

Insights from AutoResearch for Solar Panel Segmentation

AutoML4EO Workshop

Justinas Lekavičius Kürşat Kömürcü Valentas Gružasuskas Linas Petkevičius

Vilnius University, Institute of Computer Science
IRISA, Université Bretagne Sud
European Commission Joint Research Centre

AutoML4EO Workshop, 1 October 2026
AutoML 2026 · Ljubljana, Slovenia

AIML at Vilnius University

Artificial Intelligence Methods Lab



Trustworthy, reproducible AI methods

Explainable AI, multimodal learning
and agentic systems

Earth observation research

Climate monitoring and environmental change
Peatland moisture and urban heat
Satellite and in-situ data fusion

Supported by VU high-performance computing



Open to collaboration in AI for Earth observation

Joint research and EU proposals, shared datasets, and applied pilots

Valentas Gružas - valentas.gruzas@mif.vu.lt - aiml.lt

AutoML for Earth observation

Two proposed research projects



AWARE-EO

MSCA Postdoctoral Fellowship proposal

Automated data preparation and model development

Build validated multisensor data cubes
Select and adapt foundation models
Generate and test model architectures

Use cases

Peatland moisture, water quality
and solar-panel mapping

VERITAS

ERC Starting Grant proposal

Automated construction and verification of simulations

Build EO-grounded digital twins
Check physical and economic consistency
Benchmark reliability with VERITAS-Bench

Use cases

Peatlands and urban climate risk,
with food-risk stress tests

We are looking for research and application partners

EO data and ground truth, AutoML methods, verification, and joint pilots

Valentas Gružas - valentas.gruzas@mif.vu.lt | aiml.it

EO segmentation is experimentally sensitive

- Loss, schedule, optimiser and augmentation interact with class imbalance, GSD and GPU memory
- Published training configurations are sized for someone else's hardware and epoch budget
- Transferring them is not always effective, and manual search is slow

AutoResearch [Kar25]

- A coding LLM edits the training program, runs it under a fixed wall-clock budget, keeps the edit only if the metric improves
- Built for single-GPU LM pretraining (~ 300 s per idea); we port it to EO segmentation

Research question

Is AutoResearch a useful AutoML operator for EO segmentation, and do the changes it keeps **transfer across hardware**?

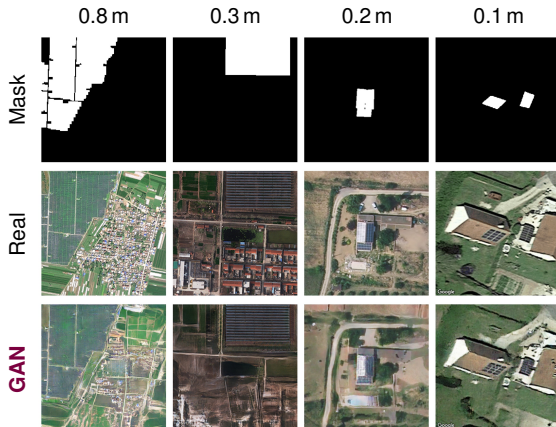
Contributions

- 1 Port AutoResearch from LM pretraining to binary PV segmentation, with an immutable evaluator and a fixed real-image split
- 2 Three independent 24-experiment campaigns on two GPU classes; compare the retained modifications
- 3 Score the best configuration once on the held-out test split, next to the published reference [LG24]

Selected data set and specifics

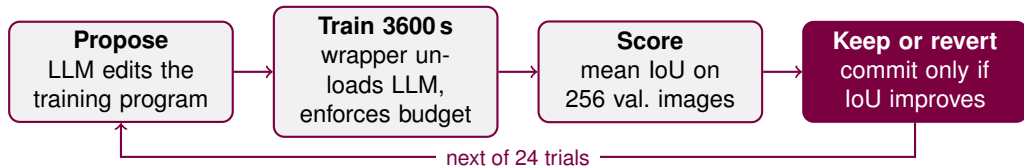


- PV panel testbed of [LG24]
- 640 image–mask pairs from each of four GSDs
- Resampled to 0.1 m, 512×512 crops
- Split **2048 / 256 / 256** (train / val / test), seed 35
- **Real images only:** the reference's 1228 pix2pix images (bottom row) are never loaded



Adapted from Fig. 3 of [LG24], CC BY. GAN row: used by the reference only.

The AutoResearch loop



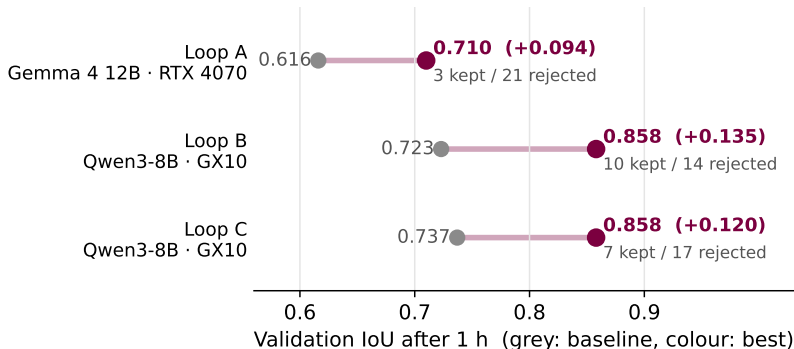
The agent may change: loss, optimiser, LR schedule, auxiliary head, gradient clipping, train-time augmentation

Frozen: DeepLabV3–ResNet-50 [CPS⁺17], data split, evaluator, time budget; failed runs = rejections

	Loop A	Loop B	Loop C
Coding LLM	Gemma 4 12B	Qwen3-8B	Qwen3-8B
GPU	RTX 4070 (12 GB)	GX10	GX10
Batch	2	8	8→16
Trials	24	24	24
Test scored	no	once	no

GX10 = ASUS Ascent GX10. Architecture, split, seed and 1 h budget shared.

Every campaign beats its one-hour baseline



- Most of the gain comes from the first one or two kept changes
- Later accepted deltas are $\sim 10^{-3}$ IoU: within single-run noise the protocol does not estimate

Kept changes do not transfer across hardware



Modification	Loop A (4070)	Loop B (GX10)	Loop C (GX10)
Soft Dice + BCE [MNA16]	rejected	kept +0.103	kept +0.076
Cosine schedule	kept +0.092	rejected	rejected
Polynomial LR 0.9	rejected	kept +0.012	—
AdamW vs. Adam	rejected	kept +0.005	—
Gradient clip 1.0	rejected (−0.359)	kept +0.002	—
Horizontal flip	rejected	kept +0.005	rejected, then kept
Lower peak LR	rejected	rejected	kept
Rotation / freeze / EMA	rejected	rejected	rejected

Deltas vs. the configuration each change was stacked on. — = not proposed.

B and C reach the same 0.858 by different paths: a one-hour plateau, not a universal recipe.

Held-out test: real images, one hour



Setup	Data	Budget	IoU	F ₁
Reference no_aug [LG24]	real	≤ 100 ep., A100	0.801	0.853
Reference basic_aug	real	≤ 100 ep., A100	0.813	0.865
Reference gan60 (best)	real + 60 % GAN	≤ 100 ep., A100	0.833	0.880
Loop B (ours)	real	3600 s, GX10	0.836	0.891

Same 256 test images as Table 5 of the reference. Scored once, on the best validation checkpoint.

- **+0.003 IoU over gan60**: a tie, reached without 1228 synthetic images
- Not compute-matched; val. $\rightarrow$ test gap ≈ 0.022 , so 0.858 is not a test number
- Specific to this split, backbone and GPU: not evidence that AutoResearch beats GAN augmentation in general

- 1 All three AutoResearch campaigns improve a one-hour DeepLabV3 baseline on a frozen real-image split
- 2 Retained modifications are hardware-specific: Dice on GX10, cosine decay under 12 GB
- 3 The Qwen3-8B configuration matches the reference test IoU (0.836 vs. 0.833) without synthetic images
- 4 The protocol works under a declared budget *if* the evaluator stays frozen and kept changes are not treated as universal defaults

Limits: one seed per trial; coding LLM confounded with GPU class; Loops A and C are validation-only.

Code: <https://github.com/VU-AIML/automl4eo-autoresearch-segmentation>

Thank you for your attention!

- [CPS⁺17] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [Kar25] Andrej Karpathy. Autoresearch.
<https://github.com/karpathy/autoresearch>, 2025.
- [LG24] Justinas Lekavičius and Valentas Gružas. Data augmentation with generative adversarial network for solar panel segmentation from remote sensing images. *Energies*, 17(13):3204, 2024.
- [MNA16] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, 2016.