Plan Overview

A Data Management Plan created using DMPonline

Title: Mixed Methods

Creator: Roger Carlsson

Principal Investigator: Riksbankens Jubileumsfond

Affiliation: Halmstad University

Template: Halmstad University template

Project abstract:

Deep convolutional neural networks (CNNs) have been shown to predict poverty and development indicators from satellite images with surprising accuracy. This paper presents a first attempt at analyzing the CNNs responses in detail and explaining the basis for the predictions. The CNN model, while trained on relatively low resolution day- and night-time satellite images, is able to outperform human subjects who look at high-resolution images in ranking the Wealth Index categories. Multiple explainability experiments performed on the model indicate the importance of the sizes of the objects, pixel colors in the image, and provide a visualization of the importance of different structures in input images. A visualization is also provided of type images that maximize the network prediction of Wealth Index, which provides clues on what the CNN prediction is based on.
Index Terms—Poverty prediction, Deep Convolutional Neural Networks, Satellite Images, Explainable AI

ID: 141077

Start date: 01-01-2020

End date: 29-12-2024

Last modified: 17-04-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Mixed Methods

Administrativa projektuppgifter / Administrative project information

Version

1.0

Datum / Date (Version)

2024-01-31

Projektnamn / Project name

Mixed Methods

Huvudman / Head of project

Ola Hall, Lunds Universitet och Thorsteinn Rögnwaldsson Halmstad Högskola

Medverkande organisationer / Participating organisations

Lunds University and Halmstad University.

Primärforskare / Primary researcher

Ola Hall, Lunds University and Thorsteinn Rögnwaldsson Halmstad University
Mattias Ohlsson Lunds University, Roger Carlsson och Hamid Sarmadi, Halmstad University.

Medverkande forskare / Participating researcher

?

Kontaktperson / Contact person

Ola Hall, Lunds University

Forskningsfinansiär / Research funder

Riksbankens Jubileumsfond, Sweden

Projekt-/Diarienummer / Project diary number

F 2019/71

Projektbeskrivning / Project description

To investigate Demographic Health Surveys (DHS) points-data versus stellite images for given geographical corresponding areas in order to predict and evaluate wealth; given position data and images for both daylight and nighttime light.

Beskrivning av data / Description of data

Hur kommer data samlas in, skapas eller återanvändas? /*How will data be collected, created or reused?*

The data is primarily divided in two sets: Images and DHS-points with Meta-data. Images is in turn divided as satellite daylight and nighttime images. The images is downloaded using the tool Google Earth Engine using own Python code. The images originates from various defined areas primarily Tanzania and Some surrounding areas. Daylight images are postprocessed after download of geographical big images down to smaller images 224x224 color images suitable for Neural Network (NN) training.
Then two other datasets are generated: Labeldata as lightvalues calculated from VIIRS satellite night-time images and datasets with daylight images combined with DHS-data corresponding to DHS geographical points. These datasets are the feed for NN trainings.
Reuse are the subdivided images, the Labeldata (if not regenerated) and the DHS-datasets (if no regenerated).
The data resides in the folder: Mixed_Methods. Under that folder daytime images (224x224x3 pixels) are found in: Sentinel_p25_224px_images and in Sentinel_p45_224px_images. Nighttime light images are found in folder: VIIRS_images; Generated datasets are found in folder: Datasets and generated Labeldata in folder: NN_labels.

Vilka typer av data kommer skapas och/eller samlas in, vad gäller dataformat och mängd/volymer data? /*What types of data will be created and/or collected, considering data format and volume of data?*

The data created is: Images in .tif format and csv files (Labeldata and datasets). Daylight images postprocessed is in .tif (Geotiff RGB) format and nigttime images as .tif (black and white format), the program tiffinfo can be used to investigate the format. The volume of all data is approximately up to 1 TB.

Dokumentation och datakvalitet / Documentation and data quality

Hur kommer materialet att dokumenteras och beskrivas med tillhörande metadata om struktur, standarder och format för beskrivningar av innehållet, insamlingsmetod m.m.? / *How will the material be documented and described with associated meta data about structure, standards and format for describing the content, method of collection etc.?*

Current data, source code and results will be locally stored in "Box" SUNET with version control.
Datahandling and procedures are documented using UML with the program: Visual Paradigm as a .vpp file.

Hur kommer kvalitet hos data säkerställas och dokumenteras (till exempel repeterade mätningar, validering av datainmatning, m.m.)? / *How will the quality of data be controlled and documented (for example repeated samples of measurements, validation of data collection etc.)?*

The data stored is original data extracted and processed with own code.

Lagring och säkerhetskopiering / Storage and backups

Hur säkerställs lagring och säkerhetskopiering av data och metadata under forskningsprocessen? /*How is the storage and safety back up of the data and metadata controlled during the research process?*

The basic storage is under central research storage at servers at Halmstad University. Most of that data will be available in SUNET.

Har IT eller berörd forskningsdataadministratör/data manager kontaktats i dessa frågor? /*Have IT or affected research data administrator / data manager been contacted in these matters?*

- Nej / No

Rättsliga och etiska aspekter / Legal and ethical matters

Hur säkerställs att data hanteras enligt de rättsregler som gäller till exempel hantering av personuppgifter, sekretess och immaterialrätt? / *How do we ensure that data is handled in accordance with applicable legal rules that apply to, for example the handling of personal data, confidentiality and intellectual property rights?*

Not applicable, all data is open.

På vilket sätt säkerställs det att data hanteras på rätt sätt utifrån etiska aspekter? /*In what way it is ensured that data is handled correctly based on ethical aspects?*

There are no ethical aspects with this data.

Finns etikprövning eller kommer det att göras en ansökan? /*Are there any ethical review or will there be an application?*

- Nej / No

Kommer projektet innebära att du hanterar personuppgifter? /*Will the project mean that you will handle personal data?*

- Nej / No

Tillgängliggörande och långtidsbevarande / Accessibility and long term preservation

Hur, när och var kommer forskningsdata eller information om data (metadata) att tillgängliggöras? Finns det eventuella villkor, embargon och begränsningar kring tillgång till och återanvändning av data att ta hänsyn till? / *How, when and where will the research data or information about data (metadata) be made available? Are there any terms, embargos and limitations regarding access to and reuse of data?*

The data will be for usage stored in SUNET database as long as possible...

På vilket sätt säkerställs långsiktigt bevarande och av vem? Hur kommer urval av data för långtidsbevarande att göras? / *In what way is long-term storage safeguarded, and by whom? How will the selection of data for long-term storage be made?*

The data is stored in SUNET as long as SUNET will.

Kommer det att krävas särskilda system, mjukvaror, källkod eller andra typer av tjänster för att kunna förstå, ta del av eller använda/analysera data långsiktigt? / *Will this require special systems, software, source code or other types of services to understand, access or use / analyze data in the long term?*

Python public libraries can be used, especially Google Earth Engine.

På vilket sätt säkerställs användning av beständiga identifierare (PID) till exempel DOI? /*In what way are the use of persistant identifiers (PID) ensured?*

Not applicable.

Ansvar och resurser / Responsibility and resources

Vem ansvarar för datahanteringen och eventuellt stödjer arbetet med detta under arbetet med forskningsprojektet? Vem har ansvar för fortsatt förvaltning, och långtidsbevarande av data efter projektavslut? / *Who is responsible for data management and (possibly) supports the work with this while the research project is in progress? Who is responsible for data management, ongoing management and long-term storage after the research project has ended?*

Ola Hall, Lund University.

Vilka resurser (kostnad, arbetsinsats eller annat) kommer att krävas för datahantering (inklusive lagring, säkerhetskopiering, tillgängliggörande och hantering för långtidsbevarande)? Vilka resurser kommer behövas för att tillse att data uppfyller FAIR-principerna? (FAIR = Findable, Accessible, Interoperable, Readable, internationella krav för datahantering) / *What resources (costs, labour input or other) will be required for data management (including storage, back-up, provision of access and processing for long-term storage)? What resources will be needed to ensure that data complies the FAIR-principles?*

Nothing special.