



HAL
open science

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

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

► **To cite this version:**

Robin Jarry, Marc Chaumont, Laure Berti-Equille, Gérard Subsol. Predicting Socio-economic Indicator Variations with Satellite Image Time Series and Transformer. MVEO 2024 - Workshop on Machine Vision for Earth Observation and Environment Monitoring, Nov 2024, Glasgow, United Kingdom. lirmm-04895134v1

HAL Id: lirmm-04895134

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-04895134v1>

Submitted on 17 Jan 2025 (v1), last revised 20 Jan 2025 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

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

Robin JARRY¹, Marc CHAUMONT^{1,2},
Laure BERTI-ÉQUILLE³ Gérard SUBSOL¹
LIRMM, Univ Montpellier, CNRS ¹, Univ Nîmes², 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

Introduction

State-Of-The-Art

Our proposition

Experiments & Results

Conclusions and perspectives

Context

We^a want to produce world maps that report:

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

^aSocio-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/>

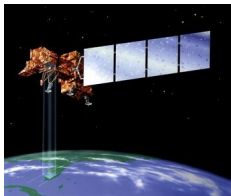
Problem and Solution

Problem:

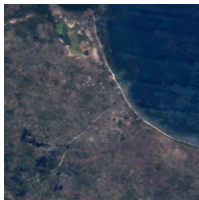
Need to conduct surveys in many places and very often
 → This is costly and time-consuming

Solution:

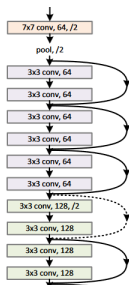
Use satellite images and deep learning.



Landsat 7 Satellite



Village in Tanzania



Outline

Introduction

State-Of-The-Art

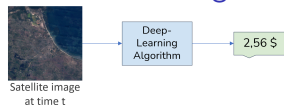
Our proposition

Experiments & Results

Conclusions and perspectives

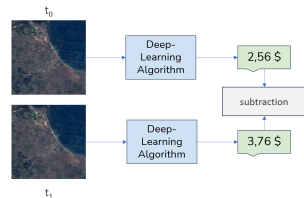
State-Of-The-Art

Prediction at a given date



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

But we are looking at a variation (i.e. \approx 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

Introduction

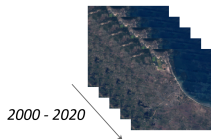
State-Of-The-Art

Our proposition

Experiments & Results

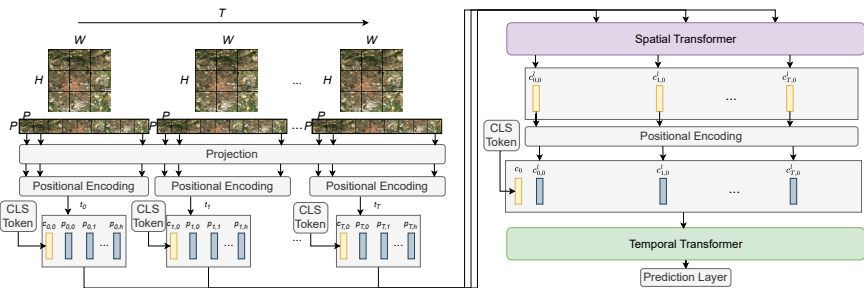
Conclusions and perspectives

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 ($\approx 6.72 \text{ km}^2$),
- ▶ 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



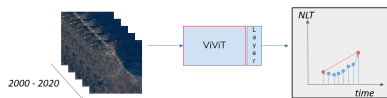
Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lucic, and Cordelia Schmid.

Vivit: A video vision transformer. ICCV'2021.

Ingredient 3: (1) A pretraining

FIRST: A pretraining:

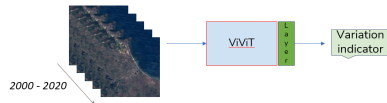
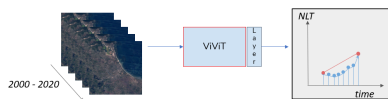
- ▶ On the whole 2000 - 2020 duration,
- ▶ To predict nighttime light time (NLT) **series**.
Note: NLT is correlated to our socio-economic indicator.



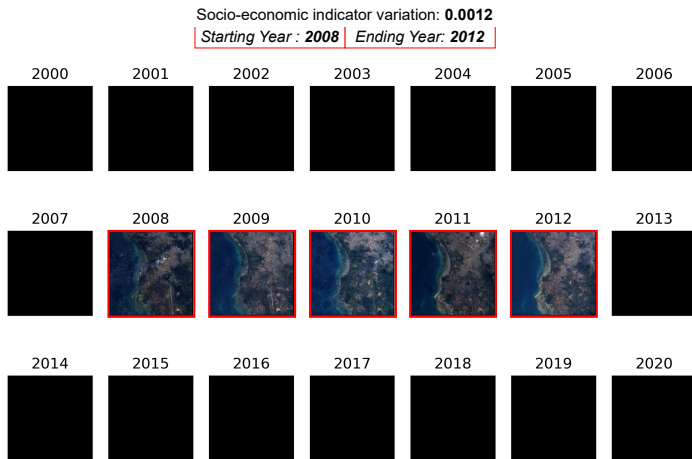
Ingredient 3: (2) Finetuning

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

Introduction

State-Of-The-Art

Our proposition

Experiments & Results

Conclusions and perspectives

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= 5×10^{-4} , ...,
- ▶ ViVit 10 millions parameters, 4 Nvidia V100 GPU.

Results:

	$MAE \downarrow$	$RMSE \downarrow$	$r^2 \uparrow$	$R^2 \uparrow$
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	<u>$0.460^{\pm 0.013}$</u>	<u>$0.366^{\pm 0.020}$</u>	<u>$0.328^{\pm 0.063}$</u>	<u>$0.319^{\pm 0.065}$</u>

- ▶ 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

Introduction

State-Of-The-Art

Our proposition

Experiments & Results

Conclusions and perspectives

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...