AI-assisted Accurate School Mapping in Developing Countries

Zhuang-Fang Yi, [1] Naroa Zurutuz, [2] Ruben Mendoza, [1] Dohyung Kim, [2]
Martha Morrissey,[1] Drew Bollinger, [1] Jeevan Farias [1]

[1] Development Seed
[2] UNICEF
1226 9th Street NW, 2nd floor
Washington, DC 20001
nana@developmentseed.org

Abstract

UNICEF and Development Seed are working to leverage machine learning, high-resolution imagery, and inexpensive cloud computing to create a comprehensive map of school at the global scale. This paper details a case study using satellite imagery to detect previously un-mapped schools throughout parts of South America, Africa, and South East Asia in partnership with UNICEF Project Connect. This paper focuses on results of school detection models from South America and the Caribbean, but work is in progress extending these models for school detection in Africa and South East Asia. This paper illustrates that a combination of machine learning with satellite imagery for school detection, and humans in the loop validation process can be effective to discover and map schools. After training the South American and Caribbean school model, inference yielded 73,000 predicted school tiles which the Development Seed Data Team validated. From those 73,000 tiles, almost 11,000 schools were added to the map in Colombia and the Caribbean islands, of which approximately 7,000 were unmapped schools. Accurate data about school locations is critical to provide quality education.

Introduction

Development Seed is pleased to present a paper detailing a case study/deployed project about using satellite imagery to detect previously unmapped schools throughout parts of South America, Africa, and Central Asia in partnership with UNICEF Project Connect. This paper focuses on results of school detection models from South America and the Caribbean, but we are working on extending these models for school detection in Africa and South East Asia.

Accurate data about school locations is critical to provide quality education and promote lifelong learning, listed as UN sustainable development goal 4 (SDG4), to ensure equal access to opportunity (SDG10) and eventually, to reduce poverty (SDG1) (UN 2020). However, in many countries educational facilities' records are often inaccurate, incomplete or non-existent. Understanding the location of schools can help governments and international organizations gain critical insights around the needs of vulnerable populations, and better prepare and respond to exogenous shocks such as disease outbreaks or natural disasters (UNICEF 2020).Unfortunately, some national governments still don't know where all the schools in their country are. Crowdsourcing can be used as an alternative for these cases: e.g. through OpenStreetMap (OpenStreetMap 2020), volunteers can map schools at their crowdsourcing platform. These kinds of maps are hard to verify and keep up-to-date, leading to inaccurate, outdated information. UNICEF and Development Seed are working to leverage machine learning, high-resolution imagery, and inexpensive cloud computing to create a comprehensive map of school at the global scale.

Despite their varied structure, many schools have identifiable overhead signatures that make them possible to detect in high-resolution imagery with modern deep learning techniques. One goal of this project was to test the generalizability of these models. We believed that the digital signature of schools in Colombia was close enough to that of neighboring countries that a model trained on Colombia would perform well there. To test this, we used the model developed in Colombia to detect schools in eleven Eastern Caribbean nations: Anguilla, Antigua, Barbuda, British Virgin islands, Dominica, Grenada, The Grenadines, Montserrat, St Kitts and Nevis, St Lucia, St Vincent.

We trained a binary school classifier with existing and cleaned schools dataset in Colombia. We created a Convolution Neural Network (CNN) school classifier based on Xception (Chollet 2017), modified for use on overhead imagery and tuned with a new, highly accurate school dataset that we created for Colombia. We then selected the best-performed model–0.94 area under the ROC curve and 9 percent of the false positive rate–from our nearly 200 training iterations. We then applied this model across Colombia and the Eastern Caribbean islands to detect potentially unmapped schools. Running the model over such a large area of high resolution imagery is a huge task, requiring processing over 52 million DigitalGlobe Vivid imagery tiles (DigialGlobe 2020). Development Seed created an open source tool, chip-n-scale-queue-arranger, to manage high volume satellite imagery inference tasks.

After the model inference, we downloaded the model predicted schools from our inference pipeline. The Develop-

AAAI Fall 2020 Symposium on AI for Social Good.

ment Seed Data Team validated 73,000 predicted school tiles from machine learning. We added about 11,000 schools to the map in Colombia and the Caribbean islands, of which approximately 7,000 were unmapped schools.

Training Data Set

A high-quality training dataset is essential for machine learning models to accurately learn object features. Our first step was to prepare a set of verified school locations, as well as a set of locations that are verified not to contain schools. The CNN will use its knowledge of both to build a model of determining whether any given image contains a school.

School Locations

UNICEF was able to provide a preliminary list of 44,665 school locations in Colombia. Five expert mappers from Development Seed's Data Team, reviewed that dataset, and compared it to the DG Vivid RBG imagery. The Data team classified each location into those where 1) overhead imagery clearly contain schools ('confirmed'), 2) overhead imagery clearly does not contain a school ('not-school'), 3) it is uncertain whether or not overhead imagery contains a school ('unrecognized'). Confirmed schools are observed from the high-resolution satellite imagery and have very clear school features, e.g. building size, shape, and facilities. The unrecognized schools referred to geolocations that were part of the original 44,665 school dataset but that had no clear school features, especially in urban areas with high building density or, in rural areas that can't be distinguished from residential buildings. Another case of unrecognized schools is school building(s) that can not be seen on DG Vivid because of cloud/tree cover. These locations were not used in training the model. The not-schools refers to locations from the original 44,665 school dataset where the expert mappers could not find any school looking buildings at the provided school geolocations 1. As an example, some of the schools were mislocated in the middle of the ocean. This can be because the school geolocation was recorded incorrectly or because the DG Vivid imagery has been updated in particular areas of Colombia after schools were built.

Tasks	Confirmed	Unrecognized	not-school	Total
Data Cleaning	10,951	26,638	7,066	**44,655 schools**
Training dataset for the 1st session	5,904	-	9,092	**14,996 tiles**
Training dataset for the 2nd session	8,716	-	11,192	**19,908 tiles**

Figure 1: School data cleaning and training dataset creation for machine learning.

.

Two categories of datasets, 'school' and 'not-school', were generated as the training dataset for the machine learning model 1. We randomly sampled half of the school geolocations from the validated "confirmed schools", and generated 5,904 tiles as "school" training dataset in the following table. The not-school category is not as trivial. This category can contain forest, grassland or agricultural fields without any buildings, among others. It can also be a building complex or facility that looks very similar to "school" from space, e.g. hospitals, market places, courthouses etc.

To enrich the "not-school" category, and allow our machine learning algorithm to detect real schools more accurately, we queried available hospitals, farmlands, parks, courthouses and marketplaces from OpenStreetMap and added them to the "not-schools" category. We ended up with 9,092 'not-school' tiles, see the above table. Through the machine learning model training iterations, we learned that the model was over-confident in certain areas, and therefore we adjusted the training dataset accordingly. For instance, we randomly included more schools as well as purposely added more regular building tiles that are not schools.

To assess the model performance fairly, our training dataset which contains two categories, school and not-school tiles, was then split into 70:20:10 ratio as train, validation and test datasets. We randomly selected 70% of tiles to train the model, the remaining 20% was used to validate the model. The training set was seen and used by the model intensively to train the school classifier, and the validation set is occasionally seen and used to fine-tune the classifier. The last 10% of the test dataset was not seen by the model, and is used to evaluate the model performance. To explore the geo-diversity of the 'schools' and 'not-schools' in the data set we used t-SNE to group the RBG pixels of the datasets, as shown in Figure 2. t-SNE is one helpful way to start to understand the geographic diversity of a dataset to show schools that occur in heavily forested regions vs. schools that occur in desert regions, and the spectrum in between.

Figure 2: Kazakhstan Schools t-SNE

.

The first round of models detailed in the results section were generated with zoom 18 tiles[5], current school models under development are exploring using a zoom 17 super-tile, which is (512, 512, 3) instead of (256, 256, 3), so it covers the same area as a zoom 17 tile, but without sacrificing the resolution of a zoom 18 tiles. We anticipate super-tiles yielding and improvement in performance because sometimes schools got split between multiple zoom 18 tiles.

Methodology

At a high level our methods involved 5 key steps 3:

- Create map of known schools
- Generate a training dataset
- Iterate and tune a school CNN model
- Run our model over all of Colombia and the Eastern Caribbean to identify candidate unmapped schools
- Validate these predictions with Development Seed expert mappers

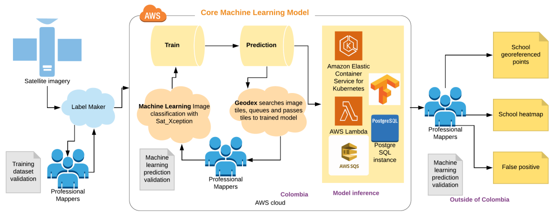

Figure 3: School detection high level overview

.

Selecting a machine learning framework

We did initial testing on two promising ML frameworks: Xception and MobileNetV2(Chollet 2017), two pre-trained models built in our Sat-Xception. To facilitate hyperparameter search and efficient model training we use the Kubeflow tool Katib. Xception is one of current state-of-the-art CNN architectures and pre-trained models on top of ImageNet (Russakovsky et al. 2015). It's a high performing and efficient network compared to other pre-trained networks. MobileNetV2[14] on the other hand, is a model that is slightly less accurate compared to Xception. However, it's very light-weight, fast, and easy to tune when limited resources are available. We broke the training sessions into two sessions. The first session was designed to test the feasibility of using Sat-Xception to train a well-performed school classifier in Colombia. The model was over-confident in rural Colombia in the first session, leading to too many false predictions in the area. To overcome the issue, the expert mapper team created a new training dataset that was slightly different from the training dataset in the first session. In the second training session, 2,048 'not-school' buildings were added. In addition, for the "school" category, we only kept rural schools that have very clear school features. We also randomly selected another 2,500 confirmed school tiles to add to the category. 4

We trained about 200 model iterations on two separate AWS EC2 P3.2xlarge(Services 2020). They are AWS's deep learning AMI machines that have deep learning virtual environment setup, e.g. python3 with Tensorflow GPU version pre-installed in our case, and ready-to-use [16]. We found

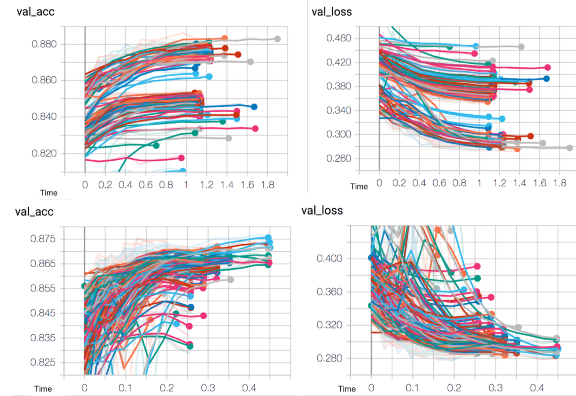

Figure 4: Model performance. First row two figures represent the model trained with Xception and the second row are model validation accuracy and loss that trained with MobileNetV2.

.

the best-performing model from MobileNetV2 with a validation accuracy of 0.88. However, Xception reached a validation accuracy of 0.89, and therefore, we picked the model trained with Xception. We packaged the best-trained Xception model with Tensorflow Serving (Olston et al. 2017). Tensorflow Serving helps to package the Keras (Chollet et al. 2015) and Tensorflow(Abadi et al. 2015) model as a Docker image. The image can serve as an endpoint for large spatial scale model inference, which allows us to run model inference on tens of millions of image tiles per hour.

Large Scale ML Inference

We applied our model to 52 million zoom 18 tiles of DigitalGlobe Vivid Basemap, representing all of Colombia and eleven Eastern Caribbean nations. To run the inference at a country-wide scale on high-resolution imagery, we developed an open source library called chip-n-scale-queue-arranger 5.

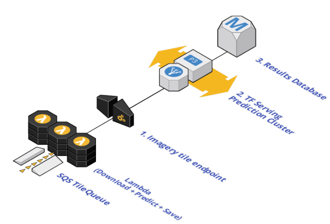

Figure 5: chip-n-scale: to run machine learning models over satellite imagery at scale.

.

Chip-n-Scale is a collection of AWS CloudFormation templates deployed by kes, lambda functions, and utility scripts for monitoring and managing the project. Key steps that chip-n-scale helps facilitate inference at scale include:

- A user sends ‘x/y/z‘ tile indices to an AWS SQS queue to indicate which geographic region to run our model over.
- Each SQS message triggers the Lambda function ‘DownloadAndPredict‘ which downloads images, posts to a prediction cluster (via a Load Balancer endpoint), and saves the result to an RDS database.
- The prediction cluster on ECS runs the TensorFlow Serving image to predict each tile. All instances are behind an Application Load Balancer which will dynamically register new instances that appear on the cluster and allocate jobs to them evenly.
- A user manually downloads predictions from RDS after the full inference process is complete.

The predictions were converted to a shapefile for expert mappers to validation using a map editor. In the editor, the mappers can overlay the predicted school tiles only and focus their attention on confident predictions, avoiding the tedious task of reviewing the entire area for Colombia and the eastern Caribbean islands.

ML Prediction Validation

In the validation process, expert mappers validated each of predicted schools as either “confirmed”, “unrecognized”, and “not-school”, based on the learned school features from the cleaned school dataset. With the increase of the threshold, e.g. from 0.44 to 0.99, we would limit the false prediction but we will also lose an increasing proportion of correct prediction. When the threshold is set to 0.92, we have 73,717 tiles that are predicted school tiles for our expert mappers to go through. With the validation speed of 10,000 tiles per day, we were able to complete the predicted school validation within eight working days.

Total Tiles Reviewed	Confirmed
Confirmed ML School Tiles	73,717
Confirmed and Added School Geolocations	12,250
ML School Unrecognized	60,568
ML School No	899

Table 1: DevSeed Data Team validated 73,000 predicted school tiles from machine learning. We added about 11,000 schools to the map in Colombia and the Caribbean islands, and around 7,000 of them were unmapped schools.

Results and Discussion

Our machine learning model took individual tiles as input and provided an output in the form of a probability between zero and one for each. By running the model prediction over the test set (about 2000 tiles) we plotted a ROC curve 6. The ROC curve gives us an overall model performance and guidance on threshold cutoff.

The ROC curve indicates the area under the curve is up to 0.94 from the test set with the given threshold of 0.44 -

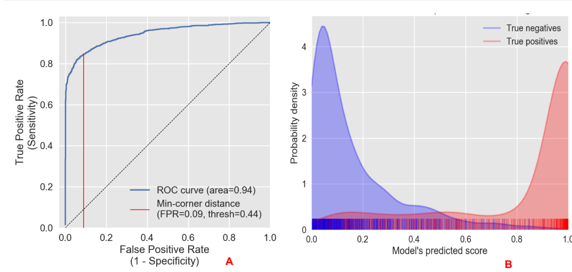

Figure 6: The ROC curve and false positive rate for the school classifier in Colombia.

.

we got a false positive rate of 9 percent. The ROC curve (A) shows that the model we selected was a high-performed model, where only 9 percent of detected schools are expected to be false positive.

The ROC curve and false positive rate for the school classifier in Colombia. The ROC curve indicates the area under the curve is up to 0.94 from the test set with the given threshed of 0.44 In initial testing, we found that the Xception model was slightly more accurate than MobileNetV2. However we also noticed that MobileNetV2 was much faster for each model training iteration. Specifically, MobileNetV2 only used a quarter of time per training iteration on exactly the same training set. During the validation process, the expert mappers validate each predicted school tile and tag it as “yes”, “unrecognized” and “no” based on the school features part of the selection criteria defined during the initial data cleaning process. With the increase of the threshold (e.g. from 0.44 to 0.99) we would limit the false predictions but in the process, we will also lose an increasing proportion of correct predictions.

With a threshold score of 0.92 our model predicted that 73,717 tiles across Colombia and the eastern Caribbean islands contained schools. While this set certainly contained false positives and unverifiable tiles, it also significantly reduced the search space for schools in these countries to less than 0.15 percent of 52 million tiles. This shifted what was previously a nearly impossible task to one that could be done by 5 expert mappers in eight days.

With a validation speed of 10,000 tiles per day, the mappers identified 10,998 school geolocations, where 6,954 of them are unmapped schools (schools that were not part of the initial dataset of 44,665 schools). 60,568 predicted school tiles were tagged as “unrecognized” by our expert mappers 1. These tiles do not have clear school features. During the machine learning prediction validation, we found that schools in rural areas are hard to verify as schools because residential houses may be used as school locations.

Machine learning model generalizability is an active research area (1, 2), and in our study, the school classifier we trained in Colombia generalized well in the Eastern Caribbean islands. We added 262 schools to the islands that had not been mapped before. Our school classifier, the Tensorflow Serving image(GPU version), lives on DockerHub now. It is open-source and free to run as an end-point to users

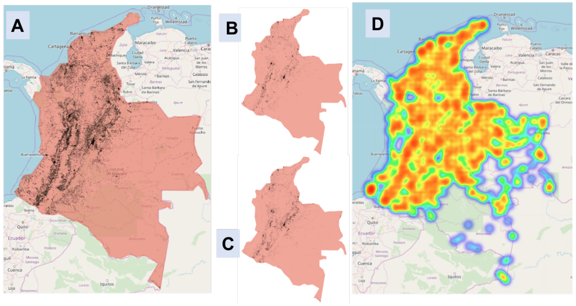

Figure 7: Schools are represented as black dots in figures A - C. UNICEF provided 44,655 school locations (A). Of those 10,951 schools were confirmed as schools by expert mappers (B.). 70% of these verified schools were used to train the machine learning model. After the validation of our model predictions, 10,988 schools were added to the map (C.), around 7000 of these are previously unmapped schools that we recently added after the ML validation. The school heatmap was created from the machine learning predicted and validated as "unrecognized" by our expert mappers (D). The heatmap is an interactive map that the field agents can use to prioritize ground validation of school locations.

.

who want to send zoom 18 images tiles to classify schools in their area of interest.

Conclusion

The school classifier explored in this work, including model training and model inference, can be fully automated and scaled up to much larger geographic areas in the future. These results suggest considerable potential for mapping schools at a scale, quickly with human mappers' in the loop for validation. This will support the improvement of education information management systems, reduce gaps in access to information and opportunity, improve the quality of education, and further disaster response to vulnerable populations. Thereby aid the achievement of the relevant UN Sustainable Development Goals of equal access to quality education.

References

Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; and Zheng, X. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. URL https://www.tensorflow.org/. Software available from tensorflow.org.

Chollet, F. 2017. Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv:1610.02357 [cs]* URL http://arxiv.org/abs/1610.02357. ArXiv: 1610.02357.

Chollet, F.; et al. 2015. Keras. https://keras.io.

DigialGlobe. 2020. DigitalGlobe | Imagery mosaics support best foundation. URL http://www.digitalglobe.com/products/mosaics.

Olston, C.; Fiedel, N.; Gorovoy, K.; Harmsen, J.; Lao, L.; Li, F.; Rajashekhar, V.; Ramesh, S.; and Soyke, J. 2017. TensorFlow-Serving: Flexible, High-Performance ML Serving.

OpenStreetMap. 2020. OpenStreetMap. URL https://www.openstreetmap.org/.

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; and Fei-Fei, L. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115(3): 211–252. doi:10.1007/s11263-015-0816-y.

Services, A. W. 2020. Amazon EC2 P3 – Ideal for Machine Learning and HPC - AWS. URL https://aws.amazon.com/ec2/instance-types/p3/.

UN. 2020. Sustainable Development Goals .:. Sustainable Development Knowledge Platform. URL https://sustainabledevelopment.un.org/?menu=1300.

UNICEF. 2020. School Mapping. URL https://www.unicef.org/innovation/school-mapping.