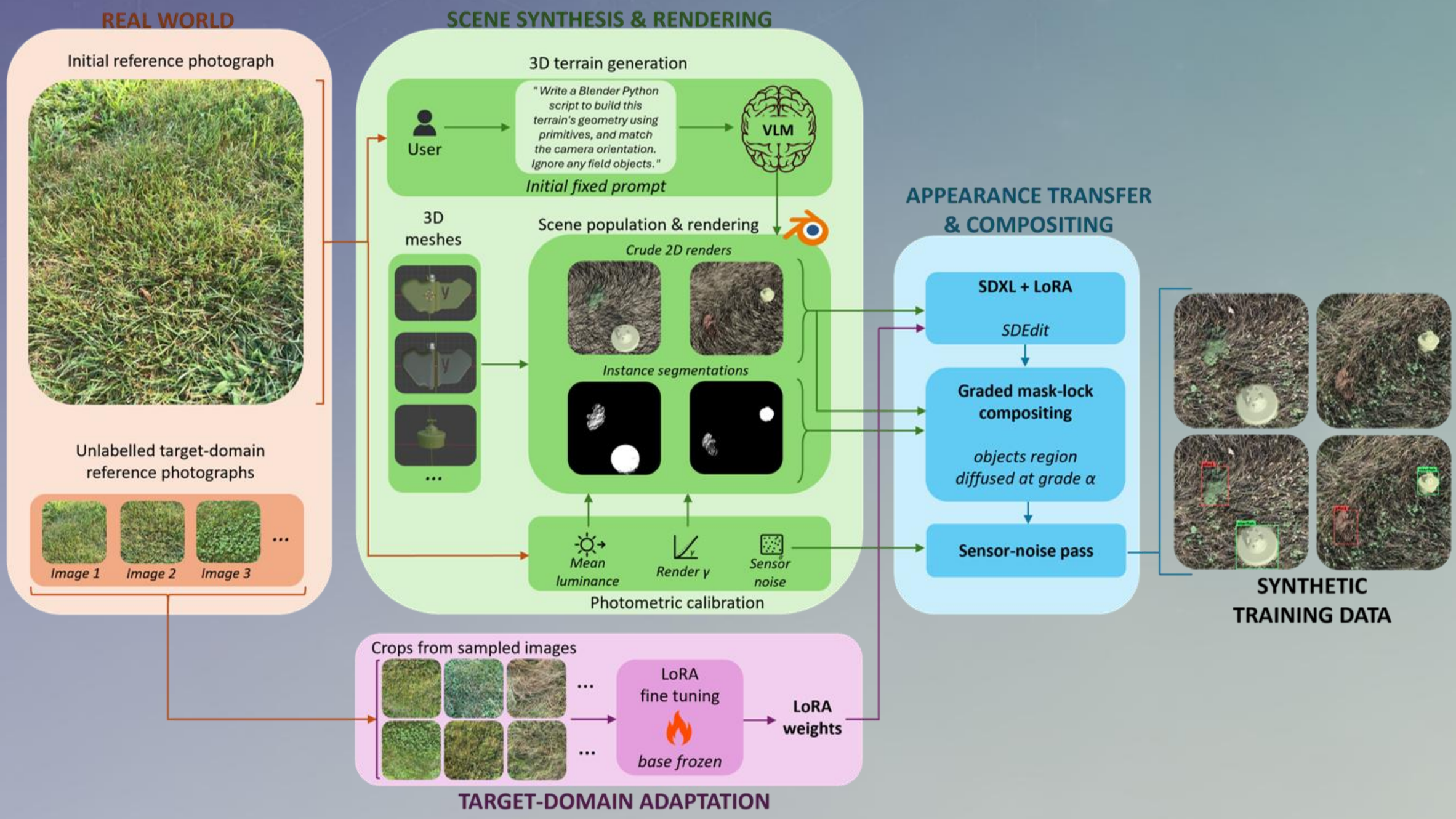
Diffuse the object, keep its label: curating detector training data from a few unlabeled photographs via VLM-built 3D vegetation scenes

Mario Malizia[1,2] · Marnix Enting[1] · Rob Haelterman[1] · Ken Hasselmann[1]

Can a few unlabeled target-site photographs become detector training data, without manual annotation?

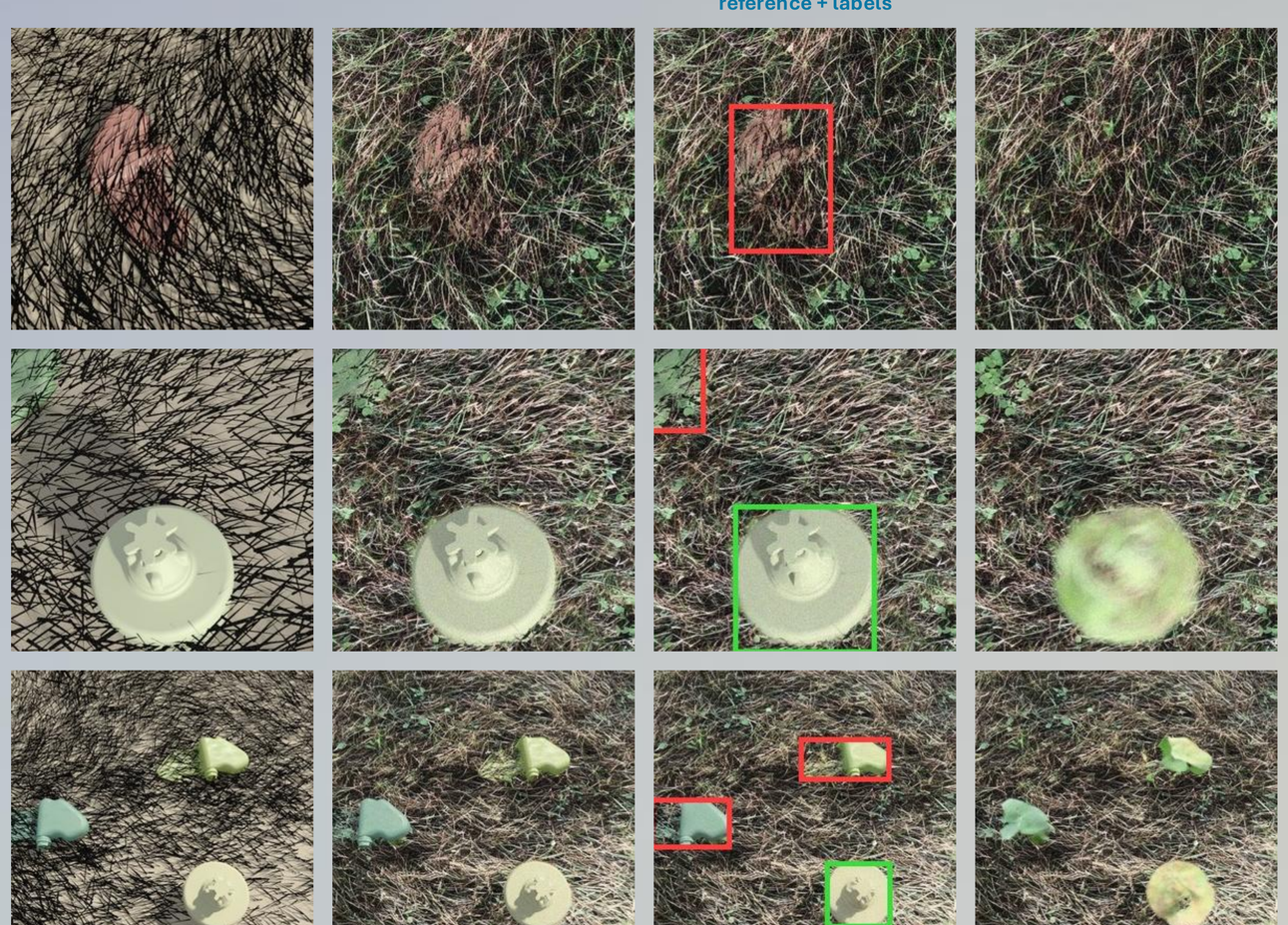
We build site-specific 3D vegetation scenes, derive labels by construction, and use diffusion-based curation to reduce the synthetic-to-real appearance gap.



Qualitative behavior across target domains

Same scene and diffusion seed; only the mask-lock grade α changes.

SULAND - grassland



Rows illustrate occluded PFM-1, PMA-2 close-up, and a multi-landmine scene.

Key quantitative results

0.584 mF1

graded mask-lock, α = 0.10

+0.191 mF1

over the real labeled cross-site baseline

0.393 mF1 · real labeled baseline

Cross-site generalization on SULAND

ITA / synthetic to USA split · five seeds · larger is better

Configuration	mF1	R PFM-1	R PMA-2	mAP50
Real labeled (22.7k)	0.393 ± 0.021	0.377	0.217	0.327 ± 0.015
Raw CG, no diffusion	0.194 ± 0.111	0.208	0.185	0.136 ± 0.095
Hard lock, α = 0	0.441 ± 0.051	**0.579**	0.196	0.346 ± 0.037
Graded, α = 0.05	0.503 ± 0.060	0.571	0.285	0.401 ± 0.049
Graded, α = 0.10	**0.584 ± 0.025**	**0.577**	**0.418**	**0.481 ± 0.024**
Unlocked, α = 1	0.466 ± 0.030	0.378	0.339	0.376 ± 0.034

How much data is enough?

Lowest-budget spot-checks at the reference lock α = 0.10

1 target photo	**0.562 ± 0.011 mF1**
2 patches / photo	**0.564 ± 0.023 mF1**
100 synthetic images	**0.514 ± 0.042 mF1**
Reference: 10 photos · 14 patches/photo · 1,000 images	**0.584 ± 0.025 mF1**

Quantitative evaluation was conducted on the SULAND USA split.

Key takeaways

- Graded diffusion at α = 0.10 gives the strongest cross-site generalization.
- Fully unlocking the object gives back much of the gain.
- Within the tested range, curation matters more than data quantity.

Selected references

[1] Vivoli et al., "Deep Learning-Based Real-Time Detection of Surface Landmines Using Optical Imaging," *Remote Sensing*, 2024.
[2] Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models," *ICLR*, 2022.
[3] Meng et al., "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations," *ICLR*, 2022.
[4] Podell et al., "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis," *ICLR*, 2024.

This work was supported by the **Belgian Defense** under **Grant DAP 23/08**.

[1] Royal Military Academy · [2] KU Leuven

Workshop on Curated Data for Efficient Learning (CDEL)

ECCV 2026 | MALMÖ | SEPT 8–12