Supplementary Material for Morphing Through Time: Diffusion-Based Bridging of Temporal Gaps for Robust Alignment in Change Detection

In this supplementary document, we provide extended qualitative results, ablations, and implementation details that complement the main paper. We focus on three benchmark datasets, LEVIR-CD, WHU-CD, and DSIFN-CD, and present detailed visual comparisons of (1) registration baselines, (2) the effect of refinement when applied to direct versus composed flows, (3) downstream change detection outputs, and (4) the impact of the number of intermediate frames in DiffMorpher.

A. Registration Baselines across Datasets

Figures S1, S2, and S3 show warped results on LEVIR-CD, DSIFN-CD, and WHU-CD, respectively. Each figure compares multiple dense registration backbones (RoMa, MASt3R, SuperPoint+SuperGlue, LoFTR, RAFT) against our composed and refined flows.

Across datasets, RAFT achieves low local EPE but struggles with global consistency, introducing shearing artifacts. MASt3R suffers severe deformation under strong viewpoint variation. LoFTR preserves broad structure but blurs fine boundaries. In contrast, our composed RoMa variant, aided by DiffMorpher intermediates, yields geometrically faithful reconstructions. The refinement stage further sharpens alignment and reduces residual drift.

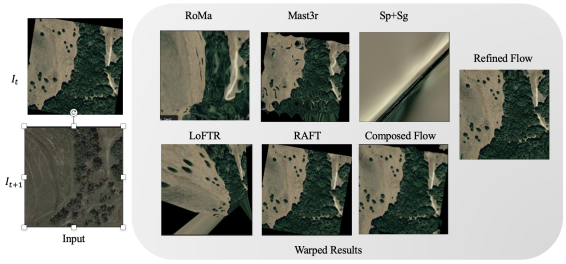

Figure S1. **LEVIR-CD registration results.** From left to right: unaligned inputs, warped results under different strategies, and CD predictions. Our composed+refined pipeline recovers structures and reduces false positives.

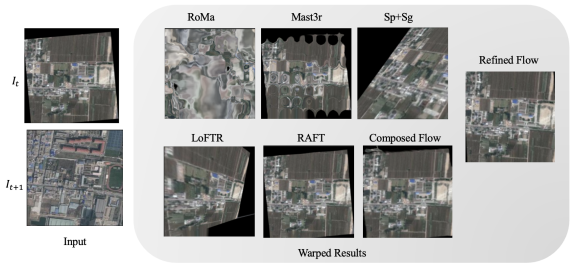

Figure S2. **DSIFN-CD registration results.** DSIFN contains fine-grained changes. Baselines distort these regions, while our composed+refined pipeline preserves small buildings and roads, improving recall of subtle changes.

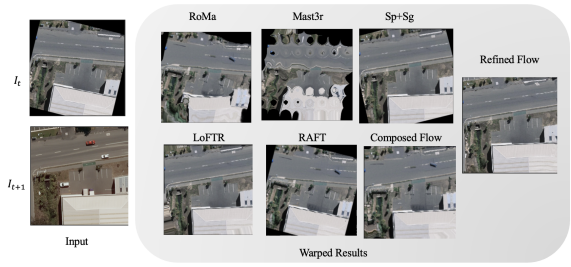

Figure S3. **WHU-CD registration results.** WHU involves viewpoint differences and seasonal variation. Baselines misalign rooftops and vegetation, while our refined pipeline restores global consistency and yields cleaner masks.

B. Flow Refinement Analysis

Figure S4 examines refinement applied to direct RoMa flows (top) versus composed RoMa flows (bottom). Refinement on direct RoMa improves alignment modestly but remains limited by large distortions. Refinement on composed flows, initialized with DiffMorpher intermediates, corrects structure more effectively, keeping building edges crisp and tree clusters coherent. This shows refinement is most effective when paired with morph-based composition.

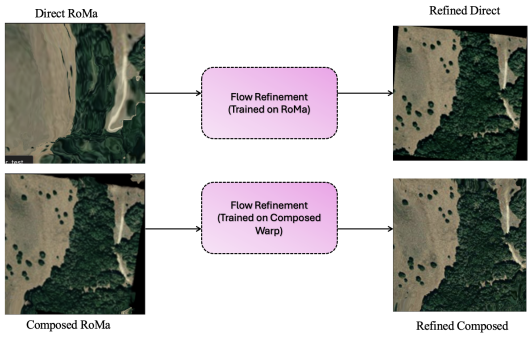

Figure S4. **Flow refinement analysis.** Top: refinement on direct RoMa flows. Bottom: refinement on composed flows. Refinement on composed flows produces sharper details and reduced distortion.

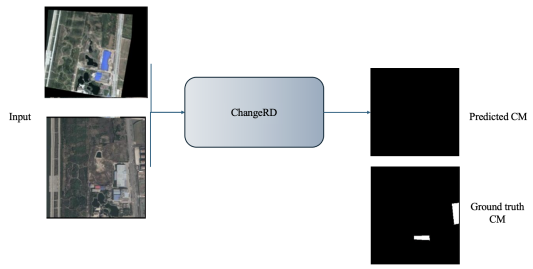

Figure S5. **DSIFN-CD CD outputs.** Left: input pair. Middle: predictions. Right: ground truth. Our refined pipeline enhances both precision and recall.

C. Change Detection Outputs

Figures S5, S6, and S7 present CD results on DSIFN-CD, LEVIR-CD, and WHU-CD. In each case, unaligned inputs yield noisy masks, while refined alignment produces predictions closer to the ground truth.

On DSIFN-CD, our method recovers subtle urban changes. On LEVIR-CD, refinement reduces false alarms and restores building contours. On WHU-CD, which has large viewpoint shifts, our pipeline suppresses spurious detections in vegetation and roads. These confirm the quantitative gains in the main paper.

D. Number of Generated Frames in DiffMorpher

We studied the effect of varying the number of intermediate frames N synthesized by DiffMorpher. With $N = 3$, intermediates add little new information, and accuracy quickly saturates. Increasing to $N = 5$ provides the best trade-off, producing sharper warps and improved downstream CD accuracy. At $N = 7$, results are similar to $N = 5$ but runtime and memory usage grow substantially. We therefore adopt

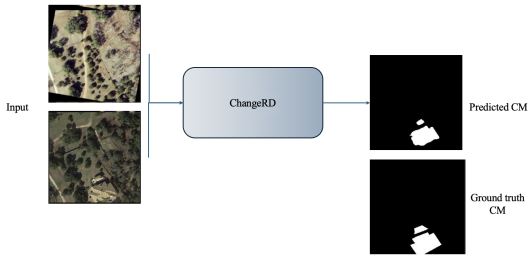

Figure S6. **LEVIR-CD CD outputs.** Refinement reduces false alarms and restores missing structures, yielding masks closer to ground truth.

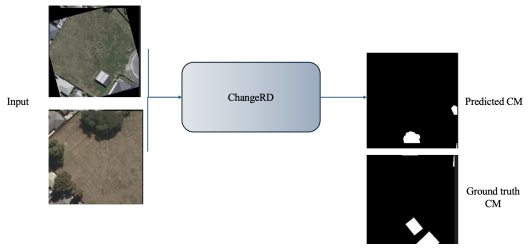

Figure S7. **WHU-CD CD outputs.** Severe misalignment produces fragmented masks. Our refined alignment yields cleaner boundaries and more stable predictions.

$N = 5$ in all main experiments.

N	PSNR↑	SSIM↑	mIoU↑
3	24.5	0.812	0.79
5	**26.1**	**0.834**	**0.83**
7	26.0	0.831	0.82

Table S1. Effect of intermediate frame count N on registration and CD. $N = 5$ gives the best accuracy-efficiency trade-off.

E. Reproducibility Notes

Hardware. All experiments used a single NVIDIA RTX A5500 GPU.
Datasets & splits. Datasets split 60/20/20 with seed 42; split files under `splits/{train,val,test}.txt`.
Preprocessing. Images resized to 256×256 RGB; only `ToTensor()` applied. Each sample loads `00.png`, `04.png`, and `roma_flow.npy`.
Model. Residual Refiner predicts residual flows added to RoMa flows (bilinear up/downsampling, skip connections, SE/fusion blocks).
Optimization. Adam optimizer, LR=1×10^{-4}, Smooth L1 loss, no schedule/decay.
Batching. Batch size 4. Training shuffled; val/test fixed.
Training length. $\sim$40 epochs per dataset with fixed

hyperparameters.

Checkpointing. Checkpoints saved each epoch; resume supported.

Evaluation. Scripts, metrics, and preprocessing identical to main paper. Figures exported from same evaluation path.