Leveraging Transformers, Vision and Multimodal Models to support SDGs

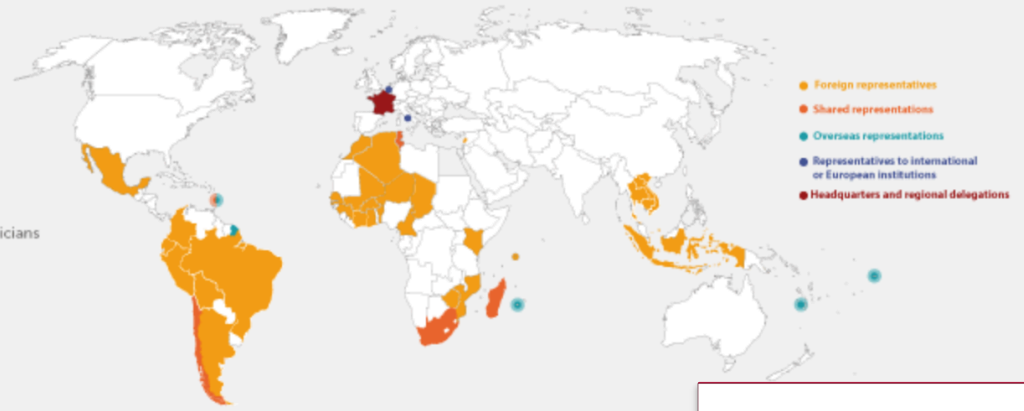
Laure Berti-Equille

IRD ESPACE-DEV
Montpellier, France
laure.berti@ird.fr
https://laureberti.github.io/website/

IRD: the French Research Institute on Sustainable Development

80 years of multidisciplinary research

In 2023



2,306 AGENTS

including **855** researchers and **1,194** engineers and technicians

52% MEN **48%** WOMEN

24% OF AGENTS WORKING OUTSIDE MAINLAND FRANCE

157 SOUTH-NORTH MOBILITIES

64% OF CO-PUBLICATIONS WITH A PART

5,621 PUBLICATIONS OF THE IRD RESE

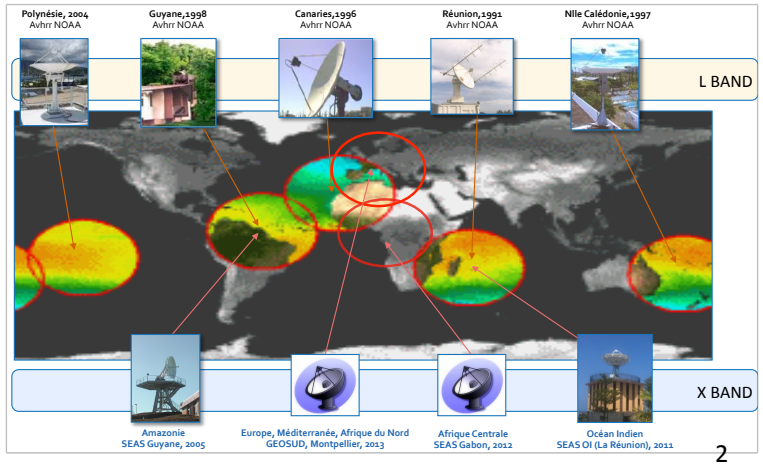
80 RESEARCH UNITS

98 SOUTHERN RESEARCH TEAMS

€ **277 M (2023)**

162 PhD students

RESEARCH CONTRACTS RESOURCES: € **49,5 M (2023)**



2030 Sustainable Development Goals (SDGs)

2030 Agenda for Sustainable Development: 17 goals, 169 targets, 232 Indicators

New norms to integrate the principles of sustainable development into country policies and programs

https://sdgs.un.org/2030agenda

https://sdgs.un.org/goals

Example of SDG#1 Targets

UN TARGETS AND INDICATORS FOR GOAL 1		
	Targets	**Indicators**
	1.1 By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day.	1.1.1 Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural).
	1.2 By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions.	1.2.1 Proportion of population living below the national poverty line, by sex and age. 1.2.2 Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions.
	1.3 Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable.	1.3.1 Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable.
	1.4 By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance.	1.4.1 Proportion of population living in households with access to basic services. 1.4.2 Proportion of total adult population with secure tenure rights to land, (a) with legally recognized documentation, and (b) who perceive their rights to land as secure, by sex and type of tenure.

https://knowsdgs.jrc.ec.europa.eu/sdg/1

Applying ML to SD

Transdisciplinary research, reasoning, and discovery from interconnected multimodal information

Data Engineering

- Information extraction
- Data linking, integration
- Multiscale data fusion
- Data curation/preparation
- Data analytics
- …

Model Engineering

- Feature engineering
- Spatiotemporal & dynamic models
- Uncertainty quantification
- Constraint reasoning
- Optimization
- …
-

Citizen Science

- Crowd sensing/sourcing
- Behavioral game theory
- HITL
- …

Requires constant interaction with a wide range of non-CS disciplines: Earth, Environmental, Social sciences, etc.

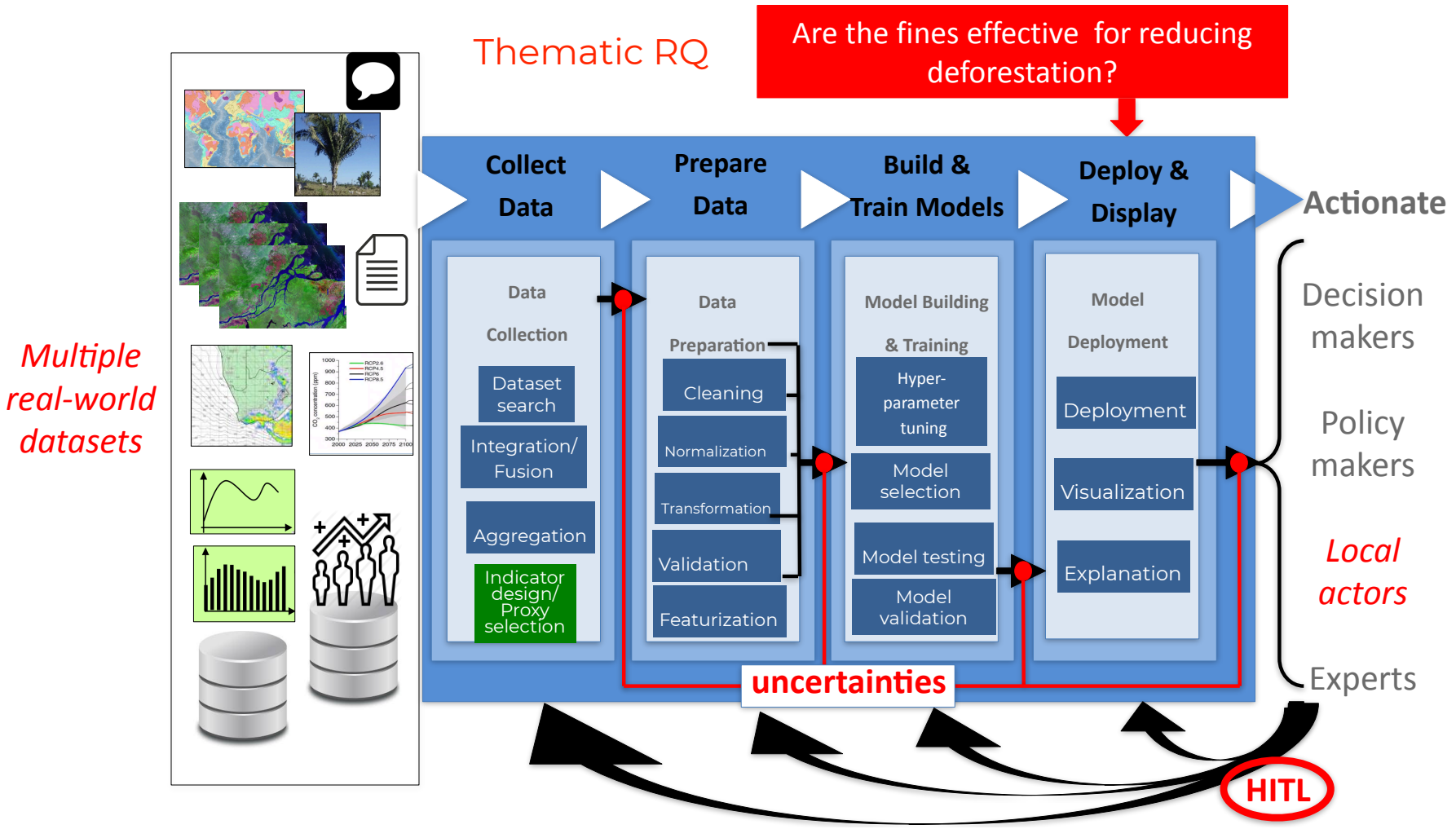
SDGs as Optimization Problems

- Maximizing the probability of achieving an SD target
- Minimizing the degradations of environmental and human conditions

Complexity:

- Multi-objective: improve the quality of human life, preserve the Earth's diversity, minimize the depletion of non-renewable resources
- Multi-disciplinary: environmental, social sciences, etc.
- Multi-scale : global, national, regional, local
- Multi-actor: civil society, private companies, government

Generic Data Science Pipeline



Challenges of ML applied to SDGs

DATA

1. Data comes with various modalities, various spatial and temporal resolutions and scales
2. Representative ground truth and labeled data are scarce
3. Data is imperfect with data quality issues, uncertainty, label noise, bias...
4. Data is non I.I.D.

MODEL

5. It is easy to pick inappropriate features and inadequate preprocessing
6. Architecture search and hyperparameter optimization are hard

PEOPLE

7. It is difficult to conciliate non-interpretable ML features and domain-expert hand-made indicators
8. It is easy to generate non actionable and misleading outputs
9. It is difficult to measure social impact of ML-based solutions on the field

...

Outline

Introduction

- IRD Presentation
- SDGs and applied ML
- Building SD data science pipelines
- Main Challenges of ML Applied to SDGs

Overview of Our Research

- Estimate poverty evolution from satellite images using transformers
- Detection of favelas from satellite images using CROMA
- Automated annotation of coral reef images with hierarchical classification

Conclusions & Perspectives

Outline

Introduction

- IRD Presentation
- SDGs and applied ML
- Building SD data science pipelines
- Main Challenges of ML Applied to SDGs

Overview of Our Research

- Estimate poverty evolution from satellite images using transformers
- Detection of favelas from satellite images using CROMA
- Automated annotation of coral reef images with hierarchical classification

Conclusions & Perspectives

Overview of Our Research (1/3)

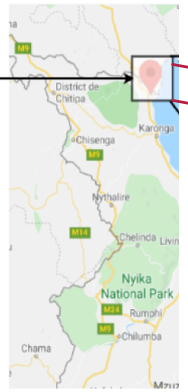
Estimate poverty evolution from satellite images using transformers

Ph.D. of Robin Jarry, co-supervised with Marc Chaumont & Gérard Subsol
(LIRMM), Montpellier, France
ANR MPA-Poverty

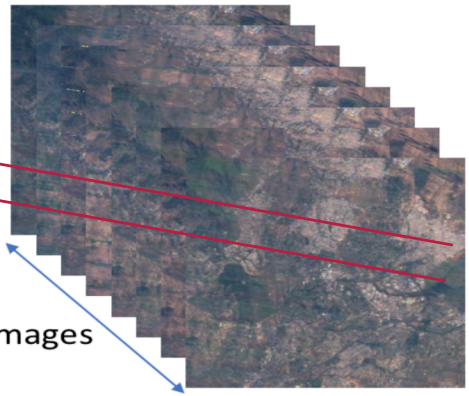
Main Idea: Mapping socioeconomic indicators with satellite image features

- Socioeconomic Indicators

Latitude	Longitud	Consumption expenditures ($/day)	Name of the village
-9.7298955	33.859230	1.8056488929595411	Kaporo
-9.715316	33.88666	1.4157905044417405	Karonga district
-11.95341	33.367064	1.957231017237896	Mzimba disctrict
-11.632151	34.235546	1.3677132387498063	Nkhata Bay



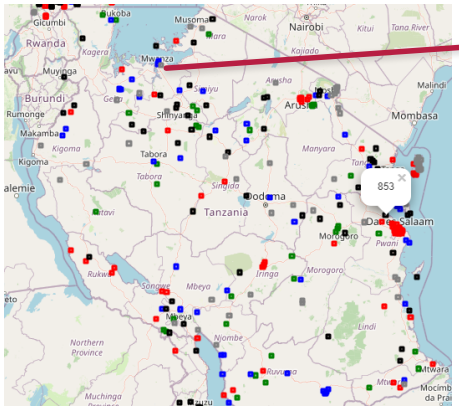
- Satellite Images



Deep Learning Model

Predictions

Socioeconomic Indicators



Census:

- ☑ Number of people
- ☑ Access to water/electricity
- ☑ Number of goods
- ☑ Agricultural lands
- ☑ Healthcare services
- ☑ Transportation

Construction of indicators [SMITS, 2015]

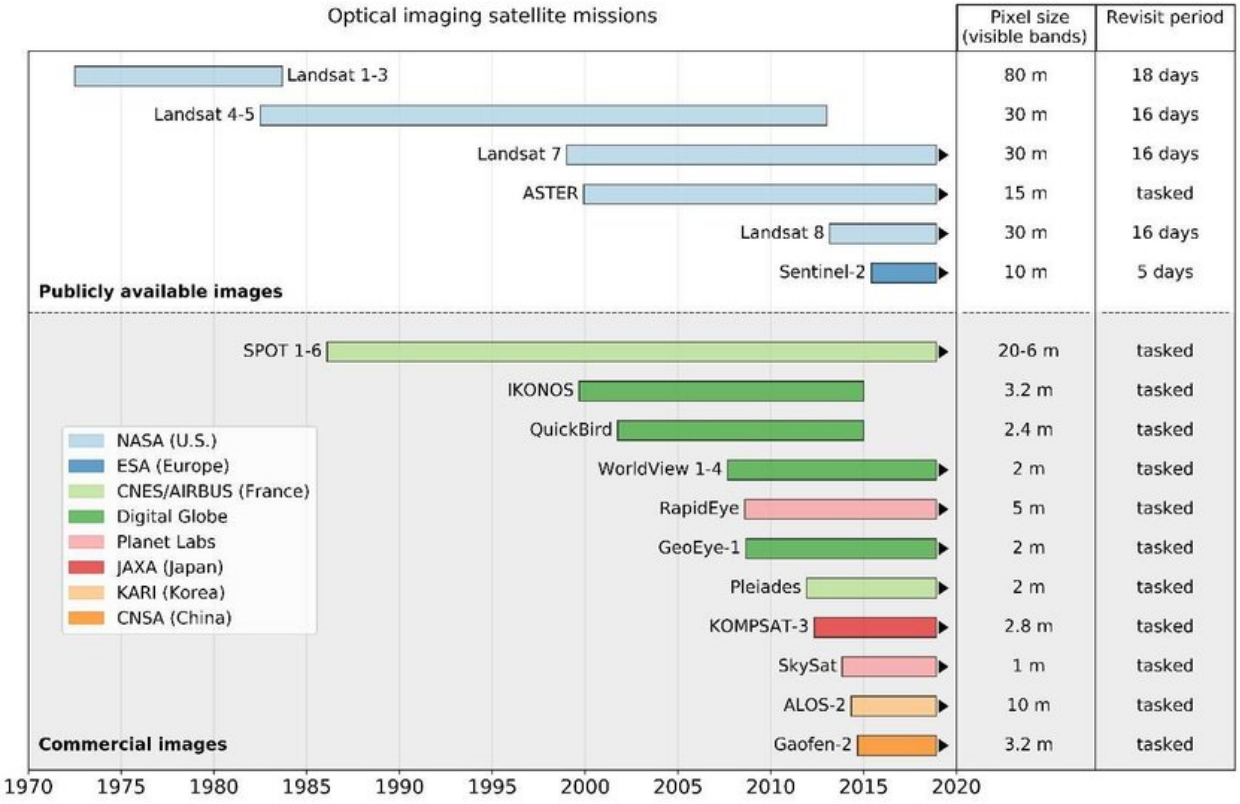
- DHS[1] and LSMS[2]
  - Tanzania: 17 surveys between 1992 and 2020
- Panel surveys: *revisit*:
  - Tanzanie: 4 surveys between 1992 and 2020
- Anonymization [BURGERT, 2013]
- Incertainty [BURKE, 2021]

Main Challenges

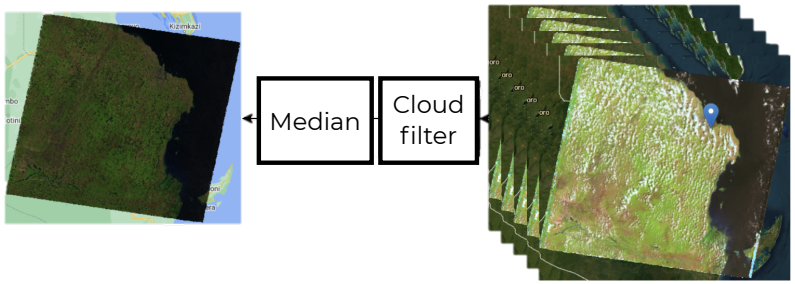
- **Sparse coverage in time and space**
- **Noise/uncertainty on the locations and indicator values**

Smits et al., (2015). Social Indicators Research. "The International Wealth Index (IWI)."
Burgert et al., (2013). Technical Report. "Geographic displacement procedure and georeferenced data release policy for the demographic and health surveys."
Burke et al., (2021). Science. "Using satellite imagery to understand and promote sustainable development."
[1]*Demographic and Health Surveys (DHS) USAID*
[2]*Living Standards Measurement Sudies (LSMS) World Bank*

Various Sources of Satellite Images



- Compromise:
  - Time period (> 20 ans)
  - Spatial Resolution (30 m)
  - Composite images

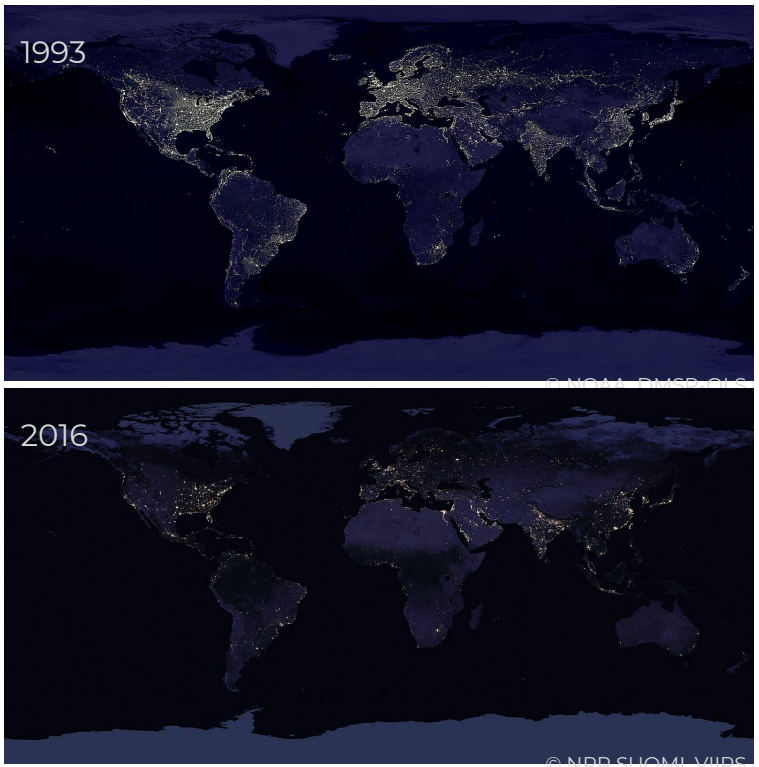


- Regular acquisitions
- Presence of clouds
- Various aggregation methods [Qiu, 2023]

Vos et al., (2019). Coastal Engineering. "Sub-annual to multi-decadal shoreline variability from publicly available satellite imagery."

Qiu et al., (2023). Remote Sensing of Environment. "Evaluation of Landsat image compositing algorithms. Remote Sensing of Environment."

NLI: A Proxy of Economic Activity



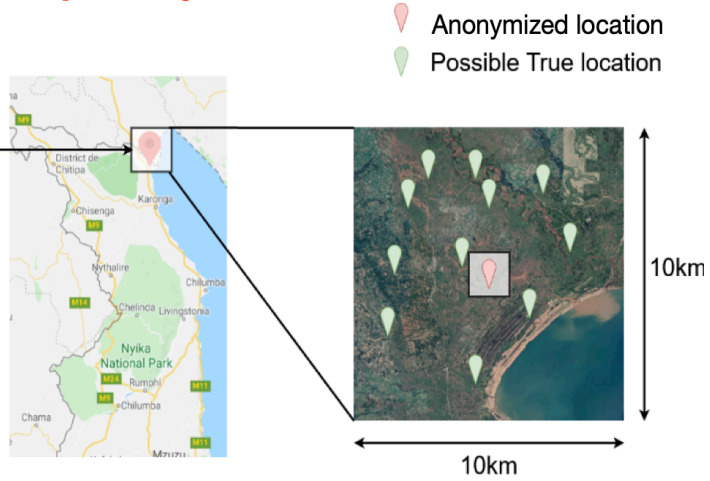
- "The lighter, the richer"
  - true at the country level [Elvidge, 1997]
  - not so true locally [Noor, 2008]
- Time period : from 1991
- Resolution: 500 to 1000 meters
- Two satellites: DMSP-OLS and VIIRS
  - Need for harmonization [Li, 2020 ; Chen, 2021]: source of errors

Elvidge et al., (1997). International Journal of Remote Sensing. "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption."
Noor et al.,(2008). Population Health Metrics. "Using remotely sensed night-time light as a proxy for poverty in Africa."
Li et al., (2020). Scientific Data. "A harmonized global nighttime light dataset 1992–2018."
Chen et al., (2021). Earth System Science Data."An extended time series (2000–2018) of global NPP-VIIRS-like nighttime light data from a cross-sensor calibration."

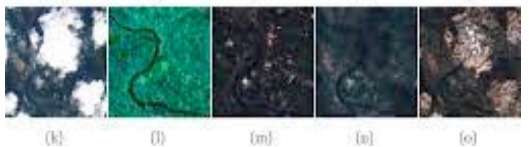
Sources of Uncertainty (1/2)

- **"Anonymized" locations for surveyed poverty indicators**

Latitude	Longitude	Consumption expenditures ($/day)	Name of the village
-9.7298955	33.859230	1.8056488929595411	Kaporo
-9.715316	33.88666	1.4157905044417405	Karonga district
-11.95341	33.367064	1.957231017237896	Mzimba disctrict
-11.632151	34.235546	1.3677132387498063	Nkhata Bay

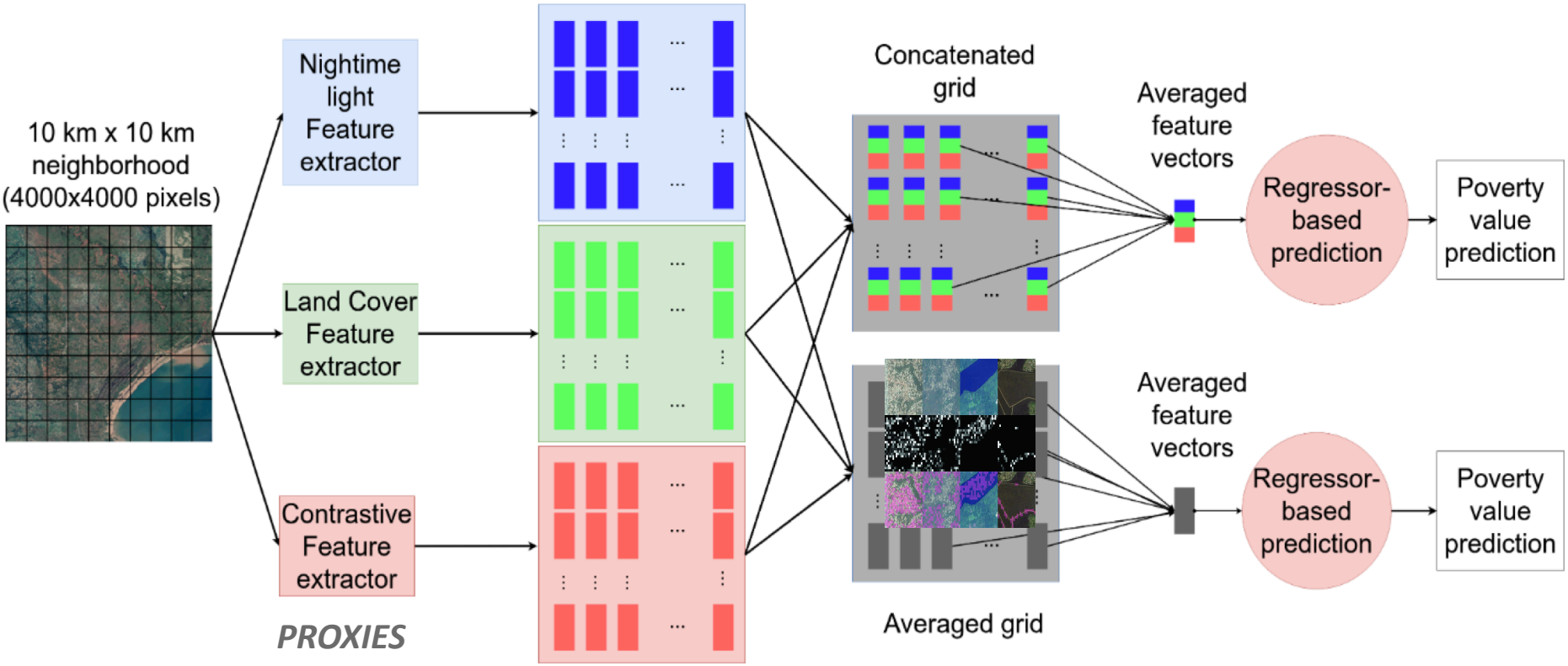


- **Image resolution, resolution mapping, and image preprocessing/aggregation**



Sources of Uncertainty (2/2)

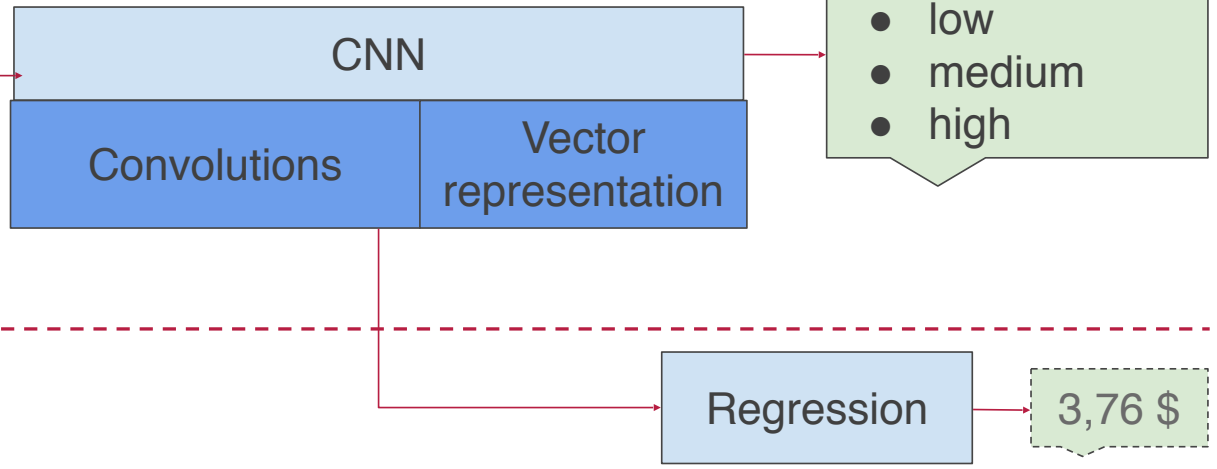
- **Architecture and pipeline design choices**



Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. In: Science, 353(6301) :790–794.

SOTA: Pioneer work of JEAN et al., 2016

- Pretraining

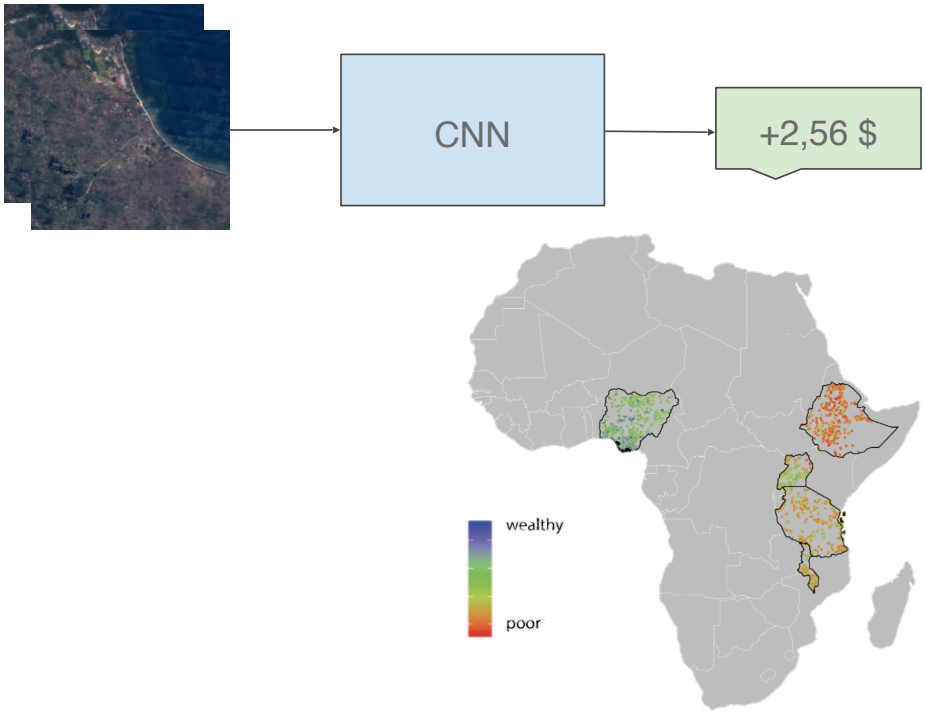


- Transfer learning

- Data required
  - **Image satellites**
  - **Socioeconomic Indicators (SEI)**

Jean et al., Science (2016). *"Combining satellite imagery and machine learning to predict poverty."*

YEH et al., 2020: Poverty Evolution

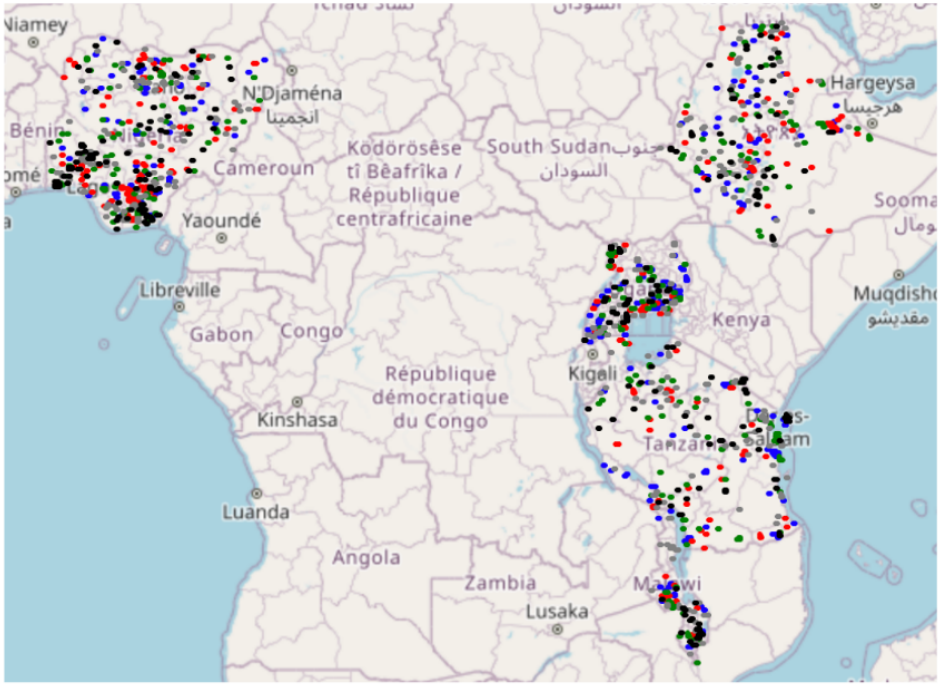


- Bi-temporal Method
- 5 countries in Africa
- 2009-2016

- Results:
  - the prediction of poverty evolution is difficult

Yeh et al., (2020). Nature Communication. "Using publicly available satellite imagery and deep learning to understand economic well-being in Africa."

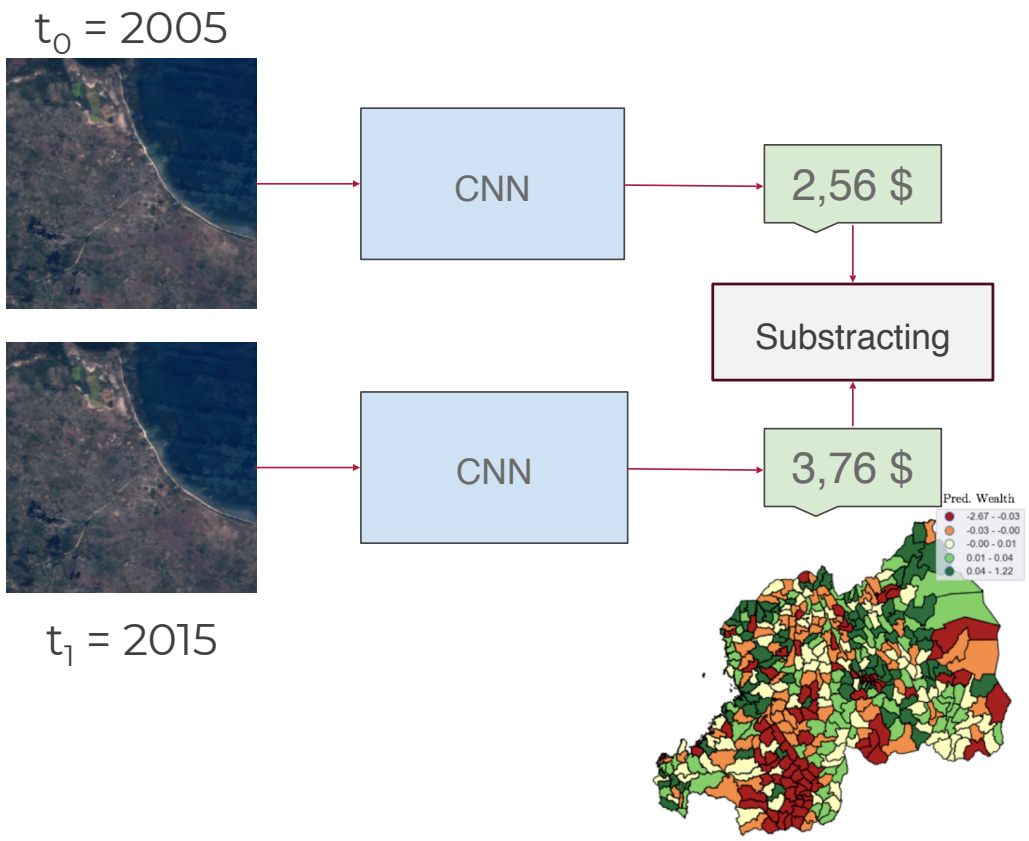
YEH et al., 2021: *SustainBench*



- 5 countries in Africa
- 2005-2016
- 1,665 locations
- Spatial overlaps
- Evolution indicator
  - Differences between 2 surveys
  - PCA

Yeh et al., (2021). NeurIPS, Track on Datasets and Benchmarks 1. "SustainBench: Benchmarks for Monitoring the Sustainable Development Goals with Machine Learning."

KONDMANN et al., 2020: Poverty Evolution



- Methodology from JEAN et al., 2016
- Successive Predictions
- Rwanda 2005-2015

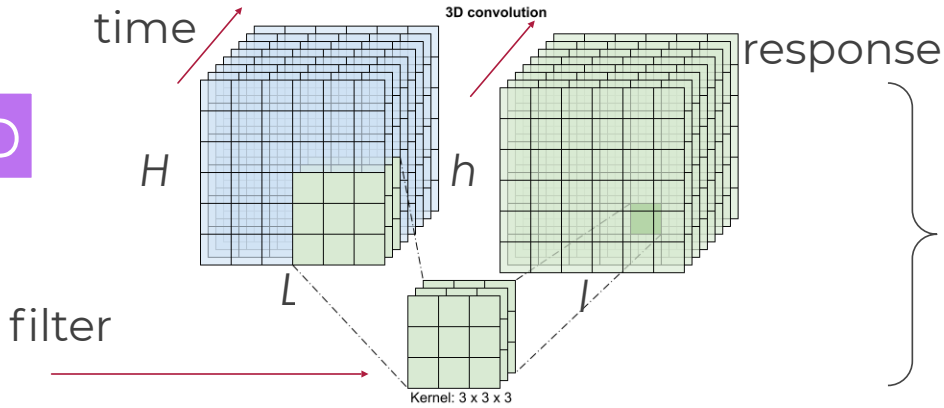
- Results:
  - Heterogeneity of regions
  - No evolution at the country level

Variations in the image series not studied

Kondmann et al., (2020). ICLR Proceedings. "Measuring Changes in Poverty with Deep Learning and Satellite Imagery."

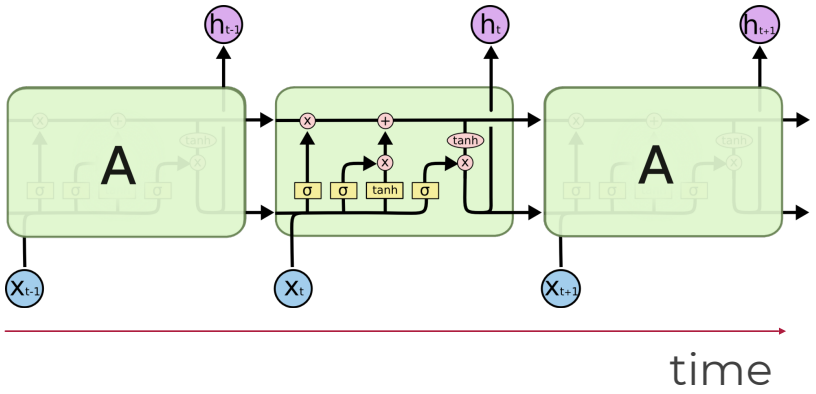
Limitations of architectures for image sequences

CNN 3D



- Temporal dependencies disappear

LSTM

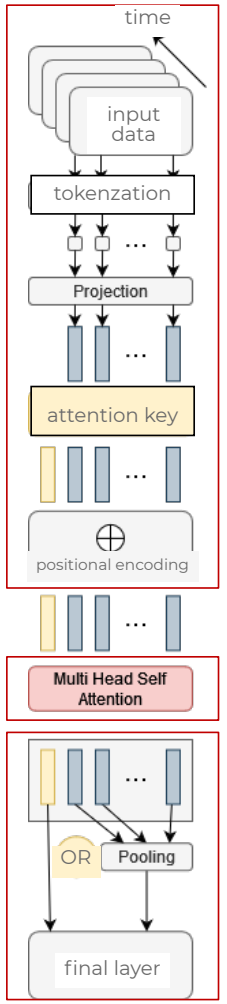


- Temporal dependencies are preserved
- Complex computation

Miller et al., (2024). IEEE Geoscience and Remote Sensing Magazine. "Deep learning for satellite image time-series analysis: A review."

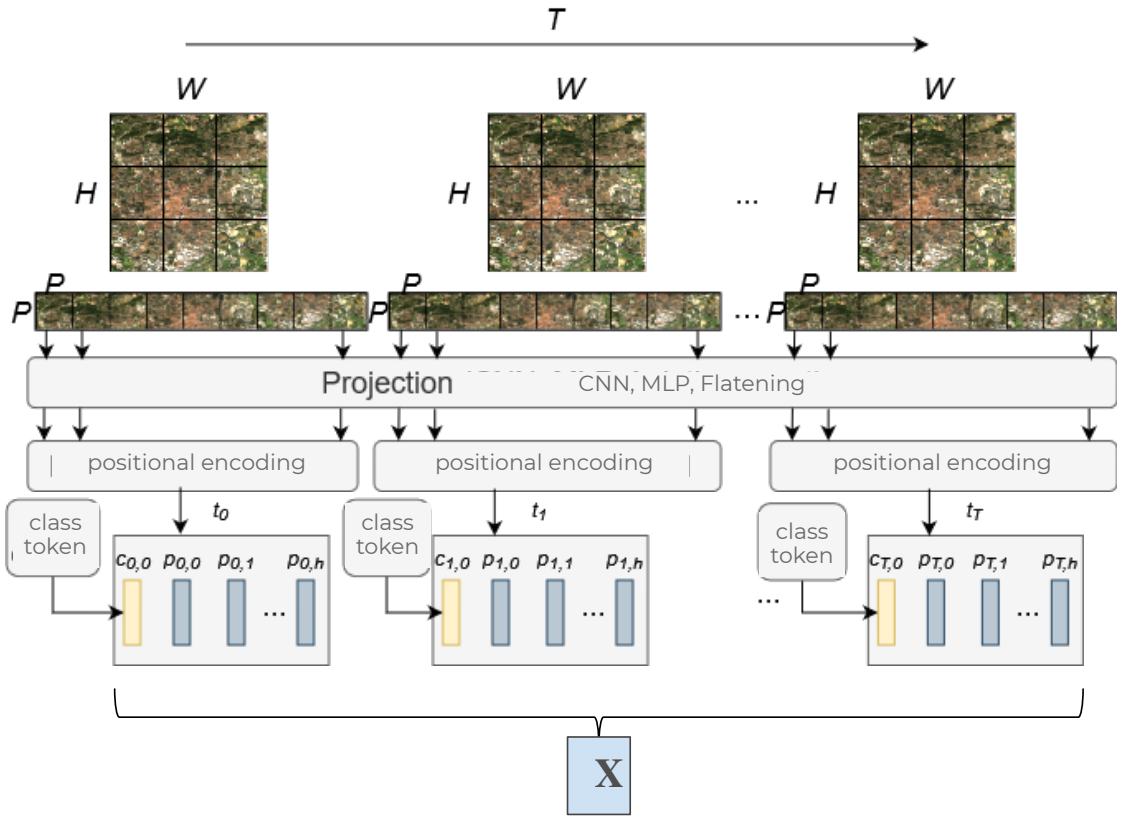
Our Proposition: A Spatiotemporal Transformer

- Preprocessing:
  - Tokenization
  - Projection
  - Attention key
  - Positionnal Encoding
- Learning:
  - Multi-head self-attention
- Output:
  - MLP



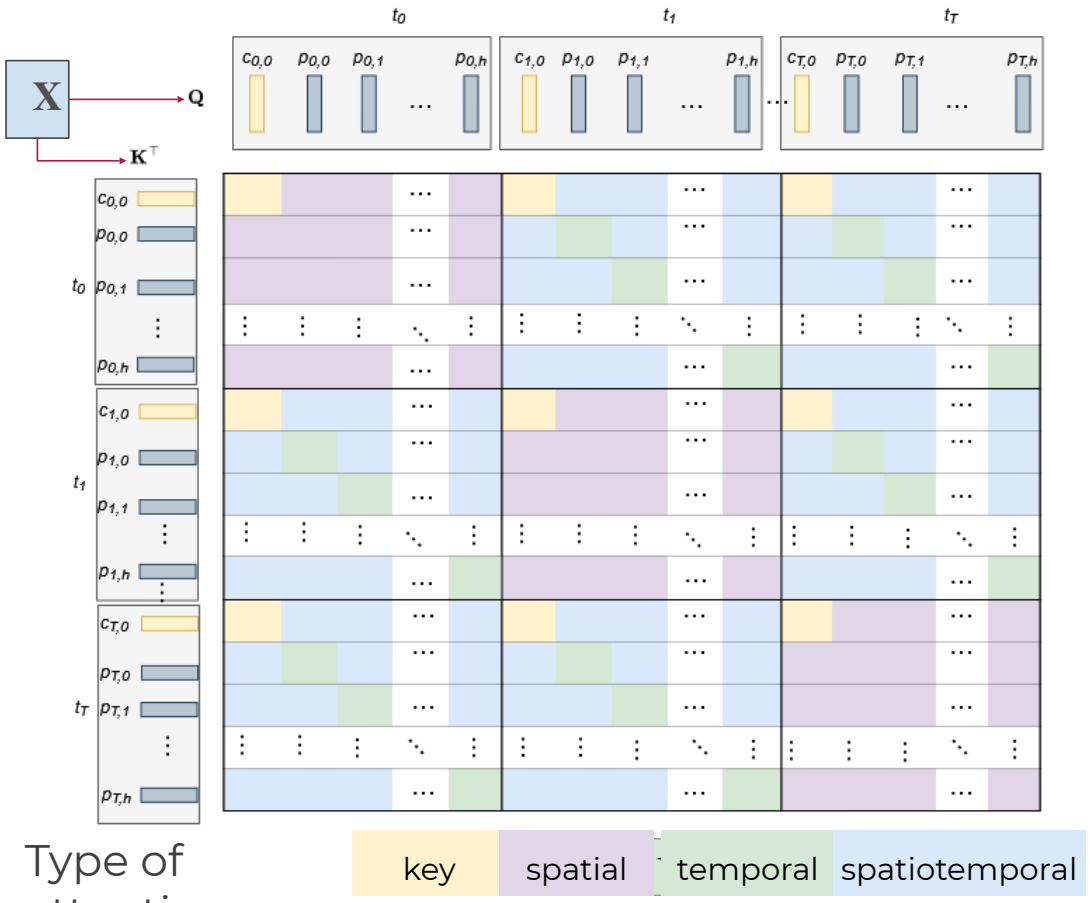
Vaswani et al., (2017). NeurIPS. "Attention is all you need."
Dosovitskiy et al., (2020). ICLR. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale."
Arnab et al., (2021). ICCV. "Vivit: A video vision transformer."
Rußwurm and Körner., (2020). ISPRS Journal of Photogrammetry and Remote Sensing. "Self-attention for raw optical Satellite Time Series Classification."
Tarasiou et al., (2023). CVPR. "Vits for sits: Vision transformers for satellite image time series."

From Raw Data to Sequences of Vectors



- **Tokenization:** transformation into sequence
- **Projection:** transformation into vector
- **Positional encoding:** temporal information
- **Token of class:** representative of the sequence of vectors
- **Output:** sequence of vectors

Multi-Head Self-Attention



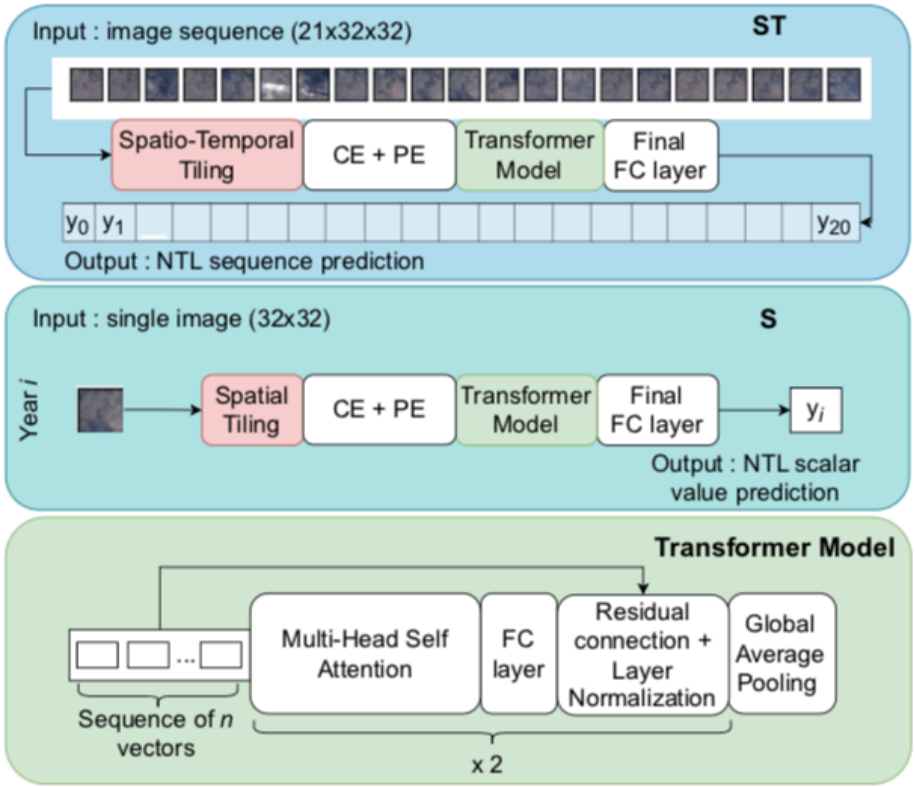
- **Parallelizable**
- Compute all the types of the relationships:
  - **spatial**
  - **temporal**
  - **spatiotemporal**
- Long-term dependencies are **preserved**

Comparison of Transformer Architectures

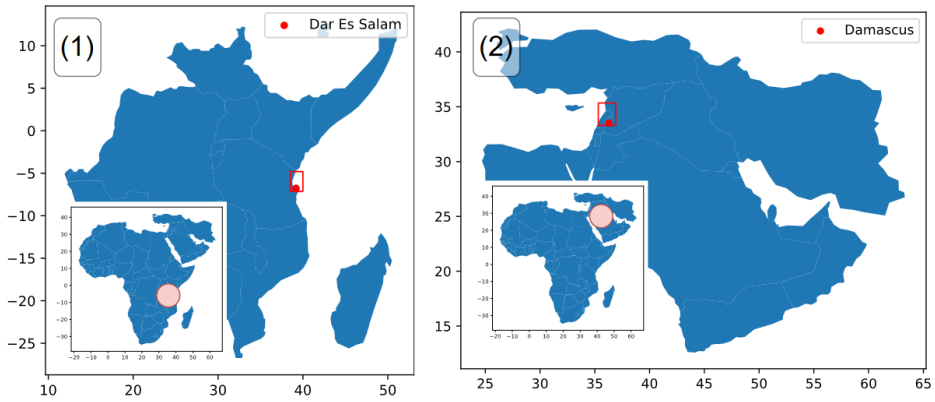
spatio-temporal Transformer (ST)

spatial Transformer (S)

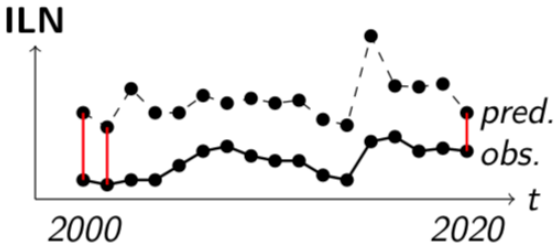
architecture *Transformer*



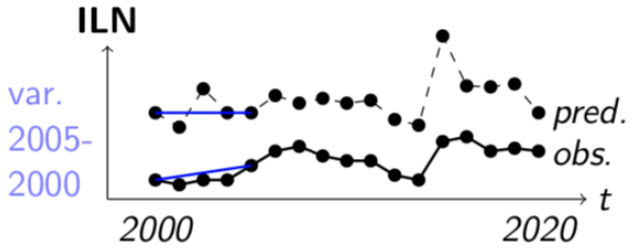
Experiments: Tanzania and Syria



- 8000x5000 pixels (50 000 km²)
- Sequences of 21 images
- 32x32 pixels
- 6 bands (B, G, R, NIR, SWIR-1 et -2)
- Zanzibar excluded
- Cross Validation (5 folds)



per Year



per Variation range

$$MAE = \frac{1}{N}\sum_{i=0}^{N}|y_i - \hat{y}_i|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

Experimental Results

Tanzania (1) and Syria (2)

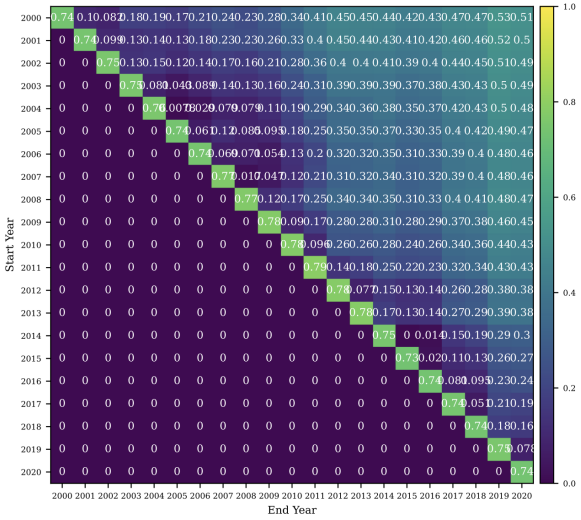
Score		$MAE \downarrow (1)$	$R^2 \uparrow (1)$	$MAE \downarrow (2)$	$R^2 \uparrow (2)$
per Year	ST	$\mathbf{0.085}^{\pm 0.010}$	$\mathbf{0.695}^{\pm 0.063}$	$\mathbf{1.011}^{\pm 0.061}$	$\mathbf{0.543}^{\pm 0.014}$
	S	$0.098^{\pm 0.008}$	$0.591^{\pm 0.040}$	$1.262^{\pm 0.104}$	$0.431^{\pm 0.028}$
$\Delta t = 1$	ST	$\mathbf{0.033}^{\pm 0.002}$	$\mathbf{0.123}^{\pm 0.062}$	$\mathbf{0.453}^{\pm 0.020}$	$\mathbf{0.128}^{\pm 0.010}$
	S	$0.087^{\pm 0.009}$	$-4.664^{\pm 1.390}$	$0.685^{\pm 0.066}$	$-1.240^{\pm 0.439}$
$\Delta t = 10$	ST	$\mathbf{0.103}^{\pm 0.005}$	$\mathbf{0.322}^{\pm 0.042}$	$\mathbf{1.083}^{\pm 0.043}$	$\mathbf{0.218}^{\pm 0.013}$
	S	$0.140^{\pm 0.010}$	$-0.291^{\pm 0.083}$	$1.287^{\pm 0.056}$	$-0.150^{\pm 0.099}$
$\Delta t = 15$	ST	$\mathbf{0.137}^{\pm 0.009}$	$\mathbf{0.439}^{\pm 0.048}$	$\mathbf{1.069}^{\pm 0.041}$	$\mathbf{0.278}^{\pm 0.020}$
	S	$0.163^{\pm 0.010}$	$-0.032^{\pm 0.129}$	$1.314^{\pm 0.059}$	$-0.123^{\pm 0.071}$

Spatio-temporal (ST) **Spatial (S)**

- Per Year: similar performances
- $\Delta t = 1$:

  both models have low performances
- $\Delta t = 10 : S \ll ST$
- $\Delta t = 15 : S \ll ST$
- ST model have better performances when variation range increases

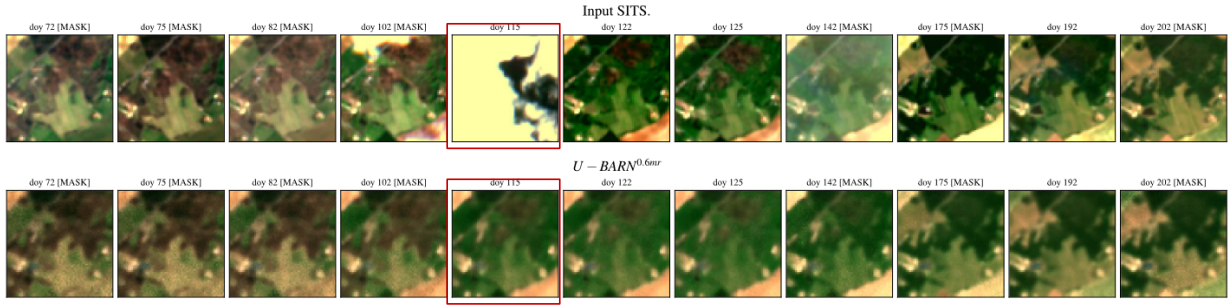
Contributions

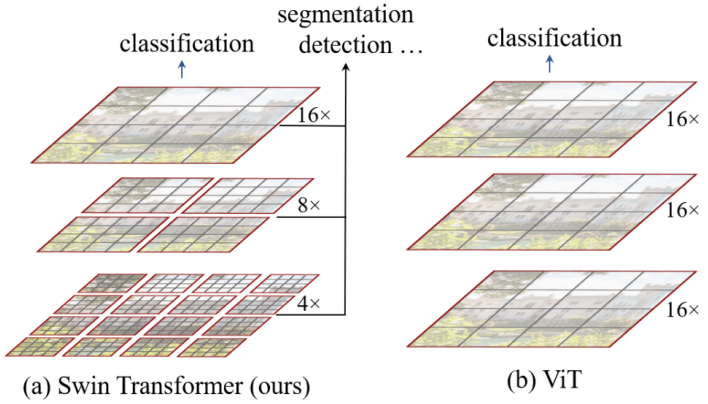


- Comparison of paradigms: spatial vs spatiotemporal
- Transformer applied to the prediction of SEI evolution
- Issues in generalization

- *R. Jarry, M. Chaumont, L. Berti-Équille, G. Subsol. Comparer le paradigme spatial au spatio-temporel pour estimer l'évolution d'indicateurs socioéconomiques à partir d'images satellites. In: CNIA 2024 - Plateforme Intelligence artificielle - Conférence nationale en intelligence artificielle, La Rochelle*
- *R. Jarry, M. Chaumont, L. Berti-Équille, G. Subsol. Comparing Spatial and Spatio-Temporal Paradigms to Estimate The Evolution of Socio-Economical Indicators from Satellite Images. In: IGARSS 2023 - IEEE International Geoscience and Remote Sensing Symposium (p. 5790-5793). IEEE.*
- *R. Jarry, M. Chaumont, G. Subsol, L. Berti-Équille. Predicting Socio-economic Indicator Variations with Satellite Image Time Series and Transformer. In: British Machine Vision Conference 2024, workshop on Machine Vision for Earth Observation and Environment Monitoring*

Perspectives





- Issues in generalization:
  - Self-supervised Learning[1]
- Try more recent Transformer architectures:
  - Slided Window Transformer[2]

[1]*Dumeur et al., (2024). IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. "Self-Supervised Spatio-Temporal Representation Learning of Satellite Image Time Series."*
[2]*Liu et al., (2021). ICCV. "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows."*

Overview of Our Research (2/3)

Detection of favelas from satellite images using CROMA

M.Sc. Internship of Thomas Hallopeau supervised by N. Dessay, L. Demagistri, J. Guérin (IRD ESPACE-DEV), Montpellier, France
MATHIS Project funded by CNES TOSCA

Image Classification for Favela Monitoring

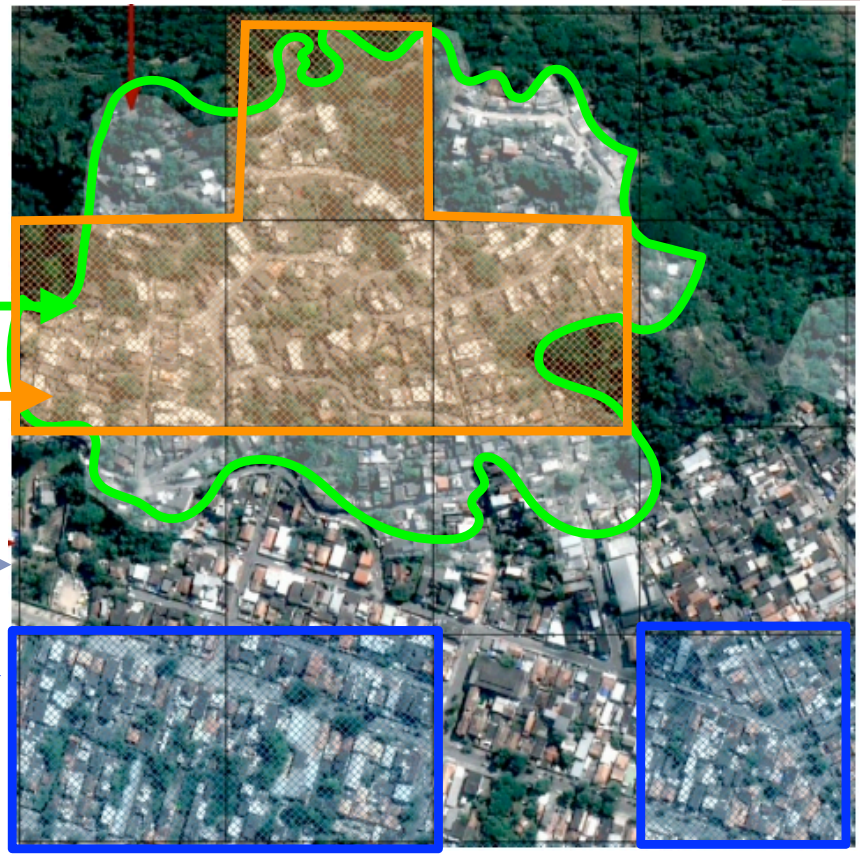
Random Forest on 6 variables from OpenStreet Map, Digital Terrain Models, Vegetation index

Ground truth provided by IBGE

Detected Cell **class F** : Favela

Disgarded cell

Detected Cell **class R** : urbain no favela



Sentinel-2 Resolution ***150m***

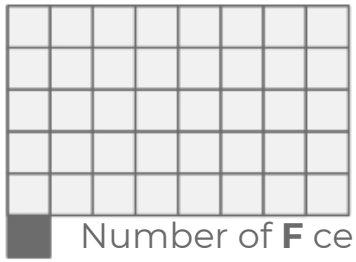
Reuß, Felix. (2017). Detection of favelas in Brazil using texture parameters and machine learning. MSc. thesis https://elib.dlr.de/115220/1/Reuss_MA.pdf

Complex Intra-Urban Classification Task

- Selection and computation of variables for RF is hard and time-consuming
- High variability of slum features
- High similarity with urban neighbourhoods
- Unbalanced dataset

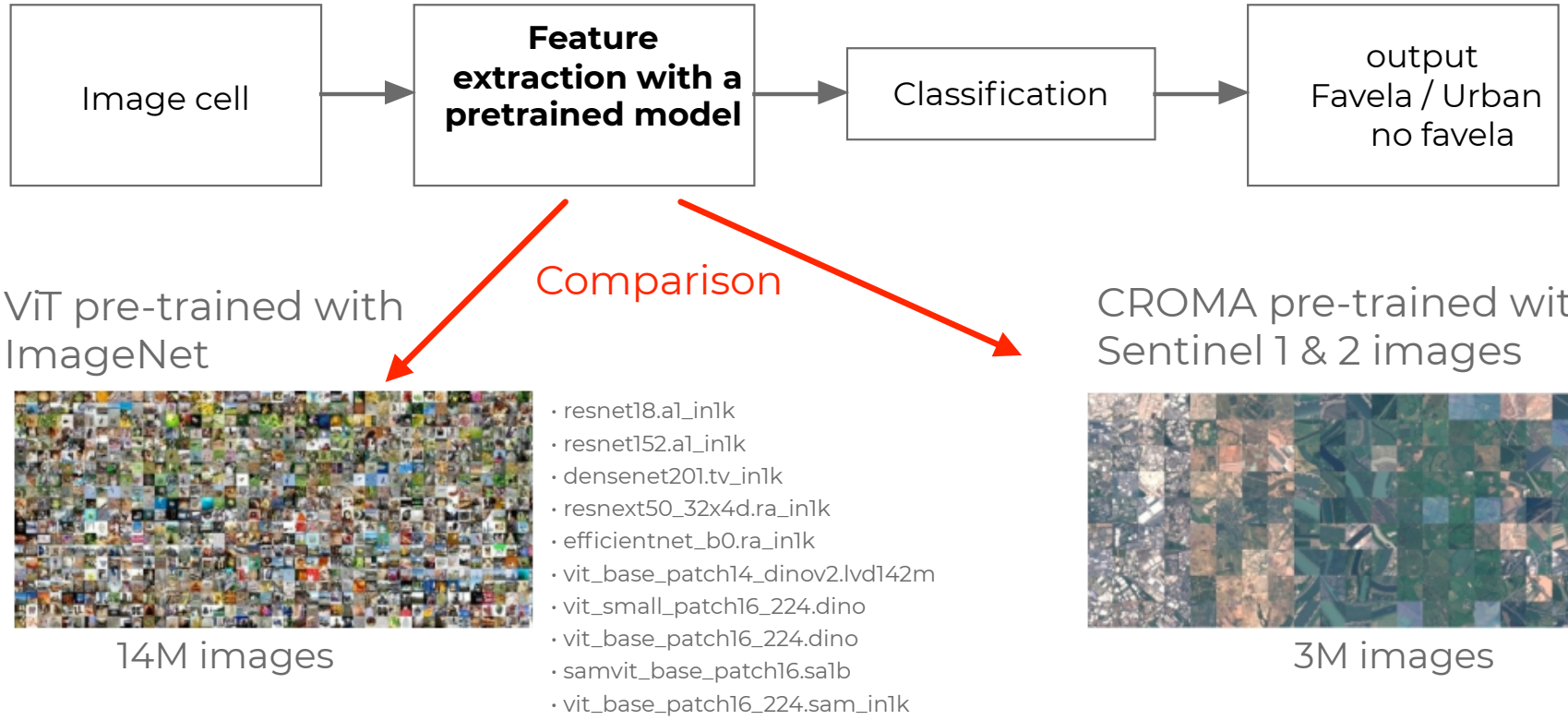
Class F

Class R



What could be the added-value of pre-trained models ?

Proposed Approach

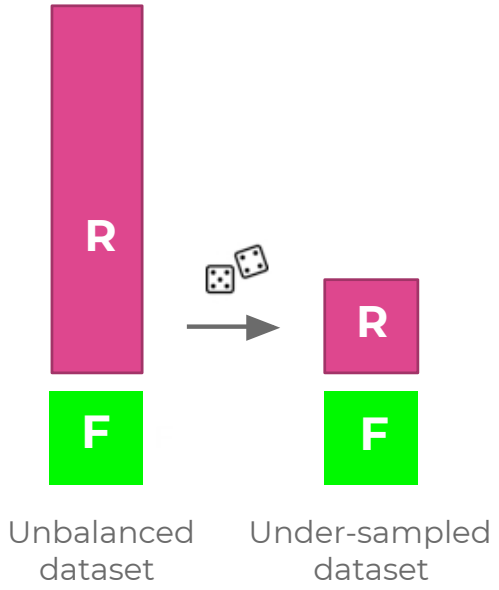


Experimental Results (1/3)

	ViT pretrained on ImageNet *	CROMA
Precision	.70+/- 0.05	.81 +/- 0.03
Recall	.78 +/- 0.08	.81 +/- 0.05
F1-score	.74 +/- 0.03	.81 +/- 0.02

* vit_base_patch16_224.sam_in1k

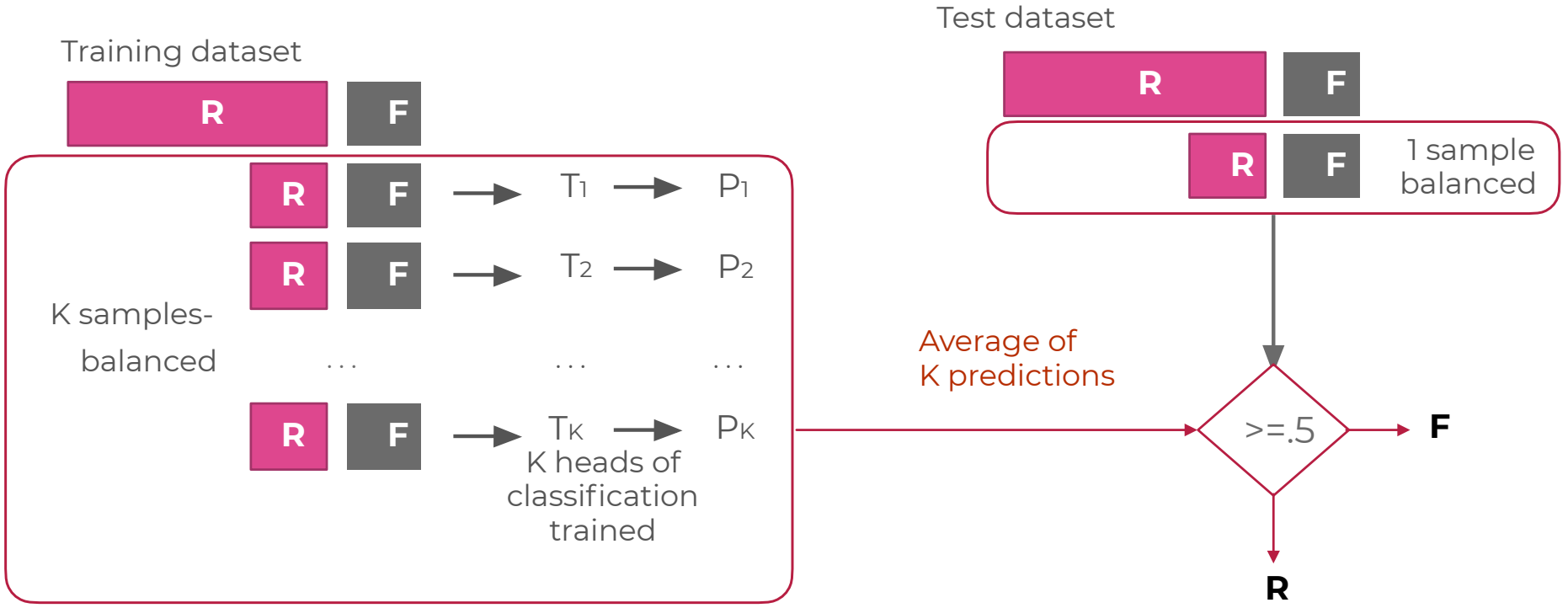
Class F : Positive
Class R : Negative



Room for improvement: **Undersampling**

- The classification head focuses on the majority class
- High variabily between samples

Ensembling balanced samples for training



Zhongbin Sun et al. « A novel ensemble method for classifying imbalanced data ». In : Pattern Recognition 48 (mai 2015). doi : 10.1016/j.patcog.2014.11.014.

Experimental Results (2/3)

	Random Forest	CROMA	CROMA with sample ensembling
Precision	.84	.83 +/- 0.06	.86 +/- 0.05
Recall (%)	.82	.80 +/- 0.07	.84 +/- 0.07
F1-score (%)	.75	.81 +/- 0.06	.85 +/- 0.06

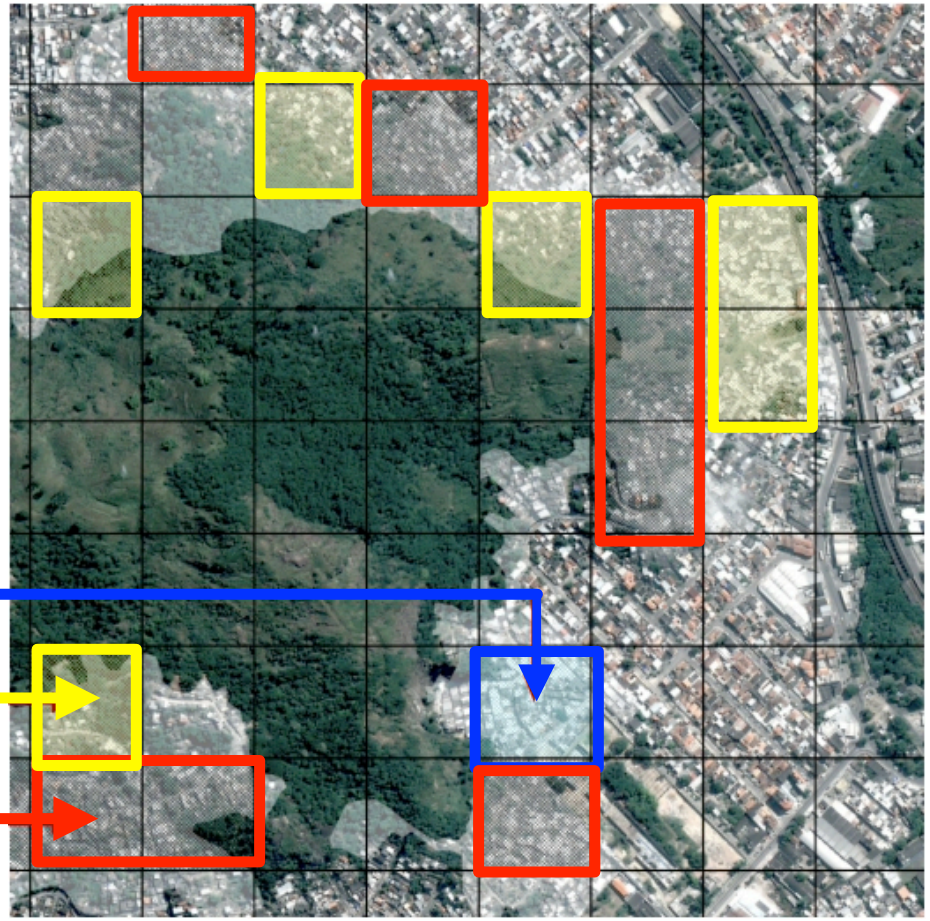
Comparison of the two approaches

Some favelas are better detected by RF, others by deep learning

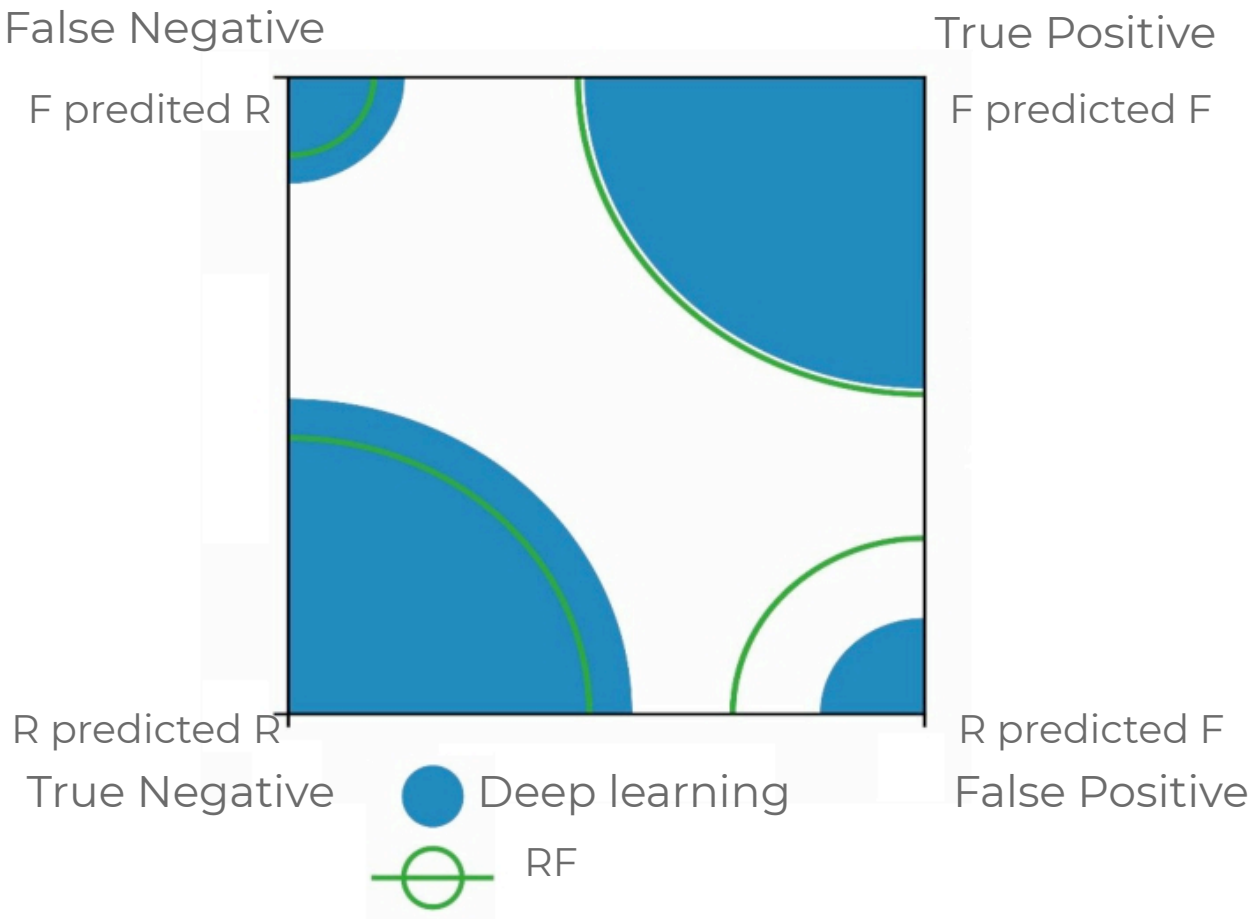
Correct prediction with RF but incorrect with deep learning

Correct prediction with deep learning but incorrect with RF

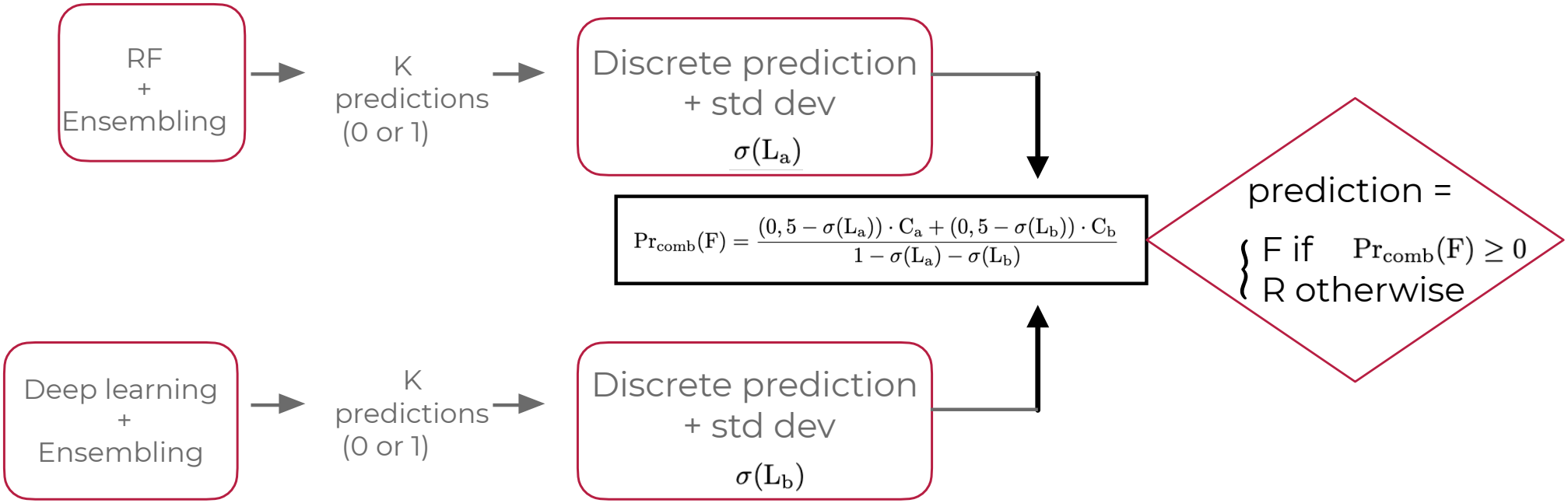
Same prediction by the two approaches (true or false)



Confusion Matrix: Complementary Methods



Our Proposition: An Hybrid Approach



Experimental Results (3/3)

	Random Forest	CROMA	CROMA + ensembling	RF on Concatenation of CROMA (768) and RF (6) variables	RF on weighted combinaison of predictions	Hybrid approach with ensembling
Precision	.84	.83	.86	.84	.88	.83
Recall	.82	.80	.84	.82	.85	**.91**
F1-score	.75	.81	.85	.83	.86	**.87**

Perpectives:

- Semantic segmentation of images
- Dynamic monitoring with a series of satellite images

Overview of Our Research (3/3)

Automated annotation of coral reef images with hierarchical classification

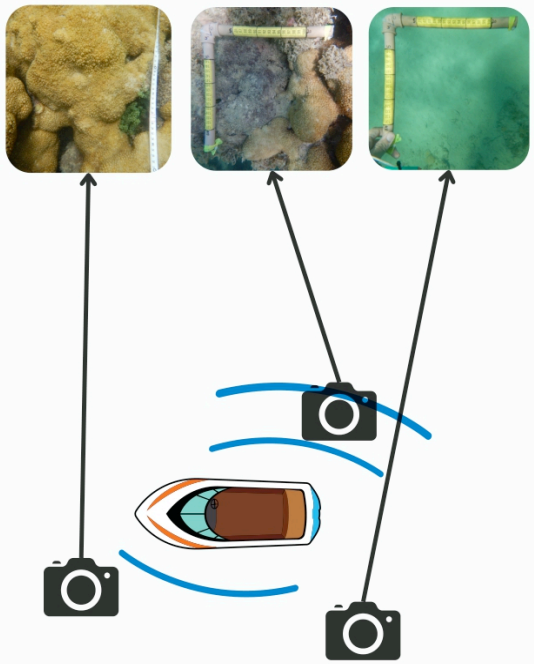
M.Sc. Internship of Celia Blondin co-supervised with J. Guérin (IRD ESPACE-DEV), Montpellier, France

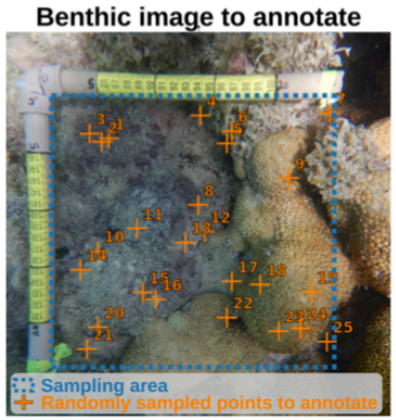
Automating the annotation of coral reef images

Data collection

Annotation

Ecosystem Indicators





Manual annotation

Expert annotator

Automatic annotation

Image patches centered around annotation points

Deep neural network

Annotations

Turf Filamentous
1, 2, 3, 5, 10, 11, 14

Calcarous Turf
4

Dictyota
6

Palythoa
7, 9, 12, 17, 18, 19, 22, 23, 24, 25

Calcyfying calcareous crustose algae
8, 13, 15, 16, 20, 21

Coral cover

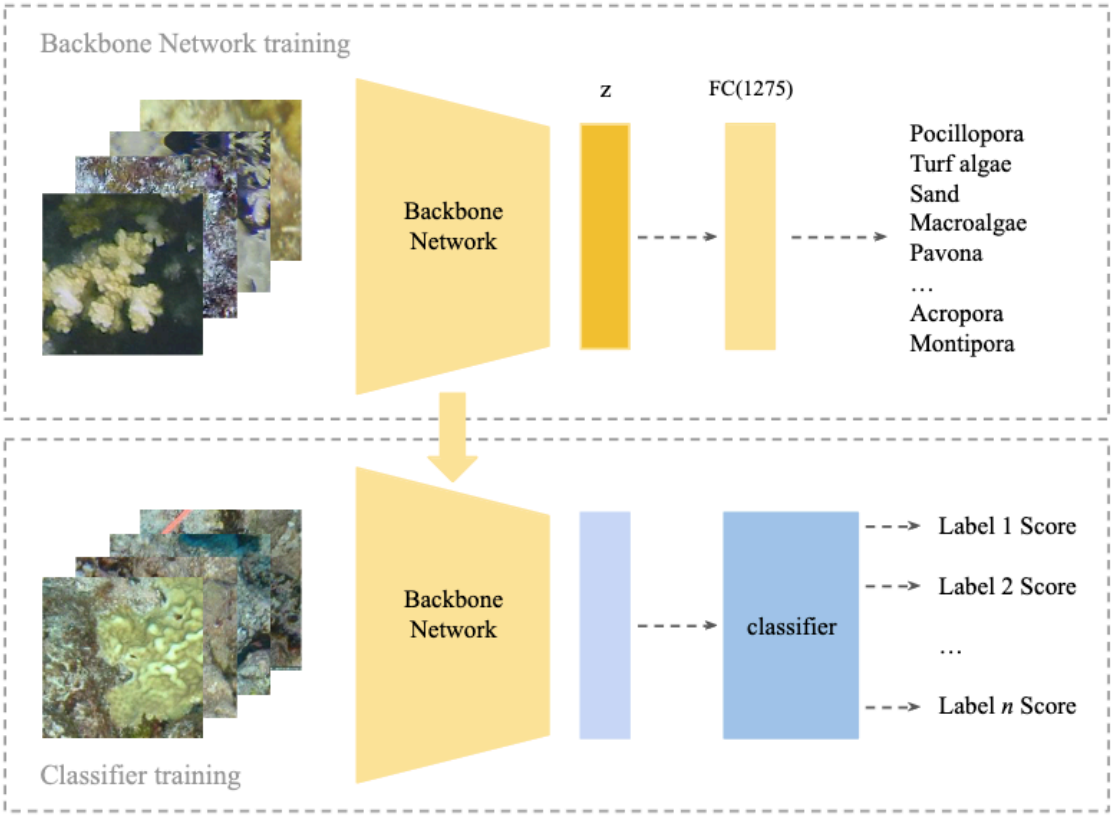
Limitations of automated annotation methods for biodiversity images

- Need to exploit knowledge in ecology/biology to use less training data [2]
- Hard to build a dataset large enough with high-quality annotations [1,2,3]
- Biodiversity datasets are unbalanced [4]
- Existing models are highly energy and resource-intensive [1]
- Hard to adapt models to other biodiversity datasets

[1] M. M. Adnan, M. S. M. Rahim, A. Rehman, Z. Mehmood, T. Saba and R. A. Naqvi, "Automatic Image Annotation Based on Deep Learning Models: A Systematic Review and Future Challenges," in IEEE Access, vol. 9, pp. 50253-50264, 2021, doi: 10.1109/ACCESS.2021.3068897
[2] B. G. Weinstein, "A computer vision for animal ecology," Journal of Animal Ecology, vol. 87, no. 3, pp. 533–545, 2018.
[3] Saleh, A., Sheaves, M., & Rahimi Azghadi, M. (2022). Computer vision and deep learning for fish classification in underwater habitats: A survey. Fish and Fisheries, 23, 977–999. https://doi.org/10.1111/faf.12666
[4] Eerola, T., Batrakhanov, D., Barazandeh, N.V. et al. Survey of automatic plankton image recognition: challenges, existing solutions and future perspectives. Artif Intell Rev 57, 114 (2024). https://doi.org/10.1007/s10462-024-10745-y
[5] Keller, A. A. (2017). Multi-objective optimization in theory and practice i: classical methods. Bentham Science Publishers.

CoralNet

https://coralnet.ucsd.edu/



- Nearly 3,000 registered users
- 1,741,855 images
- from 2,040 distinct sources
- with over 65 million annotations

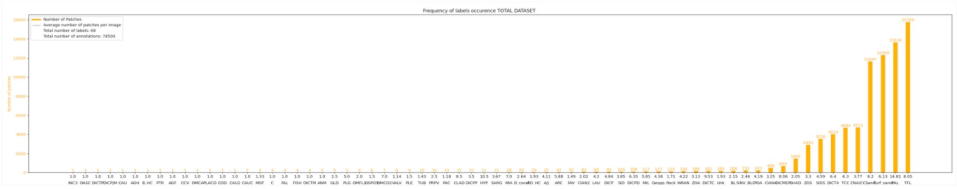
V1	V2
432, 489 images	**591, 604** images
15, 137, 977 annotated points	**16, 533, 651** annotated points
254 sources for training the backbone network with **1, 279** classes in common	304 sources for training the backbone network with **1, 275** classes in common
for both versions: 26 sources for training the classifiers, each randomly split into 80/20 for training and testing	

Q. Chen, O. Beijbom, S. Chan, J. Bouwmeester, D. Kriegman. "A New Deep Learning Engine for CoralNet" . International Conference on Computer Vision (ICCV) Workshops, 2021.
https://openaccess.thecvf.com/content/ICCV2021W/OceanVision/papers/Chen_A_New_Deep_Learning_Engine_for_CoralNet_ICCVW_2021_paper.pdf

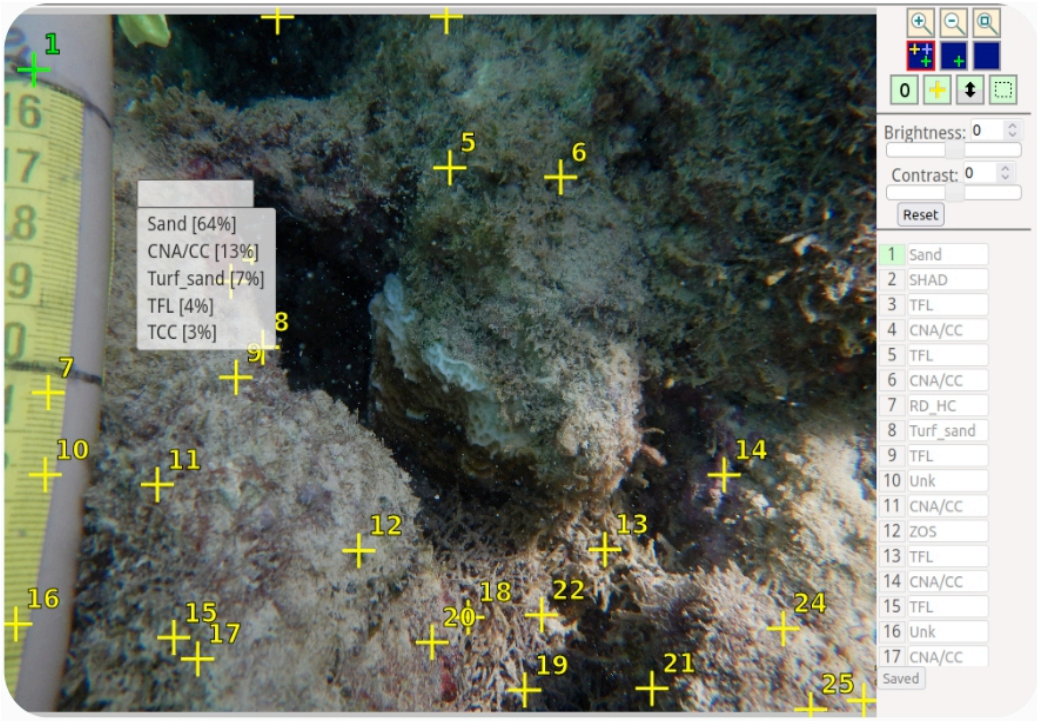
Main Limitations

- 3000+ **NEW** images x 25 points x 146 labels
- **LABOR-INTENSIVE manual annotation**
- CoralNet Classifier **precision** of **0.70**

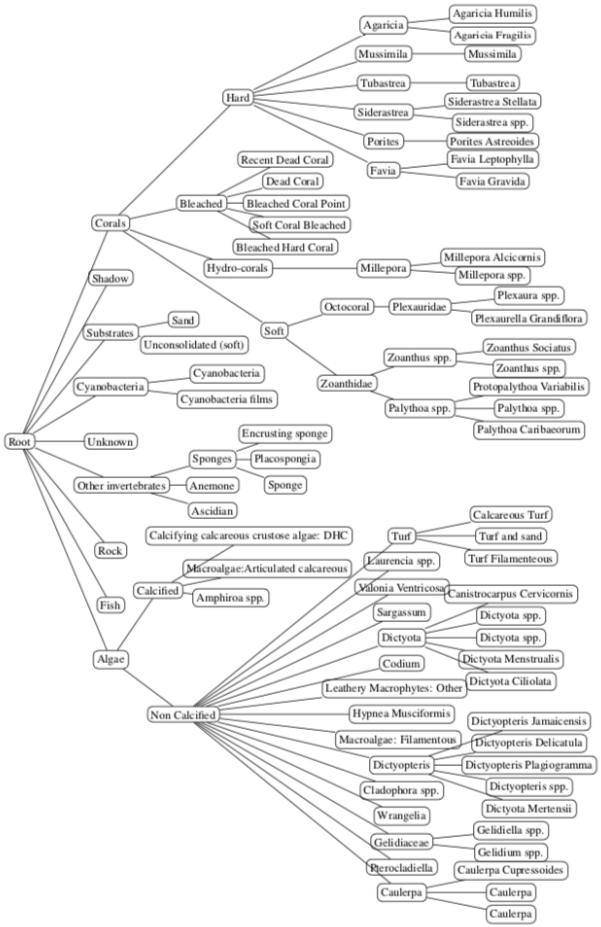
  dominated by the majority class



When can we trust the classifer and stop manual annotation?



Our Proposition: Leveraging Hierarchy



Advantages for ML:

- Using uncertainty to differentiate various species
- Compare metrics and use the ones that are adapted for hierachical datasets

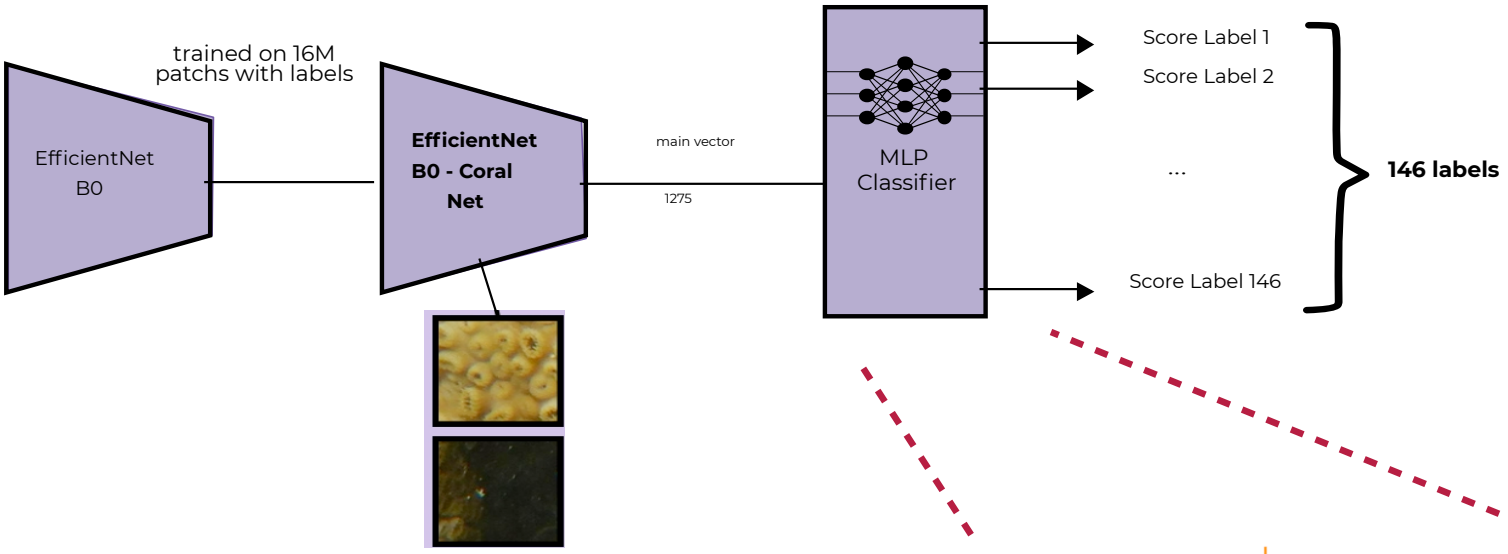
Advantages for Ecology:

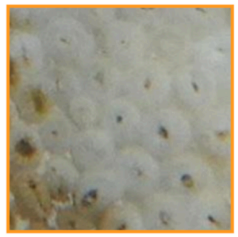
- Improve the confidence in the detected class
- Estimate the coral cover at different levels of the hierachy (multi-level prediction)

Collaboration between ML experts and marine ecologists

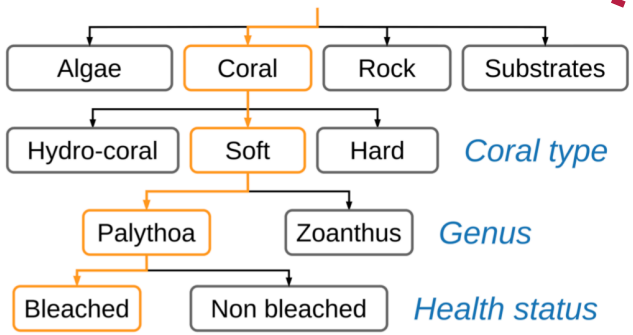
Silla, C.N., Freitas, A.A. A survey of hierarchical classification across different application domains. Data Min Knowl Disc 22, 31–72 (2011). https://doi.org/10.1007/s10618-010-0175-9

Our proposition: Hierarchical Classification

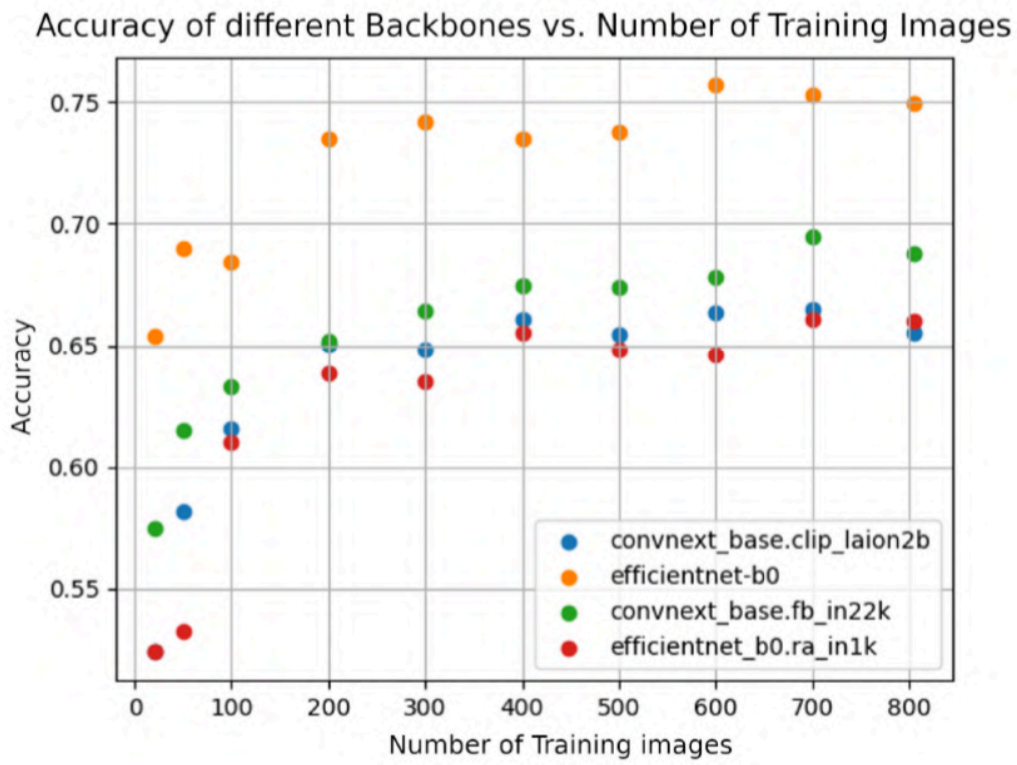




Label:
Bleached Palythoa



Experimental Results (1/2)



Experimental Results (2/3): Hierarchical Metrics

$$hP = \frac{\sum_i |\alpha_i \cap \beta_i|}{\sum_i |\alpha_i|},$$

$$hR = \frac{\sum_i |\alpha_i \cap \beta_i|}{\sum_i |\beta_i|},$$

$$hF = \frac{2 \times hP \times hR}{hP + hR}.$$

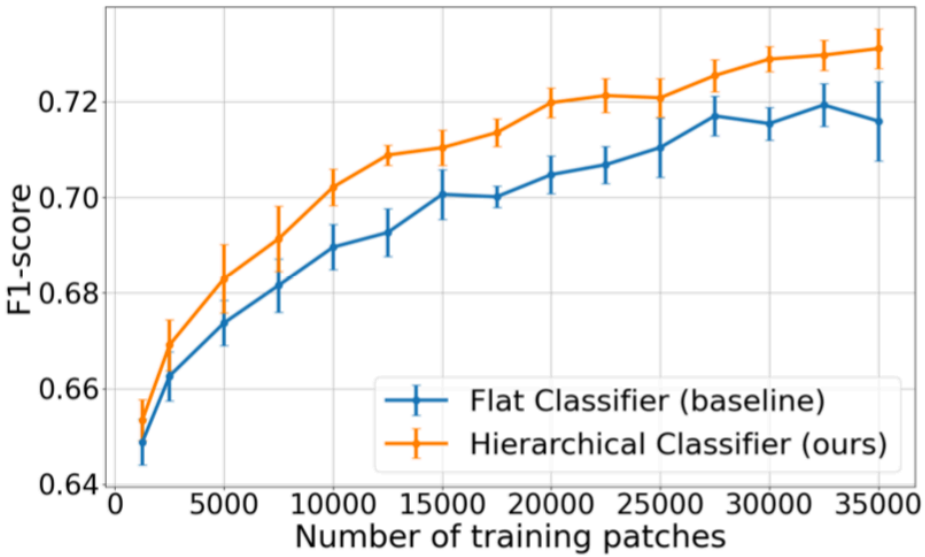
α_i the set of the most specific classes **predicted** for each test example i, and all of their ancestor classes;

β_i the set of the **true** most specific classes of test example i, and all their ancestor classes;

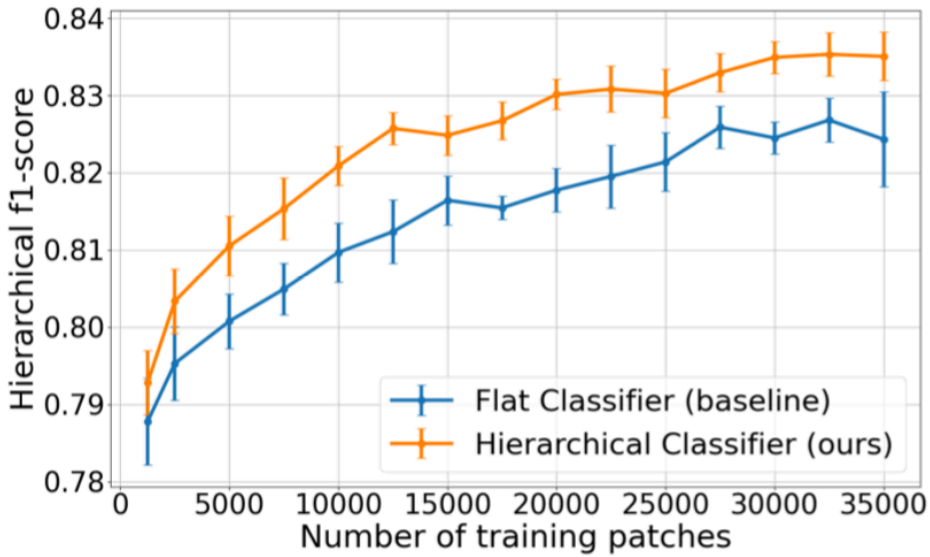
each summation is computed **over all of the test set examples**.

A. Kosmopoulos, I. Partalas, E. Gaussier, G. Paliouras, and I. Androutsopoulos, "Evaluation measures for hierarchical classification: a unified view and novel approaches," Data Mining and Knowledge Discovery, vol. 29, pp. 820–865, 2015.

Experimental Results (3/3)



(a) F1-score



(b) Hierarchical F1-score

A. Kosmopoulos, I. Partalas, E. Gaussier, G. Paliouras, and I. Androutsopoulos, "Evaluation measures for hierarchical classification: a unified view and novel approaches," Data Mining and Knowledge Discovery, vol. 29, pp. 820–865, 2015.

Outline

Introduction

- IRD Presentation
- SDGs and applied ML
- Building SD data science pipelines
- Main Challenges of ML Applied to SDGs

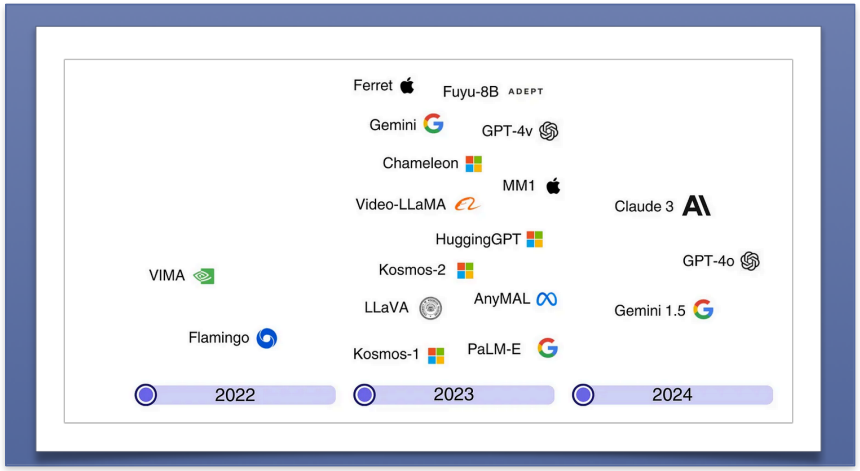
Overview of Our Research

- Estimate poverty evolution from satellite images using transformers
- Detection of favelas from satellite images using CROMA
- Automated annotation of coral reef images with hierarchical classification

Conclusions & Perspectives

Conclusions

- New era of multimodal models and MLLMs for CV
- Many exciting research directions to advance
  - **uncertainty quantification**,
  - **model fine-tuning**,
  - **HIL** and **human/agent orchestration**
  - **Architecture seach**
  - **Actionability**



Source Medium: https://medium.com/@tenyks_blogger/multimodal-large-language-models-mllms-transforming-computer-vision-76d3c5dd267f

Thanks!

Laure Berti-Equille
contact: laure.berti@ird.fr
https://laureberti.github.io/website/

References

- R. Jarry, M. Chaumont, L. Berti-Équille, G. Subsol. Comparer le paradigme spatial au spatio-temporel pour estimer l'évolution d'indicateurs socioéconomiques à partir d'images satellites. In: CNIA 2024 - Plateforme Intelligence artificielle - Conférence nationale en intelligence artificielle, La Rochelle
- R. Jarry, M. Chaumont, L. Berti-Équille, G. Subsol. Comparing Spatial and Spatio-Temporal Paradigms to Estimate The Evolution of Socio-Economical Indicators from Satellite Images. In: IGARSS 2023 - IEEE International Geoscience and Remote Sensing Symposium (p. 5790-5793). IEEE.
- R. Jarry, M. Chaumont, L. Berti-Équille, G. Subsol. Assessment of CNN-based methods for poverty estimation from satellite images. In: Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VII (p.550-565). Presented In IAPR PRRS 2021 Workshop.
- R. Jarry, M. Chaumont, G. Subsol, L. Berti-Équille. Predicting Socio-economic Indicator Variations with Satellite Image Time Series and Transformer. In: British Machine Vision Conference 2024, workshop on Machine Vision for Earth Observation and Environment Monitoring
- K. K. Owen et D. W. Wong. « An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics ». eng. In : (2013). Accepted : 2013-10-23T09 :29 :44Z. issn : 0143-6228. doi : 10.1016/j.apgeog.2012.11.016.
- Yann LeCun, Yoshua Bengio et Geoffrey Hinton. « Deep learning ». en. In : Nature 521.7553 (mai 2015). Publisher : Nature Publishing Group, p. 436-444. issn : 1476-4687. doi : 10.1038/nature14539.
- Alexey Dosovitskiy et al. An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale. arXiv :2010.11929 [cs]. Juin 2021. doi : 10.48550/arXiv.2010.11929.
- Mathilde Caron et al. Emerging Properties in Self-Supervised Vision Transformers. arXiv :2104.14294 [cs]. Mai 2021. doi : 10.48550/arXiv.2104 . 14294.
- Maxime Oquab et al. DINOv2 : Learning Robust Visual Features without Supervision. arXiv :2304.07193 [cs]. Fév. 2024. doi : 10.48550/arXiv.2304 . 07193.
- Jia Deng et al. « ImageNet : A large-scale hierarchical image database ». In : 2009 IEEE Conference on Computer Vision and Pattern Recognition. ISSN : 1063-6919. Juin 2009, p. 248-255. doi : 10 . 1109 / CVPR . 2009 . 5206848.
- Alexander Kirillov et al. Segment Anything. arXiv :2304.02643 [cs]. Avr. 2023. doi : 10.48550/arXiv.2304.02643.
- Anthony Fuller, Koreen Millard et James R. Green. CROMA : Remote Sensing Representations with Contrastive Radar-Optical Masked Autoencoders. arXiv :2311.00566 [cs]. Nov. 2023. doi : 10.48550/arXiv.2311 . 00566.
- Zhongbin Sun et al. « A novel ensemble method for classifying imbalanced data ». In : Pattern Recognition 48 (mai 2015). doi : 10.1016/j.patcog.2014.11.014.