‘The Road Not Taken’ - Using satellite imagery for road detection in rural areas

Portfolio project at Data Science Retreat, Berlin (www.datascienceretreat.com)
Batch #15, June 16 - September 13, 2018

Lisa Heße
Kiran Prakash
Harald Hentschke

The following slides are a detailed project summary for both technical and nontechnical audiences, based on our presentation slides on Community Day in Berlin (Sep 11, 2018).

Introduction

Background

Data Science Retreat in Berlin is a three-month intensive training, geared to anyone with programming skills and a strong will to put Data Science into practice. Each participant (or group of participants) is required to work on a project of reasonable complexity and novelty, to be presented at the end of the training.

The trigger for this project was a letter by W.F. Laurance in the scientific journal nature in June 2018, instigating the AI community to come up with an automated method of detecting roads in satellite imagery of rural areas (see excerpt). The emphasis here is on rural areas; published material on road detection in urban areas abounds.

We took on this challenge by devising a machine learning approach, with data being kindly provided by the group of Prof. Laurance (detailed description in the following). Data are from Southwestern Borneo, hence the focus on this region.

Correspondence

Arctic collaboration transcends tensions

Of all world regions, the Arctic is the most sensitive to climate change and drives feedbacks that amplify the effects of global warming around the planet. Understanding the Arctic relies on developing a better knowledge of the hugely expansive Russian Arctic regions, which offer unique opportunities to study landscape systems across large latitudinal gradients, linked by major river networks.

However, these regions have been something of a blind spot for the international community of Arctic scientists. This is due to access difficulties and to research findings going unrecognized because of language barriers. Happily, at a time of mounting political tension between Russia and the progress over the decades to come.

Sammie Buzzard *University College London, UK.*
Joseph Cook *University of Sheffield, UK.*
Alexey Maslakov *Lomonosov Moscow State University, Russia*
s.buzzard@ucl.ac.uk

Road mapping needs AI experts

As road building expands globally, an automated system for detecting and mapping roads in near-real time is urgently needed to plan land use and conservation management. Machine-learning or artificial-intelligence (AI) specialists must help to meet this formidable challenge.

Current road data are grossly inadequate (see W. F. Laurance *et al. Nature* **513**, 229–232; 2014), of tackling this huge problem effectively.

William F. Laurance *James Cook University, Cairns, Australia.*
bill.laurance@jcu.edu.au

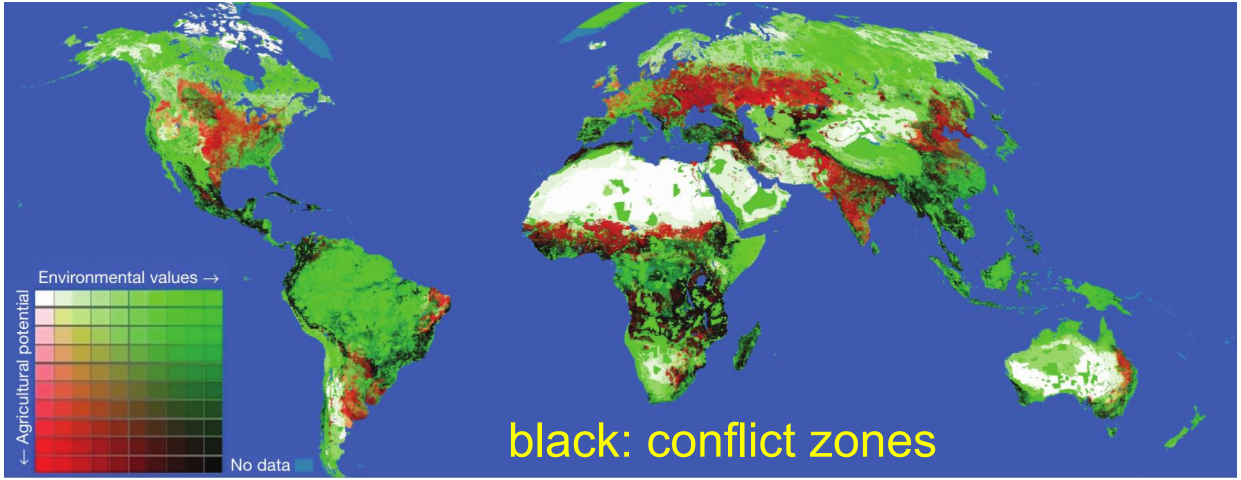
Circular economy creates new jobs

Governments are anticipating that people will be displaced from factory and service jobs as intelligent systems are increasingly deployed. Smart environmental enterprises could offer a more sustainable approach than solutions such as universal pay, and provide employment.

In a circular economy (see www.nature.com/thecirculareconomy), commercial enterprises that reverse environmental damage, for example, are needed to deliver value in a new guise.

nature International journal of science 558:30, June 2018

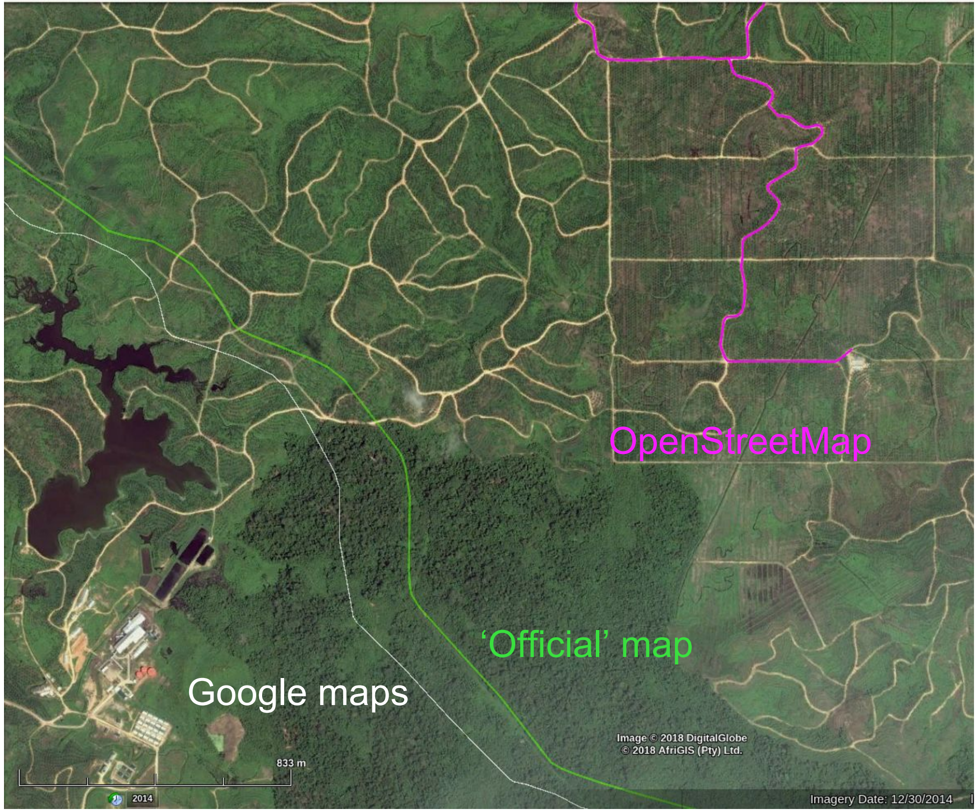
Why do we care about road detection in rural areas?



Laurance, W.F. et al. (2014). A global strategy for road building. Nature *513*, 229–232.

In the briefest of terms, road construction is bound to increase on a massive scale globally (2010 - 2050: anticipated expansion of the global road network by 60%). A substantial part of these activities will occur in environmentally sensitive areas and in an unplanned manner, with according negative repercussions on the environment and local communities. In order to discover and regulate these activities, efficient ways of detecting roads in satellite imagery are needed.

A glimpse at rural road maps



For a graphical introduction to the problem, the image on the left shows a patch of land in Southwestern Borneo, viewed with Google Earth. Road labels from various sources are overlaid as colored lines. It is plain to see that none of the sources of road maps are adequate, although it has to be stated that in our experience OpenStreetMaps covers not-too-remote regions of Southwestern Borneo relatively well. Nonetheless, none of the maps is nearly complete, and recently built roads are certain not be covered.

Definition of the task. i) Principal approach

- Pixel-level classification task
  - Binary: [no road, road]
  - Ternary: [no road, unpaved road, paved road]

Our machine learning approaches are of the ‘supervised’ category: we employed neuronal networks that are able to learn which pixels of any given satellite image belong to a road, and which ones belong to anything else. A road vs. no road dichotomy (=binary classification approach) was pursued first because it was more likely to lead to success. Subsequently, we also trained a network to discriminate between no road, paved roads and unpaved roads (ternary classification).

Definition of the task. ii) Regions and imagery

- Regions: Borneo and Harz (~500 km² each)

We trained the networks on two quite different rural regions, one in Southwestern Borneo (‘3093’), the other in the Harz (Germany, ca. 300 km southwest of Berlin). The rationale was that i) for the latter region we were more certain to obtain road labels with the required precision, and ii) the different appearances of the landscapes in both regions would further the networks’ capability of abstraction. We did train networks on both regions separately, but all results shown here are from models trained on both areas.

- Satellite imagery (Planet.com)
  - 4-band (RGB & infrared)
  - 3 m/pixel

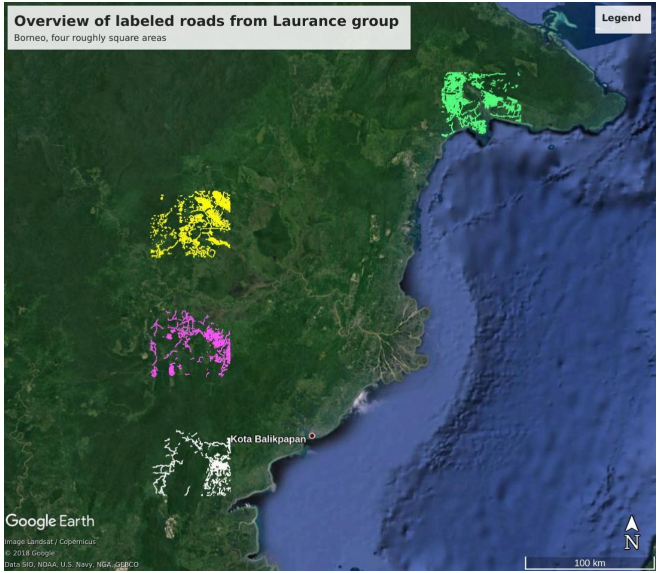
We opted for ‘Planetscope’ satellite imagery, kindly provided by Planet.com (Planet team, 2017). The resolution of the images at about ⅓ pixels/m (with matching optical resolution) seemed just about right for our purpose. This is not to say that freely available imagery from e.g. the EU’s Copernicus program (best resolution, depending on satellite/image/product, at ⅕ pixels/m) wouldn’t be suitable, but we wanted to be on the safe side.

Planet Team (2017). Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https://api.planet.com

Definition of the task. iii) Labels

Correct **labels** in the 'training' data set are critical for success with any supervised approach. In our case, we ultimately needed images of the same size as the satellite imagery in which e.g. no road-pixels are zero (would appear as black in an image viewer) and road pixels have any other value (e.g. 255, so would appear white). Road labels, however, usually come in the form of concatenations of linear segments, as visualized in the image on the left, showing an unpaved road in the process of being manually labeled in Google Earth. Tapping the vast arsenal of Python geospatial libraries we converted these vectorized road labels into label images, assuming a fixed road width (examples shown in the results slides further below). This is a simplistic approach which can certainly be improved, but it worked.
It needs to be stressed that for a machine-learning based approach to work, labels must be as precise as possible (e.g. no cutting of corners, no mis-labeling of non-road structures) and also complete (no unlabeled roads should exist).

Definition of the task. iii) Labels (continued)

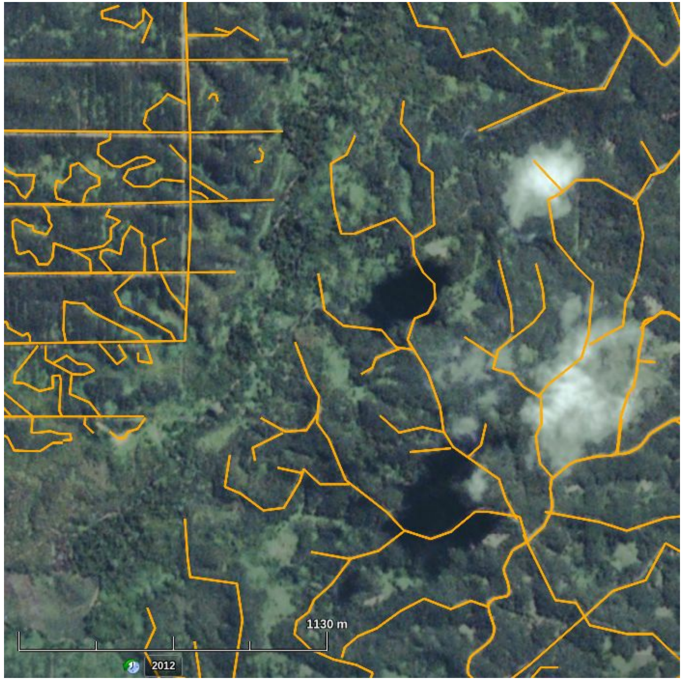


We were kindly provided with labels of roads for four regions in Southwestern Borneo by the Laurance lab/GlobalRoadMap initiative (www.global-roadmap.org), see overview image to the left. These are road labels generated by volunteers in painstaking manual work with Google Earth (GE). As this initiative's focus is on rural roads, defined as those outside of human settlements, it was clear that additional labels would be required. Apart from that, we found some issues with the labels with regard to their usage for our purposes (which is not exactly surprising as they were never intended for that purpose):

- Labels were often too coarse
- In a few rare instances, rivers were traced (which, admittedly, can look deceivingly similar to roads in regular RGB imagery)
- Labels were partly incomplete, even in remote areas

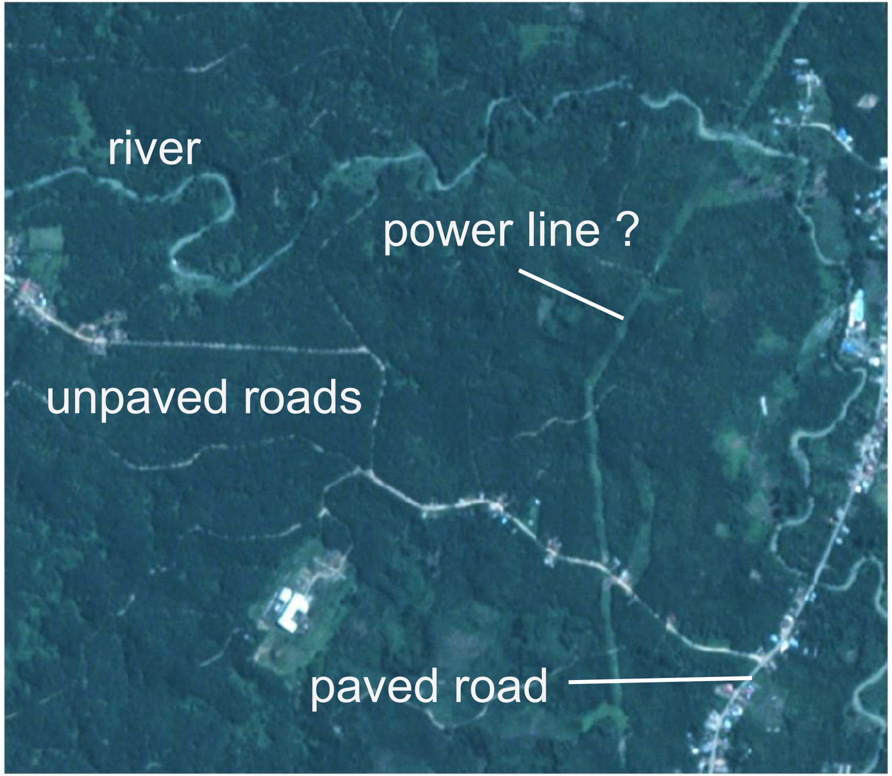
The last point is likely due to the nature of road construction in rural areas and the workings of GE, as we experienced ourselves. Road construction is an ongoing process, so road labeling had best be based on satellite imagery from a limited time period (say, 6 months), in order to obtain a snapshot of the situation. In GE, however, the user is presented with a mish-mash of imagery from various sources taken at different dates. There is no full control of the date/recency of the satellite imagery displayed, and in some instances, depending on the zoom level, the exact date is not even displayed.

Definition of the task. iii) Labels (continued)



Hence, we decided to combine labels from different sources and also create some labels on our own, avoiding redundancies as far as possible. Prior to combination, provided labels were rectified to some degree in GE. Rectification and combination was based on cloud-free (as far as possible) satellite imagery from a defined, limited time span (2017-08-01 to 2018-08-14) which was loaded into GE and displayed on top of the default GE imagery. Narrower time spans proved not to work due to the high incidence of clouds over Borneo; see https://medium.com/planet-stories/one-of-the-worlds-largest-reservoirs-lies-hidden-in-the-mountainous-center-of-borneo-1300215f26bf
. The sources of labels were i) Laurance lab, ii) OpenStreetMaps (which conveniently covered all roads in more ‘urban’ areas).
The resulting labels were still not perfect, but represented a working compromise between quality and quantity (see example image, left).

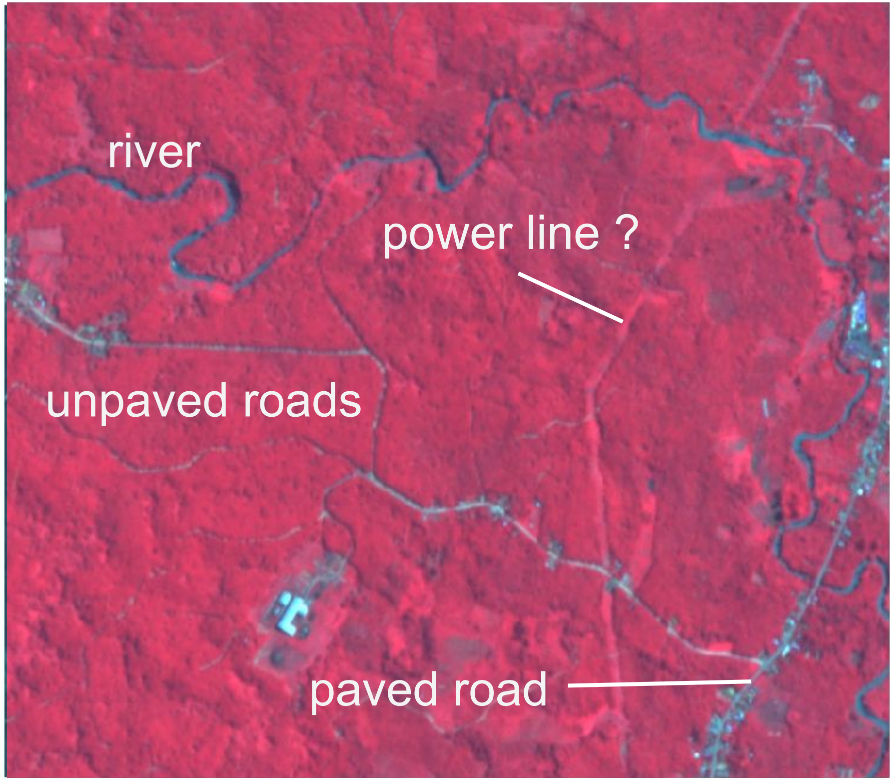
The task: challenges



For a human observer, some of the challenges of road detection in satellite imagery are plain to see, as in the excerpt of a satellite image from rural Southwestern Borneo depicted on the left:

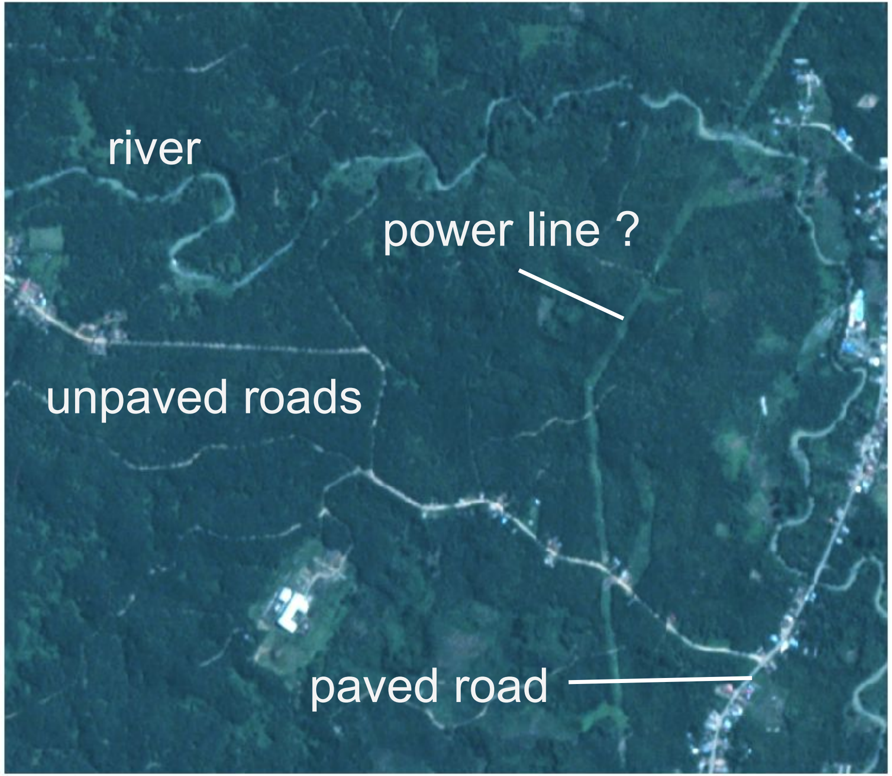
- Paved and unpaved roads look similar
- Partial occlusion of roads by trees, vehicles (and clouds, in case we haven't mentioned that yet...)
- Rivers and other road-like structures exist

The task: challenges (continued)



Rivers, luckily, stand out from roads in the infrared (IR) band, as illustrated in the false-color version of the previous image, which depicts intensity in the IR band in red.

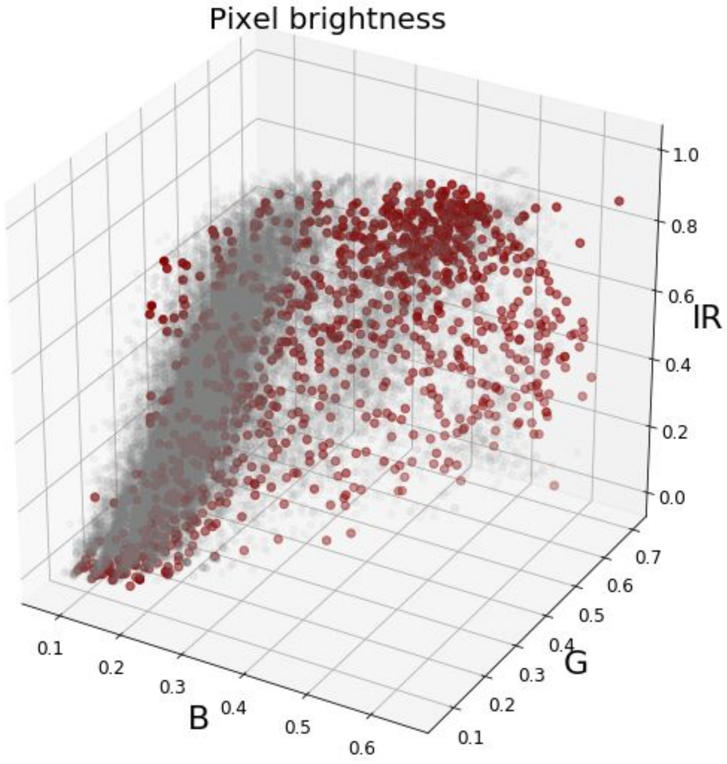
The task: challenges (continued)



From a machine learning point of view, among the challenges we faced are the following:

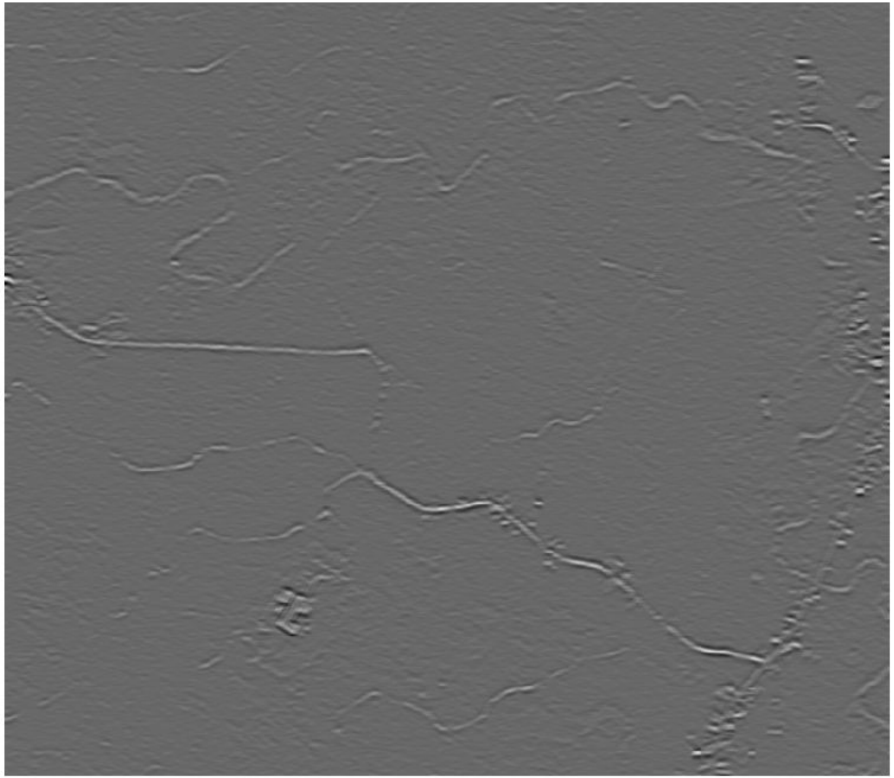
- Imbalance of classes (< 1 % of pixels belong to roads)
- Different hues of images
- Imperfection of road labels mentioned above

The task: challenges (continued)



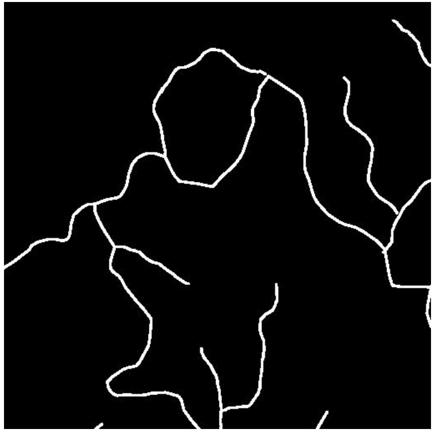
In a nutshell, in pixel-level classification tasks (aka image segmentation tasks) it is less the pixel intensity values that count, but rather the **context**. The first point is illustrated in the image to the left, showing the intensity values of an arbitrary selection of image pixels in the B(lue) - G(reen) IR (infrared) space (‘no road’- pixels in translucent gray, ‘road’- pixels in dark red). Road pixels do show some structure, but they do not form a cluster that could be easily (or at all) separated from the no road pixels.

The job: challenges (continued)



When image context matters, convolutional networks are among the first choice, and this is what we opted for. Such networks perform 2D convolutions. To illustrate the concept of 2D convolutions, we present the same excerpt of a satellite image as above, convoluted with a small patch of 7 by 7 pixels featuring a central horizontal bar. (The resulting image was also converted to grayscale). In the resulting image, much fine-grained structural detail is gone, and instead linear structures with a more or less horizontal orientation pop out. Regardless of the exact details of what is done in the actual networks we used, here is the essence of convolutional networks: abstraction of spatial features.

Data Preprocessing



Overview of data preprocessing steps:

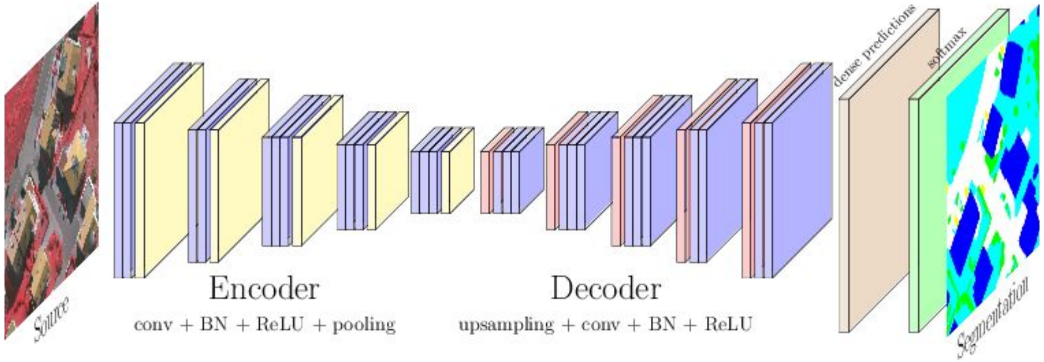
- excision of 512 by 512 pixel image tiles from satellite imagery (left image)
- creation of road label maps of the same size, aligned with satellite image tiles (right image)
- Image Augmentation:
  - 25 % overlap of edges in tiles
  - Rotations and horizontal & vertical flips
- Fully automated
  - Repeatable data pipeline
  - Infrastructure provisioning

HashiCorp Terraform   Google Cloud   HashiCorp Packer

Training

We trained our networks on 60% of our image tiles, set aside 20% of the tiles for validation, and a further 20% for testing; only results from the latter set are reported here. The rationale for this standard procedure is that a meaningful assessment of the performance of the networks can only be obtained from a set of data which the networks have never ‘seen’.

Segnet

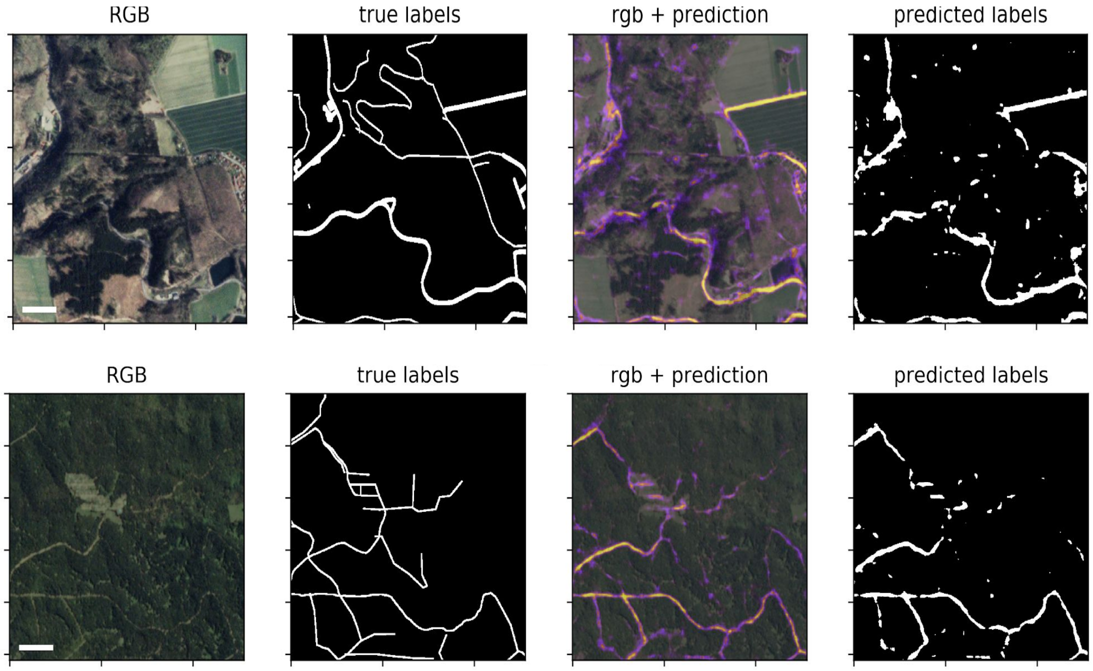


Segnet is a comparatively simple convolutional network which we used as a ‘baseline’ approach (a simple approach that provides a lower bound of what is feasible).

- 3 Convolution layers each for encoding - decoding
- Pixel wise classification using sigmoid/softmax
- Output shape is same as Input
- Batch Normalization
- Cross entropy Loss
- ~5 Million parameters

Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, Senior Member, IEEE

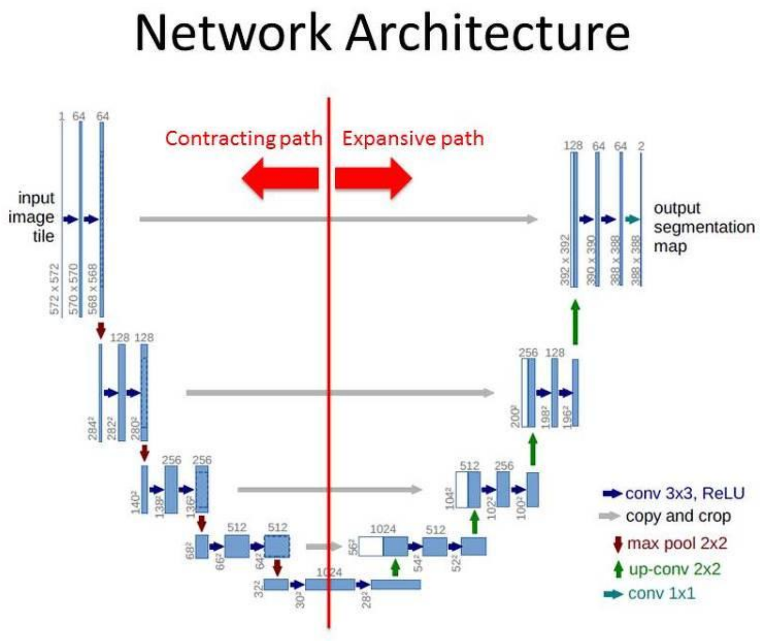
Segnet - Results (binary classification)



As expected, the simple Segnet detects roads alright in not too complicated scenes, but overall lacks in performance.

Left to right: RGB image tile, corresponding labels, prediction score of trained network (brighter colors=higher scores), predicted labels based on thresholded prediction scores. Top row, exemplary scene from Harz; bottom row, exemplary scene from Borneo

Main approach: **U-Net**



- end-to-end encoder-decoder
- architecture build upon FCN:

↘ downsampling: **semantic** information

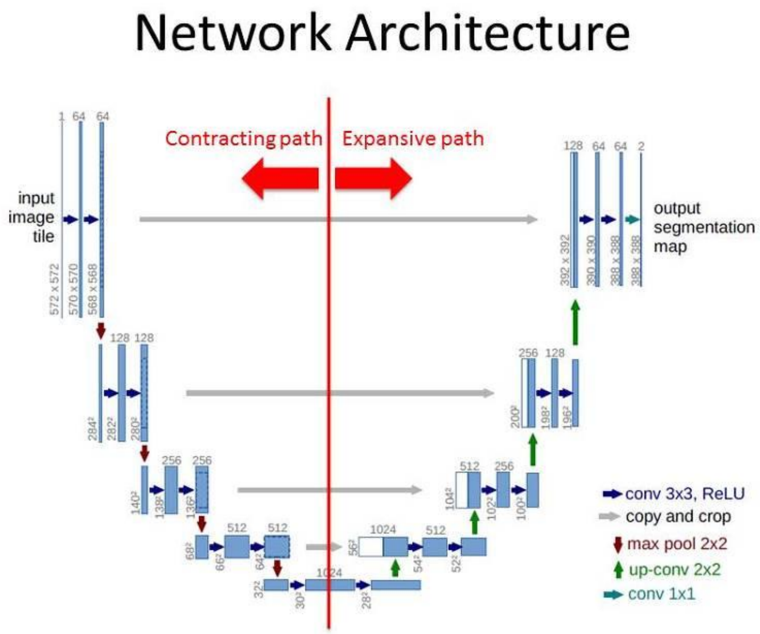
➲ bottleneck: max. pooling

↖ upsampling: **spatial** information

⤇ **bypass**: concatenate ↘ with ↖
+ better localization
+ learn representation

Ronneberger *et al.*, MICCAI, Springer, LNCS, **9351**, 234 - 241 (2015)

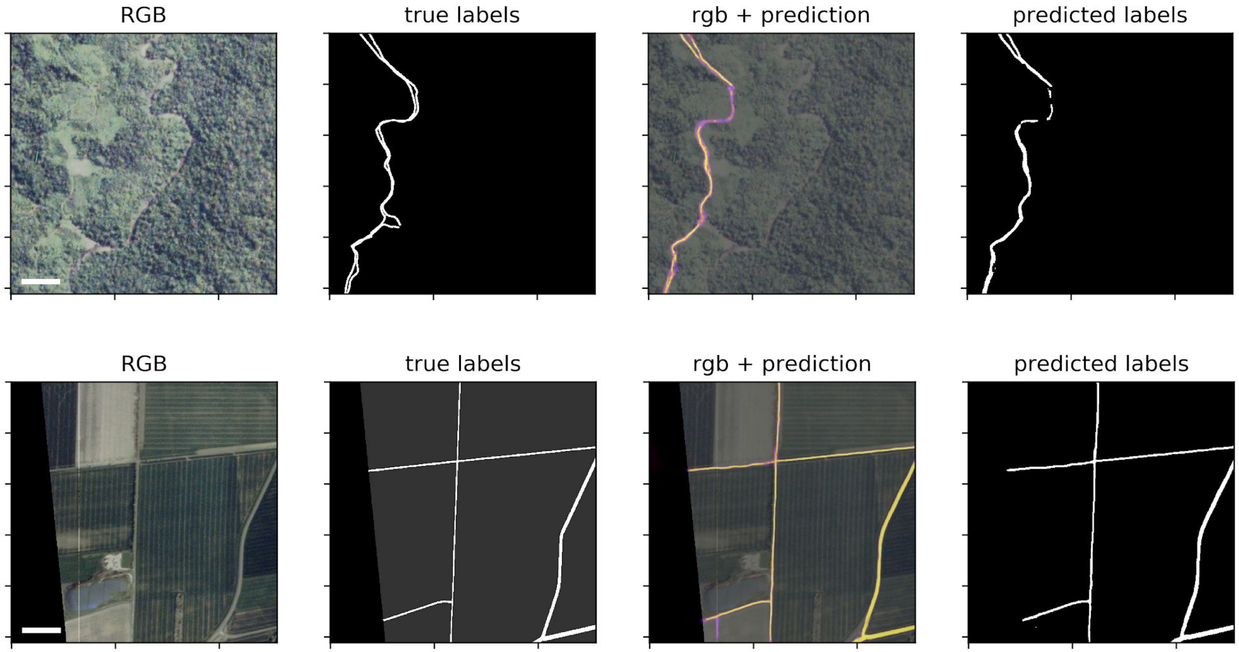
Main approach: **U-Net**



- very powerful architecture
- even for small sample sizes
- fast learner

Ronneberger *et al.*, MICCAI, Springer, LNCS, **9351**, 234 - 241 (2015)

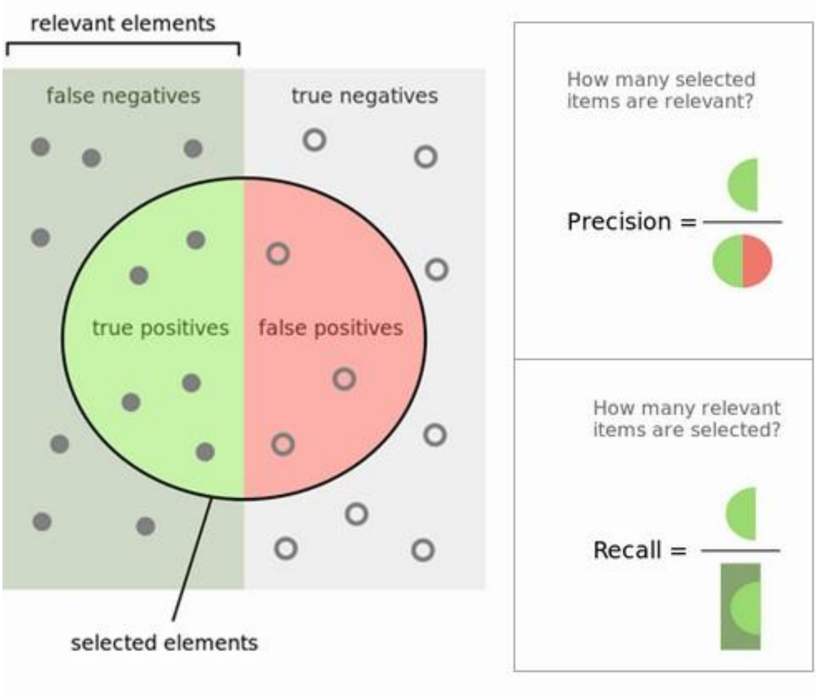
U-Net Binary Classification - Results



Left to right: RGB image tile, corresponding labels, prediction score of trained network (brighter colors=higher scores), predicted labels based on thresholded prediction scores. Top row, exemplary scene from Borneo; bottom row, exemplary scene from Harz.

The U-Net performs much better than the simple Segnet.

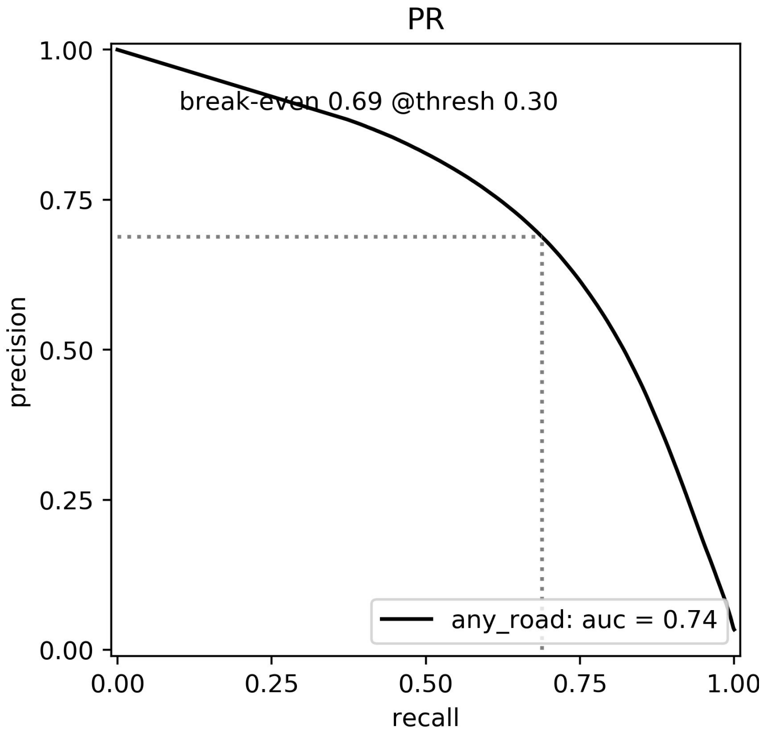
Metrics



In order to evaluate the performance of our models in an objective way we computed metrics based on precision and recall (see illustration on left). These are common metrics in classification tasks with class imbalances: as mentioned before, in our case, the true negatives (pixels not belonging to roads) by far outnumber the true positives.

Precision and recall are computed by setting a threshold value for the prediction score (say, 0.5) and and classifying all pixels (in all images of the test set) with prediction scores equal to or above that value as positives (road pixels), and all other pixels as belonging to the no road class. By comparing the true labels of each pixel, we arrive at values of precision and recall.

U-Net Binary Classification - Results

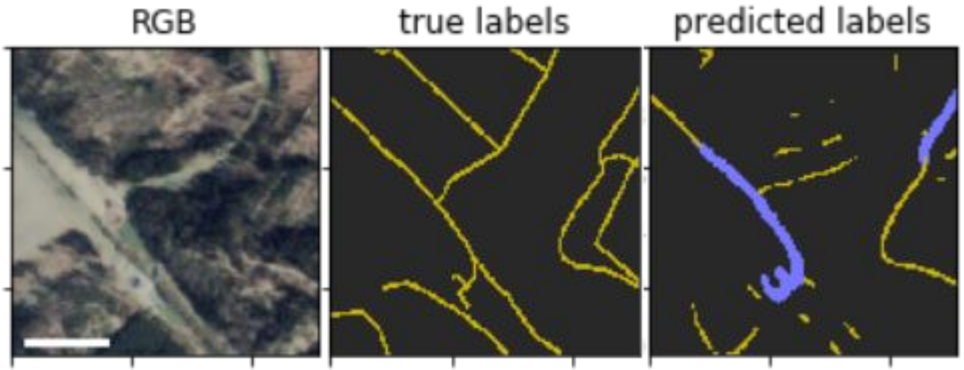


Values of the threshold are systematically varied from zero to the maximal value of the prediction score; thus, we obtain series of precision and recall values. Precision is then plotted versus recall, which is shown in the figure to the left for the U-Net with the binary classification task. The higher the precision and the higher the recall, the better (so, the curve should bend as far to the upper right corner as possible). The integral under the curve (auc) is usually used as a scalar summary metric of the discriminability of the model; the closer the value to unity, the better. 'Break-even' is the point at which precision == recall; the value of the threshold at this point was used for generating prediction labels from prediction scores as shown in the previous figures.

Loss-function modelling:

- **noisy** (improper) labels $\tilde{\boldsymbol{y}}$

$$p(\boldsymbol{y}|\text{image}) = P(\boldsymbol{y}|\tilde{\boldsymbol{y}})\,p(\tilde{\boldsymbol{y}}|\text{image})$$

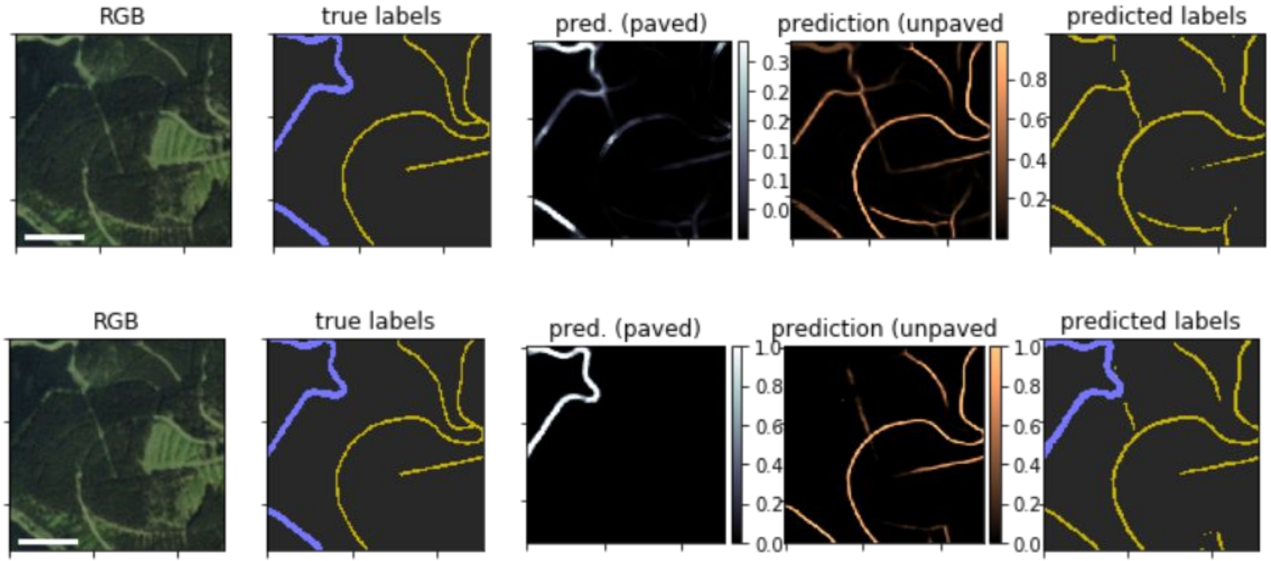


- less penalty if *wrong* $\hat{\boldsymbol{y}}$ but confident
- increase learning of multiclass labels

Next, we pursued ternary classification. When trained with a standard ‘loss function’ (a metric used internally by the networks for learning) the networks tended to mistake paved roads for unpaved roads. One approach to overcome this problem (as well as mislabeling in general) is to assume that the labels are ‘noisy’ and adjust the loss functions accordingly. This is what we have pursued next.

V. Mnih, G. Hinton, ICML-12, 567 - 572 (2012)

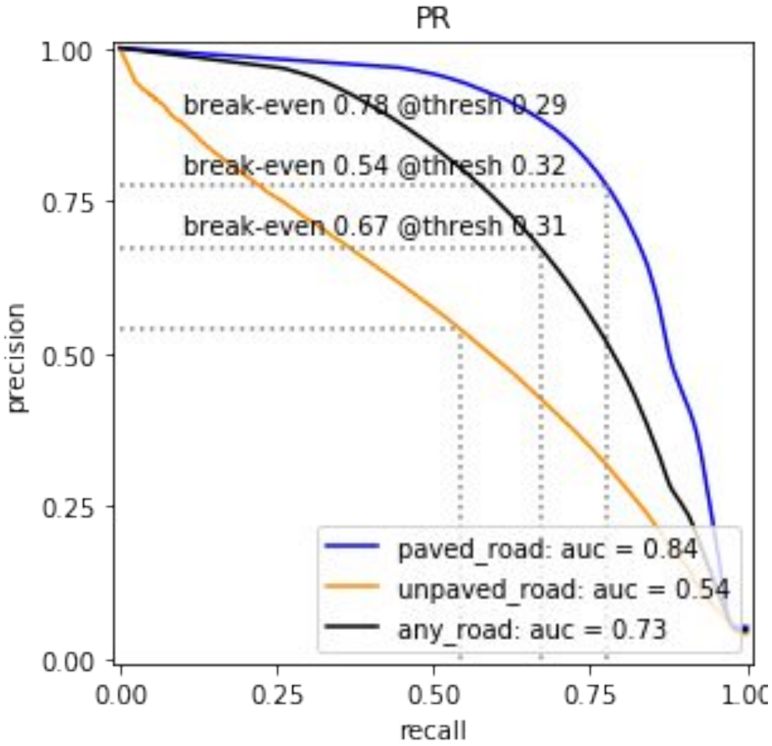
U-Net Ternary Classification - Results



Inspection of individual tiles revealed that the ‘noisy label’ approach more successfully detected paved roads.

Left to right: RGB image tile, corresponding labels (blue=paved roads, orange=unpaved roads), prediction score for paved roads,prediction score for unpaved roads, predicted labels based on thresholded prediction scores. Top row, results with regular categorical crossentropy; bottom row, results with noisy label approach)

U-Net Ternary Classification - Results



With the given ‘plausible guesses’ of parameters we needed to specify for the noisy label approach, paved roads were detected quite successfully, but at the expense of discriminability of unpaved roads. When the results were downscaled to the binary case by combining paved and unpaved roads into a ‘any road’ class (black curve), performance of the model was very close to the natively binary model shown above.

Takeaways

- **Road detection in rural areas works**
- Image segmentation: context counts
- U-net delivers good results with only a few samples (~ dozens)
- Ternary classification is more complex and requires parameter fine-tuning
- Room for improvement: quality and quantity of labels

  … in the meanwhile: loss function modulation

Thanks

Srinivas Sridharan
Greenlake Research, Berlin
Mentor

William Laurance & coworkers
James Cook University
Inspiration & data

Planet.com
Free satellite imagery

Pierre Ibisch & Monika Hoffmann
TU Eberswalde
Pointers to data

Chris Armbruster, José Quesada, Batch 15

Good times @