University of Ferrara – INFN section of Ferrara

Andrea Miola

Crop Classification using High-Performance Deep Learning Predictive Models for Agricultural Yields Analysis

2nd AI-INFN Advanced Hackathon

November 24, 2025

Introduction

Design, implementation and performance assessment of a DNN model for crop classification using remote sensing satellite time-series images

Use cases:

- Agricultural monitoring
- Food security analysis
- Impact of environmental hazards
- Climate change impact assessment

Key challenges:

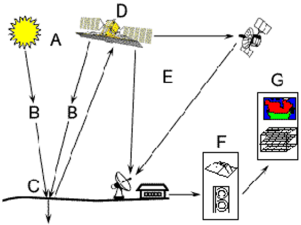
- Spectral similarity between different crops
- Temporal variability (growth stages, phenology)
- Cloud cover and missing observations
- Compute intensive models and high data volumes

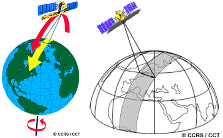
Satellite Imagery

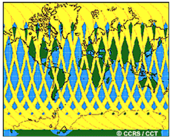
Remote Sensing

the process of using electromagnetic energy to observe and measure properties of the Earth's environment via sensor systems positioned at a distance from the region of interest (Chapman and Gasparovic, 2022)

- Satellites like **Sentinel-2** capture Earth's surface across multiple spectral bands.
- Each band records the **reflectance** of surface materials at different wavelengths.



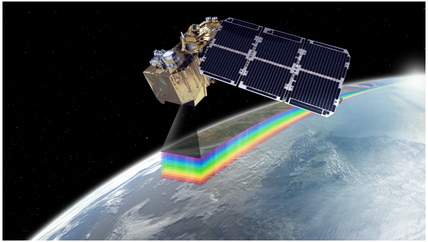




Sentinel-2

To perform crop classification, I used observations collected with Copernicus' Sentinel-2[1] constellation of satellites.

- 2 polar-orbiting satellites;
- launched in 2015;
- 786 km altitude;
- 290 km orbital swath width;
- 3 to 10 days revisit time;
- 13 spectral bands (optical, NIR and SWIR);
- Free access data.



[1] https://sentinel.esa.int/web/sentinel/missions/sentinel-2

Image Segmentation Task

Image segmentation is a computer vision technique that, for any given image, creates a mask or a matrix with various elements that specify the object class or instance to which each pixel belongs.

Time Series Importance

Crop mapping studies can't be properly made using only single date observations:

- different crops have different cycles, both in length and time-frame;
- the date of interest could be cloudy;
- fields are not equal.

High classification accuracy is only achievable using time-series of the crops growth cycle.

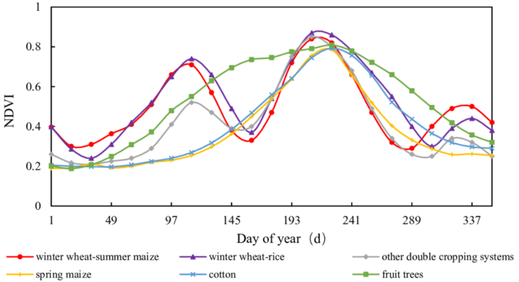
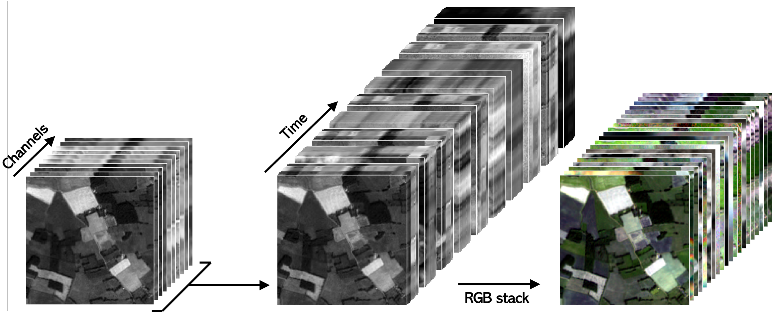


Figure 1: Example of NDVI values across the year. From Yang et. al., 2023

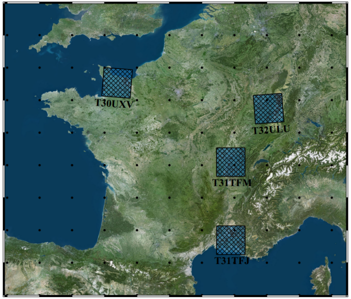
Dataset

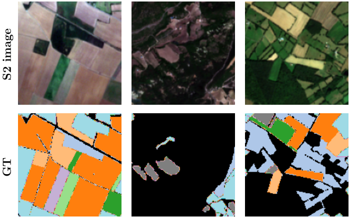
Typical approaches use composite vegetation indices (e.g. NDVI)

Two datasets of multispectral Satellite Image Time-Series (SITS) patches from Sentinel-2 observations



Dataset – PASTIS

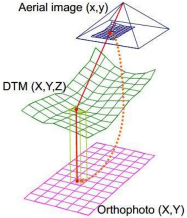




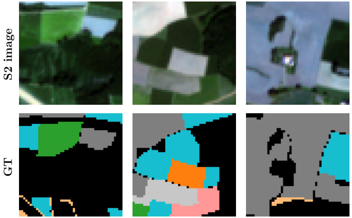
OrthoPASTIS dataset (based on PASTIS[a]):

- ≈2 400 128 × 128 px patches within France extracted from 4 Sentinel-2 L2A tiles
- 10 spectral channels
- 38 to 61 observations (Sep 2018 – Oct 2019)
- Around 124 000 field parcels
- Ground Truth contains:
  - 18 crops
  - 1 background

[a] https://github.com/VSainteuf/pastis-benchmark



Dataset – Munich

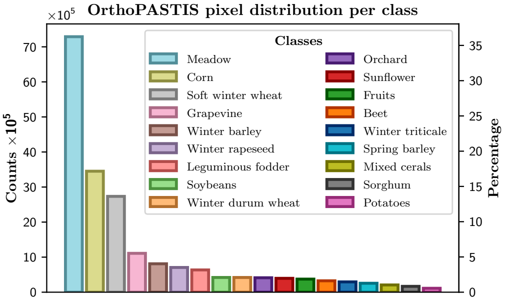


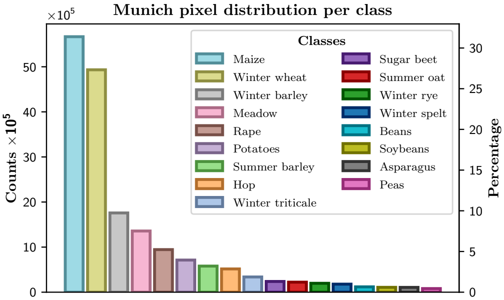
Munich dataset[a]:

- ≈28 000 48 × 48 px patches retrieved from Sentinel-2 L1C tiles around Munich
- 7 spectral channels (B02, B03, B04, B08, B10, B11, B12)
- Up to 52 observations for 2016 and 2017, respectively
- Ground Truth contains:
  - 17 crops
  - 1 background

[a] Rußwurm M. and Körner M., 2018 – https://github.com/MarcCoru/MTLCC

Dataset – Distribution



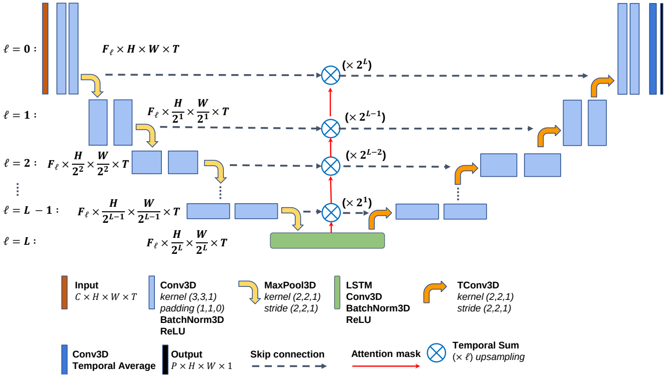


Extremely imbalanced!

Architecture – AgrUNet

I developed AgrUNet, a modular, configurable UNet-like architecture. It features 3D convolutions in the encoder and decoder branches and a Temporal Attention Encoder (e.g. ConvLSTM) in the bottleneck layer.

PyTorch Lightning



- Configurable depth
- "Space first, time second" strategy
- Multi-GPU ready (DDP)
- Mixed-precision support
- Deep Supervision
- Dice + Cross Entropy loss
- 3 accuracy metrics (Dice, IoU, Overall Accuracy)

HPC Infrastructure

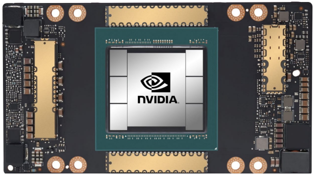
COKA Cluster @ UniFE

dual 32-core Intel Gold 6242 CPUs

394 GB RAM

4x Nvidia V100 32 GB GPUs

NVLink technology

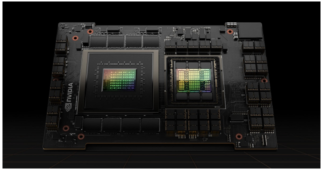


DGX @ CINECA

2x 64-core AMD EPYC 7742 CPUs

320 GB RAM

8x Nvidia A100 40 GB GPUs



GH200 @ UniFE

1x 72 core Grace ARM CPU

580 GB RAM

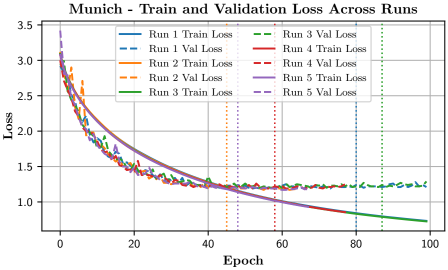
1x Nvidia H100 96 GB GPU

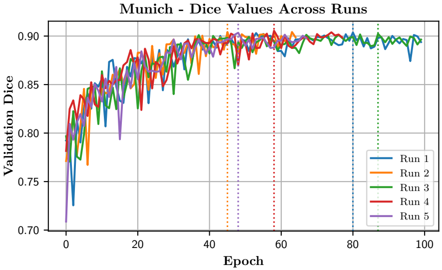
NVLink-C2C connection

Training phase

- k-folding cross-validation with $k = 5$
- 100 epochs
- Validation Dice score
- Early stopping criterion at 20 epochs
- On-the-fly Z-score channel-wise normalization
- No data augmentation
- Hyperparameters tuned with *Optuna*

	Munich	**OrthoPASTIS**
Learning Rate	1.1×10^{-3}	9×10^{-4}
Downsamplings	3	2
Filters	[64, 128, 256]	[64, 128]
Batch size	32	4





Segmentation results

	OrthoPASTIS			Munich		
	Dice	mIoU	OA	Dice	mIoU	OA
ConvRNN	-	-	-	-	-	0.896
3D FPN	-	-	-	-	-	0.936
TSViT	-	0.651 *	0.834 *	-	0.848	0.950
UNet-3D	0.801	0.699	0.912	0.921	0.856	0.965
U-TAE	0.821	0.723	0.921	0.929	**0.888**	**0.976**
AgrUNet	**0.862**	**0.777**	**0.944**	**0.932**	0.880	0.971
Difference	0.041	0.054	0.023	0.003	−0.008	−0.005

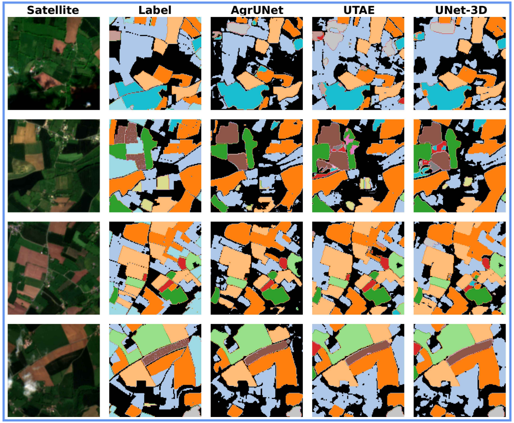
Results averaged across 5 folds.

Values marked with (*) are obtained from a different database configuration.

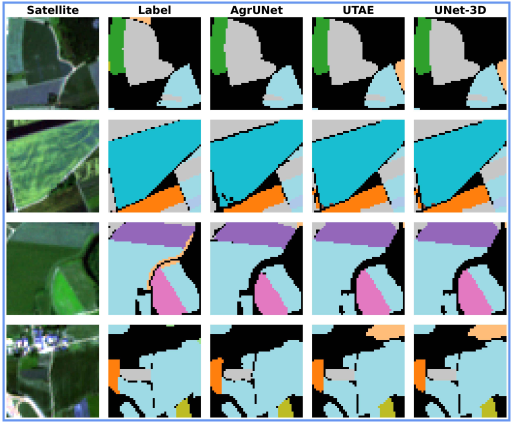
State-of-the-art-like results!

Segmentation results

OrthoPASTIS



Munich



Computational results – Full time-series lenght

Throughput measured in img/s (higher is better) – V100 HPC system

TRAINING		OrthoPASTIS		Munich	
	GPUs	T_{FP32}	T_{AMP}	T_{FP32}	T_{AMP}
UNet-3D	1	2.95		24.83	
U-TAE	1	6.51		40.94	
AgrUNet	1	2.52	9.08	23.85	63.99
AgrUNet	2	5.01	17.71	45.16	123.82
AgrUNet	3	7.51	26.29	67.93	186.26
AgrUNet	4	10.01	35.29	90.23	247.50

	OrthoPASTIS	Munich
	n. Params	n. Params
UNet-3D	1.55 M	1.55 M
U-TAE	1.09 M	1.08 M
AgrUNet	2.02 M	6.28 M

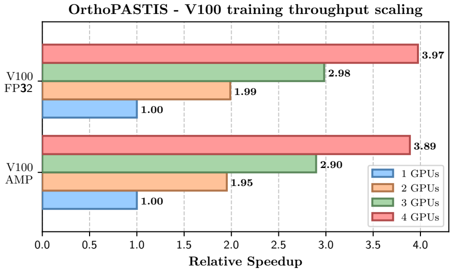
Computational results – Full time-series lenght

Throughput measured in img/s (higher is better) – V100 HPC system

TRAINING		OrthoPASTIS		Munich	
	GPUs	T_{FP32}	T_{AMP}	T_{FP32}	T_{AMP}
UNet-3D	1	2.95		24.83	
U-TAE	1	6.51		40.94	
AgrUNet	1	2.52	9.08	23.85	63.99
AgrUNet	2	5.01	17.71	45.16	123.82
AgrUNet	3	7.51	26.29	67.93	186.26
AgrUNet	4	10.01	35.29	90.23	247.50

	OrthoPASTIS		Munich	
	n	Params	n	Params
AgrUNet		2.02 M		6.28 M

Up to 6x training speedup!



Computational results – Full time-series lenght

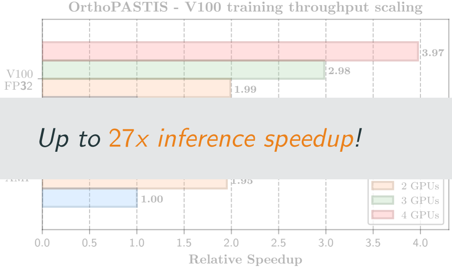
Throughput measured in img/s (higher is better) – V100 HPC system

TRAINING	GPUs	OrthoPASTIS T_{FP32}	OrthoPASTIS T_{AMP}	Munich T_{FP32}	Munich T_{AMP}
UNet-3D	1	2.95		24.83	
U-TAE	1	6.51		40.94	
AgrUNet	1	2.52	9.08	23.85	63.99
AgrUNet	2	5.01	17.71	45.16	123.82
AgrUNet	3	7.51	26.29	67.93	186.26
AgrUNet	4	10.01	35.29	90.23	247.50

INFERENCE	GPUs	OrthoPASTIS T_{FP32}	OrthoPASTIS T_{AMP}	Munich T_{FP32}	Munich T_{AMP}
UNet-3D	1	5.41		22.40	
U-TAE	1	5.93		59.29	
AgrUNet	1	9.35	41.04	105.73	263.68
AgrUNet	2	18.65	81.55	209.94	519.39
AgrUNet	4	37.34	162.09	428.02	1 057.24

	OrthoPASTIS n	OrthoPASTIS Params	Munich n	Munich Params
AgrUNet		2.02 M		6.28 M

Up to 6x training speedup!



Up to 27x inference speedup!

Computational results – Time-series optimization

How to improve performance?

Due to shortage of good data, time-series reduced to the January-July period

avg training throughput comparison on the OrthoPASTIS dataset on the January–July and the full-year periods

TRAINING	T_{AMP}			T_{FP32}		
GPUs	T_{year}	T_{JJ}	Δ	T_{year}	T_{JJ}	Δ
1	9.08	15.30	1.7	2.52	4.52	1.8
4	35.29	60.28	1.7	10.01	18.08	1.8
INFERENCE	41.04	86.05	2.1	9.35	16.69	1.8

Nearly 2x speedup in training and inference time!

Dice score obtained: 0.865

Dice score full year: 0.862

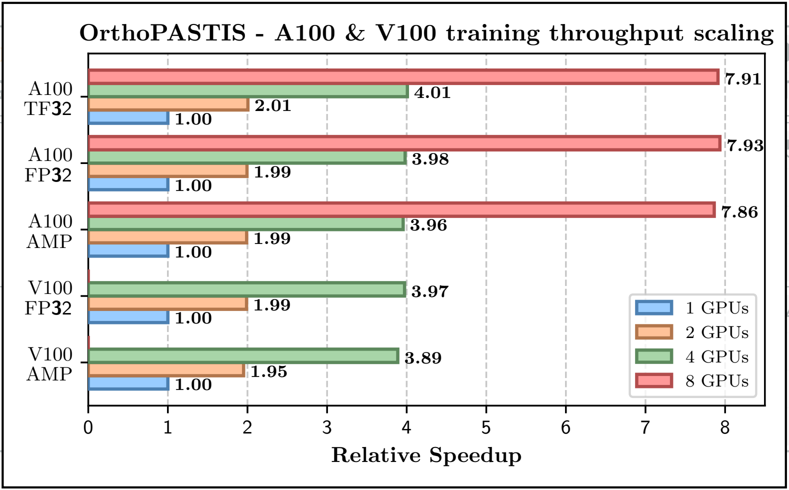
Computational results – different GPUs

Training throughput (img/s) comparison for the 3 HPC infrastructure considered

TRAINING	GPUs	Nvidia V100 T_{FP32}	Nvidia V100 T_{AMP}	Nvidia A100 T_{FP32}	Nvidia A100 T_{TF32}	Nvidia A100 T_{AMP}	Nvidia H100 T_{FP32}	Nvidia H100 T_{TF32}	Nvidia H100 T_{AMP}
OrthoPASTIS	1	4.52	15.30	5.61	19.65	27.85	12.71	37.62	59.49
	2	9.02	30.07	11.19	39.42	55.41			
	4	18.08	60.28	22.34	78.78	110.17			
	8			44.54	155.47	218.99			
Munich	1	38.09	87.85	46.29	126.96	141.09	64.91	185.25	243.35
	2	74.24	169.94	92.15	252.40	273.75			
	4	147.93	337.19	183.84	504.61	545.16			
	8			366.25	999.31	1 076.18			

Computational results – different GPUs

Training throughput (img/s) comparison for the 3 HPC infrastructure considered



Computational results – different GPUs

Inference throughput (img/s) comparison for the 3 HPC infrastructure used

INFERENCE		Nvidia V100		Nvidia A100			Nvidia H100		
	GPUs	T_{FP32}	T_{AMP}	T_{FP32}	T_{TF32}	T_{AMP}	T_{FP32}	T_{TF32}	T_{AMP}
OrthoPASTIS	1	16.69	86.05	16.37	75.64	125.93	75.48	264.41	605.53
	2	33.23	159.30	32.58	125.41	230.84			
	4	66.44	270.30	64.23	232.05	434.00			
	8			120.40	340.30	735.75			
Munich	1	153.30	373.65	207.77	479.90	740.92	280.56	897.16	1 649.10
	2	304.65	735.77	405.32	908.62	1 406.66			
	4	609.10	1 455.93	705.76	1 645.55	2 779.95			
	8			1 505.78	3 166.60	5 450.09			

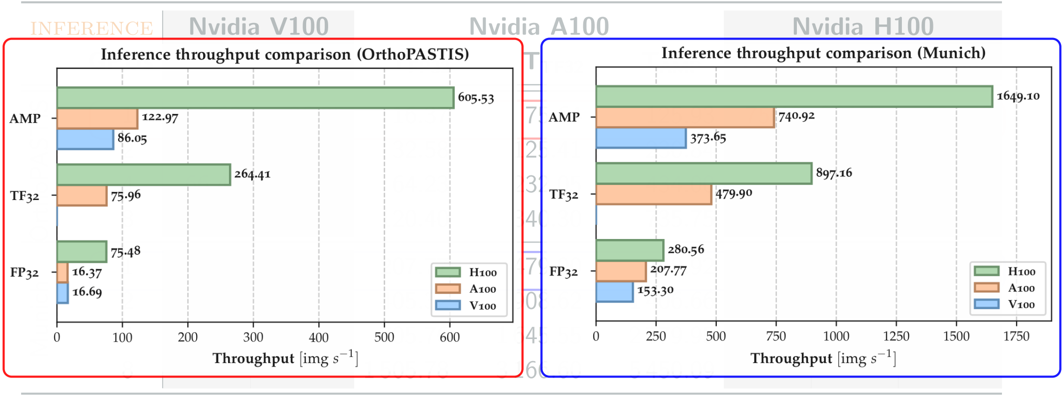
Computational results – different GPUs

Inference throughput (img/s) comparison for the 3 HPC infrastructure used

INFERENCE		Nvidia V100		Nvidia A100			Nvidia H100		
	GPUs	T_{FP32}	T_{AMP}	T_{FP32}	T_{TF32}	T_{AMP}	T_{FP32}	T_{TF32}	T_{AMP}
OrthoPASTIS	1	16.69	86.05	16.37	75.64	125.93	75.48	264.41	605.53
	2	33.23	159.30	32.58	125.41	230.84			
	4	66.44	270.30	64.23	232.05	434.00			
	8			120.40	340.30	735.75			
Munich	1	153.30	373.65	207.77	479.90	740.92	280.56	897.16	1 649.10
	2	304.65	735.77	405.32	908.62	1 406.66			
	4	609.10	1 455.93	705.76	1 645.55	2 779.95			
	8			1 505.78	3 166.60	5 450.09			

Computational results – different GPUs

Inference throughput (img/s) comparison for the 3 HPC infrastructure used



Expanding AgrUNet: EU-MESEO project

EU-MESEO is a collaborative research project, designed to contribute to developing and deploying global, space-based services, applications and data and to foster competitiveness of the European space sector (exp. Nov 2026).

My role in the project:

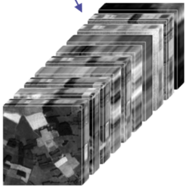
- Design an End-to-End pipeline simulating the downstream path of a user request, from the acquisition of images on-board the satellite, up to the processing of the products on ground for the thematic use cases
- integrate AgrUNet as one of the thematic processors

EU-MESEO – Orchestration

Orchestrator Input	Data Retrieval	Data Preprocessing	Pretrained model selection	Inference	Post Processing	Final Output
Job Order XML file	XML parsing, locate the images	Resolution normalization, AOI cropping, etc.	Selection based on the AOI geographical location	Time-series parsing	Predictions vectorization, computation statistics, accuracy results	Zip of results and log files

Job Order XML file containing:

- <id_request>
- AOI coordinates (EPSG:4326)
- start/end processing dates in yyyy-MM-DD format
- input folder path, containing the TIF files of the time-series ("MESEO_<id_request>_<yyyy-MM-DD>.tif")
- Sentinel-2 L1C images have 7 bands
- output folder path, containing the processed products in a ZIP file



Conclusion

I developed a new DL configurable architecture for the time-series satellite image segmentation, AgrUNet, which is able to:

- fully exploit mixed precision training and multi-GPU HPC environments, up to 45x throughput wrt other models on same HPC system
- achieve state-of-the-art multiclass segmentation results on two different satellite multispectral, multitemporal satellite datasets (0.87 and 0.93 Dice score)
- achieve nearly linear throughput scaling on multi-GPU systems
- limit the time-series lenght without reducing the segmentation performance (2x speedup)

I also assessed the performance on different GPUs and HPC infrastructures and integrated it in real-world projects such as EU-MESEO.

Thank you!