PAPER ANALYSIS

Presented by Yannis He

-

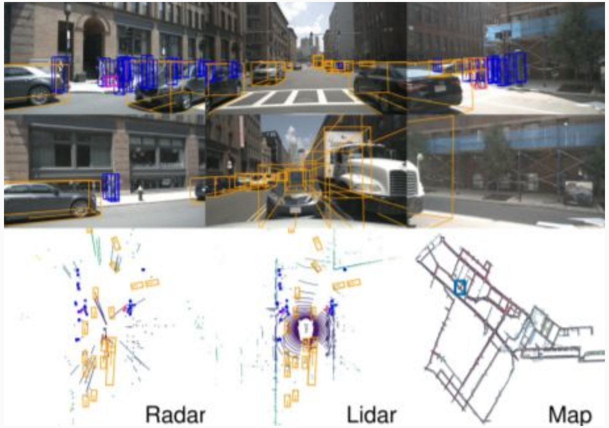
Paper: **nuScenes: A multimodal dataset for autonomous driving**

Authors: Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, Oscar Beijbom

https://arxiv.org/pdf/1903.11027.pdf

DATASET INTRODUCTION

- First dataset to carry the full autonomous vehicle sensor suite:
    - 6 cameras, 5 radars, 1 lidar
    - All with 360 degree field of view.
- Contents:
    - 1000 scenes, each 20s long and fully annotated with 3D bounding boxes for 23 classes and 8 attributes
    - 7 times more annotations and 100 times more images than KITTI dataset
- Full released in March 2019
- Purpose of multimodal:
    - Cameras: accurate measurements of edges, color, light
        - Good at classifications and localization on the image plane
        - Not good at 3D localization
    - Lidar: less semantic information but highly accurate localization in 3D
        - Data is sparse
            - Range of 50-150m
    - Radar: measure object velocity through Doppler effect
        - Data sparser than lidar
        - Less precise in terms of localization
        - Range of 200-300m
- Only dataset with radar and night/rain data included
- Recorded at Boston (Seaport and South Boston) and Singapore (SG)



Presented by Yannis He

OTHER DATASET IN THIS FIELD

Some other dataset in this field:

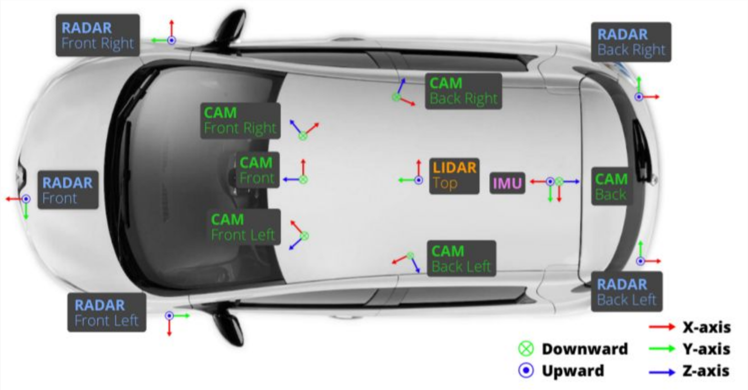
Dataset	Year	Sce-nes	Size (hr)	RGB imgs	PCs lidar††	PCs radar	Ann. frames	3D boxes	Night / Rain	Map layers	Clas-ses	Locations
CamVid [8]	2008	4	0.4	18k	0	0	700	0	No/No	0	32	Cambridge
Cityscapes [19]	2016	n/a	-	25k	0	0	25k	0	No/No	0	30	50 cities
Vistas [33]	2017	n/a	-	25k	0	0	25k	0	Yes/Yes	0	152	Global
BDD100K [85]	2017	100k	1k	100M	0	0	100k	0	Yes/Yes	0	10	NY, SF
ApolloScape [41]	2018	-	100	144k	0**	0	144k	70k	Yes/No	0	8-35	4x China
D^2-City [11]	2019	1k†	-	700k†	0	0	700k†	0	No/Yes	0	12	5x China
KITTI [32]	2012	22	1.5	15k	15k	0	15k	200k	No/No	0	8	Karlsruhe
AS lidar [54]	2018	-	2	0	20k	0	20k	475k	-/-	0	6	China
KAIST [17]	2018	-	-	8.9k	8.9k	0	8.9k	0	**Yes**/No	0	3	Seoul
H3D [61]	2019	160	0.77	83k	27k	0	27k	1.1M	No/No	0	8	SF
nuScenes	2019	**1k**	5.5	**1.4M**	**400k**	**1.3M**	**40k**	1.4M	**Yes/Yes**	**11**	**23**	Boston, SG
Argoverse [10]	2019	113†	0.6†	490k†	44k	0	22k†	993k†	**Yes/Yes**	2	15	Miami, PT
Lyft L5 [45]	2019	366	2.5	323k	46k	0	**46k**	1.3M	No/No	7	9	Palo Alto
Waymo Open [76]	2019	**1k**	5.5	1M	200k	0	**200k**‡	**12M**‡	**Yes/Yes**	0	4	3x USA
A*3D [62]	2019	n/a	**55**	39k	39k	0	**39k**	230k	**Yes/Yes**	0	7	SG
A2D2 [34]	2019	n/a	-	-	-	0	12k	-	-/-	0	14	3x Germany

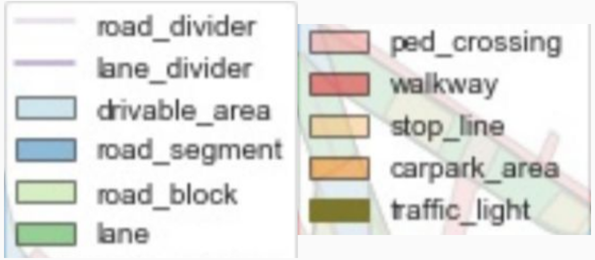
SETUP

- Renault Zoe supermini electric cars
- Localization are created with detailed HD map of lidar points in an offline step and collected with Monte Carlo from lidar odometry
  - Localization errors of < 10cm
- Maps are human-annotated
  - Original rasterized maps: two semantic: roads & sidewalk
  - Vectorized map expansion: 11 semantic classes
- 23 categories
- Around 7 pedestrians and 20 vehicles per keyframe on average

car
adult
barrier
trafficcone
truck
trailer
push/pullable
constr. veh.
bus.rigid
motorcycle
bicycle
worker
debris
bicycle racks
child
bus.bendy
stroller
animal
police
police car
wheelchair
p.mobility
ambulance

Annotation categories





Semantic segmentation label

SOME MEASUREMENTS

- Metrics:
  - Average Precision (AP)
  - True Positive Metrics (TP metrics)
  - Average Multi Object Tracking Accuracy (AMOTA)
  - Average Multi Object Tracking Precision (AMOTP)
  - nuScences Detection Score (NDS)
  - Tracking Initialization Duration (TID)
  - Longest Gap Duration (LGD)
- Which sensor is more important
  - PointPillars vs MonoDIS:
    - mAP: 30.5% vs 30.4%
    - NDS: 45.3% vs 38.4%
- Findings:
  - Importance of pre-training
  - Better detections gives better tracking