Vilniaus universitetas
Duomenų mokslo ir skaitmeninių technologijų institutas
Kognityvinių skaičiavimų grupė

Povilas Gudžius
Informatikos 09P doktorantas
Doktorantūros pradžios ir pabaigos metai: 2017–2021

Mokslinė ataskaita už 2019-2020 mokslo metus:
„Deep fully convolutional neural network architecture for object recognition in multispectral satellite imagery for low-latency algorithmic trading“

Darbo vadovas: Prof. Dr. Olga Kurasova
Konsultantas: Doc. Dr. Ernestas Filatovas

Research problem

On the back of improvements in optical Earth observation satellites such as propulsion systems, signal transmission, cost of launch and resolution; the number of CubeSats deployed to Lower Earth Orbit (LEO) has grown exponentially together with the imagery it produces [2]. Deep learning tools enable us to recognise objects in multi-spectral satellite imagery at scale and generate insights with practical industrial applications such as financial trading. According to the MIT technology review [3] and New Space Index [4], commercial satellite imagery in the next five years will reach high-resolution, near-real-time coverage of earth. In order to process it, we will have to deploy object recognition techniques that are capable not just accurately predict pixel-level classifications, but also to achieve this precision with low-latency.

We adopt semantic segmentation is also known as dense prediction due to the fact that it predicts the category of each pixel and it is more precise compared to object detection and scene interpretation [E8]. These classes are "semantically interpretable" and correspond to real-world categories. Current academic research is predominantly focused on accuracy rather that speed. In our research we have improved the architecture and hyperparameters of fully convolutional neural network (U-net) for semantic segmentation task. After conducting experiments, we were able achieve the State-Of-The-Art (SOTA) level accuracy for light vehicle object class. Additionally, we have optimised the U-net for the prediction speed by utilizing two leading GPU and TPU computational architectures.

Scientific significant research results developed in 2019-2020

During the academic year of 2019-2020 we have researched, developed and therefore propose advancements to the process of U-net design, hyperparameters tuning, training, and complexity optimisation to enhance the prediction accuracy and speed. The entire process from satellite imagery acquisition (P1) to end-signal generation and delivery to algorithmic trading system (P13) is depicted in Figure 2. Components from P5 to P10 coloured in blue represent areas of advancements proposed in this article and are described in the following subsections: 1) Network depth construction and feature extraction; 2) Computational complexity analysis; 3) Pixel frame sequencing.

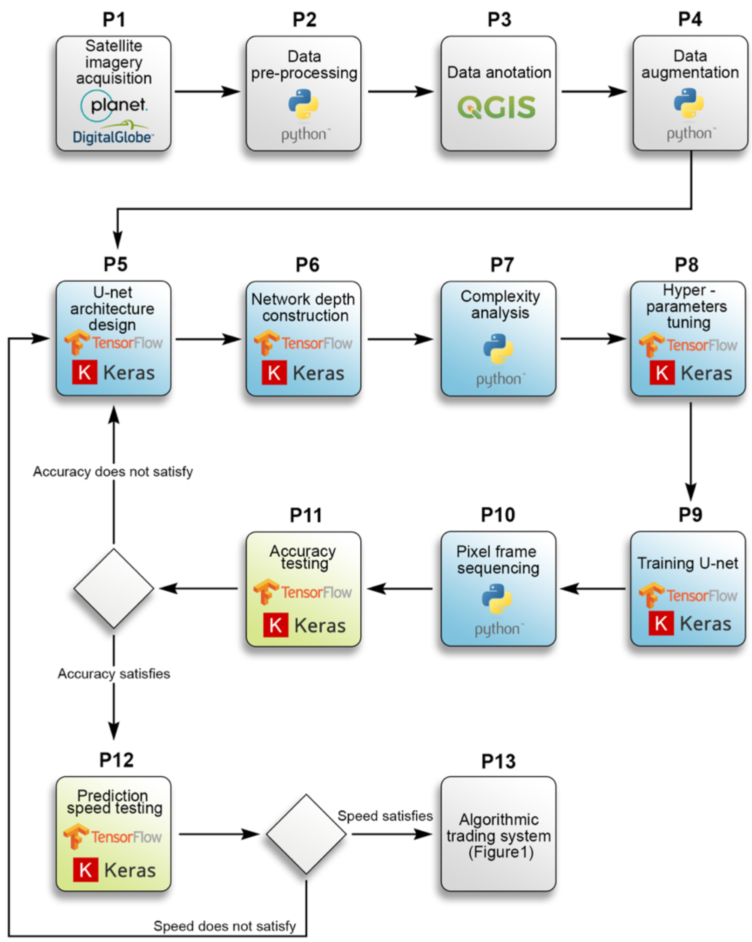


Figure 1. Object recognition in satellite imagery schematic workflow diagram

Stage P5 and P6 – Network architecture and depth

On the stage of **P5** and **P6** we have propose new U-net architectures representing the range of complexity in layers of the U-net. Each proposed U-net model consists of an even number of layers plus a single fully-connected layer with Sigmoid activation generating per-pixel semantic segmentation as an output (Figure 3). Models were initiated at fifteen convolutions and sequentially (in four groups) increased by six layers (three in the encoder and three in the decoder part) to a total of thirty-nine layers:

- *U-net_Model_1*: 21 layers in total (15 conv2d);
- *U-net_Model_2*: 27 layers in total (19 conv2d);
- *U-net_Model_3*: 33 layers in total (23 conv2d);
- *U-net_Model_4*: 39 layers in total (27 conv2d).

Figure 2. A visual representation of u-net models we developed

Beyond the layer size and complexity, we had a look at the layer composition and investigated the different concatenation differentials of maxout and concat for the different fully convolutional layers, the different between the layer output is depicted the figure 3 below:

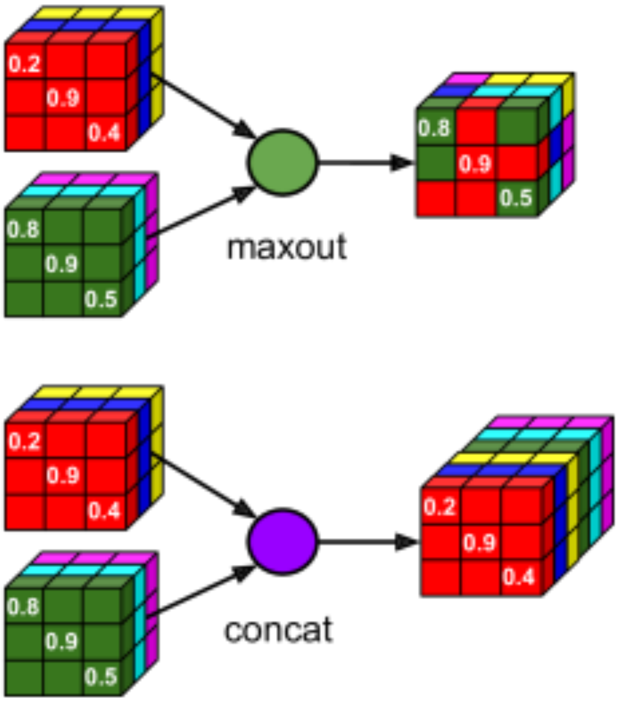


Figure 3. A visual representation maxout vs concat important for feature extraction

Stage P7 – Computational complexity

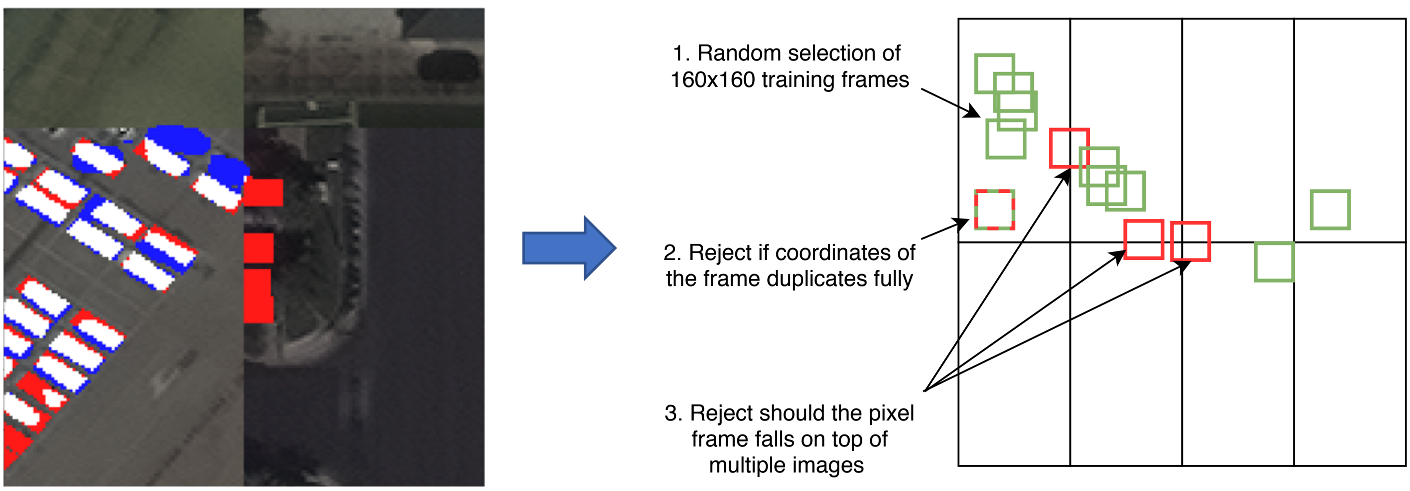
In order to establish the level of network complexity, we have adopted the FLOPs framework to the Convolutional Neural nets. For the design of efficient models, a detailed analysis of the number of floating-point operations (FLOPs) is required based on matrix operations such as matrix-matrix products (Figure 2, component P7). Matrix-Matrix Product of two matrices $A^{m\times n}$ and $C^{n\times l}$ needs mnl multiplications and $ml(n-1)$ summations, altogether $2mnl -$

ml FLOPs [48]. However, to our knowledge, there is no conventional benchmark that sets to define the computational complexity of the neural network [49]. Researches show that the number of operations in a network model can effectively estimate inference time [50]. The number of FLOPs represents how computationally expensive a model is [17]. We customize the FLOPs approach suggested by Sehgal et al. [17] to calculate the computational complexity of a neural network as defined in equation (1):

$$\textit{G-FLOPs} = \left[\sum_{e=1}^{E}\left(2 \times \left(\prod_{d=1}^{D_e} A_{ed}\right) \times F_e \times H_e \times W_e\right) + \sum_{b=1}^{B}\left(\prod_{x=1}^{X_b} P_{bx} \times \prod_{z=1}^{Z_b} O_{bz}\right)\right] / 10^9 \quad (1)$$

Stage P10 – Pixel frame sequencing

This step in the workflow of the object recognition in satellite imagery is a solution that improves the accuracy of training and prediction of the network by a significant 4.1%. The problem with the satellite imagery is that it has to be cropped and batched together missing the sceneries of the training set can creating distortion in the weights in backpropagation of the CNN. This novel method has helped to provide the “low-contextual-noise-input” type of training set, training windows for the neural net.



Stage P11 – Accuracy testing and experimentation

U-net model experimentation results of are provided in Table 1. We were able to achieve the state-of-the-art object recognition accuracy of 97.67% with *U-net_Model_2*. This network also maintained an FPO level 17.83%, and a 0.6162 Jaccard coefficient. A close second best, *U-net_model_3* has, however, provided a significant overprediction (FPO = 26.45%) rate. G-Flops metric indicates the computational complexity and *U-net_model_2* represents relatively light computational complexity with 6.9832 allowing faster prediction.

Table 1. Prediction accuracy results on the test set

	Accuracy (TPO) %	Overprediction (FPO) %	G-Flops	Jaccard coefficient
U-NET_MODEL_1	95.33	12.01	5.3218	0.6402
U-NET_MODEL_2	97.67	17.83	6.9832	0.6162
U-NET_MODEL_3	97.01	26.45	8.6443	0.5573
U-NET_MODEL_4	96.70	16.60	10.3053	0.6226

Conclusions

We have made a significant process in the FCN network design and optimization. This year we have also complete the development of the technological architecture and also proposed few unique and practical improvements in the complete workflow of object recognition satellite imagery and improvements to the accuracy of semantic segmentation. We will continue this research in finding a hybrid neural network that would provide similar accuracy of TPO's (true-positives), yet with the reduction of FPO (false positives) and also reduction for the computational complexity so that we could improve the predication speed.

Bibliography

[1] A. Monk, M. Prins and D. Rook (2019) "Rethinking Alternative Data in Institutional Investment," *The Journal of Financial Data Science,* vol. 1, no. 1, pp. 14-31

[2] A. Letizia, S. Marino, U. Ahmad and A. Alvino (2019) "Investigating the global socio-economic benefits of satellite industry and remote sensing applications," in *International Business Information Management Association Conference (IBIMA)*, Granada

[3] T. W. Gillespie, J. Chu, E. Frankenberg and D. Thomas (2007) "Assessment and prediction of natural hazards from satellite imagery," *Progress in Physical Geography,* vol. 31, no. 5, pp. 459-470

[4] E. Guirado, S. Tabik, M. L. Rivas, D. Alcaraz-Segura and F. Herrera (2019) "Whale counting in satellite and aerial images with deep learning," *Nature Scientific Reports,* vol. 9, no. 1, pp. 1-12

[5] G. Di Baldassarre, G. Schumann and P. Bates (2009) "Near real time satellite imagery to support and verify timely flood modelling," *Hydrological Processes: An International Journal,* vol. 23, no. 5, pp. 799-803

[6] S. Voigt (2016) "Global trends in satellite-based emergency mapping," *Science,* vol. 353, no. 6296, pp. 247-252

[7] A. S. Belward and J. O. Skøien (2015) "Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites," *Journal of Photogrammetry and Remote Sensing,* vol. 103, pp. 115-128

[8] D. Franco, C. Kourogiorgas, M. Marchese, A. Panagopoulos and F. Patrone (2019) "Small satellite and CubeSats: Survey of structures, architectures, and protocols," *International Journal of Satellite Communications and Networking,* vol. 4, no. 37, pp. 343-359

[9] A. Van Etten (2018) "You only look twice: Rapid multi-scale object detection in satellite imagery," *Preprint at arXiv:1805.09512*

[10] R. A. Marcum, C. H. Davis, G. J. Scott and T. W. Nivin (2017) "Rapid broad area search and detection of Chinese surface-to-air missile sites using deep convolutional neural networks," *Journal of Applied Remote Sensing,* vol. 11, no. 4

[11] O. Ronneberger, P. Fischer and T. Brox (2015) "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Cham

[12] H. Li, J. Chen, H. Lu and Z. Chi (2017) "CNN for saliency detection with low-level feature integration," *Neurocomputing,* vol. 226, pp. 212-220

[13] Z. N. Musa, I. Popescu and M. A (2015) "A review of applications of satellite SAR, optical, altimetry and DEM data for surface water modelling, mapping and parameter estimation," *Hydrology and Earth System Sciences,* vol. 9, no. 19, pp. 3755-3769