Supplementary Material for DiffRegCD: Integrated Registration and Change Detection with Diffusion Features

A. Additional Qualitative Results

In addition to the quantitative benchmarks reported in the main paper, we provide qualitative examples illustrating the joint behavior of our registration and change detection modules. Figure S1 shows one representative case: given bi-temporal satellite inputs (I_t, I_{t+1}), our model predicts dense flow fields that align closely with the ground-truth displacement, as well as binary change maps highlighting the structural modifications between the two timestamps.

Notably, the predicted flow captures the dominant global motion (e.g., large-scale translation), which facilitates accurate warping before differencing. The corresponding change map then successfully localizes object-level changes (e.g., newly constructed regions), while suppressing spurious differences arising from misalignment. This qualitative evidence supports our central claim that unified registration and change detection produces more robust results than treating the tasks independently.

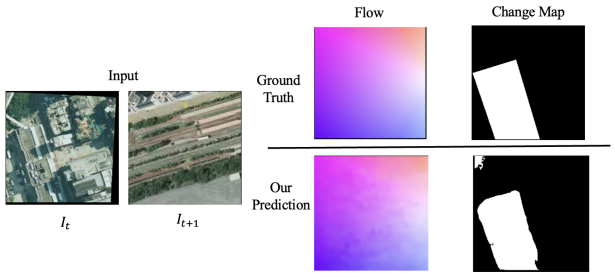

Figure S1. **Qualitative example of joint registration and change detection.** Given bi-temporal inputs (I_t, I_{t+1}), the model predicts dense optical flow and a binary change map. Compared with the ground truth, the predicted flow captures the main geometric displacement, enabling more accurate change localization.

To complement the quantitative evaluation, we provide qualitative comparisons across multiple datasets (e.g., WHU-CD, DSIFN-CD, and CMU-CD). These visualizations illustrate how our framework performs under diverse conditions, including strong misalignment, structural changes, and background clutter.

Figure S2 shows a representative case from the WHU-CD dataset. Baseline methods such as ChangeRD and DDPM-CD partially localize the change region but miss structural boundaries. BiFA suffers from noisy grid artifacts, while RoMa+DDPM-CD (pre-warp) reduces some misalignment but yields fragmented outputs. In contrast, our method produces a compact, accurate mask that aligns well with ground truth.

We include additional examples from other datasets in Figures S5, S3 and S4, where similar trends hold: unified registration and change detection consistently reduces spurious detections and recovers complete object-level changes.

Table S1 extends the robustness analysis from the main paper. BiFA, ChangeRD, and BIT were retrained on perturbed training data using the same augmentation protocol, while registration-only methods (RoMa, MASt3R) were applied sequentially with a CD head. DiffRegCD was trained once with synthetic flow supervision and tested directly.

Table S1. Robustness on VL-CMU-CD under induced misalignment (supplementary). BiFA, ChangeRD, and BIT were retrained on perturbed data.

Method	Low	Medium	High
BiFA	12.4	11.0	9.8
BIT	54.5	46.8	44.5
ChangeRD	73.1	67.4	58.9
MASt3R + CD	68.3	61.3	49.7
RoMa + CD	78.5	73.2	65.1
Ours	**82.1**	**79.4**	**72.0**

B. Additional Registration Results

To provide a more complete picture of registration performance, we report two complementary sets of results that were omitted from the main paper due to space constraints.

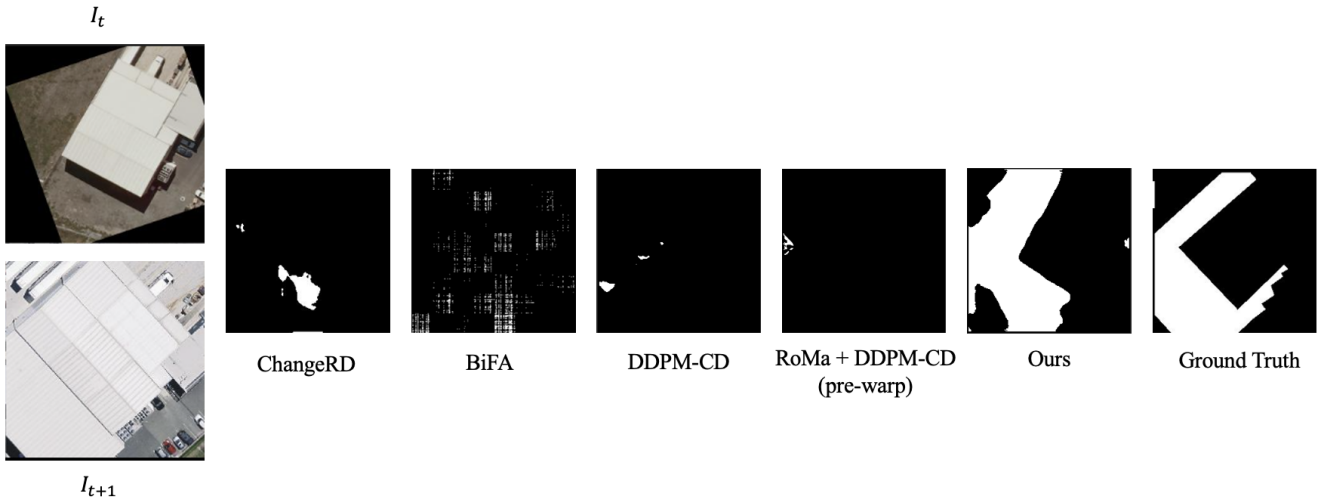

Figure S2. **Qualitative comparison on WHU-CD.** Our method generates more accurate and complete change masks under challenging geometric misalignment, compared to baselines (ChangeRD, BiFA, DDPM-CD, RoMa+DDPM-CD).

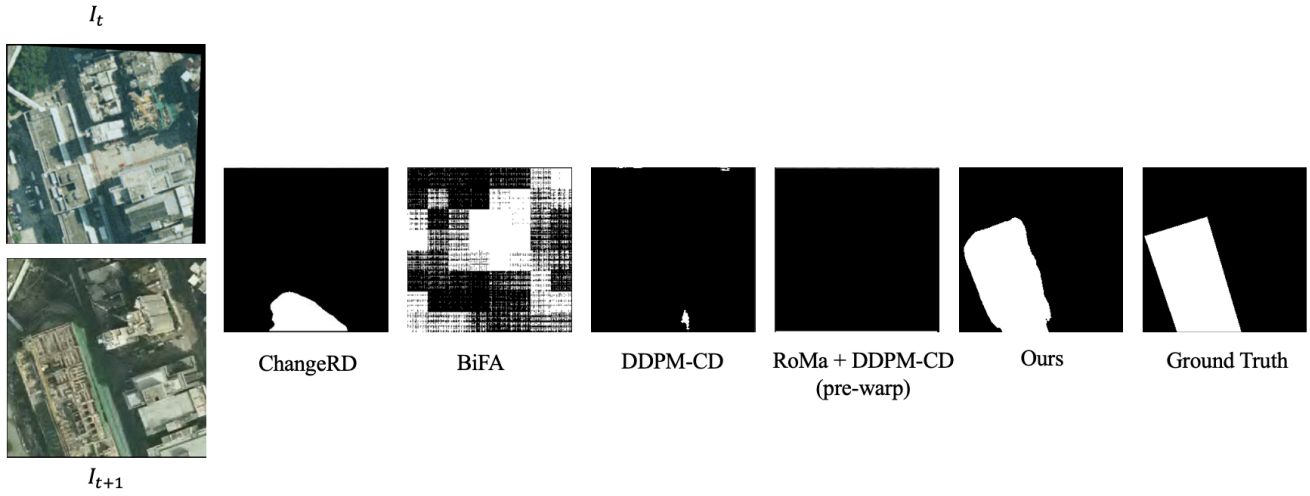

Figure S3. **Qualitative comparison on SYSU-CD.** Our method generates more accurate and complete change masks under challenging geometric misalignment, compared to baselines (ChangeRD, BiFA, DDPM-CD, RoMa+DDPM-CD).

B.1. Baseline Comparison (ECC/EPE)

Table S2 compares DiffRegCD against strong registration baselines across three datasets using **Enhanced Correlation Coefficient (ECC; lower is better)** and **Endpoint Error (EPE, px; lower is better)**. These metrics capture global alignment quality and pixel displacement accuracy. DiffRegCD ("RoMa+Refined") consistently achieves the lowest error, highlighting the effectiveness of our GP-smoothed transformer decoder.

Table S2. Registration comparison across LEVIR-CD, WHU-CD, and DSIFN-CD. Reported are ECC (lower is better) and EPE (lower is better).

Method	LEVIR (ECC/EPE)	WHU (ECC/EPE)	DSIFN (ECC/EPE)
SP+SG	.23 / 308.27	.38 / 243.96	.08 / 397.21
LoFTR	.24 / 229.34	.34 / 217.52	.10 / 275.00
RAFT	1.58 / 23.03	1.63 / 24.65	1.27 / 17.25
MASt3R	.48 / 57.60	.43 / 53.84	.67 / 82.66
RoMa	.89 / 29.41	.87 / 25.78	.99 / 21.76

B.2. Extended DiffRegCD Evaluation (AEPE/PCK)

We additionally evaluate DiffRegCD alone across four satellite datasets (LEVIR-CD, WHU-CD, DSIFN-CD, SYSU-CD) using **Average Endpoint Error (AEPE)** and **Percentage of Correct Keypoints at 3px (PCK@3)**. These metrics emphasize local correspondence precision

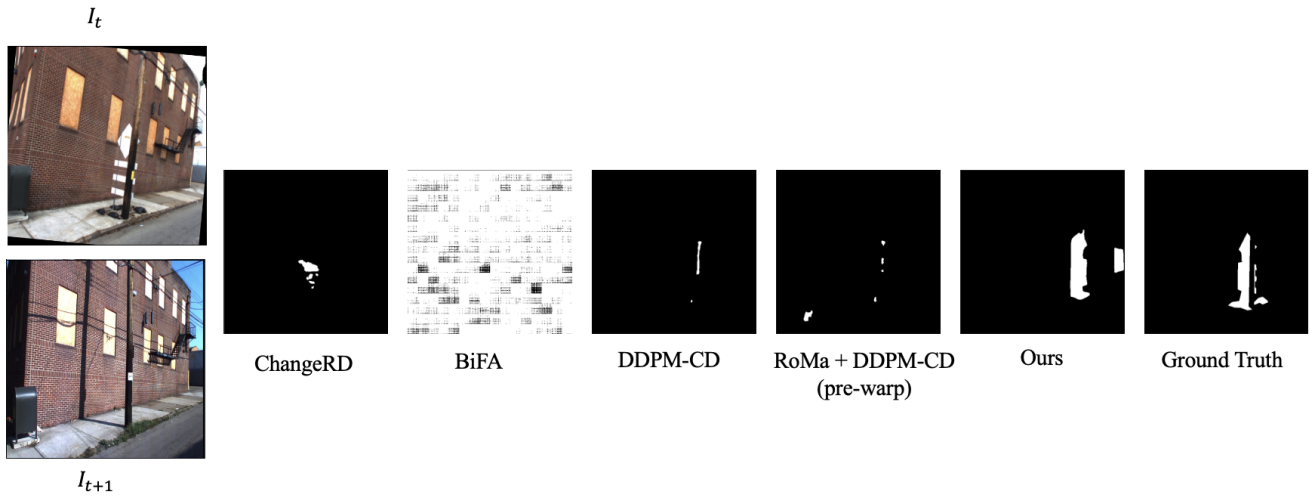

Figure S4. **Qualitative comparison on VL-CMU-CD.** Our method generates more accurate and complete change masks under challenging geometric misalignment, compared to baselines (ChangeRD, BiFA, DDPM-CD, RoMa+DDPM-CD).

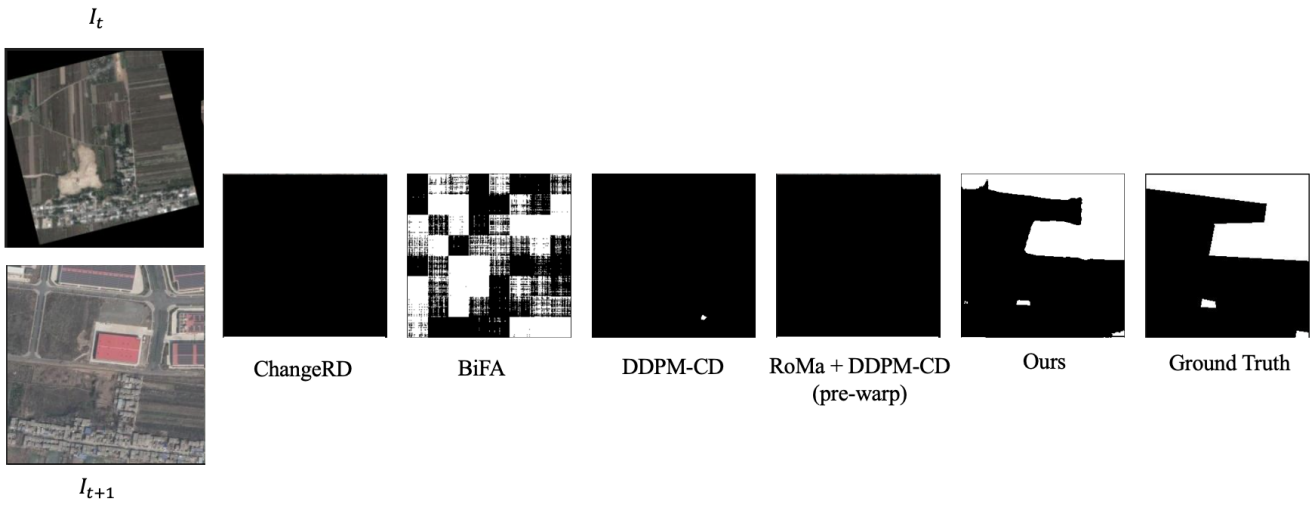

Figure S5. **Qualitative comparison on DSIFN-CD.** Our method generates more accurate and complete change masks under challenging geometric misalignment, compared to baselines (ChangeRD, BiFA, DDPM-CD, RoMa+DDPM-CD).

Table S3. Extended registration evaluation of DiffRegCD across four datasets. Reported are AEPE (px; lower is better) and 1-PCK@3 (%; lower is better).

Dataset	AEPE ↓	1- PCK@3 ↓
LEVIR-CD	21.73	0.008
WHU-CD	22.24	0.007
DSIFN-CD	21.44	0.006
SYSU-CD	21.48	0.008

and robustness. Results are summarized in Table S3.

The two tables provide complementary insights. ECC/EPE highlights *relative performance* against established registration methods, showing that DiffRegCD outperforms baselines. AEPE/PCK offers a *dataset-level breakdown* of our model alone. We report $1 - \text{PCK@3}$, which measures the fraction of correspondences whose endpoint error exceeds 3 pixels (lower is better). These results confirm that DiffRegCD produces flows that are globally smooth and locally reliable, enabling robust downstream change detection.

C. Model Complexity Analysis

To provide a comprehensive view of computational requirements, we report model complexity in terms of parameter

counts (millions, M) and floating-point operations (billions, G) at an input resolution of 256×256. This analysis is split into two parts: (i) a component-level breakdown of our framework, and (ii) a comparative study against representative change detection baselines.

Component-level breakdown. Table S4 details the contribution of each component in our architecture. The diffusion backbone (netG) constitutes the majority of parameters, accounting for approximately 391M parameters. The registration head (netReg) adds 143M parameters, while the change detection head (netCD) remains relatively lightweight at 46M parameters. Notably, the combined registration + CD branch contributes 189M parameters with 211G FLOPs at 256×256 input size, highlighting the efficiency of the downstream modules compared to the frozen backbone. The full model (backbone + Reg + CD) totals 580M parameters.

Component	Params (M)	FLOPs (G) @ 256×256
Diffusion backbone (netG)	391.048	N/A
Registration head (netReg)	143.212	N/A
Change detector (netCD)	46.405	N/A
Reg + CD (combined)	**189.617**	**211.278**
Total (netG + Reg + CD)	**580.665**	N/A

Table S4. Model complexity by component. Parameters in millions; FLOPs in billions.

Comparison with baselines. We further benchmark our model against widely used change detection architectures in Table S5. Methods such as BIT, VcT, and ResNet18+Transformer are lightweight (3–4M parameters) but operate with limited expressive power. Mid-sized architectures such as BiFA (5.6M, 53G) and SNUNet (10.2M, 176G) balance parameter count with FLOPs but lack robust registration capability. ChangeFormer (41M, 203G) is among the heaviest baselines. By contrast, our registration + CD branch ($\sim$190M parameters, 211G FLOPs) is substantially larger, reflecting the integration of explicit flow reasoning into the pipeline. Despite the higher complexity, our unified model achieves superior alignment and change detection accuracy across datasets , demonstrating that performance gains justify the computational cost.

Discussion. While our method exhibits a larger parameter footprint, two factors mitigate this concern. First, the diffusion backbone is frozen during training and inference, meaning the computational overhead primarily stems from a one-time forward pass rather than gradient updates. Second, the majority of FLOPs are concentrated in the registration + CD branch, which is well-optimized for 256×256 inputs and scales linearly with resolution. Therefore, although heavier than prior CD-only baselines, the design strikes a balance between complexity and accuracy, making it suitable for practical deployment in scenarios where robustness to misalignment is critical.

Method	Params (M)	FLOPs (G)
BIT	3.50	10.63
VcT	3.57	10.64
ResNet18+Transformer	3.64	11.74
STANet (BAM)	16.89	6.58
STANet (PAM)	16.93	26.32
BiFA	5.58	53.00
SNUNet	10.21	176.36
ChangeFormer	41.03	202.85
Ours (Reg + CD)	**189.617**	**211.278**

Table S5. Complexity comparison with baselines. Parameters in millions; FLOPs in billions.

Robustness Evaluation Protocol and Results

Metric. Robustness is measured using the F1-score of the predicted change maps against the ground-truth masks:

$$\mathrm{F1} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}},$$

where precision and recall are computed on the binary change masks. A higher F1 indicates that the model correctly captures true changes while suppressing false positives introduced by misalignment.

Perturbation setup. Standard CD benchmarks (e.g., LEVIR, VL-CMU) are pre-aligned, but real deployments often involve viewpoint and geometric shifts. We simulate these by applying synthetic transformations to the post-change image I^B:

- *Translation*: uniform shifts in x, y.
- *Rotation*: in-plane rotation about the image center.
- *Scaling*: isotropic zoom in/out.
- *Homography*: projective warps combining the above.

Each perturbation is sampled within ranges corresponding to three difficulty levels:

- **Low**: $\leq$4 px translation, $\pm2^\circ$ rotation, $\pm2\%$ scale, mild homographies.
- **Medium**: $\leq$8 px translation, $\pm5^\circ$ rotation, $\pm5\%$ scale.
- **High**: $\leq$16 px translation, $\pm10^\circ$ rotation, $\pm10\%$ scale, strong homographies.

All perturbations are applied at test time.

Fairness of comparison. To ensure fairness, BiFA, BIT, and ChangeRD are retrained on perturbed data using the same augmentation protocol. Registration methods (SuperPoint+SuperGlue, LoFTR, RoMa, MASt3R) are applied as

pre-processing before CD. DiffRegCD is trained once with synthetic flow supervision and tested directly under perturbations.

Results. Table S1 reports F1-scores on VL-CMU-CD. CD-only baselines degrade rapidly with displacement despite retraining (BiFA collapses almost entirely). Sequential pipelines (RoMa+CD, MASt3R+CD) are more robust but still lose accuracy under severe warps. DiffRegCD preserves over 80% of its aligned performance even in the high regime, confirming the effectiveness of explicit flow-guided alignment.