Satellite Image Feature Detection

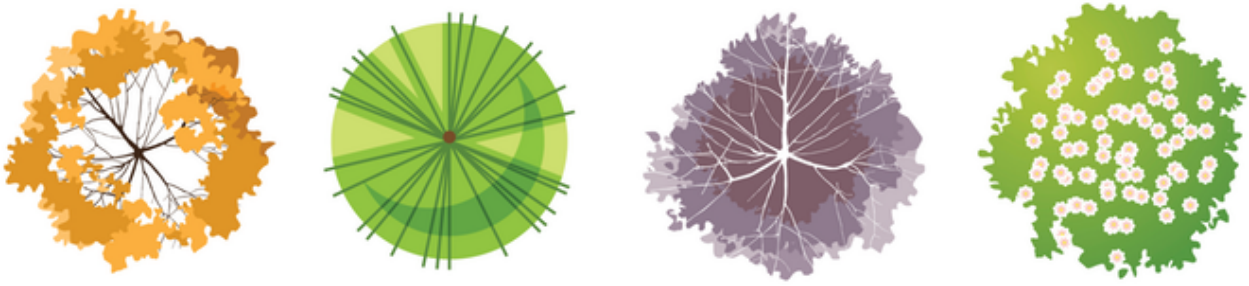
Author: Aakash Chaudhary

Table of Contents

Abstract 2
Problem Statement 2
Objective 2
Feature Classification 2
Type of Image Features 3
Edges *3*
Corners *3*
Data Description 4
Object types *4*
Geo Coordinates *5*
Strategy 5
Statistic 7
Model 8
A U-net model *8*
Training Augmentation 9
Testing Augmentation 10
Prediction 11
Future Development 13
References 13

Abstract

The rapid increase of satellite imagery has given us a drastically improved understanding of our planet. It has empowered us to all the more likely accomplish everything from mobilizing resources during disasters to monitoring effects of global warming. What is frequently underestimated is that advancements such as these have relied on labeling features of significance like building footprints and roadways fully by hand or through imperfect semi-automated methods.



Problem Statement

As these huge, complex datasets keep on expanding exponentially in number, the Defense Science and Technology Laboratory (Dstl) is looking for novel answers for reduce the weight on their picture investigators.

Objective

The goal of this paper, which I believe I have come to, was to build up an Algorithm to accurately classify features in overhead imagery that is quick, vigorous, sensible simple and efficient with a moderately straightforward and straightforward algorithms and methods.

Feature Classification

There is no all-inclusive or accurate meaning of what comprises a feature, and the specific definition frequently relies upon the issue or the sort of use. By and by, A

feature is commonly characterized as an "intriguing" some portion of an Image, and feature are utilized as a beginning stage for many computer vision Algorithm. Since feature are utilized as the beginning stage and fundamental natives for ensuing Algorithm, the general Algorithm will regularly just be in the same class as its feature detector. Subsequently, the attractive characteristic for a feature detector is repeatability: regardless of whether a similar feature will be identified in at least two distinct pictures of a similar scene.
There are many computer vision algorithms that use feature detection as the beginning stage, so as a result, a very large number of feature detectors have been developed. These vary widely in the kinds of feature detected, the computational complexity and the repeatability.

Type of Image Features

Edges

Edges are focuses where there is a boundary (or an edge) between two picture districts. All in all, an edge can be of practically self-assertive shape, and may incorporate intersections. By and by, edges are normally characterized as sets of focuses in the picture which have a strong gradient magnitude. Besides, some common Algorithm will at that point chain high gradient together toward structure a progressively complete depiction of an edge.

Corners

The terms corners and intrigue focuses are utilized fairly conversely and allude to point-like feature in a picture, which have a nearby two-dimensional structure. The name "Corner" emerged since early calculations originally performed edge recognition, and afterward broke down the edges to discover quick alters in course (corners). These calculations were then grown so express edge identification was not, at this point required, for example by searching for significant levels of ebb and flow in the picture inclination. It was then seen that the purported corners were likewise being recognized on parts of the picture which were not corners in the customary sense (for example a little brilliant spot on a dull foundation might be identified). These focuses are as often as possible known as intrigue focuses, however the expression "corner" is utilized by convention

Data Description

In this paper, we have 1km x 1km satellite images in both 3-band and 16-band formats. As we know the goal is to detect and classify the types of objects found in these regions. 3- and 16-bands images, we have two types of imagery spectral content. The 3-band images are the traditional RGB natural color images. The 16-band images contain spectral information by capturing wider wavelength channels. This multi-band imagery is taken from the multispectral (400 – 1040nm) and short-wave infrared (SWIR) (1195-2365nm) range. All images are in Geo Tiff format and require Geo Tiff viewers (such as QGIS) to view.

Object *types*

In a satellite image, we have lots of different objects like roads, buildings, vehicles, farms, trees, water ways, etc. Dstl has labeled 10 different classes:

- Building-large building, residential, non-residential, fuel storage facility, fortified building
- Misc. Manmade structures
- Road
- Track - poor/dirt/cart track, footpath/trail
- Track - poor/dirt/cart track, footpath/trail
- Trees - woodland, hedgerows, groups of trees, standalone trees

- Crops - contour ploughing/cropland, grain (wheat) crops, row (potatoes, turnips) crops
- Waterway
- Standing water
- Vehicle Large - large vehicle (e.g. lorry, truck, bus), logistics vehicle
- Vehicle Small - small vehicle (car, van), motorbike

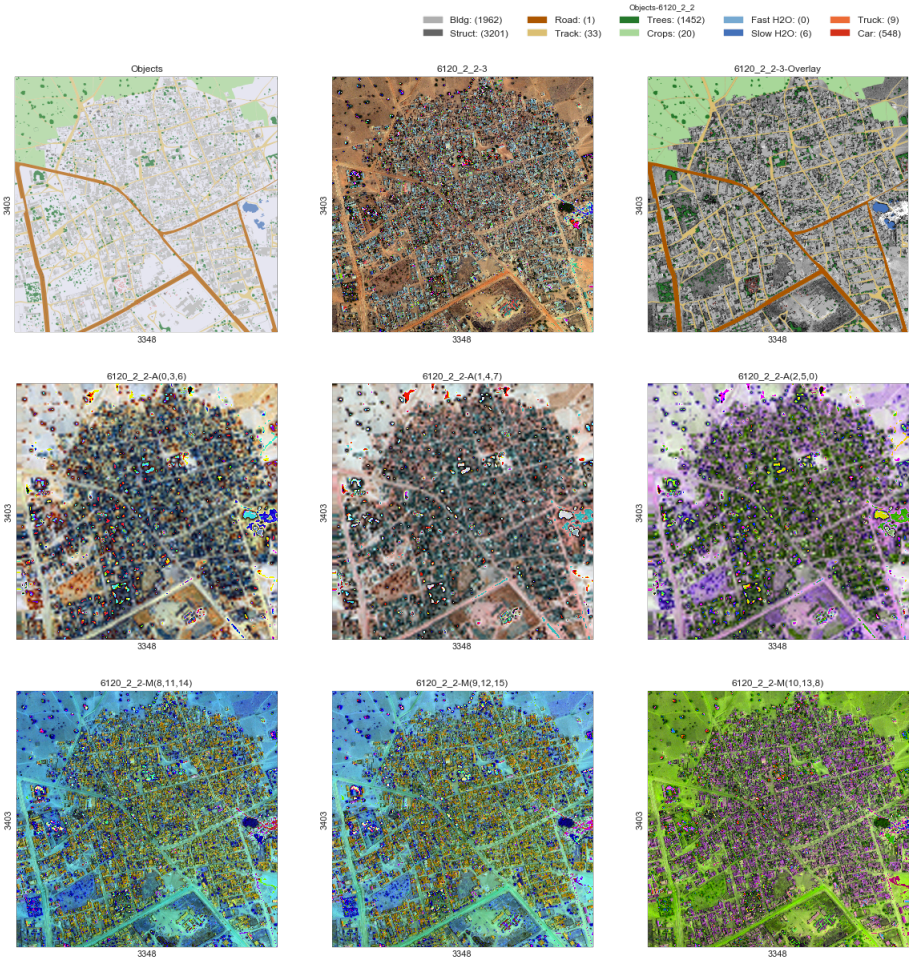
Geo Coordinates

In this dataset, we have a set of geo-coordinates that are in the range of x = [0,1] and y = [-1,0]. These coordinates are transformed such that we obscure the location of where the satellite images are taken from. The images are from the same region on Earth.

To utilize these images, we have the grid coordinates of each image to scale them and align them with the images in pixels

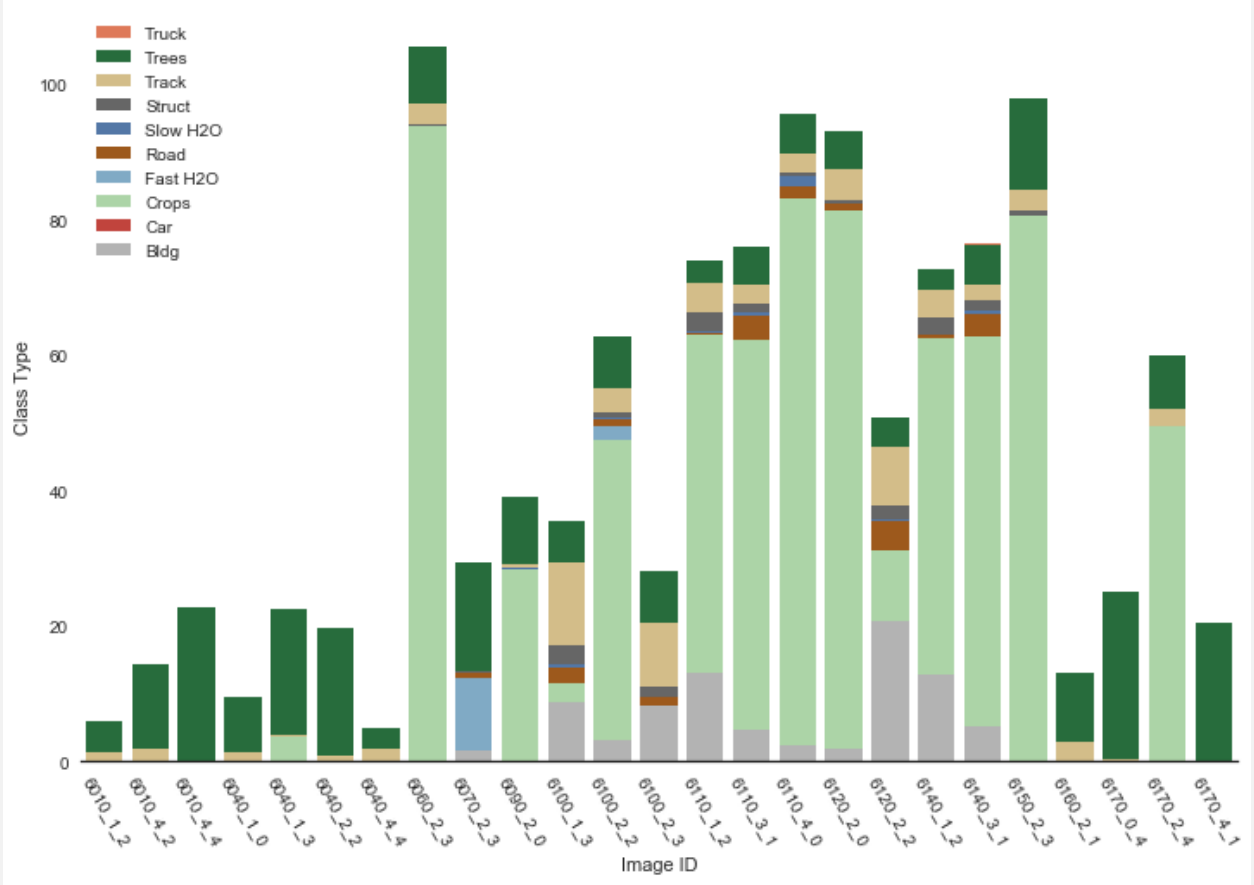
Strategy

Our procedure is to initially change over the forms into mask, and afterward to train a pixel-wise binary classifier for each class of object. The accompanying figures show instances of the 20 channels and the names. Perceptibly, the A band (the center line beneath) has a lot of lower resolution than different groups, and in this way isn't legitimately utilized in model. The M band shows slight spatial move comparative with the 3-band and is interjected to a similar resolution as and further enrolled to the 3-band. We likewise made 4 different indexes, CCCI, NDWI, EVI, SAVI, in light of the given 3, A and M groups. These indexes have been appeared to associate well with specific classifications of object in conventional GIS and can likely make the feature learning simpler. In this way, altogether, we have 16 channels (3 band (3 channels) + M band (8 channels) + P band (1 channel) + 4 incorporated channels) in the information.


Objects-6120_2_2
Bldg: (1962)
Struct: (3201)
Road: (1)
Track: (33)
Trees: (1452)
Crops: (20)
Fast H2O: (0)
Slow H2O: (6)
Truck: (9)
Car: (548)
Objects
6120_2_2-3
6120_2_2-3-Overlay
6120_2_2-A(0,3,6)
6120_2_2-A(1,4,7)
6120_2_2-A(2,5,0)
6120_2_2-M(8,11,14)
6120_2_2-M(9,12,15)
6120_2_2-M(10,13,8)
3403
3348

Statistic

The labels in this Data are on the pixel level, Therefore, each pixel is a training example, and the effective size of this data is 3300*3300*25 = 270 million. Statistics of percentage area of all classes of all training images shows the high imbalance between the true and false labels for most classes, and the heavily imbalanced distribution of true labels among classes. Therefore, I have decided a train a binary classifier for each class, instead of a single model to predict all.



Model

A U-net model

A U-net with bunch standardization is created in TensorFlow and fills in as the arrangement model. The model was prepared for 9000 bunches, each group containing 60 picture patches. Each picture fix is a 144*144 harvest from the first pictures. Like the first U-net paper, the shortfall is just determined on the inside 80*80 district, since the edge pixels just get halfway data.

Loss function: two sorts of loss capacities were looked at, including weighted cross entropy, and delicate Jaccard file, the two of which can represent the valid and bogus names awkwardness. Weighted cross entropy was found to yield wavering execution on the cross-approval information during the preparation. Given the assessment metric is the mean Jaccard file over classes, it is perfect to utilize Jaccard index as the Loss function. Lamentably, Jaccard index isn't differentiable. Delicate Jaccard index rather is differentiable and is near Jaccard file in

exceptionally sure forecasts. A blend of cross entropy (H) and delicate Jaccard index (J), L = H - log(J), is utilized as the loss function.

Optimizer: Adam optimizer with an underlying learning pace of 0.0001 is utilized. In spite of the fact that Adam optimizer should normally perform step size strengthening, we discovered learning rate rot to its 0.1 for each 4000 clusters improved the preparation.

Batch size: the first U-net model uses a solitary picture in each bunch for preparing, which is clearly not suitable for this issue. As can be seen from the class details, the class dispersion fluctuates a great deal across pictures. A huge bunch size (60 here) is attractive, to ensure the measurements of each compelling cluster (assessing the energy) in concurrence with the insights of entire preparing informational index. The group size additionally needs to bargain with the viable preparing district, since the edge pixels must be disposed of. Too huge group size will bring about little community districts for preparing.

Two separate strings are made on CPUs for information preprocessing, and the preprocessed information is driven into two lines and further took care of into preparing and cross approval on GPU individually. This decreases the weight on the GPU and accelerates the preparation. All preparation information is preloaded into RAM to maintain a strategic distance from the moderate record I/O during the preparation.

The task was created via preparing on 21 pictures, utilizing the other 4 pictures for cross approval. Also, the last model was prepared with each of the 25 pictures.

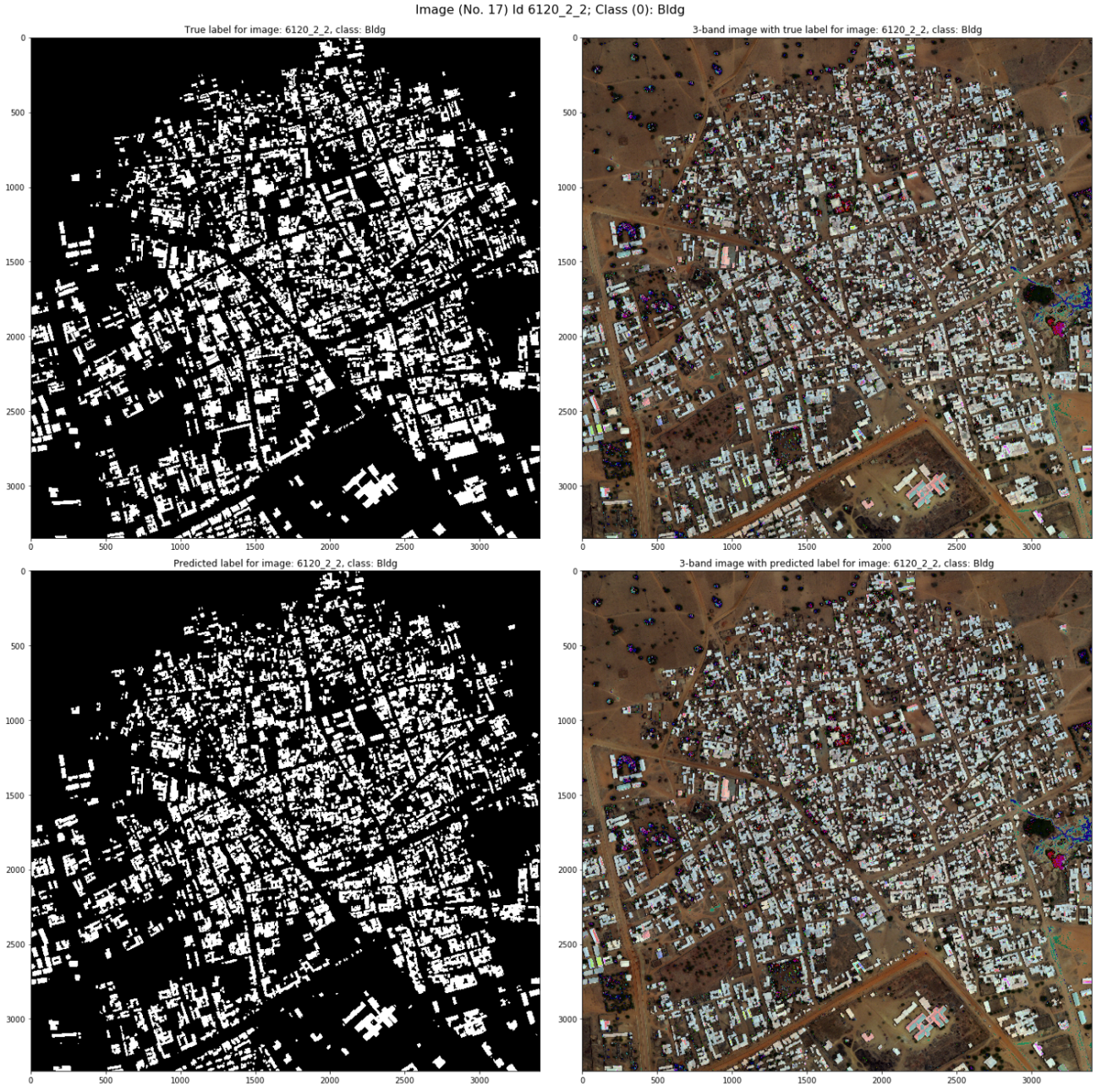
Training Augmentation

Training data augmentation includes random horizontal and vertical reflection and translation, and 360 degrees rotation (1-degree step size). The rotation breaks the translational symmetry of CNN. 1-degree step size is chosen, because for crop size of 144, rotation by 1 degree around its center creates a shift of ~1.3 pixels on the edge, yielding a different input image for CNN. It would be interesting to experimentally determine at what step size the rotation stops improving the performance

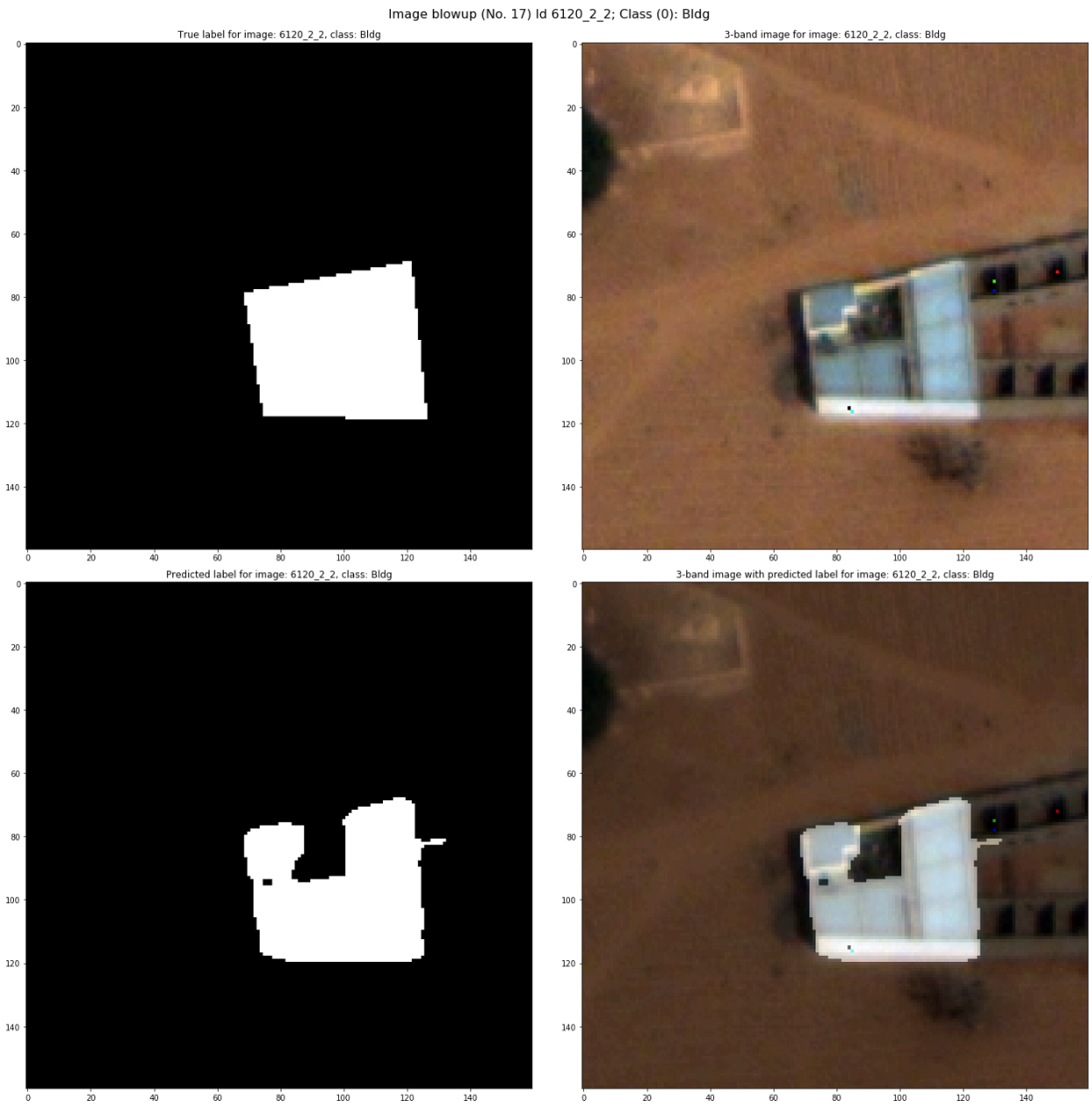
Testing Augmentation

One preferred position of U-net model is that it is completely convolutional, and accordingly the spatial elements of info information for preparing and deduction don't need to be the equivalent. As referenced above, in preparing, generally little picture patches and enormous cluster size are alluring. Nonetheless, for induction, it is valuable to make the picture fixes as extensive as conceivable to decrease the overhead made by disposing of expectations on the edge. Constrained by GPU memory, we could just take care of about a fourth of a picture (~1900*1900) into the U-net for derivation each time. Each quarter of a picture is turned by 0, 90, 180 and 360 degrees, and furthermore evenly and vertically flipped, for increase. Furthermore, the last expectation cover is the number juggling normal more than 8 veils, to diminish the change in the forecasts. Like the first U-net paper, we applied reflections on the fringe to improve the forecasts on the edge pixels.

Prediction



Some intriguing perceptions: the model is capable right some uncertain label from the training data. As appeared in the figures beneath, the 3-band picture appears to have a little "Empty" territory, loaded up with shadow made by regular light, in the center of the structure. While the "true" name incorporates the entire zone as Buildings class, the anticipated cover can accurately bar the shadow region. We have discovered two or three such models in the expectations.



The water classes: we have seen that the U-net would absolutely overfit whenever prepared with just 2 pictures for the Buildings class yet sums up well with ≥21 preparing pictures. It is intriguing to realize where the specific limit is and why. What's more, it additionally infers that profound learning won't work for the water classes in this challenge, which have cases just on not very many pictures. We have discovered that thresholding CCCI is better than a profound learning model for water classes.

Future Development

This model was essentially evolved dependent on its exhibition on class 0, which is the Buildings. The outcome can be additionally improved, by tweaking model boundaries for each class.

Our model absolutely neglects to anticipate the enormous and little vehicle classes, presumably because of the substantial valid and bogus names irregularity and too little spatial size of vehicles. Strategies, for example, over inspecting on obvious marks and preparing with extremely enormous bunch size may help with the previous.

References

U-Net: Convolutional Networks for Biomedical Image Segmentation: Olaf Ronneberger, Philipp Fischer, Thomas Brox