INFORMATIK

augmented VISION

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Seminar 3D Computer Vision & Augmented Reality
SS 2021
Survey on Data Sets for Semantic Segmentation (S13)

Adelya Kamaleeva

kamaleev@rhrk.uni-kl.de

Supervisor: René Schuster

Outline

- Motivation
- Semantic segmentation
- Methodology
- First classification of data sets
- Second classification of data sets
- Creation of class hierarchy
- Conclusion

Motivation

- **The aim of this work is**
  - to make the unification of class labels
  - to make classifications according to the common properties of the data sets.

SUN RGB-D

COCO Common Objects in Context

Mapillary

ADE20K

Semantic segmentation

- **Image segmentation can be formulated as a classification problem of pixels with semantic labels. Semantic segmentation performs labeling in the pixel-level (e.g. bus, human, table) for all pixels of the image.**

https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html

https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/shkf_eccv2012.pdf

https://www.cityscapes-dataset.com/

https://arxiv.org/pdf/1604.01685.pdf

Methodology for finding data sets

1.

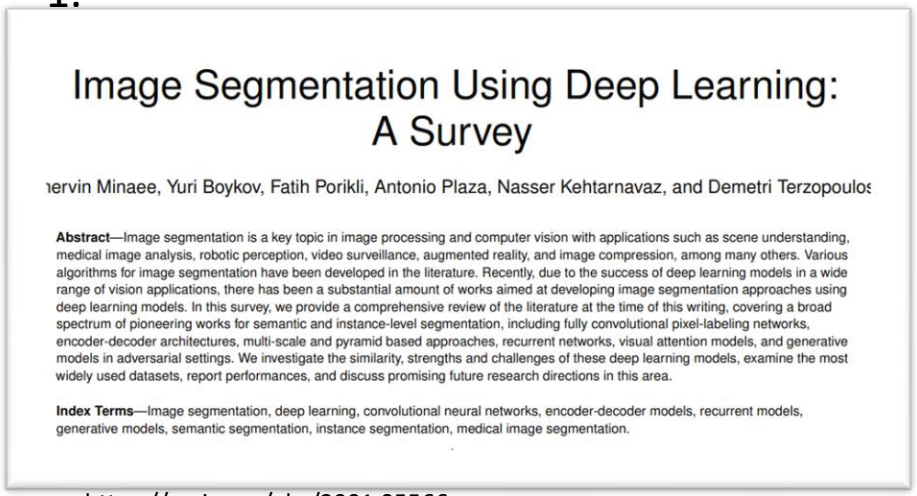


Image Segmentation Using Deep Learning: A Survey

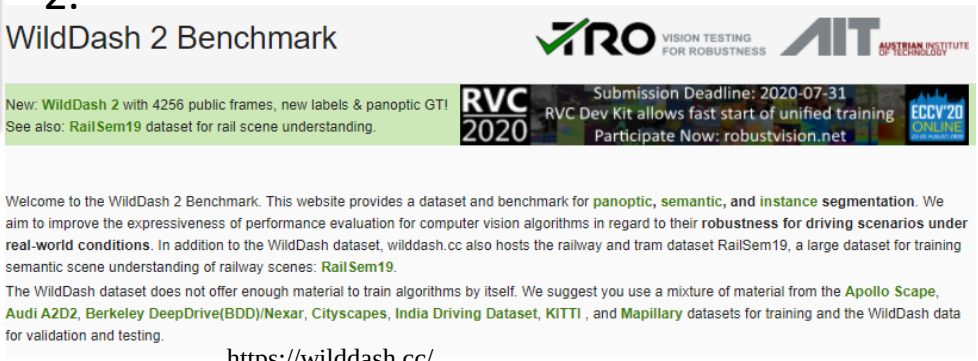
ıervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos

Abstract—Image segmentation is a key topic in image processing and computer vision with applications such as scene understanding, medical image analysis, robotic perception, video surveillance, augmented reality, and image compression, among many others. Various algorithms for image segmentation have been developed in the literature. Recently, due to the success of deep learning models in a wide range of vision applications, there has been a substantial amount of works aimed at developing image segmentation approaches using deep learning models. In this survey, we provide a comprehensive review of the literature at the time of this writing, covering a broad spectrum of pioneering works for semantic and instance-level segmentation, including fully convolutional pixel-labeling networks, encoder-decoder architectures, multi-scale and pyramid based approaches, recurrent networks, visual attention models, and generative models in adversarial settings. We investigate the similarity, strengths and challenges of these deep learning models, examine the most widely used datasets, report performances, and discuss promising future research directions in this area.

Index Terms—Image segmentation, deep learning, convolutional neural networks, encoder-decoder models, recurrent models, generative models, semantic segmentation, instance segmentation, medical image segmentation.

https://arxiv.org/abs/2001.05566

2.



WildDash 2 Benchmark

VRO VISION TESTING FOR ROBUSTNESS AIT AUSTRIAN INSTITUTE OF TECHNOLOGY

New: **WildDash 2** with 4256 public frames, new labels & panoptic GT!
See also: **RailSem19** dataset for rail scene understanding.

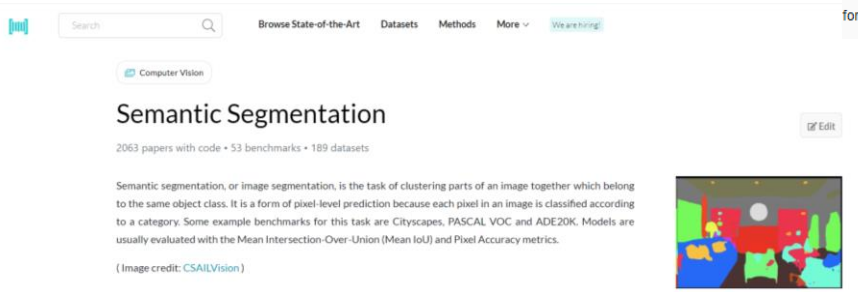
RVC 2020 Submission Deadline: 2020-07-31
RVC Dev Kit allows fast start of unified training
Participate Now: robustvision.net
ECCV'20 ONLINE

Welcome to the WildDash 2 Benchmark. This website provides a dataset and benchmark for **panoptic**, **semantic**, **and instance segmentation**. We aim to improve the expressiveness of performance evaluation for computer vision algorithms in regard to their **robustness for driving scenarios under real-world conditions**. In addition to the WildDash dataset, wilddash.cc also hosts the railway and tram dataset RailSem19, a large dataset for training semantic scene understanding of railway scenes: **RailSem19**.

The WildDash dataset does not offer enough material to train algorithms by itself. We suggest you use a mixture of material from the **Apollo Scape**, **Audi A2D2**, **Berkeley DeepDrive(BDD)/Nexar**, **Cityscapes**, **India Driving Dataset**, **KITTI** , and **Mapillary** datasets for training and the WildDash data for validation and testing.

https://wilddash.cc/

3.



Search

Browse State-of-the-Art Datasets Methods More We are hiring!

Computer Vision

Semantic Segmentation

Edit

2063 papers with code • 53 benchmarks • 189 datasets

Semantic segmentation, or image segmentation, is the task of clustering parts of an image together which belong to the same object class. It is a form of pixel-level prediction because each pixel in an image is classified according to a category. Some example benchmarks for this task are Cityscapes, PASCAL VOC and ADE20K. Models are usually evaluated with the Mean Intersection-Over-Union (Mean IoU) and Pixel Accuracy metrics.

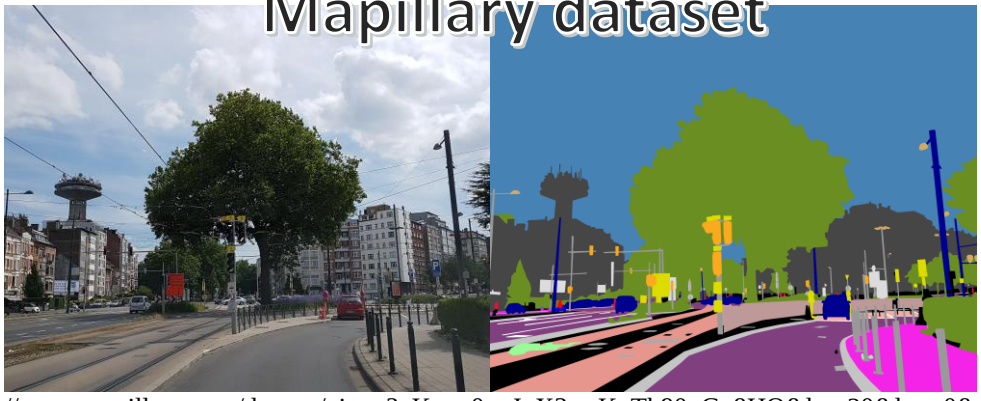
(Image credit: CSAILVision)

https://paperswithcode.com/task/semantic-segmentation

First classification

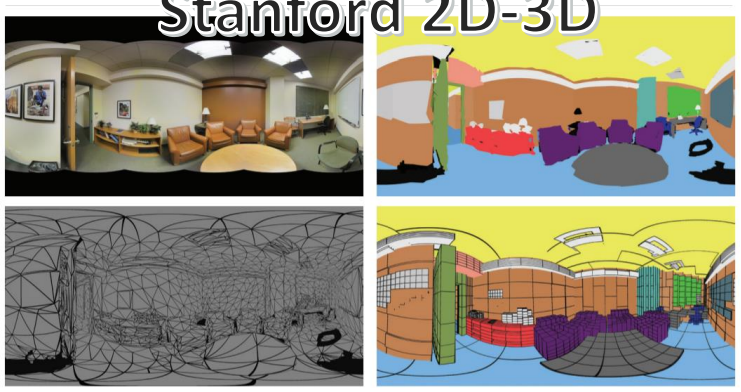
- **First classification is based on place where photographs were taken.**

Mapillary dataset



https://www.mapillary.com/dataset/vistas?pKey=0_xJqX3-c-KyTb90oG_8HQ&lat=20&lng=0&z=1.5

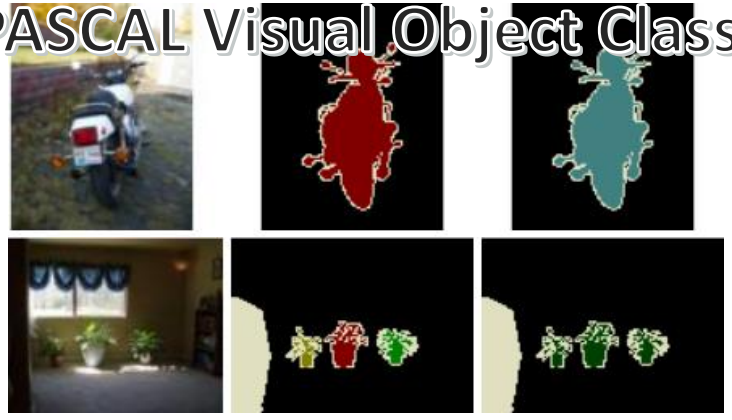
Stanford 2D-3D



https://cvgl.stanford.edu/resources.html

https://arxiv.org/pdf/1702.01105.pdf

PASCAL Visual Object Classes (VOC)



https://homepages.inf.ed.ac.uk/ckiw/postscript/ijcv_voc09.pdf

Road datasets

	Name of dataset	Year	№ of classes	Resolution (pixels)	Segmentation type	№ of images
1	**CamVid**	2008	30	960 × 720	Semantic	700
2	**DUS**	2014	5	1024 × 440	Semantic	5 000
3	**Cityscapes**	2016	30	2048 × 1024	Semantic, instance	15 000
4	**Mapillary (Mapillary Vistas Dataset)**	2017	124	1920 × 1080	Semantic	5 000
5	**SYNTHIA**	2017	13	1280 × 960	Semantic	2 224
6	**Playing for Benchmarks**	2017	23	1920 × 1080	Semantic, instance	254 064
7	**KITTI**	2018	8	1392 × 512	Pixel, instance	400
8	**WildDash 2**	2018	28	896 × 896	Semantic, instance	4 256
9	**ApolloScape**	2018	36	3384 × 2710	Semantic, instance	146 997
10	**India Driving Dataset**	2019	34	720 × 720, 1080 × 1080	Semantic, instance	10 000
11	**BDD100K**	2020	19	720 × 720	Semantic, instance	10 000
12	**A2D2: Audi Autonomous Driving Dataset**	2020	38	1920 × 1208	Semantic	41 277

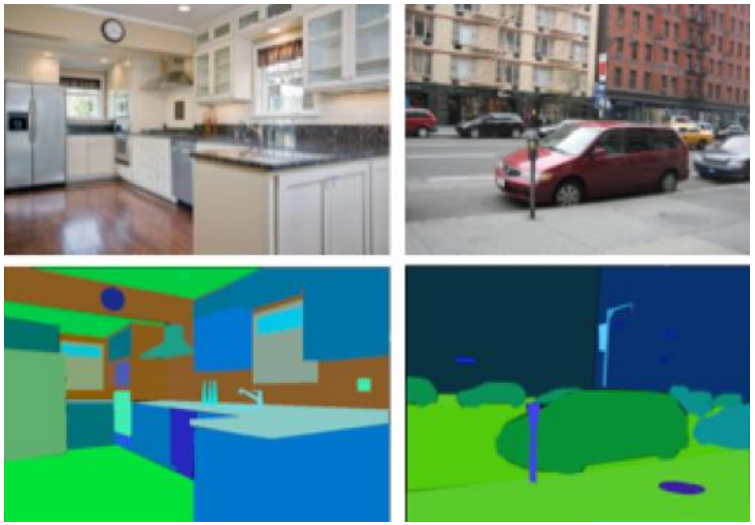
Indoor datasets

	Name of dataset	Year	№ of classes	Resolution (pixels)	Segmentation type	№ of images
13	**NYU-D V2**	2012	13	640 × 480	Semantic	408 473
14	**SUN-3D**	2013	29	640 × 480	Semantic	
15	**SUN RGB-D**	2015	29	1920 × 108, 640 × 480	Semantic	10 335
16	**Stanford 2D-3D**	2017	12	1080 × 1080	Semantic	70 000

https://rgbd.cs.princeton.edu/ https://rgbd.cs.princeton.edu/paper.pdf

Objects (indoor + outdoor) datasets

	Name of dataset	Year	№ of classes	Resolution (pixels)	Segmentation type	№ of images
17	**PASCAL Visual Object Classes (VOC)**	2012	21	650 × 590	Semantic, instance	9 993
18	**Microsoft Common Objects in Context (MS COCO)**	2017	80	640 × 480	Instance	123 287
19	**ADE20K /MIT Scene Parsing (SceneParse150)**	2017	150	640 × 480	Instance, part	25 574



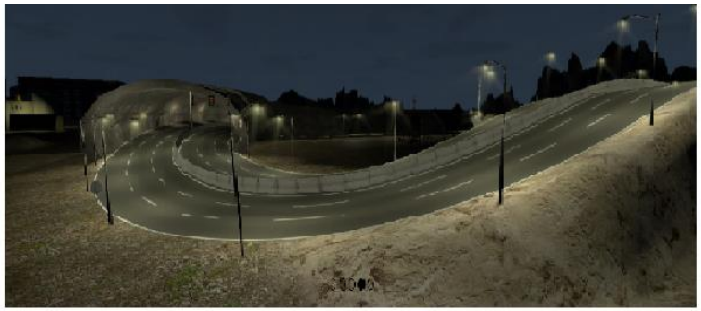
https://groups.csail.mit.edu/vision/datasets/ADE20K/

https://people.csail.mit.edu/bzhou/publication/scene-parse-camera-ready.pdf

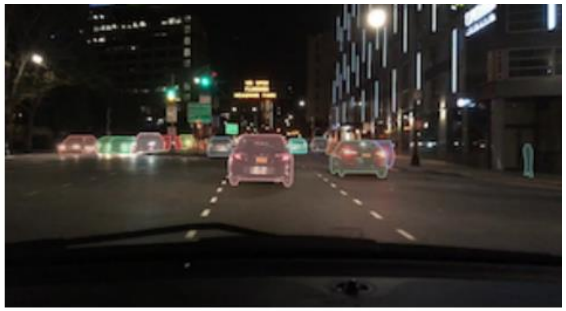
Second classification

	day	night	dawn	dusk	spring	summer	fall	sunny	rainy	snowy	cloudy	foggy	various weather
KITTI	+												
CamVid	+												
DUS	+												
India Driving Dataset	+												
ApolloScape	+												+
BDD100K	+												+
A2D2	+												+
SYNTHIA	+												+
Cityscapes	+				+	+	+						
Mapillary	+	+	+	+				+	+		+	+	
WildDash 2	+			+					+	+			
Playing for Benchmarks	+	+		+					+	+			

https://wilddash.cc/



https://synthia-dataset.net/



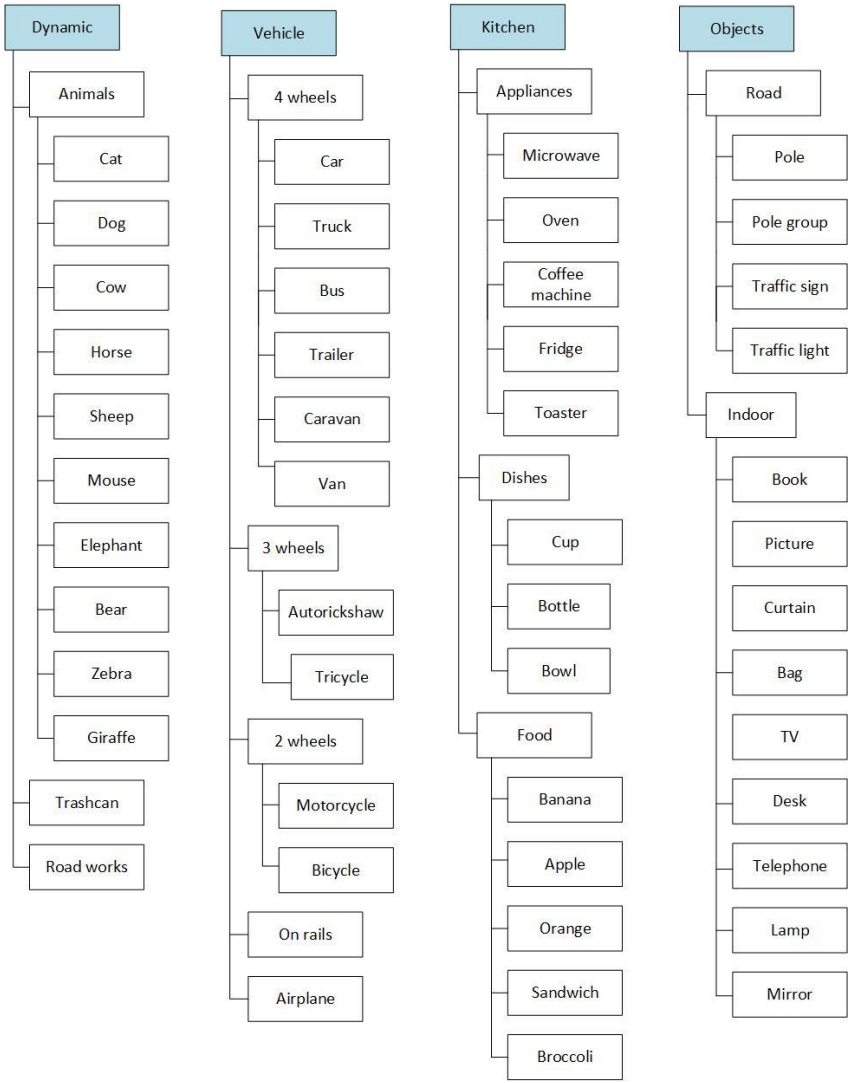
https://www.bdd100k.com/

Creation of class hierarchy

	City scapes [12]	KITTI [14]	Cam Vid [10]	MVD [20]	DUS [23]	Wild Dash 2 [9]	Apollo Scape [16]	IDD [25]	BDD 100K [27]	A2D2 [15]	Synthia [22]	PfB [21]	Stanford 2D-3D [11]	NYUD V2 [19]	SUN RGBD [24]	SUN 3D [26]	Pascal VOC [13]	MS COCO [17]	ADE 20K [28]	Res
car	+	+	+	+	+	+	+	+	+	+	+	+					+	+	+	15
truck	+	+	+	+	+	+	+	+	+	+	+	+						+	+	14
person	+	+	+	+	+	+	+		+		+	+			+		+	+	+	14
traffic light	+		+	+		+	+	+	+	+	+	+						+	+	12
bus	+		+	+		+	+	+	+		+	+					+	+	+	12
motorcycle	+		+	+		+	+	+	+		+	+					+	+	+	12
sky	+		+	+	+	+	+		+	+	+	+							+	11
bicycle	+			+		+	+	+	+	+	+						+	+	+	11
rider	+	+	+	+	+	+	+	+	+		+									10
building	+		+	+	+	+	+		+		+	+							+	10
road	+		+	+		+	+		+	+	+	+							+	10
sidewalk	+		+	+		+	+		+	+	+	+							+	10
traffic sign	+			+		+	+	+	+	+	+	+						+		10
on rails	+	+	+	+		+		+	+		+									
fence	+		+	+		+	+		+		+									
pole	+		+	+		+	+		+	+	+									
vegetation	+		+	+		+	+		+		+									
wall	+		+	+		+	+		+		+									
animal			+	+				+		+										
terrain	+			+		+			+		+									
bridge	+		+	+		+	+													
vehicle				+	+			+		+										

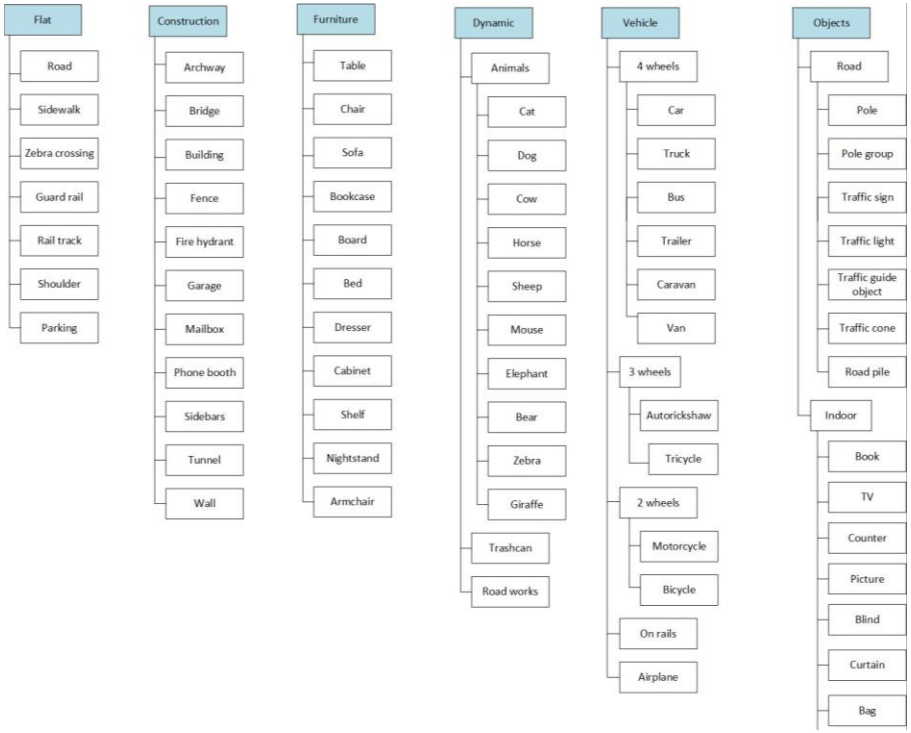
chair													+	+	+	+	+	+	+	7
sofa													+	+	+	+	+	+	+	7
table													+	+	+		+	+	+	6
TV														+	+	+	+	+	+	6
ceiling													+	+	+	+			+	5
floor													+	+	+	+			+	5
wall													+	+	+	+			+	5
bed														+	+	+		+	+	5
window													+	+	+				+	4
door													+		+	+			+	4
book														+	+			+	+	4
counter														+	+	+			+	4
bag															+	+		+	+	4
column													+	+					+	3
picture															+	+			+	3
curtain															+	+			+	3
pillow															+	+			+	3
mirror															+	+			+	3
fridge															+			+	+	3
sink																+		+	+	3
toilet																+		+	+	3
towel															+	+			+	3
bottle																	+	+	+	3

Creation of class hierarchy



Conclusion

Dataset	Year	№ of classes	Resolution (pixels)	Segmentation type	№ of images	Category
CamVid [10]	2008	30	960 × 720	Semantic	700	road
DUS [23]	2014	5	1024 × 440	Semantic	5 000	road
Cityscapes [12]	2016	30	2048 × 1024	Semantic, instance	15 000	road
Mapillary (Mapillary Vistas Dataset, MVD) [20]	2017	124	1920 × 1080	Semantic	5 000	road
SYNTHIA [22]	2017	13	1280 × 960	Semantic	2 224	road
Playing for Benchmarks (PfB) [21]	2017	23	1920 × 1080	Semantic, instance	254 064	road
KITTI [14]	2018	8	1392 × 512	Pixel, instance	400	road
WildDash 2 [9]	2018	28	896 × 896	Semantic, instance	4 256	road
ApolloScape [16]	2018	36	3384 × 2710	Semantic, instance	146 997	road
India Driving Dataset [25]	2018	34	720 × 720, 1080 × 1080	Semantic, instance	10 000	road
BDD100K [27]	2020	19	720 × 720	Semantic, instance	10 000	road
A2D2: Audi Autonomous Driving Dataset [15]	2020	38	1920 × 1208	Semantic	41 277	road
NYU-D V2 [19]	2012	13	640 × 480	Semantic	408 473	indoor
SUN-3D [26]	2013	29	640 × 480	Semantic		indoor
SUN RGB-D [24]	2015	29	1920 × 108, 640 × 480	Semantic	10 335	indoor
Stanford 2D-3D [11]	2017	12	1080 × 1080	Semantic	70 000	indoor
PASCAL Visual Object Classes (VOC) [13]	2012	21	650 × 590	Semantic, instance	9 993	object
Microsoft Common Objects in Context (MS COCO) [17]	2017	80	640 × 480	Instance	123 287	object
ADE20K /MIT Scene Parsing (SceneParse150) [28]	2017	150	640 × 480	Instance, part	25 574	object



References for road data sets

- 1. http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/
- 2. https://paperswithcode.com/dataset/dus
- 3. https://arxiv.org/pdf/1604.01685.pdf; https://www.cityscapes-dataset.com/
- 4. https://www.mapillary.com/dataset/vistas?pKey=0_xJqX3-c-KyTb90oG_8HQ&lat=20&lng=0&z=1.5
- 5. https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Ros_The_SYNTHIA_Dataset_CVPR_2016_paper.pdf; https://synthia-dataset.net/
- 6. https://playing-for-benchmarks.org/overview/
- 7.http://www.cvlibs.net/publications/Geiger2013IJRR.pdf; http://www.cvlibs.net/datasets/kitti/index.php
- 8. https://wilddash.cc/
- 9. http://apolloscape.auto/scene.html
- 10. https://idd.insaan.iiit.ac.in/
- 11. https://arxiv.org/pdf/1805.04687.pdf; https://www.bdd100k.com/
- 12. https://arxiv.org/pdf/2004.06320.pdf; https://www.a2d2.audi/a2d2/en/dataset.html

References for indoor data sets

- 13.https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/shkf_eccv2012.pdf; https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html
- 14.https://vision.cs.princeton.edu/projects/2013/SUN3D/paper.pdf; http://sun3d.cs.princeton.edu/
- 15. https://rgbd.cs.princeton.edu/paper.pdf; https://rgbd.cs.princeton.edu/
- 16.https://arxiv.org/pdf/1702.01105.pdf; https://cvgl.stanford.edu/resources.html

References for object data sets

- 17. https://homepages.inf.ed.ac.uk/ckiw/postscript/ijcv_voc09.pdf; http://host.robots.ox.ac.uk/pascal/VOC/
- 18. https://arxiv.org/pdf/1405.0312.pdf; https://cocodataset.org/#home
- 19. https://people.csail.mit.edu/bzhou/publication/scene-parse-camera-ready.pdf; https://groups.csail.mit.edu/vision/datasets/ADE20K/

Thank you for attention!

Do you have any questions?