Introduction to Modern LiDAR for Machine Perception

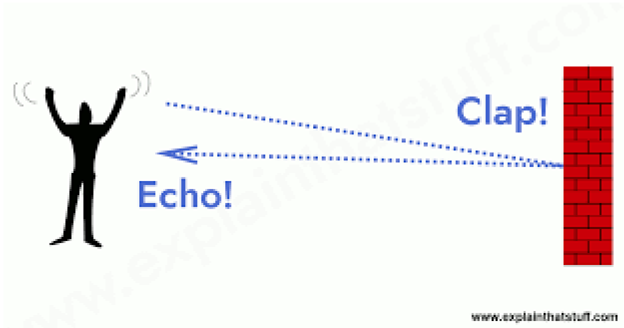
Robert Laganière
CEO / Professor
Sensor Cortek inc. / University of Ottawa
Canada

Presentation outline

- LiDAR fundementals
- LiDAR principles
- LiDAR taxonomy
- LiDAR processing (AI)

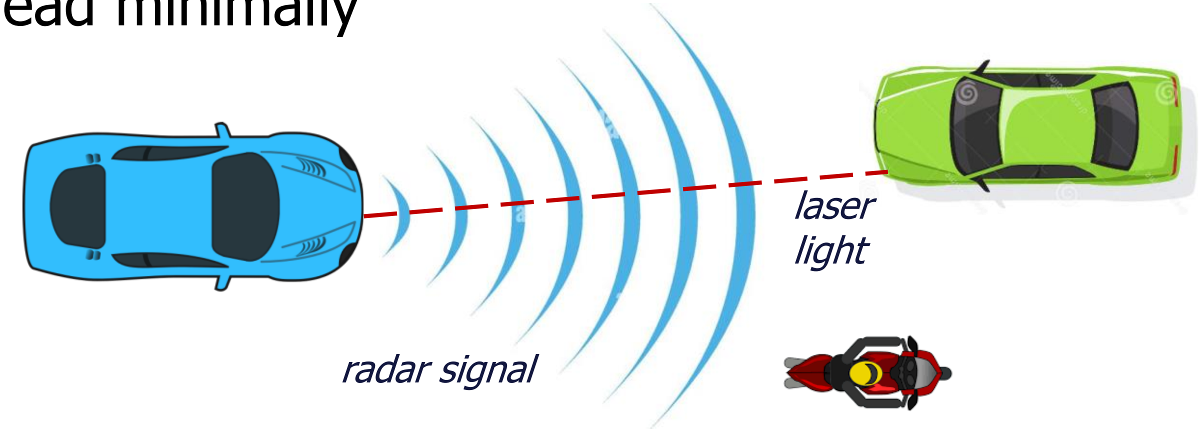
LiDAR: Light Detection and Ranging

- LiDAR (or CoLiDAR) name derived from the RADAR acronym
- LiDAR refers to the technology that uses a laser to sense the environment
- Same fundamental principle as radar (or sonar)
  - Emitting a signal and analyzing the bounced back signal (the echo)



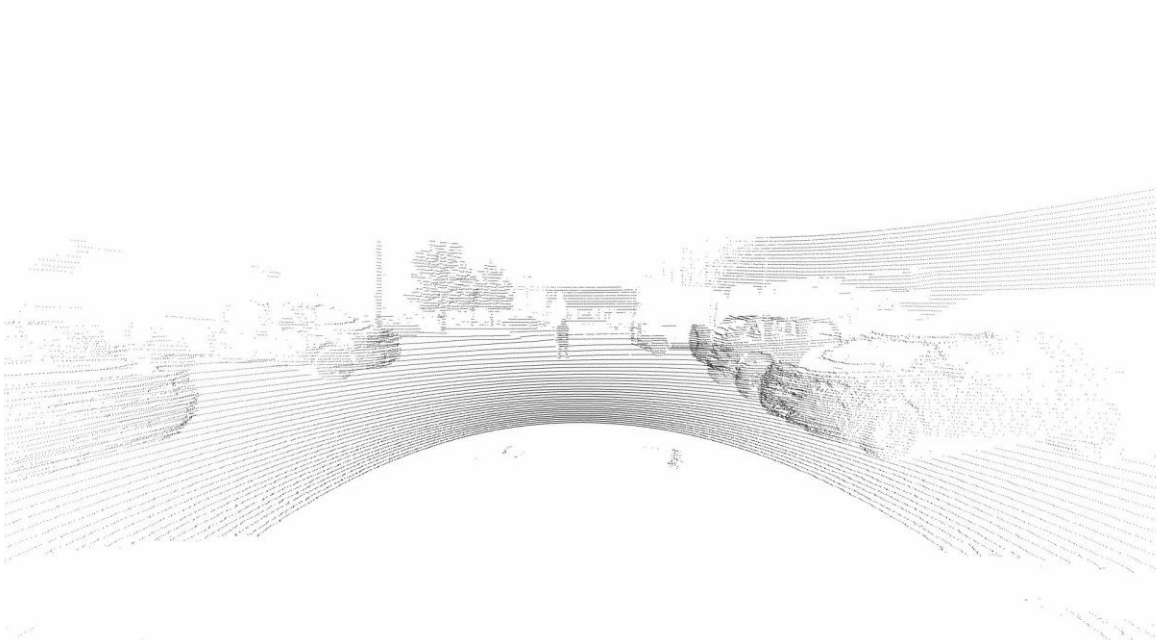
Main benefit of using coherent collimated light beam (i.e., laser)

- The laser light used by LiDAR has two interesting properties:
  - It is coherent: all emitted light rays have the same frequency and phase
  - It is collimated: the beam of light has parallel rays and spread minimally



SENSOR CORTEK

LiDAR sequence



Corresponding camera sequence

did you perceive the same things?

What makes LiDAR an attractive sensor?

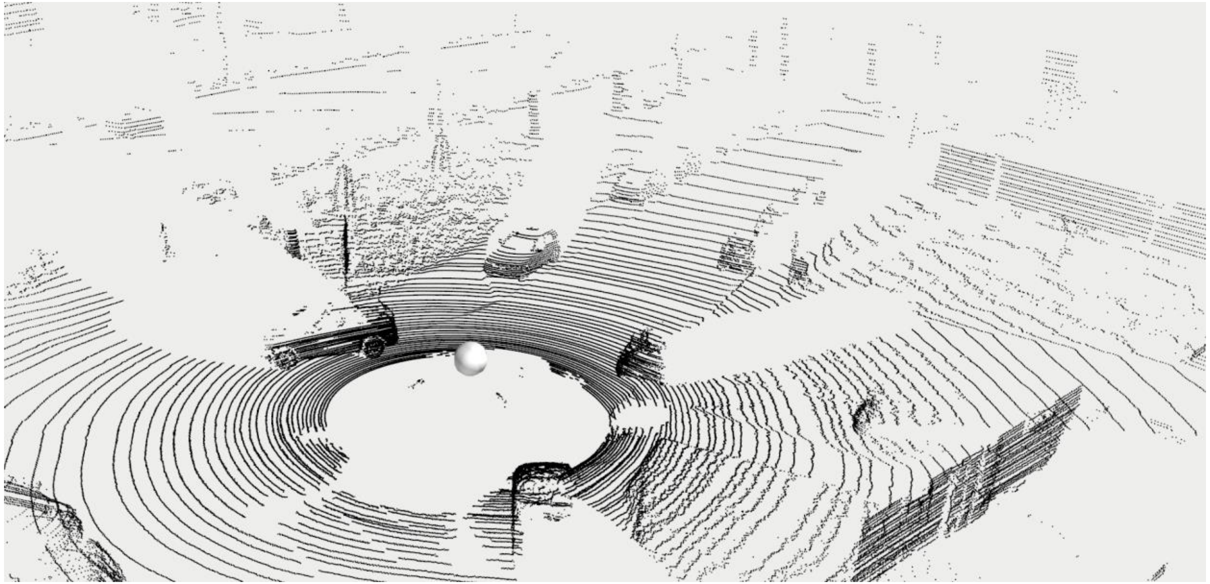
- It produces direct 3D information
- It provides accurate 3D measurements
  - From under a mm to few cm depending on the distance
- It's an active sensor that operates day and night
- It can capture information at long range
  - ~200 m
- It has a large FoV
  - even 360°

However, LiDAR is not perfect

- It captures shape but not appearance
- It produces sparse data
  - Sometimes only few points on an object
- It becomes noisy under fog, snow and rain
- It is still an expensive sensor
  - Some are thousands of $
  - While radars and cameras could be less than $100
- It often includes mechanical parts

The LiDAR point cloud

- A collection of sparse 3D points
  - A LiDAR frame
    - Points not captured at exactly the same time….

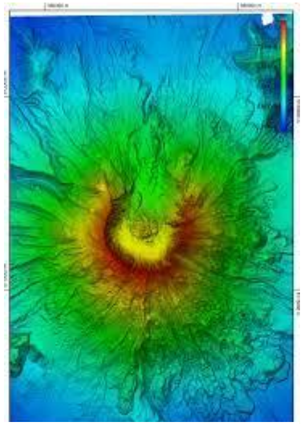


History of Laser

- 1900s: Planck, Einstein and others "discovered" the photon
- May 16th 1960: first laser light produced (T. Maiman)
  - A laser is a device that generates an intense beam of coherent monochromatic light
  - **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation
- Laser differs from other light sources because it emits coherent collimated light
  - Laser light is very narrow, making it possible to see the smallest details with high resolution at relatively long distances.

History of LiDAR

- 1960s: LiDAR initially developed for metrology and atmospheric research
  - But the idea of probing the atmosphere with light can be dated back to the 1930s
- 1970s: LiDAR used for terrain mapping
  - Apollo missions used laser to accurately measure Earth-Moon distances
- 1980s: with the advent of GPS and inertial measurement units (IMUs), LiDAR became very accurate
- 2005: first AV to complete the DARPA Grand Challenge (142 mile desert course) was equipped with a LiDAR



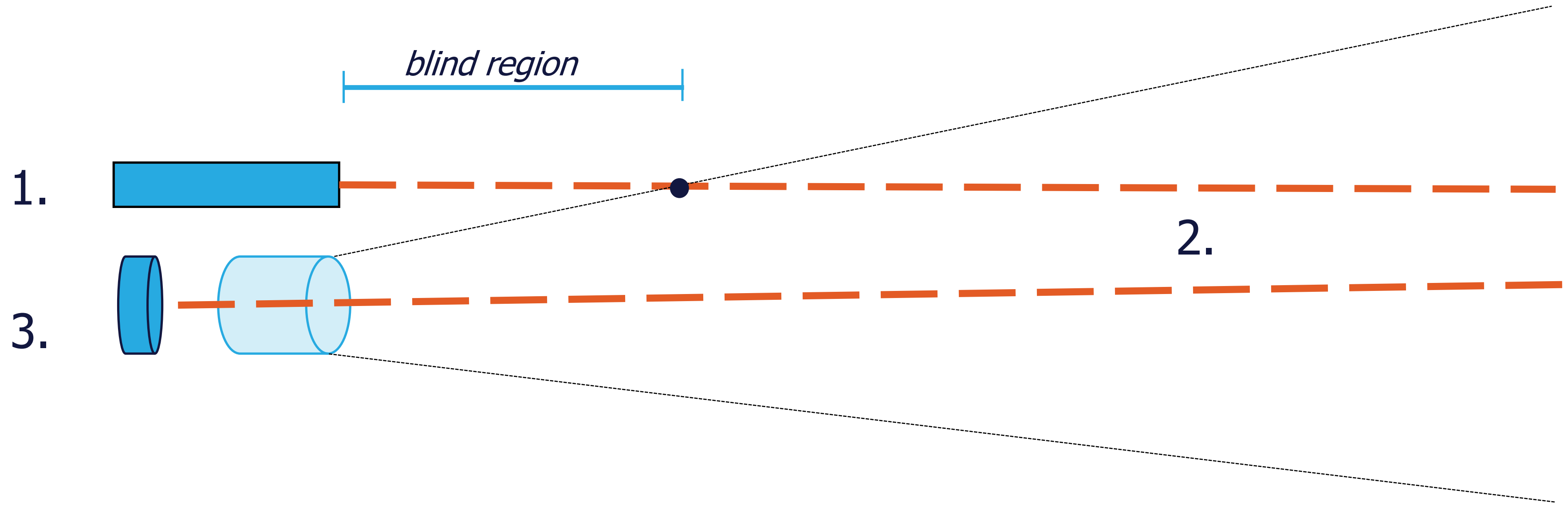
Mt St-Helens
Wikimedia commons

How does LiDAR work? Pulsed LiDAR

1. The LiDAR emits pulsed light waves
2. The light potentially hits a surrounding object and bounces back to the sensor
3. The sensor reads the bounced signal, estimates the time it took to return to the LiDAR and measures the reflected light energy

- Simple technology; almost instantaneous!
- Potential interference from the sun and other LiDARs

How does LiDAR work?

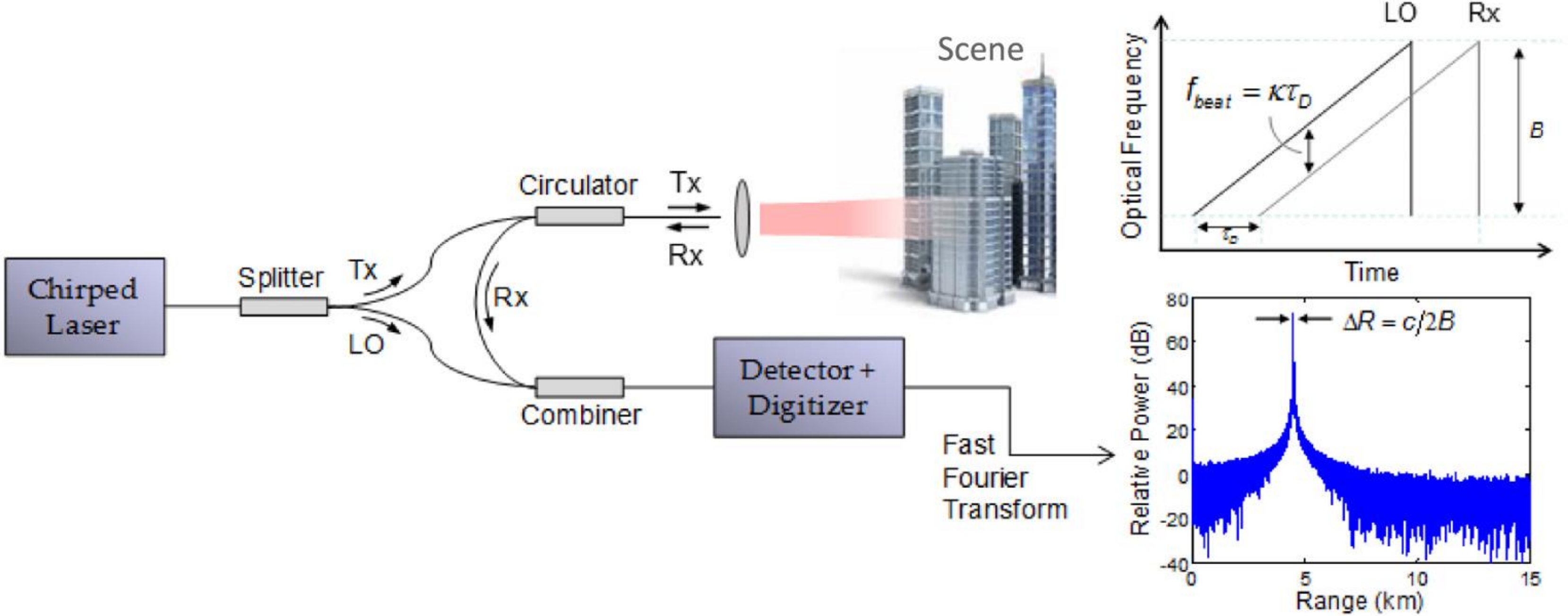


This configuration is called *bistatic* (most common and cheaper)
A *monostatic* optical system aligns the Tx and Rx for better detection

How does LiDAR work? FMCW Lidar

1. LiDAR can also use continuous waves
   - Frequency Modulated Continuous Wave
2. The phase of the bounced back signal will differ from the emitted signal
3. The change in phase is used to extract the distance information
   - This is done by mixing the emitted and received signals (as done in radar)
   - Velocity is a bonus!
   - Virtually no interference
   - You must read a longer signal (stay longer at each point)

How does LiDAR work? FMCW Lidar



www.bridgerphotonics.com/blog/frequency-modulated-continuous-wave-fmcw-lidar

LiDAR mathematics

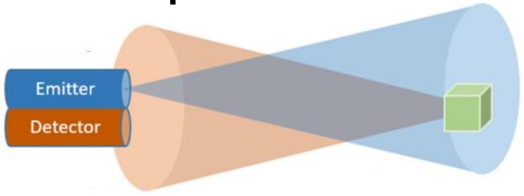
- LiDAR physics is governed by one simple equation

Distance of the object = (Speed of Light x Time of Flight) / 2

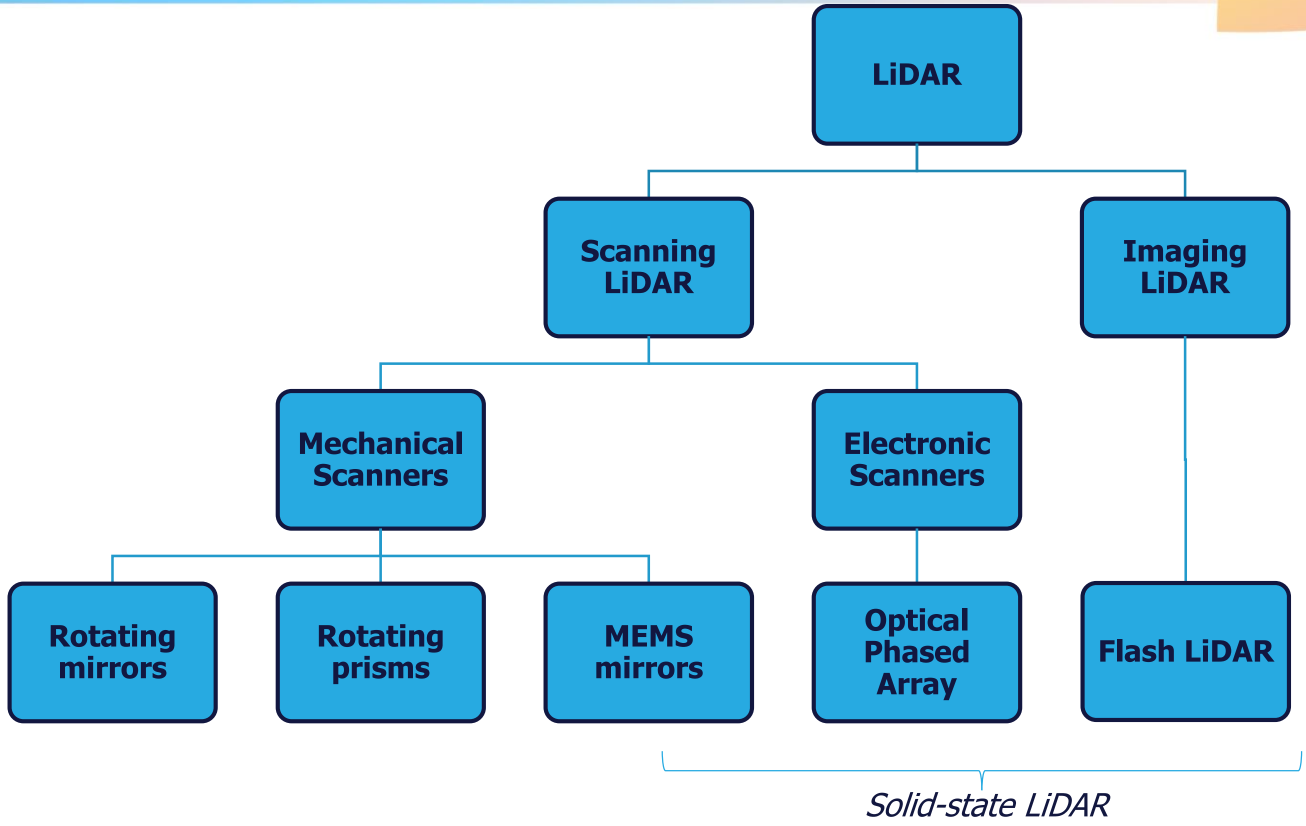
- But to be able to read the received light, you need power (i.e., enough photons bouncing back)

$$\text{Power received} \approx \text{Power transmitted} \times \frac{\text{Cross Section}}{\text{Distance}^2} \times \frac{\text{Optic area}}{\text{Distance}^2}$$

- The laser cross section is the average amount of optical power returned by the target



LiDAR sensor taxonomy



Capturing a scene with LiDAR

- Three strategies:
  - Scanning LiDAR
    - A laser scans the scene and a single photodetector is used to read the returned photons
  - Flash LiDAR
    - The entire field of view is illuminated and a photodetector array captures the received photons
  - Optical Phased Arrays
    - Several transmitters emitting laser light at different phases enabling the steering of the beam (constructive/destructive interference)

Flash LiDAR

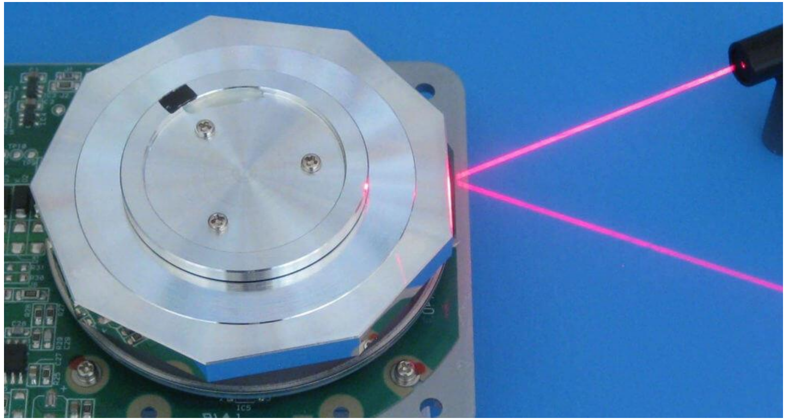
- Shorter range – higher frame rate
- Costly focal plane array
- Limited FoV
- Light is distributed across the FoV – more noisy
- Pixels are small – more power required
- Angular resolution determined by the pixel density
- No motion distortion
- No moving parts

Scanning LiDAR

- Longer range – smaller frame rate
- Expensive scanning mechanism
  - e.g., spinning mirrors, MEMS mirrors
  - or rotate everything
- Can be bulky
- LiDAR motion must be compensated
- Less tolerant to mechanical vibrations

Scanning with rotating mirrors (or rotate everything)

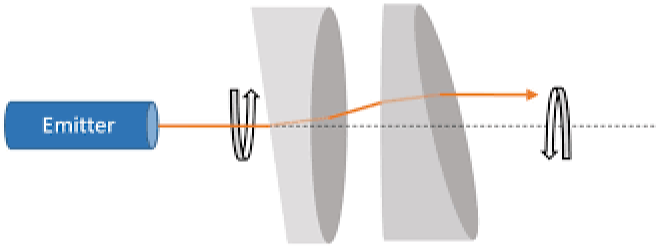
- Today's most popular solution
- Heavier than other solutions
- Vulnerable to vibrations
- Generally includes stack of photodetectors to scan in several horizontal layers



precisionlaserscanning.com/2017/12/mems-mirrors-vs-polygon-scanners-for-lidar-in-autonomous-vehicles/

Scanning with rotating prisms

- Uses two (or more) sequential prisms to steer the beam
- Shape of prism and speed of rotation determines the scan pattern
- Limited FoV



Lidars for vehicles: from the requirements to the technical evaluation, Z. Dai et al., Conference: 9th International Forum on Automotive Lighting, 2021

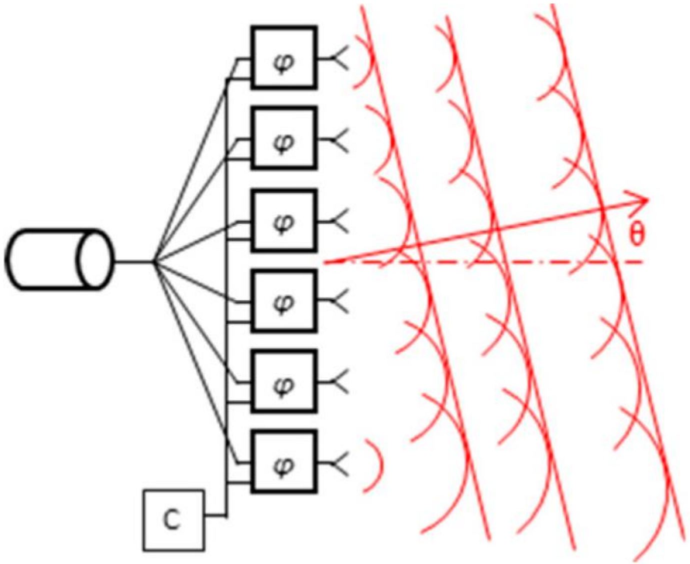
Scanning with MEMS mirrors

- Micro-Electro-Mechanical System
- Quasi solid-state
- Programmable scan patterns
- Limited FoV
- Requires careful calibration

preciseley.com/product/mems-scanning-mirror/

Optical phased array

- Lower resolution – high frame rate
- Solid state
- No moving parts
- Smart zooming capability
- Interference from the antenna lobes limits the angular resolution
- Complex design
- Lower production cost
- Probably the solution of the future



MEMS Mirrors for LiDAR: A Review,
D. Wang, C. Watkins, H. Xie. *Micromachines* 2020, *11*(5)

Lidar technologies

	Rotating	Prisms	MEMS	OPA	Flash
Range	Long	Long	Long	Medium	Low
Frame rate	Low	Low	Low	High	High
FoV	Large	Limited	Limited	Limited	Limited
Resolution	High	High	High	Adaptive	Low
Power	High	High	Low	Low	High
Solid Sate	No	No	Quasi	Yes	Yes
Vulnerability	High	High	High	Low	Low
Complexity	Low	Low	High	High	Low
Cost	High	High	High	Low	High

Some LiDAR specs

- **LiDAR operates in the near IR spectrum**
  - 780 nm to 3000 nm
  - Be careful about eye safety!
- **Typical field of view:**
  - 90°, 180° or 360°
- **Depth resolution determined by the temporal sampling frequency**
  - $\Delta D = c / 2f$
  - e.g., a depth resolution of 1 cm requires a 1.5 GHz sampling rate
- **Angular resolution determined by the scanning point rate**
  - e.g., 0.1° corresponds to 18 cm at 100 m
- **Pulse frequency**
  - e.g., 2 ns corresponds to a range resolution of 3 cm
- **Pulse frequency determines the number of points per second per layer**
  - e.g., 40 kHz

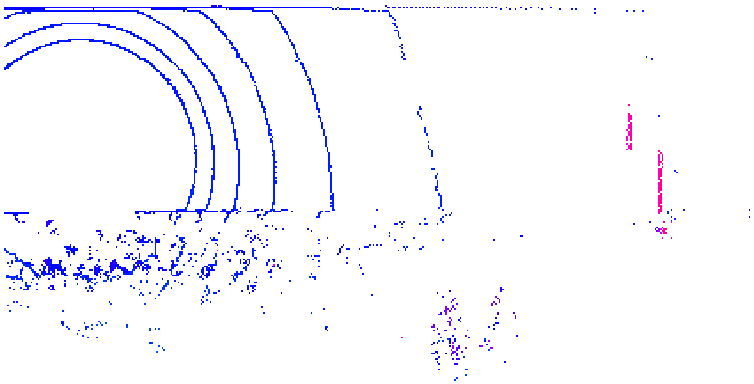
Specs are usually given for 80% Lambertian reflectivity

LiDAR motion compensation using IMU

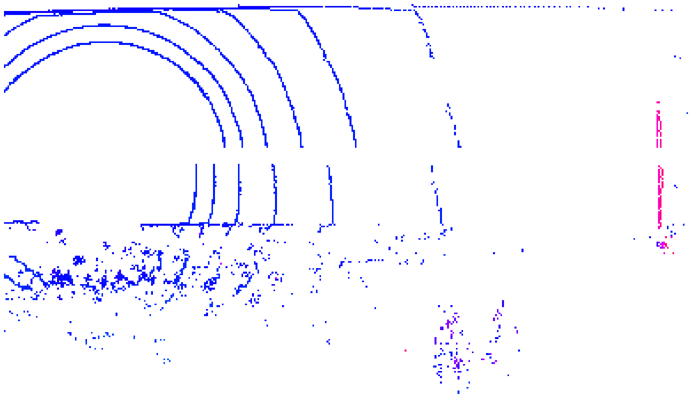
- When the vehicle moves, the LiDAR is scanning a moving scene
  - Which will distort the point cloud
  - This distortion is proportional to the vehicle’s speed and inversely proportional to laser scanning rate
- Solution: using an inertial measurement unit in order to compensate for the sensor motion
  - IMUs provide acceleration and angular velocity
  - Assumption: the ego vehicle motion has constant angular and linear velocities

LiDAR motion compensation using IMU

Before motion compensation



After motion compensation



https://www.mathworks.com/help/lidar/ug/motion-compensation-in-lidar-point-cloud.html

Processing LiDAR data for detection

- A LiDAR sensor produces a frame of 3D points
  - A point cloud
- This point cloud has to be processed and analyzed in order to interpret the scene
  - e.g., detect objects on the road
- But LiDAR data is sparse and unstructured
  - Signal processing (convolution) prefers regular grid sampling
  - LiDAR needs to be preprocessed to build a more suitable representation

LiDAR representation strategies

- Possible representations:
  - Point sets
  - Voxelization
  - Bird’s eye view
  - Point pillars encoding
  - Frontal image generation
  - Sparse convolution

LiDAR representations

- **Point sets**
  - Can we work directly on the point cloud?
    - Point-based representation
- Voxelization
- Bird's eye view
- Point pillars encoding
- Frontal image generation
- Sparse convolution

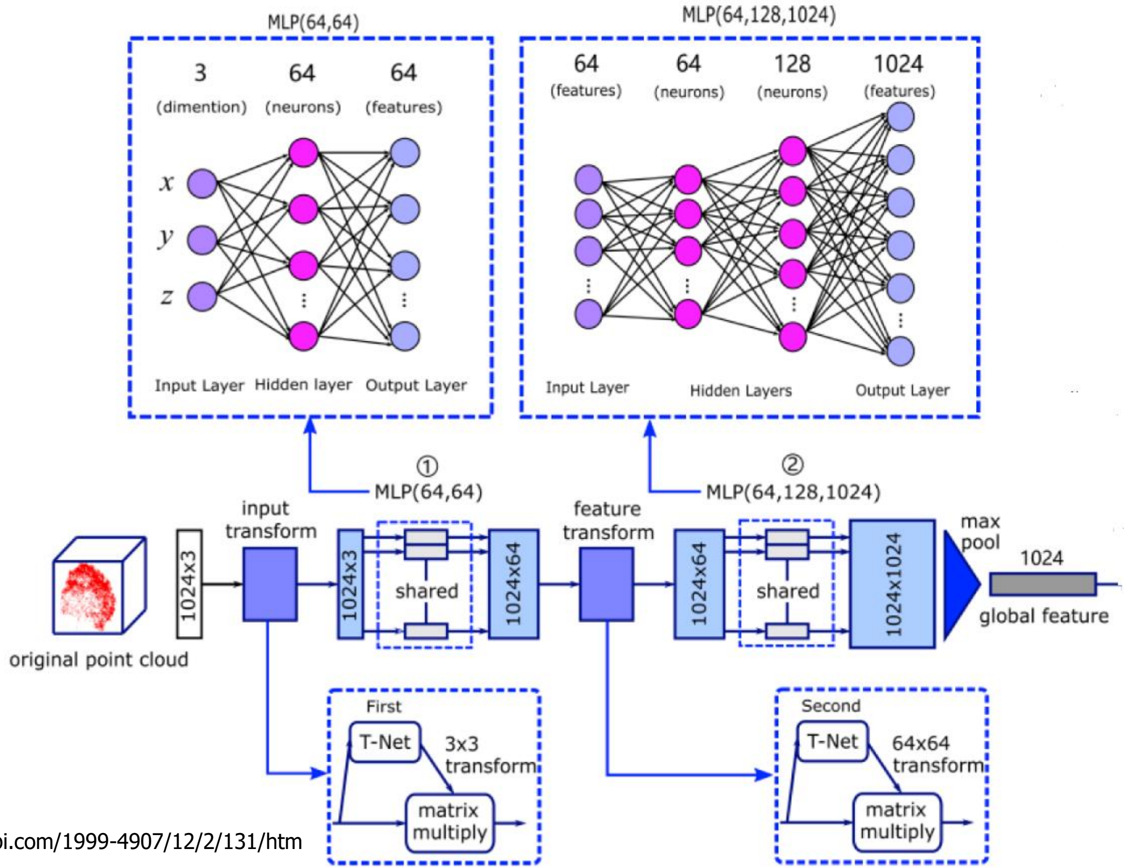
Point-based representations

- The idea is to consume an unordered set of 3D points
  - Transformations are learned to normalize the data
  - Global point features are learned from the set
  - Point cloud has to be segmented into small regions of interest
    - More difficult to apply in a complex scene composed of many objects

C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. Proc. Computer Vision and Pattern Recognition (CVPR) 2017

Point-based representations

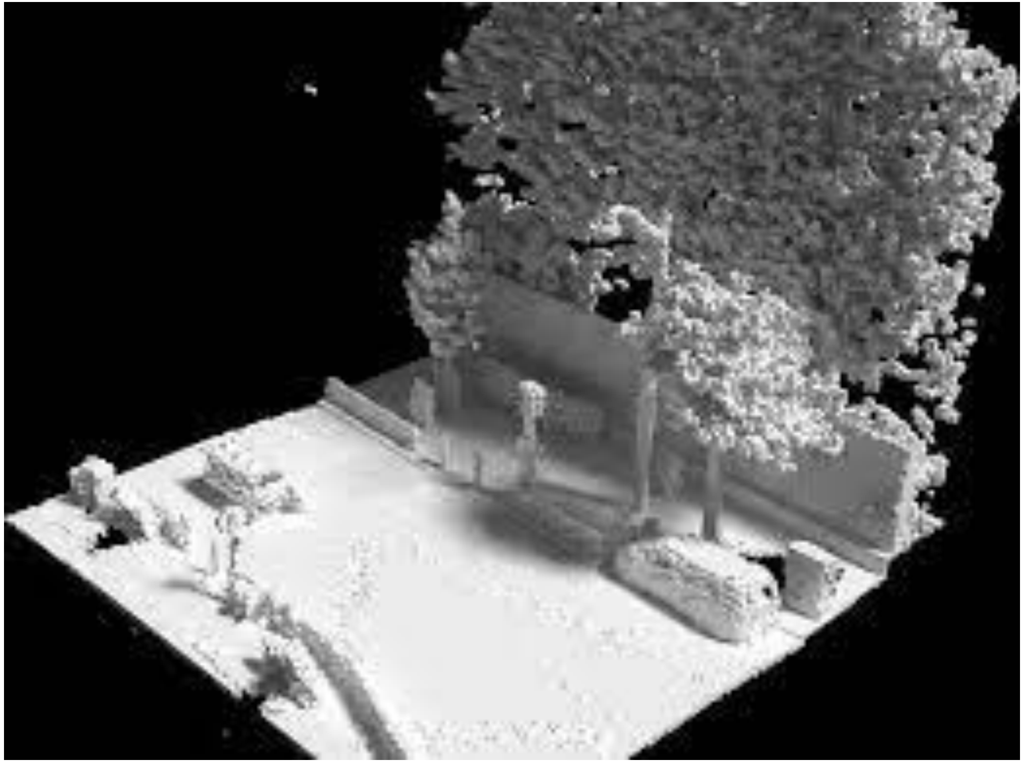
- The idea is to consume an unordered set of 3D points



www.mdpi.com/1999-4907/12/2/131/htm

LiDAR representations

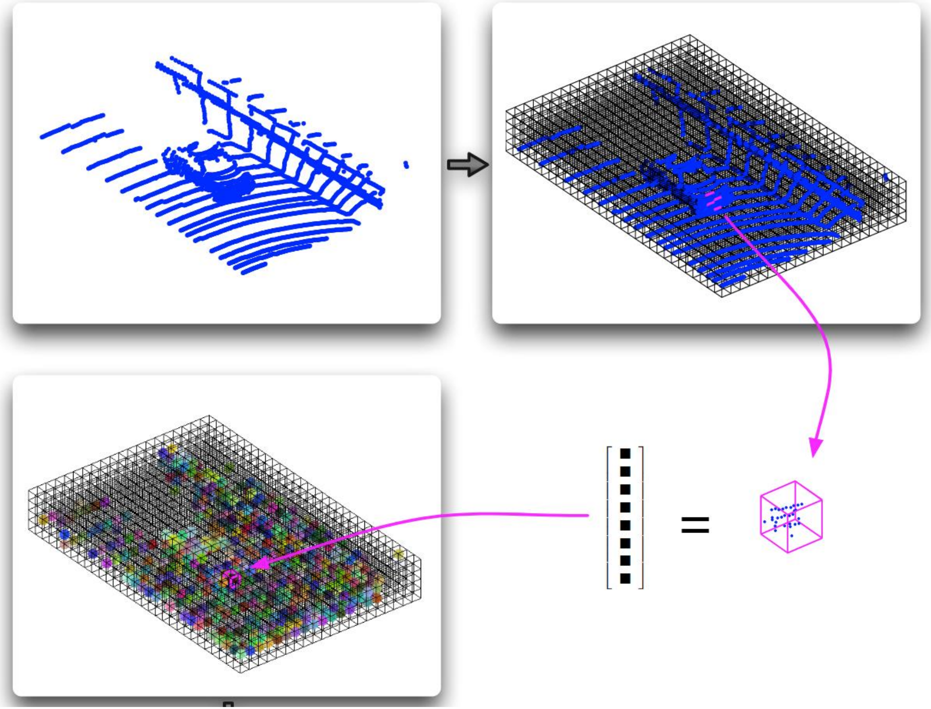
- Point sets
- **Voxelization**
  - e.g., occupancy grid
- Bird's eye view
- Point pillars encoding
- Frontal image generation
- Sparse convolution



Attribute Filtering of Urban Point Clouds Using Max-Tree on Voxel Data, F. Guillotte, Mathematical Morphology and Its Applications to Signal and Image Processing, 2019.

Voxelization

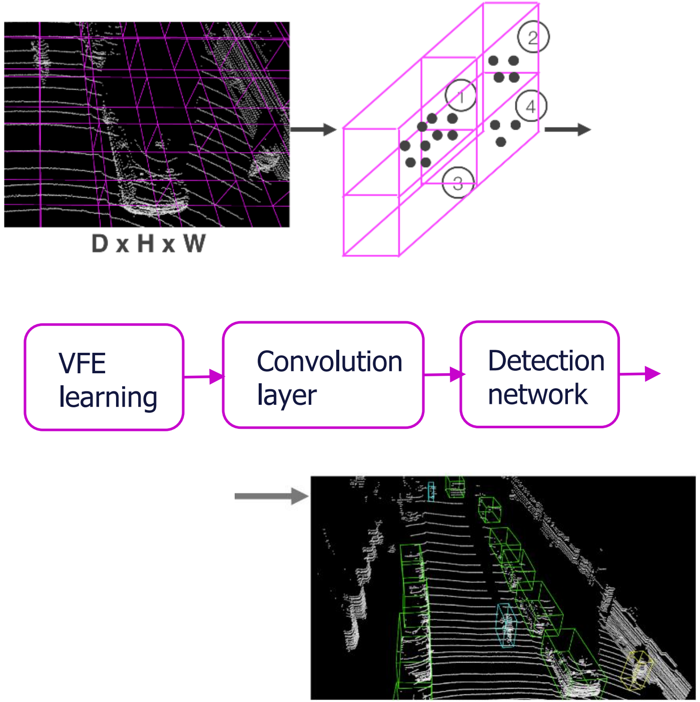
- A 3D voxel grid is created in which each voxel contains:
  - A scalar value
  - A vector made of statistics computed from the points inside the voxel
    - Mean, variance, reflectance, …
- By nature, the occupied voxels are very sparse
  - 3D convolution expensive and inefficient



Wang, D.Z.; Posner, I. Voting for voting in online point cloud object detection. In Proceedings of the Robotics: Science and Systems, Rome, Italy, 13–17 July 2015

Voxelization: VoxelNet

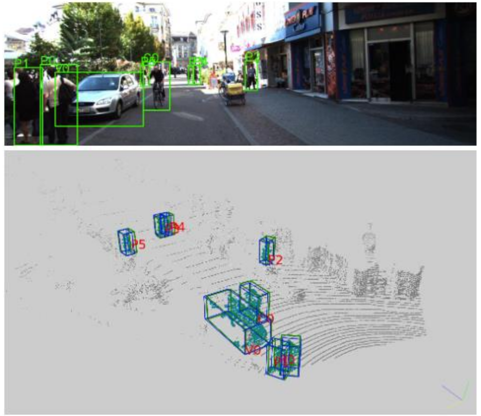
- Voxel feature encoding (VFE) is used
  - from randomly sampled points in each voxel
  - using point coordinates and reflectance
  - and a fully connected network transformation
- 3D convolution is applied on the VFE



Y. Zhou and O. Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In CVPR, 2018

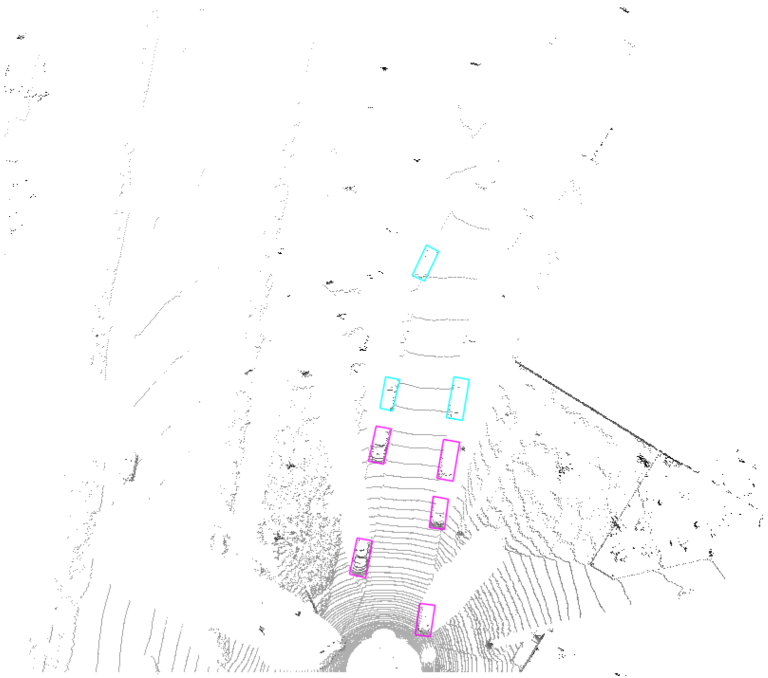
Vehicle Detection results – KITTI Dataset

	Mean Average Precision (%)		
	Easy	**Medium**	**Hard**
VoxelNet	89.35	79.26	77.39



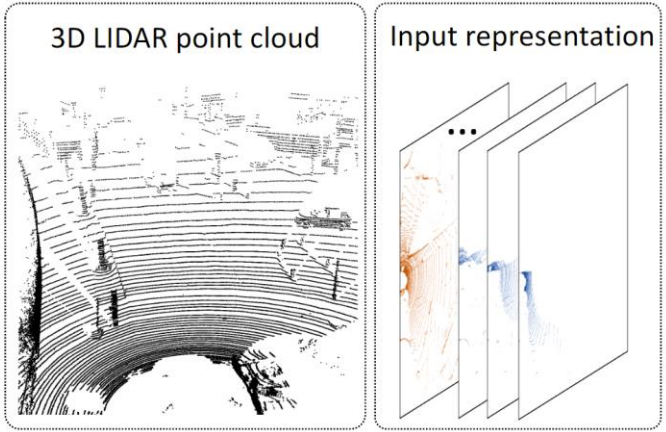
LiDAR representations

- Point sets
- Voxelization
- **Bird’s eye view**
- Point pillars encoding
- Frontal image generation
- Sparse comvolution



Bird's eye view network : Pixor

- Computationally efficient
- Preserve the metric space for objects on the road
- The representation produces an image
  - Height becomes a channel
  - Image detection network can be used
  - Objects are not occluded



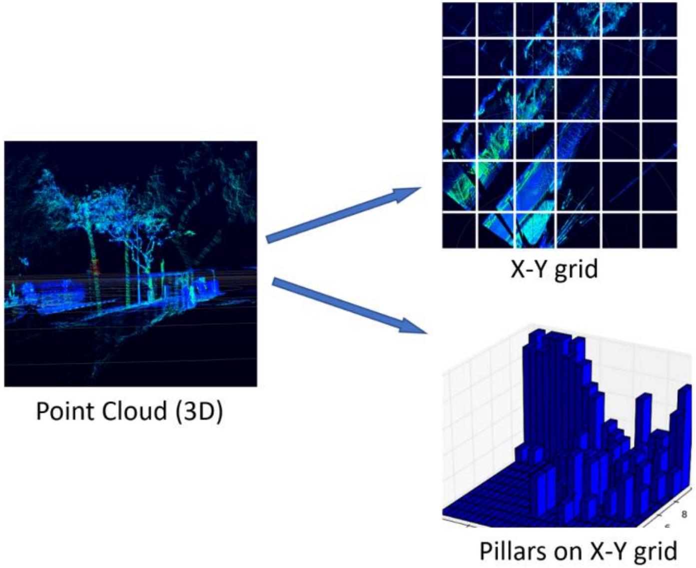
Yang, B., Luo, W., Urtasun, R.: PIXOR: real-time 3D object detection from point clouds. CVPR (2018)

Vehicle Detection results – KITTI Dataset

	Mean Average Precision (%)		
	Easy	**Medium**	**Hard**
VoxelNet	**89.35**	**79.26**	**77.39**
Pixor	81.7	77.05	72.95

LiDAR representations

- Voxelization
- Bird's eye view
- **Pillars encoding**
  - Each pillar encodes point distance to centroid and reflectance
  - Simplified PointNet is used
- Frontal image generation
- Sparse convolution



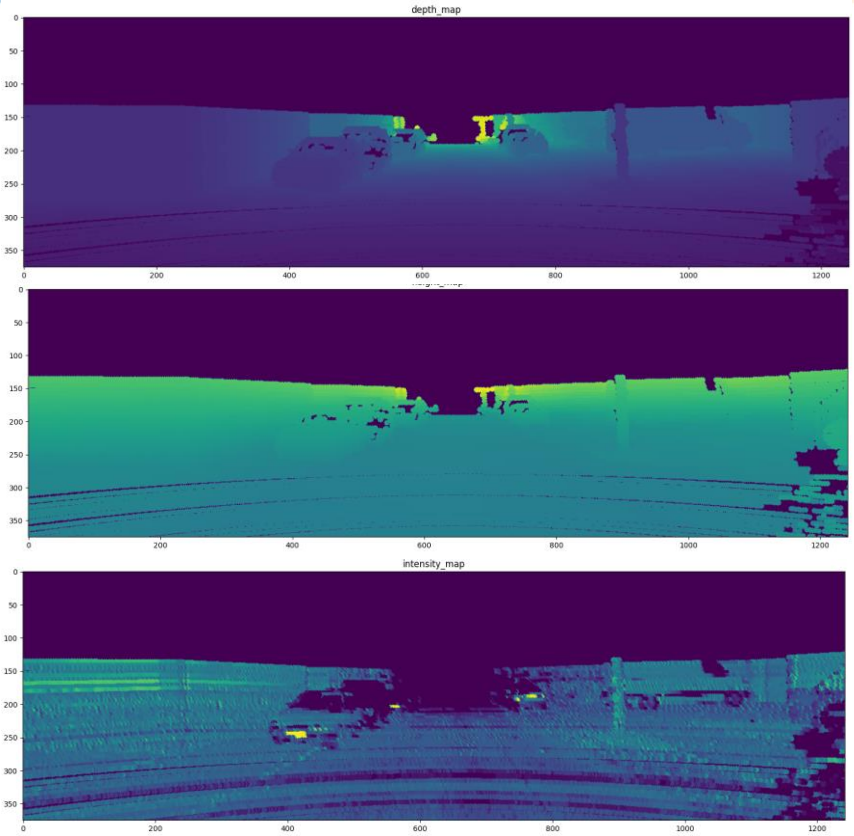
becominghuman.ai/pointpillars-3d-point-clouds-bounding-box-detection-and-tracking-pointnet-pointnet-lasernet-67e26116de5a

Vehicle Detection results – KITTI Dataset

	Mean Average Precision (%)		
	Easy	**Medium**	**Hard**
VoxelNet	89.35	79.26	77.39
Pixor	81.7	77.05	72.95
PointPillar	**92.07**	**87.74**	**86.65**

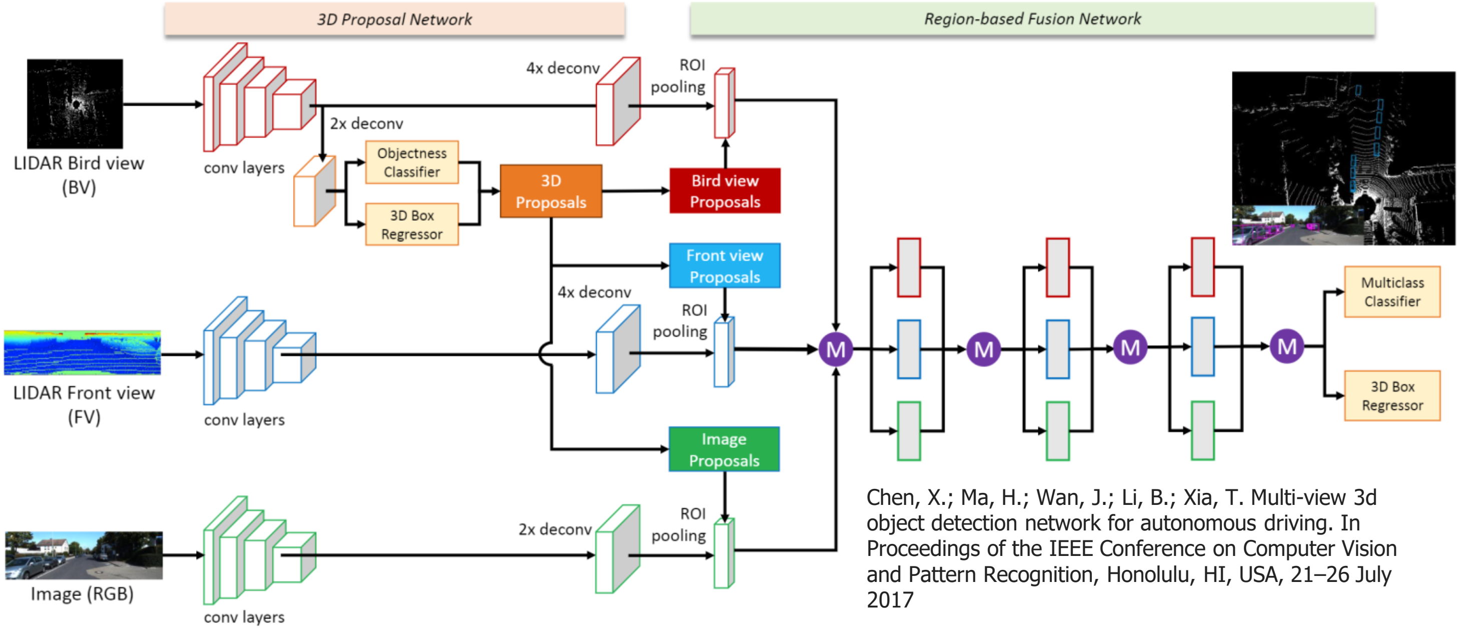
LiDAR representations

- Voxelization
- Bird's eye view
- Point pillars encoding
- **Frontal image generation**
- Sparse convolution



Object-based feature extractor: MV3D

- BEV + LiDAR project + Camera
- BEV is used to propose potential objects
- Multiview features are then uses to predict objects



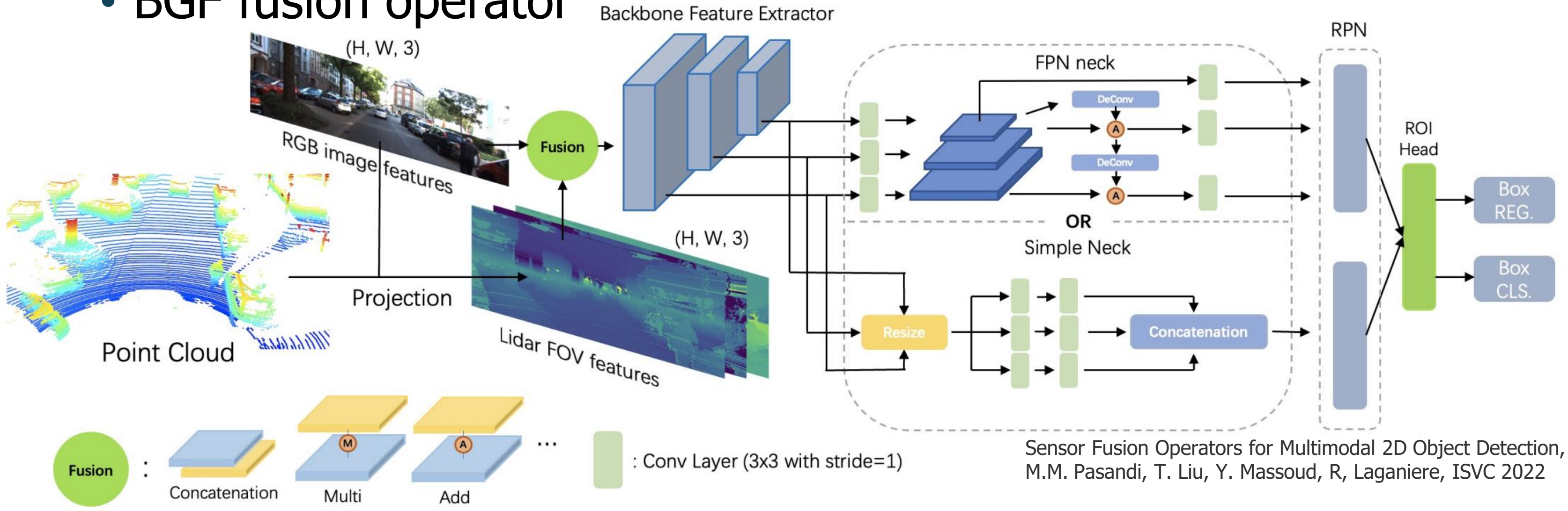
Chen, X.; Ma, H.; Wan, J.; Li, B.; Xia, T. Multi-view 3d object detection network for autonomous driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017

Vehicle Detection results – KITTI Dataset

	Mean Average Precision (%)		
	Easy	**Medium**	**Hard**
VoxelNet	89.35	79.26	77.39
Pixor	81.7	77.05	72.95
PointPillar	**92.07**	**87.74**	**86.65**
MV3D	86.55	78.1	76.67

Frontal view fusion

- All images are fused together
  - Including the camera image
  - Lidar/camera fusion
    - BGF fusion operator



Sensor Fusion Operators for Multimodal 2D Object Detection,
M.M. Pasandi, T. Liu, Y. Massoud, R, Laganiere, ISVC 2022

Vehicle Detection results – KITTI Dataset

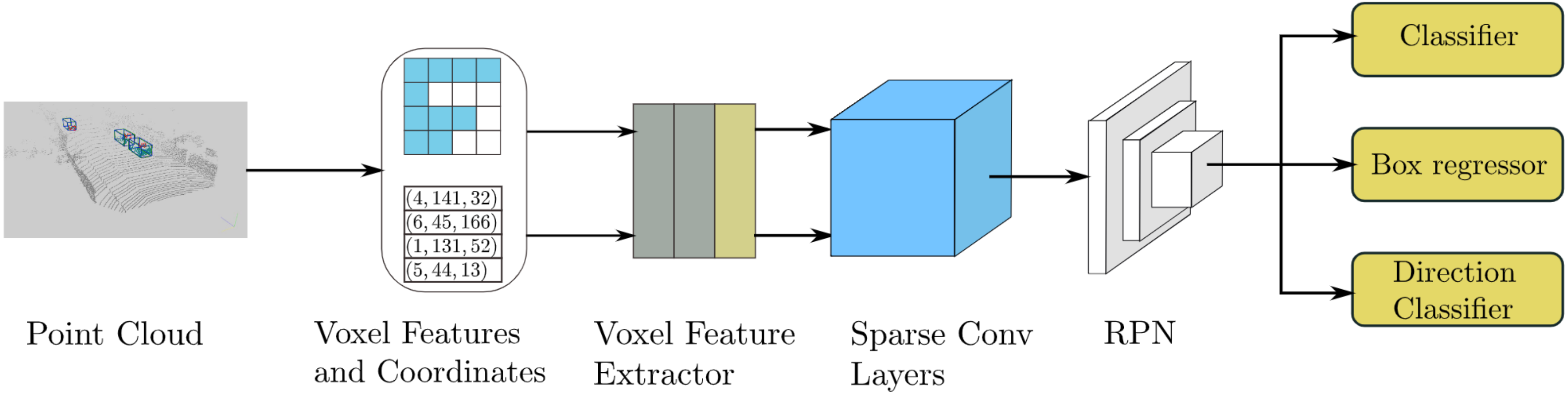
	Mean Average Precision (%)		
	Easy	**Medium**	**Hard**
VoxelNet	89.35	79.26	77.39
Pixor	81.7	77.05	72.95
PointPillar	92.07	87.74	**86.65**
MV3D	86.55	78.1	76.67
BGF Fusion	**94.90**	**88.40**	78.38

SENSOR CORTEK

LiDAR representations

- Voxelization
- Bird’s eye view
- Pillars encoding
- Frontal image generation
- **Sparse convolution**
  - When the data is very sparse, regular convolution becomes very inefficient
  - The idea is to compress the representation by ignoring zero values
  - To this end, look-up tables are often used

SECOND: Sparse convolution on point features

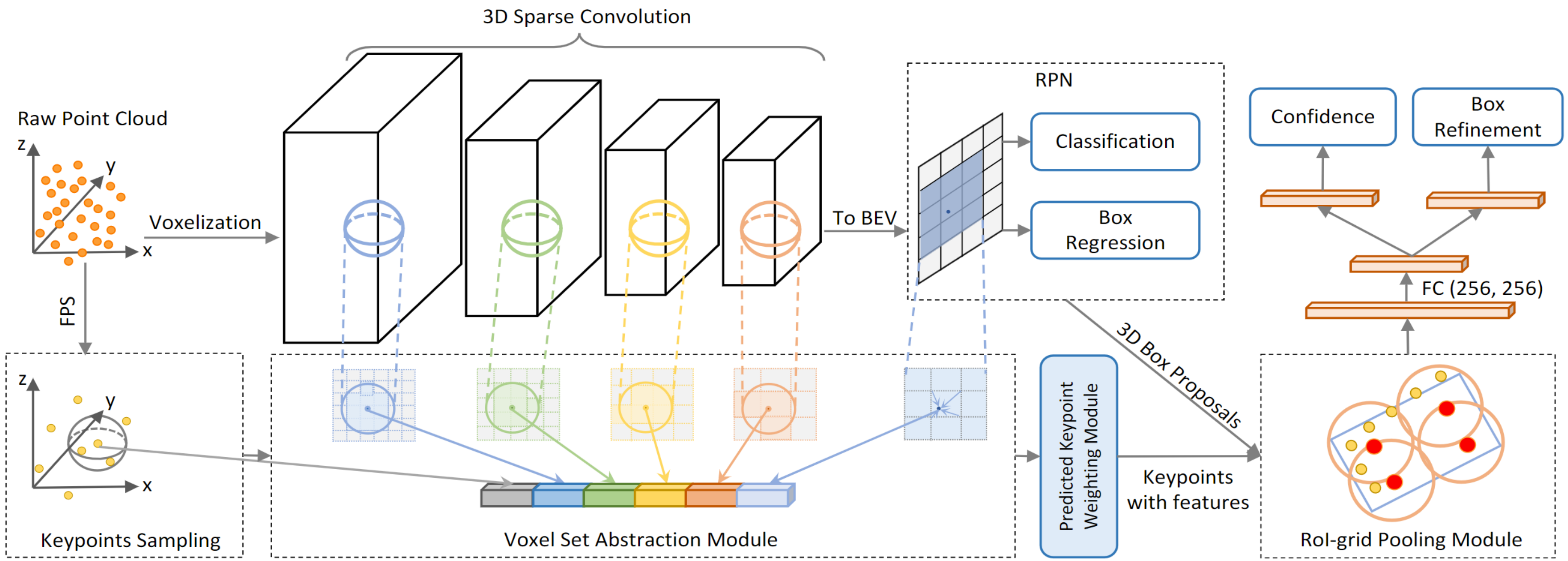


SECOND: Sparsely Embedded Convolutional Detection by Yan Yan, Yuxing Mao, Bo Li, Sensors 2018.

Vehicle Detection results – KITTI Dataset

	Mean Average Precision (%)		
	Easy	**Medium**	**Hard**
VoxelNet	89.35	79.26	77.39
Pixor	81.7	77.05	72.95
PointPillar	92.07	87.74	**86.65**
MV3D	86.55	78.1	76.67
BGF Fusion	**94.90**	**88.40**	78.38
SECOND	91.92	87.92	85.39

PV-RCNN: Point-based + Voxels



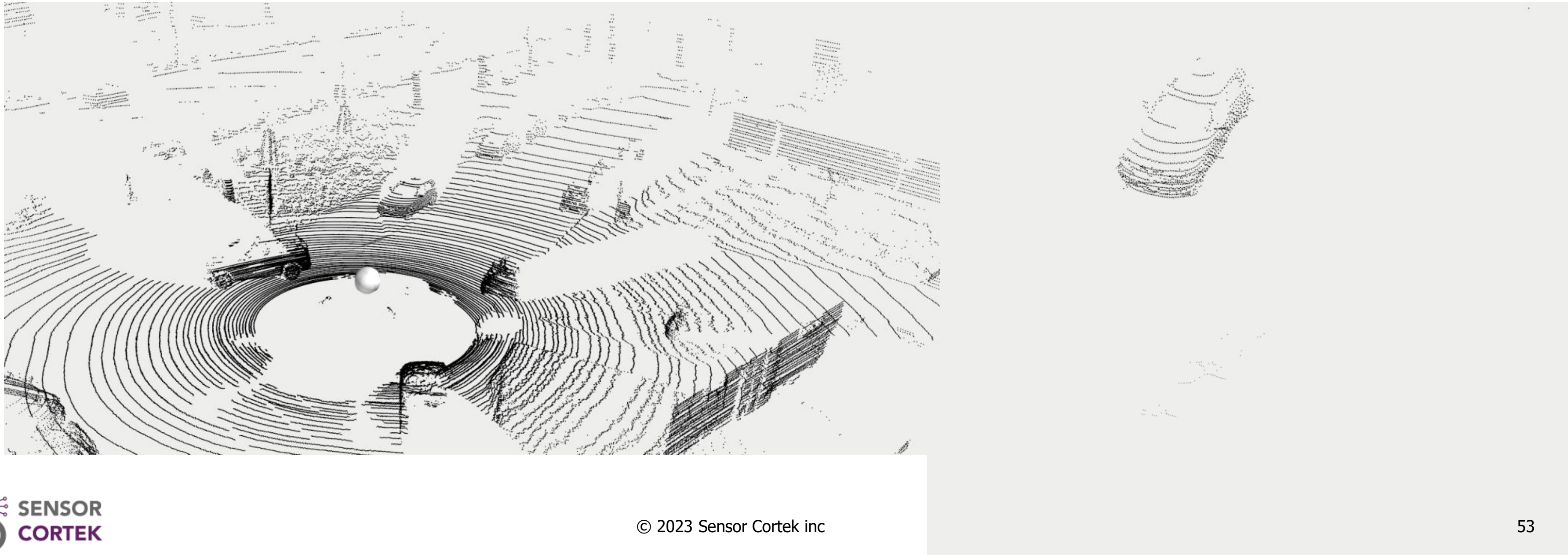
PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection (2020)
Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, Hongsheng Li, IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Vehicle Detection results – KITTI Dataset

	Mean Average Precision (%)		
	Easy	**Medium**	**Hard**
VoxelNet	89.35	79.26	77.39
Pixor	81.7	77.05	72.95
PointPillar	92.07	87.74	86.65
MV3D	86.55	78.1	76.67
BGF Fusion	**94.90**	88.40	78.38
SECOND	91.92	87.92	85.39
PV-RCNN	92.86	**88.93**	**88.74**

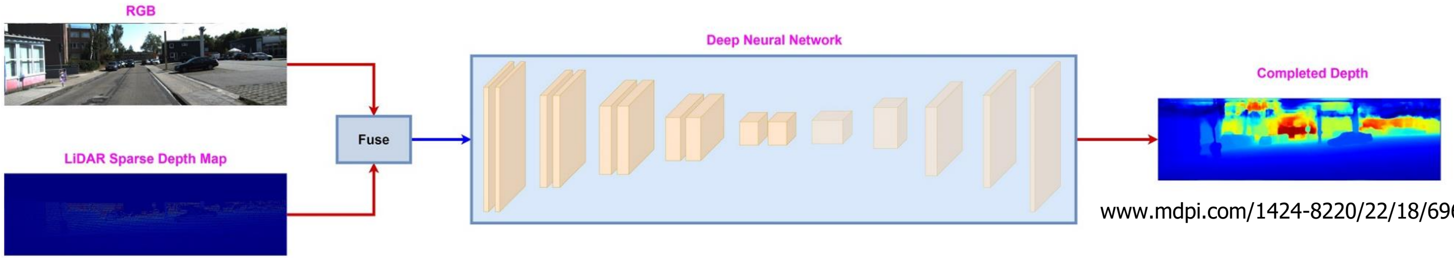
How to further improve detection?

- Solving the LiDAR sparsity problem
  - The farther the object, the sparser the point density



Pseudo LiDAR

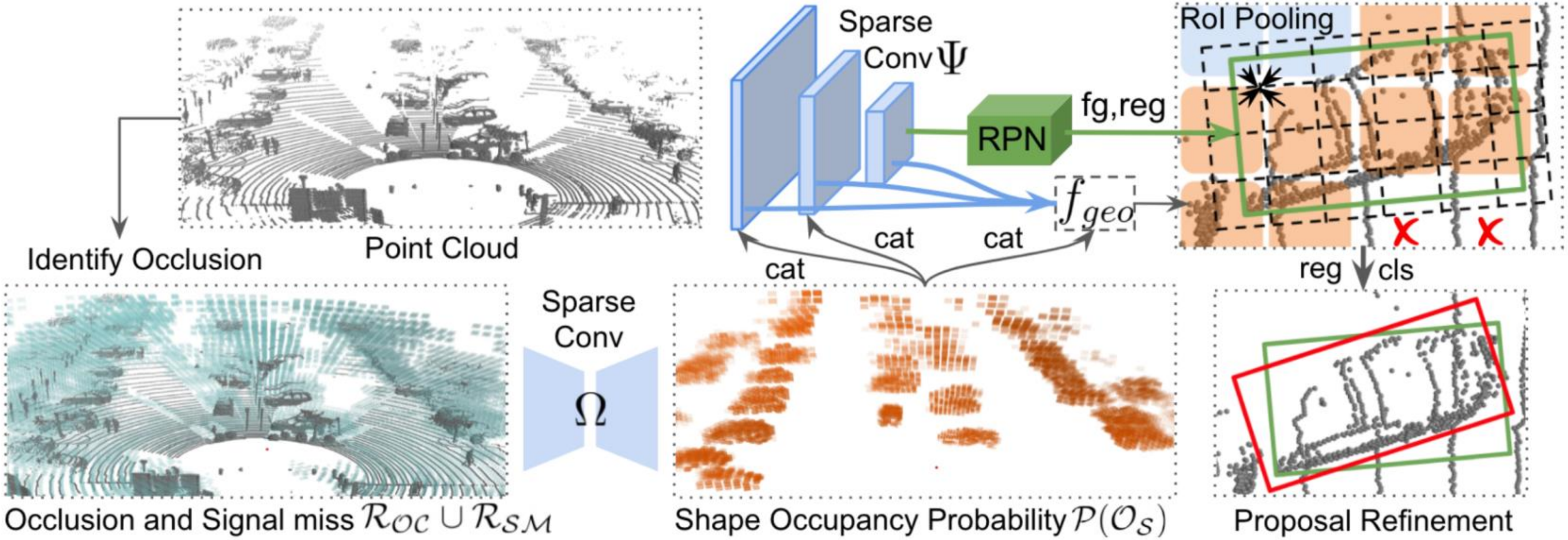
- Depth completion network
  - Can be used to densify the point cloud

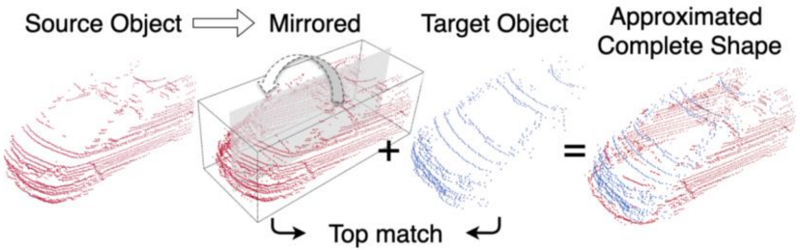


www.mdpi.com/1424-8220/22/18/6969

Depth Completion with Twin Surface Extrapolation at Occlusion Boundaries, Saif Imran, Xiaoming Liu, Daniel Morris, CVPR 2021

BTC: using shape completion





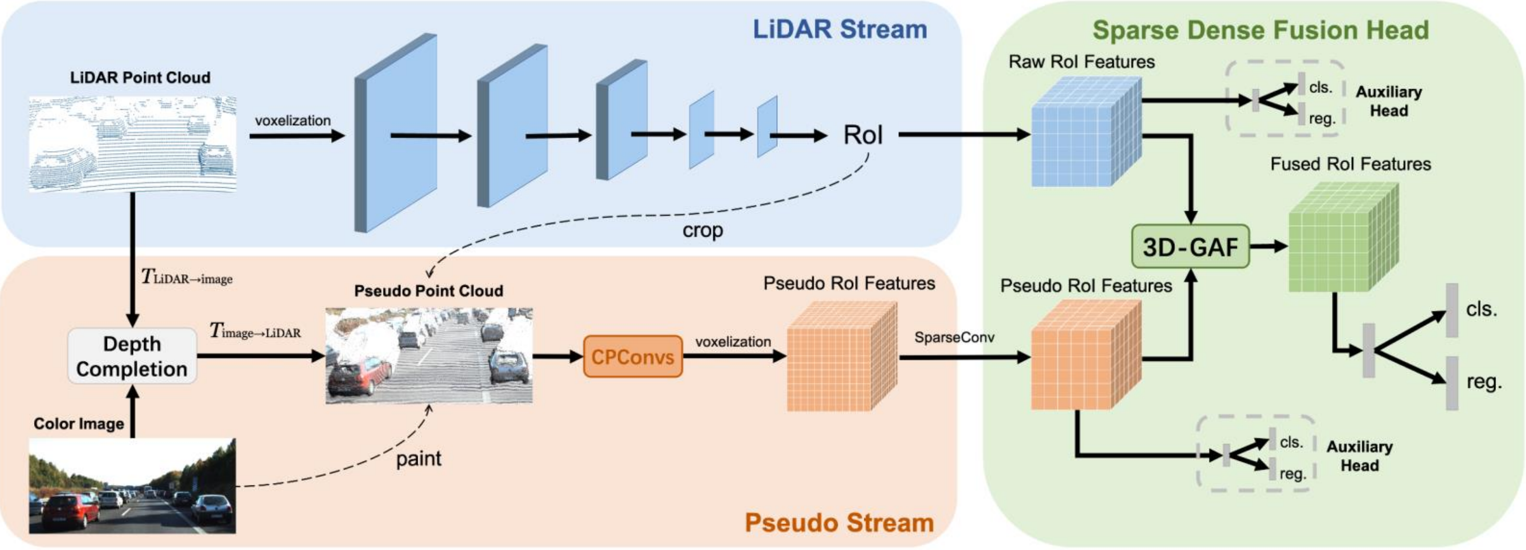
Behind the Curtain: Learning Occluded Shapes for 3D Object Detection, Qiangeng Xu, Yiqi Zhong, Ulrich Neumann, AAAI 2022

Vehicle Detection results – KITTI Dataset

	Mean Average Precision (%)		
	Easy	**Medium**	**Hard**
VoxelNet	89.35	79.26	77.39
Pixor	81.7	77.05	72.95
PointPillar	92.07	87.74	86.65
MV3D	86.55	78.1	76.67
BGF Fusion	**94.90**	88.40	78.38
SECOND	91.92	87.92	85.39
PV-RCNN	92.86	88.93	88.74
BTC	93.46	**89.53**	**87.44**

LiDAR and pseudo-LiDAR: SFD

- Lidar/camera fusion



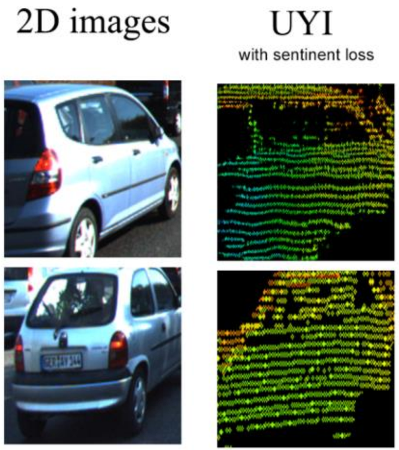
Sparse Fuse Dense: Towards High Quality 3D Detection with Depth Completion, Wu, Xiaopei and Peng, Liang and Yang, Honghui and Xie, Liang and Huang, Chenxi and Deng, Chengqi and Liu, Haifeng and Cai, Deng}, 2022

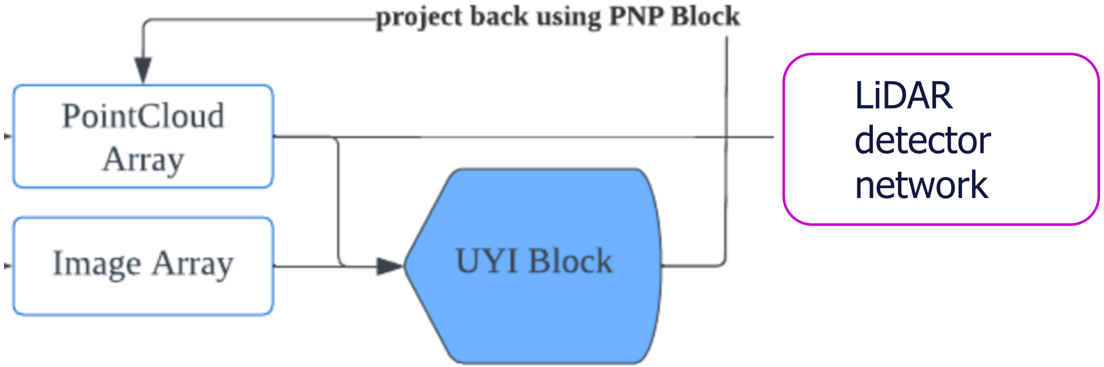
Vehicle Detection results – KITTI Dataset

	Mean Average Precision (%)		
	Easy	**Medium**	**Hard**
VoxelNet	89.35	79.26	77.39
Pixor	81.7	77.05	72.95
PointPillar	92.07	87.74	86.65
MV3D	86.55	78.1	76.67
BGF Fusion	94.90	88.40	78.38
SECOND	91.92	87.92	85.39
PV-RCNN	92.86	88.93	88.74
BTC	93.46	89.53	87.44
SFD	**95.85**	**91.92**	**91.41**

LiDAR densification: UYI

- Pseudo LiDAR module





Use Your Imagination: A Detector-Independent Approach For LiDAR Quality Booster, Z. Zhang, T. Liu, R. Laganiere, 2023

LiDAR densification: UYI

- Vehicle detection performance boosting

Method	Reference	Metrics	UYI Easy	UYI Medium	UYI Hard	Baseline Easy	Baseline Medium	Baseline Hard	Improvement Easy	Improvement Medium	Improvement Hard
SECOND [19]	SENSORS 2018	BEV_AP	**92.27**	87.91	**86.63**	91.92	87.92	85.39	*+0.35*	*-0.01*	*+1.24*
		3D_AP	**88.88**	**79.15**	**76.07**	88.45	78.75	75.72	*+0.43*	*+0.4*	*+0.35*
PointPillar [41]	CVPR 2019	BEV_AP	92.04	**87.93**	85.39	92.04	87.74	86.65	*0*	*+0.19*	*-1.26*
		3D_AP	**88.51**	**78.78**	**75.72**	88.24	78.25	75.32	*+0.27*	*+0.53*	*+0.40*
PartA2 [42]	TPAMI 2020	BEV_AP	**92.96**	**88.21**	**87.56**	90.86	86.26	85.80	*+2.10*	*+1.95*	*+1.76*
		3D_AP	**91.83**	**82.06**	**79.57**	89.93	80.24	77.89	*+1.90*	*+1.82*	*+1.68*
PV-RCNN [18]	CVPR 2020	BEV_AP	**94.63**	**90.28**	88.43	92.86	88.93	88.74	*+1.77*	*+1.35*	*-0.31*
		3D_AP	**92.97**	**83.11**	**82.38**	91.99	82.86	80.40	*+0.98*	*+0.25*	*+1.98*
VOXEL-RCNN [17]	AAAI 2021	BEV_AP	**95.76**	**91.06**	**88.84**	95.35	90.83	88.64	*+0.41*	*+0.23*	*+0.20*
		3D_AP	**92.31**	**83.09**	**82.57**	92.06	82.64	80.10	*+0.25*	*+0.45*	*+2.47*
BtcDet [13]	AAAI 2022	BEV_AP	**93.89**	**91.79**	**89.28**	93.46	89.53	87.44	*+0.43*	*+2.26*	*+1.84*
		3D_AP	**92.81**	**85.80**	**83.08**	92.17	82.39	80.96	*+0.64*	*+3.41*	*+2.12*
SFD [21]	CVPR 2022	BEV_AP	**96.10**	91.68	91.08	95.85	91.92	91.41	*+0.25*	*-0.24*	*-0.33*
		3D_AP	**95.64**	**88.42**	**85.70**	95.01	88.31	85.69	*+0.63*	*+0.11*	*+0.01*

Use Your Imagination: A Detector-Independent Approach For LiDAR Quality Booster, Z. Zhang, T. Liu, R. Laganiere, 2023

Conclusion

- LiDAR provides accurate 3D detection
  - An essential component in ADAS/AV
- LiDAR technology will continue to evolve
  - Improve density
  - Lower power
  - Solid-state
  - Lower cost
- Radar / LiDAR Convergence

References

Reading list on LiDAR and AI

- C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. CVPR 2017.
- Yang, B., Luo, W., Urtasun, R.: PIXOR: real-time 3D object detection from point clouds. CVPR 2018.
- X. Chen, H. Ma, J. Wan, B. Li, T. Xia, Multi-view 3d object detection network for autonomous driving. CVPR 2017.
- Yan Yan, Yuxing Mao, Bo Li. SECOND: Sparsely Embedded Convolutional Detection. Sensors 2018.
- S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, H. Li. PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection. CVPR 2020.
- X. Wu, L. Peng, H. Yang, L. Xie, C. Huang, C. Deng, H. Liu, D. Cai. Sparse Fuse Dense: Towards High Quality 3D Detection with Depth Completion, CVPR 2022.
- Q. Xu, Y. Zhong, U. Neumann. Behind the Curtain: Learning Occluded Shapes for 3D Object Detection. AAAI 2022.