dataset A); (c,d) probe-induced motion (Datasets B and D); (e) externally induced motion (Dataset B). Orange/yellow dashed lines indicate visual alignment guides for I_0 , I_1 . The color bar indicates flow direction (x/y).

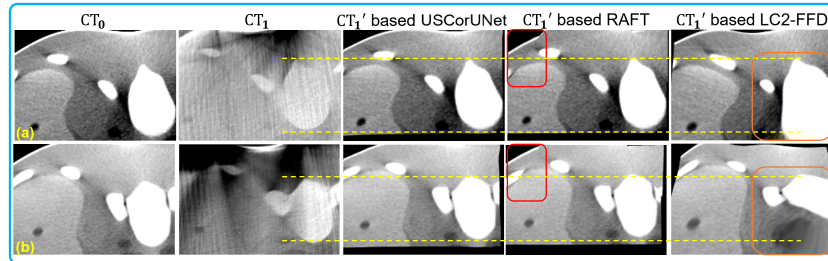


Fig. 4: CT_1' update results on two representative abdominal phantom cases (Dataset D) using USCorUNet, RAFT, and LC2-FFD (CT_0 : source; CT_1 : target). Red/orange boxes indicate structural artifacts/severe deformation regions.

4 Discussion and Conclusion

In this paper, we introduced a deformation-aware CBCT updating framework that integrates rigid calibration, registration, and deformation field estimation for real-time end-to-end CBCT slice updating and enhanced intraoperative visualization. The proposed USCorUNet efficiently estimates deformation fields from adjacent ultrasound frames while preserving structural consistency.

Compared to RAFT and LC2-FFD baselines, our method achieves a superior trade-off between registration accuracy and computational efficiency. Future work will extend the current slice-level CBCT updating framework toward multi-plane or volumetric ultrasound guidance, and may further incorporate semantic segmentation [1] into the registration pipeline. While the current approach relies on intensity and structural features, integrating semantic information could further refine deformation details in complex anatomical regions, ultimately paving the way for more reliable intraoperative guidance.

Acknowledgments. This work was supported in part by the Multi-Scale Medical Robotics Center, AIR@InnoHK, Hong Kong; in part by the NVIDIA Academic Grant Program.

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

References

1. Bi, Y., Jiang, Z., Clarenbach, R., Ghotbi, R., Karlas, A., Navab, N.: Mi-segnet: Mutual information-based us segmentation for unseen domain generalization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 130–140. Springer (2023)
2. Bi, Y., Jiang, Z., Duelmer, F., Huang, D., Navab, N.: Machine learning in robotic ultrasound imaging: Challenges and perspectives. *Annual Review of Control, Robotics, and Autonomous Systems* **7** (2024)
3. Bi, Y., Su, Y., Navab, N., Jiang, Z.: Gaze-guided robotic vascular ultrasound leveraging human intention estimation. *IEEE Robotics and Automation Letters* (2025)
4. Bi, Y., Zhou, Y., Liu, P., Li, F., Jiang, Z., Navab, N.: From scanning guidelines to action: A robotic ultrasound agent with llm-based reasoning. *arXiv preprint arXiv:2603.14393* (2026)
5. Bruhn, A., Weickert, J., Schnörr, C.: Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International journal of computer vision* **61**(3), 211–231 (2005)
6. Curiale, A.H., Vegas-Sánchez-Ferrero, G., Aja-Fernández, S.: Influence of ultrasound speckle tracking strategies for motion and strain estimation. *Medical image analysis* **32**, 184–200 (2016)
7. Fuerst, B., Wein, W., Müller, M., Navab, N.: Automatic ultrasound–mri registration for neurosurgery using the 2d and 3d lc2 metric. *Medical image analysis* **18**(8), 1312–1319 (2014)

8. Greer, H., Kwitt, R., Vialard, F.X., Niethammer, M.: Icon: Learning regular maps through inverse consistency. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 3396–3405 (2021)
9. Janai, J., Guney, F., Ranjan, A., Black, M., Geiger, A.: Unsupervised learning of multi-frame optical flow with occlusions. In: *Proceedings of the European conference on computer vision (ECCV)*. pp. 690–706 (2018)
10. Jiang, Z., Bi, Y., Zhou, M., Hu, Y., Burke, M., Navab, N.: Intelligent robotic sonographer: Mutual information-based disentangled reward learning from few demonstrations. *The International Journal of Robotics Research* **43**(7), 981–1002 (2024)
11. Jiang, Z., Salcudean, S.E., Navab, N.: Robotic ultrasound imaging: State-of-the-art and future perspectives. *Medical image analysis* **89**, 102878 (2023)
12. Jiang, Z., Zhou, Y., Bi, Y., Zhou, M., Wendler, T., Navab, N.: Deformation-aware robotic 3d ultrasound. *IEEE Robotics and Automation Letters* **6**(4), 7675–7682 (2021)
13. Jiang, Z., Zhou, Y., Cao, D., Navab, N.: Defcor-net: Physics-aware ultrasound deformation correction. *Medical Image Analysis* **90**, 102923 (2023)
14. Karamalis, A., Wein, W., Klein, T., Navab, N.: Ultrasound confidence maps using random walks. *Medical image analysis* **16**(6), 1101–1112 (2012)
15. Karius, A., Leifeld, L.M., Strnad, V., Fietkau, R., Bert, C.: First implementation of an innovative infra-red camera system integrated into a mobile cbct scanner for applicator tracking in brachytherapy—initial performance characterization. *Journal of Applied Clinical Medical Physics* p. e14364 (2024)
16. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 4015–4026 (2023)
17. Li, F., Bi, Y., Huang, D., Jiang, Z., Navab, N.: Robotic cbct meets robotic ultrasound. *International Journal of Computer Assisted Radiology and Surgery* pp. 1–9 (2025)
18. Li, F., Bi, Y., Song, T., Jiang, Z., Navab, N.: Ultrasound-guided real-time spinal motion visualization for spinal instability assessment. *arXiv preprint arXiv:2602.12917* (2026)
19. Monfardini, L., Gennaro, N., Orsi, F., Della Vigna, P., Bonomo, G., Varano, G., Solbiati, L., Mauri, G.: Real-time us/cone-beam ct fusion imaging for percutaneous ablation of small renal tumours: a technical note. *European Radiology* **31**(10), 7523–7528 (2021)
20. Shen, Z., Han, X., Xu, Z., Niethammer, M.: Networks for joint affine and non-parametric image registration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 4224–4233 (2019)
21. Song, T., Li, F., Bi, Y., Karlas, A., Yousefi, A., Branzan, D., Jiang, Z., Eck, U., Navab, N.: Intelligent virtual sonographer (ivs): Enhancing physician-robot-patient communication. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 287–297. Springer (2025)
22. Song, T., Li, F., Pabst, F., Gafencu, M.A., Bi, Y., Eck, U., Navab, N.: Comparative study of ultrasound shape completion and cbct-based ar workflows for spinal needle interventions. *International Journal of Computer Assisted Radiology and Surgery* pp. 1–9 (2026)
23. Tanaka, M., Schol, J., Sakai, D., Sako, K., Yamamoto, K., Yanagi, K., Hiyama, A., Katoh, H., Sato, M., Watanabe, M.: Low radiation protocol for intraoperative robotic c-arm can enhance adolescent idiopathic scoliosis deformity correction accuracy and safety. *Global Spine Journal* **14**(5), 1504–1514 (2024)

24. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: European conference on computer vision. pp. 402–419. Springer (2020)
25. Varadhan, R., Karangelis, G., Krishnan, K., Hui, S.: A framework for deformable image registration validation in radiotherapy clinical applications. *Journal of applied clinical medical physics* **14**(1), 192–213 (2013)
26. Wein, W., Ladikos, A., Fuerst, B., Shah, A., Sharma, K., Navab, N.: Global registration of ultrasound to mri using the lc2 metric for enabling neurosurgical guidance. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 34–41. Springer (2013)