Generating Interpretable Poverty Maps using Object Detection in Satellite Images

Analysis of the work of Kumar Ayush

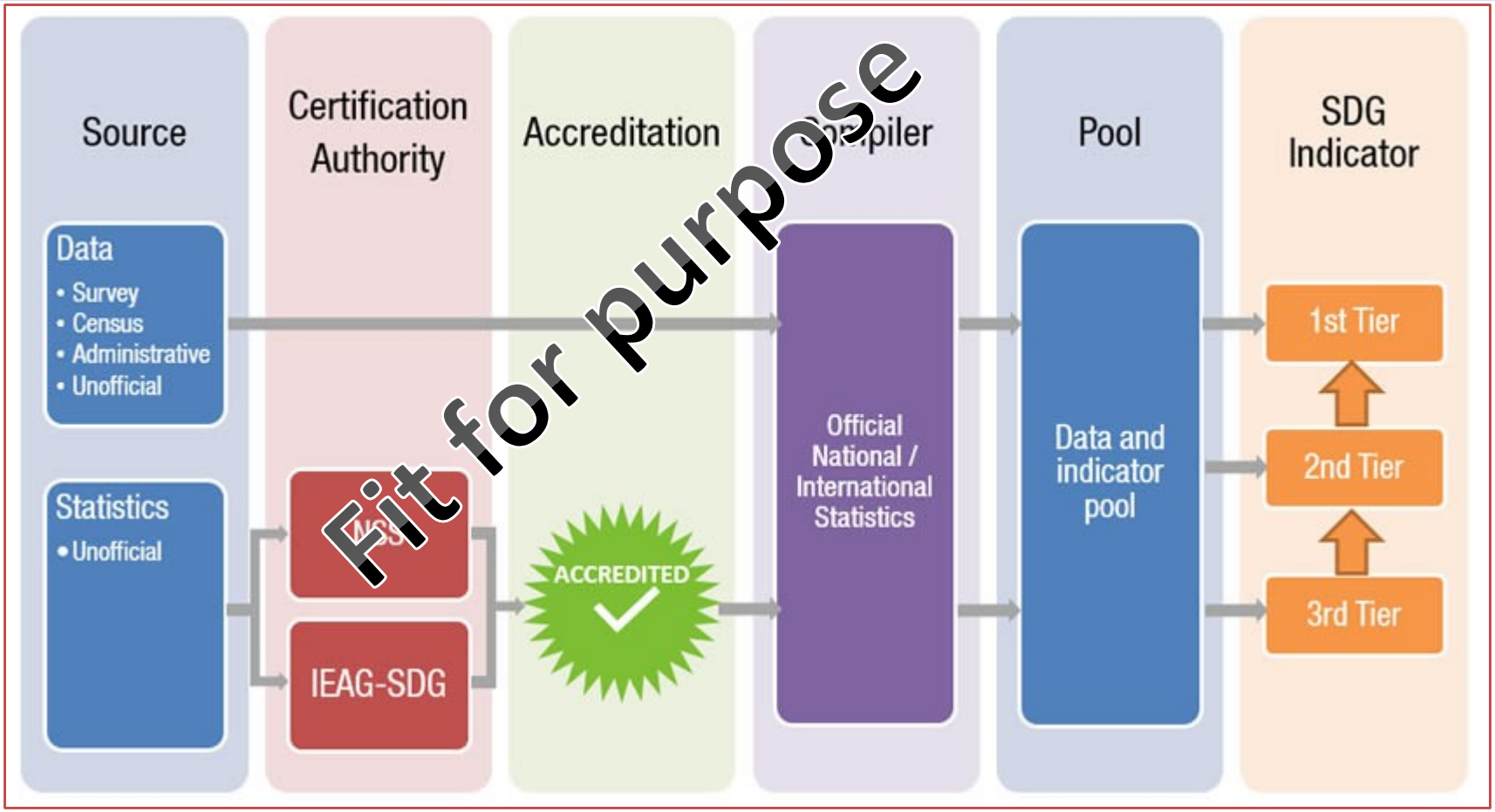
Stanford University

United Nations

GOAL 2: Zero Hunger
GOAL 3: Good Health and Well-being
GOAL 4: Quality Education
GOAL 5: Gender Equality
GOAL 6: Clean Water and Sanitation
GOAL 7: Affordable and Clean Energy
GOAL 8: Decent Work and Economic Growth
GOAL 9: Industry, Innovation and Infrastructure
GOAL 10: Reduced Inequality
GOAL 11: Sustainable Cities and Communities
GOAL 12: Responsible Consumption and Production
GOAL 13: Climate Action
GOAL 14: Life Below Water
GOAL 15: Life on Land
GOAL 16: Peace and Justice Strong Institutions
GOAL 17: Partnerships to achieve the Goal

Data Revolution for SDG

In March 2015, presidents and prime ministers around the world signed up to the United Nations' 2030 Agenda for Sustainable Development. That agenda is the most ambitious development plan ever conceived by the UN. The 169 targets cover just about every dimension of development imaginable, including no less than the total eradication of extreme poverty by 2030. This programme foresees that 232 statistical indicators or performance metrics will be produced by every country in the world to benchmark progress towards the Sustainable Development Goals (SDGs).



Meaning of Fit for Purpose

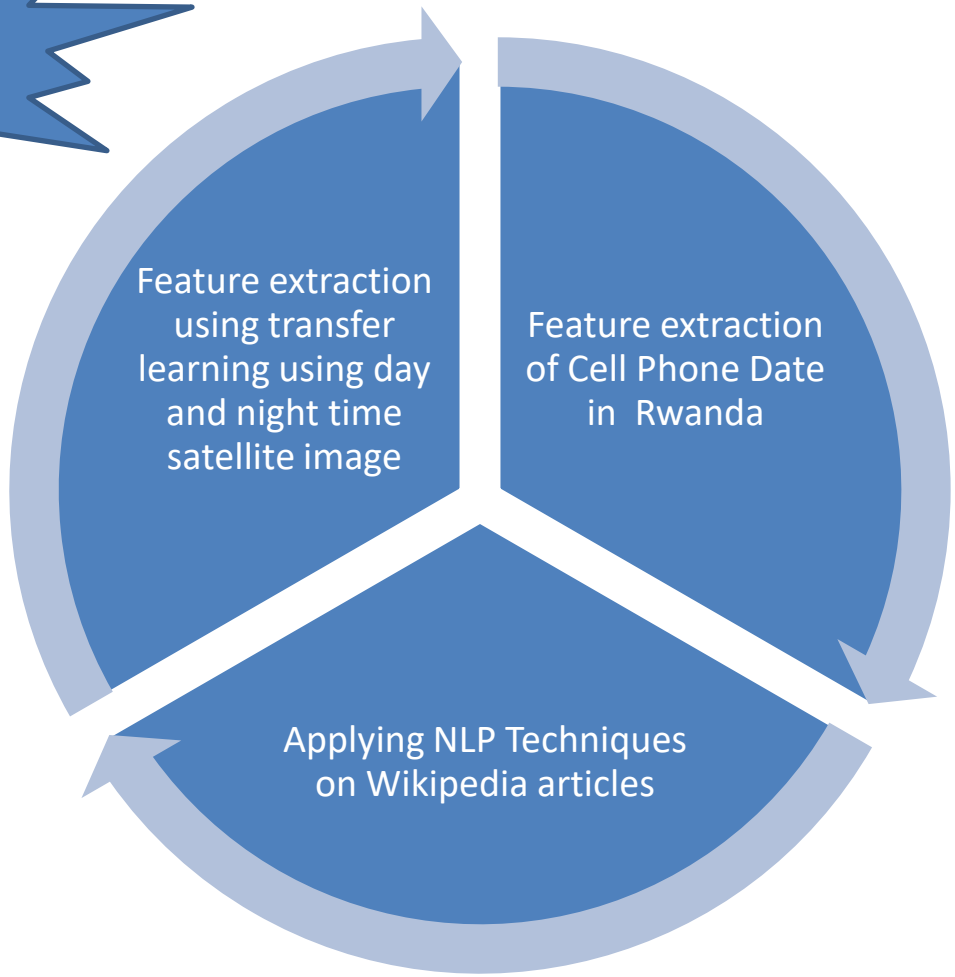
- Fit for purpose means that an indicator or statistic adheres to pre-defined quality and metadata standards. These standards would be set in advance and open to all. Prospective compilers of official SDG indicators would be required to guarantee that they can supply those indicators for, at least, the lifetime of Agenda 2030. In practical terms, this means being able to supply, at a minimum, the statistic on an annual basis for the years 2010–2030. Their input data and methodologies must also be non-proprietary, available to all and open to scrutiny (subject to sensible confidentiality constraints).

The Core Problem Statement and immediate impact

- Lack of frequent, reliable local-level information/data on economic livelihoods
- This impacts the ability of governments and other organizations to predict poverty appropriately.

Few known approaches to solve the problem

Only done in African Countries



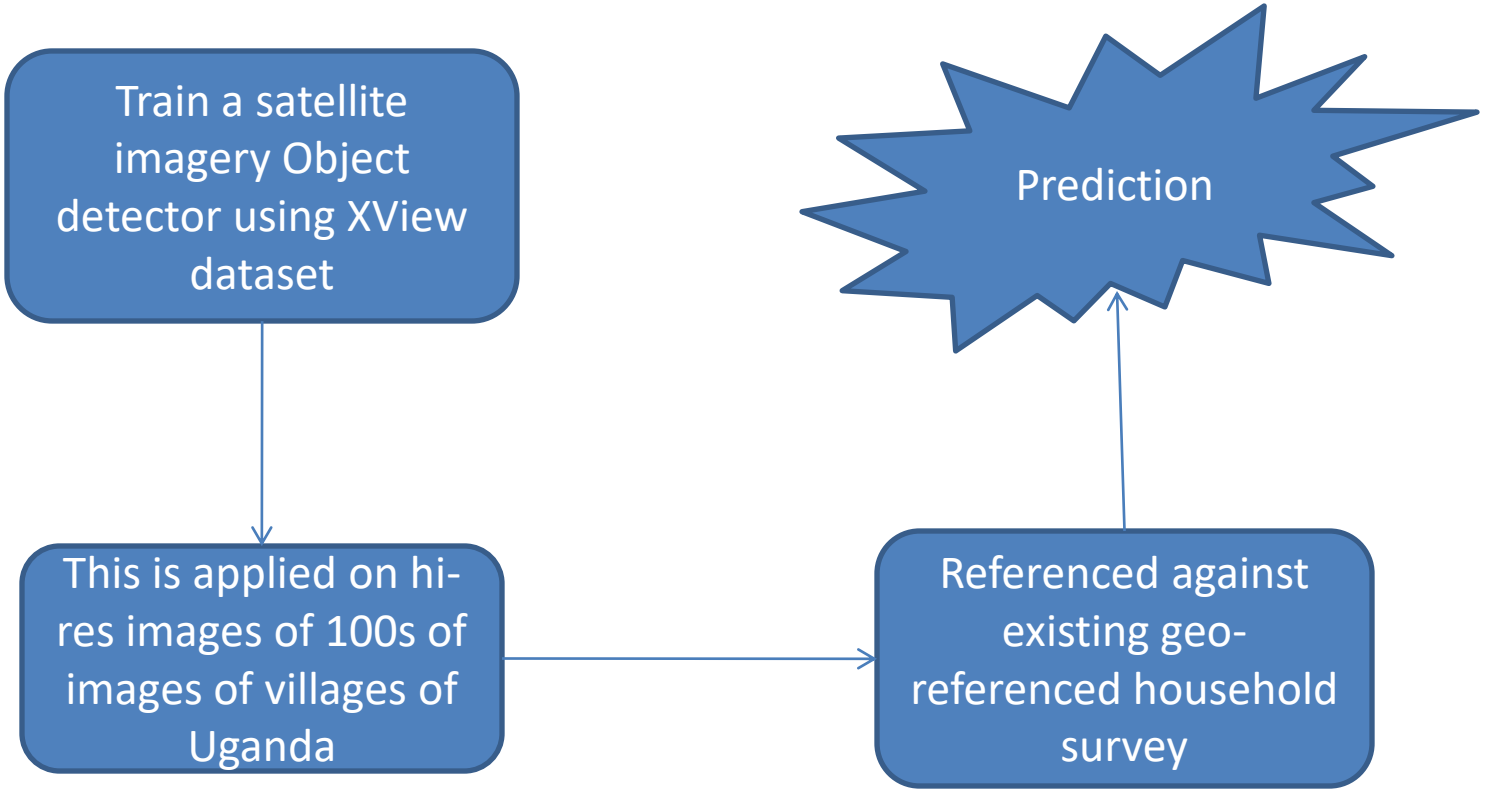
Issues with the Current Approaches

- The existing approaches does not focus on consumption patterns, rather it focuses on parameters like asset wealth.
- Interpretability of model parameters which is relevant for decision makers.

Scope of the paper

- Interpretable computational framework for predicting local level consumption expenditure.
- This is done using high resolution satellite image taken during daytime [30 cm]

High Level Approach taken in the paper



xView is one of the largest and most diverse publicly available overhead imagery datasets for object detection

Detailed Problem Statement

Let

- $\{(H_i, L_i, y_i, c_i)\}^{N}_{i=1}$ -> Set of N villages surveyed.
- $c_i = (c^{lat}_i, c^{long}_i)$ -> Latitude and Longitude of cluster i.
- $Y_i \in R$ where R is the poverty index for a given year

Detailed Problem Statement contd ..

- For each cluster *i*
  - *A high resolution satellite image [at cost] is obtained for the survey year H*
    - $H_i \in R^{WXHXB}$ is a WXH image with B channels
  - *A Low resolution satellite image [free] is obtained for the survey year H.*
    - $L_i \in R^{(W/D)X(H/D)X(B/D)}$ is a W/DXH/D image with B channels and D is the scalar differential of resolution between the High and Low Resolution

The Problem Goal

- To learn a regressor f_t to predict the poverty index y_i using L_i *and parts of* H_i *(selected)*
- Find an adaptive data acquisition scheme based on L_i. This adaptive scheme should do the following
  - *Maximize* f_t
  - *Minimize cost [This means minimized number of* L_i

Information on LSMS Dataset

- Data from 2716 households in Uganda [Per LSMS report of Uganda Bureau of Statistics 2011-12]
- Number of clusters N = 320
- For each Cluster : Poverty is measured as daily consumption in USD [aka *LSMS Poverty Score*]

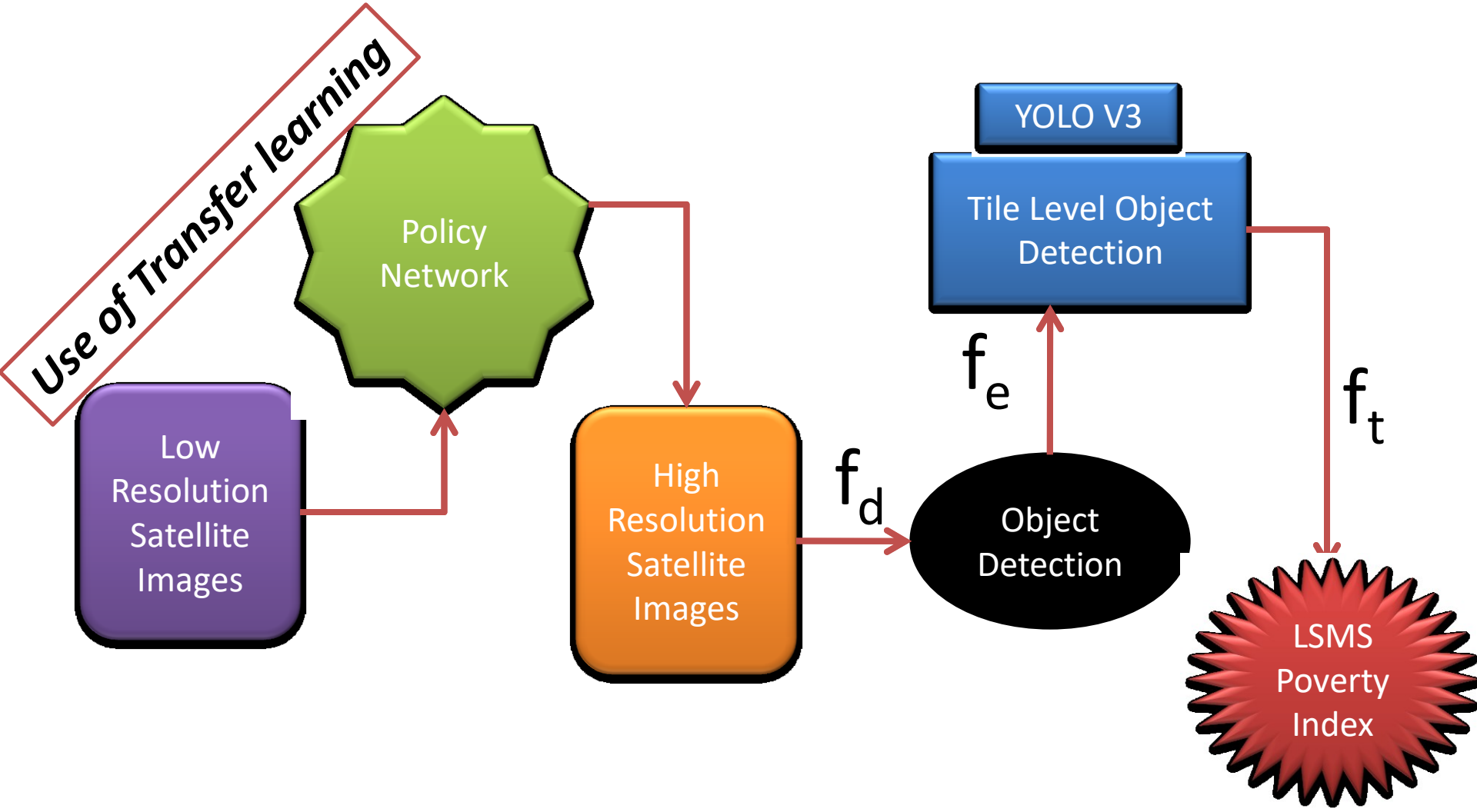
Information about Hi Res Images

- $H_i = \{ H_i \}_{j=1}^{T}$ *is a sequence of high resolution images*
- *T=34×34=1156 images of 1000×1000 pixels each with 3 channels, arranged in a 34×34 square grid.*
- *This corresponds to a 10km×10km spatial neighbourhood centred at the cluster.*

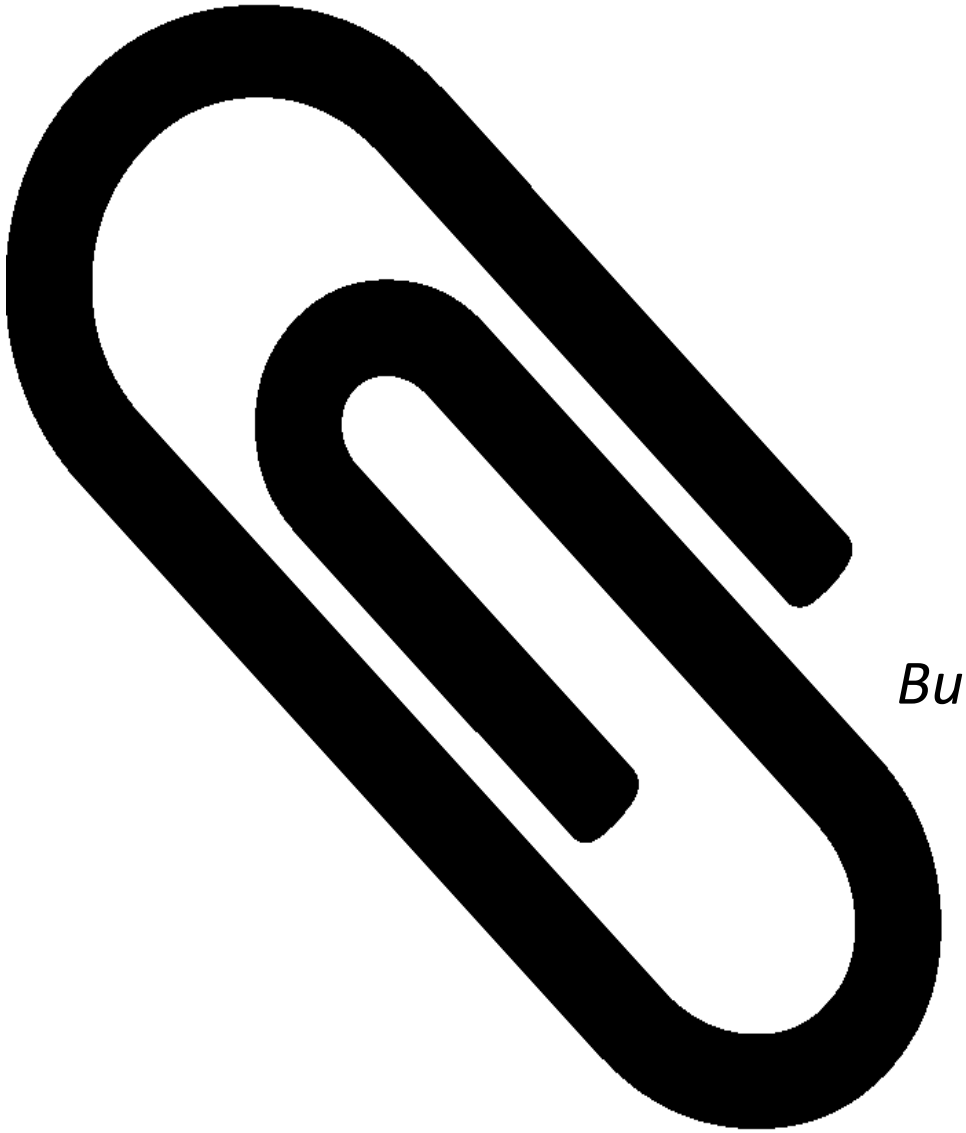
Information about Low Res Images

- *L_i=single image of 1014×1014 pixels each with 3 channels.*
- *This corresponds to a 10km×10km spatial neighbourhood centred at the cluster.*

Solution Approach – General Idea



** Lack of un-annotated data

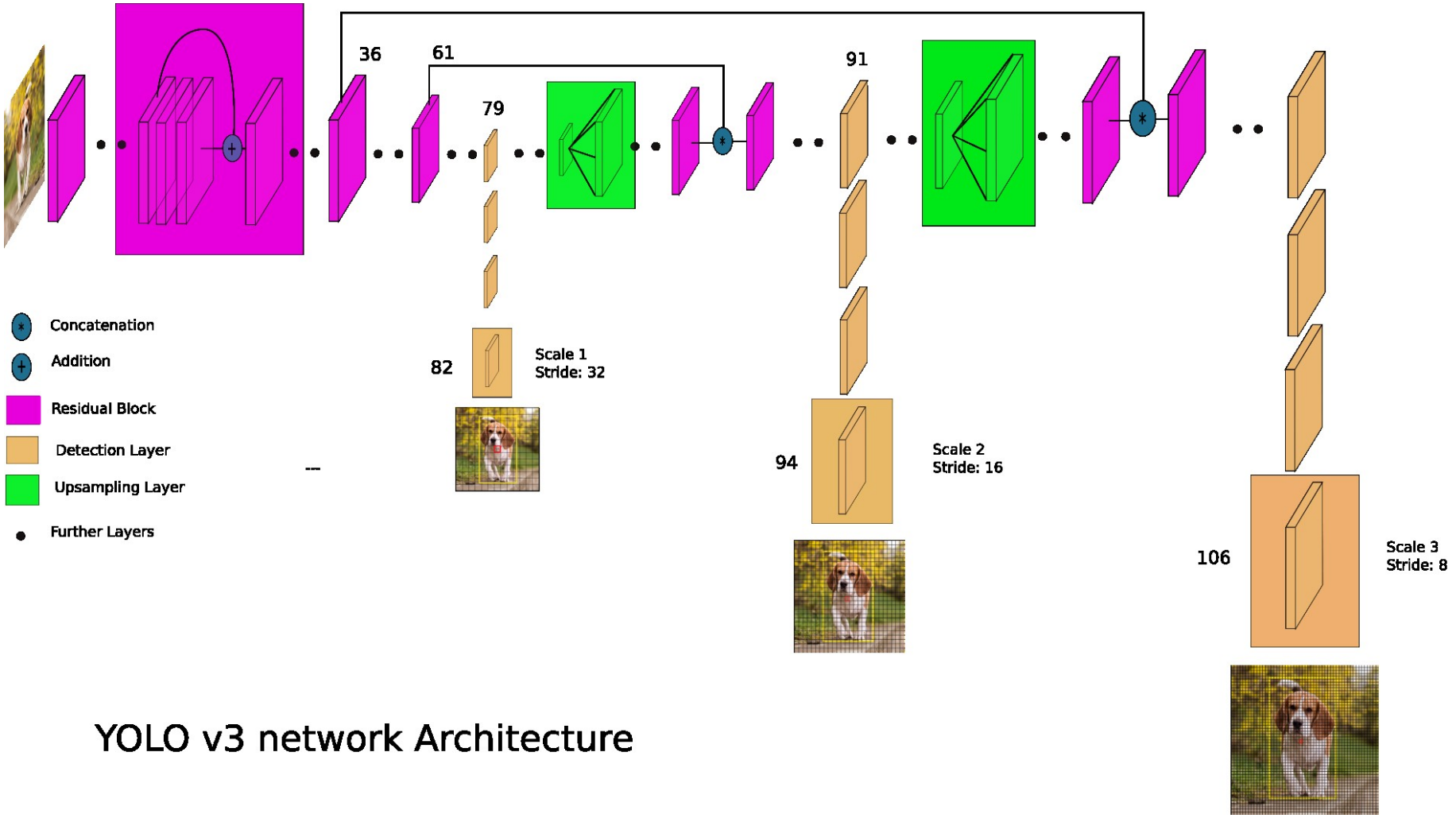


Appendix

Building ideas for the target paper

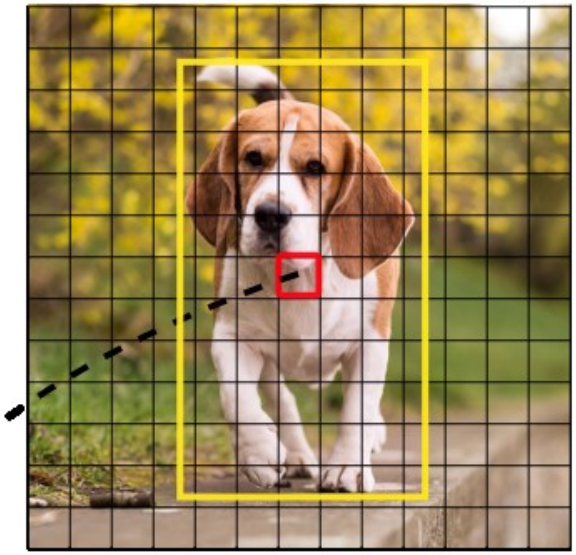
NASA Night Image of West Bengal

2016-17

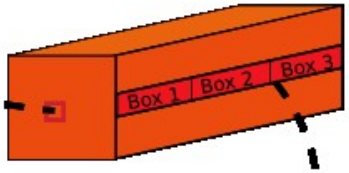


YOLO v3 network Architecture

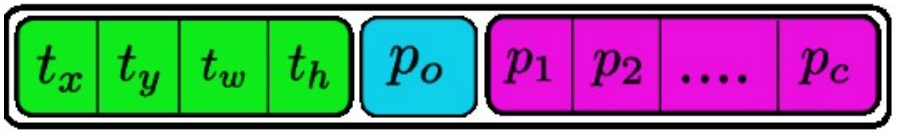
Image Grid. The Red Grid is responsible for detecting the dog



Prediction Feature Map



Attributes of a bounding box



x B

Box Co-ordinates

Objectness Score

Class Scores