CNNs for Road Surface Semantic Segmentation

Ondřej Pešek
Czech Technical University in Prague, Faculty of Civil Engineering

Telč, 2023

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

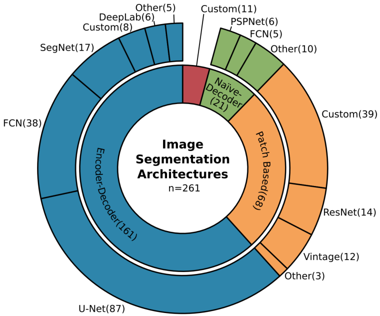


Figure 1: Overview of the most frequently used convolutional neural networks for semantic segmentation in the field of remote sensing. Source: [6]

Use Case Definition

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

Results

Table 1: Overall accuracy values in per cent for various configurations and architectures. The values were computed over the validation dataset. Rows abbreviations explained: `dN%` means dropout, `N` indicates the dropout ratio, and `rnN` indicates the ResNet depth in the DeepLabv3+ backbone model. Columns abbreviations explained: `fb` means the full-band images dataset, `rgb` means the dataset comprising the red, green, and blue bands, `rgb_ndvi` means the `rgb` dataset enhanced by the NDVI, and the _a suffix means that the dataset was augmented. The best result is highlighted in bold.

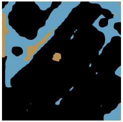
Architecture	fb	fb_a	rgb	rgb_a	rgb_ndvi	rgb_ndvi_a
FCN_d00	89.5	89.7	89.4	87.0	89.2	90.2
FCN_d50	71.4	83.2	80.9	67.8	74.2	73.0
U-Net_d00	**91.1**	89.2	87.6	90.3	90.9	89.6
U-Net_d50	77.6	67.7	69.1	71.2	79.8	66.8
SegNet_d00	90.3	88.1	86.8	87.8	89.6	88.8
SegNet_d50	66.7	83.7	80.1	75.4	79.2	84.6
DeepLabV3+_rn50_d00	81.2	79.7	81.3	86.9	75.4	84.7
DeepLabV3+_rn50_d50	67.2	69.3	70.9	81.9	66.5	63.1
DeepLabV3+_rn101_d00	75.9	82.6	80.0	84.9	83.9	86.2
DeepLabV3+_rn101_d50	69.0	65.7	83.2	81.7	62.6	48.1
DeepLabV3+_rn152_d00	77.4	85.9	84.4	83.2	80.8	74.0
DeepLabV3+_rn152_d50	75.3	69.8	79.3	82.1	63.8	70.4
random_forests	59.4	59.4	65.5	66.5	60.5	60.5

Results

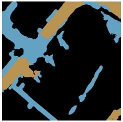
(a) A patch of the aerial imagery.

(b) Ground-truth mask.

(c) FCN — D 0%. OA 89.5%

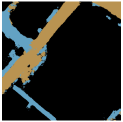


(d) FCN — D 50%. OA 71.4%

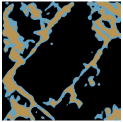


(e) U-Net — D 0%. OA 91.1%

(f) U-Net — D 50%. OA 77.6%

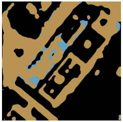


(g) SegNet — D 0%. OA 90.3%

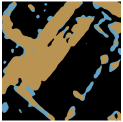


(h) SegNet — D 50%. OA 66.7%

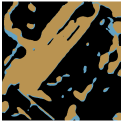
(i) DeeplabV3+ with ResNet-50 — D 0%. OA 81.2%



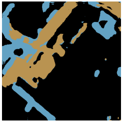
(j) DeeplabV3+ with ResNet-50 — D 50%. OA 67.2%



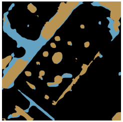
(k) DeeplabV3+ with ResNet-101 — D 0%. OA 75.9%



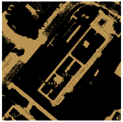
(l) DeeplabV3+ with ResNet-101 — D 50%. OA 69.0%



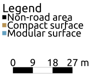
(m) DeeplabV3+ with ResNet-152 — D 0%. OA 77.4%



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



(o) Random forests. OA 59.4%

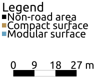


(p) Legend.

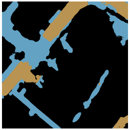
Results

(a) A patch of the aerial imagery.

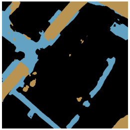
(b) Ground-truth mask.



(c) Legend.

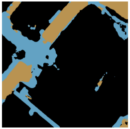


(d) After a training using full-band input.



(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.

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 — Desired Extensions

- 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.