CLIN-FuseDiff++: A Clinically-Aligned, ROI-Aware, and Uncertainty-Calibrated Diffusion Framework for Multimodal Medical Image Fusion

Pakawat Phasook

CVPR 2026 Research Proposal (One-Column)

Abstract

Multimodal medical image fusion (e.g., MRI–CT; MRI–SPECT) is essential for surgical planning and image-guided intervention, yet translation is limited by *domain-centric* constraints rather than architecture-specific ones. We identify four pain points: (**P1**) residual *misregistration* that introduces millimetric boundary distortion in clinical workflows; (**P2**) *modality-incompatible physics* (soft-tissue vs. high-density) rendering global IQA metrics (PSNR/SSIM) only weakly aligned with clinical utility; (**P3**) lack of *calibrated spatial uncertainty* for risk-aware decisions; and (**P4**) insufficient *clinician controllability* to prioritize modality dominance by region and disease. In response to problem-aware validation guidance, we propose **CLIN-FuseDiff++**, a diffusion-based fusion framework with *ROI-aware guidance*, *registration-aware robustness*, and *uncertainty-aware sampling*, coupled with a primary ROI metric suite (Dice/NSD/HD95 for lesions; SSIM/FSIM in brain ROI vs. MRI; PSNR/SSIM in bone ROI vs. CT), secondary global IQA, and downstream task probes. The approach targets measurable ROI-wise gains, robustness under plausible misregistration, and practical inference via few-step sampling and distillation.

1. Motivation and Domain-Centric Pain Points

P1. Residual misregistration. Clinical MRI–CT alignment often leaves small but consequential boundary errors that propagate into fusion and downstream tasks (e.g., RT planning).
P2. Modality-incompatible physics. CT encodes electron density; MRI encodes relaxation—global PSNR/SSIM insufficiently reflect clinical goals; problem-aware, boundary-/ROI-aware metrics are recommended.
P3. Safety via uncertainty. Deployment requires spatially calibrated uncertainty (epistemic/aleatoric) to avoid overconfident decisions.
P4. Controllability. Clinicians require region-wise preference (e.g., CT-dominant for skull; MRI-dominant for brain parenchyma) with disease-specific presets.

2. Problem Setting

Given co-registered MRI I_{M} and CT I_{C} and ROI masks $M_{\mathrm{brain}}, M_{\mathrm{bone}}, M_{\mathrm{les}}$ (anatomy via robust auto-segmentation; lesion via expert curation), synthesize a fused image F such that:

- Brain ROI (M_{brain}): F mirrors MRI-like contrast and soft-tissue detail.
- Bone ROI (M_{bone}): F preserves CT-like edges/density.
- Lesion ROI (M_{les}): F retains geometrically faithful shape/borders within clinical tolerance.

Algorithm 1 ROI-aware guided reverse diffusion (inference)

Inputs: $I_{\mathrm{M}}, I_{\mathrm{C}}, M_{\mathrm{brain}}, M_{\mathrm{bone}}, M_{\mathrm{les}}$; hyperparams $\eta, \eta_u, \alpha, \beta, \gamma, \lambda_{1:3}, \tau$.
$x_T \sim \mathcal{N}(0, \mathbf{I})$; $c \leftarrow \{\phi(I_{\mathrm{M}}), \phi(I_{\mathrm{C}}), M.\}$.
for $t = T, \ldots, 1$ **do**
  $F \leftarrow \mathrm{denoise}(x_t)$
  $\mathcal{L}_{\mathrm{brain}} \leftarrow 1 - \mathrm{SSIM}(F, I_{\mathrm{M}} \mid M_{\mathrm{brain}})$
  $\mathcal{L}_{\mathrm{bone}} \leftarrow 1 - \mathrm{SSIM}(F, I_{\mathrm{C}} \mid M_{\mathrm{bone}})$
  $\hat{S} \leftarrow \mathrm{LesionHead}(F)$
  $\mathcal{L}_{\mathrm{les}} \leftarrow \lambda_1 \mathcal{L}_{\mathrm{Dice}}(\hat{S}, S^\star) + \lambda_2 \mathcal{L}_{\mathrm{NSD}_\tau}(\hat{S}, S^\star) + \lambda_3 \mathcal{L}_{\mathrm{HD95}}(\hat{S}, S^\star)$
  $\mathcal{L}_{\mathrm{ROI}} \leftarrow \alpha\, \mathcal{L}_{\mathrm{brain}} + \beta\, \mathcal{L}_{\mathrm{bone}} + \gamma\, \mathcal{L}_{\mathrm{les}}$
  $x_{t-1} \leftarrow g_\theta(x_t, t \mid c) - \eta \nabla_{x_t} \mathcal{L}_{\mathrm{ROI}} - \eta_u \nabla_{x_t} \mathcal{U}$
end for
return $F = x_0$

3. Method: CLIN-FuseDiff++

3.1 Conditioned reverse diffusion with ROI guidance. Let x_t be the noisy fused image at step t; $\epsilon_\theta(x_t, t \mid c)$ predicts noise under conditioning $c = \{\phi(I_{\mathrm{M}}), \phi(I_{\mathrm{C}}), M.\}$ with light encoders $\phi(\cdot)$. We augment the DDPM update by a composite guidance:

$$x_{t-1} \;=\; g_\theta(x_t, t \mid c) \;-\; \eta\, \nabla_{x_t} \mathcal{L}_{\mathrm{ROI}}(x_t) \;-\; \eta_u\, \nabla_{x_t} \mathcal{U}(x_t), \tag{1}$$

where $\eta, \eta_u > 0$ control strengths. The (*frozen*) lesion head $\hat{S}$ supplies lesion predictions $\hat{S}(F)$ for boundary-aware terms.

3.2 Clinical composite ROI loss. We map clinical priorities into optimization via:

$$\begin{aligned}\mathcal{L}_{\mathrm{ROI}} \;&= \alpha \cdot \Big(1 - \mathrm{SSIM}(F, I_{\mathrm{M}} \,|\, M_{\mathrm{brain}})\Big) + \beta \cdot \Big(1 - \mathrm{SSIM}(F, I_{\mathrm{C}} \,|\, M_{\mathrm{bone}})\Big) \\ &\quad + \gamma \cdot \Big[\lambda_1\, \mathcal{L}_{\mathrm{Dice}}(\hat{S}(F), S^\star) + \lambda_2\, \mathcal{L}_{\mathrm{NSD}_\tau}(\hat{S}(F), S^\star) + \lambda_3\, \mathcal{L}_{\mathrm{HD95}}(\hat{S}(F), S^\star)\Big],\end{aligned} \tag{2}$$

where $S^\star$ is the reference lesion mask; τ is a tolerance (mm). The total objective is

$$\mathcal{L} \;=\; \mathbb{E}_{t,\epsilon}\, \|\epsilon - \epsilon_\theta(x_t, t \mid c)\|_2^2 \;+\; \lambda_{\mathrm{roi}}\, \mathbb{E}\, \mathcal{L}_{\mathrm{ROI}}. \tag{3}$$

Clinician presets: We expose (α, β, γ) and $(\lambda_1, \lambda_2, \lambda_3)$ for disease scenarios (e.g., bone tumor $\Rightarrow$ larger β).

3.3 Registration-aware robustness. We (i) compute boundary metrics with tolerance bands and (ii) incorporate small, plausible warp jitter to $I_{\mathrm{M}}, I_{\mathrm{C}}$ during training/sampling, desensitizing the sampler to misregistration.

3.4 Uncertainty-carrying diffusion. We propagate per-voxel variance via sampling ensembles/posterior maps and calibrate spatial uncertainty (e.g., ECE/Brier). Uncertainty modulates guidance strengths in low-agreement regions:

$$\eta^{\mathrm{eff}}(p) \;=\; \eta \cdot \sigma\big(\kappa(1 - \mathrm{conf}(p))\big), \quad \mathrm{conf}(p) = 1 - \mathrm{uncert}(p). \tag{4}$$

3.5 Efficiency. We target few-step sampling and score-distillation refinements for practical latency while preserving fidelity (distilled student updates the denoising field to approximate multi-step teacher).

4. Primary Evaluation: ROI-Aware Metrics

Lesion ROI: Dice, NSD@ τmm, HD95 between $\hat{S}(F)$ and $S^\star$ (*boundary-/tolerance-aware*).
Brain ROI: $\mathrm{SSIM/FSIM}(F, I_{\mathrm{M}} \mid M_{\mathrm{brain}})$ (MRI-like soft-tissue).
Bone ROI: $\mathrm{PSNR/SSIM}(F, I_{\mathrm{C}} \mid M_{\mathrm{bone}})$ (CT-like edges/density).
Secondary (global): PSNR/SSIM/FSIM/FMI for backward comparability.
Downstream: segmentation/localization trained on F; sensitivity to misregistration.

5. Datasets, ROI Masks, and Ethics

We begin with multimodal brain pairs (MRI–CT / MRI–SPECT). Anatomy masks ($M_{\mathrm{brain}}, M_{\mathrm{bone}}$) derive from robust auto-segmentation; lesion masks (M_{les}) are curated by experts. All data are de-identified with IRB compliance. We will release code, configs, and an evaluation toolkit (ROI-SSIM/FSIM, NSD/HD95) with a model card.

6. Baselines and Ablations

Baselines: State-of-the-art diffusion fusion and strong non-diffusion fusion such as TTTFusion and others.
Ablations: Remove ROI guidance; vary (α, β, γ); auto vs. corrected masks; uncertainty modulation on/off; sampling steps; registration jitter amplitudes.

7. Expected Outcomes

(i) ROI-wise fidelity gains (brain/bone) and lesion boundary accuracy; (ii) robustness under plausible misregistration; (iii) spatially calibrated uncertainty overlays useful in reader studies; (iv) practical latency via few-step sampling/distillation.

References

[1] L. Maier-Hein *et al.*, "Metrics Reloaded: Recommendations for Image Analysis Validation," *Med. Image Anal.*, 2024.
[2] L. Maier-Hein *et al.*, "Metrics Reloaded," *arXiv:2206.01653*, 2022.
[3] S. Masitho *et al.*, "Impact of MRI Acquisition Setup and Registration Method on Brain Radiotherapy Planning," *Front. Oncol.*, 2022.
[4] J. Wasserthal *et al.*, "TotalSegmentator: Robust Segmentation of 104 Anatomical Structures in CT," *RSNA AI*, 2023.
[5] T. A. D'Antonoli *et al.*, "TotalSegmentator MRI: Sequence-Independent Segmentation," *arXiv:2405.19492*, 2024.
[6] MONAI Consortium, "Metrics: Normalized Surface Dice (NSD)," Docs, 2024.
[7] D. Karimi, S. E. Salcudean, "Generalized Surface Loss for Reducing HD," *arXiv:2302.03868*, 2023.
[8] B. Fei *et al.*, "A Diffusion Model for Universal Medical Image Enhancement," *Commun. Med.*, 2025.
[9] W. Wang *et al.*, "Diffusion Models for Medical Image Denoising, Reconstruction, and Translation," *CMIG*, 2025.
[10] Z. Huang *et al.*, "FusionDiff: A Unified Image Fusion Network Based on Diffusion Models," *CVIU*, 2024.
[11] X. Yi *et al.*, "Diff-IF: Multi-modality Image Fusion via Diffusion," *Inf. Fusion*, 2024.