جهـــاز التخطيـــط والإحصـــاء
Planning and Statistics Authority
State of Qatar • دولـــة قطـــر

Regional Workshop on the Changing Role of Official Statistics in the State of Qatar:
Why Data Culture Matters

21st -22nd September 2022, Doha, Qatar
Pullman Hotel, Al Thuraya Ballroom - West Bay

ورشـة العمل الإقليمية حول الـدور المتغير للإحصاءات الرسـمية فـي دولة قطر:
ثقافة البيانات مهمة

٢١-٢٢ سبتمبر ٢٠٢٢، الدوحة، قطر
فندق بولمان الدوحة، قاعة الثريا - الخليج الغربي

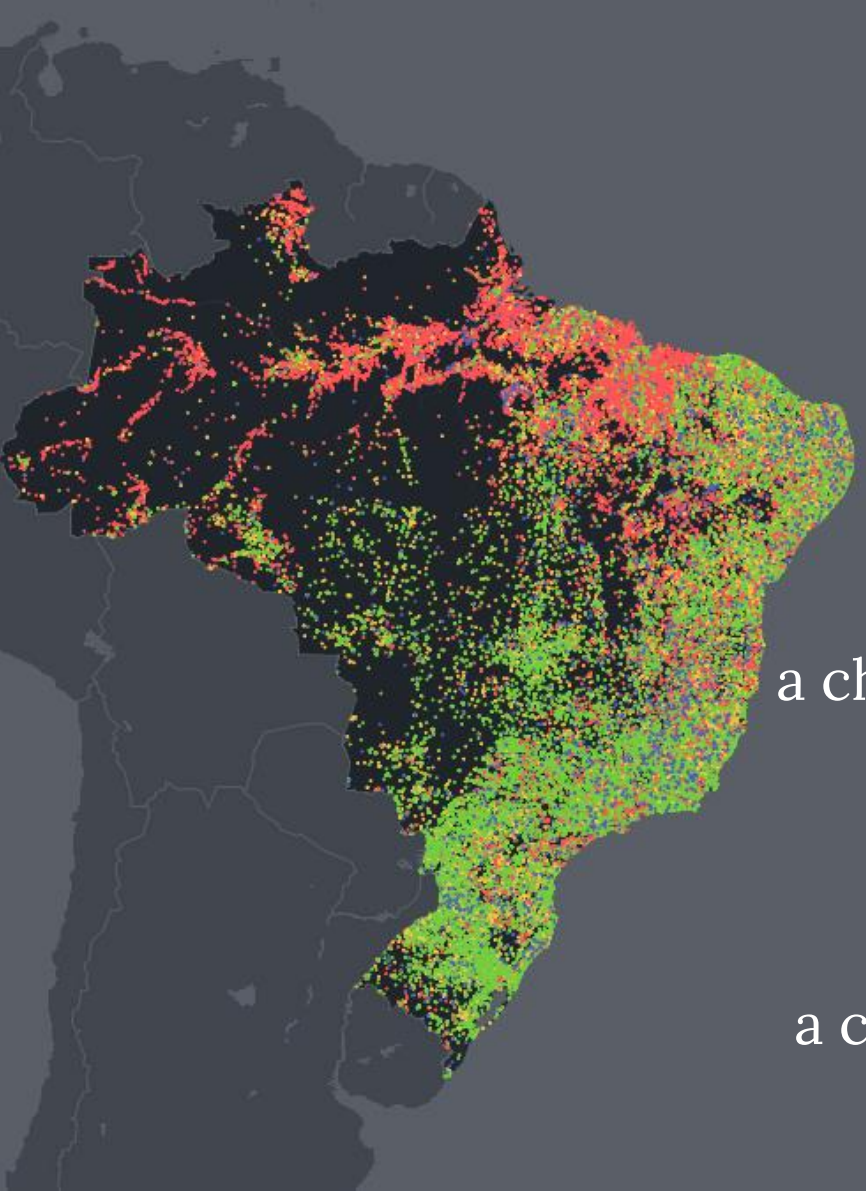
Scalable AI for School Mapping

September 6th, 2022

Introduction

Giga is a partnership between UNICEF and ITU.
Together we are working to connect every young person to the internet.

giga



Each red dot is...

a school unconnected to the internet

denying a child the opportunity to learn and realize their dreams

isolating a community from the world

green dot is...

a school connected to the internet

giving a child the opportunity to learn and realize their dreams

integrating a community with the world

Giga is already one of the most impactful UN initiatives ever, with 1.3 Mn+ students connected globally in 3 years

Giga's traction and global footprint since establishment in 2019

1.3 Mn+
students connected

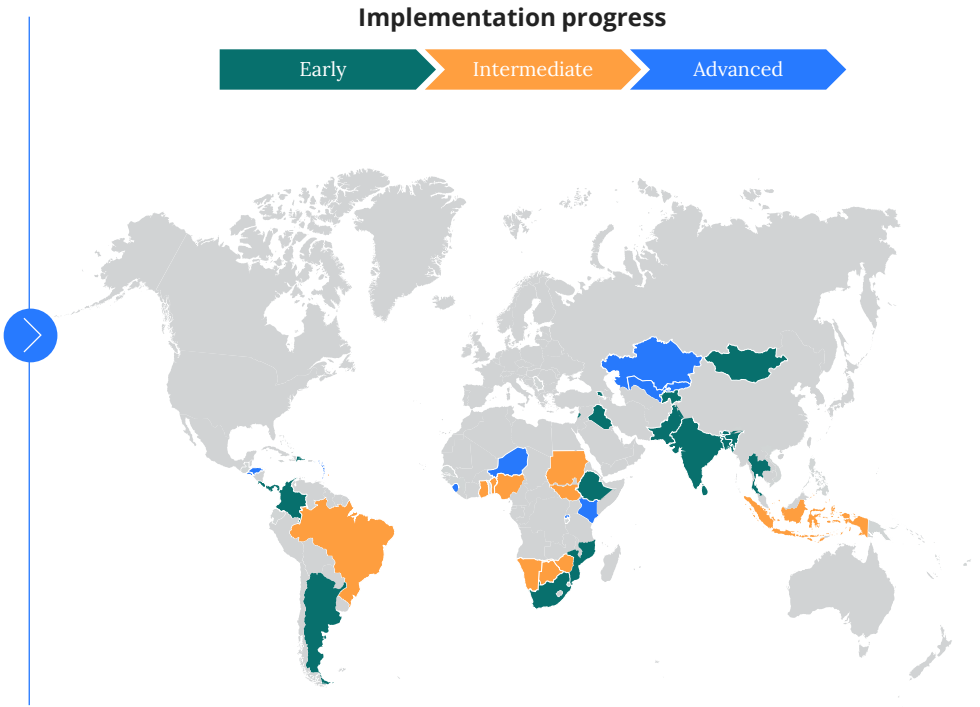
3,715+
schools connected

1.1 Mn+
schools mapped

$25 Mn+
raised

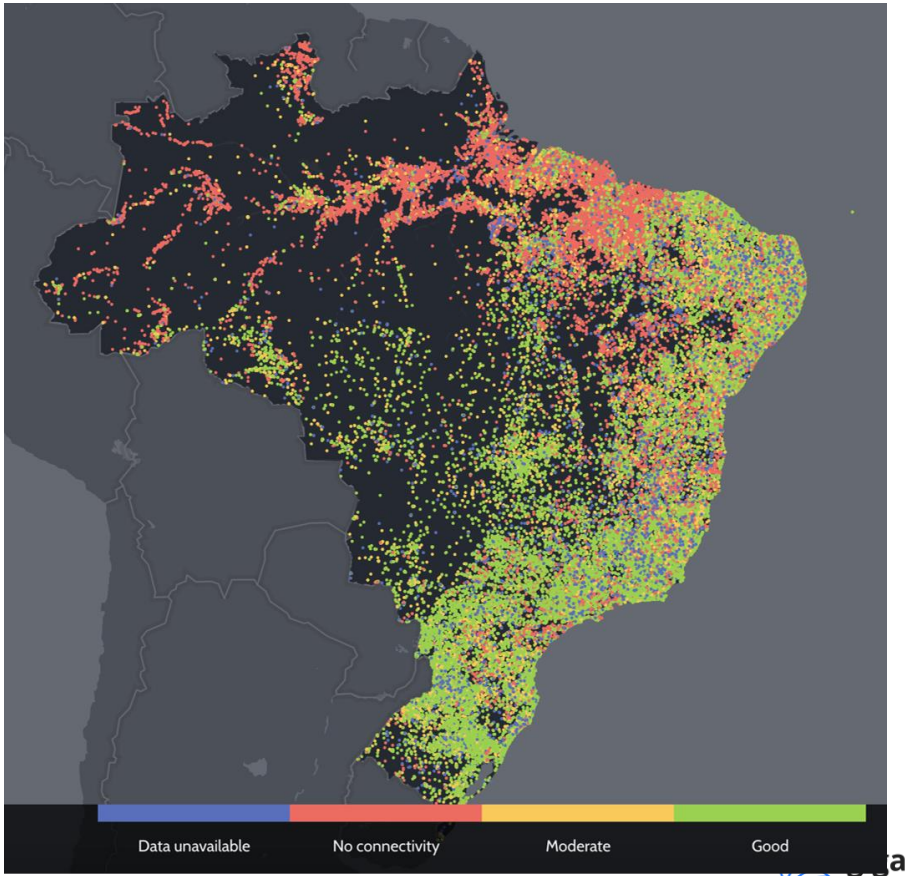
$200 Mn+
funding mobilized

14
partners



School Mapping

- Giga has been developing global database of school locations
- AI has been tested in a few countries
- We have tested different methods – classification and object detection
- For Sudan, we made a good results from the object detection methods
- Based on Sudan project's success, we aim to build a scalable approach



School Mapping in Sudan

- Co-operation with UNICEF Sudan Country Office
- Accurate school location data are not available - Ground data collection is not feasible
- About 6,000 school locations falls within Sudan country boundary were provided
- 2,000 school locations are verified and used in this work

AI Model Development for Sudan

giga

Resources

- High resolution imagery
  - MAXAR/US State Department
  - 0.6m spatial resolution
  - Globally available

- High Performance Computing clusters with GPU support -Dell

MAXAR's cloud free mosaic image

Dell's Rattler cluster spec

Component	Configuration
Servers	48x PowerEdge XE8545 servers 6x PowerEdge C4140 5x PowerEdge R750xa 1x PowerEdge R740
Processors	Intel Xeon Scalable and AMD EPYC™ processors
Accelerators	NVIDIA GPUs
Memory	30TB
Operating System	Red Hat Enterprise Linux
HCA card	NVIDIA Quantum High Data Rate (HDR) 100 and 200 InfiniBand
Storage	931TB HPC BeeGFS High Capacity 1.9PB HPC BeeGFS High Capacity

giga

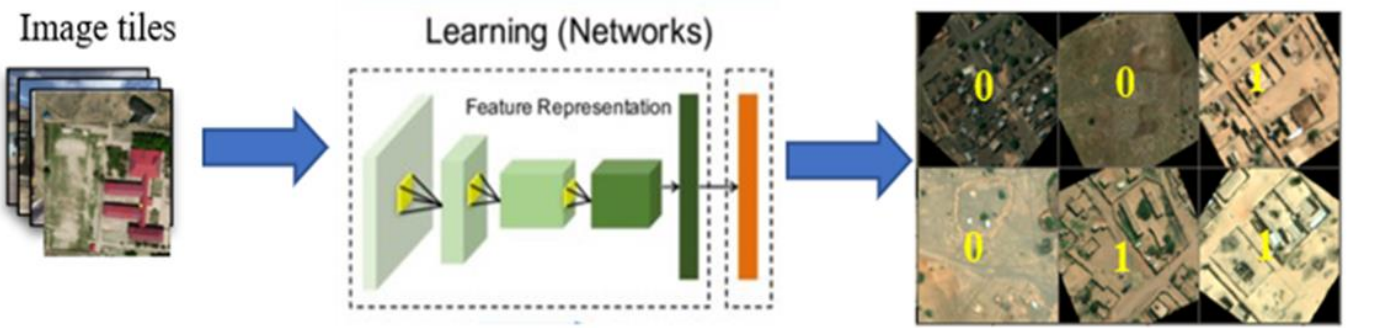
AI Models

- Country models :

Sierra Leone, Niger, Sudan, Honduras, Colombia, Kazakhstan and Kenya
7,000 new schools mapped in Colombia, over 20k new schools mapped in 8 countries

- Regional models : West Africa, East Africa

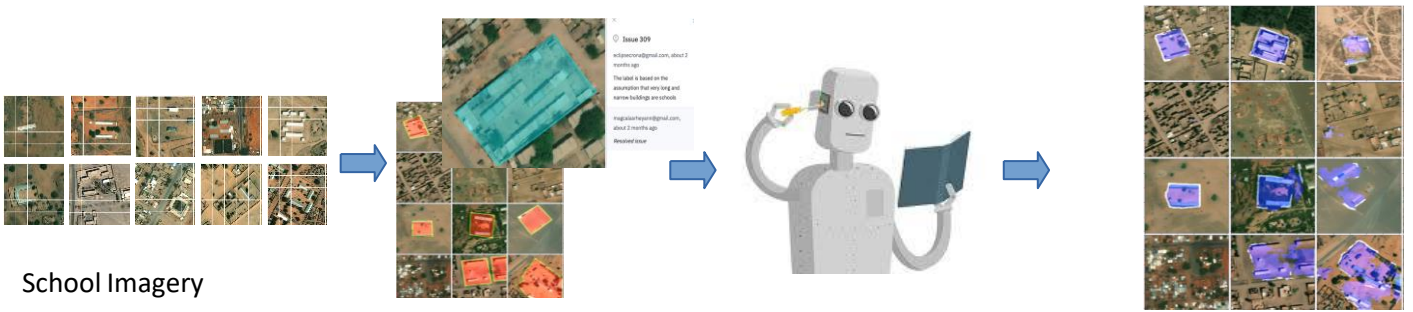
Tile-based Classification model



- Validation accuracy: overall F1 score > 0.9

AI Models

- Object Detection model (YOLO V5) for Sudan



School Imagery

Labeling

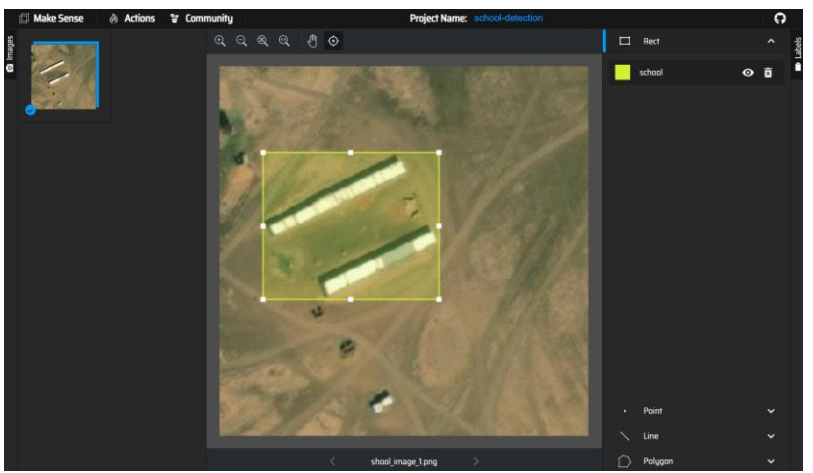
Object based learning

Predictions

- Able to map the exact locations of schools in each image tile

Training Samples

- Decide the accuracy of AI models
- Most labor intensive – major bottleneck of AI scalability
- For Sudan, 2,000 school samples, 3,000 non-school samples are used



- First phase – delineation of exact boundary
- Second phase – simple bounding box

Model Development – First Phase

- YOLOv5 Object detection model
- Trained over 50 epochs
- Precision, recall value of 0.80 and 0.83 on validation data

	Train	Validation
School	4178	522
Non-school	4178	521
Total	8356	1043

Number of Train and Validation samples

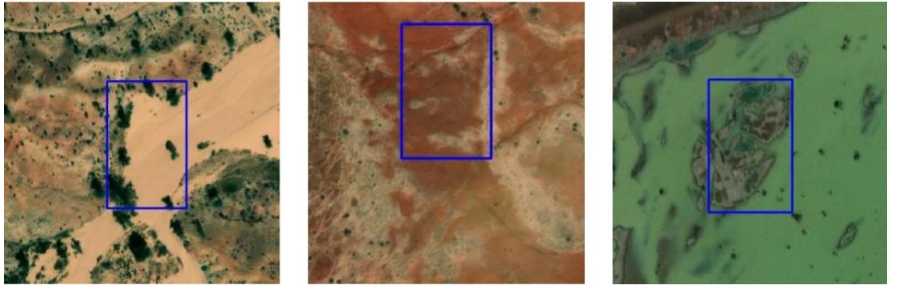
Prediction Results

Error analysis shows many false positives

Industrial buildings

Other buildings with fences

Agricultural fields



Desert area and water body

Model Development – Second Phase

- Data-centric model development approach

Typification of school samples

(a) Parallel double bars

(b) Parallel bars in groups

(c) Buildings with left wing and right wing

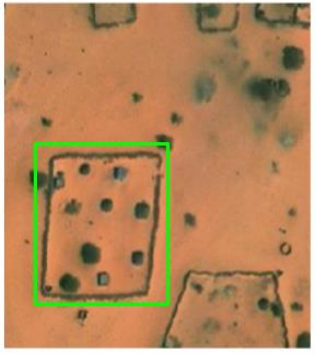
(d) C shape buildings

Model Development – Second Phase

Removing samples causing errors

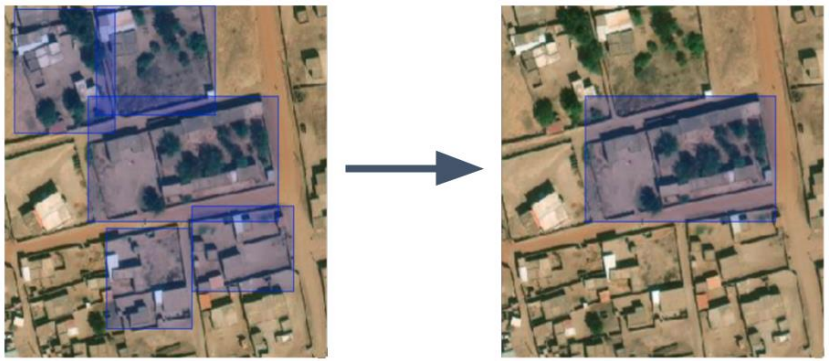
Single-bar school buildings

schools that look like regular buildings

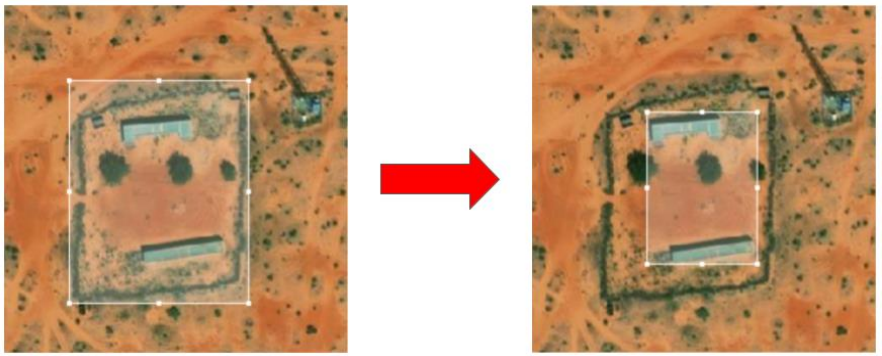


Model Development – Second Phase

correction of wrong labels



modification of existing labels



Model Development – Second Phase

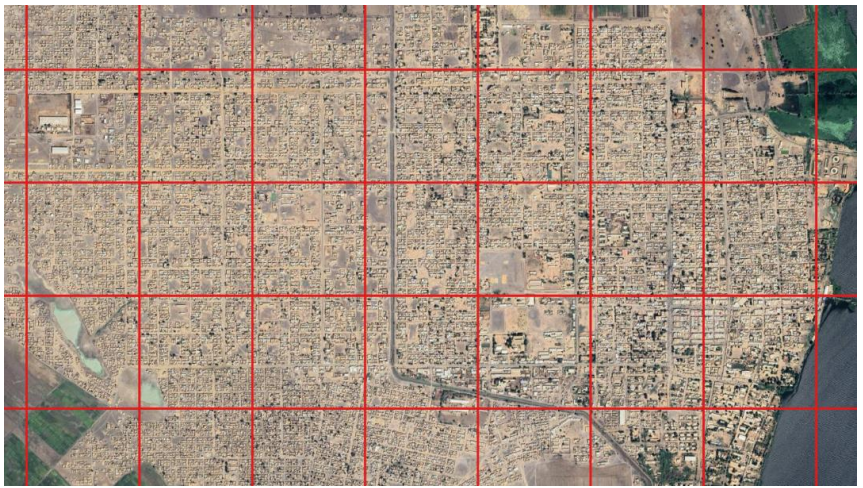
- Trained over 50 epochs
- Precision, recall value of 0.884 and 0.916 on validation data

	Train	Validation
School	1,935	491
Non-school	3,654	917
Total	5,589	1,408

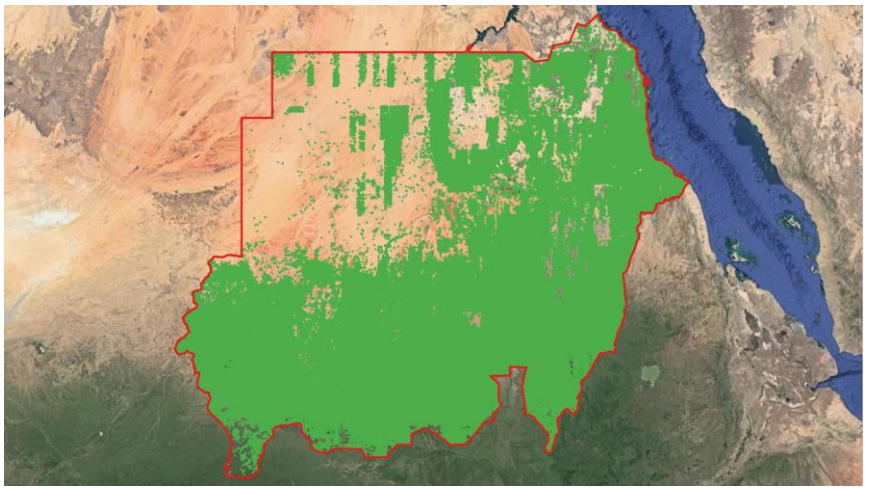
Number of Train and Validation samples

Area Selection

- Large desert area - 5,694,227 grids of 600X600m for the entire country
- Only settlement areas are selected using available building/settlement data
- 1,282,586 grid cells are selected



Grid cells with 600x600m size (1000x1000 pixels)



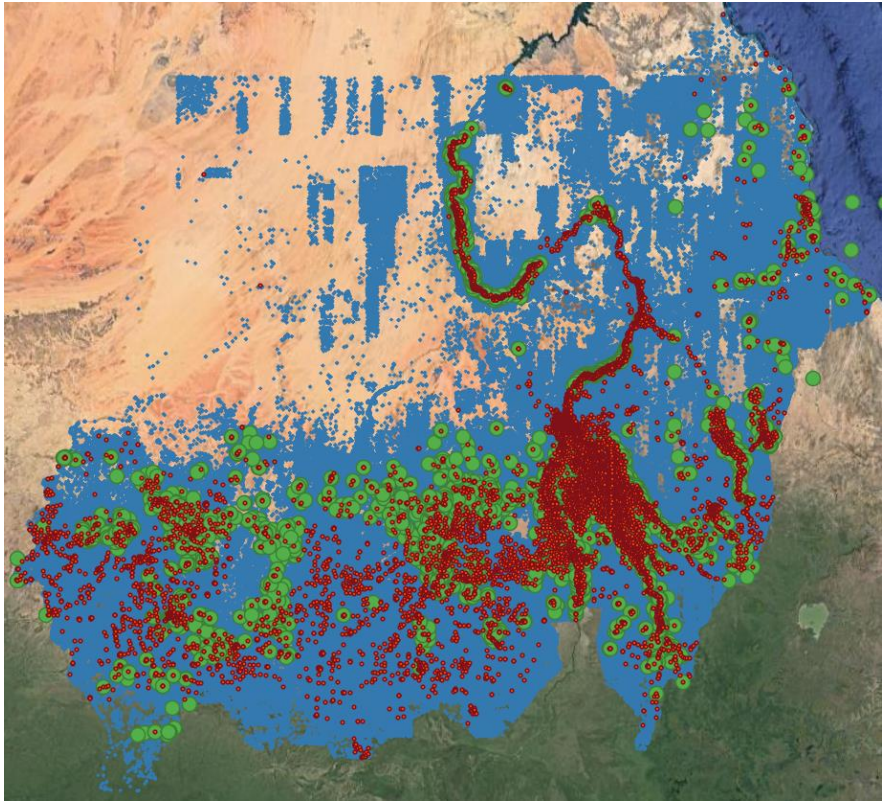
Selected are for image download (green)

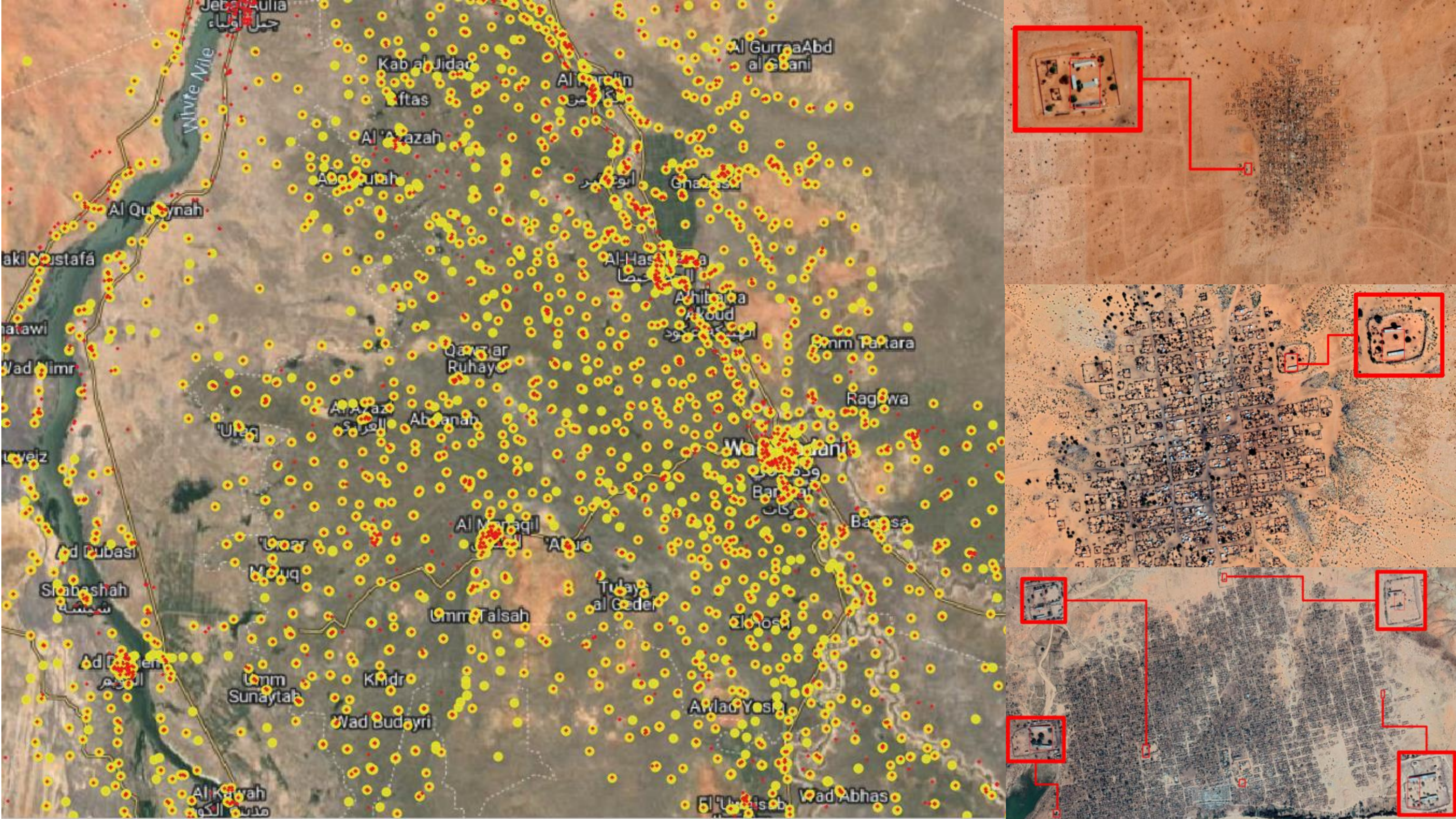
Results

Schools Mapped

20,125 schools (red) are detected out of 1,282,586 satellite image tiles (blue)

Green dots indicates original sample locations



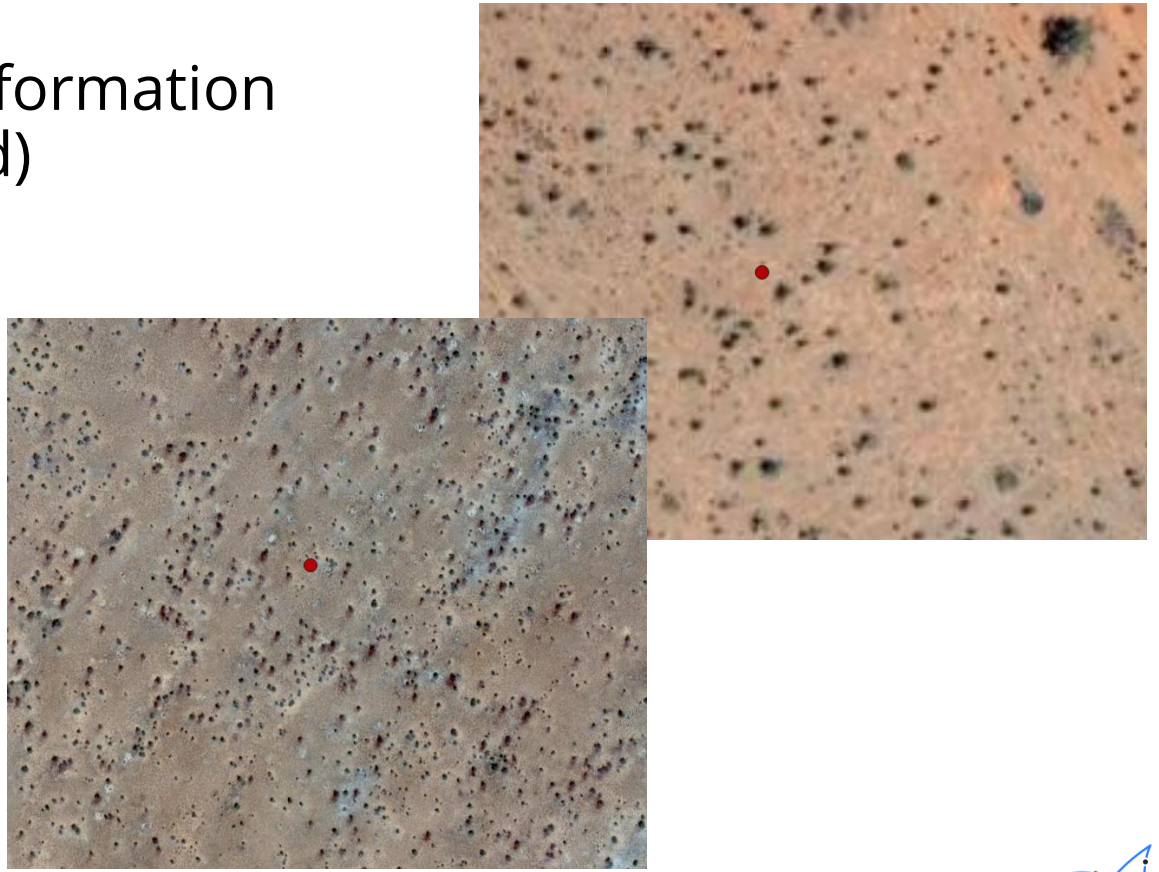

White Nile
Kab al Jidad
Al 'Azazah
Al Qutaynah
Wad Nimr
'Ulaq
Abu Hanab
Umm Tartara
Ad Dubasi
Umm Talsah
Khidr
Umm Sunaytah
Wad Budayri
Awlad Yosif
Wad Abhas
Al Kawah

Results

Singles school (red)
→ clustered multiple schools (green)

Results

Detect incorrect location information from the original data (red)



Results

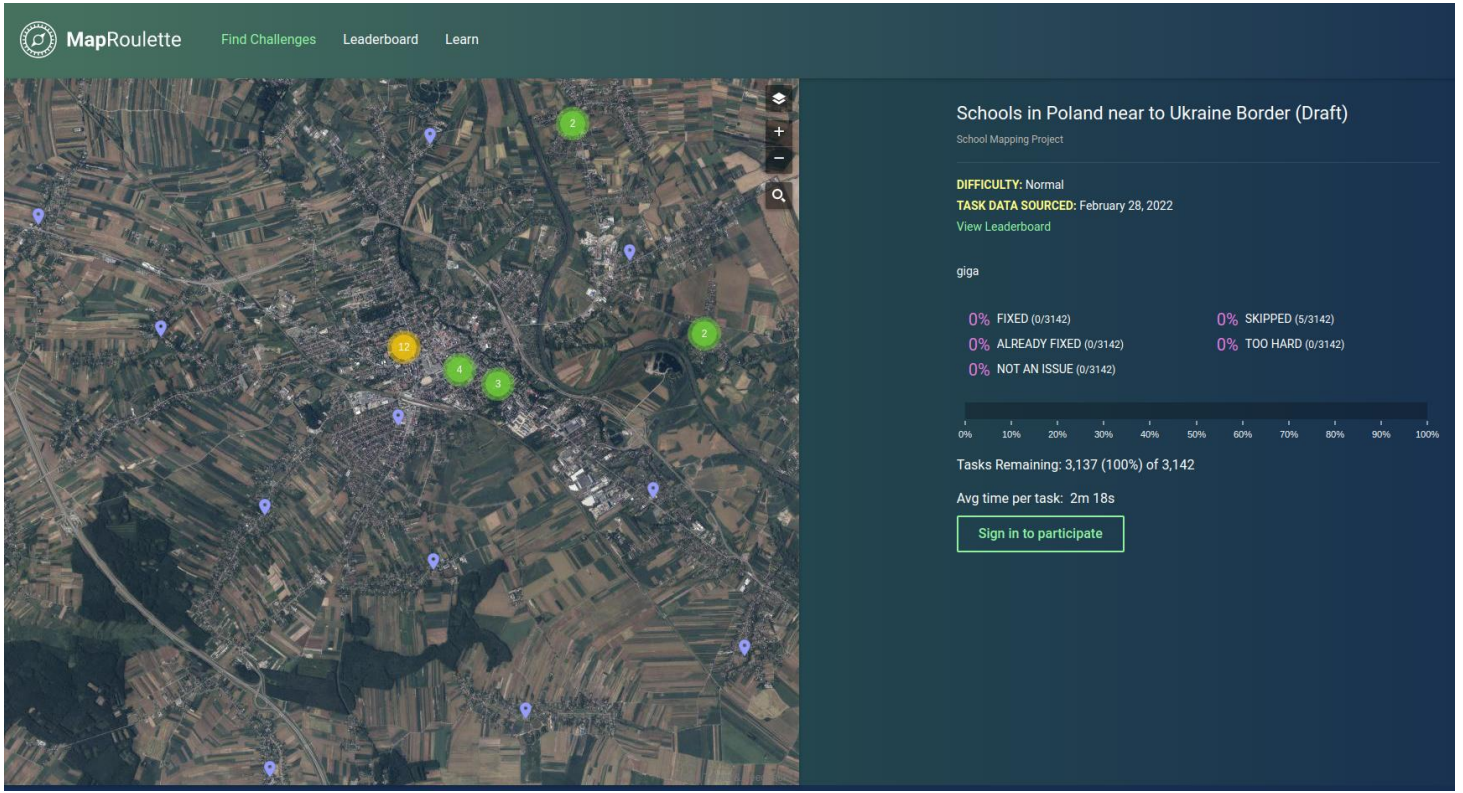
More accurate location information

Next Steps and directions

giga

Cloud Mapping Based Validation

- USAID
- Youth Mappers



Brazil School Mapping

- About 30 % of Brazil school locations are not correct
- Current progress
  - Typifying schools
  - Labeling for training set (With Scale AI)

A sample of typical urban schools

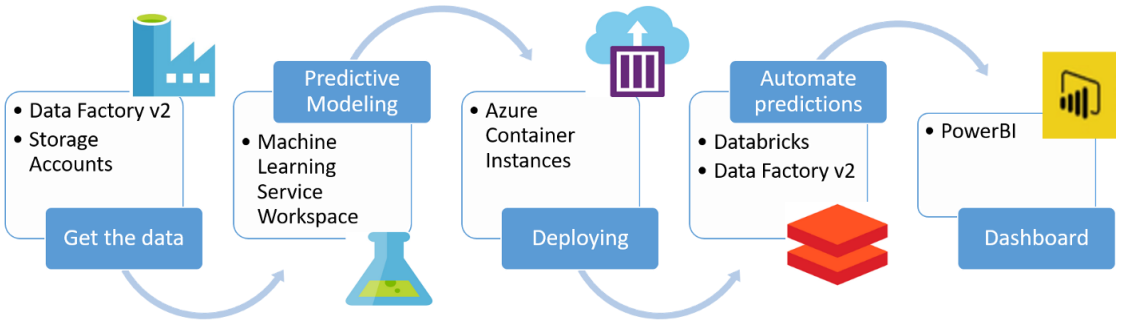
A sample of typical rural schools

Future Directions

AI pipeline

- Automate data processing

- API for Model deployment



Future Directions

Scalable AI

Less human input

Applicable in larger areas

Global school location training data base

Future Directions

Semi-supervised learning for object detection

- Train model with small amount of labeled data
- Generate pseudo-labeled data

A country model to a regional model

Retrain model with pseudo labels acquired in neighboring countries

Example Schools in Colombia, Brazil and Paraguay

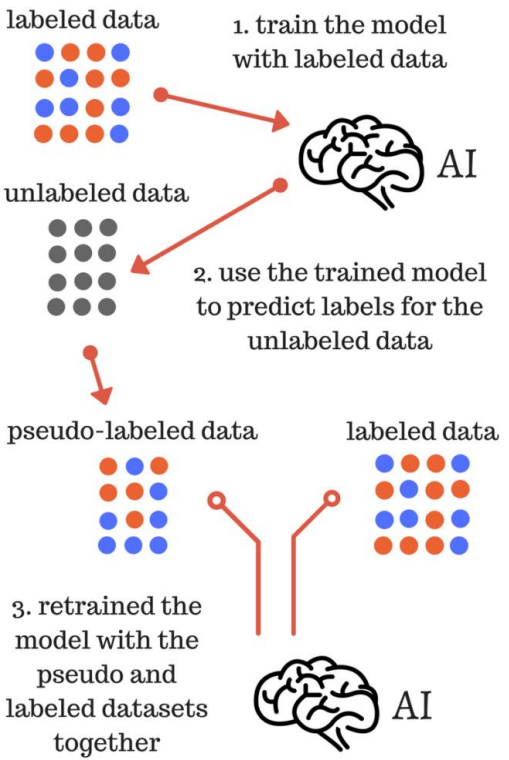
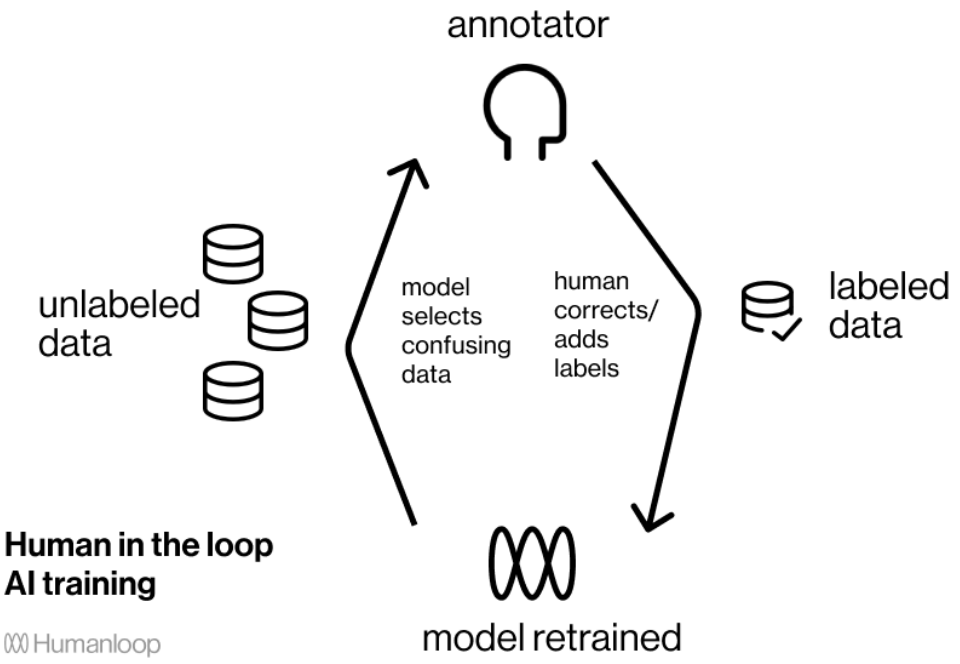


Image source: Deeplearning.net

Future Directions

- Human in the loop modules



Thank You!

Giga
Naroa Zurutuza nzurutuza@unicef.org
Dohyung Kim, Ph.D. dokim@unicef.org
Iyke Maduako, Ph.D. imaduako@unicef.org
Aye Nyein Thaw athaw@unicef.org

UNICEF Sudan
Rasha Obaid robaid@unicef.org