Building an Operational Shallow Sentinel-2 Single Image Super-Resolution Network for EVOLAND Prototypes

Lessons Learned

Julien MICHEL, Ekaterina KALINICHEVA

CESBIO, Université de Toulouse, CNES/CNRS/INRAe/IRD/UPS, Toulouse, FRANCE

SUREDOS24 conference, Frascati, 29-31 may 2024

Super-Resolution in EVOLAND context

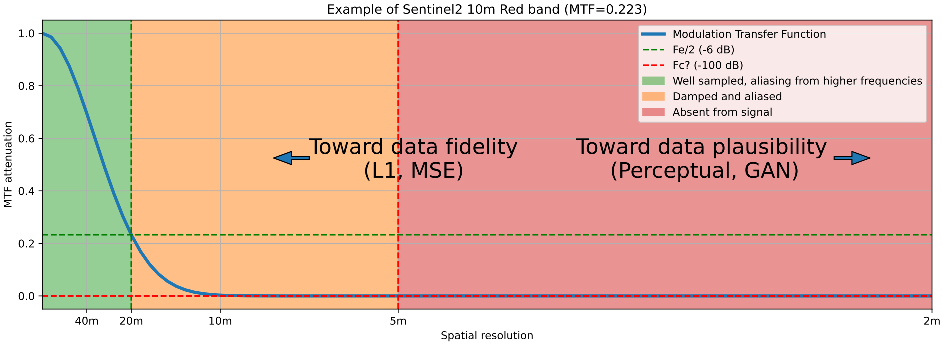
About EVOLAND (evo-land.eu)

- 3-years Horizon Europe Project
  - Coordinated by Vito
  - Partners: IIASA, CLS, UT3, SINERGISE, Joaneum Research, CNES, DLR, GAF and Eventflow
- Evolution of Copernicus Land Monitoring Service
  - 11 new candidate products for Copernicus Portfolio
  - Agriculture, water, forests, urban, general land cover
- Methods development (WP2) ⇒ Prototypes development (WP3)

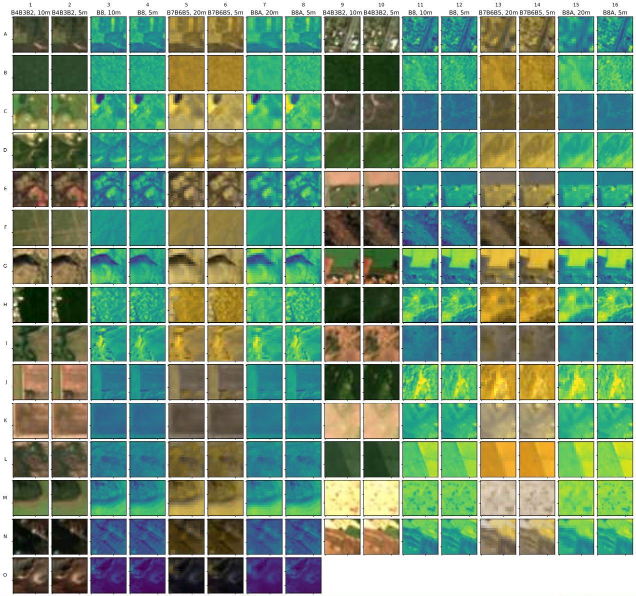
SISR in EVOLAND: why, how ?

- Task T2.5 : improved spatial/spectral/temporal resolution led by UT3 (CESBIO)
  - Task T2.5.1 : Improved resolution – Single Image Super-Resolution applied to Sentinel-2
- To be used by CLMS prototypes :
  - Process all useful bands
  - Radiometric Faithfulness is critical
  - Inference time is critical
  - Super-resolved image size vs. upscaling factor

Super-resolution or super-restoration ?



The sen2venµs dataset - extended



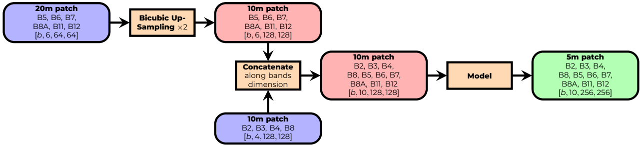
Description [1]

- Open Dataset on Zenodo: `https://zenodo.org/record/6514159`
- 133k cloud-free patches, 29 locations
- B2, B3, B4, B5, B6, B7, B8, B8A
- from 10m/20m (S2) to 5m (Venµs)
- L2A products from theia.cnes.fr

Extension

- 20m B11 and B12 patches
- Also L2A from theia.cnes.fr
- No 5m reference !

Models and architecture

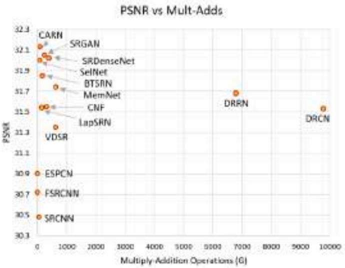


Models and parameters

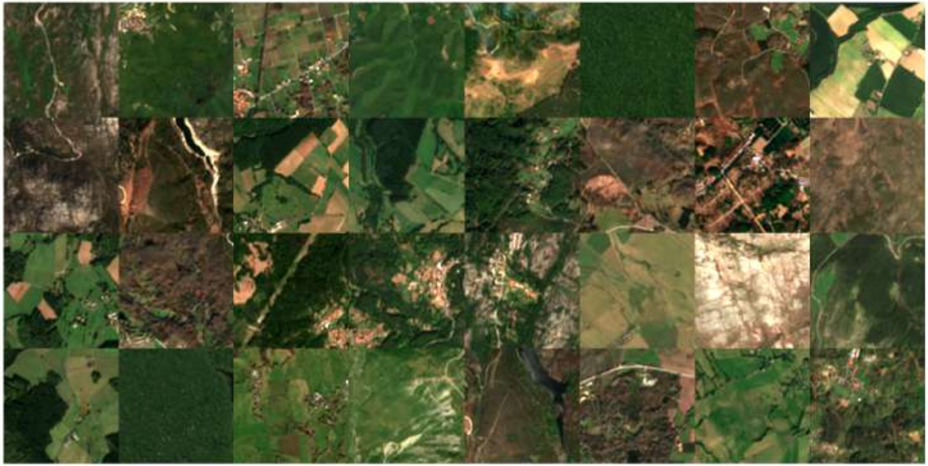
- CARN [2] + ESRGAN (RRDBNet) [3]
- Shallow: only 3 to 4 residual blocks

Model	#params	Recept. Field
`carn_1x3x64g4sw`	1 243 530	29
`carn_2x3x64g4sw`	1 867 914	17
`carn_3x3x64g4sw`	2 492 298	41
`carn_4x3x64g4sw`	3 116 682	53
`carn_3x3x64g4`	4 659 594	41
`carn_3x3x64g1sw`	5 146 506	41
`esrgan_3x64x32`	2 280 662	36

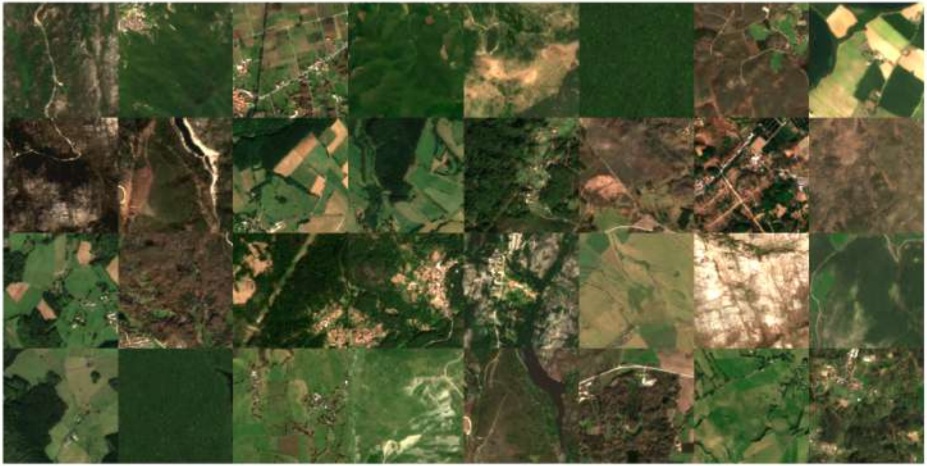
Complexity vs. perf. [4]



Discrepancies in the dataset

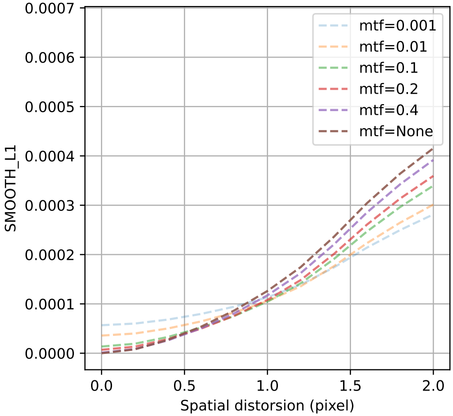


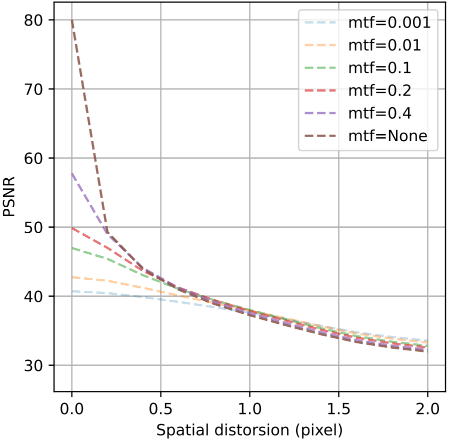
Discrepancies in the dataset



Impact on training and evaluation - spatial registration

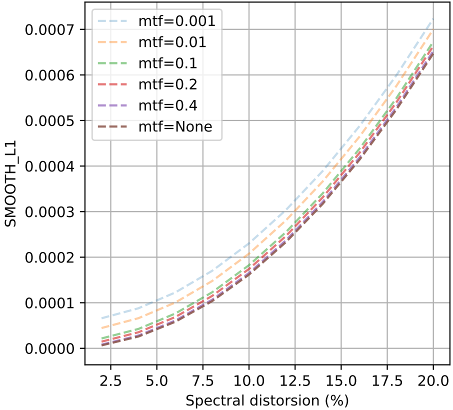
- L_1 loss will drive models towards warping of the input
- At 1 pix. of registration error, L_1 loss no longer favors sharpness
- at 1 pix. registration error, PSNR can not distinguish sharp images from blurry ones

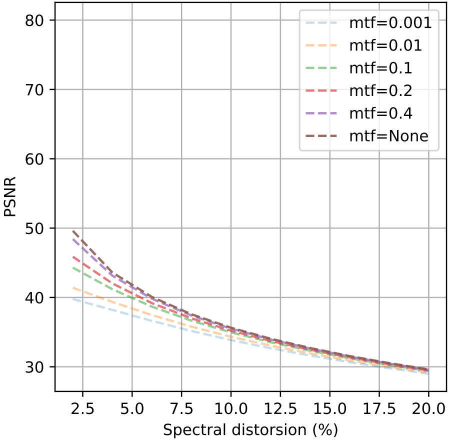




Impact on training and evaluation - spectral consistency

- L_1 loss will drive models towards spectral distorsion
- If there is spectral distorsion, PSNR can not distinguish sharp images from blurry ones





Pre-training stage (20 epochs): use simulated data

- Simulate Sentinel-2 patches from Venµs

$$P^{input}_{sim}(n,i\ldots) = \begin{cases} \left(P^{target}(n,i\ldots) \circledast \phi_{\sigma^{spat}(2,m)}\right)\downarrow_2 + \epsilon_i \sim \mathcal{N}(0,\sigma_i^{spect}) & i = 1\ldots 4 \\ \left(P^{target}(n,i\ldots) \circledast \phi_{\sigma^{spat}(4,m)}\right)\downarrow_4 + \epsilon_i \sim \mathcal{N}(0,\sigma_i^{spect}) & i = 5\ldots 8 \\ P^{input}(n,i\ldots) & i = 9,10 \end{cases}$$

- Simulate downscaled S2 patches (x4) to train B11 and B12 (Wald trick)

$$P^{input}_{wald}(n,i\ldots) = \left(P^{input}(n,i\ldots) \circledast \phi_{\sigma^{spat}(4,m)}\right)\downarrow_4 + \epsilon_i \sim \mathcal{N}(0,\sigma_i^{spect})$$

$$P^{target}_{wald}(n,i\ldots) = \begin{cases} \left(P^{input}(n,i\ldots) \circledast \phi_{\sigma^{spat}(2,m)}\right)\downarrow_2 + \epsilon_i \sim \mathcal{N}(0,\sigma_i^{spect}) & i = 1...4 \\ P^{input}(n,i\ldots) & i = 5...10 \end{cases}$$

- Total loss

$$\mathcal{L}_{pretrain}(\theta) = \mathcal{L}_{FR}(\theta) + \mathcal{L}_{Wald}(\theta)$$

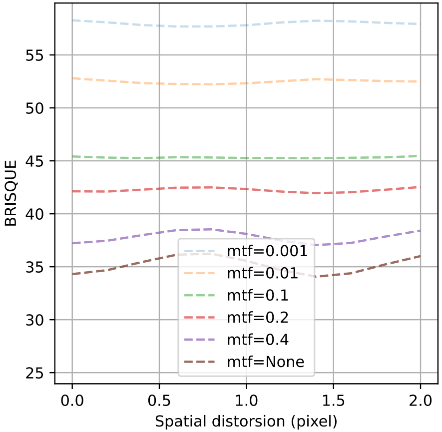
GAN fine-tuning stage (10 epochs): use real data

GAN setup

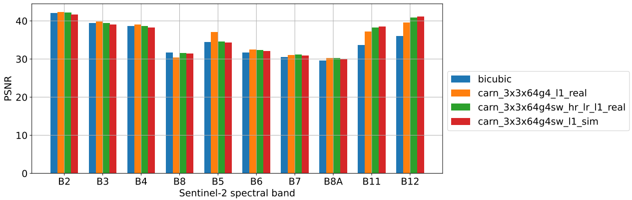
- Standard UNet discriminator with High Pass Filtering of inputs
- Relativistic GAN formulation
- Checkpointing using BRISQUE no reference metric [5]

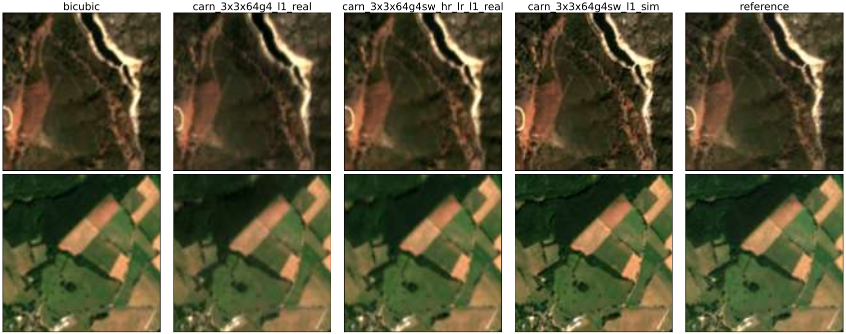
$$\mathcal{L}_G^{Ra} = \mathbb{E}_{x_f}[log(\mathcal{D}_{Ra}(x_f, x_r))] - \mathbb{E}_{x_r}[log(1 - \mathcal{D}_{Ra}(x_r, x_f))]$$

$$\mathcal{L}_{finetune}(\theta) = \mathcal{L}_{pretrain}(\theta) + \lambda \cdot \mathcal{L}_G^{Ra}(\theta)$$

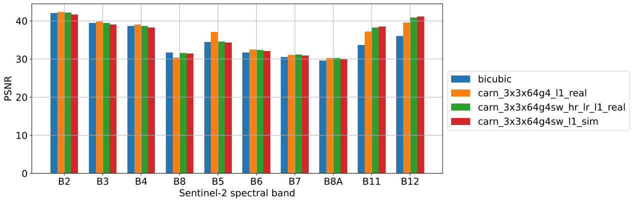


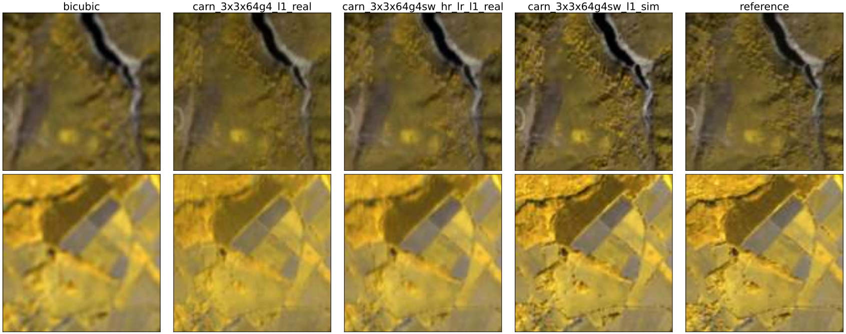
The PSNR Failure, illustrated



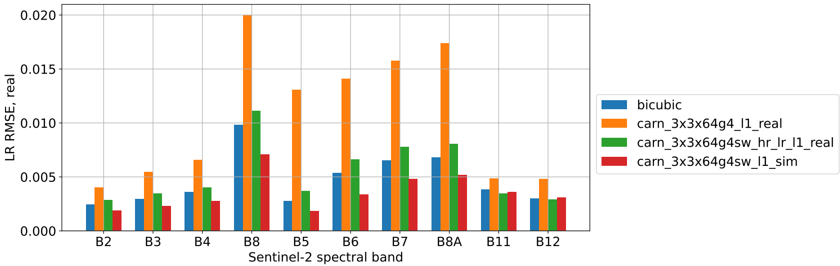


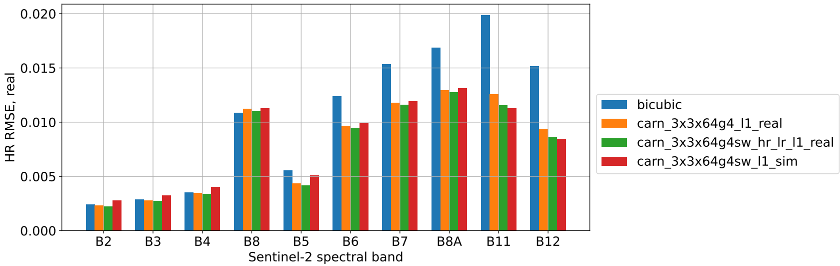
The PSNR Failure, illustrated



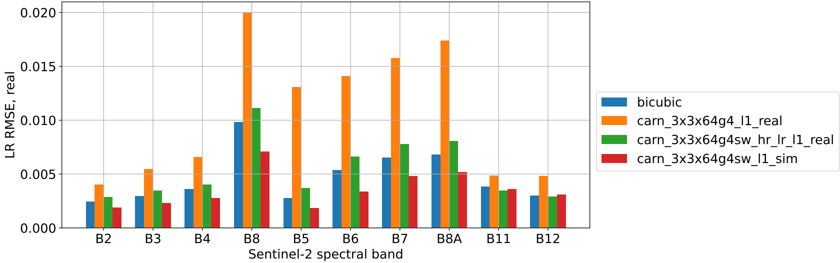


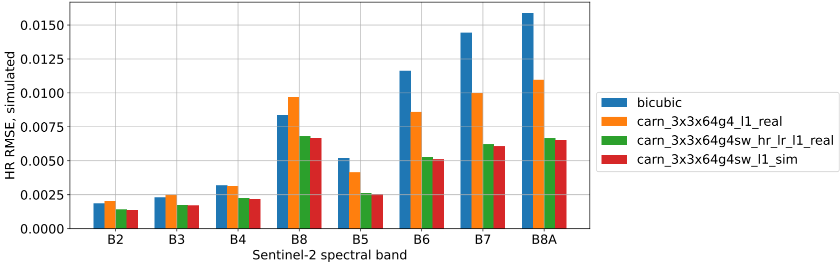
Pre-training performances



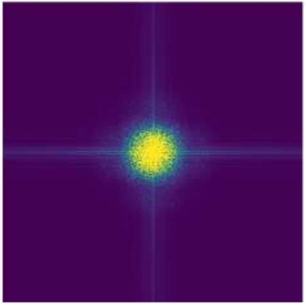


Pre-training performances

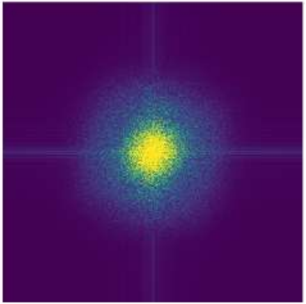




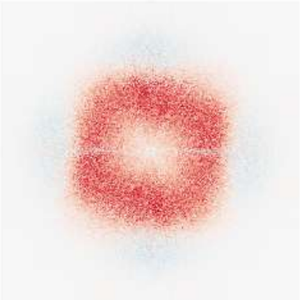
And in Fourier domain (B7)



Bicubic FFT

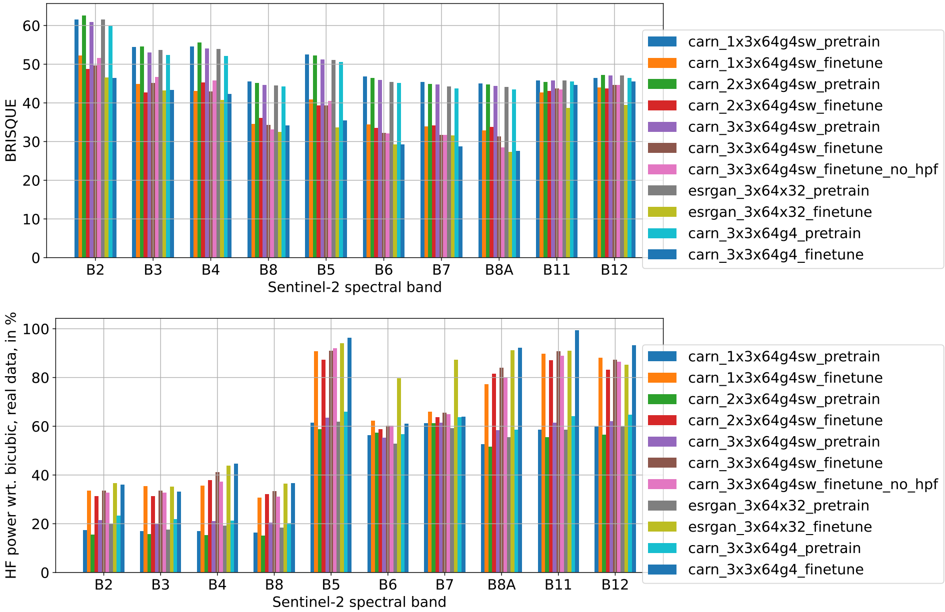


super-resolution FFT

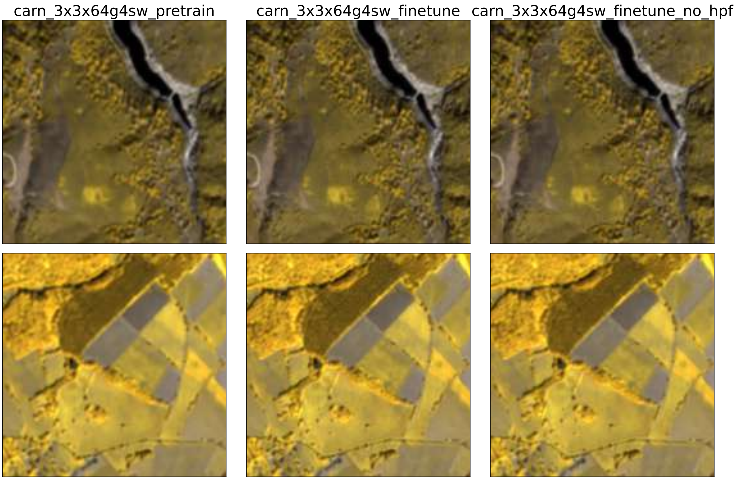


FFT difference

Performance gain from GAN fine-tuning



Fine-tuning visual assessment: minor improvements



Conclusions

Lessons learned

- Spatial and spectral discrepancies in multi-sensors dataset impairs training and evaluation
- Learning on simulated data allow to overcome this issue, with limitations
- Proper pre-training achieves a solid part of the final perf. wrt. GAN (for our shallow models)

EVOLAND

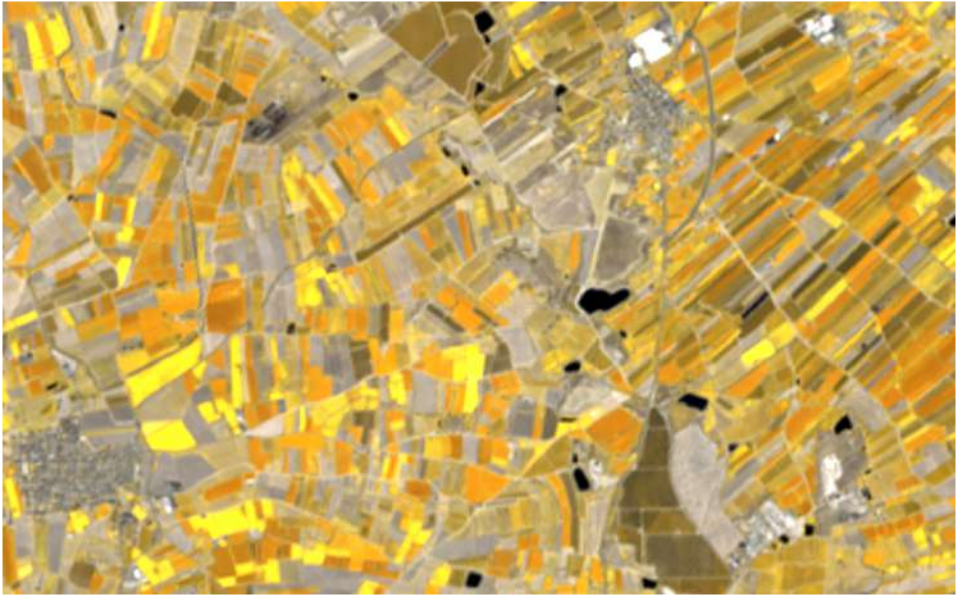
- Waiting for feedback from prototypes
- Now focusing on T2.5.2 and T2.5.3 (spatio-temporal fusion)
- Integration as a User Defined Function in OpenEO

Model available for inference on Sentinel-2 L1C and L2A products

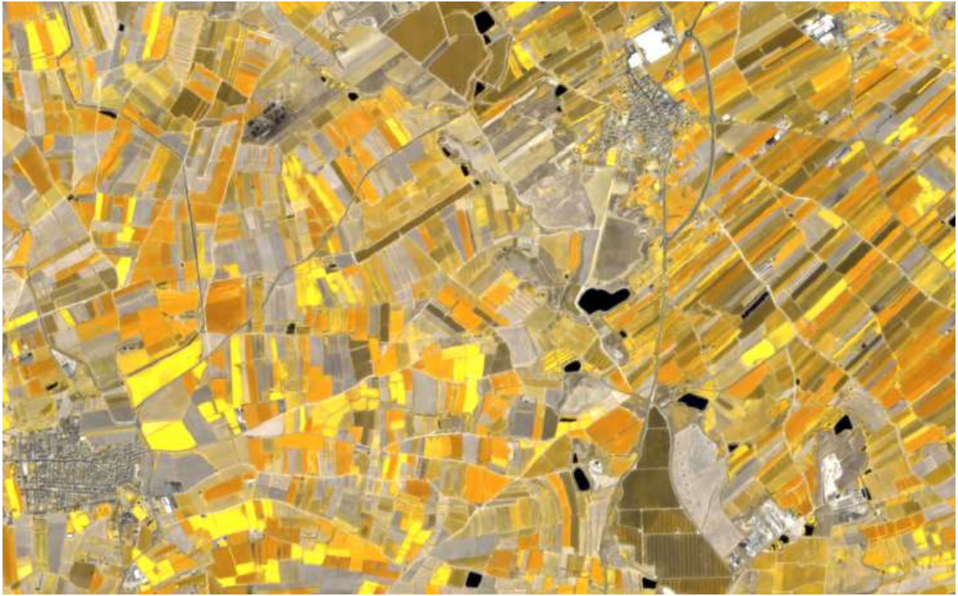
`https://github.com/Evoland-Land-Monitoring-Evolution/sentinel2_superresolution`

	CPU (1 core)	CPU (8 cores)	GPU (A100)
L1C	6 hours	1 hour	6 minutes
L2A	5 hours	50 minutes	5 minutes

Example: 31TCG, 2020.08.08, B5, B6, B7, 5m bicubic



Example: 31TCG, 2020.08.08, B5, B6, B7, 5m SISR



Thank you

THANK YOU! Any questions ?

- This work was partly performed using HPC resources from GENCI-IDRIS (Grant 2023-AD010114835)
- This work was partly performed using HPC resources from CNES.

Funded by the European Union

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

References I

- J. Michel, J. Vinasco-Salinas, J. Inglada, and O. Hagolle, “SEN2VENµS, a dataset for the training of Sentinel-2 super-resolution algorithms,” May 2022.
- N. Ahn, B. Kang, and K.-A. Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 252–268, 2018.
- X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, “Esrgan: Enhanced super-resolution generative adversarial networks,” in *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0, 2018.
- S. Anwar, S. Khan, and N. Barnes, “A deep journey into super-resolution: A survey,” *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.
- A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.