0.1 Comparison with earlier work

Comparison on the ISPRS Potsdam dataset

Model	Proposed Model	SBD
Resolution used	**30 cm/pixel**	12.5 cm/pixel
Pixelwise F1 score	0.851	**0.884**
Vehiclewise F1 score	**0.800**	0.773
Mean prediction error counting cars	**1.67 %**	3.57 %
Evaluation time per tile*	**0.19 seconds**	28.19 seconds

Table 1: Shows the comparison between the proposed model and the Segment before you Detect (SBD) model [1] on the Potsdam dataset.
* The SBD model was evaluated on a Tesla K20 which can at maximum perform $3.52 * 10^{12}$ 32 bit floating point operations per second. The proposed model was evaluated on a Tesla K80 wich can perform at maximum $8.74 * 10^{12}$ 32 bit floating point operations per second. Therefore the evaluation time on the SBD model was multiplied with $3.52/8.74 \approx 0.4027$ to make fair comparisons. The evaluation time should therefore not be regarded as exact but as an indication of the speed difference between the two models.

Comparison on the Vedai dataset

Model	**Detection time per image***	**Vehiclewise F1 score**
Faster R-CNN (Z&F)	0.1998	0.212
Faster R-CNN (VGG-16)	0.2248	0.225
Fast R-CNN (VGG-16)	3.1465	0.224
CCNN	0.2736	0.305
Proposed Model	$\approx$ **0.19/4**	**0.554**

Table 2: Shows the comparison between the proposed model and the Faster R-CNN (Z&F), Faster R-CNN (VGG-16), Fast R-CNN (VGG-16) [2] and the Cascaded Convolutional Neural Networks (CCNN) [3] on the Vedai dataset. * The other models were evaluated on a Titan X which can at maximum perform $11{*}10^{12}$ 32 bit floating point operations per second. The proposed model was evaluated on a Tesla K80 wich can perform at maximum $8.74 * 10^{12}$ 32 bit floating point operations per second. Therefore the evaluation time on the compared models were multiplied with $11/8.74 \approx 1.2586$ to make fair comparisons. The evaluation time should therefore not be regarded as exact but as an indication of the speed difference between the two models.

Bibliography

[1] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. "On the usability of deep networks for object-based image analysis". In: *arXiv:1609.06845 [cs]* (Sept. 2016). URL: http://arxiv.org/abs/1609.06845 (visited on 03/09/2018).

[2] Matthew D. Zeiler and Rob Fergus. "Visualizing and Understanding Convolutional Networks". en. In: *Computer Vision – ECCV 2014*. Lecture Notes in Computer Science. Springer, Cham, Sept. 2014, pp. 818–833. ISBN: 978-3-319-10589-5 978-3-319-10590-1. DOI: 10.1007/978-3-319-10590-1_53. URL: https://link.springer.com/chapter/10.1007/978-3-319-10590-1_53 (visited on 04/19/2018).

[3] Jiandan Zhong, Tao Lei, and Guangle Yao. "Robust Vehicle Detection in Aerial Images Based on Cascaded Convolutional Neural Networks". en. In: *Sensors* 17.12 (Nov. 2017), p. 2720. DOI: 10.3390/s17122720. URL: http://www.mdpi.com/1424-8220/17/12/2720 (visited on 04/19/2018).