CNNs for Remote Sensing Semantic Segmentation

Ondřej Pešek, Martin Landa
Czech Techincal University in Prague, Faculty of Civil Engineering

At: Institut plánování a rozvoje, Praha

7th May 2025

Table of Contents

Methodology

Architectures Examined

- FCN [1]
- U-Net [2]
- SegNet [3]
- DeepLabv3+ [4] with three backbone models [5]
  - ResNet-50
  - ResNet-101
  - ResNet-152

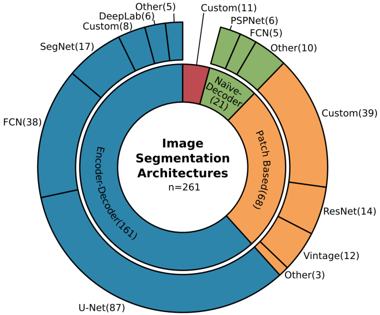


Figure 1: Overview of the most frequently used convolutional neural networks for semantic segmentation in the field of remote sensing. Source: [6]

Approaches Examined

- Different features of interest
  - Clouds
  - Vegetation
  - Urban features
- Different approaches
  - Land cover classification
  - Land use classification
- Different spatial resolution
  - VENμS – 5 to 10 metres
  - Sentinel-2 – 10 to 60 metres
  - Aerial imagery – 10 centimetres
- Different input bandsets
  - All bands provided in the data source
  - RGB
  - All bands enhanced by NDVI
- Different class number
  - Binary classification
  - Multi-class classification
- Different dataset sizes
- Compare with random forests

Approaches Examined — the effect of common anti-overfitting methods

- Dropout [7]
- Data augmentation

Use cases

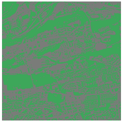
Use Cases — Urban Greenery

Use case definition:

- Published in Remote Sensing Applications: Society and Environment journal: https://doi.org/10.1016/j.rsase.2024.101238
- Training dataset published on Zenodo server: https://zenodo.org/records/8413116
- Using Sentinel-2 data as input
- Using 10-metre imagery is rare for urban greenery detection
- The first study using CNNs for urban green land use classification
- Very limited training dataset
- Three levels
  - Level 1: Non-vegetated pixels, vegetation
  - Level 2: Non-vegetated pixels, non-recreational vegetation, recreational vegetation
  - Level 3: Non-vegetated pixels, non-recreational vegetation, low recreational vegetation, high recreational vegetation

Use Cases — Urban Greenery

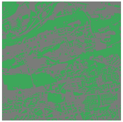
(a) A patch of a Sentinel-2 image tile.



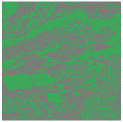
(b) Ground-truth mask.

(c) FCN — D 0%. OA 72.8%

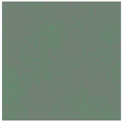
(d) FCN — D 50%. OA 72.2%



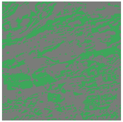
(e) U-Net — D 0%. OA 91.8%



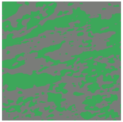
(f) U-Net — D 50%. OA 93.6%



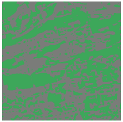
(g) SegNet — D 0%. OA 51.7%



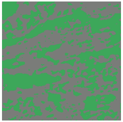
(h) SegNet — D 50%. OA 82.2%



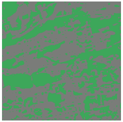
(i) DeeplabV3+ with ResNet-50 — D 0%. OA 75.5%



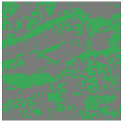
(j) DeeplabV3+ with ResNet-50 — D 50%. OA 80.2%



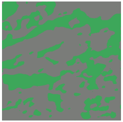
(k) DeeplabV3+ with ResNet-101 — D 0%. OA 75.0%



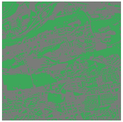
(l) DeeplabV3+ with ResNet-101 — D 50%. OA 74.3%



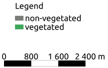
(m) DeeplabV3+ with ResNet-152 — D 0%. OA 76.4%



(n) DeeplabV3+ with ResNet-152 — D 50%. OA 74.4%



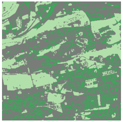
(o) Random forests. OA 92.7%



(p) Legend.

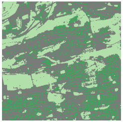
Use Cases — Urban Greenery

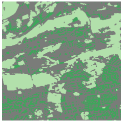

(a) A patch of a Sentinel-2 image tile.

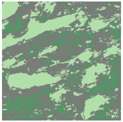

(b) Ground-truth mask.

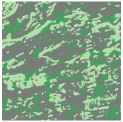

(c) FCN — D 0%. OA 64.8%

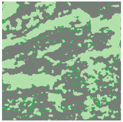

(d) FCN — D 50%. OA 64.5%

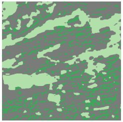

(e) U-Net — D 0%. OA 87.1%

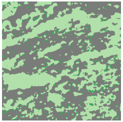

(f) U-Net — D 50%. OA 82.4%

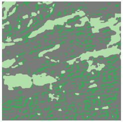

(g) SegNet — D 0%. OA 72.8%

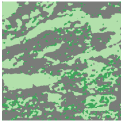

(h) SegNet — D 50%. OA 65.0%

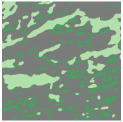

(i) DeeplabV3+ with ResNet-50 — D 0%. OA 67.6%

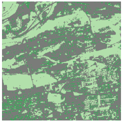

(j) DeeplabV3+ with ResNet-50 — D 50%. OA 73.3%

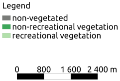

(k) DeeplabV3+ with ResNet-101 — D 0%. OA 65.1%


(l) DeeplabV3+ with ResNet-101 — D 50%. OA 71.5%


(m) DeeplabV3+ with ResNet-152 — D 0%. OA 66.2%


(n) DeeplabV3+ with ResNet-152 — D 50%. OA 64.4%


(o) Random forests. OA %84.3

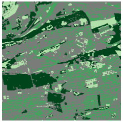


(p) Legend.

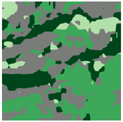
Ondřej Pešek

Use Cases — Urban Greenery

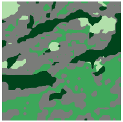
(a) A patch of a Sentinel-2 image tile.



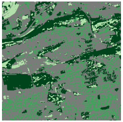
(b) Ground-truth mask.



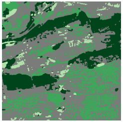
(c) FCN — D 0%. OA 59.1%



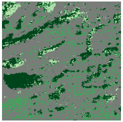
(d) FCN — D 50%. OA 58.5%



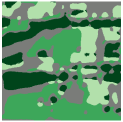
(e) U-Net — D 0%. OA 80.6%



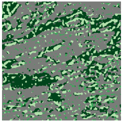
(f) U-Net — D 50%. OA 73.9%



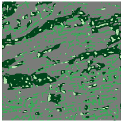
(g) SegNet — D 0%. OA 60.6%



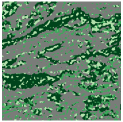
(h) SegNet — D 50%. OA 45.8%



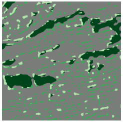
(i) DeeplabV3+ with ResNet-50 — D 0%. OA 59.1%



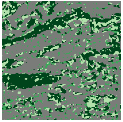
(j) DeeplabV3+ with ResNet-50 — D 50%. OA 65.9%



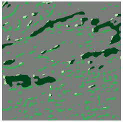
(k) DeeplabV3+ with ResNet-101 — D 0%. OA 58.7%



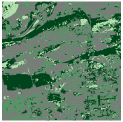
(l) DeeplabV3+ with ResNet-101 — D 50%. OA 60.0%



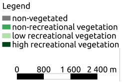
(m) DeeplabV3+ with ResNet-152 — D 0%. OA 57.3%



(n) DeeplabV3+ with ResNet-152 — D 50%. OA 60.7%



(o) Random forests. OA 80.9%

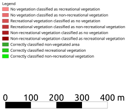


(p) Legend.

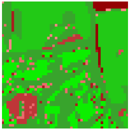
Use Cases — Urban Greenery

(a) Bing Maps aerial photo covering the sample area.

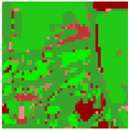
(b) Ground-truth mask.



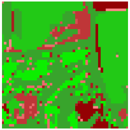
(c) Legend.



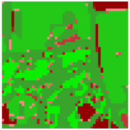
(d) After a training using full-band input.



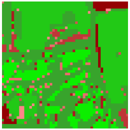
(e) Using only RGB.



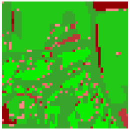
(f) Using full-band input enhanced with NDVI.



(g) Using augmented full-band input.



(h) Using only augmented RGB.



(i) Using augmented full-band input enhanced with NDVI.

Use Cases — Urban Greenery

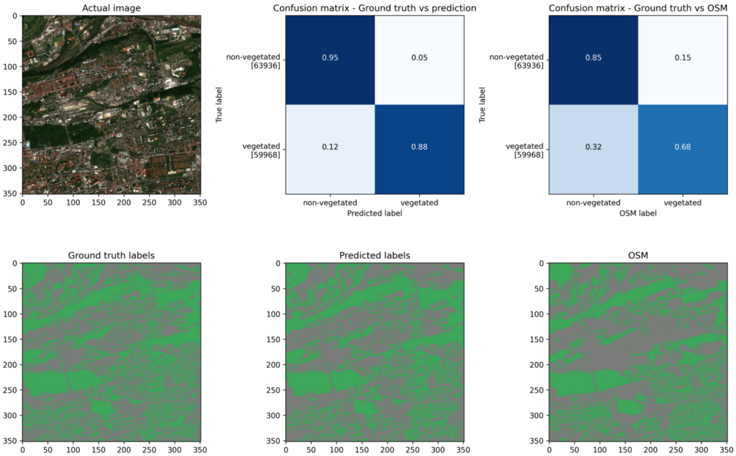


Figure 6: Comparison of OSM data and results from the level-1 detection of U-Net without dropout layers trained on the full-band dataset. The validation patch.

Use Cases — Urban Greenery

Conclusions:

- CNNs reached good resuls
  - Vegetation detection: Highest overall accuracy over 95%
  - Vegetation land use classifciation: Highest overall accuracy over 90%
- CNNs can handle even the land use task
- Results reached using CNNs are more accurate than available open datasets
- Extra bands and NDVI improve results
- Dropout did not improve results
- Data augmentation improved results
- The best results reached by U-Net

Use Cases — Road Surface Classification

Use case definition:

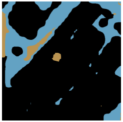
- Submitted for publishing, waiting for the second round of peer reviews... still
- Training dataset published on Zenodo server: `https://zenodo.org/records/10602515`
- Using 10-cm aerial data as input
- The first study using CNNs to distinguish between modular and compact road surface on remotely sensed data

Use Cases — Road Surface Classification

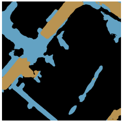
(a) A patch of the aerial imagery.

(b) Ground-truth mask.

(c) FCN — D 0%. OA 89.5%

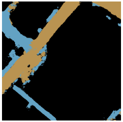


(d) FCN — D 50%. OA 71.4%

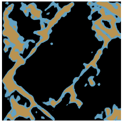


(e) U-Net — D 0%. OA 91.1%

(f) U-Net — D 50%. OA 77.6%

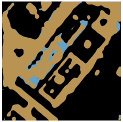


(g) SegNet — D 0%. OA 90.3%

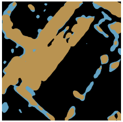


(h) SegNet — D 50%. OA 66.7%

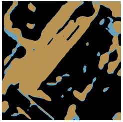
(i) DeeplabV3+ with ResNet-50 — D 0%. OA 81.2%



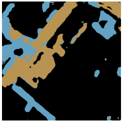
(j) DeeplabV3+ with ResNet-50 — D 50%. OA 67.2%



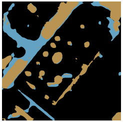
(k) DeeplabV3+ with ResNet-101 — D 0%. OA 75.9%



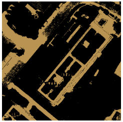
(l) DeeplabV3+ with ResNet-101 — D 50%. OA 69.0%



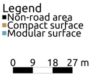
(m) DeeplabV3+ with ResNet-152 — D 0%. OA 77.4%



(n) DeeplabV3+ with ResNet-152 — D 50%. OA 75.3%



(o) Random forests. OA 59.4%

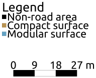


(p) Legend.

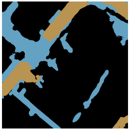
Use Cases — Road Surface Classification

(a) A patch of the aerial imagery.

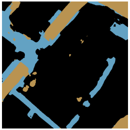
(b) Ground-truth mask.



(c) Legend.

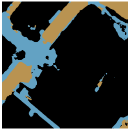


(d) After a training using full-band input.



(e) Using only RGB.

(f) Using full-band input enhanced with NDVI.



(g) Using augmented full-band input.

(h) Using only augmented RGB.

(i) Using augmented full-band input enhanced with NDVI.

Use Cases — Road Surface Classification

Conclusions:

- CNNs can deal with the task
- Extra bands improved results, NDVI did not
- Dropout did not improve results
- Ambivalent effect of simple data augmentation
- The best results reached by U-Net

Conclusions

- U-Net was always the best performing of the tested architectures
- In all the presented use cases, U-Net overperformed existing approaches and open datasets
- FCN and DeepLabv3+ are training data-greedier than the other architectures
- The number of classes did not have any influence on the performance of the CNN
- Extra bands and NDVI (usually) improve the performance but even the results reached on pure RGB outperform existing methods
- Thanks to its context consciousness, CNNs can deal even with challenging tasks as land use or clouds over desert detection

Conclusions — Desired Extensions

- More use cases
- Research on the relationship between batch normalisation and dropout
- Deeper research on various data augmentation techniques
- More architectures
- Make GIS addons out of the accompanying source code

References I

[1] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *IEEE conference on Computer Vision and Pattern Recognition CVPR*, pages 3431–3440, 2015.

[2] O. RONNEBERGER, P. FISCHER, and T. BROX. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI*, pages 234–241, 2015.

[3] V. BADRINARAYANAN, A. KENDALL, and R. CIPOLLA. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[4] L. CHEN, Y. ZHU, G. PAPANDREOU, F. SCHROFF, and H. ADAM. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European Conference on Computer Vision ECCV*, pages 833–851, 2018.

[5] K. HE, X. ZHANG, S. REN, and J. SUN. Deep residual learning for image recognition. pages 770–778, 2016.

[6] T. HOESER and C. KUENZER. Object detection and image segmentation with deep learning on earth observation data: A review-part ii: Applications. *Remote Sensing*, 11(10), 2020.

[7] G. E. HINTON, N. SRISTAVA, A. KRIZHEVSKY, I. SUTSKEVER, and R. R. SALAKHUTDINOV. Improving neural networks by preventing co-adaptation of feature detectors, 2012. URL `http://arxiv.org/abs/1207.0580`.

Thank you for your attention.