EASE-CR: EVIDENCE-AWARE MULTIMODAL REPRESENTATION LEARNING FOR CLOUD RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Clouds obscure critical information in optical satellite imagery, making reliable reconstruction fundamentally dependent on how complementary observations are interpreted. Existing CR (cloud-removal/reconstruction) methods primarily fuse optical and synthetic aperture radar (SAR) observations without explicitly reasoning about whether the available evidence is reliable, degraded or mutually inconsistent. We introduce EaSe-CR (Evidence-aware Sensing experts for Cloud Reconstruction), a reconstruction model that jointly learns modality reliability, cross-modal evidence interaction, disagreement and latent uncertainty, allowing the reconstruction process to adapt to the quality and consistency of available evidence. Our analysis show that the learnt representation responds systematically to changes in evidence quality and cross-modal consistency, while external evaluation demonstrates that EaSe-CR retains an advantage over fixed-fusion baselines under zero-shot, geographically diverse conditions spanning Kerala, Rajasthan and Ladakh.

1 INTRODUCTION

Reconstructing cloud-free optical imagery from multimodal satellite observations is challenging because the available evidence is inherently incomplete, degraded and potentially inconsistent across sensing modalities. Multimodal cloud reconstruction is commonly approached as a fusion problem, given a partially observed optical image and a complementary SAR observation, learn a mapping to the corresponding clear optical scene. This formulation has produced a progression from SAR-to-optical generative translation (Bermudez et al., 2018; 2019; Yang et al., 2022) to residual and multimodal fusion (Meraner et al., 2020; Ebel et al., 2021; Xu et al., 2022), and more recent attention- and diffusion-based reconstruction (Pan et al., 2024; Liu et al., 2024; Zou et al., 2024; Tu et al., 2025). While these approaches increasingly improve how information from multiple sensors is combined, they largely treat the available observations as fixed sources of complementary information.

This assumption becomes problematic when evidence quality varies. Cloud obscuration can progressively remove useful optical information, SAR observations can become degraded and the two modalities can become inconsistent despite remaining jointly available. Thus, the usefulness of a modality is not determined solely by its presence. More broadly, multimodal learning has studied missing modalities, latent modality structure and adaptive mixture-of-experts (Jiang et al., 2023; Hu et al., 2024; Yun et al., 2024; Xin et al., 2025), while uncertainty-aware cloud removal has considered variability in reconstruction predictions (Ebel et al., 2023). However, these directions do not directly establish how evidence quality and cross-modal consistency affect the *internal reconstruction representation*. This leaves an important question open: **When the available evidence changes, does the reconstruction process change with it?**

We address this question by treating multimodal reconstruction as an *evidence-dependent inference problem*. We first perform controlled interventions on modality availability, evidence quality and cross-modal consistency, and analyze their effects on reconstruction, learnt reliability and the shared representation. The resulting observations motivate **EaSe-CR (Evidence-aware Sensing Experts for Cloud Reconstruction)**, an evidence-conditioned multimodal mixture-of-experts architecture that explicitly models modality reliability, cross-modal agreement and disagreement, representation uncertainty, and spatially adaptive expert specialization.

Our contributions are:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

- We establish a controlled empirical framework for studying **modality contribution, evidence degradation and cross-modal inconsistency** in multimodal cloud reconstruction.
- We introduce **EaSe-CR**, which conditions expert routing on an explicit evidence state incorporating reliability, interaction, disagreement and representation uncertainty.
- We show that the shared representation responds systematically to evidence perturbations and that its variability contains information about reconstruction difficulty.
- We evaluate on an independent, geographically diverse zero-shot collection spanning Kerala, Ladakh and Rajasthan, to test transfer beyond the primary benchmark distribution.
- Our analysis show that the learnt representation responds systematically to changes in evidence quality and cross-modal consistency.

2 Related Work

SAR–Optical Cloud Reconstruction. Early work explored SAR-to-optical synthesis for cloud removal using conditional GANs (Bermudez et al., 2018; 2019; Yang et al., 2022; Wang et al., 2022; Noa Turnes et al., 2022; Zhao et al., 2022; 2025), followed by SAR–optical fusion and residual reconstruction (Meraner et al., 2020; Ebel et al., 2021; 2020; 2022; Xu et al., 2022; Wen et al., 2023; Xu et al., 2024; Xia et al., 2024; Wang et al., 2024; Pan et al., 2024; Liu et al., 2024; Tu et al., 2025; Liu et al., 2025; Yang et al., 2025; Wang et al., 2025a; Hu et al., 2025; He et al., 2026; Zhang et al., 2026; Xu et al., 2026; Wang et al., 2026).

Multimodal Representation and Evidence. Beyond reconstruction, multimodal learning has studied modality translation, cross-modal representation structure, missing modalities andflexible expert-based fusion (Chen & Bruzzone, 2022; Liu et al., 2022; Jiang et al., 2023; Zhu et al., 2024; Hu et al., 2024; Yun et al., 2024; Xin et al., 2025; Mustafa et al., 2022; Cheng et al., 2024; Han et al., 2024; Dufumier et al., 2025; Wang et al., 2025b), while remote-sensing foundation representations increasingly exploit multimodal and radar–optical observations (Fuller et al., 2023; Cong et al., 2022; Guo et al., 2024; Zhou et al., 2024).

Uncertainty and Representation Analysis. Uncertainty-aware reconstruction has been explored through stochastic cloud removal (Ebel et al., 2023), while predictive uncertainty, calibration and representation similarity provide broader foundations for analyzing model uncertainty and internal representations (Kendall & Gal, 2017; Lakshminarayanan et al., 2017; Guo et al., 2017; Kornblith et al., 2019; Chandu et al., 2025).

3 From Reconstruction to Evidence: Controlled Empirical Investigation

3.1 How Does Reconstruction Depend on Evidence?

Multimodal reconstruction implicitly assumes that the sensing channels provide complementary information, but their contribution need not be symmetric or remain constant as evidence quality changes. We therefore evaluate both modality contribution and degradation sensitivity under controlled inference. Removing either modality degrades reconstruction substantially, with optical removal producing the larger change across spatial, spectral andperceptual measures, while SAR removal also yields a statistically significant degradation. The joint model therefore benefits from **complementary evidence rather than simple redundancy**, with optical observations providing the dominant direct reconstruction signal and SAR contributing additional information beyond optical evidence alone.

Table 1: Evidence Ablation Results. Removing either modality degrades reconstruction, with a substantially larger effect for optical removal across spatial and spectral measures.

Evidence	MSE ↓	MAE ↓	PSNR ↑	SSIM ↑	SAM ↓
Optical + SAR	**0.001345**	**0.024505**	**29.520**	**0.8450**	**7.289°**
Optical only	0.016854	0.094035	18.557	0.5186	46.645°
SAR only	0.002402	0.036168	26.743	0.7979	12.729°

We next vary evidence quality while retaining the complementary modality. SAR is progressively degraded from 0-90%, while optical observations undergo controlled cloud-obscuration severities from 10-90%. As shown in Fig. 1, SAR degradation produces a pronounced nonlinear decline in reconstruction fidelity, whereas increasing optical obscuration results in a more gradual deterioration. Thus, the contribution of an observation cannot be adequately represented as simply present or absent, its utility changes with the quality of the available evidence. This motivates the next question of **whether modality reliability can be explicitly estimated from the evidence itself**.

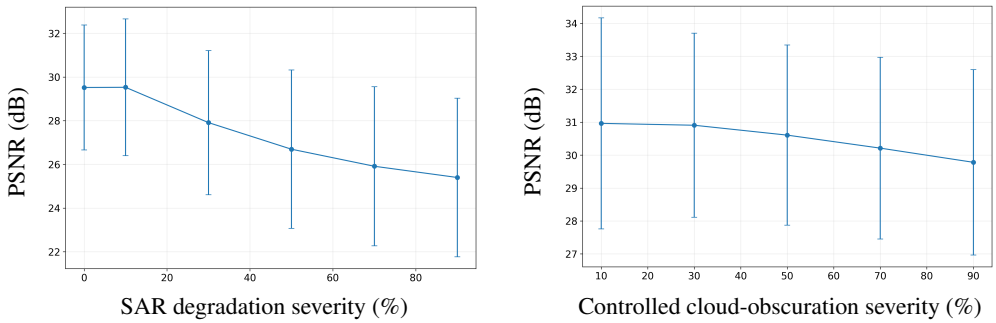

Figure 1: **Reconstruction fidelity under controlled evidence degradation. Left:** PSNR as SAR evidence is progressively degraded. **Right:** PSNR under increasing controlled cloud-obscuration severity. Error bars denote 95% bootstrap confidence intervals. SAR degradation produces a pronounced nonlinear decline in reconstruction fidelity, whereas increasing optical obscuration results in a substantially more gradual deterioration.

3.2 Can Evidence Reliability be Estimated?

The preceding analysis shows that modality utility changes with evidence quality, motivating a more direct question, **Can the usefulness of each modality be estimated from the observations themselves?** We therefore augment the fusion model with modality-specific reliability heads and evaluate whether their predictions track empirical modality utility. The predicted scores exhibit substantial correspondence with empirical reliability, with Pearson correlations of **0.617** for optical and **0.745** for SAR, while the predicted modality ordering agrees with the empirically more useful modality in 74.8% of samples. Calibration is non-trivial but measurable, with **ECE/Brier/NLL** of **0.126/0.193/0.570** for **optical** and **0.113/0.210/0.608** for **SAR**. Under controlled SAR degradation, the predicted SAR reliability decreases sharply while optical reliability remains approximately stable, providing further evidence that the reliability estimates respond selectively to changing evidence quality.

Table 2: Reliability estimation and fusion. Predicted reliability is correlated with empirical modality utility, but directly using the scores as scalar fusion weights does not consistently improve reconstruction: MSE decreases marginally, while MAE, PSNR, SSIM andSAM degrade.

Method	MSE ↓	MAE ↓	PSNR ↑	SSIM ↑	SAM ↓
Equal fusion	0.001512	0.024940	29.131	0.8468	7.575°
Reliability-weighted	0.001431	0.025397	29.104	0.8270	7.998°

The learnt scores contain useful information about evidence quality, directly using them to determine modality contribution does not consistently improve reconstruction fidelity. This suggests that **evidence quality should not merely determine how much a modality contributes, it may need to influence how the modalities are represented and interact before reconstruction**.

3.3 Does the Shared Representation Change When Evidence Becomes Unreliable?

If one evidence source becomes unreliable or inconsistent, does the shared representation change accordingly? We test this by independently corrupting the optical and SAR observations and measuring displacement of the resulting shared latent from the clean representation across increasing

corruption severity. We additionally introduce cross-modal mismatch by pairing an observation with SAR from a different sample, separating evidence degradation from inconsistent multimodal evidence.

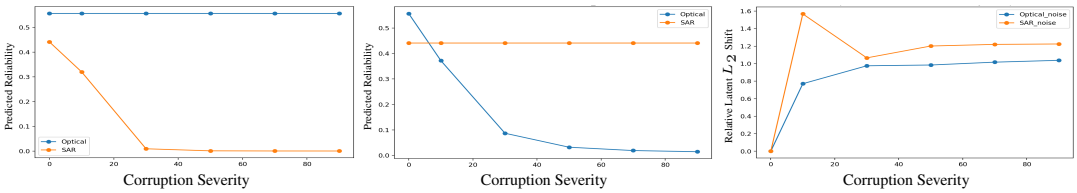

Figure 2: **Representation response to evidence perturbation. Left:** Reliability under SAR degradation. **Center:** Reliability under optical degradation. **Right:** Latent representation shift under modality corruption. Latent displacement increases sharply under modality corruption, while reliability selectively decreases for the perturbed modality and remains approximately stable for the complementary modality. The cross-modal mismatch produces a non-zero latent displacement (0.196 L_2) despite leaving the mean modality reliabilities essentially unchanged, distinguishing evidence inconsistency from simple degradation.

The shared representation responds systematically to evidence perturbation, latent-shift correlations with corruption severity are **0.732** for optical and **0.443** for SAR, while reliability selectively decreases for the perturbed modality. Cross-modal mismatch also produces a measurable latent displacement (**0.196** L_2), despite largely unchanged mean reliability. These results indicate that the **shared representation is sensitive to evidence conditions rather than invariant to modality corruption, providing empirical evidence for evidence-conditioned representation learning.**

3.4 What Does the Shared Representation Encode?

The preceding experiments establish that evidence quality affects reconstruction and reliability, but not whether these changes are represented internally. We therefore analyze the shared latent directly, asking **whether it preserves information about the prevailing evidence state while maintaining a common multimodal representation**. Across eight evidence states, the shared representation shows strong alignment with the optical and SAR representations (CKA **0.908** and **0.753**, respectively), while optical–SAR alignment is lower (**0.576**), indicating substantial shared structure alongside modality-specific information. CCA likewise reveals strong correspondence across representations, with the first eight canonical correlations ranging from **0.997** to **0.943**.

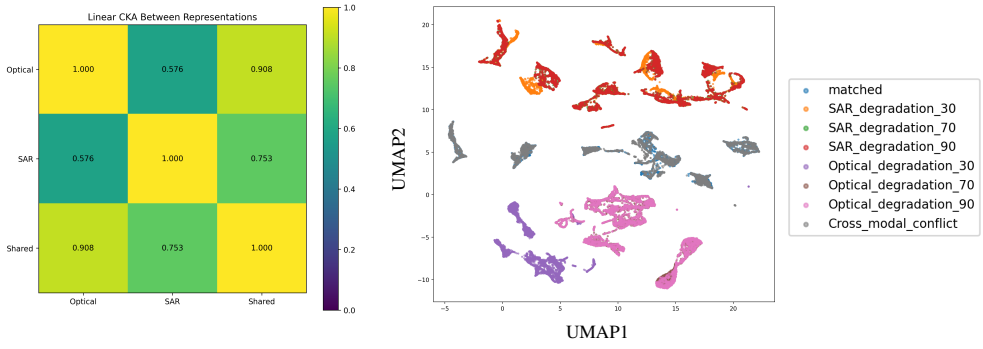

Figure 3: **Shared representation and evidence-state geometry. Left:** Linear CKA alignment between modality-specific and shared representations. **Right:** UMAP visualization of the shared latent across eight evidence states. The representation retains strong cross-modal alignment while encoding structured, overlapping evidence-state variation rather than forming discrete clusters.

The representation also retains measurable evidence-state structure, the first 50 PCA components explain 97.66% of the variance, while grouped 7-NN classification reaches 76.62%, compared with 12.5% chance. At the same time, the low silhouette score (0.012) indicates that these states do not

form cleanly separated clusters. Latent displacement nevertheless varies systematically with degradation, with mixed-effects slopes of 0.00617 for SAR and 0.00742 for optical severity. The shared latent therefore exhibits a continuous, overlapping geometry of evidence conditions, **it retains common multimodal structure while remaining sensitive to changes in the available evidence**.

3.5 Representation Variability and Reconstruction Error: Does Representation Uncertainty Track Reconstruction Difficulty?

The preceding analysis shows that the shared representation changes with evidence conditions, we next ask whether this variation provides a useful sample-level uncertainty signal. We use latent variance as an empirical representation-uncertainty proxy and test whether it tracks reconstruction difficulty across samples, rather than assuming uncertainty must increase monotonically with an externally imposed corruption level. Across 4,934 samples, latent-variance uncertainty is positively associated with reconstruction error (Pearson r : 0.299, Spearman ρ : 0.334; Holm-adjusted $p < 10^{-128}$), with progressively higher reconstruction error across the uncertainty quartiles.

Importantly, uncertainty does not increase monotonically under controlled SAR degradation, its mean rises from 1.29×10^{-4} at 0% to 1.56×10^{-4} at 20%, then returns to approximately 1.30×10^{-4} at higher severities. Thus, the proxy captures sample-dependent representation variability, but should not be interpreted as a direct measure of corruption severity or calibrated probabilistic uncertainty. The result instead supports a narrower claim: representation uncertainty contains measurable information about reconstruction difficulty, motivating its use as an internal diagnostic of evidence-dependent inference.

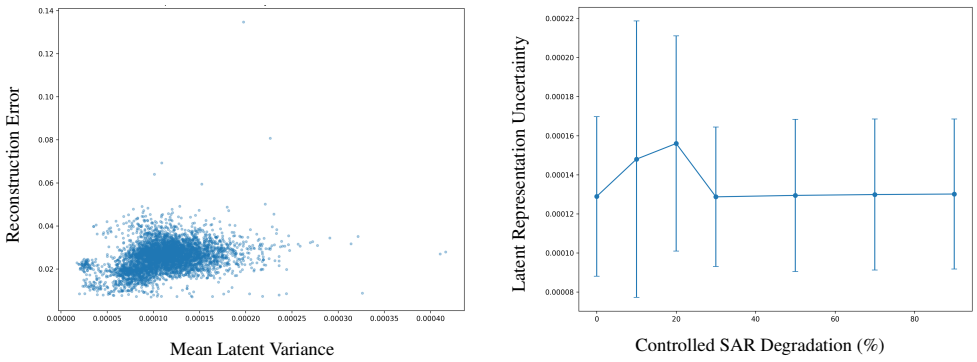

Figure 4: **Representation uncertainty as an empirical diagnostic. Left:** Latent-variance uncertainty exhibits a positive association with reconstruction error across samples ($r = 0.299$, $\rho = 0.334$). **Right:** Uncertainty does not vary monotonically with SAR degradation, indicating that the proxy captures sample-dependent representation variability rather than simply encoding corruption severity.

4 EaSe-CR: Evidence-Aware Sensing Experts for Cloud Reconstruction

The controlled analysis above show that SAR and optical observations provide complementary but asymmetric evidence, that their effective reliability varies across samples, and, that the shared representation changes systematically as evidence is degraded or becomes inconsistent. These observations suggest that multimodal reconstruction should not treat fusion as a fixed operation, the reconstruction pathway itself should depend on the state of the available evidence. We therefore introduce **EaSe-CR**, an evidence-conditioned multimodal mixture-of-experts architecture designed to transform modality-specific evidence into an adaptive reconstruction representation. Rather than assigning a fixed contribution to each sensing modality, EaSe-CR (Refer Fig. 5) explicitly models modality reliability, cross-modal agreement, representation uncertainty and expert specialization before reconstruction.

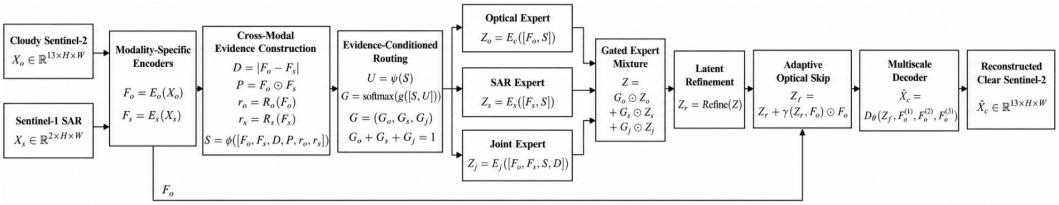

Figure 5: **Overview of EaSe-CR.** Evidence-conditioned routing of optical, SAR andjoint experts for clear-S2 reconstruction.

4.1 Evidence-conditioned representation and expert routing

Let $X_o \in \mathbb{R}^{13\times H\times W}$ denote the cloudy Sentinel-2 observation and $X_s \in \mathbb{R}^{2\times H\times W}$ the co-registered Sentinel-1 observation. Two independent encoders preserve modality-specific representations,

$$F_o = E_o(X_o), \qquad F_s = E_s(X_s),$$

where $F_o, F_s \in \mathbb{R}^{256\times 16\times 16}$. From these representations, modality-specific reliability fields are estimated as

$$r_o = R_o(F_o), \qquad r_s = R_s(F_s),$$

with $r_o, r_s \in [0,1]^{1\times 16\times 16}$. These fields are not imposed as externally defined confidence scores, they are jointly learnt with reconstruction and provide spatial evidence descriptors. We then construct an explicit evidence state by combining modality content, cross-modal agreement and disagreement:

$$D = |F_o - F_s|, \qquad P = F_o \odot F_s,$$
$$S = \phi([F_o, F_s, D, P, r_o, r_s]).$$

Here, D exposes locations where the two sensing representations diverge, while P captures their multiplicative interaction. The resulting state $S \in \mathbb{R}^{256\times 16\times 16}$ therefore represents not only what each modality observes, but also how their evidence relates at the current spatial location. From S, EaSe-CR predicts a non-negative empirical uncertainty field,

$$U = \psi(S), \qquad U \geq 0,$$

and uses both state and uncertainty to produce spatial expert-routing probabilities,

$$G = \mathrm{softmax}(g([S, U])), \qquad G_o + G_s + G_j = 1.$$

The uncertainty variable is used here as a learnt representation of evidence variability.

4.2 Modality-specialized reconstruction

The routing mechanism controls three deliberately different reconstruction pathways:

$$Z_o = \mathcal{E}_o([F_o, S]),$$
$$Z_s = \mathcal{E}_s([F_s, S]),$$
$$Z_j = \mathcal{E}_j([F_o, F_s, S, D]).$$

The first two experts retain modality-specific reconstruction pathways conditioned on the shared evidence state, whereas the joint expert receives both modalities together with their explicit disagreement representation. The routed latent is therefore

$$\boxed{Z = G_o \odot Z_o + G_s \odot Z_s + G_j \odot Z_j}$$

where the gates are spatial fields rather than sample-level scalar coefficients. Consequently, the model can continuously interpolate between optical-dominant, SAR-dominant andjoint reconstruction regimes instead of committing to a single sensing source. This formulation also avoids reducing

multimodal evidence to a single scalar reliability weight, reliability contributes to the evidence state, while the resulting state and uncertainty jointly determine which reconstruction computation is appropriate.

The routed representation is subsequently refined,

$$Z_r = \mathcal{R}(Z),$$

and combined with the deep optical representation through an adaptive skip,

$$Z_f = Z_r + \gamma(Z_r, F_o) \odot F_o.$$

The resulting latent is decoded using the first three optical encoder scales,

$$\hat{X}_c = D_\theta(Z_f, F_o^{(1)}, F_o^{(2)}, F_o^{(3)}),$$

to recover the 13-channel clear Sentinel-2 image. A final residual refinement produces the bounded reconstruction $\hat{X}_c \in [0,1]^{13\times256\times256}$. The complete forward mapping is therefore summarized as

$$\boxed{(X_o, X_s) \to (F_o, F_s) \to (S, U, G) \to (Z_o, Z_s, Z_j) \to Z_f \to \hat{X}_c.}$$

The implementation follows this formulation directly: the final model contains separate 13- and 2-channel encoders, three explicit experts, spatial reliability and uncertainty heads, a three-way softmax router, latent refinement, adaptive optical skip gating and a multiscale decoder, with 22.46M trainable parameters.

4.3 Fidelity-aware reconstruction objective

Because the target is multispectral reconstruction, optimizing only pointwise error does not fully reflect structural and spectral fidelity. We therefore train EaSe-CR using a composite objective

$$\mathcal{L} = \mathcal{L}_1 + 0.25\mathcal{L}_{\text{MSE}} + 0.30(1 - \text{SSIM}) + 0.15\mathcal{L}_{\text{spec}} + 0.02\mathcal{L}_{\text{grad}},$$

where $\mathcal{L}_{\text{spec}}$ is the spectral cosine discrepancy between reconstructed and target pixel spectra and $\mathcal{L}_{\text{grad}}$ measures horizontal and vertical gradient disagreement. Thus, the objective jointly constrains radiometric error, structural similarity, spectral direction and local spatial variation.

EaSe-CR therefore couples two levels of adaptation: **the evidence state determines what information is available and how reliable it appears**, while **expert routing determines which reconstruction pathway should process that evidence**. This directly operationalizes the empirical findings preceding the model, multimodal reconstruction is treated not as fixed feature fusion, but as evidence-conditioned reconstruction under heterogeneous sensing quality.

5 Experimental Setup and Evaluation Protocol

We evaluate cloud reconstruction on SEN12MS-CR Spring (Schmitt et al., 2019), using the multimodal setting in which cloudy Sentinel-2 observations and Sentinel-1 SAR observations are used to reconstruct the corresponding clear Sentinel-2 target. The dataset contains 29,117 paired samples, split into 20,301/3,882/4,934 training, validation and test samples across 28/6/7 ROIs, respectively. Each sample contains a 256×256 Sentinel-2 observation with 13 channels and a 2-channel Sentinel-1 SAR observation. We deliberately retain this single-season benchmark as a controlled training distribution, without external pretraining or additional seasonal data, allowing the evaluation to focus on the representation and inductive bias of the reconstruction architecture. Model selection is performed exclusively on the validation set, while all primary results are reported on the complete 4,934-sample test set.

Reconstruction fidelity is evaluated using MSE, MAE, PSNR, SSIM and SAM, covering pixel-wise, structural and spectral fidelity. All metrics are computed between the reconstructed and ground-truth clear Sentinel-2 observations over the complete test set. Controlled evidence experiments evaluate modality removal and progressive evidence degradation at inference while keeping the reconstruction procedure otherwise unchanged. Paired statistical tests are performed over corresponding samples, with Bonferroni correction for multiple comparisons, while degradation curves report 95% bootstrap confidence intervals. To assess generalization beyond the standard benchmark distribution, we additionally construct an independent validation collection spanning **Kerala, Ladakh and**

Rajasthan, representing substantially different geographic, terrain and environmental conditions. After modality matching, the collection contains 337 paired 256×256 image patches with corresponding cloudy optical, clear optical and SAR observations. The collection is used strictly for zero-shot evaluation, with no fine-tuning, adaptation or parameter updates, thereby testing whether the learnt reconstruction function transfers beyond the benchmark's geographic distribution.

6 RESULTS

6.1 RECONSTRUCTION QUANTIFICATION PERFORMANCE

We first compare EaSe-CR with the evaluated modality-specific and multimodal fusion baselines under the common evaluation protocol. The results are summarized in Table 3. EaSe-CR achieves the highest PSNR and SSIM among the evaluated multimodal models, while also obtaining the lowest MSE. Relative to the strongest fixed-fusion baseline, EarlyFusion, EaSe-CR improves PSNR from 29.52 to **29.72** dB and SSIM from 0.8450 to **0.8756**, while reducing MSE from 1.345×10^{-3} to 1.325×10^{-3}. The improvement in structural fidelity is accompanied by a small trade-off in MAE and SAM, indicating that the gain is not uniform across all reconstruction criteria.

Table 3: Quantitative reconstruction performance of EaSe-CR and the evaluated baselines.

Method	Input	MSE ↓	MAE ↓	PSNR ↑	SSIM ↑	SAM ↓
Pix2Pix	Optical	0.002170	0.030873	27.453	0.7988	9.722°
SAR2OPT	SAR	0.002736	0.036251	26.061	0.7514	12.039°
Early Fusion	Optical + SAR	0.001345	0.024505	29.520	0.8450	7.289°
Late Fusion	Optical + SAR	0.001458	0.025369	29.021	0.8331	8.216°
Cross-Attn.	Optical + SAR	0.001836	0.027235	28.687	0.8322	7.847°
EaSe-CR	**Optical + SAR**	**0.001325**	0.024728	**29.720**	**0.8756**	7.451°

6.2 EXTERNAL VALIDATION AND GENERALIZATION

We further evaluate the frozen models on an independently constructed external set derived from Sentinel-2 Level-1 observations collected between 2018 and 2025. For each sample, cloudy and clear observations are temporally matched within a maximum separation of five days, with corresponding Sentinel-1 SAR observations paired to each sample, providing geographically distinct evaluation pairs outside the primary benchmark. Table 4 summarizes the results. EaSe-CR achieves 24.302 dB PSNR, 0.7375 SSIM and 10.636° SAM, compared with 23.835 dB, 23.674 dB and 23.729 dB PSNR for Early Fusion (Meraner et al., 2020), Late Fusion (Ebel et al., 2021) and Cross-Attention (Xu et al., 2022), respectively. It also yields substantially lower SAM than the fixed-fusion models, indicating improved spectral fidelity under the external distribution.

Table 4: Zero-Shot External validation performance on geographically distinct matched samples from Kerala, Ladakh and Rajasthan

Method	MSE ↓	MAE ↓	PSNR ↑	SSIM ↑	SAM ↓
Cloudy Input	0.019993	0.079470	21.283	–	–
DiffCR	0.024176	0.082279	22.481	0.5174	19.460°
Early Fusion	0.021419	0.073919	23.835	0.7211	15.982°
Late Fusion	0.023067	0.076153	23.674	0.7047	16.301°
Cross-Attn.	0.022460	0.075425	23.729	0.7112	16.111°
EaSe-CR	0.023348	0.075683	**24.302**	**0.7375**	**10.636°**

Table 5: Regional external validation of EaSe-CR across geographically distinct regions.

Region	PSNR ↑	SSIM ↑	SAM ↓
Kerala	30.066	0.8746	9.452°
Rajasthan	26.605	0.8628	10.310°
Ladakh	17.312	0.5368	26.532°

Regional evaluation further reveals substantial geographic variation, as shown in Fig. 6. EaSe-CR achieves 30.066 dB PSNR in Kerala, 26.605 dB in Rajasthan and 17.312 dB in Ladakh, with corresponding SAM values of 9.452°, 10.310° and 26.532°. This variation reflects the heterogeneous difficulty of reconstruction across geographic conditions, while the external evaluation provides an additional test of transfer beyond the primary benchmark.

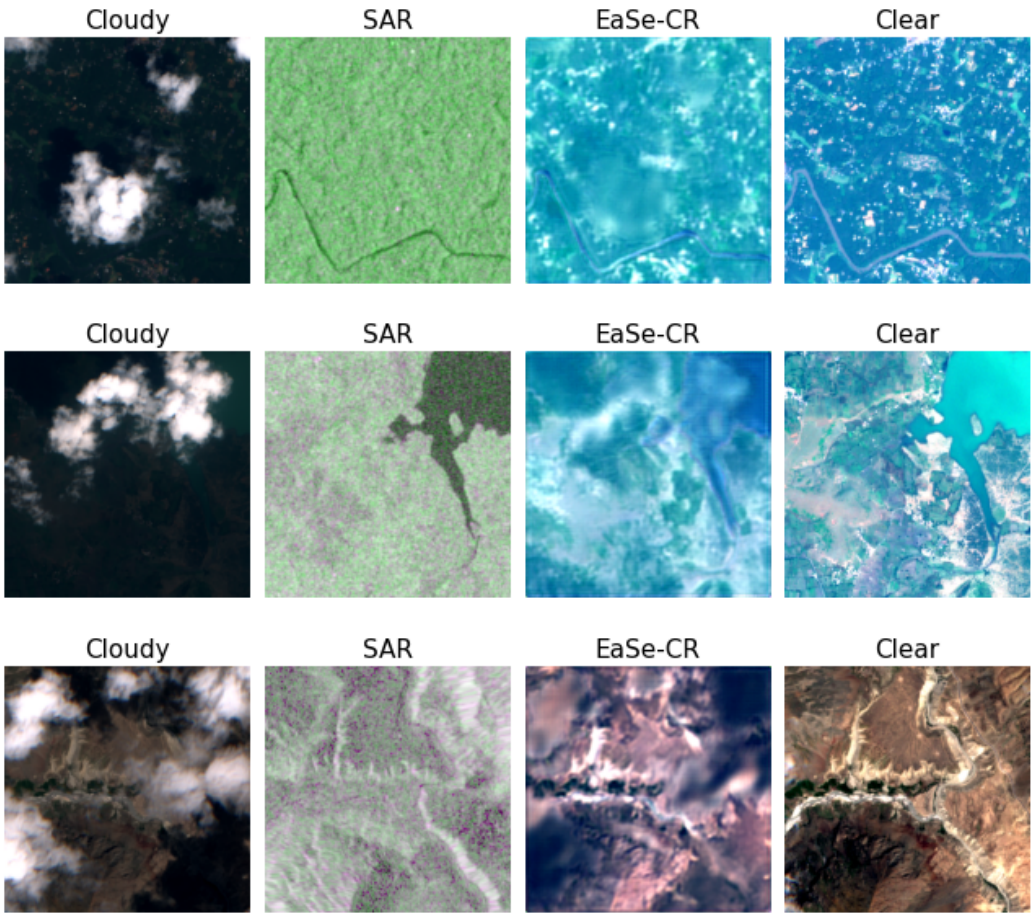

Figure 6: **Qualitative cloud reconstruction across geographically diverse regions.** Kerala shows a reconstruction PSNR of 34.43 dB, Rajasthan achieves 33.10 dB, and Ladakh achieves 29.07 dB. Each patch tile presents the cloudy optical observation, SAR observation, EaSe-CR reconstruction and corresponding clear reference, demonstrating reconstruction performance across geographically distinct regions.

7 Conclusion

We presented EaSe-CR, an evidence-aware framework for multimodal SAR-optical cloud reconstruction that conditions the reconstruction pathway on modality reliability, cross-modal interaction, disagreement and representation variability. Our experiments show that reconstruction depends not only on the availability of complementary observations, but also on the quality and consistency of the evidence they provide. The learnt representation exhibits corresponding sensitivity to these evidence conditions, while EaSe-CR achieves the strongest multimodal reconstruction performance among the evaluated methods and retains an advantage under zero-shot external evaluation across geographically diverse conditions.

Future work will extend the framework to broader temporal and geographic settings, incorporate additional contextual observations such as temporal information and digital elevation data and investigate calibrated probabilistic uncertainty.

REFERENCES

Jose D. Bermudez, Patrick N. Happ, Dario A. B. Oliveira, and Raul Q. Feitosa. Sar to optical image synthesis for cloud removal with generative adversarial networks. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume IV-1, pp. 5–11, 2018. doi: 10.5194/isprs-annals-IV-1-5-2018.

Jose D. Bermudez, Patrick N. Happ, Raul Q. Feitosa, and Dario A. B. Oliveira. Synthesis of multispectral optical images from sar/optical multitemporal data using conditional generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, 16(8):1220–1224, 2019. doi: 10.1109/LGRS.2019.2894734.

Khyathi Chandu, Linjie Li, Anas Awadalla, Ximing Lu, Jae Sung Park, Jack Hessel, Lijuan Wang, and Yejin Choi. Certainlyuncertain: A benchmark and metric for multimodal epistemic and aleatoric awareness. In *International Conference on Learning Representations*, 2025.

Yuxing Chen and Lorenzo Bruzzone. Self-supervised sar-optical data fusion of sentinel-1/-2 images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022. doi: 10.1109/TGRS.2021.3128072.

Ying Cheng, Yang Li, Junjie He, and Rui Feng. Mixtures of experts for audio-visual learning. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-0007.

Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David B. Lobell, and Stefano Ermon. Satmae: Pre-training transformers for temporal and multispectral satellite imagery. In *Advances in Neural Information Processing Systems*, volume 35, pp. 197–211, 2022. doi: 10.52202/068431-0015.

Benoit Dufumier, Javiera Castillo Navarro, Devis Tuia, and Jean-Philippe Thiran. What to align in multimodal contrastive learning? In *International Conference on Learning Representations*, 2025.

Patrick Ebel, Michael Schmitt, and Xiao Xiang Zhu. Cloud removal in unpaired sentinel-2 imagery using cycle-consistent gan and sar-optical data fusion. In *2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2065–2068, 2020. doi: 10.1109/IGARSS39084.2020.9324060.

Patrick Ebel, Andrea Meraner, Michael Schmitt, and Xiao Xiang Zhu. Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):5866–5878, 2021. doi: 10.1109/TGRS.2020.3024744.

Patrick Ebel, Andrea Meraner, Michael Schmitt, and Xiao Xiang Zhu. Sen12ms-cr-ts: A remote-sensing data set for multimodal multitemporal cloud removal. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 2022. doi: 10.1109/TGRS.2022.3146246.

Patrick Ebel, Vivien Sainte Fare Garnot, Michael Schmitt, Jan Dirk Wegner, and Xiao Xiang Zhu. Uncrtaints: Uncertainty quantification for cloud removal in optical satellite time series. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2086–2096, 2023.

Anthony Fuller, Koreen Millard, and James R. Green. Croma: Remote sensing representations with contrastive radar-optical masked autoencoders. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 1321–1330, 2017.

Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, Huimei He, Jian Wang, Jingdong Chen, Ming Yang, Yongjun Zhang, and Yansheng Li. Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27672–27683, 2024. doi: 10.1109/CVPR52733.2024.02613.

Xing Han, Huy Nguyen, Carl Harris, Nhat Ho, and Suchi Saria. Fusemoe: Mixture-of-experts transformers for fleximodal fusion. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-2167.

Zhongchen He, Peng Wang, Yunkun Zou, Bo Huang, Daiyin Zhu, Harry F. Lee, and Henry Leung. Dadigan: A dual attention blocks-based disentangled iterative generative adversarial network for cloud and shadow removal on sar and optical images. *Information Fusion*, 125:103487, 2026. doi: 10.1016/j.inffus.2025.103487.

Lianyu Hu, Tongkai Shi, Wei Feng, Fanhua Shang, and Liang Wan. Deep correlated prompting for visual recognition with missing modalities. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Yuyang Hu, Suhas Lohit, Ulugbek S. Kamilov, and Tim K. Marks. Multimodal diffusion bridge with attention-based sar fusion for satellite image cloud removal. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 2025. doi: 10.1109/TGRS.2025.3604654.

Qian Jiang, Changyou Chen, Han Zhao, Liqun Chen, Qing Ping, Son Dinh Tran, Yi Xu, Belinda Zeng, and Trishul Chilimbi. Understanding and constructing latent modality structures in multimodal representation learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7661–7671, 2023.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 3519–3529, 2019.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Rui Liu, Siqi Meng, Yini Peng, and Xin Tian. Transfusion-cr: Two-phase sar-to-optical translation and deep feature fusion for cloud removal. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 2024. doi: 10.1109/TGRS.2024.3439854.

Xun Liu, Danfeng Hong, Jocelyn Chanussot, Baojun Zhao, and Pedram Ghamisi. Modality translation in remote sensing time series. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 2022. doi: 10.1109/TGRS.2021.3079294.

Yi Liu, Wengen Li, Jihong Guan, Shuigeng Zhou, and Yichao Zhang. Effective cloud removal for remote sensing images by an improved mean-reverting denoising model with elucidated design space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17851–17861, 2025.

Andrea Meraner, Patrick Ebel, Xiao Xiang Zhu, and Michael Schmitt. Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:333–346, 2020. doi: 10.1016/j.isprsjprs.2020.05.013.

Basil Mustafa, Carlos Riquelme, Joan Puigcerver, Rodolphe Jenatton, and Neil Houlsby. Multimodal contrastive learning with limoe: the language-image mixture of experts. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0695.

Javier Noa Turnes, Jose David Bermudez Castro, Daliana Lobo Torres, Pedro Juan Soto Vega, Raul Queiroz Feitosa, and Patrick N. Happ. Atrous cgan for sar to optical image translation. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. doi: 10.1109/LGRS.2020.3031199.

Jun Pan, Jiangong Xu, Xiaoyu Yu, Guo Ye, Mi Wang, Yumin Chen, and Jianshen Ma. Hdrsa-net: Hybrid dynamic residual self-attention network for sar-assisted optical image cloud and shadow removal. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218:258–275, 2024. doi: 10.1016/j.isprsjprs.2024.10.026.

Michael Schmitt, Lloyd H. Hughes, Chen Qiu, and Xiao Xiang Zhu. Sen12ms – a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W7:153–160, 2019. doi: 10.5194/isprs-annals-IV-2-W7-153-2019.

Kelong Tu, Chao Yang, Yaxian Qing, Kunlun Qi, Nengcheng Chen, and Jianya Gong. Cloud removal with optical and sar imagery via multimodal similarity attention. *ISPRS Journal of Photogrammetry and Remote Sensing*, 226:116–126, 2025. doi: 10.1016/j.isprsjprs.2025.05.004.

Jun Wang, Xinhao Xu, Changqing Lin, and Feng Wang. Causally regularized full-resolution channel attention and frequency refinement for cloud removal in satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 235:133–152, 2026. doi: 10.1016/j.isprsjprs.2026.02.036.

Peng Wang, Yongkang Chen, Bo Huang, Daiyin Zhu, Tongwei Lu, Mauro Dalla Mura, and Jocelyn Chanussot. Mtgan: A sar-to-optical image translation method for cloud removal. *ISPRS Journal of Photogrammetry and Remote Sensing*, 225:180–195, 2025a.

Yuxi Wang, Bing Zhang, Wenjuan Zhang, Danfeng Hong, Bin Zhao, and Zhen Li. Cloud removal with sar-optical data fusion using a unified spatial–spectral residual network. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–20, 2024. doi: 10.1109/TGRS.2023.3339210.

Zehan Wang, Ziang Zhang, Minjie Hong, Hang Zhang, Luping Liu, Rongjie Huang, Xize Cheng, Shengpeng Ji, Tao Jin, Hengshuang Zhao, and Zhou Zhao. Omnibind: Large-scale omni multimodal representation via binding spaces. In *International Conference on Learning Representations*, 2025b.

Zhaobin Wang, Yikun Ma, and Yaonan Zhang. Hybrid cgan: Coupling global and local features for sar-to-optical image translation. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 2022. doi: 10.1109/TGRS.2022.3212208.

Zhenyu Wen, Jiahui Suo, Jie Su, Bingning Li, et al. Edge-sar-assisted multimodal fusion for enhanced cloud removal. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. doi: 10.1109/LGRS.2023.3307240.

Yu Xia, Wei He, Qi Huang, Guoying Yin, Wenbin Liu, and Hongyan Zhang. Crformer: Multimodal data fusion to reconstruct cloud-free optical imagery. *International Journal of Applied Earth Observation and Geoinformation*, 128:103793, 2024. doi: 10.1016/j.jag.2024.103793.

Jiayi Xin, Sukwon Yun, Jie Peng, Inyoung Choi, Jenna L. Ballard, Tianlong Chen, and Qi Long. I^2moe: Interpretable multimodal interaction-aware mixture-of-experts. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 68870–68888, 2025.

Fang Xu, Yilei Shi, Patrick Ebel, Lei Yu, Gui-Song Xia, Wen Yang, and Xiao Xiang Zhu. Glf-cr: Sar-enhanced cloud removal with global–local fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192:268–278, 2022. doi: 10.1016/j.isprsjprs.2022.08.002.

Fang Xu, Yilei Shi, Patrick Ebel, Wen Yang, and Xiao Xiang Zhu. Multimodal and multiresolution data fusion for high-resolution cloud removal: A novel baseline and benchmark. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. doi: 10.1109/TGRS.2023.3337845.

Meng Xu, Ruihua Zhang, Shuyi Liu, Xiuping Jia, Qingquan Li, and Sen Jia. Nam-cr: Noise-aware multisensor cloud removal for optical remote sensing imagery by fusing sar data. *Information Fusion*, 133:104295, 2026. doi: 10.1016/j.inffus.2026.104295.

Chen Yang, Qinxu Gao, Guanhua Chen, and Huiying Li. Dbd-cr: Dual-branch diffusion residual reconstruction for cloud removal in optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 2025. doi: 10.1109/TGRS.2025.3550422.

Xi Yang, Jingyi Zhao, Ziyu Wei, Nannan Wang, and Xinbo Gao. Sar-to-optical image translation based on improved cgan. *Pattern Recognition*, 121:108208, 2022. doi: 10.1016/j.patcog.2021.108208.

Sukwon Yun, Inyoung Choi, Jie Peng, Yangfan Wu, Jingxuan Bao, Qiyiwen Zhang, Jiayi Xin, Qi Long, and Tianlong Chen. Flex-moe: Modeling arbitrary modality combination via the flexible mixture-of-experts. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-3135.

Zaiyan Zhang, Jie Li, Yuanqi Liang, Jining Yan, Yi Xiao, Xin Su, and Qiangqiang Yuan. Ecrformer: An efficient cloud removal transformer with semantic-decoupled learning for multimodal satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 237:323–338, 2026. doi: 10.1016/j.isprsjprs.2026.04.009.

Wenbo Zhao, Nana Jiang, Xiaoxin Liao, and Jie Zhu. Hvt-cgan: Hybrid vision transformer cgan for sar-to-optical image translation. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 2025. doi: 10.1109/TGRS.2024.3523040.

Yitao Zhao, Turgay Celik, Nanqing Liu, and Heng-Chao Li. A comparative analysis of gan-based methods for sar-to-optical image translation. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5, 2022. doi: 10.1109/LGRS.2022.3177001.

Hangyu Zhou, Chia-Hsiang Kao, Cheng Perng Phoo, Utkarsh Mall, Bharath Hariharan, and Kavita Bala. Allclear: A comprehensive dataset and benchmark for cloud removal in satellite imagery. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-1696.

Ye Zhu, Yu Wu, Nicu Sebe, and Yan Yan. Vision + x: A survey on multimodal learning in the light of data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9102–9122, 2024. doi: 10.1109/TPAMI.2024.3420239.

Xuechao Zou, Kai Li, Junliang Xing, Yu Zhang, Shiying Wang, Lei Jin, and Pin Tao. Diffcr: A fast conditional diffusion framework for cloud removal from optical satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024. doi: 10.1109/TGRS.2024.3365806.

A Appendix

A.1 Implementation Details and Data Collection

All experiments were implemented in Google Colab, with datasets, checkpoints and results stored in Google Drive. Experiments were run on NVIDIA A100 and L4 GPUs. For external evaluation, DiffCR was evaluated using the official author-released pretrained checkpoint without retraining. All experiments used the corresponding official implementations where available. All other baseline models were trained from scratch on our respective data split using the same training and evaluation protocol.

A.2 Evidence Degradation Results

The paper reports these experiments as reconstruction-fidelity curves, here we provide the complete numerical results across all evaluated metrics. Each condition is evaluated on the full test set.

Table 6: Detailed reconstruction performance under controlled SAR degradation. The optical observation is kept fixed while SAR evidence is progressively degraded.

SAR Deg. (%)	MSE ↓	MAE ↓	PSNR ↑	SSIM ↑	SAM ↓	UIQI ↑
0	0.001345	0.024505	29.5197	0.8450	7.2887°	0.2820
10	0.001381	0.024962	29.5281	0.8444	7.2893°	0.2803
30	0.002023	0.030623	27.9116	0.8192	9.1794°	0.2252
50	0.002839	0.037017	26.6948	0.7966	11.1752°	0.1937
70	0.003350	0.040319	25.9138	0.7836	12.6410°	0.1733
90	0.003695	0.042543	25.3998	0.7751	13.7590°	0.1604

Table 7: Detailed reconstruction performance under controlled optical cloud-obscuration severity.

Cloud Severity (%)	MSE ↓	MAE ↓	PSNR ↑	SSIM ↑	SAM ↓	UIQI ↑
10	0.001008	0.021096	30.9635	0.8603	6.6513°	0.4179
30	0.000972	0.021097	30.9066	0.8578	6.5507°	0.3881
50	0.001037	0.021702	30.6064	0.8546	6.5875°	0.3535
70	0.001139	0.022644	30.2116	0.8510	6.7709°	0.3236
90	0.001262	0.023778	29.7814	0.8472	7.0787°	0.2967

A.3 Reliability Estimation Results

The paper reports the overall correlation between predicted and empirical modality utility and the resulting reliability-weighted fusion performance. We provide the corresponding calibration, ranking and controlled degradation results below.

Table 8: Correlation between predicted modality reliability and empirical modality utility.

Modality	Pearson correlation ↑
Optical	0.617037
SAR	0.745409

Table 9: Accuracy of predicted modality ordering relative to empirical modality utility.

Metric	Accuracy ↑
Modality ranking accuracy	0.748326

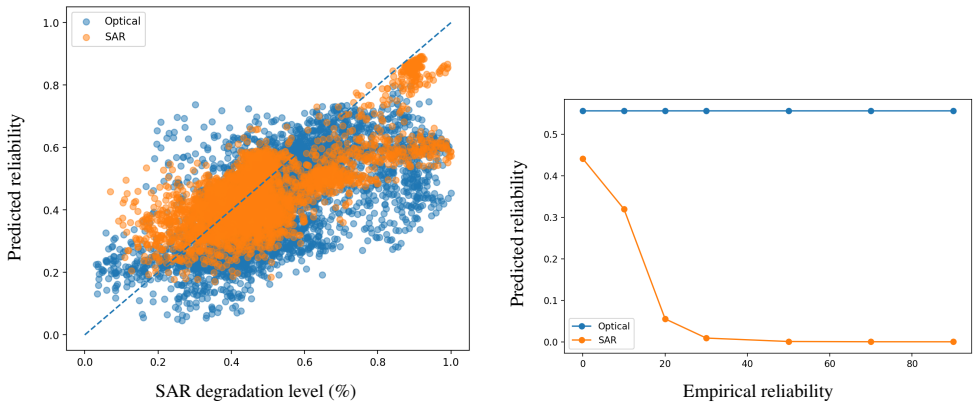

Figure 7: **Evidence reliability estimation.** Left: Predicted modality reliability tracks empirical utility across samples, with stronger correspondence for SAR. Right: Reliability responds selectively to controlled SAR degradation: SAR reliability decreases sharply while optical reliability remains comparatively stable. These results indicate that reliability is an observable property of the evidence, rather than a fixed modality attribute.

A.4 Representation Response to Evidence Perturbation

The paper reports the resulting trends and latent-shift correlations; the complete numerical results are provided here.

Table 10: Shared-representation response to controlled modality corruption. Latent shifts are measured relative to the unperturbed representation. Reliability values denote the corresponding mean predicted modality reliability.

Perturbation	Severity (%)	Latent L2 Shift	Cosine Shift	Optical Reliability	SAR Reliability
Optical noise	0	0.0000	0.0000	0.5559	0.4410
Optical noise	10	0.7707	0.2761	0.3718	0.4410
Optical noise	30	0.9754	0.4492	0.0870	0.4410
Optical noise	50	0.9843	0.4598	0.0321	0.4410
Optical noise	70	1.0168	0.4672	0.0190	0.4410
Optical noise	90	1.0376	0.4714	0.0144	0.4410
SAR noise	0	0.0000	0.0000	0.5559	0.4410
SAR noise	10	1.5675	0.4791	0.5559	0.3198
SAR noise	30	1.0657	0.1593	0.5559	0.0091
SAR noise	50	1.2010	0.1593	0.5559	0.0008
SAR noise	70	1.2196	0.1603	0.5559	0.0002
SAR noise	90	1.2244	0.1605	0.5559	0.0001
SAR sample mismatch	100	0.1962	0.0370	0.5559	0.4410

Table 11: Correlation between perturbation severity and shared-representation displacement.

Perturbation	Severity–Shift Correlation ↑
Optical noise	0.7324
SAR noise	0.4433

A.5 Shared Representation Analysis

Section-3.4 analyzes whether the shared representation preserves common multimodal structure while remaining sensitive to the prevailing evidence state. We provide the complete representation-similarity, canonical correlation and evidence-state geometry results below.

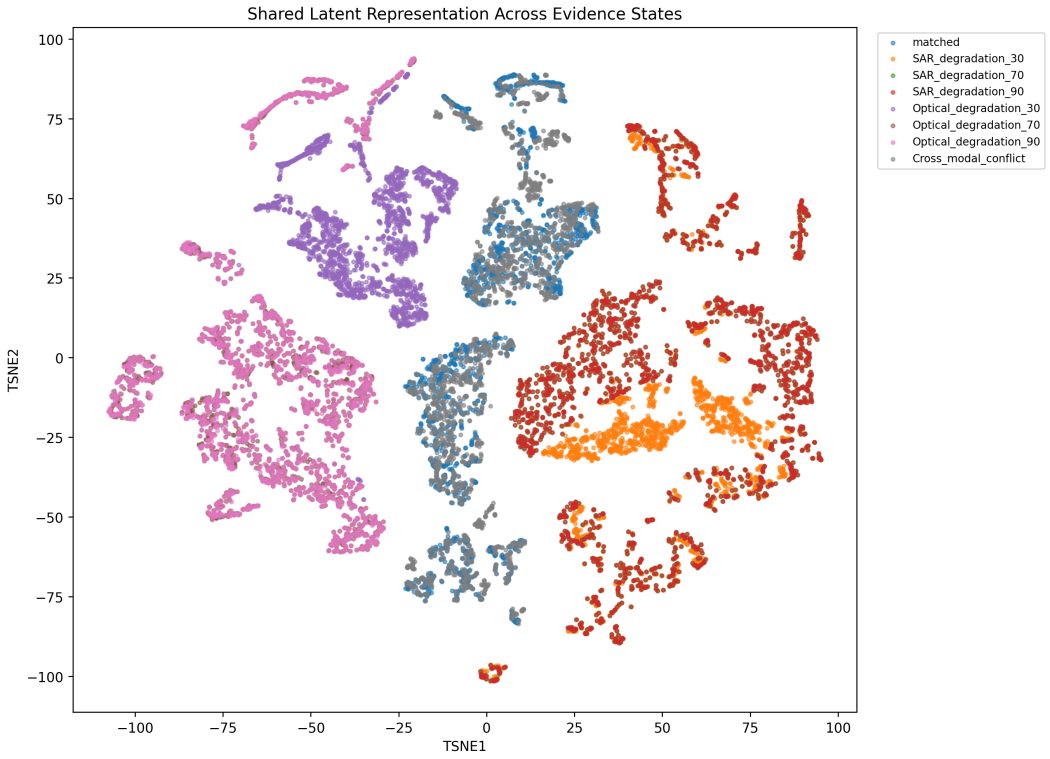

Figure 8: t-SNE visualization of the shared latent representation across matched, modality-degraded, and cross-modal-conflict evidence states. The overlapping geometry indicates that evidence states remain embedded within a common multimodal representation rather than forming completely separated clusters.

Table 12: Canonical correlations between modality-specific and shared representations.

Component	Canonical Correlation ↑
1	0.9974
2	0.9922
3	0.9799
4	0.9785
5	0.9719
6	0.9605
7	0.9443
8	0.9434

Table 13: Linear CKA similarity between modality-specific and shared representations.

Representation Pair	CKA ↑
Optical–SAR	0.5759
Optical–Shared	0.9076
SAR–Shared	0.7527

Table 14: Evidence-state structure in the shared representation. PCA variance is computed over the shared representation, while grouped 7-NN evaluates evidence-state predictability using the PCA representation.

Measure	Value
Variance explained by first 50 PCs	97.6563%
Grouped 7-NN accuracy	76.6205%
Chance accuracy	12.5000%
Silhouette score	0.0120

A.6 Representation Uncertainty Analysis

Refer Section-3.5. We provide the complete correlation, quartile and degradation results below.

Table 15: Correlation between representation uncertainty measures and reconstruction error over the 4,934-sample test set.

Uncertainty Measure	Pearson $r \uparrow$	Pearson p	Spearman $\rho \uparrow$	Holm-adjusted p
Mean latent variance	0.2990	1.85×10^{-102}	0.3341	1.23×10^{-128}
Median latent variance	0.4253	5.91×10^{-216}	0.4767	7.87×10^{-278}
Latent variance L2	0.2179	3.97×10^{-54}	0.2615	5.69×10^{-78}
Mean latent standard deviation	0.3970	6.32×10^{-186}	0.4242	3.10×10^{-214}

Table 16: Reconstruction error across representation-uncertainty quartiles.

Uncertainty Quartile	Mean Latent Variance	Mean Reconstruction Error	Median Reconstruction Error
Q1 (low)	0.0000721	0.021648	0.020453
Q2	0.0001013	0.026021	0.025876
Q3	0.0001208	0.026888	0.026720
Q4 (high)	0.0001579	0.027031	0.026766

Table 17: Representation uncertainty under controlled SAR degradation.

SAR Deg. (%)	Mean Latent Uncertainty	Median	Std. Dev.
0	0.000129	0.000126	0.000041
10	0.000148	0.000134	0.000071
20	0.000156	0.000148	0.000055
30	0.000129	0.000127	0.000036
50	0.000129	0.000128	0.000039
70	0.000130	0.000128	0.000039
90	0.000130	0.000129	0.000038

A.7 External Validation

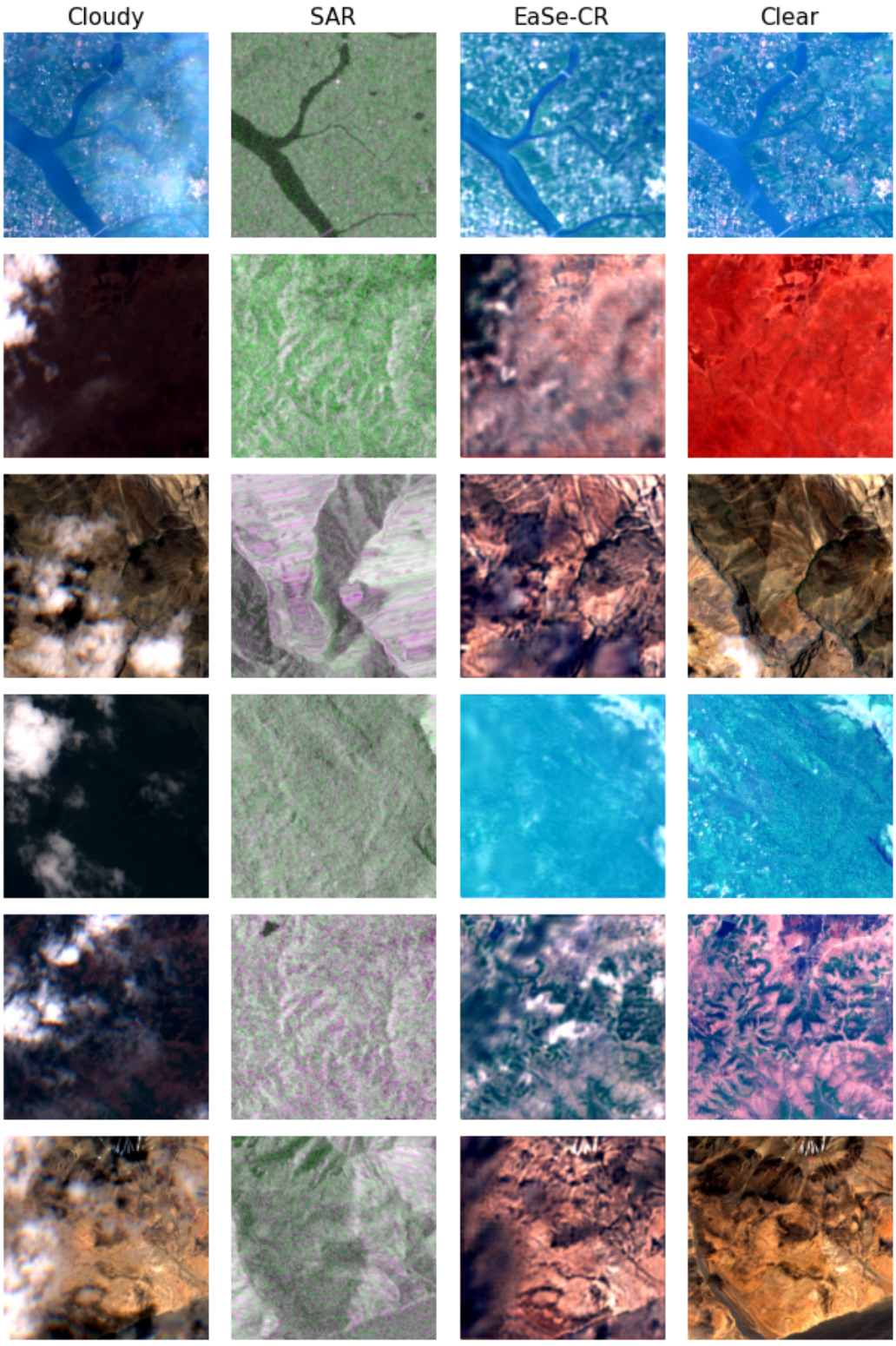

Figure 9: Qualitative cloud reconstruction across geographically diverse regions.

A.8 AI POLICY

AI tools were used solely as research-assistance tools for grammatical and formal polishing, literature retrieval and organization and limited portions of evaluation and debugging. All research decisions, methodological development, experimentation, analysis and conclusions were performed and verified by the authors, AI was used to assist these processes, not to replace them.