Prediction of Urban growth from satellite images

Lubor Pacak

April 26, 2021

1 Introduction

This report summarizes development of model for prediction of urban growth evolution in two different areas. Provided are 2 time series of satellite images from 2013 to 2015 for which we would like to predict how and whether urban-covered areas in these 2 locations are increasing. Some of the input images are annotated - in each pixel of annotated image is denoted, whether the area, corresponding to this pixel is urban covered, non-urban covered or cloud covered. These annotated pictures are used as training sample for the model and when the model is developed, it will be used for inference to pictures, where labels are missing. Location 1 contains 35 pictures and 31 of them are annotated. Location 2 contains 97 pictures and 15 of them is annotated.

2 Input data and pre-processing

Images were captured by Landsat 8 satellite, therefore each picture contains 12 channels (more about this satellite and channels can be found here. For our case, only Red, Green, Blue and near-infrared channels will be used for model development.

Annotation pictures (masks) are 3 channel pictures, where each channel contains binary values for corresponding classes - urban, non-urban and clouds.

Pictures from first location have resolution 339x338, pictures from second location have resolution 343x343. To have both groups of pictures of the same size, pictures are cropped to the size of 336x336.

For training purposes, few image augmentation techniques are applied.

- Original picture
- Left-right flipped original picture
- Up-down flipped original picture
- Affine transformation of picture
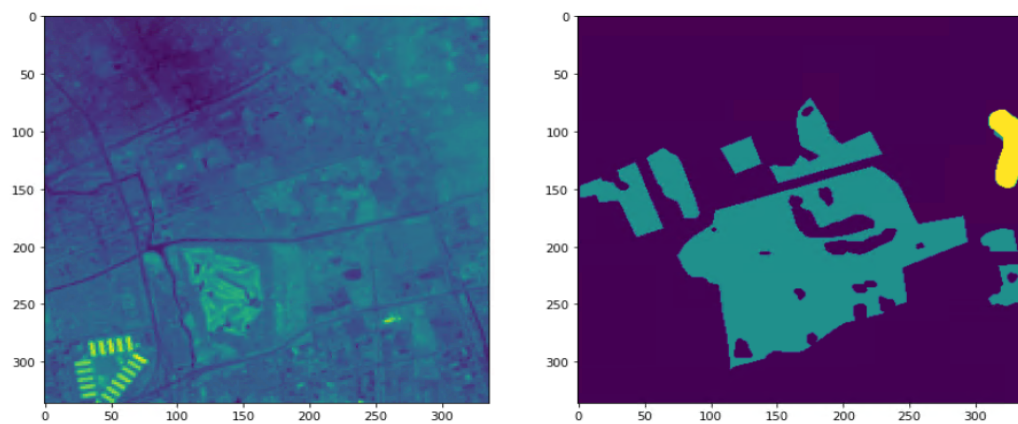- Affine transformation of picture with different parameters
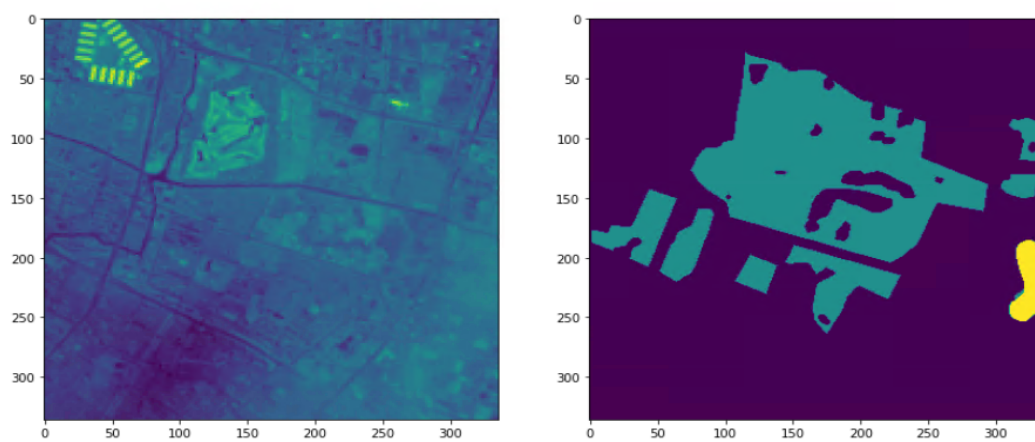
Figure 1: Original picture

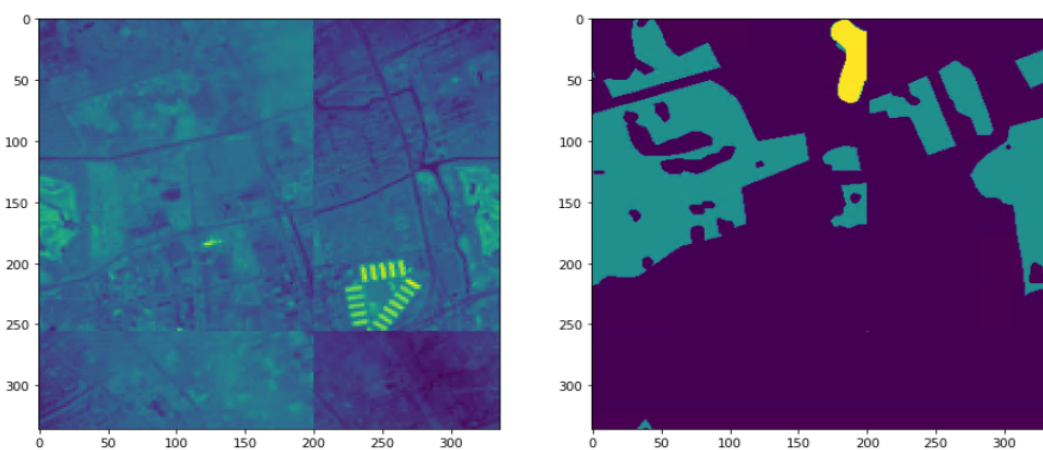Figure 2: Up-down flipped picture picture

Figure 3: Picture after application of one of affine transformation

3 Model Architecture

Since this task is about semantic image segmentation, i.e. we have input image and pixel-wise annotation of the image to several categories (3 in our case), it is necessary to select appropriate model architecture. One of the most famous model architecture is so-called U-net. U-net is convolutional neural network, with 2 parts - contraction and expansion. Sometimes it can be called encoder part and decoder part. Contraction follows typical architecture of CNN - creating lot of feature maps using convolutional and pooling layers. Exapansion path consists of upsampling with up-convolution and concatenation with corresponding cropped feature mask from contraction path.
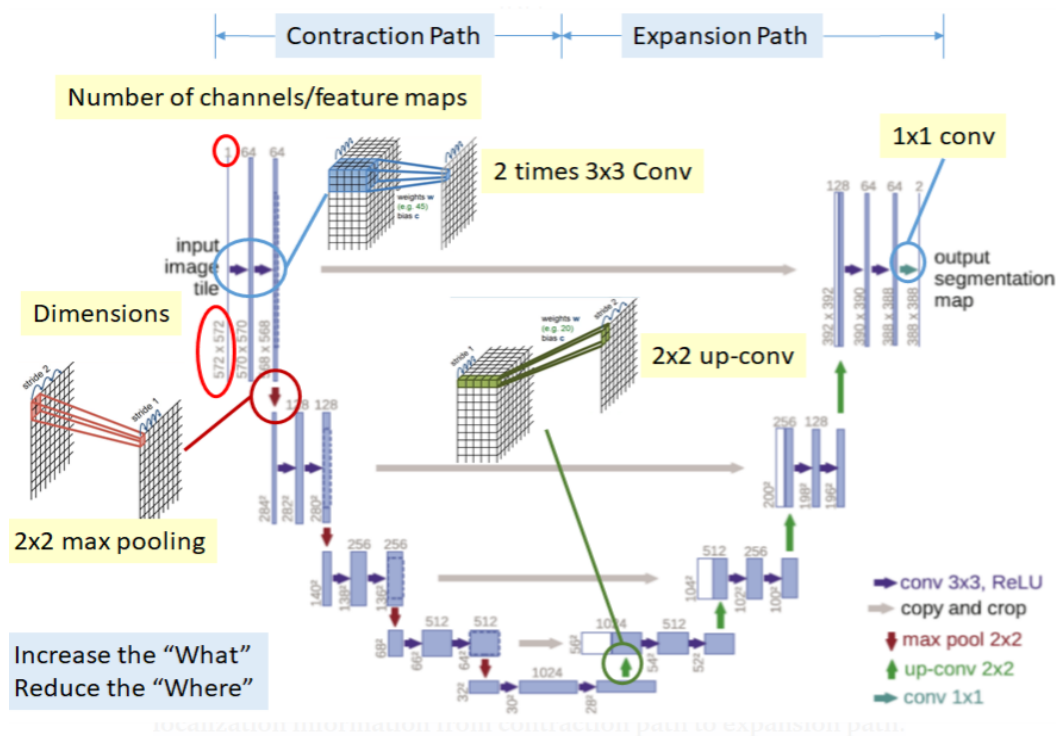

Figure 4: Example of U-NET model architecture

U-net used in my solution has depth of 3 and has following architecture:

1. Input is 4 channel image with sizes 336x336

2. 2x (2D convolution, BatchNorm, Relu) with Max pooling afterwards -> Output is 32 channels image with sizes 168x168

3. 2x (2D convolution, BatchNorm, Relu) with Max pooling afterwards -> Output is 64 channels image with sizes 84x84

4. 2x (2D convolution, BatchNorm, Relu) with Max pooling afterwards -> Output is 128 channels image with sizes 42x42

5. 2x (2D convolution, BatchNorm, Relu) with up convolution afterwards -> Output is 64 channels image with sizes 84x84

6. Concatenation results of previous and results of point 3.

7. 2x (2D convolution, BatchNorm, Relu) with up convolution afterwards -> Output is 32 channels image with sizes 168x168

8. Concatenation results of previous and results of point 2.

9. 2x (2D convolution, BatchNorm, Relu) with up convolution afterwards -> Output is 3 channels image with sizes 336x336

Optimizer used in my model is SGD with constant learning rate =0.01. Loss function used is standard categorical crossentropy with softmax as final activation function. Pixel-wise accuracy is measured.

4 Training of cross-validation model

Since we are facing task of model development with only few provided labeled training samples, I decided to train 4-fold Cross- Validation model. Each training image is therefore used 3 times in training sample and once in validation sample. 4 models are developed therefore. For final prediction (each pixel has 4 prediction, one from each model), majority voting is used.

Average validation accuracy was about 0.92. Learning curves of one of the folds is here:
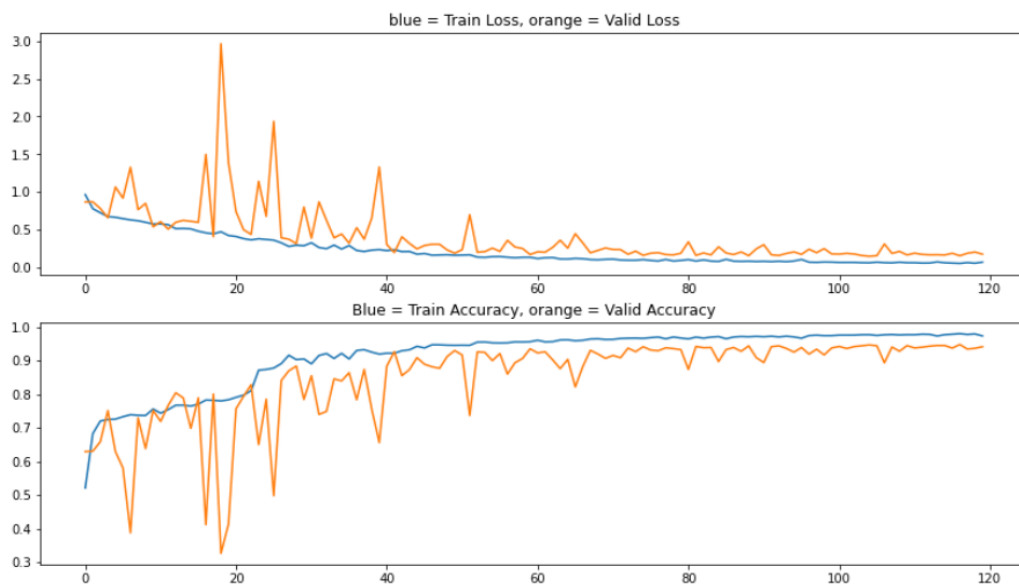

Figure 5: Training and validation losses and accuracies during model training. Horizontal axis is number of epoch.

You can see from the picture lot of spikes, mainly in the beginning of training phase. This is in my opinion caused by lack of validation samples. Other methods and parameters that were tried, but without any significant effect on performance of spikes smoothing: Pretrained Unet or Unet++ with resnet18 backbone from pytorch segmentations model package, Adam optimizer, various learning rates, scheduler of learning rates, dice loss. Please see attached .html or .ipynb files for learning curves of all 4 folds. It does not make much sense to compare predictions against real masks, because all available masks were used for training and there was no holdout testing dataset. Comparison with unseen images without provided masks are in the next section.

5 Prediction of Urban growth

Now comes the reason why we are doing all of this. Prediction of our model to all, even unlabeled pictures. Afterwards Urban growth, or urban growth index can be measured as percentage of urban-covered area in the picture. Unfortunately, there are clouds on lot of pictures and it is impossible to say, whether under the cloud was built or destroyed some house or road.

5.1 Prediction of Urban growth for the first area

First area where 31 pictures out o 35 are labeled is less interesting in my opinion, because solid urban growth prediction could be made even without the model, only from masks. But this was not the purpose of the task, therefore all the pictures are scored with trained model regardless the fact whether mask is provided or not.

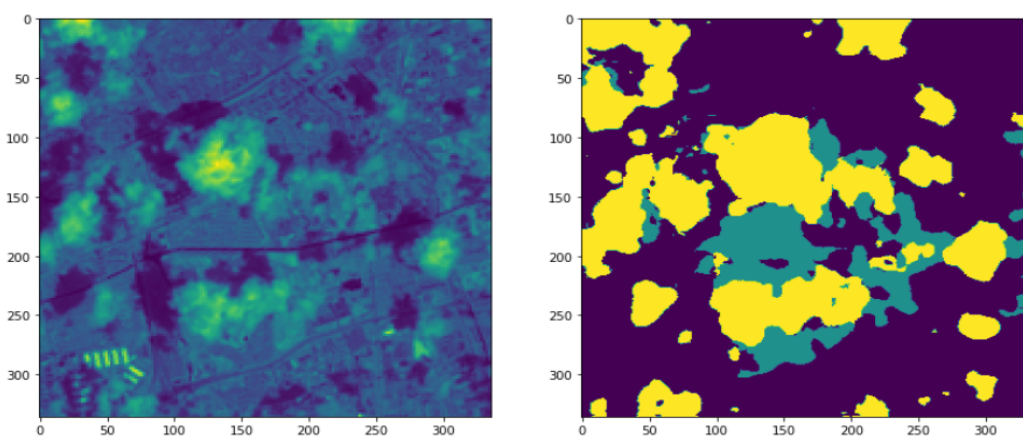Let's compare 2 prediction of the model with the real images, where label was not given.

Figure 6: Picture from 2014-07-31, where predictions of area in percentages are Non-urban 0.11, Urban 0.61, Clouds 0.27.
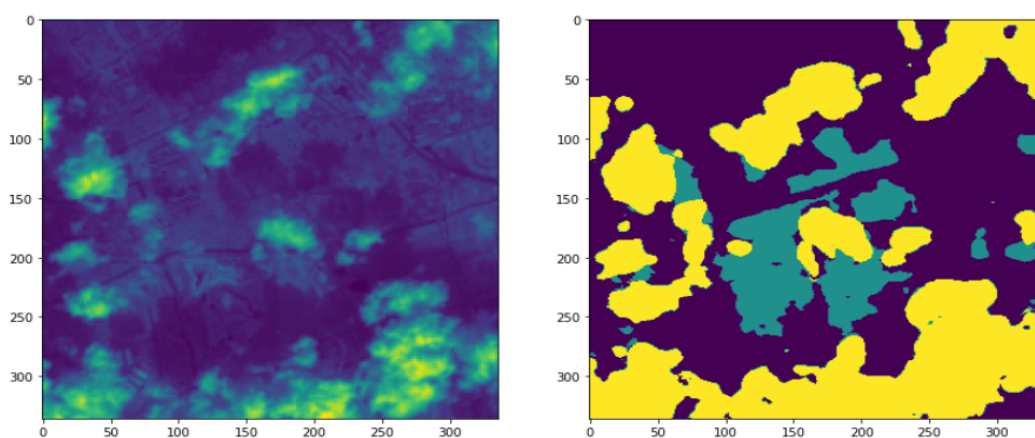

Figure 7: Picture from 2015-10-22, where predictions of area in percentages are Non-urban 0.11, Urban 0.53, Clouds 0.35.
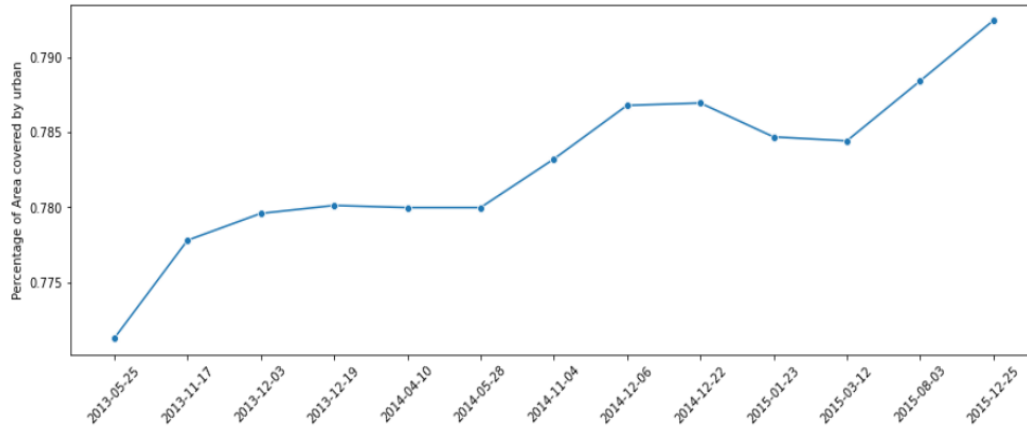
Figure 8: Evolution of Urban-covered area in time. Images where predictions of cloud-covered are was higher than 0.005 were excluded.

From these 2 provided predictions, it is obvious that model can captures clouds quite well. From the last plot, that denotes urban growth, it is obvious that even if there is positive trend in coverage by urban area, the progress is pretty small. Area covered by urban raises from 0.77 to 0.795.

5.2   Prediction of Urban growth for the second area

Second area is much more interesting, because only 15 of the 97 images were labeled. Completely same approach as in the previous section is applied.

Let's compare 3 prediction of the model with the real images, where label was not given. To see prediction for all images, please see enclosed .html or .ipynb files.
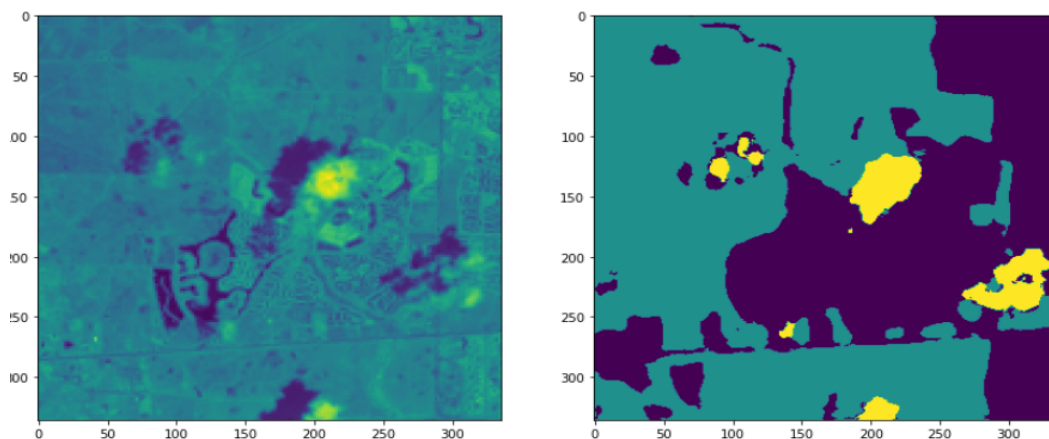

Figure 9: Picture from 2013-05-13, where predictions of area in percentages are Non-urban 0.56, Urban 0.40, Clouds 0.04.
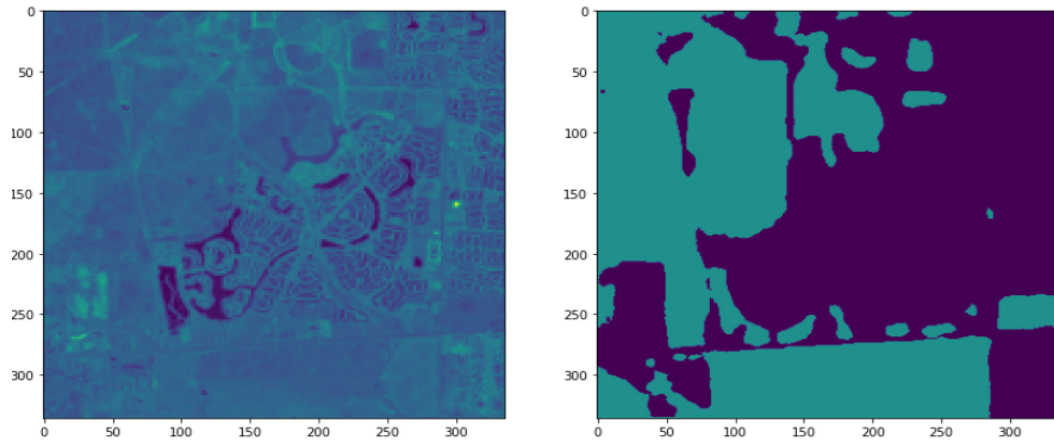
Figure 10: Picture from 2014-11-24, where predictions of area in percentages are Non-urban 0.44, Urban 0.56, Clouds 0.
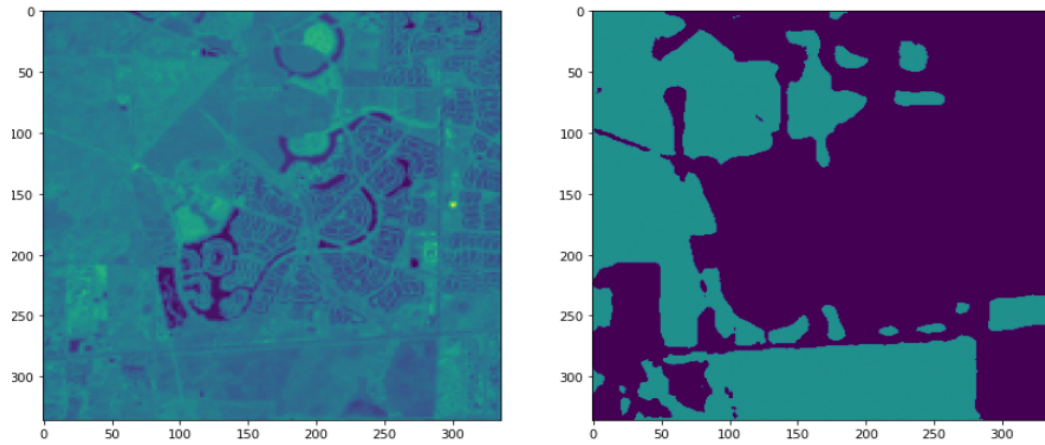

Figure 11: Picture from 2015-10-10, where predictions of area in percentages are Non-urban 0.38, Urban 0.62, Clouds 0.
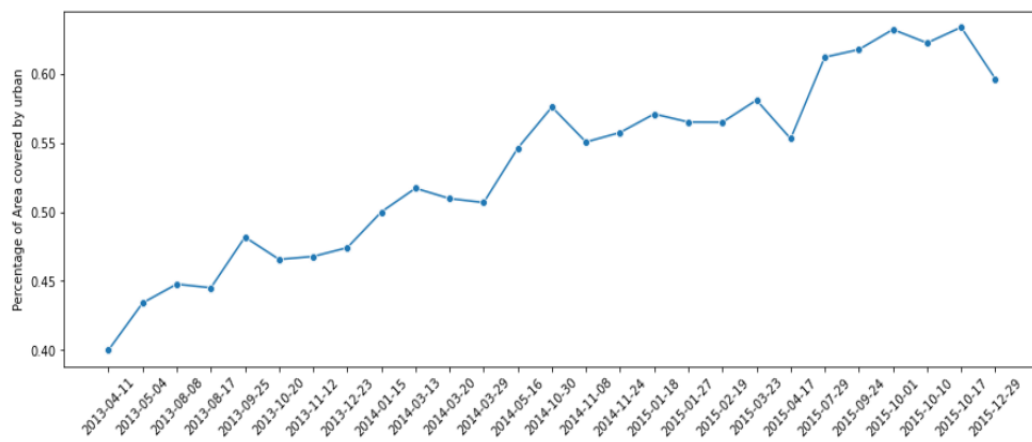

Figure 12: Evolution of Urban-covered area in time. Images where predictions of cloud-covered are was higher than 0.005 were excluded.

In can be seen that area covered by Urban is rapidly increasing.

6 Conclusion

U-net model for prediction of area in satellite images was developed and evaluated with average cross-validation accuracy 0.92. Consequently, the model was applied on time-series of images from 2 areas to detect whether the area covered by Urban is increasing or not. Positive increasing trend was observed in both areas, but much steeper in the second one. Used percentage can be used as valid econometric index. As the area for improvements it would be nice to have the urban-covered area more segmented - for example to big buildings, houses, roads, parks,.. Then all of these segments can be predicted and indicate something like wealth - index.