A Fine-Grained Dataset and its Efficient Semantic Segmentation for Unstructured Driving Scenarios

technik autonomer systeme

Kai A. Metzger, Peter Mortimer, and Hans-Joachim Wuensche

Introduction

We introduce TAS500, a novel semantic segmentation dataset for autonomous driving in unstructured outdoor environments. Using TAS500 as our training dataset, we study how a more fine-grained labeling policy affects the overall performance of the semantic segmentation.

Legend:
asphalt (■), building (■), bush (■), car (■), ego vehicle (■), forest (■), gravel (■), high grass (■), low grass (■), miscellaneous vegetation (■), person (■), soil (■), sky (■), tree crown (■), tree trunk (■)

Data Collection

The data was collected using the autonomous vehicle MuCAR-3 [1]. Our vision system is mounted on a multifocal active/reactive camera platform called MarVEye [2]. This allows us to pan the camera when turning into a corner. We cut off most of the sky and ego vehicle hood from all images. The final images have a resolution of 620 px × 2026 px.

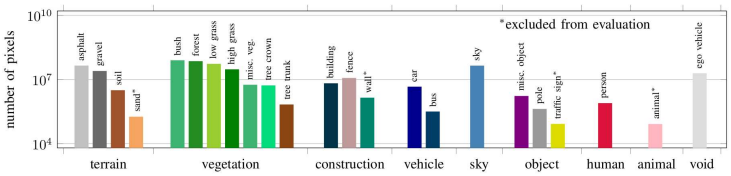
Dataset Comparison

Laying our focus on annotating scenes from unstructured outdoor environments naturally leads to the observation that the scenes in our TAS500 dataset contain more terrain and vegetation pixels per image than other urban scene understanding datasets [3, 4]. We also compare TAS500 to the multimodal Freiburg Forest dataset [5], which also contains scenes in unstructured outdoor environments.

Dataset	Resolution ($H \times W$)	Terrain (%)	Vegetation‡ (%)
Cityscapes†	1024×2048	38.0	14.1
KITTI†	375×1242	10.0	30.3
FreiburgForest	487×880	9.3	25.7 + 39.9
TAS500 (**ours**)	620×2026	18.4	35.3 + 23.3

† The terrain class is composed of the road and sidewalk classes.
‡ Vegetation is split into drivable and non-drivable where applicable.

Semantic Classes



TAS500 provides a fine-grained segmentation for the terrain and vegetation categories.

Boundary Jaccard

The common IoU evaluation metric is known to be biased towards semantic classes that cover a large image area. A fine-grained segmentation leads to many smaller semantic classes. The Boundary Jaccard (BJ) measure [6] evaluates scene segmentations along the class borders, which is more informative for a fine-grained segmentation of a scene.

The area of interest evaluated by the BJ metric is within θ pixels to a class border in the ground truth and the prediction. In our experiments, we set $\theta = 10$.

Fine-Grained Semantic Segmentation

We analyzed the segmentation performance for different class granularities for the categories terrain and vegetation. We increased the label granularity in each step. Training Fast-SCNN [7] for the different class granularities has shown a drop in segmentation performance when separating the visually similar low grass and high grass. The same holds for the fine-grained segmentation of non-drivable vegetation.

# C	Level of granularity	Val mIoU (%)	Val mBJ (%)
14	Baseline (terrain, vegetation)	53.00	52.79
17	Terrain → *Asphalt, gravel, soil, sand*	53.54 (+0.54)	**54.29 (+1.50)**
18	Vegetation → *Drivable, non-drivable*	**54.77 (+1.77)**	53.56 (+0.77)
19	Drivable vegetation → *Low grass, high grass*	52.30 (-0.70)	42.88 (-9.91)
23	Non-drivable vegetation → *Bush, forest, misc. vegetation, tree crown, tree trunk*	49.70 (-3.30)	50.12 (-2.67)

In brackets we compare the performance to the 14 class baseline.

Qualitative Comparison

Legend:
asphalt (■), building (■), car (■), drivable vegetation (■), ego vehicle (■), gravel (■), non-drivable vegetation (■), obstacle (■), person (■), pole (■), soil (■), sky (■)

Download the TAS500 dataset at mucar3.de/icpr2020-tas500.

[1] M. Himmelsbach, T. Luettel, F. Hecker, F. Hundelshausen, and H.-J. Wuensche, "Autonomous Off-Road Navigation for MuCAR-3," KI-Künstliche Intelligenz, vol. 25, no. 2, pp. 145–149, 2011.
[2] A. Unterholzner and H.-J. Wuensche, "Hybrid Adaptive Control of a Multi-Focal Vision System," in IEEE Intelligent Vehicles Symposium, San Diego, CA, USA, Jun. 2010, pp. 534–539.
[3] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[4] H. Alhaija, S. Mustikovela, L. Mescheder, A. Geiger, and C. Rother, "Augmented reality meets computer vision: Efficient data generation for urban driving scenes," International Journal of Computer Vision (IJCV), 2018.
[5] A. Valada, G. Oliveira, T. Brox, and W. Burgard, "Deep Multispectral Semantic Scene Understanding of Forested Environments using Multimodal Fusion," in International Symposium on Experimental Robotics (ISER), Tokyo, Japan, Oct. 2016.
[6] E. Fernandez-Moral, R. Martins, D. Wolf, and P. Rives, "A new metric for evaluating semantic segmentation: leveraging global and contour accuracy," in IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018, pp. 1051–1056.
[7] R. P. Poudel, S. Liwicki, and R. Cipolla, "Fast-SCNN: Fast Semantic Segmentation Network," arXiv preprint arXiv:1902.04502, 2019.

www.unibw.de/tas
peter.mortimer@unibw.de

Autonomous Systems Technology
University of the Bundeswehr Munich, Neubiberg, Germany