L. Beltrame[1,4] // J. Salzinger[1] // F. Svoboda[2] // P. Fanta-Jende[1] // J. Lampert[1] // Pascal Leon Thiele[1] // R. Timofte[3] // M. Körner[4,5,6]

[1] Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria.
[2] University of Cambridge, William Gates Building, 15 JJ Thomson Ave., CB3 0FD Cambridge, UK.
[3] University of Würzburg, John-Skilton-Str. 4a, Hubland Nord, 97074 Würzburg, Germany.
[4] Technical University of Munich, TUM School of Engineering and Design, Chair of Remote Sensing Technology, Arcisstraße 21, 80333 Munich, Germany.
[5] Technical University of Munich, Munich Data Science Institute, 85748 Garching, Germany.
[6] ELLIS Unit Jena, Friedrich Schiller University of Jena, 07743 Jena, Germany.

GEOMETRY- AND PHYSICS-AWARE DATASET CREATION FOR SHADOW REMOVAL IN HIGH-RESOLUTION SATELLITE IMAGERY

ENSURE MONITORING IN ADVERSE CONDITIONS

Shadows are a major limitation in satellite imagery analysis. They degrade tasks such as classification, detection, and 3D reconstruction, especially in urban an mountainous regions.

However, **no public dataset provides geometry-consistent paired shadowed and shadow-free satelli images**, making supervised learning infeasible.

Existing datasets:

- Focus on **shadow detection**, not removal
- Lack **true shadow-free references**
- Ignore **multi-view geometry and seasonal variation**

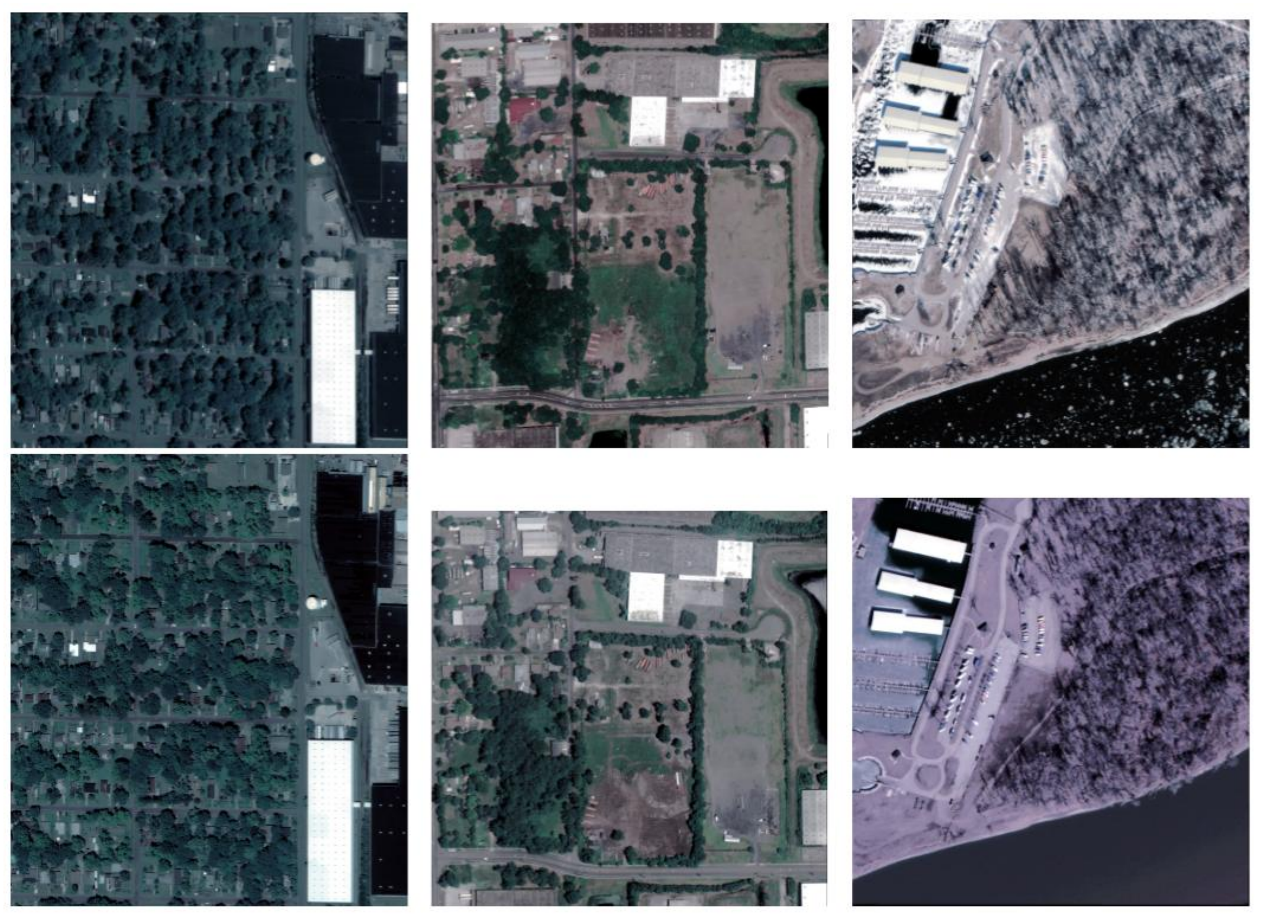
Our goal:

Transform existing Earth observation datasets into **paired training data for shadow removal**. We use S-EO [1] as a demonstration.

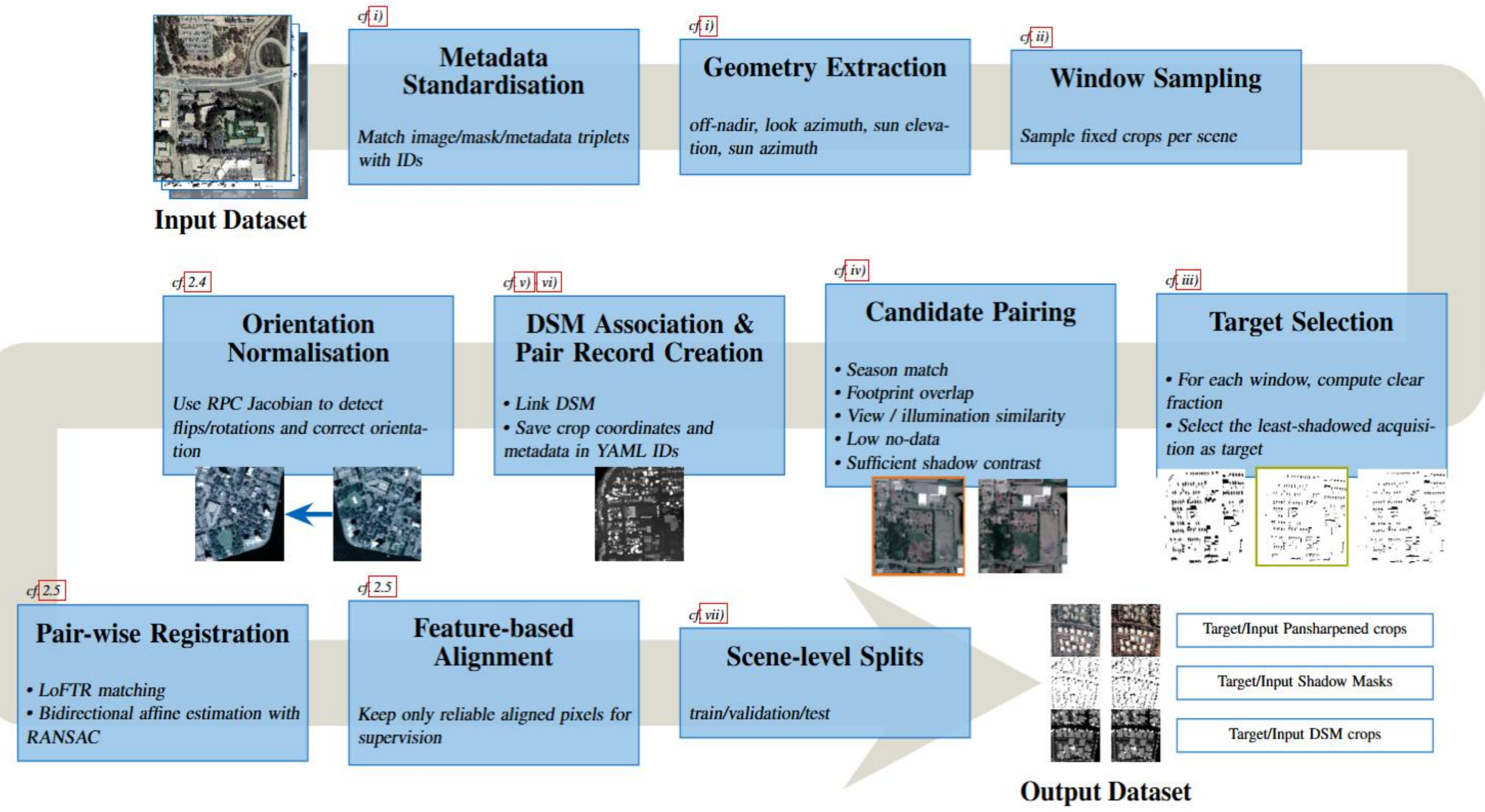
QUALITY CONTROL

Pairs are filtered by footprint overlap, camera/solar geometry, missing data, seasonal gap, and minimum shadow contrast.

Rejected pairs include excessive *azimuth difference* (a), *off-nadir mismatch* (b), *and seasonal change* (c).



(a) (b) (c)



deSEO pipeline [4]

PIPELINE DESCRIPTION

i) Match image–mask–metadata IDs and extract acquisition geometry: off-nadir, look azimuth, sun elevation, and sun azimuth.

ii) Sample fixed crop windows to define where input–target pairs are created.

iii) Select the least-shadowed acquisition as the weak shadow-free reference for each crop.

iv) Filter candidate inputs using season, footprint overlap, viewing geometry, no-data, and shadow-contrast constraints.

v-vi) Link DSM priors and save crop coordinates, metadata, and pair IDs for reproducible training.

2.4) Correct orientation using RPC Jacobian information to detect flips and rotations.

2.5) Register pairs with LoFTR–RANSAC [2,3] to align the shadowed input with the weak reference target.

vii) Use only reliably aligned pixels for supervision and apply **scene-level train/validation/test splits** to avoid spatial leakage.

How to test the pipeline?

A **conditional GAN** with a residual **U-Net generator** and local **PatchGAN** discriminator performs satellite-specific deshadowing:

- **RGB + DSM input**
- **Residual U-Net** predicts shadow-reduced RGB
- **PatchGAN + shadow attention** focuses on shadow-transition regions
- **Validity masks** restrict losses to well-aligned pixels
- **Perceptual + colour losses** reduce blur and colour drift

Ablation Study on Validation Set

Mean +/- standard deviation. Metrics computed on RGB channels.

Configuration	PSNR up	RMSE down	SSIM up	L1 down	VGG19 L1 down	LPIPS down
Baseline	18 +/- 2	32 +/- 5	**0.6 +/- 0.1**	**0.09 +/- 0.01**	**0.44 +/- 0.04**	**0.41 +/- 0.07**
RGB only (remove DSM)	9 +/- 2	90 +/- 20	0.18 +/- 0.06	0.30 +/- 0.05	0.57 +/- 0.06	0.71 +/- 0.02
No G self-attention	17 +/- 2	37 +/- 8	0.5 +/- 0.1	0.10 +/- 0.02	0.47 +/- 0.05	0.42 +/- 0.07
Gamma features disabled	17 +/- 2	35 +/- 9	0.6 +/- 0.1	0.10 +/- 0.03	0.46 +/- 0.04	0.42 +/- 0.06
No HV regularisation	16 +/- 2	39 +/- 9	0.4 +/- 0.1	0.11 +/- 0.03	0.46 +/- 0.05	0.45 +/- 0.08
L1-only (no perceptual)	**18 +/- 1**	**31 +/- 5**	**0.6 +/- 0.1**	**0.09 +/- 0.01**	0.55 +/- 0.05	0.47 +/- 0.04
Reduced latent channels	17 +/- 2	35 +/- 7	0.5 +/- 0.1	0.10 +/- 0.01	0.46 +/- 0.04	0.44 +/- 0.06
No spectral normalisation	16 +/- 2	39 +/- 7	0.5 +/- 0.1	0.12 +/- 0.02	0.49 +/- 0.05	0.46 +/- 0.07
Shadow attention disabled	**18 +/- 1**	34 +/- 3	**0.6 +/- 0.1**	0.10 +/- 0.01	0.45 +/- 0.05	0.42 +/- 0.06
No Pretraining	17 +/- 2	36 +/- 8	0.5 +/- 0.1	0.10 +/- 0.01	0.46 +/- 0.05	0.42 +/- 0.06
No dropout	17 +/- 2	37 +/- 7	0.6 +/- 0.1	0.10 +/- 0.02	0.47 +/- 0.05	0.42 +/- 0.06

Bold values indicate the best reported configuration for each metric.

Key ablation results: DSM conditioning and perceptual supervision improve visually coherent shadow reduction; L1-only training tends to preserve shadows.

(a) Input (b) Target (c) Full model (d) L_1 only

Link to the full paper:

References:
[1] Masquil, Elías, et al. "S-EO: A Large-Scale Dataset for Geometry-Aware Shadow Detection in Remote Sensing Applications." Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.
[2] Sun, Jiaming, et al. "LoFTR: Detector-free local feature matching with transformers." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
[3] Fischler, Martin A., and Robert C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." Communications of the ACM 24.6 (1981): 381-395.
[4] L. Beltrame et al., "deSEO: Physic-Aware Dataset Creation for High Resolution Satellite Images Shadow Removal," in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2026.