Home » Benchmark Suite

Benchmark Suite

We offer a benchmark suite together with an evaluation server, such that authors can upload their results and get a ranking regarding the different tasks (pixel-level, instance-level, and panoptic semantic labeling). If you would like to submit your results, please register, login, and follow the instructions on our submission page.

Pixel-Level Semantic Labeling Task

The first Cityscapes task involves predicting a per-pixel semantic labeling of the image without considering higher-level object instance or boundary information.

Metrics

To assess performance, we rely on the standard Jaccard Index, commonly known as the PASCAL VOC intersection-over-union metric IoU = TP / (TP+FP+FN) [1], where TP, FP, and FN are the numbers of true positive, false positive, and false negative pixels, respectively, determined over the whole test set. Owing to the two semantic granularities, i.e. classes and categories, we report two separate mean performance scores: $IoU_{category}$ and IoU_{class}. In either case, pixels labeled as void do not contribute to the score.

It is well-known that the global IoU measure is biased toward object instances that cover a large image area. In street scenes with their strong scale variation this can be problematic. Specifically for traffic participants, which are the key classes in our scenario, we aim to evaluate how well the individual instances in the scene are represented in the labeling. To address this, we additionally evaluate the semantic labeling using an instance-level intersection-over-union metric iIoU = iTP / (iTP+FP+iFN). Again iTP, FP, and iFN denote the numbers of true positive, false positive, and false negative pixels, respectively. However, in contrast to the standard IoU measure, iTP and iFN are computed by weighting the contribution of each pixel by the ratio of the class' average instance size to the size of the respective ground truth instance. It is important to note here that unlike the instance-level task below, we assume that the methods only yield a standard per-pixel semantic class labeling as output. Therefore, the false positive pixels are not associated with any instance and thus do not require normalization. The final scores, $iIoU_{category}$ and $iIoU_{class}$, are obtained as the means for the two semantic granularities.

Results

Detailed results

Detailed results including performances regarding individual classes and categories can be found here be found here.

Usage

Use the buttons in the first row to hide columns or to export the visible data to various formats. Use the widgets in the second row to filter the table by selecting values of interest (multiple selections possible). Click the numeric columns for sorting.

Column visibility | CSV | Excel | PDF

name | fine | coarse

16-bit | depth | video | sub | code

Show 20 entries

Search:

name	fine	coarse	16-bit	depth	video	sub	IoU class	iIoU class	IoU category
Hyundai Mobis AD Lab	yes	yes	no	no	no	no	83.8	65.0	92.4
HRNetV2 + OCR (w/ ASP)	yes	yes	no	no	no	no	83.7	64.8	92.4
iFLYTEK-CV	yes	yes	no	no	no	no	83.6	64.7	92.1
Improving Semantic Segmentation via Video Propagation and Label Relaxation	yes	yes	no	no	yes	no	83.5	64.4	92.2
GALD-Net	yes	yes	yes	yes	no	no	83.3	64.5	92.3
HRNetV2 + OCR	yes	yes	no	no	no	no	83.3	62.0	92.1
NV-ADLR	yes	yes	no	no	no	no	83.2	64.2	92.1
GGCF	yes	yes	no	no	no	no	83.2	63.0	92.0
GALD-net	yes	yes	no	no	no	no	83.1	63.5	92.2
Tencent AI Lab	yes	yes	no	no	no	no	82.9	63.9	91.8
DRN_CRL_Coarse	yes	yes	no	no	no	no	82.8	61.1	91.8
NAVINFO_DLR	yes	yes	no	no	no	no	82.8	63.1	91.9
Gated-SCNN	yes	no	no	no	no	no	82.8	64.3	92.3
Valeo DAR Germany	yes	yes	no	no	no	no	82.8	62.9	92.0
DPC	yes	yes	no	no	no	no	82.7	63.3	92.0
SRC-B-MachineLearningLab	yes	yes	no	no	no	no	82.5	60.7	91.8

RelationNet_Coarse	yes	yes	no	no	no	no	82.4	61.9	91.8
SSMA	yes	yes	no	yes	no	no	82.3	62.3	91.5
GFF-Net	yes	no	no	no	no	no	82.3	62.1	92.0
DDAR	yes	yes	no	no	no	no	82.2	62.7	91.9

Showing 1 to 20 of 173 entries

Previous Next

Instance-Level Semantic Labeling Task

In the second Cityscapes task we focus on simultaneously detecting objects and segmenting them. This is an extension to both traditional object detection, since per-instance segments must be provided, and pixel-level semantic labeling, since each instance is treated as a separate label. Therefore, algorithms are required to deliver a set of detections of traffic participants in the scene, each associated with a confidence score and a per-instance segmentation mask.

Metrics

To assess instance-level performance, we compute the average precision on the region level (AP [2]) for each class and average it across a range of overlap thresholds to avoid a bias towards a specific value. Specifically, we follow [3] and use 10 different overlaps ranging from 0.5 to 0.95 in steps of 0.05. The overlap is computed at the region level, making it equivalent to the IoU of a single instance. We penalize multiple predictions of the same ground truth instance as false positives. To obtain a single, easy to compare compound score, we report the mean average precision AP, obtained by also averaging over the class label set. As minor scores, we add $AP^{50\%}$ for an overlap value of 50 %, as well as AP^{100m} and AP^{50m} where the evaluation is restricted to objects within 100 m and 50 m distance, respectively.

Results

Detailed results

Detailed results including performances regarding individual classes can be found here.

Usage

Use the buttons in the first row to hide columns or to export the visible data to various formats. Use the widgets in the second row to filter the table by selecting values of interest (multiple selections possible). Click the numeric columns for sorting.

Column visibility | CSV | Excel | PDF

name | fine

coarse | 16-bit | depth | video | sub | code

Show 20 entries Search:

name	fine	coarse	16-bit	depth	video	sub	AP	AP 50%	AP 100m	AP 50m
NJUST	yes	no	no	no	no	no	38.9	64.1	53.0	55.4
iFLYTEK-CV	yes	no	no	no	no	no	38.0	65.4	51.6	55.0
Sogou_MM	yes	no	no	no	no	no	37.2	64.5	51.1	54.5
PANet [COCO]	yes	no	no	no	no	no	36.4	63.1	49.2	51.8
NV-ADLR	yes	no	no	no	no	no	35.3	61.5	49.3	53.5
UPSNet	yes	no	no	no	no	no	33.0	59.6	46.8	50.7
BshapeNet+ [COCO]	yes	no	no	no	no	no	32.9	58.8	47.3	50.7
TCnet	yes	no	no	no	no	no	32.6	59.0	45.0	47.8
AdaptIS	yes	no	no	no	no	no	32.5	52.5	48.2	52.1
RUSH_ROB	yes	no	no	no	no	no	32.1	55.5	45.2	46.3
Mask R-CNN [COCO]	yes	no	no	no	no	no	32.0	58.1	45.8	49.5
PANet [fine-only]	yes	no	no	no	no	no	31.8	57.1	44.2	46.0
SegNet	yes	yes	no	no	no	no	29.5	55.6	43.2	45.8
Instance Segmentation by Jointly Optimizing Spatial Embeddings and Clustering Bandwidth	yes	no	no	no	no	no	27.7	50.9	37.8	37.3
GMIS: Graph Merge for Instance Segmentation	yes	yes	no	no	no	no	27.6	44.6	42.7	47.9
BshapeNet+ [fine-only]	yes	no	no	no	no	no	27.3	50.4	40.5	43.1
Mask R-CNN [fine-only]	yes	no	no	no	no	no	26.2	49.9	37.6	40.1
PolygonRNN++	yes	no	no	no	no	no	25.5	45.5	39.3	43.4

PolygonRNN++	yes	no	no	no	no	no	25.5	45.5	39.3	43.4
SGN	yes	yes	no	no	no	no	25.0	44.9	38.9	44.5
Deep Coloring	yes	no	no	no	no	no	24.9	46.2	39.0	44.0

Showing 1 to 20 of 40 entries Previous Next

Panoptic Semantic Labeling Task

The third Cityscapes task was added in 2019 and combines both, pixel-level and instance-level semantic labeling, in a single task called "panoptic segmentation". The challenge as well as the evaluation metrics are described in [4].

Results

Detailed results

Detailed results including performances regarding individual classes can be found here.

Usage

Use the buttons in the first row to hide columns or to export the visible data to various formats. Use the widgets in the second row to filter the table by selecting values of interest (multiple selections possible). Click the numeric columns for sorting.

Column visibility CSV Excel PDF

name fine coarse

16-bit depth video sub code

Show 20 entries Search:

name	fine	coarse	16-bit	depth	video	sub	PQ	SQ	RQ	Runtime [s]	coc
Seamless Scene Segmentation	yes	no	no	no	no	no	62.6	82.1	75.3	n/a	yes
TASCNet-enhanced	yes	no	no	no	no	no	60.7	81.0	73.8	n/a	no
Pixelwise Instance Segmentation with a Dynamically Instantiated Network	yes	yes	no	no	no	no	55.4	79.7	68.1	n/a	no
Sem2Ins	yes	no	no	no	no	no	52.3	78.9	65.2	n/a	no

HANet	yes	no	no	no	no	no	51.2	77.7	63.9	n/a	no

Showing 1 to 5 of 5 entries

Previous Next

Meta Information

In addition to the previously introduced measures, we report additional meta information for each method, such as timings or the kind of information each algorithm is using, e.g. depth data or multiple video frames. Please refer to the result tables for further details.

References

[1] M. Everingham, A. S. M. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The Pascal Visual Object Classes challenge: A retrospective,” *IJCV*, vol. 111, iss. 1, 2014.
[2] B. Hariharan, P. Arbeláez, R. B. Girshick, and J. Malik, ”Simultaneous detection and segmentation,” in *ECCV*, 2014.
[3] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and L. C. Zitnick, “Microsoft COCO: Common objects in context,” in *ECCV*, 2014.
[4] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, “Panoptic segmentation,” in *CVPR*, 2019.

News

Panoptic Segmentation May 12, 2019

Robust Vision Challenge February 17, 2018

CVPR 2016 Paper April 6, 2016

Evaluation Server Online March 27, 2016

Initial Semantic Labeling Results March 1, 2016

Search Website

Search

Contact

Cityscapes Team

Imprint / Impressum

Data Protection / Datenschutzhinweis