Predicting Socio-economic Indicator Variations with Satellite Image Time Series and Transformer

Robin Jarry, Marc Chaumont, Laure Berti-Equille, Gérard Subsol

Predicting Socio-economic Indicator Variations with Satellite Image Time Series and Transformer

Robin JARRY[1], Marc CHAUMONT[1,2],
Laure BERTI-ÉQUILLE[3] Gérard SUBSOL[1]
LIRMM, Univ Montpellier, CNRS [1], Univ Nîmes[2], ESPACE-DEV,
Univ. Montpellier, IRD, UA, UG, UR [3],
Montpellier France

November 14, 2024

Workshop MVEO'2024 in conjunction with BMVC'2024, 25-28 Nov.24, Glasgow, UK.

Outline

Context

We[a] want to produce world maps that report:

- Consumption expenditures,
- Income per household,
- Asset index,
- Wealth index,
- ...

[a] Socio-economists, ecologists, remote sensing researchers, computer scientists, ...

Chi et al. PNAS'2022 "Micro-Estimates for all low- and middle-income countries." `http://3.15.84.96/brief/`

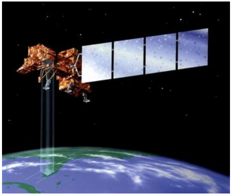
Problem and Solution

Problem:

Need to conduct surveys in many places and very often
$\rightarrow$This is costly and time-consuming

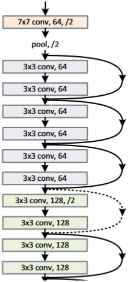
Solution:

Use satellite images and deep learning.



Landsat 7 Satellite

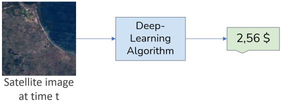
Village in Tanzania



Outline

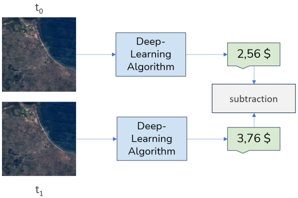
State-Of-The-Art

Prediction at a given date



[2, 4, 11, 12, 14, 15,...]

But we are looking at a variation (i.e. ≈ a math difference)



Problem: The uncertainty of each prediction value (due to noises) leads to **uncertainty** on the result of the **subtraction** which is **higher than the variation range** [12, 22].

⇒ The subtraction must not be used (= uncertainty result).

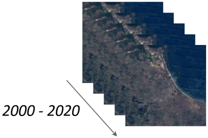
One way to improve the confidence in the predicted variation

Integration of the temporal aspect:

- Yeh et al. [22] take as inputs 2 images (at start and end time) for their CNN.
- Our proposition:
  1. Use of a sequence of images,
  2. Use a transformer,
  3. Pretrained on a related pretext task.

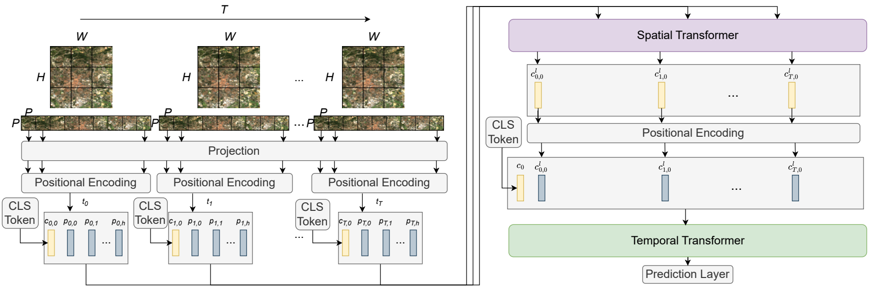
Outline

Ingredient 1: Use a Satellite Image Time Series (SITS)



- One year Landsat−7 median composite images,
- From 2000 to 2020,
- Resolution = 30 meters,
- Image size = 224×224 (≈ 6.72 km^2),
- PRETRAIN = Subset of Africa and Middle East (9795 SITS),
- TRAIN = 1665 locations (i.e 1665 SITS)
  (Nigeria, Ethiopia, Tanzania, Uganda, and Malawi).

Ingredient 2: Use a ViVit

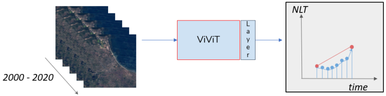


Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lucic, and Cordelia Schmid.
Vivit: A video vision transformer. ICCV'2021.

Ingredient 3: (1) A pretraining

FIRST: A pretraining:

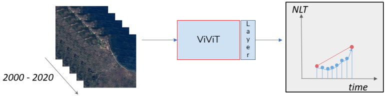
- On the whole 2000 - 2020 duration,
- To predict nighttime light time (NLT) **series**.
  Note: NLT is correlated to our socio-economic indicator.

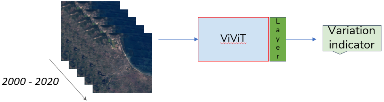


Ingredient 3: (2) Finetuning

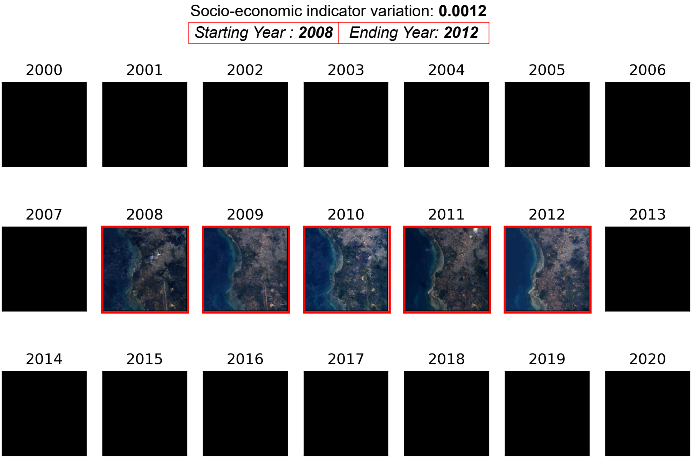
SECOND: A finetuning:

- Notion of start and end of a series,
- To predict the socio-economic **indicator** variation.



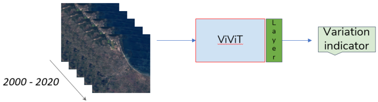


Ingredient 3: Illustration of the masking



Outline

Experimental protocol



... prediction of the Socio-economic Indicator Variations.

- 1665 pairs of (SITS, variation indicator),
- 5 countries (Nigeria, Ethiopia, Tanzania, Uganda, and Malawi),
- 5-fold cross-validation (train on 4 folds and test on 1 fold), with no location overlap between folds,
- 250 epochs, MSE loss, Batch sizes=16, LR=5x10^{-4}, ...,
- ViVit 10 millions parameters, 4 Nvidia V100 GPU.

Results:

	MAE ↓	*RMSE* ↓	r^2 ↑	R^2 ↑
Yeh et al. (2 im)	$0.528^{\pm 0.019}$	$0.687^{\pm 0.032}$	$0.182^{\pm 0.054}$	$0.122^{\pm 0.061}$
Our appr. no pretrain	$0.482^{\pm 0.015}$	$0.637^{\pm 0.020}$	$0.263^{\pm 0.057}$	$0.245^{\pm 0.054}$
Our appr. with pretrain	$\underline{0.460}^{\pm 0.013}$	$\underline{0.366}^{\pm 0.020}$	$\underline{0.328}^{\pm 0.063}$	$\underline{0.319}^{\pm 0.065}$

- MAE, RMSE, r^2, and R^2 are better,
- Note 1: Small performances due to small time range and small quantity of data ...
- Note 2: Evaluation on longer duration cannot be evaluated (no existing surveys),
- Note 3: Robustness to domain change have not been evaluated (insufficient number of surveys).

Outline

Conclusions

Take away message:

- A new approach to predict socio-economic Indicator Variations,
- Consider spatio/temporal contexts
- Better than the State-Of-The-Art.

Perspectives:

- Multi-sources and multi-modalities...