

Deep Learning Advancements for Urban Change Detection Using Multitemporal SAR Images from Heterogeneous Sensors

Luigi Russo^a, Lili Zilu^b, Silvia Liberata Ullo^c and Paolo Gamba^a

^aEngineering Department, University of Pavia, Italy, ^bAIRCAS, China, ^cEngineering Department, University of Sannio, Italy

Abstract

- As a follow-up to [1], we apply our **SAR-to-SAR framework** to Sentinel-1 (S1) and COSMO-SkyMed (CSK) data over selected cities from the OSCD dataset [2], aligning SAR acquisitions to the **spatial extent** of the original **optical image pairs**. A subset of **change labels** is leveraged to evaluate performance on **heterogeneous SAR-based change detection (CD)**.
- Notably, using **synthetic SAR pairs** improve CD performance over direct cross-sensor comparison.
- Future work** includes refining the translation process and applying it to **longer temporal sequences** for enhanced **urban change modelling**.

Introduction

- Urbanization** is a key driver of **global change**, requiring accurate **observation** for effective **urban planning** and **environmental monitoring**.
- SAR** enables all-weather, day-night observation, but observation through **heterogeneous SAR samples** poses challenges.
- This work applies a **SAR-to-SAR Deep Learning (DL)** framework [1] to improve **CD** using **synthetically aligned SAR image pairs**, simulating the scenario in which co-temporal pre/post data from the same sensor are unavailable.

Objective

- Evaluate** the effectiveness of the **SAR-to-SAR DL framework** [1] for **CD** in **conditions** in which **pre** and **post event SAR** image pairs are acquired from **different sensors** with **heterogeneous characteristics**.
- Test areas:** Rio de Janeiro and Hong Kong pre-post image pairs (see Table 1).
- The final goal is to assess whether **synthetic SAR pairs**, generated when **co-temporal acquisitions** from a **single sensor** are unavailable, can improve detection of **urban changes** such as **construction or demolition**.

Methods

1. Phase 1: SAR-to-SAR Translation

- Objective:** Learn mappings between two SAR domains ($X \rightarrow Y$ and $Y \rightarrow X$) using paired acquisitions.

- Network:** Two generators $G: X \rightarrow Y$, $F: Y \rightarrow X$ and two discriminators D_X, D_Y .

- Loss Functions:**
 - Adversarial Loss** (for G and F):
 $L_{GAN}(G, D_Y) = \mathbb{E}_Y[\log D_Y(Y)] + \mathbb{E}_X[1 - \log D_Y(G(X))]$
 - Cycle-Consistency Loss**, ensuring round-trip fidelity:
 $L_{cyc} = \mathbb{E}_X[\|F(G(X)) - X\|_1] + \mathbb{E}_Y[\|G(F(Y)) - Y\|_1]$
 - Identity Loss**: enforces colour and style consistency:
 $L_{id} = \mathbb{E}_Y[\|G(Y) - Y\|_1] + \mathbb{E}_X[\|F(X) - X\|_1]$
 - Total Loss:**
 $L_{total} = L_{GAN} + \lambda_{cyc} L_{cyc} + \gamma L_{id}$

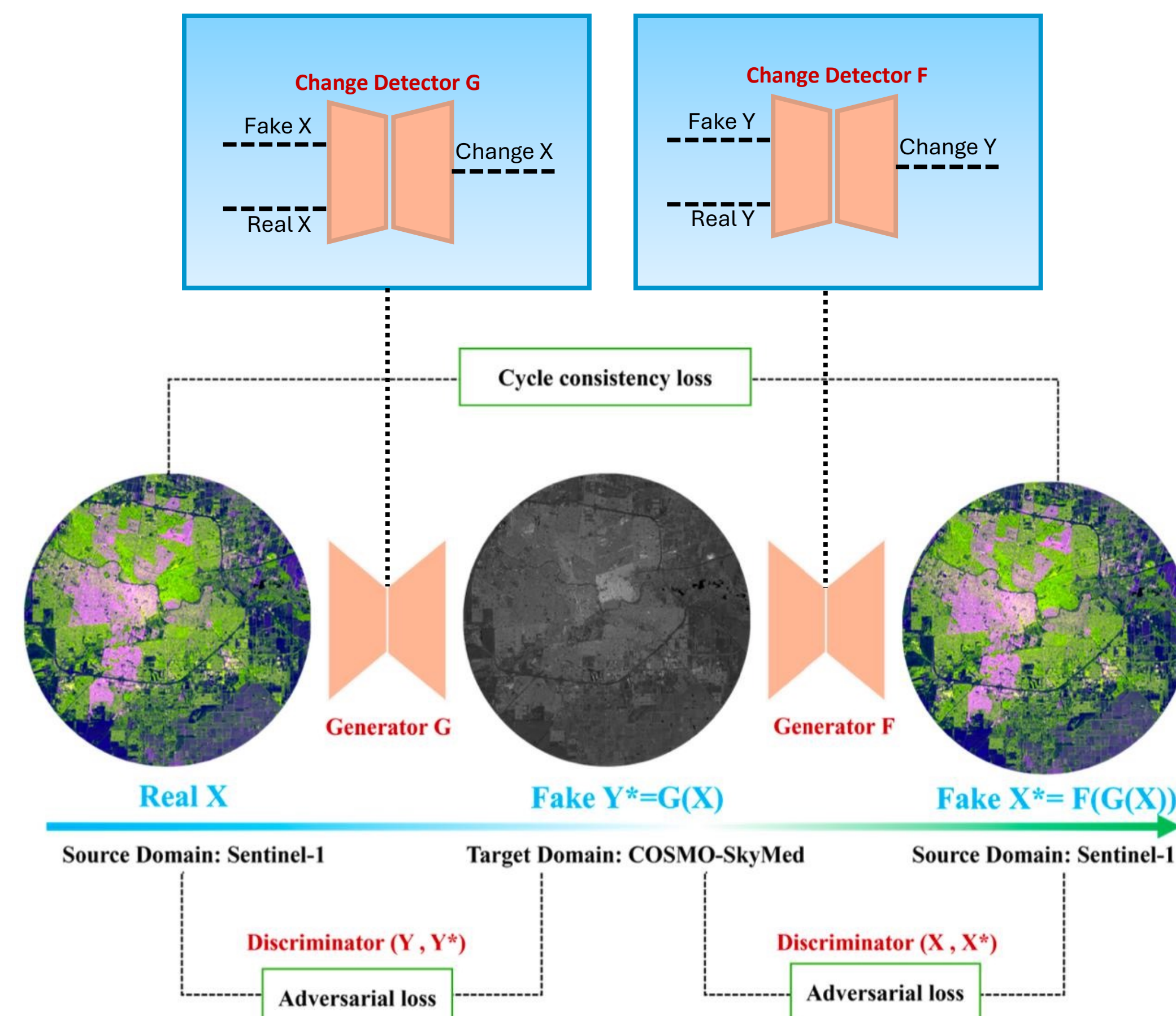
2. Phase 2: Change Detection

- Goal:** detect changes by comparing real and translated SAR pairs (e.g., x real and $G(x)$ synthetic).
- Architecture:** Siamese U-Net to extract features from image pairs and classify pixel-wise changes.
- Loss Function:** a combination of **Dice Loss** ($\mathcal{L}_{Dice}$) and **Cross-Entropy Loss** ($\mathcal{L}_{CE}$).

- Dice Loss:**
 $\mathcal{L}_{Dice} = 1 - \frac{2 * \sum p * g + \epsilon}{\sum p + \sum g + \epsilon}$, p : predicted probs.; g : targets

- Cross-Entropy Loss:**
 $\mathcal{L}_{CE} = -\frac{1}{N} \sum g * \log(p) + (1 - g) * \log(1 - p)$

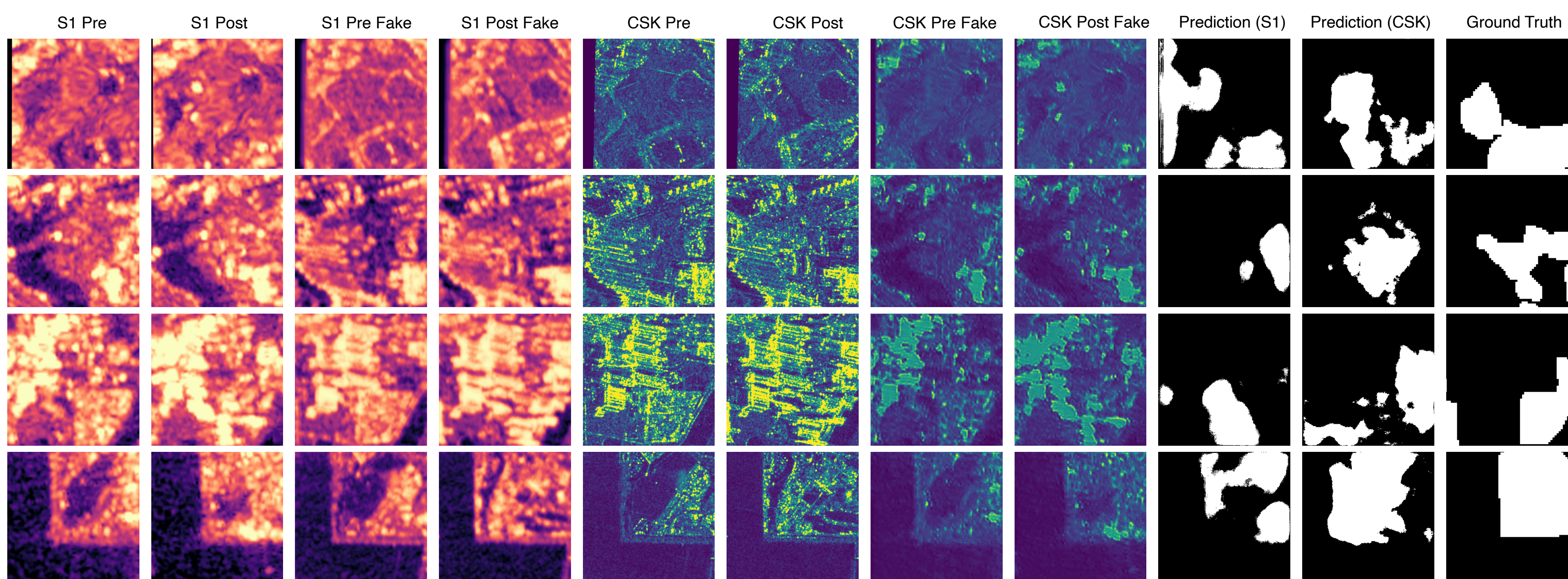
- Total Change Detection Loss:**
 $\mathcal{L}_{Total} = \mathcal{L}_{Dice} + \mathcal{L}_{CE}$



Overview of the proposed framework. At the bottom, a CycleGAN-based SAR-to-SAR translation module learns mappings between Sentinel-1 (VV) and COSMO-SkyMed (HH) domains using adversarial and cycle-consistency losses. At the top, a Change Detection (CD) module takes real/synthetic SAR pairs as input (e.g., Real X and Fake X) and predicts change masks.

Table 1: SAR acquisition details for Rio de Janeiro and Hong Kong.

City	Sensor	Acquisition Type	Date	Polarization	Imaging Mode
Rio de Janeiro	COSMO-SkyMed	Pre-event	2016-04-18	HH	Stripmap Image
	Sentinel-1	Pre-event	2016-04-20	VV	Interferometric Wide Swath
	COSMO-SkyMed	Post-event	2017-10-14	HH	Stripmap Image
	Sentinel-1	Post-event	2017-10-06	VV	Interferometric Wide Swath
Hong Kong	COSMO-SkyMed	Pre-event	2016-08-29	HH	Stripmap Image
	Sentinel-1	Pre-event	2016-09-25	VV	Interferometric Wide Swath
	COSMO-SkyMed	Post-event	2018-10-22	HH	Stripmap Image
	Sentinel-1	Post-event	2018-03-19	VV	Interferometric Wide Swath



Qualitative results of the SAR-to-SAR and Change Detection pipeline. From left to right: real and synthetic (fake) pre & post-event SAR patches for Sentinel-1 (S1) and COSMO-SkyMed (CSK), followed by change predictions and GT masks.

Table 2: Similarity metrics for real and SAR2SAR-generated image pairs.

Comparison	Metric	Value
Real CSK vs Real S1	MSE ↓	0.1156
	PSNR ↑	10.78
	SSIM ↑	0.161
Real CSK vs Fake CSK	MSE ↓	0.0285
	PSNR ↑	16.90
	SSIM ↑	0.256
	SSIM ↑	0.256
Real S1 vs Fake S1	MSE ↓	0.0377
	PSNR ↑	14.71
	SSIM ↑	0.378

Table 3: Change detection performance using S1 and CSK inputs. Hybrid configurations leverage SAR-to-SAR translation. The top section shows five-fold cross-validation with mixed training samples from Hong Kong and Rio de Janeiro. The bottom section presents the out-of-distribution (OOD) test on Hong Kong, with training performed on Rio de Janeiro only. Percentage deltas indicate improvement (+) or drop (−) of hybrid setups compared to the worst-case scenario (i.e., comparing pre/post images from heterogeneous SAR sensors).

Setting	Sensor	Configuration	IoU	F1-Score	Recall
5-fold cross-validation	CSK	Ideal (CSK Pre vs CSK Post)	0.365 ± 0.049	0.533 ± 0.052	0.605 ± 0.116
		Hybrid (CSK Pre Fake vs CSK Post)	0.086 ± 0.027 (+69%)	0.158 ± 0.044 (+63%)	0.692 ± 0.055 (+5%)
		Worst-case (S1 Pre vs CSK Post)	0.051 ± 0.016	0.097 ± 0.029	0.727 ± 0.125
	S1	Ideal (S1 Pre vs S1 Post)	0.218 ± 0.033	0.357 ± 0.043	0.403 ± 0.086
		Hybrid (S1 Pre Fake vs S1 Post)	0.088 ± 0.021 (+35%)	0.161 ± 0.036 (+34%)	0.569 ± 0.070 (+14%)
		Worst-case (CSK Pre vs S1 Post)	0.065 ± 0.020	0.121 ± 0.036	0.659 ± 0.123
OOD (Hong Kong)	CSK	Ideal (CSK Pre vs CSK Post)	0.186	0.313	0.450
		Hybrid (CSK Pre Fake vs CSK Post)	0.070 (+37%)	0.123 (+27%)	0.675 (+22%)
		Worst-case (S1 Pre vs CSK Post)	0.051	0.097	0.866
	S1	Ideal (S1 Pre vs S1 Post)	0.144	0.251	0.341
		Hybrid (S1 Pre Fake vs S1 Post)	0.060 (+30%)	0.111 (+26%)	0.583 (+29%)
		Worst-case (CSK Pre vs S1 Post)	0.046	0.088	0.816

Results

- SSIM, MSE, and PSNR values from Table 2 show that translated SAR images are much closer to target domain samples than direct cross-sensor comparisons.
- This highlights the **SAR-to-SAR translation's** ability to **reduce domain gaps** and improve alignment for subsequent downstream tasks.
- CSK→S1** translation performs best, likely due to CSK's finer resolution, which facilitates domain synthesis.
- Five-fold cross-validation** (Table 3, top): the **hybrid setup** (real-synthetic pairs) improves over direct heterogeneous comparison (worst case), with **+35–69% IoU**, **+33–63% F1**, and slightly lower recall due to fewer false positives.
- Out-of-distribution** (Hong Kong, Table 3, bottom): even under domain shift, hybrid input yields **+30–37% IoU** and **~27% F1** over the worst-case baseline.

Discussion

- In this work, we employed a **CycleGAN-based SAR-to-SAR framework** [1] to enable translation between **S1 VV** and **CSK HH** domains, facilitating **cross-polarization synthesis** when **co-temporal acquisitions** are not available.
- This translation facilitates the generation of **aligned SAR pairs** across sensors, preserving **structural content** through **cycle-consistency** and **identity constraints**.
- In the **CD phase**, we evaluate **hybrid scenarios** (real/synthetic pairs) and show that **translated images** improve the **detection of changes** over direct comparison of **heterogeneous SAR acquisitions** from different sensors.

Conclusions

- As previously demonstrated in [1], **CycleGAN-based SAR-to-SAR translation** improves the alignment between Sentinel-1 (VV) and COSMO-SkyMed (HH) domains, reducing general domain gap.
- In the **CD** phase, hybrid real/synthetic pairs outperform direct heterogeneous comparisons in both **cross-validation** and **out-of-distribution** case (Table 3).
- A key limitation is the **resolution mismatch** (CSK: 2.5 m vs. S1: 10 m), which affects **change localization** and reduces **performance** compared to **co-sensor pre/post comparisons**, especially with **limited training data**.
- Future work** includes extending the framework to **longer SAR time series** and exploring **advanced DL models** (e.g., diffusion models) to better handle temporal gaps and improve robustness of CD under varying acquisition conditions.

Major References

- [1] Russo, L., Ullo, S. L., and Gamba, P. (2024). *Changes Detection Within Urban Areas Using Multi-Temporal SAR Images from Heterogeneous Sensors*. Presented at the ESA-MOST Dragon 5 Final Symposium, Lisbon, Portugal.
- [2] R. C. Daudt, B. Le Saux, A. Boulch and Y. Gousseau, "Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018, pp. 2115-2118, doi: 10.1109/IGARSS.2018.8518015.

Acknowledgement

This work was conducted as part of the Dragon 6 cooperation programme between the European Space Agency (ESA) and the Ministry of Science and Technology (MOST) of China. The authors acknowledge the support and coordination provided by the Dragon Programme Office.