Object Detection for Autonomous Driving

Application Cases, Metrics and Data Augmentation

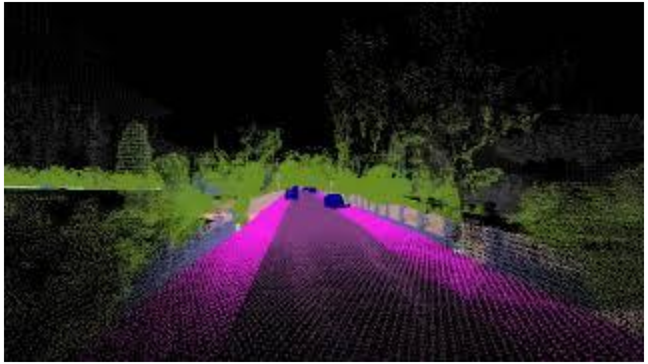
Alina Griesel, Elisa Hagensieker, Stella Hoyos Trueba, Madleen Uecker

Use Case - Object Detection for Autonomous Driving

Importance:
- Safety: ensure safe navigation and prevent collision
- Traffic efficiency: optimize traffic flow
- Regulatory compliance: adheres to traffic laws

Problems of natural data:
- Data privacy: Faces, licence plates
- Cost and time: Data collection and annotation
- Diversity and coverage: Capture all driving scenarios

Synthetic data to our help:
- Privacy friendly: no real individuals involved
- Cost effective
- Controlled environment
- Scalability



Datasets

- datasets: nuScene, Kitti, ScanNet, Waymo, S3DISCityscapes, Argoverse, CARLA, Synscapes

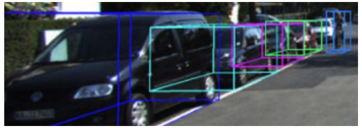
Name	Link	License	Year	3D point cloud segmentation	3D object detection & tracking	3D drivable area	2D object detection & tracking	2D segmention	2D freespace	2D drivable area	2D lane markings	Plant Classification	Motion forcasting	Image / Other Annotation Format	Lidar Annotation Format	Relevance	Scene Type	RGB	RGB-D	Lidar	Radar	FLIR / NIR	GPS / IMU	Maps	Details
DeepScene (Freiburg Forest)	http://deepscene.cs	non-commercial	2016	x										?	?	high	forest	x	x			x			Bumblebee2 Stereo
FieldSAFE	https://vision.eng.a	non-commercial	2016		x		x							-	map based + object coordinates	high	grass field	x	x	x	x	x	x		Velodynce HDL-32E, Delphi ESR Radar, Flir A65,
NREC Human Detection and T	https://www.nrec.ri.	non-commercial	2017		x		x							Pascal VOC	-	high	off-road (apple/orange field)	x	x				x		
OFFSED	http://www.dfki.uni-	non-commercial	2021	x										CVAT rgb/png files	-	high	od, farmland, construction si	x	x						Stereolabs ZED Camera, some instances labelec
OPEDD	http://www.dfki.uni-	non-commercial	2021		x									VIA json files	-	high	od, farmland, construction si	x	x						Stereolabs ZED Camera
RELLIS-3D	https://unmannedla	non-commercial	2020	x				x						rgb/png files	SemanticKITTI (.label files)	high	off-road	x	x	x			x		Ouster OS1, Velodyne Ultra-Puck 32, Karmin 2 Ste
RUGD	http://rugd.vision/	unknown	2019					x						rgb/png files	-	high	off-road	x		x			x		Velodyne HDL-32E, Proscilia 6 MP camera
SemanticUSL	https://unmannedla	non-commercial	2020	x										-	SemanticKITTI (.label files)	high	campus & off-road			x					Ouster OS1-64 Lidar
SugarBeets	http://www.ipb.uni-t	public	2016									x		?	?	high	sugar beets / field	x	x	x			x		2x Velodyne VLP-16, Camera JAI AD-130GE, Kine
YCOR		unknown	?	x										?	?	high	off-road	x							
KIT MOMA		unknown	2016				x							?	?	mid	construction sides	x							
Marulan	http://sdi.acfr.usyd.e	unknown	2009		x		x							-	-	mid	dust, smoke, rain	x		x	x	x	x		Sick LaserStarboard/Port, FMCW Radar, Raytheon
Rosario	https://www.cifasis-	unknown	2019										x	3D position GT	-	mid	soybean field		x				x		ZED Stereo Camera, LSM6DS0 6-DoF IMU
SemanticKITTI	http://semantic-kitti.	non-commercial	2019	x										-	SemanticKITTI (.label files)	mid	urban			x					Velodyne HDL-64E
RAGE	https://download.vis	unknown	2016					x	x							mid	urban simulator	x							Simulation based semantic labels
DALES	https://udayton.edu	non-commercial	2020	x												none	arial scans	x		x					airborn laser scanner
IQmulus	http://data.ign.fr/ber	non-commercial	2015	x												none	urban road scans			x					MLS (3d mobile laser scanner)
ISPRS	https://www2.isprs.	unknown	2012	x												none	arial scans			x					airborn laser scanner
Oakland 3-D Point Cloud	https://www.cs.cmu	unknown	2009	x												none	urban road scans			x					Sick LMS Laser
Paris-Lille-3D	https://npm3d.fr/pa	non-commercial	2018	x												none	urban road scans			x					Velodyne HDL-32E
Paris-rue-Madame	http://www.cmm.mi	non-commercial	2014	x												none	urban road scans			x					MLS (3d mobile laser scanner)
S3DIS		unknown	2017	x												none		x		x					
ScanNet	http://www.scan-ne	unknown	2017	x												none	indoor	x	x						
ScanNetV2	http://www.scan-ne	unknown	2018		x											none	indoor		x						
Semantic3D	https://www.seman	non-commercial	2017	x												none	urban / rural scans	x		x					Terrestrial Laser Scanner
SUN RGB-D	https://rgbd.cs.prin	unknown	2015		x											none	indoor		x						
Toronto-3D	https://github.com/V	unknown	2020	x												none	urban road scans			x					MLS (3d mobile laser scanner)
Drive&Act	https://www.drivear	non-commercial	2019													none	driver seat	x	x			x			
A*3D	https://github.com/I	non-commercial	2020		x											unknown	urban	x		x					Velodyne HDL-64E, 2x PointGrey Chameleon2 ca
ApolloScape	http://apolloscape.a	non-commercial	2018				x	x			x		x			unknown	urban	x	x						
Argoverse	https://www.argove	non-commercial	2019		x					x			x			unknown	urban	x	x	x			x	x	Velodyne Pucks, Stereo and Mono Cameras, HD-
BDK100K	https://bair.berkeley	unknown	2020				x	x		x	x					unknown	urban	x					x		HD video
Cityscape	https://www.citysca	non-commercial		x				x								unknown	urban	x	x						
Cityscape 3D	https://www.citysca	non-commercial	2020		x											unknown	urban	x	x	x			x		
Ford Autonomous Vehicle Dat	https://avdata.ford.c	university only	2020													unknown	urban	x		x			x		4x Velofyne HDL-32E, 6 Point Grey 1.3 MP Camer

1: Dosovitskiy et al.
2: Geiger et al.
3: Sun et al
4: Gaidon et al.

Overview Datasets

Dataset	KITTI[2]	Virtual KITTI[4]	nuScene	Waymo[3]	CARLA[1]
Characteristic	LiDAR and camera data	unity game engine		LiDAR and camera data	open-source simulator
Size	object detection dataset: 7481 training & 7518 test img, total: 80.256 labeled objects	50 high-resolution monocular videos (21,260 frames)		1150 scenes, each 20 sec.	
License	CC BY-NC-SA 3.0	CC BY-NC-SA 3.0		non-commercial	CC-BY / MIT
Real / Synthetic	real	synthetic		real	synthetic
Example					

Overview Datasets

Dataset	KITTI[2] human annotators	Virtual KITTI[4] automatically labeled (unity)	Waymo[3] human annotators
Characteristic	LiDAR and camera data	unity game engine	LiDAR and camera data
Size	object detection dataset: 7481 training & 7518 test img, total: 80.256 labeled objects	50 high-resolution monocular videos (21,260 frames)	1150 scenes, each 20 sec.
Classes	8 classes: number of instances (training data): 28742 car, 4487 pedestrian, 2914 van, 11627 cyclist, 1094 truck	1 class: car (main category of KITTI)	4 classes: mean count of instances per class: 30 vehicle, 14 pedestrian, 0 cyclists
License	CC BY-NC-SA 3.0	CC BY-NC-SA 3.0	non-commercial
Real / Synthetic	real,	synthetic	real
Example			

Overview Datasets - Waymo

Object Class	Mean	Median	Max
Vehicle	30	27	163
Pedestrian	14	27	192
Cyclist	0	0	11

1 (TYPE_VEHICLE)
1 (TYPE_VEHICLE)
2 (TYPE_PEDESTRIAN)
2 (TYPE_PEDESTRIAN)
2 (TYPE_PEDESTRIAN)
1 (TYPE_VEHICLE)
2 (TYPE_PEDESTRIAN)

Datasets

KITTI https://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=3d

Data collection and privacy

- Funding: KIT, TTI-C
- Privacy: academic use only (registration required, Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License)

- Footprint: equipped with Radar, LiDAR, camera data
- classes: building, tree, sky, car, sign, road, pedestrian, fence, pole, sidewalk, bicyclist
- 73.7km driving distance
- 7481 training images; 7519 test images (80256 labeled objects)
-

Datasets
Virtual KITTI

Data generation:

- Unity game engine with 5 different virtual worlds under different lightning and weather conditions
- Creative COmmons Attribution-NonCommercial-ShareAlike 3.0 License - restrictions on commercial use and distribution
- corresponds to real KITTI scenes
- “measuring the real-to-virtual gap, deep learning with virtual data, and measuring the generalization performance under changes in imaging and weather conditions”

https://github.com/VisualComputingInstitute/vkitti3D-dataset/blob/master/tools/download_raw_vkitti.sh (try out for download)

- Radar, LiDAR, camera data
- classes: building, tree, sky, car, sign, road, pedestrian, fence, pole, sidewalk, bicyclist
- 7481 training images; 7519 test images (80256 labeled objects)

Datasets
Waymo

- tfds.load('waymo_open_dataset/v1.0',data_dir='gs://waymo_open_dataset_v_1_0_0_individual_files/tensorflow_datasets')
- Creative COmmons Attribution-NonCommercial-ShareAlike 3.0 License - restrictions on commercial use and distribution; registration required for download
- objects in motion: vehicle, pedestrians, cyclists and more
- Footprint: LiDAR, Camera with annotations for scene understanding in 2D and 3D
-

Metrics

- Mean Average Precision
- Intersection over Union
- Standard accuracy measures

Virtual KITTI: MSE and Edge-Aware Smoothing loss
(https://arxiv.org/pdf/2006.04080v2)

Metrics for Object Detection
mean Average Precision (mAP)

- Single number metric (value ∈ [0,1])
- overall accuracy of object detector
- good way to evaluate models performance and to compare with other models
- metrics underlying the mAP:
  - Confusion matrix
  - intersection over union (IoU)
  - Precision
  - Recall

choose one class

set IoU threshold

determine TP - FP - FN - TN

calculate precision and recall

plot precision-recall curve

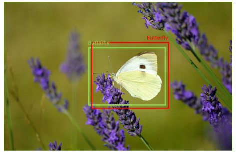
calculate AP

calculate mean of AP over all thresholds

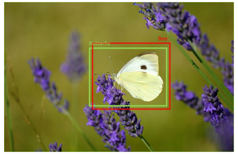
calculate mean of mAP over all classes

Metrics
mAP - explained in more Detail

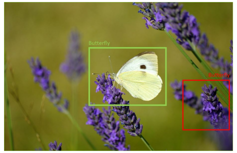
Prediction Types:



True Positive (**TP**):
- bbx aligns with gt
- label is correct



False Positive (**FP**):
- bbx aligns with gt
- label is incorrect



False Negative (**FN**):
- object detected where there is none

True Negative (**TN**):
- irrelevant here

$$\mathrm{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Intersection}}$$

- User sets threshold for IoU
- value ∈ [0,1]

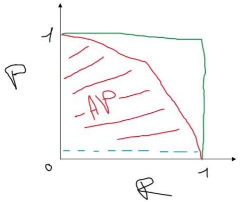
$$\text{precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$$

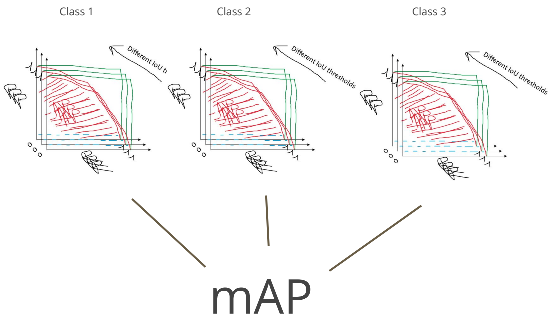
- how many of the predicted positives are TP ?

$$\text{recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$$

- have we detected all TPs ?

Precision-Recall Curve:



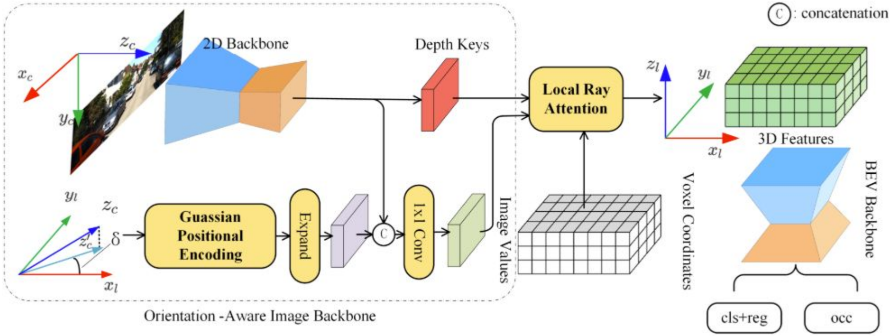


CIE

Task: Object detection & 3D representation

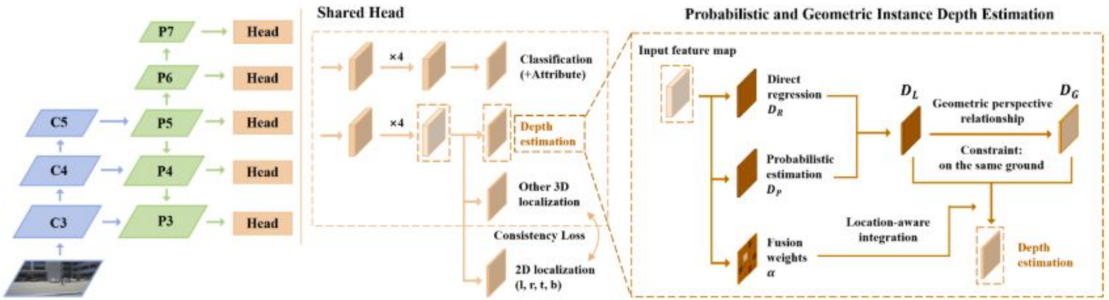
1. 2D Backbone: V2-99 extended to Feature Pyramid Network generating two feature maps:
   a. depth keys
   b. image features
2. Orientation-Aware Image Backbone:
3. Local Ray Attention Mechanism: image feature maps to 3D voxel features without point clouds
4. BEV backbone: predicts 3D occupancy map

➔ 31.55% AP



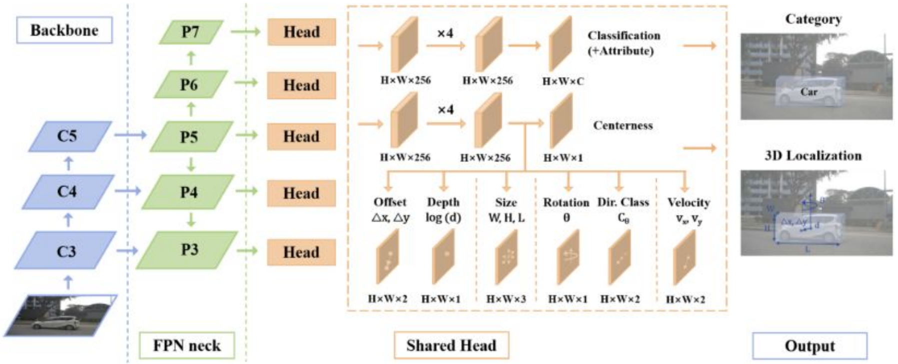
Probabilistic and Geometric Depth

https://arxiv.org/pdf/2107.14160

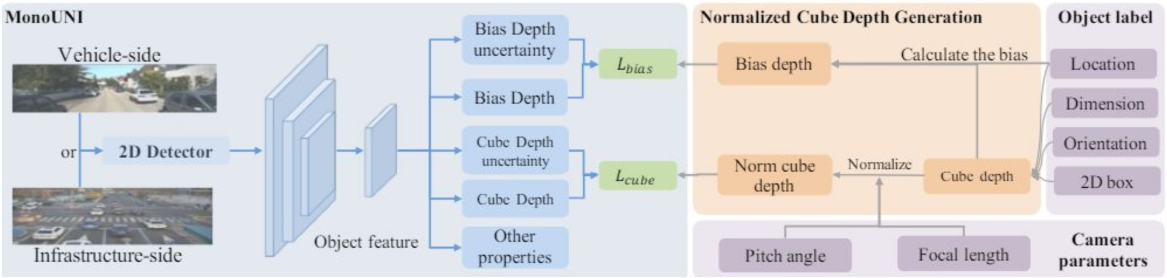


Fully Convolutional One-Stage Monocular 3D Object Detection

https://arxiv.org/pdf/2104.10956



Monocular Unified 3D Object Detection - MonoUNI



Task: Monocular 3D detection

1. BaseModel: CenterNet generates discriminative representations
2. Backbone: DLA34 for feature extraction
   - Deep Layer Aggregation for hierarchical features
3. Network Heads: prediction of various object properties:
   - category
   - 2D bounding box
   - 3D offset
   - dimension of objects
   - orientation3D normalized cube depth
   - bias depth and depth uncertainty

GANs for mixed datasets

Pros	Cons
Privacy Preservation: ● No identifiable information (i.e. faces, license plates)	Reality Gap: ● mimic complexity and variability
Scalability: ● once created it can generate unlimited amount of data	Diversity and Variability: ● capture rare cases crucial for robust object detection
Controlled Environment: ● Control various aspects of generated data (i.e. weather conditions, traffic scenarios)	Semantic Understanding: ● essential to ensure that generated scenes are semantically meaningful (i.e. spatial relationship)
	Ethical and Safety Concerns: ● additional complexities and uncertainties with the use of synthetic data

GAN Architectures:

- Deep Convolutional GAN: Convolutional layers in generator and discriminator networks to generate high-resolution images

Characteristics:

- Conditional GANs: Generate synthetic images or traffic scenarios conditioned on different weather conditions or road layouts
- Self-Attention GAN: focus on relevant spatial information of the input to keep semantic understanding of spatial relationship

THANK YOU FOR YOUR ATTENTION!

Sources

- Lightning NeRF: Efficient Hybrid Scene Representation for Autonomous Driving https://arxiv.org/pdf/2403.05907
- Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A. & Koltun, V.. (2017). CARLA: An Open Urban Driving Simulator. <i>Proceedings of the 1st Annual Conference on Robot Learning</i>, in <i>Proceedings of Machine Learning Research</i> 78:1-16 Available from https://proceedings.mlr.press/v78/dosovitskiy17a.html.
- Geiger, A., Lenz, P., & Urtasun, R. (2012, June). Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition* (pp. 3354-3361). IEEE.
    - different citations needed for different KITTI stuff !!!
- Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., ... & Anguelov, D. (2020). Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2446-2454).
- Gaidon, A., Wang, Q., Cabon, Y., & Vig, E. (2016). Virtual worlds as proxy for multi-object tracking analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4340-4349).
- Jinrang, J., Li, Z., & Shi, Y. (2024). MonoUNI: A unified vehicle and infrastructure-side monocular 3d object detection network with sufficient depth clues. *Advances in Neural Information Processing Systems*, *36*.
    - delete if model not used
- Ye, Q., Jiang, L., Zhen, W., & Du, Y. (2022). Consistency of implicit and explicit features matters for monocular 3d object detection. *arXiv preprint arXiv:2207.07933*.
    - delete if model not used
- Mao, J., Shi, S., Wang, X., & Li, H. (2023). 3D object detection for autonomous driving: A comprehensive survey. *International Journal of Computer Vision*, *131*(8), 1909-1963.

Sources

Waymo Dataset:

- https://www.tensorflow.org/datasets/catalog/waymo_open_dataset
- https://github.com/kittyschulz/Exploring-Waymo-Open-Dataset/tree/master
- https://waymo.com/open/about/

Virtual KITTI:

- https://europe.naverlabs.com/research-old2/computer-vision/proxy-virtual-worlds-vkitti-1/

KITTI:

- https://www.cvlibs.net/datasets/kitti-360/user_login.php
- https://www.tensorflow.org/datasets/catalog/kitti