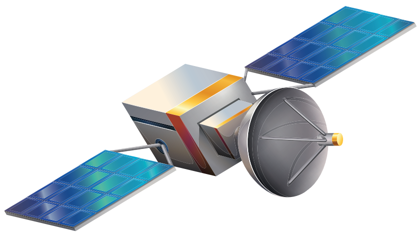
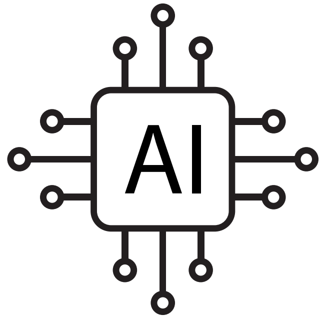
EPFL

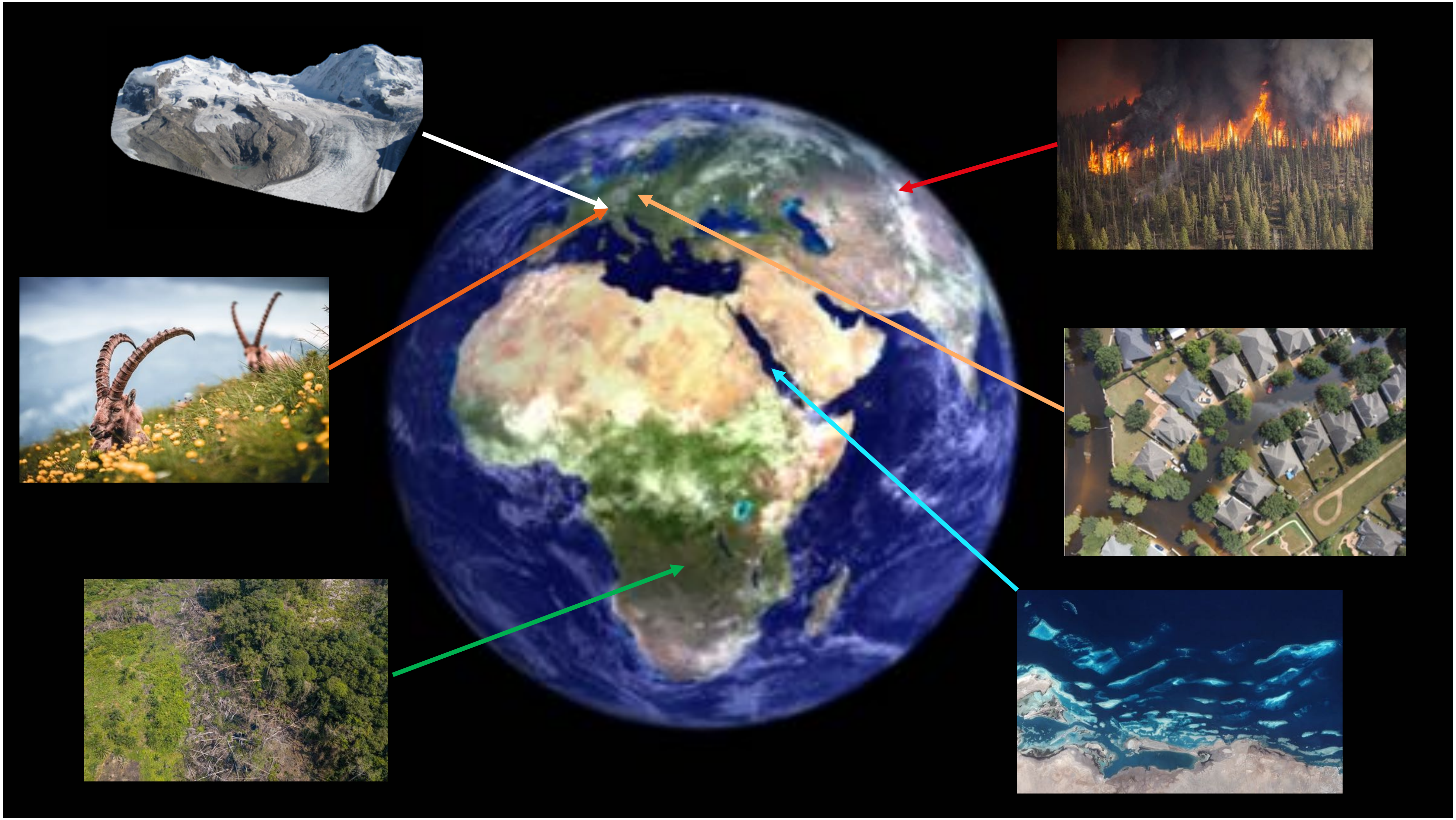
Machine learning for the environment: monitoring the pulse of our Planet with remotely sensed data

Prof. Devis Tuia, EPFL





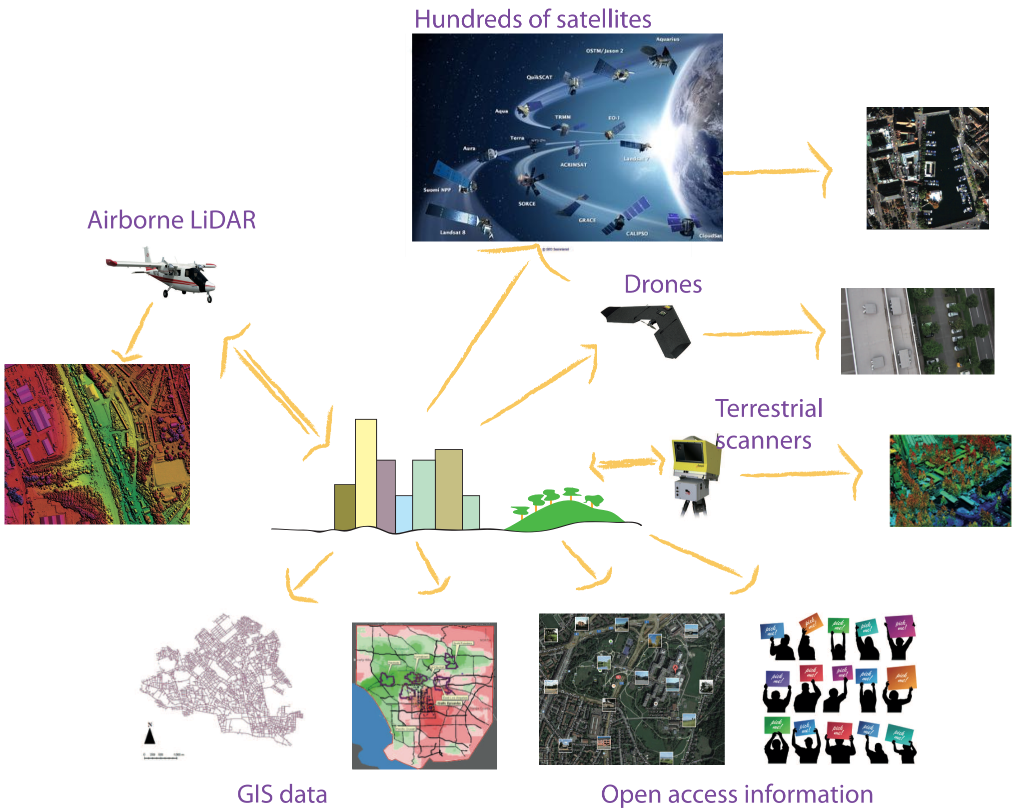
École
polytechnique
fédérale
de Lausanne



EPFL

There were many sensor data to monitor Earth in 2015

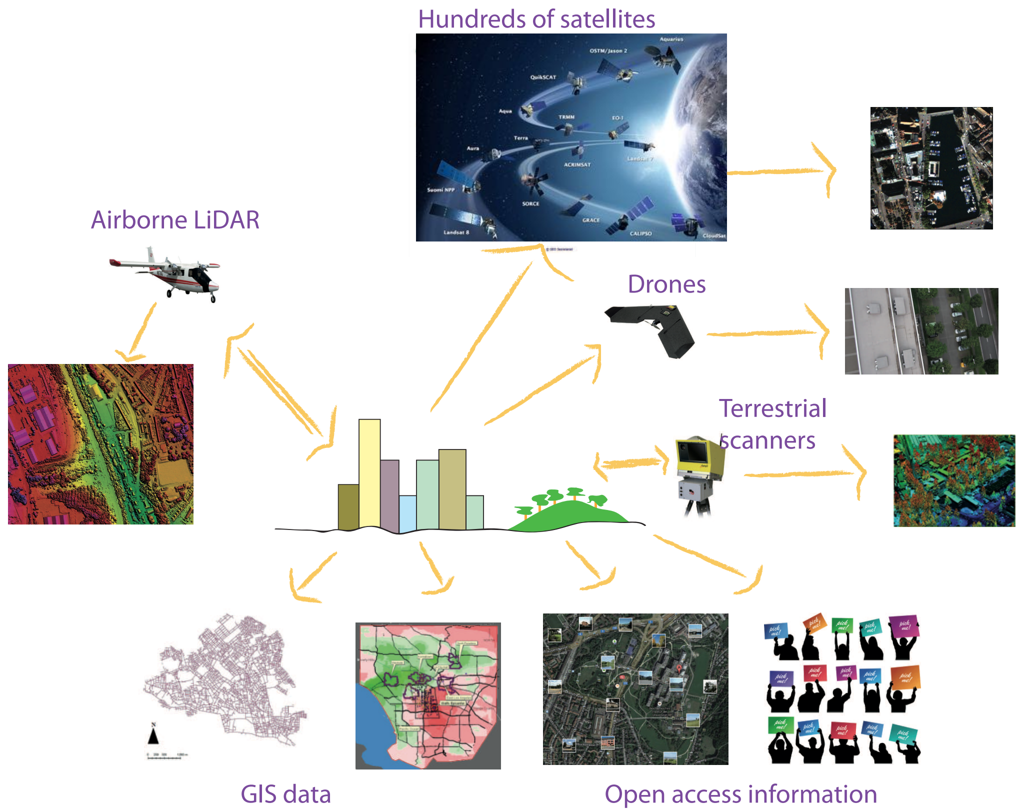
- 333 Earth Observation satellites in orbit in 2015 [ucsusa.org].
- 10'000 recreational drones registered in the U.S. by 2020 [FAA].
- 20 Pb of oblique photos in Google Street View in 2015 [Google Maps].



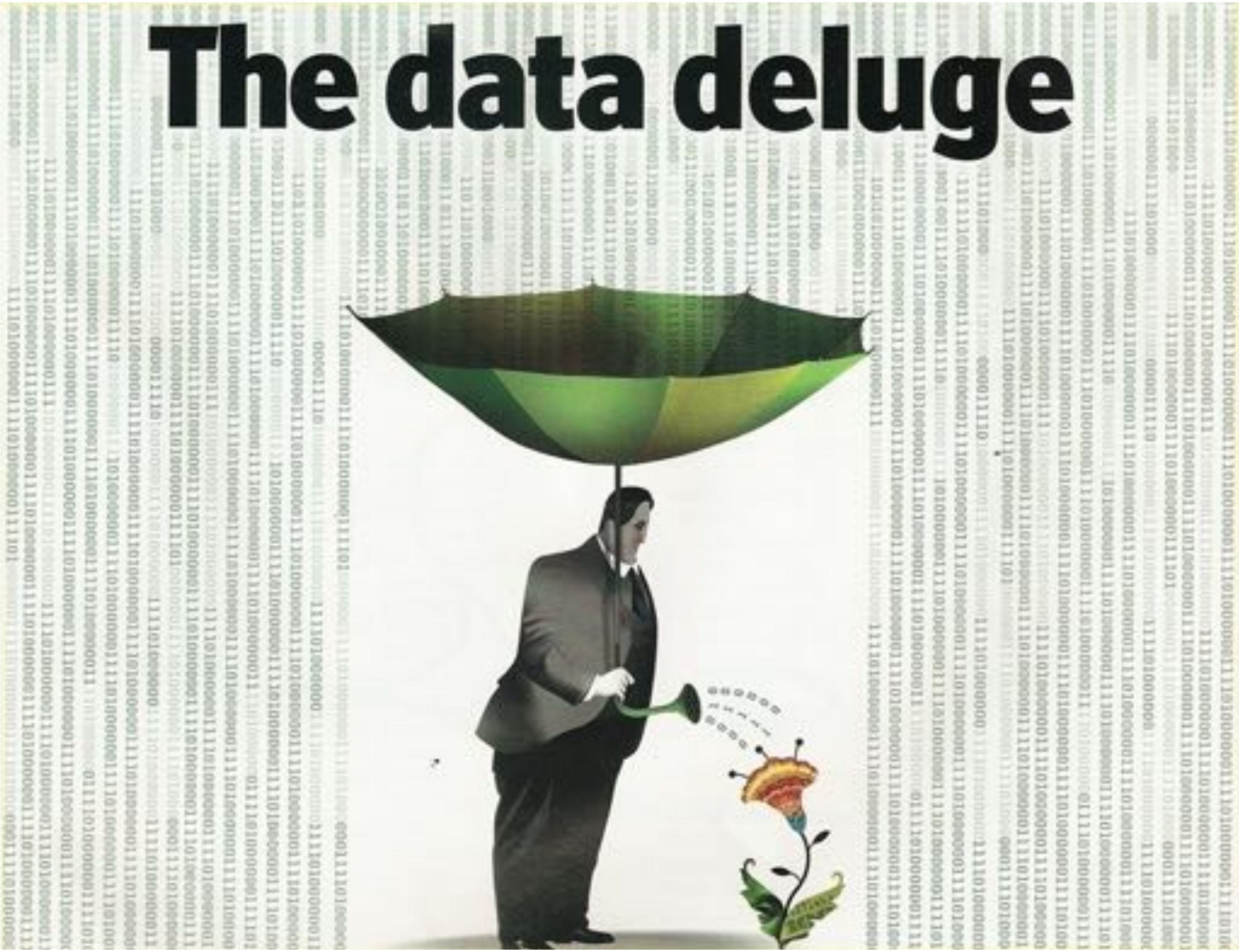
Sources of figures: usgs, astrium, aerialsurvey, sensefly, vimeo, GEOSS, Purdue, Google, thelogofactory

There are many sensor data to monitor Earth in ~~2015~~ 2023

- ~~333~~ 1'005 Earth Observation satellites in orbit in 2023 [ucsusa.org].
- ~~10'000~~ 1'100'000 recreational drones registered in the U.S in 2023. [FAA].
- 170 billions of oblique photos in Google Street View in 2020 [Google Maps].



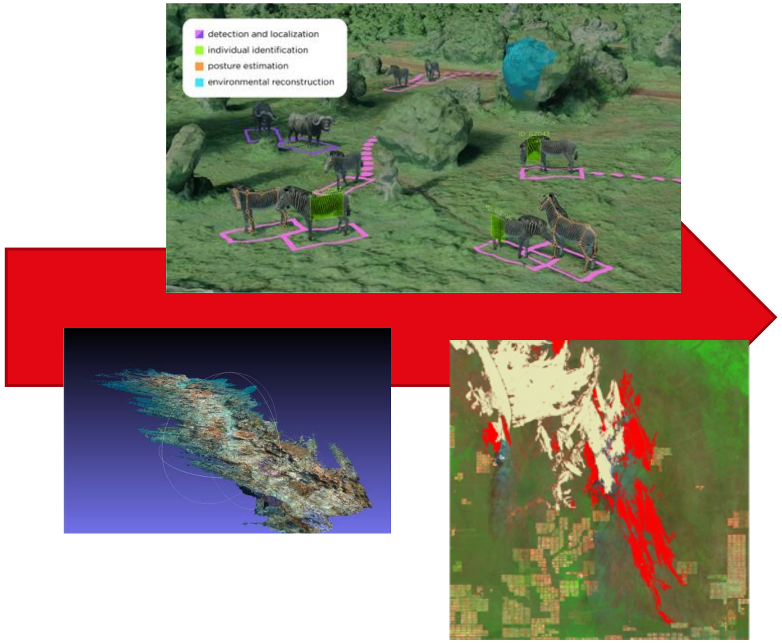
Sources of figures: usgs, astrium, aerialsurvey, sensefly, vimeo, GEOSS, Purdue, Google, thelogofactory

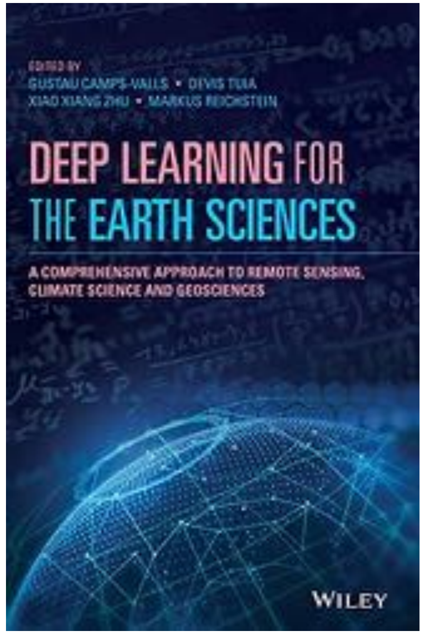

The data deluge

Why now : statistical and computational models are good enough…

- Machine learning has reached a certain maturity… and percolated in many fields of science.

2022

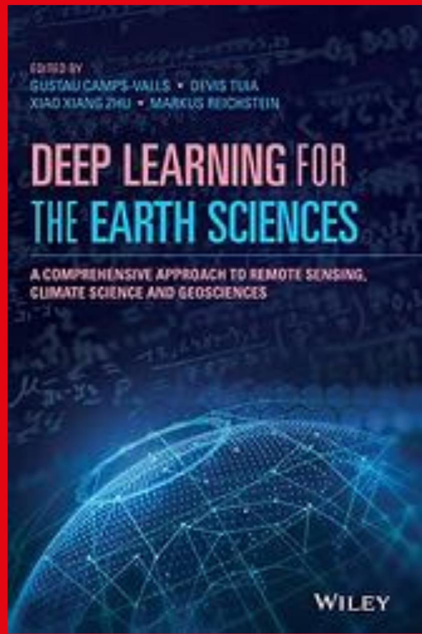




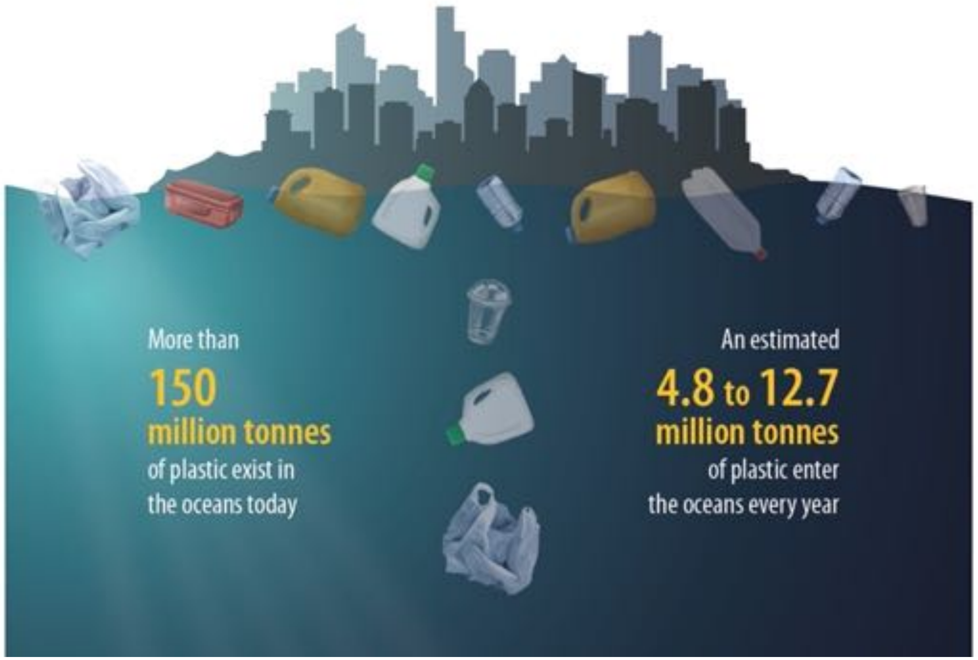
2015

With Earth observation and AI, we can

develop computational approaches
to the environmental sciences
that are accurate



Marine litter is a BIG problem



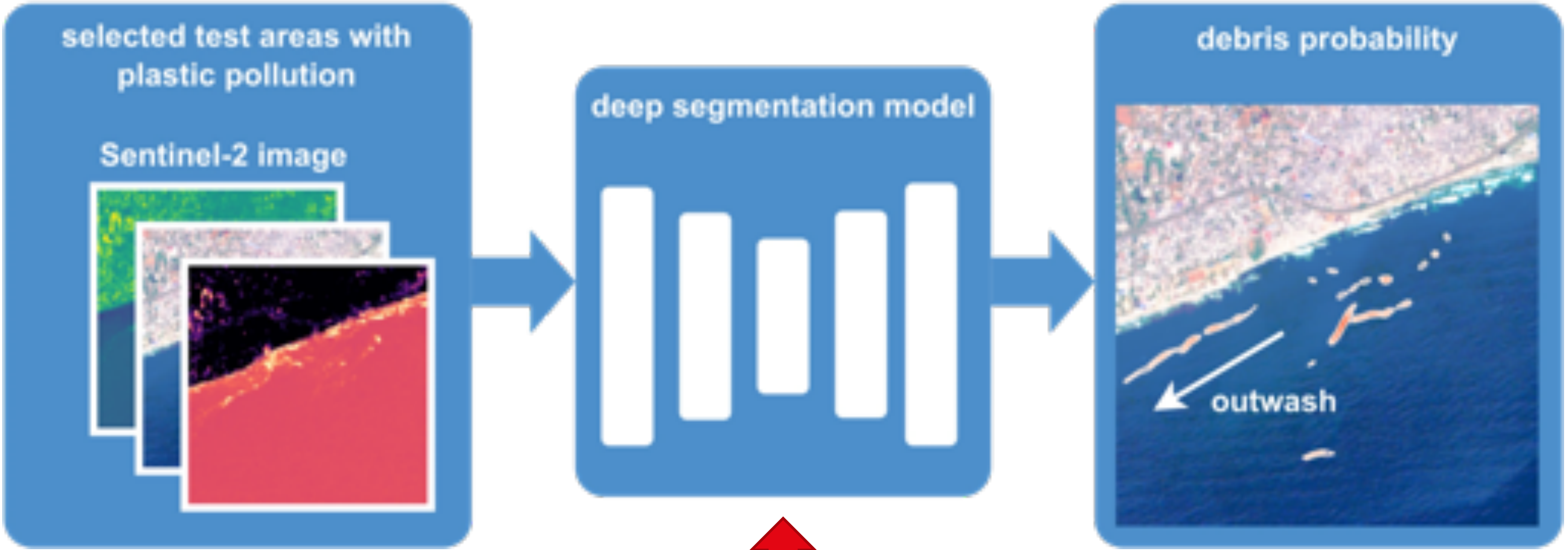
Macro-plastics decompose in microplastics that are

- a direct danger to animals

- have been found in
  - Antarctic Penguins
  - deep-sea sediments
  - human stool
  - …

with unclear and potentially harmful impact on human health

Building environmental deep learning models

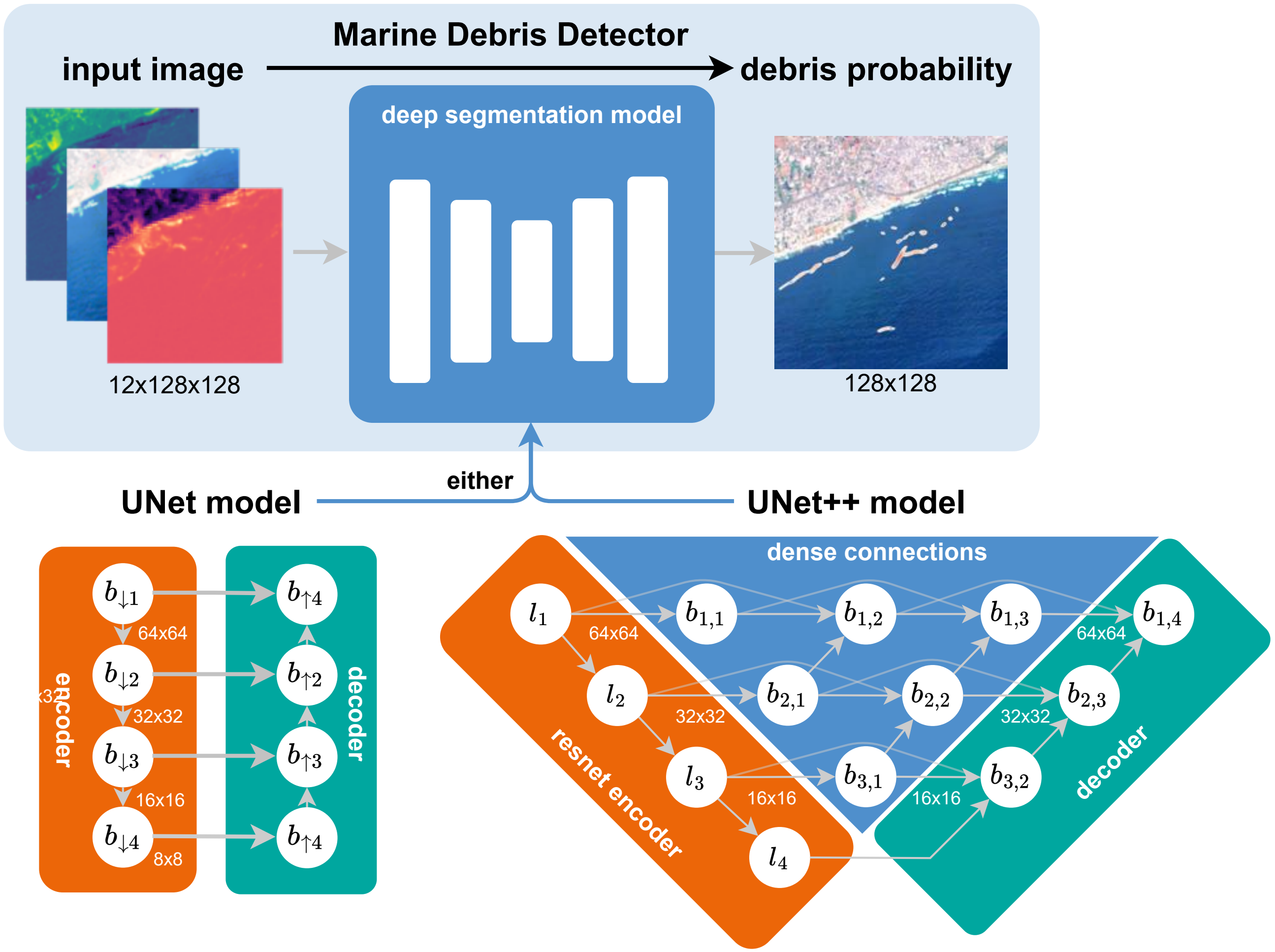


Images

Neural network
with some millions
of learnable parameters

Examples of process to be monitored

Learning Spatial Context with CNNs



EPFL

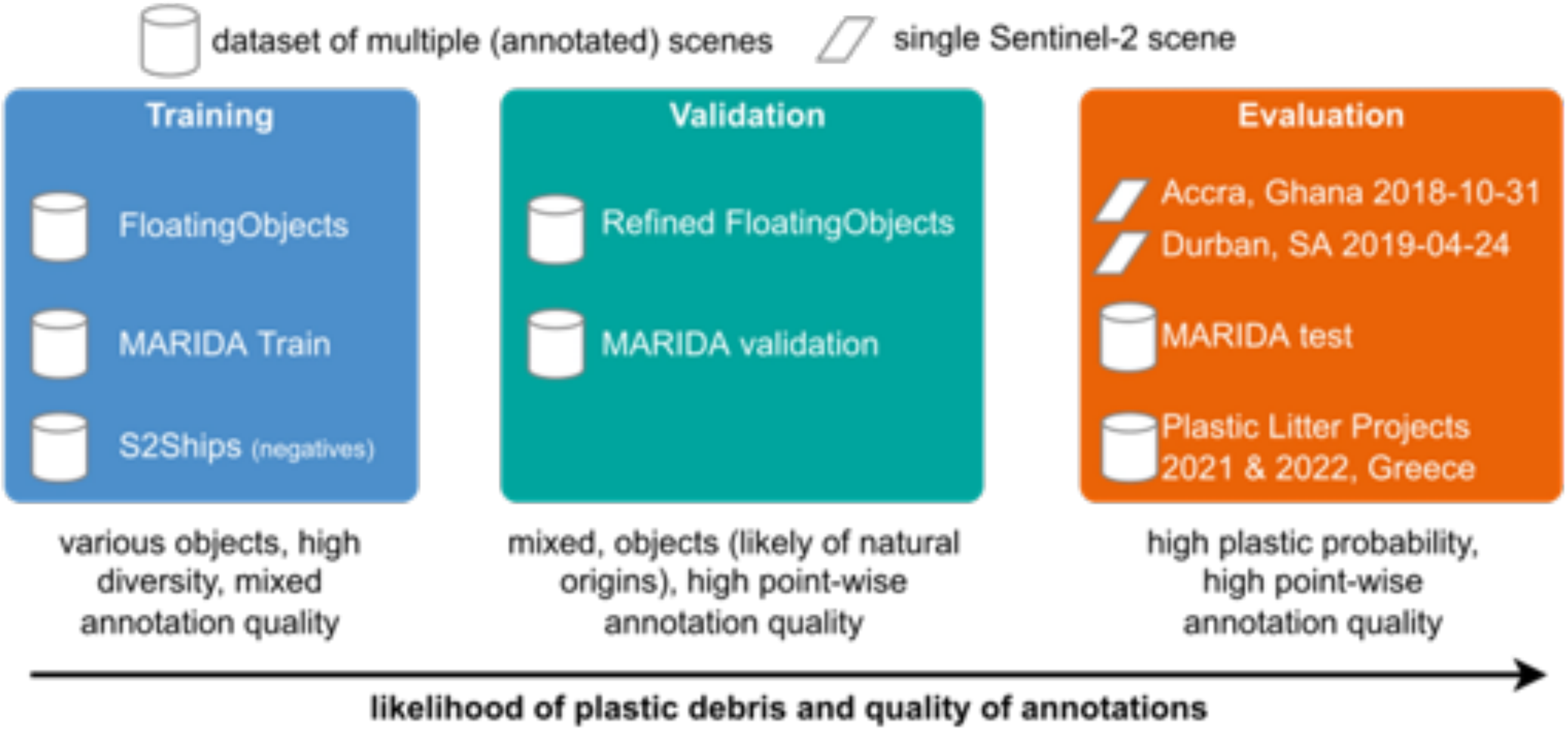
Building environmental deep learning models

- We used debris events found in news and social media, then hand labeled on images by experts.

[Mifdal et al., 2020]

Mifdal, J., Longépé, N., and **Rußwurm, M.**: Towards detecting floating objects on a global scale with spatial features using Sentinel-2, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-3-2021, 285–293,

Aggregation of large training and evaluation dataset



EPFL

Results

Takehome

- DL models outperform traditional RF

Accra trained on	original data RF	original data UNET	our train set RF	our train set UNET	our train set UNET++
ACCURACY	0.653	0.882	0.680	**0.924** ± 0.016	**0.930** ± 0.016
F-SCORE	0.464	0.871	0.545	**0.920** ± 0.018	**0.926** ± 0.018
AUROC	0.246	0.965	0.899	**0.978** ± 0.008	**0.981** ± 0.006
JACCARD	0.302	0.772	0.374	**0.852** ± 0.030	**0.862** ± 0.031
KAPPA	0.301	0.764	0.357	**0.848** ± 0.031	**0.859** ± 0.031

Durban trained on	original data RF	original data UNET	our train set RF	our train set UNET	our train set UNET++
ACCURACY	0.781	0.587	0.811	0.908 ± 0.010	**0.934** ± 0.018
F-SCORE	0.105	0.497	0.708	0.756 ± 0.032	**0.837** ± 0.053
AUROC	0.376	0.765	0.862	0.850 ± 0.030	**0.914** ± 0.018
JACCARD	0.055	0.330	0.548	0.609 ± 0.042	**0.722** ± 0.048
KAPPA	0.082	0.245	0.569	0.704 ± 0.037	**0.797** ± 0.063

Marida-test set trained on	original data RF	original data UNET	our train set RF	our train set UNET	our train set UNET++
ACCURACY	0.697	0.838	0.811	**0.865** ± 0.006	**0.867** ± 0.005
F-SCORE	0.288	0.701	0.708	**0.741** ± 0.012	**0.749** ± 0.009
AUROC	0.488	0.764	0.862	**0.738** ± 0.012	**0.746** ± 0.021
JACCARD	0.168	0.539	0.548	**0.589** ± 0.015	**0.598** ± 0.012
KAPPA	0.197	0.593	0.569	**0.654** ± 0.016	**0.661** ± 0.012

EPFL

Results

Takehome

- DL models outperform traditional RF
- Data more important than models!

FlObs+MARIDA+S2Ships

Accra	original data			our train set	
trained on	RF	UNET	RF	UNET	UNET++
ACCURACY	0.653	0.882	0.680	**0.924** ± 0.016	**0.930** ± 0.016
F-SCORE	0.464	0.871	0.545	**0.920** ± 0.018	**0.926** ± 0.018
AUROC	0.246	0.965	0.899	**0.978** ± 0.008	**0.981** ± 0.006
JACCARD	0.302	0.772	0.374	**0.852** ± 0.030	**0.862** ± 0.031
KAPPA	0.301	0.764	0.357	**0.848** ± 0.031	**0.859** ± 0.031

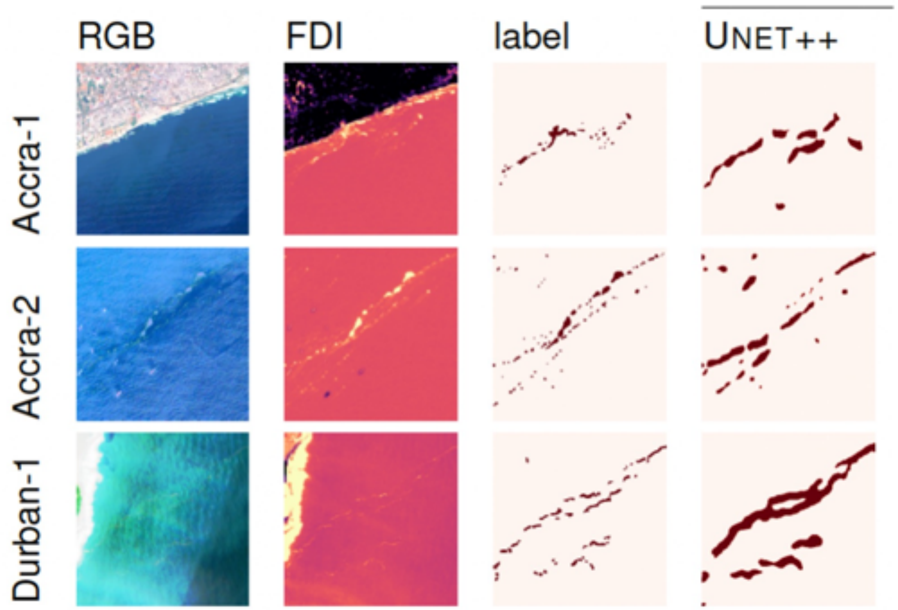
Durban	original data			our train set	
trained on	RF	UNET	RF	UNET	UNET++
ACCURACY	0.781	0.587	0.811	0.908 ± 0.010	**0.934** ± 0.018
F-SCORE	0.105	0.497	0.708	0.756 ± 0.032	**0.837** ± 0.053
AUROC	0.376	0.765	0.862	0.850 ± 0.030	**0.914** ± 0.018
JACCARD	0.055	0.330	0.548	0.609 ± 0.042	**0.722** ± 0.048
KAPPA	0.082	0.245	0.569	0.704 ± 0.037	**0.797** ± 0.063

Marida-test set	original data			our train set	
trained on	RF	UNET	RF	UNET	UNET++
ACCURACY	0.697	0.838	0.811	**0.865** ± 0.006	**0.867** ± 0.005
F-SCORE	0.288	0.701	0.708	**0.741** ± 0.012	**0.749** ± 0.009
AUROC	0.488	0.764	0.862	**0.738** ± 0.012	**0.746** ± 0.021
JACCARD	0.168	0.539	0.548	**0.589** ± 0.015	**0.598** ± 0.012
KAPPA	0.197	0.593	0.569	**0.654** ± 0.016	**0.661** ± 0.012

EPFL

Building environmental deep learning models that are accurate

- Our learning models detect plastics at sea from space with ~ 85% accuracy



M. Russwurm, Venkatesam S. J., and **D. Tuia.** Large-scale detection of marine debris in coastal areas with Sentinel-2. *Under review.*

Going further: few shot scenarios

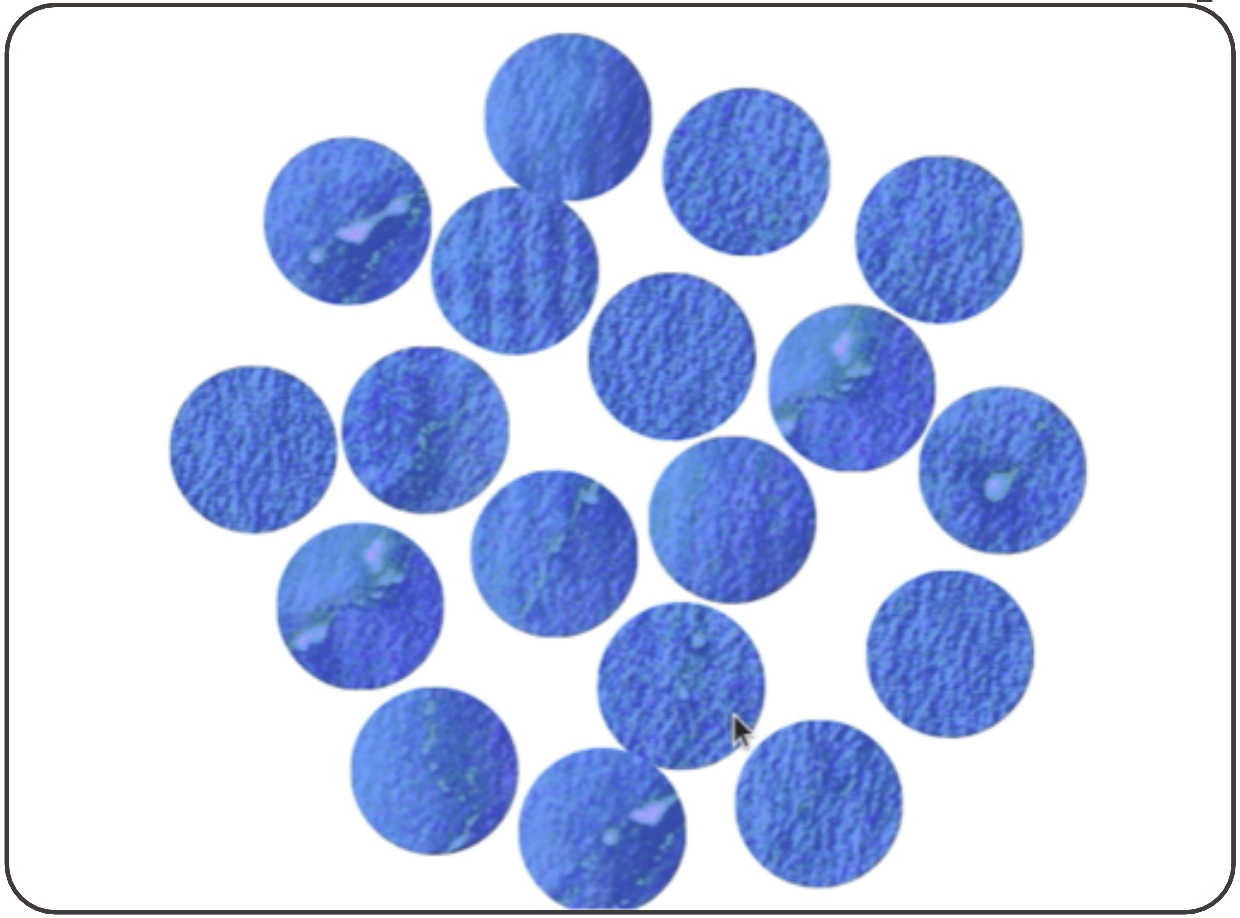
tackling geographical diversity with meta-learning

- Each region is different!
- So is each timestep!

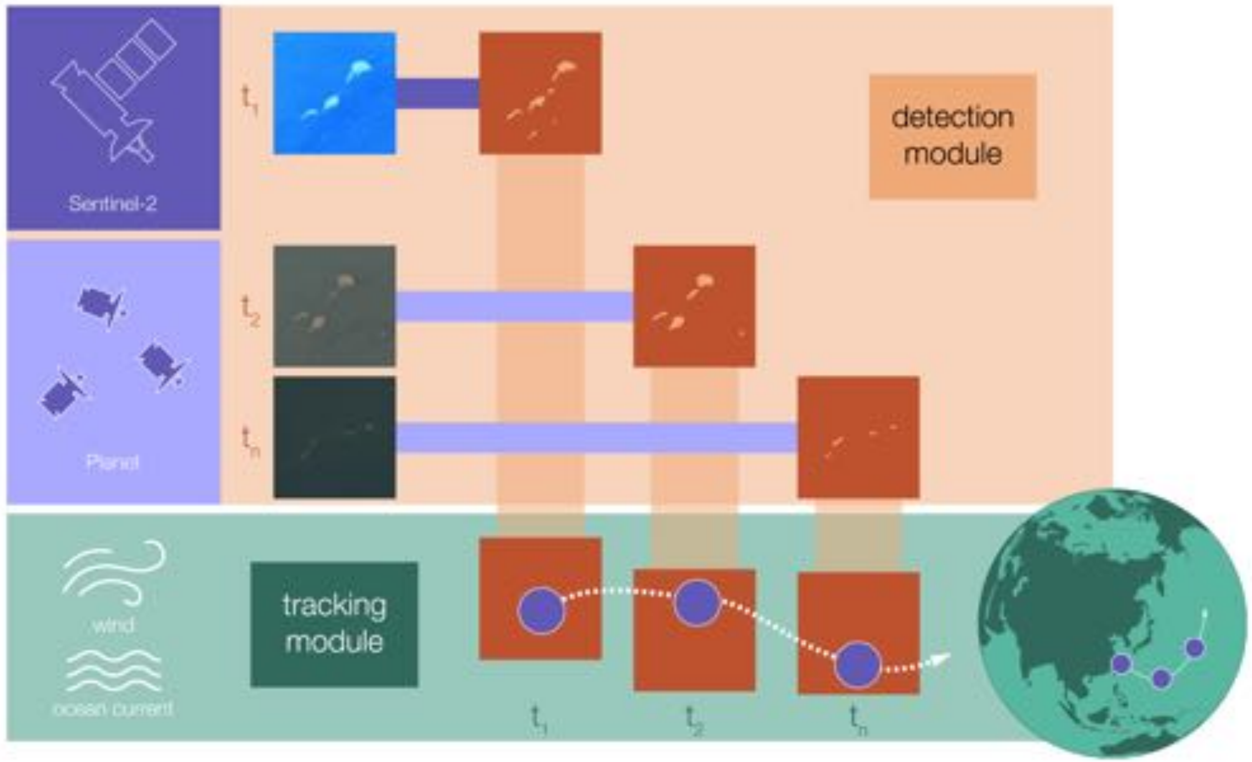
- Teaching models to adapt to new situations with just two clicks

- Using MAML

[Finn et al.]

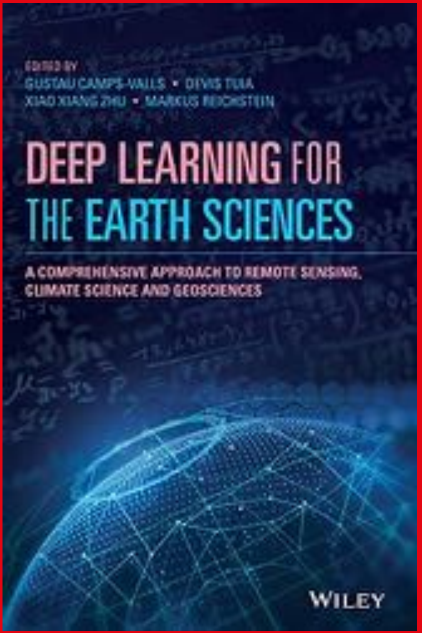


Building environmental deep learning models that are accurate and useful



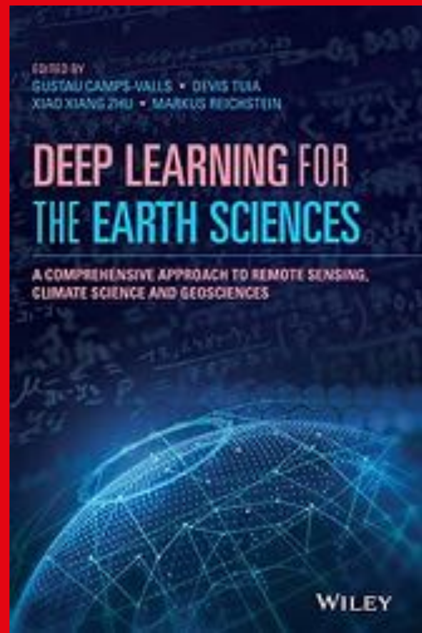
With Earth observation and AI, we can

develop computational approaches
to the environmental sciences
that are accurate



With Earth observation and AI, we can

develop computational approaches
to the environmental sciences
that are accurate, but also
scalable,
knowledge-driven and
accessible to everyone.



Towards environmental deep learning that is

Accurate

Scalable

Knowledge-driven

Accessible to anyone

EPFL

Scalable

[Mifdal et al., 2020]

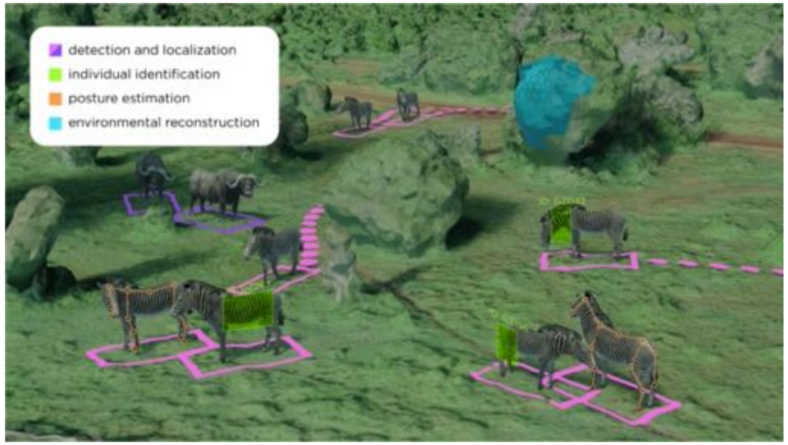
- No model should work only on
  - one image
  - one region of the world
  - one task

Mifdal, J., Longépé, N., and **Rußwurm, M.**: Towards detecting floating objects on a global scale with spatial features using Sentinel-2, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-3-2021, 285–293,

Scalable

- No model should work only on
  - one image
  - one region of the world
  - one task

[Mifdal et al., 2020]



D. Tuia, **B. Kellenberger**, S. Beery, B. Costelloe, S. Zuffi, B. Risse, A. Mathis, M. W. Mathis, F. van Langevelde, T. Burghardt, R. Kays, H. Klinck, M. Wikelski, I. D. Couzin, G. van Horn, M. C. Crofoot, C. V. Stewart, and T. Berger-Wolf. Perspectives in machine learning for wildlife conservation. *Nature Comm.*, 13(792), 2022.

Coral reefs

cover only 0.1% of oceans, but host 25% of all marine life.

protect coastlines, generate revenue (tourism, fishing, ...).

in 50 years, we managed to kill half of them.

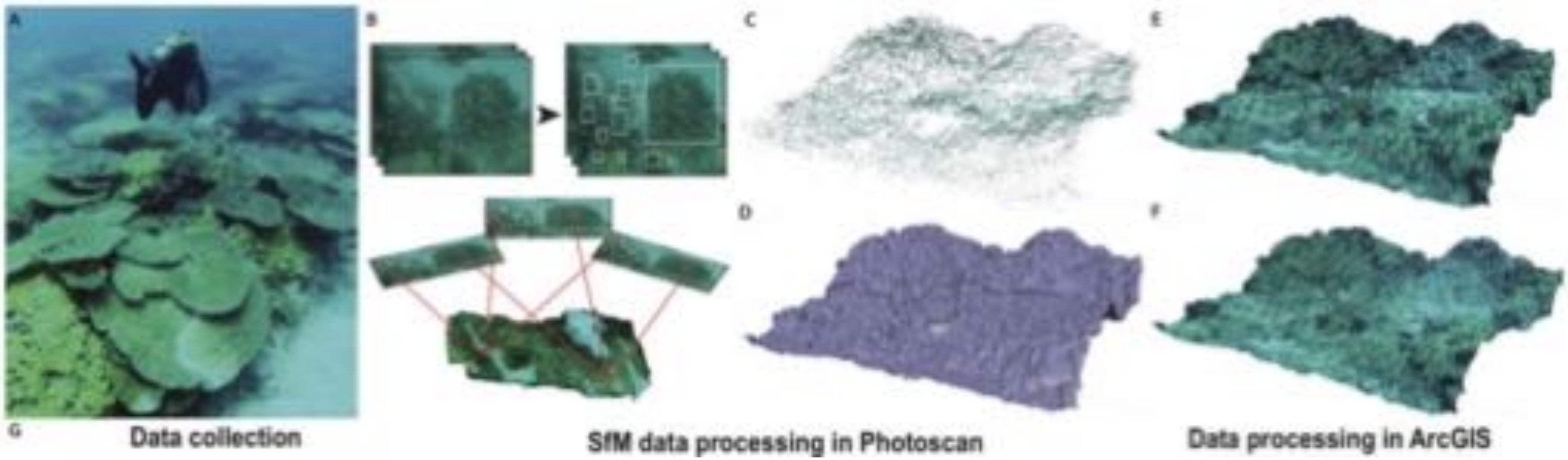

How do we monitor a large ecosystem like that?

When the technology does not scale well, monitoring is difficult.

28 x 6 m plot

1h with proprietary software

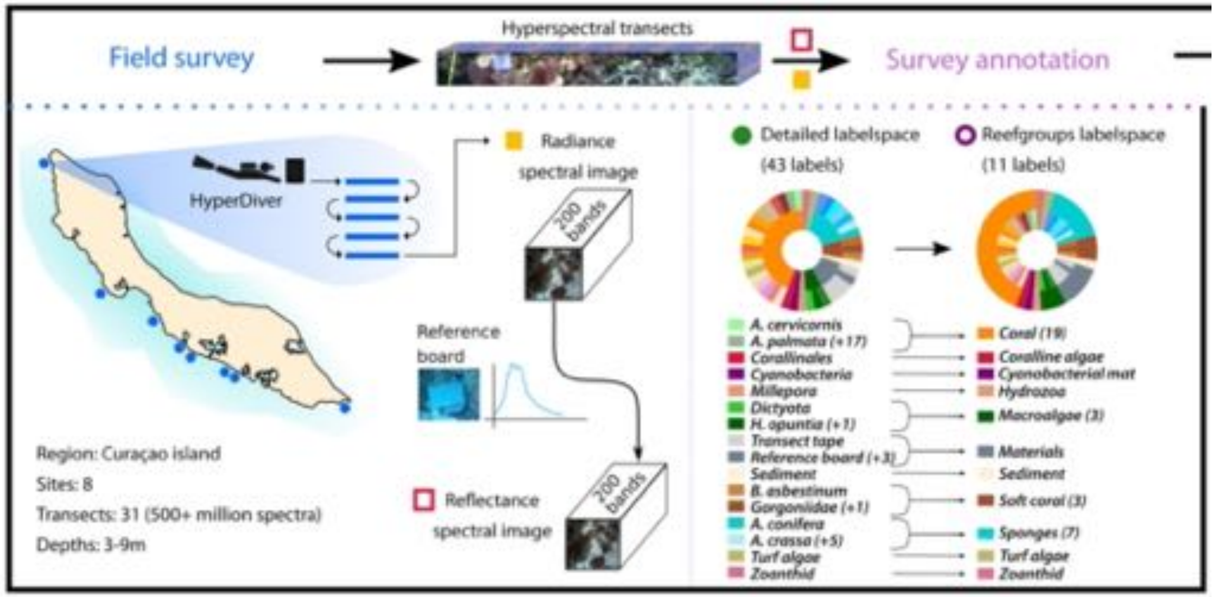
20h manual work
to extract information



[Burns et al., 2015, https://peerj.com/articles/1077/]

When the setup is unique: great results, but difficult to apply elsewhere

- Published models often rely on complex setups, very expensive
- E.g. hyperspectral sensors



Schürholz and Chennu, Methods in Ecology and Evolution, 2022

Our bet: affordable setups

- Scalable to other reefs
- Easy to acquire / replace
- Can train local communities

Mark I: March 2022 – Isreal / Jordan

Mark II: August 2022 – Djibouti

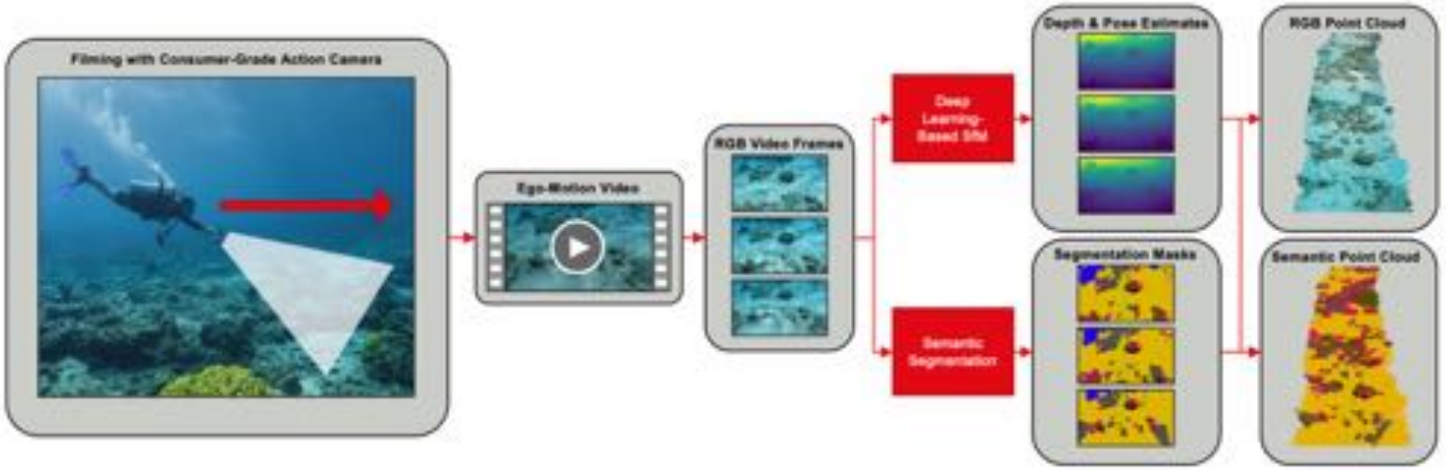
Photo credits: Guilhem Banc-Prandi, 2022

EPFL

Transnational Red Sea Center
Bridging Science & Diplomacy for the Future of Corals

Enabling scalable reef monitoring: Open source, fast, large scale.

- With custom-built, affordable imaging setup
- A model that works on videos, leveraging 2 tasks
- Once trained, 3D reconstructs a 100m transect in playing video time
- Tested in Isreal, Jordan and Djibouti in 2022

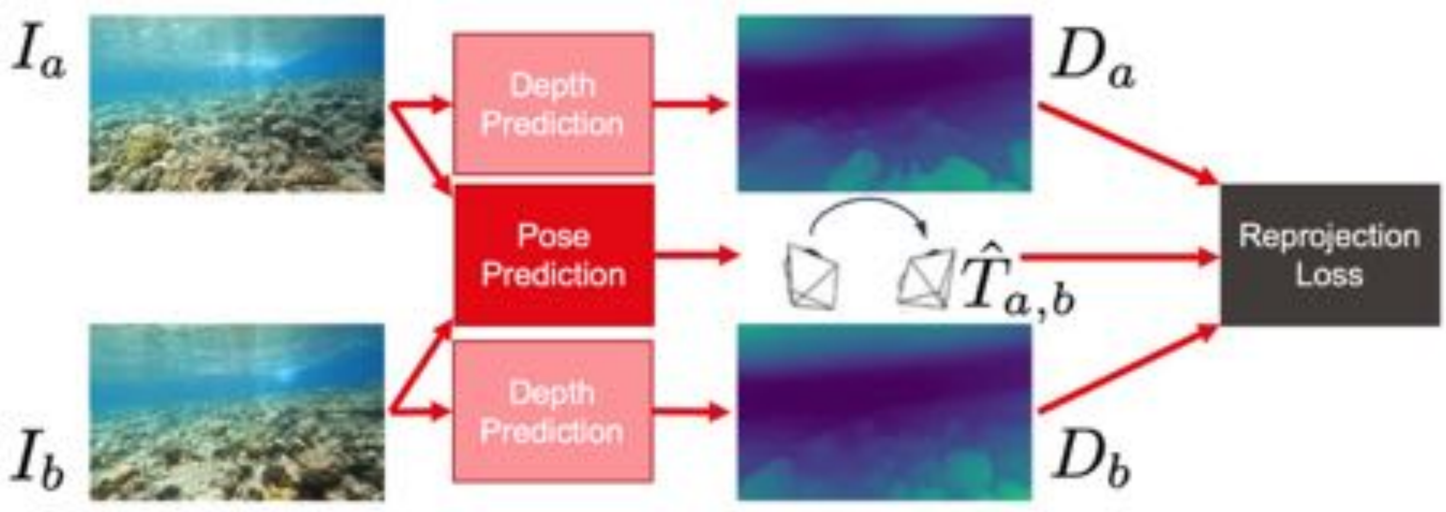


J. Sauder, G. Banc-Praudi, A. Meibom, and **D. Tuia**. Scalable semantic 3d mapping of coral reefs with deep learning. *Under review*.

EPFL

Pose and depth estimation

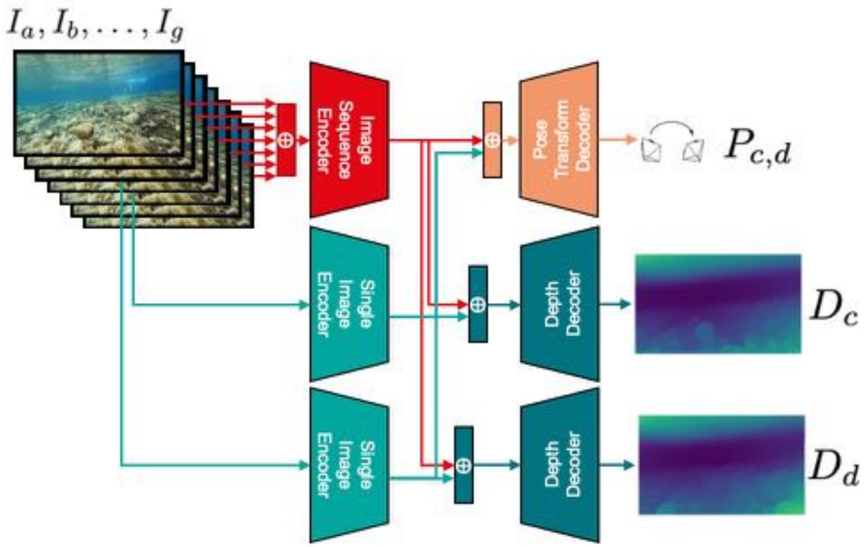
- Encoders based on ResNet-34
- Can create the 3D map at 18 frames per second



J. Sauder, G. Banc-Prandi, A. Meibom, and **D. Tuia**. Scalable semantic 3d mapping of coral reefs with deep learning. *Under review*.

Pose and depth estimation

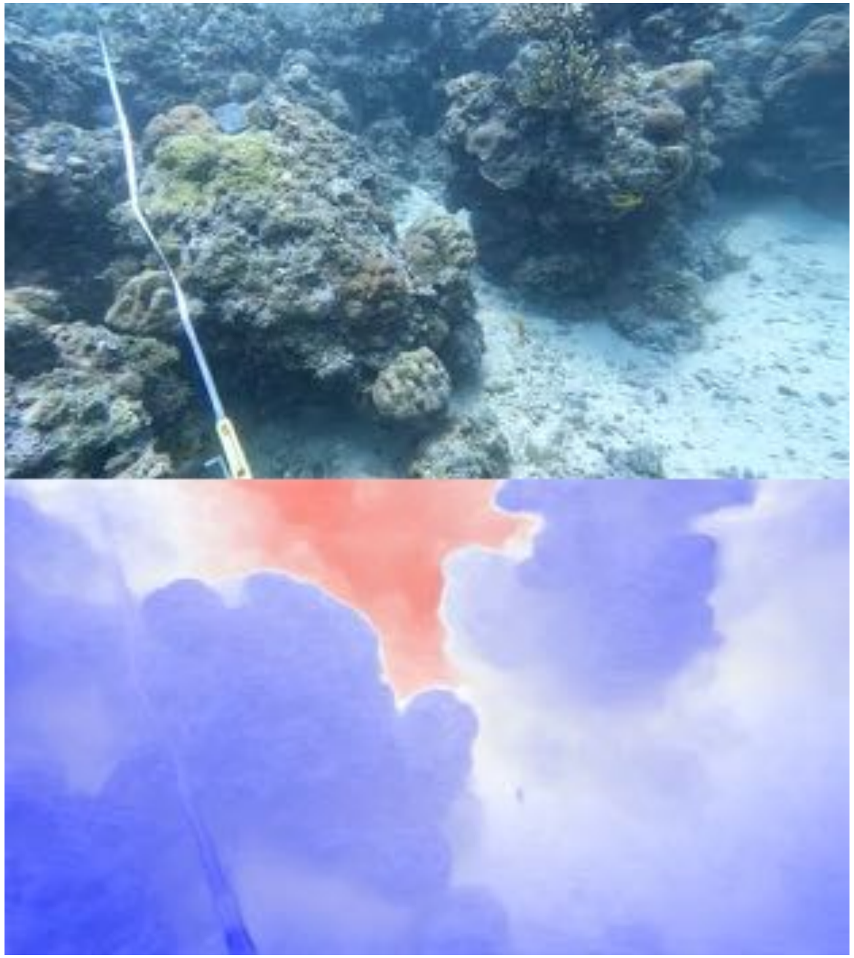
- We enforce temporal consistency by encoding 7 frames at a time
- 3 before and 3 after
- Each decoder uses the frame + the sequence features



Pose and depth estimation

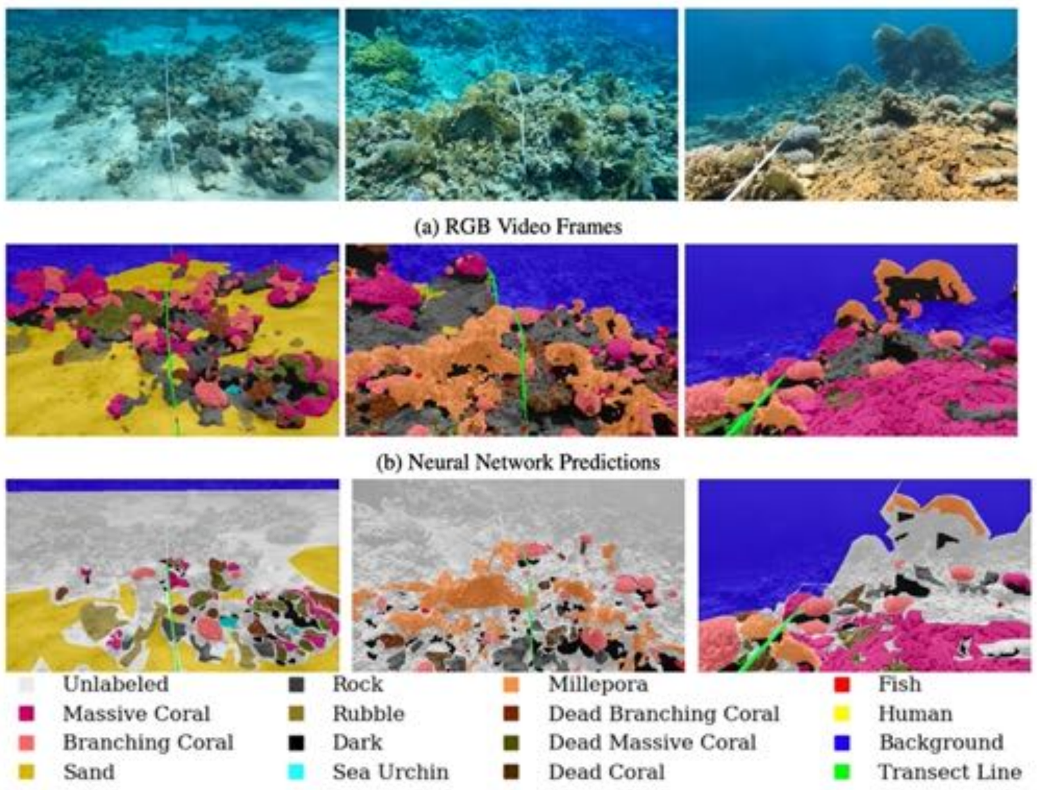
- We enforce temporal consistency by encoding 7 frames at a time
- 3 before and 3 after

- Each decoder uses the frame + the sequence features



Semantic segmentation

- Unet with ResNeXt backbone
- ~85% accurate in Jordanian and Israeli reefs

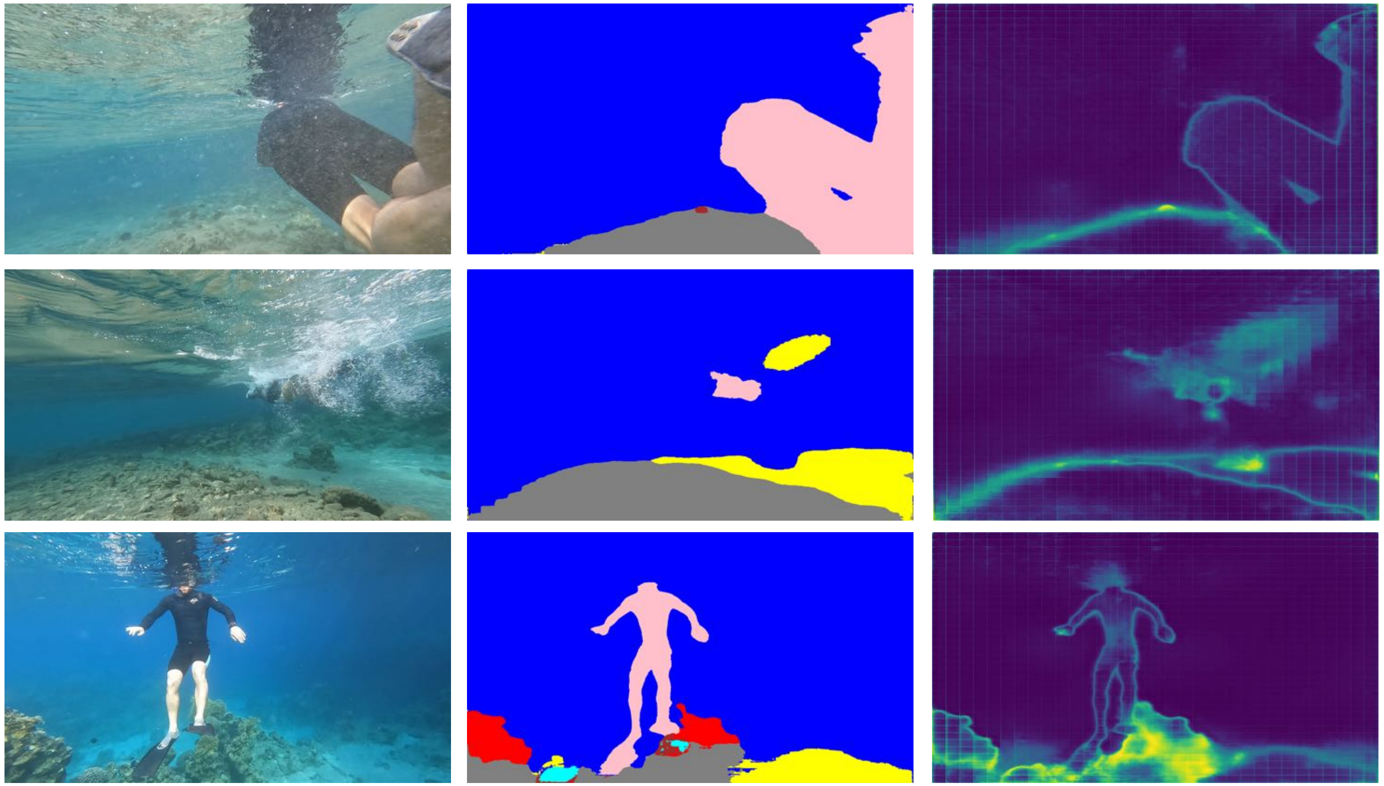


J. Sauder, G. Banc-Prandi, A. Meibom, and **D. Tuia**. Scalable semantic 3d mapping of coral reefs with deep learning. *Under review*.

Learning to detect unwanted classes

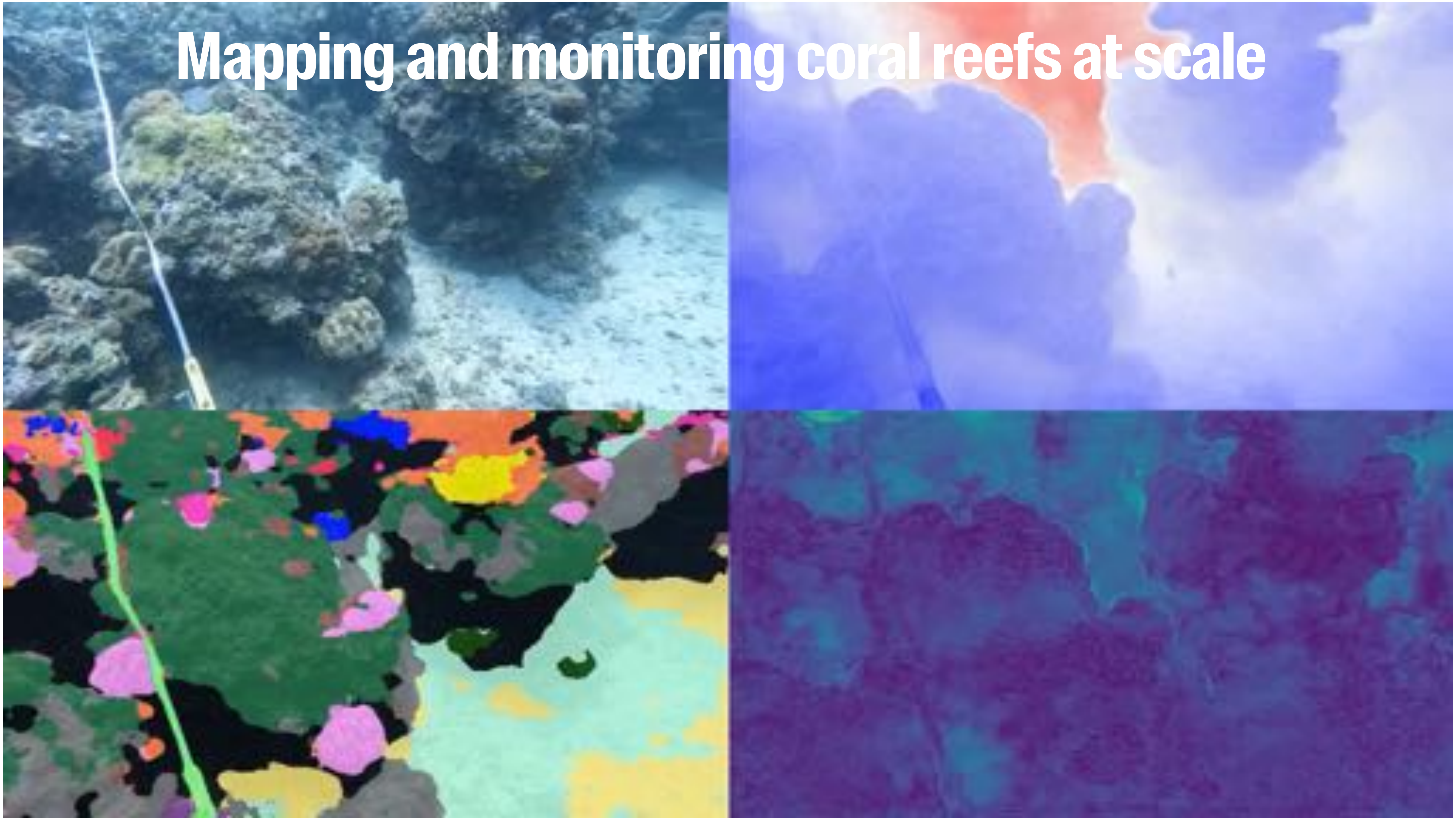
Used to remove unwanted classes prior to 3D reconstruction

- Diver body
- Fishes
- Far away pixels

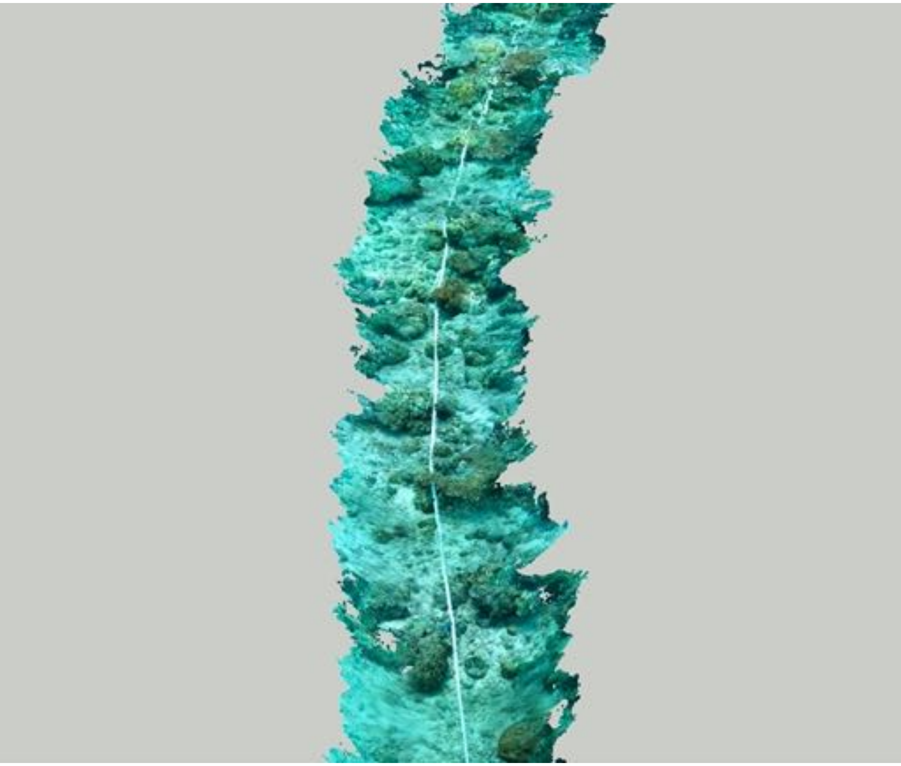


Video frame Segmentation Uncertainty

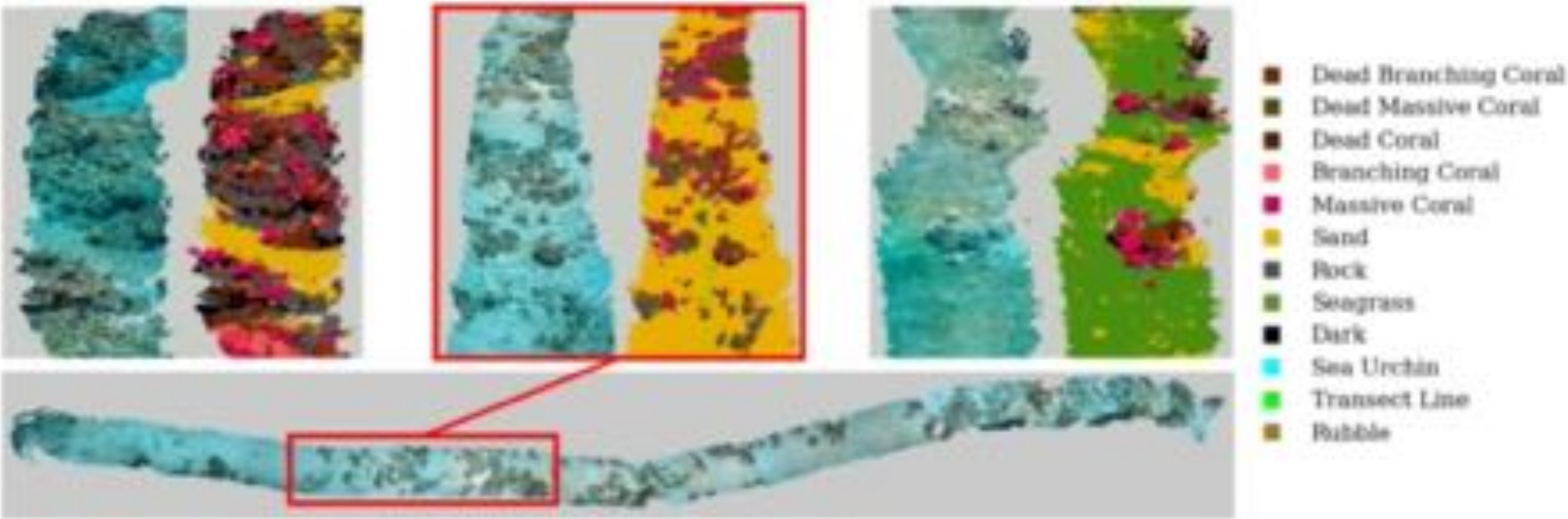
Mapping and monitoring coral reefs at scale



The multitask model allow us to create reliable 3D reconstructions of the reef



Mapping entire dive sites
(here: 100m long)



Towards environmental deep learning that is

Accurate

Scalable

Knowledge-driven

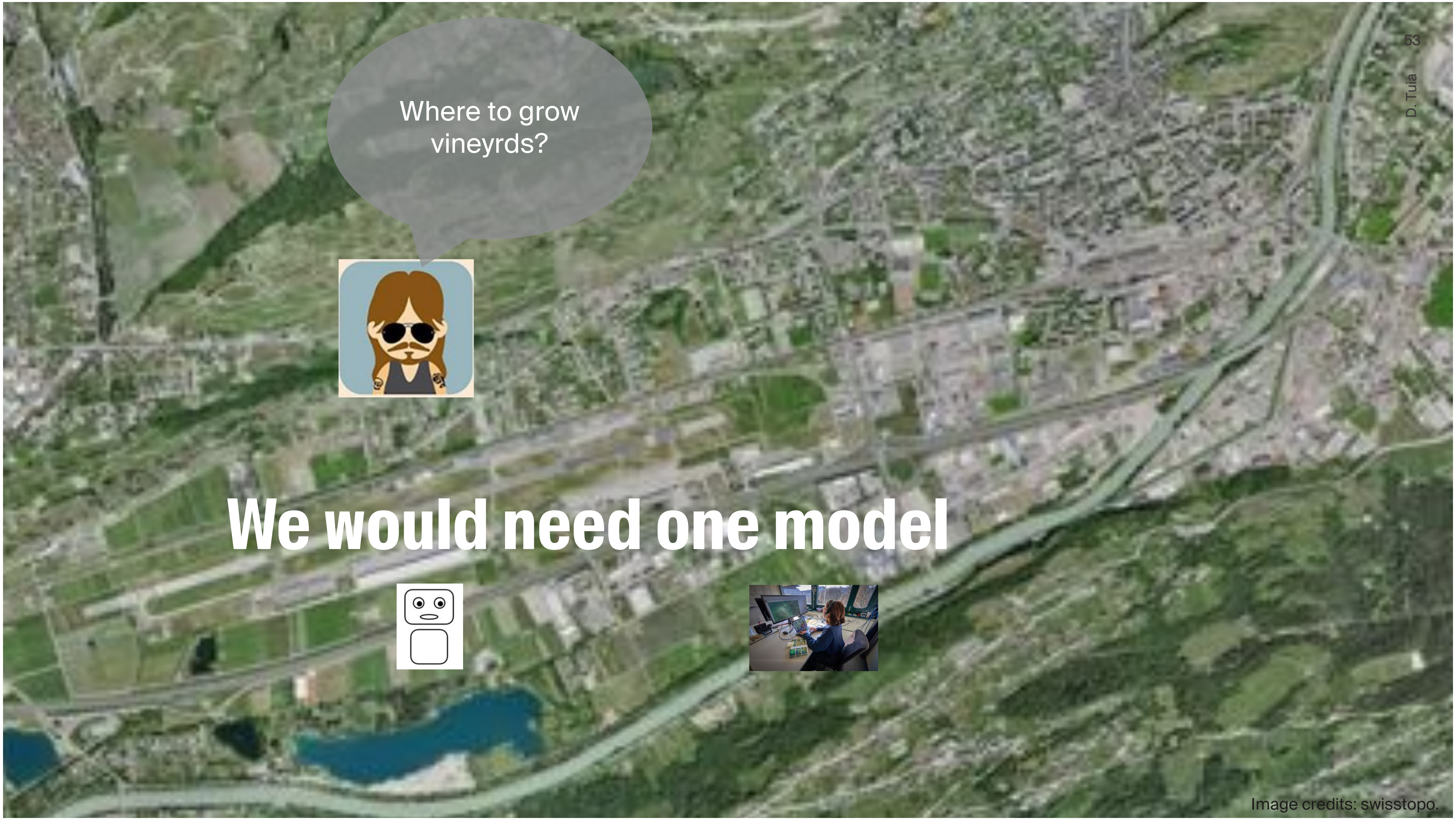
Accessible to anyone

Are our models useful and used?

- Availability of computer vision models makes it easier to use ML in our domain.
- But still we have a series of throwbacks

Source: dreamstime.com

Image credits: swisstopo.


Where to grow vineyrds?
We would need one model
Image credits: swisstopo.

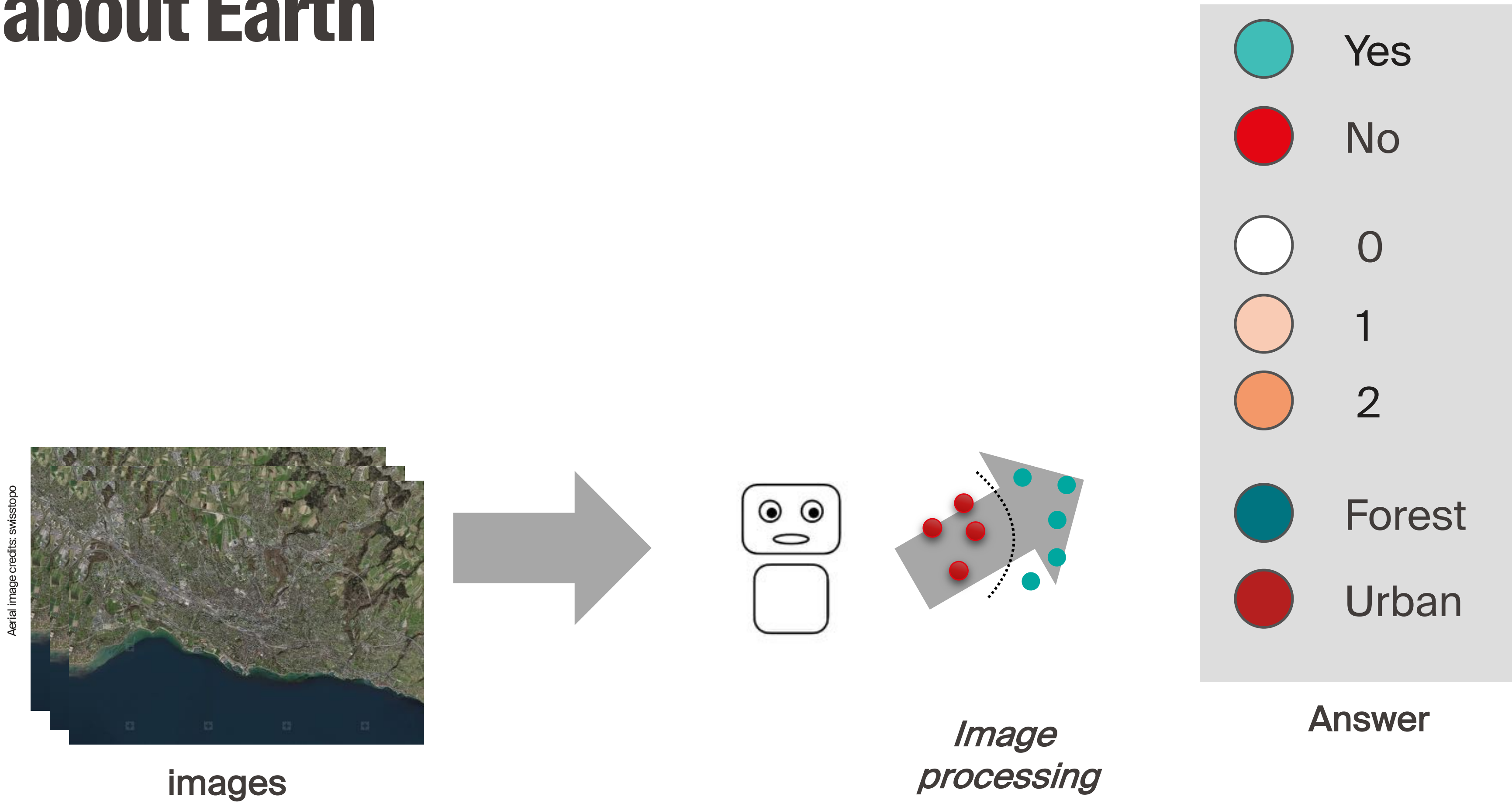

54
D. Tuia
Where to grow vineyrds?
What is the most friendly neighborhood?
We would need one model per theme
Image credits: swisstopo.

One model to answer them all

Problem 2
the tech divide

🔎 *Did forests in Lausanne grow?*

RSVQA: Answering questions about Earth

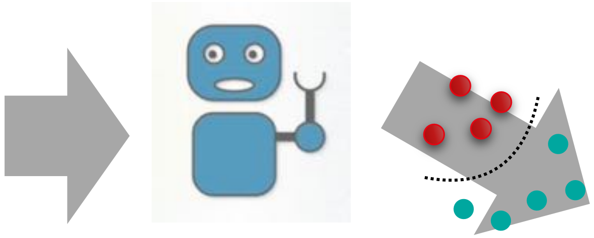


C. Chappuis, **V. Zermatten**, S. Lobry, B. Le Saux, and **D. Tuia**. Prompt–RSVQA: Prompting visual context to a language model for remote sensing visual question answering. In *CVPRW*, New Orleans, LA, 2022.

RSVQA: Answering questions about Earth

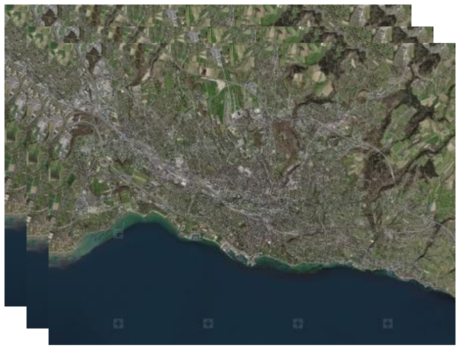
🔎 *Did forests in Lausanne grow?*

question



Language model

Aerial image credits: swisstopo



images

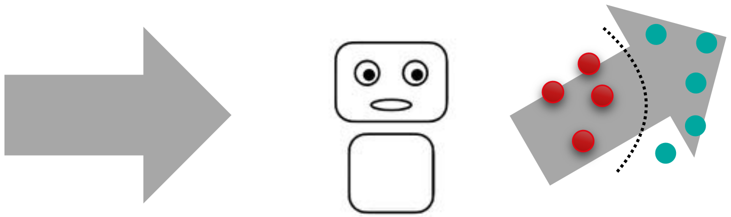
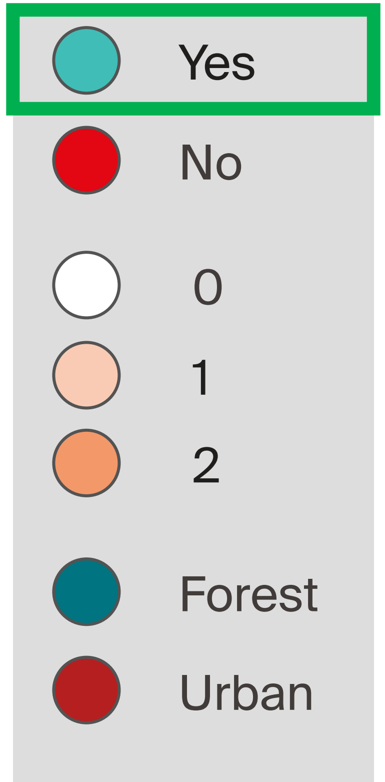


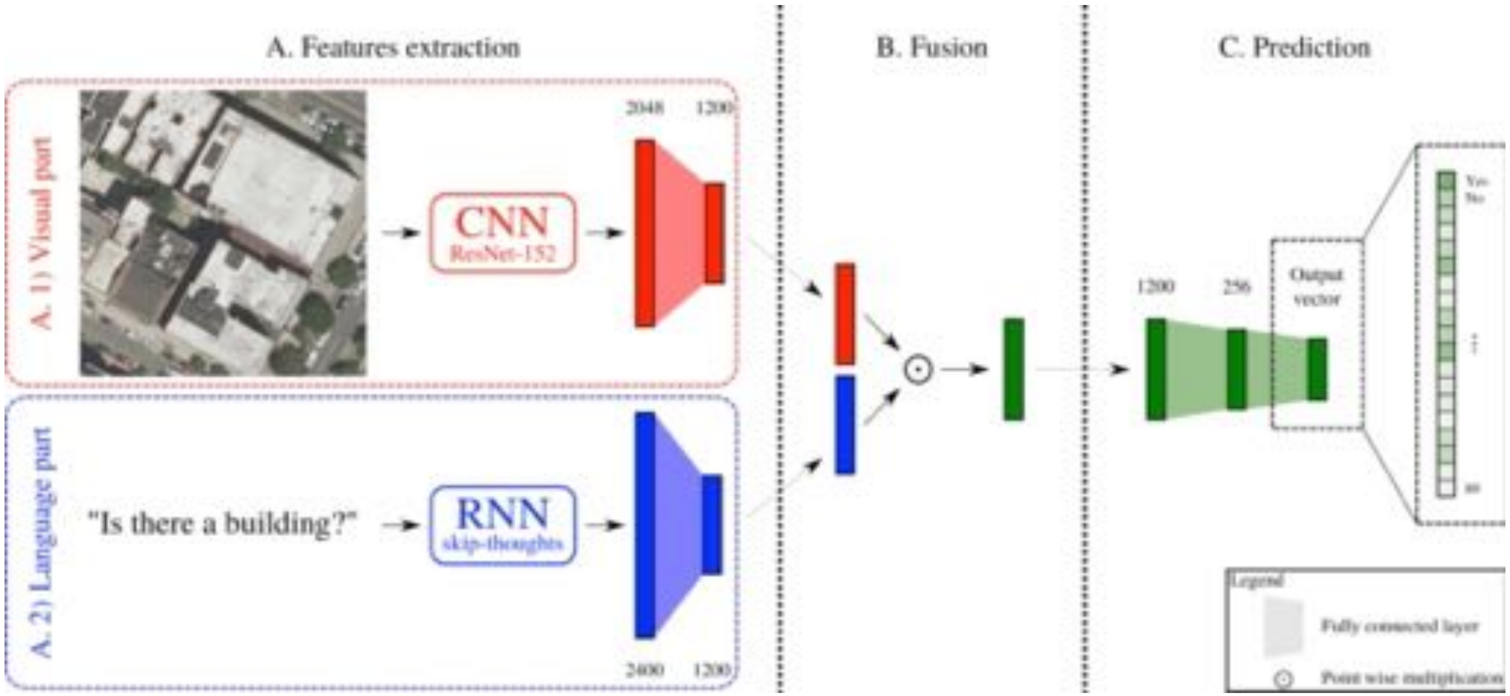
Image processing



Answer

C. Chappuis, V. Zermatten, S. Lobry, B. Le Saux, and D. Tuia. Prompt–RSVQA: Prompting visual context to a language model for remote sensing visual question answering. In *CVPRW*, New Orleans, LA, 2022.

Remote sensing visual question answering



[Lobry, Marcos, Murray, Tuia, IEEE TGRS 2020, https://arxiv.org/abs/2003.07333]

11’000
Aerial images
New York / Philadelphia

x

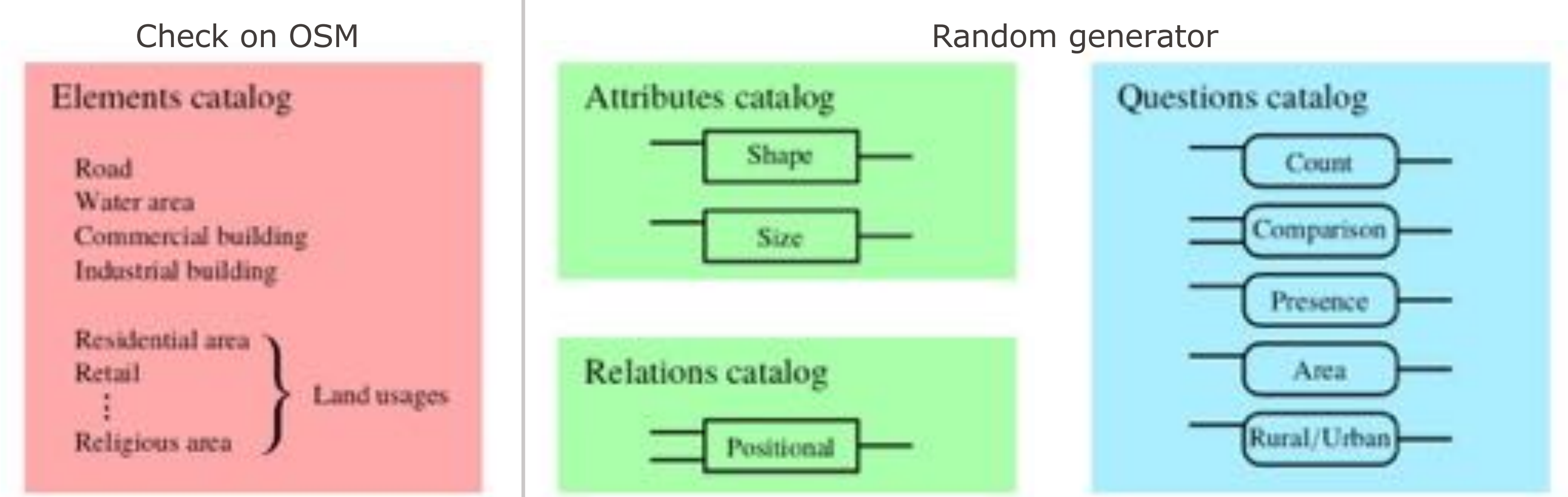
= 1’100’000 examples

11’000
Aerial images
New York / Philadelphia

100
Question
Answers

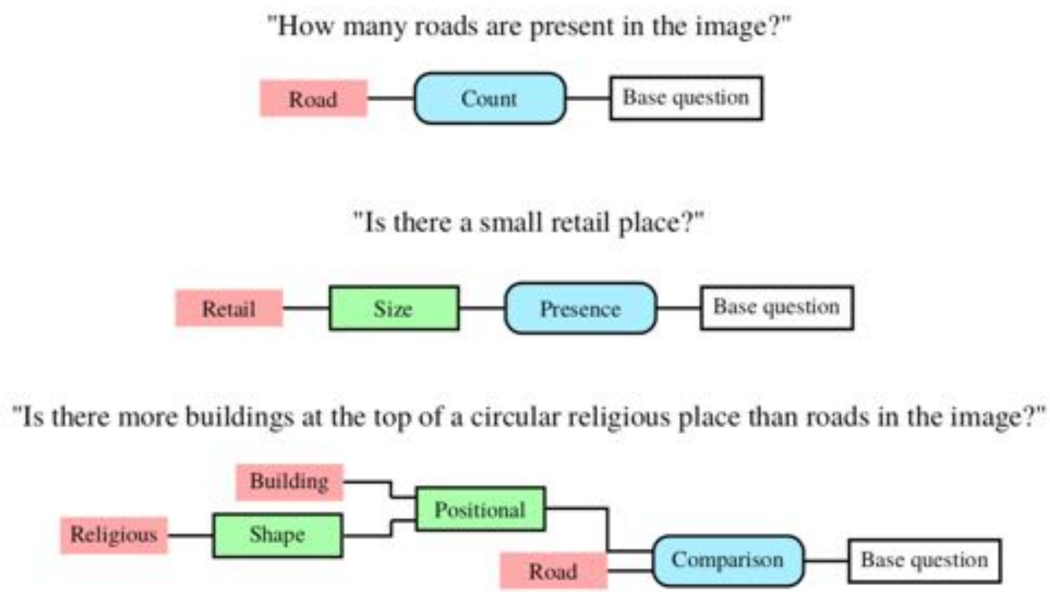
How did we generate the questions / answers?

- We generated {image, question, answer} triplets



How did we generate the questions / answers?

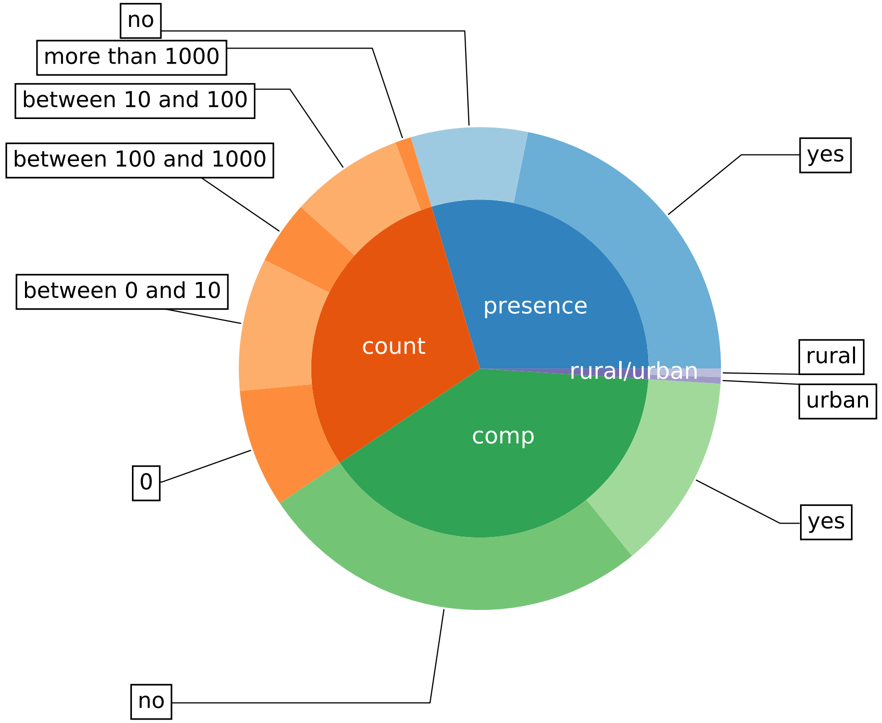
- We generated {image, **question**, answer} triplets



How did we generate the questions / answers?

- We generated {image, question, **answer**} triplets

- We again use OSM



EPFL

Results

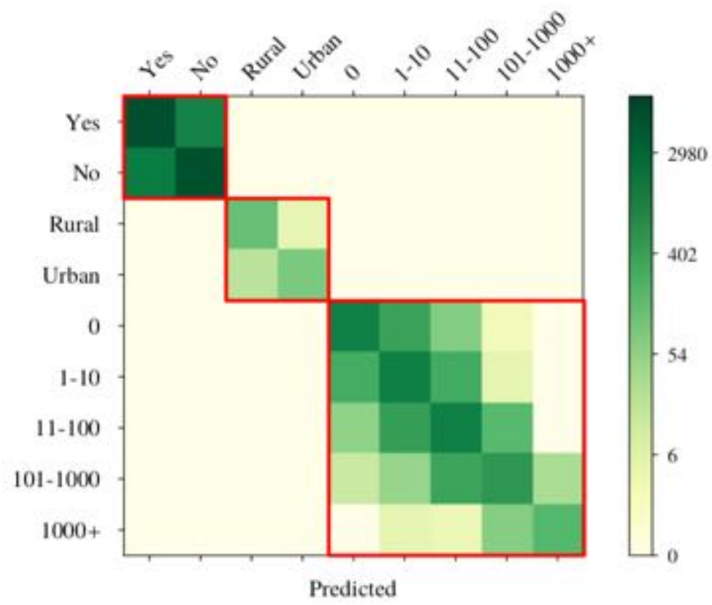
83% overall accuracy!

73% if randomizing the image part

Count questions less accurate

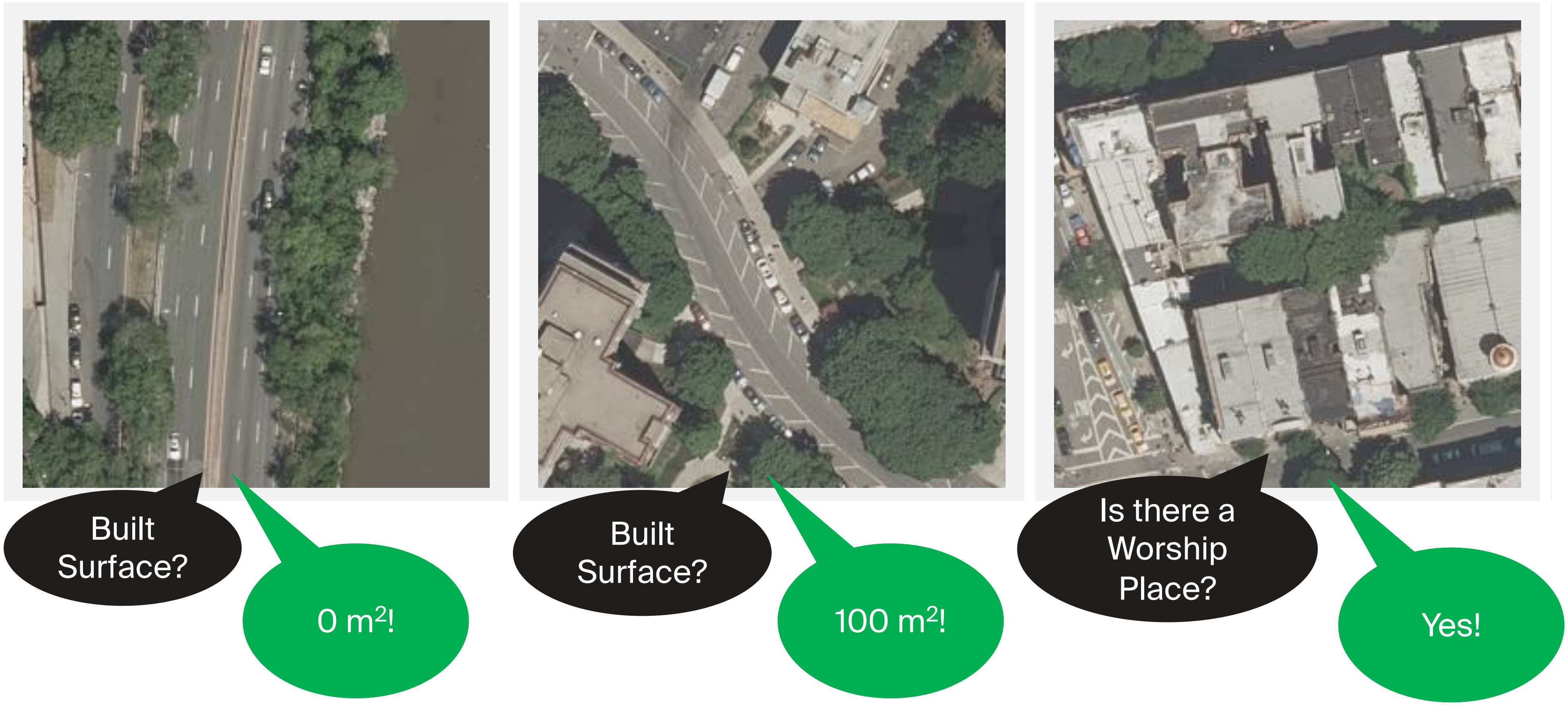
Type	New York	Philadelphia
Count	68.63% (0.11%)	61.47% (0.08%)
Presence	90.43% (0.04%)	86.26% (0.47%)
Comparison	88.19% (0.08%)	85.94% (0.12%)
Area	85.24% (0.05%)	76.33% (0.50%)
AA	83.12% (0.03%)	77.50% (0.29%)
OA	83.23% (0.02%)	78.23% (0.25%)

The model can make a good distinction between types of questions



From Lobry et al., IEEE TGRS 2020, https://arxiv.org/abs/2003.07333

Results



From Lobry et al., IEEE TGRS 2020, https://arxiv.org/abs/2003.07333

Results

How many
Buildings?

0!

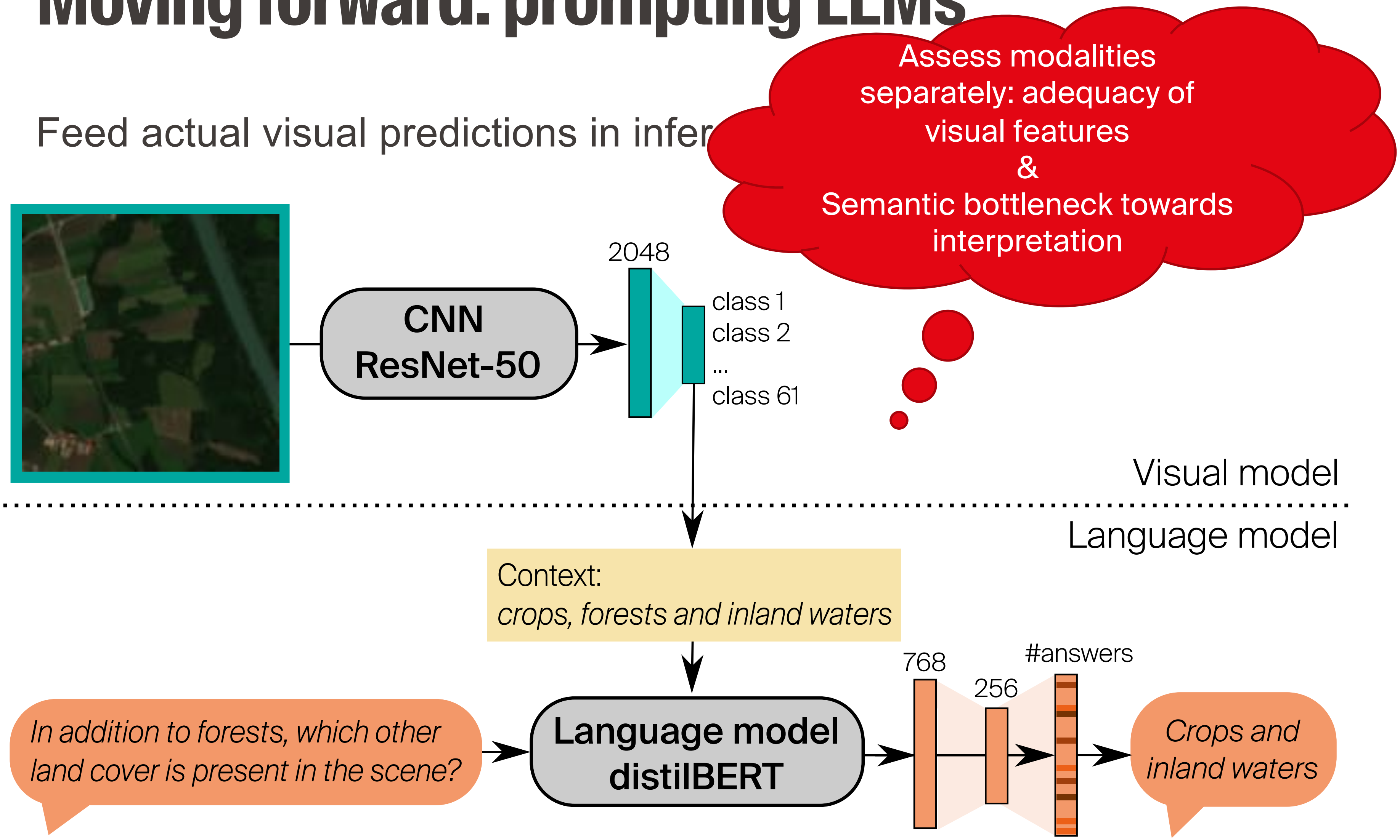
How many
buildings?

1!

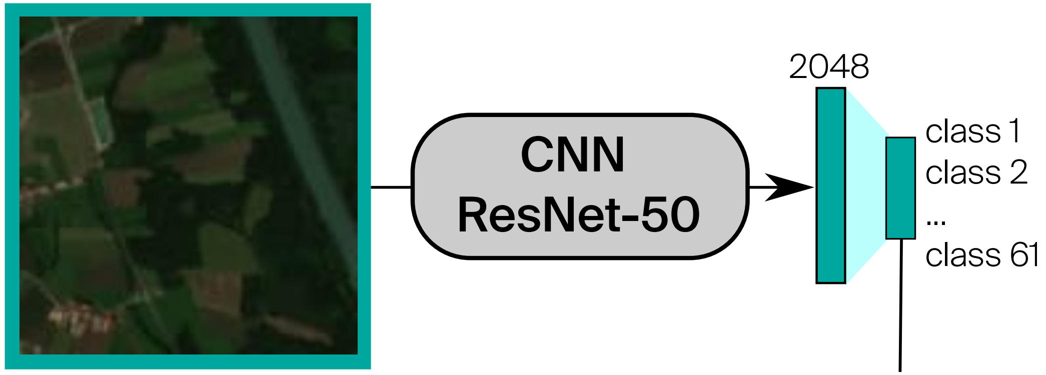
From Lobry et al., IEEE TGRS 2020, https://arxiv.org/abs/2003.07333

Moving forward: prompting LLMs

Feed actual visual predictions in infer



1. The visual part.



Learning the visual model

- Images: Sentinel 2 data
  - EU satellite, 10m resolution
  - Data every 5 days, free
- Labels (→)
  - Corine Land Cover
  - Updated every 10 years, EU-wise
  - 61 classes
- Model
  - ResNet50 from BigEarthNet *
  - Multi-label classification
  - **Predictions passed as words to the LLM**

* Sumbul, G., et al., BigEarthNet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. IGARSS 2019

2. prompting a language model

- Manipulate inputs instead of weights

- Add visual keywords to input text

Question:
In addition to forests, which other land cover is present in the scene?

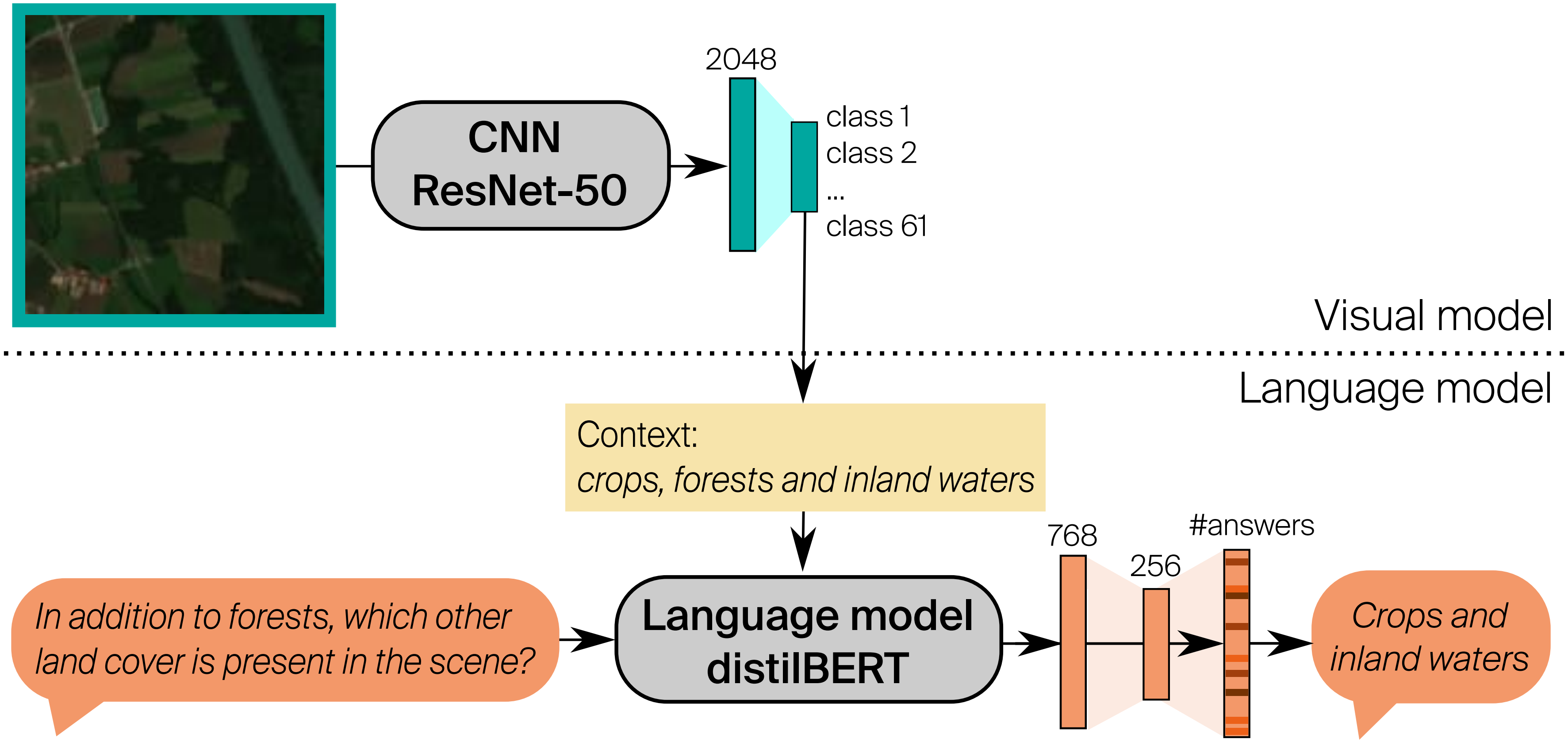
+

Context:
Land covers are crops, inland waters and forests.

- Light fine-tuning for smaller language models

Moving forward: prompting LLMs

Feed actual visual predictions in inference to a LLM



Experiment - proof of concept

RSVQA meets BigEarthNet

- From BigEarthNet, and the CORINE Land Cover inventory
- Sentinel-2 images
- 2 types of questions: land cover and yes/no

[G. Sumbul, et al., "BigEarthNet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding", IGARSS, 2019.]

Results on the final task RSVQA

Method	Accuracy		
	Global *	Yes / No	Land cover
Visual oracle	98.81%	99.90%	93.79%
Visual blind	65.36%	75.85%	17.30%
RSVQA (Lobry et al. 2021)	69.83%	79.92%	20.57%
Prompt-RSVQA (ours)	**75.40%**	**86.07%**	**26.56%**

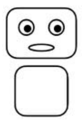
+ 6% !

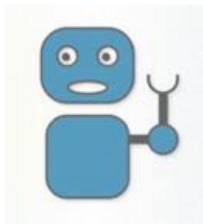
* Restriction of answer space to 1000 answers: 98.9% maximal performance on test set

Better informed language models

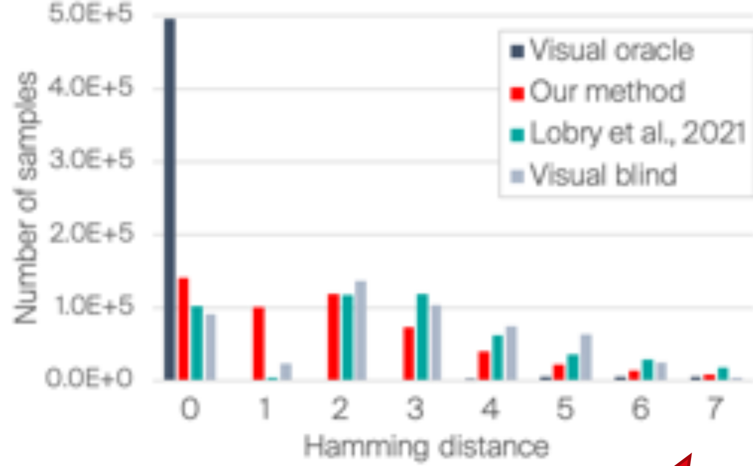
Question:
In addition to forests, which other land cover is present in the scene?

Context:
Land covers are crops, inland waters and forests.





Language model answers: «crops and inland waters»



6% improvement!
"Better errors"!

C. Chappuis, V. Zermatten, S. Lobry, B. Le Saux, and **D. Tuia**. Prompt-RSVQA: Prompting visual context to a language model for remote sensing visual question answering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, 2022.

EPFL

Language understanding for enhanced image search

A sparse residential area with a villa surrounded by forest?

Knowledge-based satellite image retrieval

[Mi et al., IJCAI workshops 2022]

Images of animals with poor body conditions

Animal health and behavior

[Gabeff et al., CVPR workshops 2023]

parc naziunal svizzer

My view on Remote sensing and AI

Advance remote sensing science to monitor and protect Earth

Interface disciplines and approaches

Bring new, open tools making EO science accessible to anyone

EPFL

eceo EPFL

@devistuia

My view on Remote sensing and AI

Advance remote sensing science to monitor and protect Earth

Interface disciplines and approaches

Bring new, open tools making EO science accessible to anyone