

3D Object Detection for Autonomous Vehicles

A snowy street scene with 3D bounding boxes around a car, a pedestrian, and another car. The scene is dark and foggy, with snow falling. A car is on the left, a pedestrian is in the center, and another car is on the right. The bounding boxes are yellow for the cars and blue for the pedestrian.

Team 8
Elham Amini
Yashaswini Joshi

What is the problem and why it is important?

Autonomous Vehicles are the significant part of the future **safe, green, affordable** and **accessible** transportation.

3 steps to have a level5 AV:

1. Perception
2. Prediction
3. Planning

Dataset:

[Lyft Perception dataset kit](#)

Sample dataset (573 MB)

Training dataset (58 GB)

Testing and Validation dataset (58 GB)

The goal is to build a model to percept objects around an AV and learn more about practical parts of deep learning.

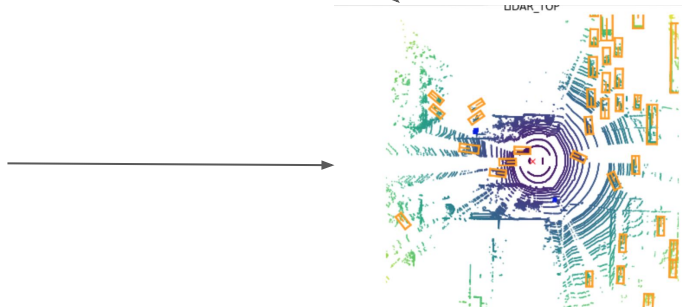


Approach

YOLO (2D)

U-NET (3d point clouds)

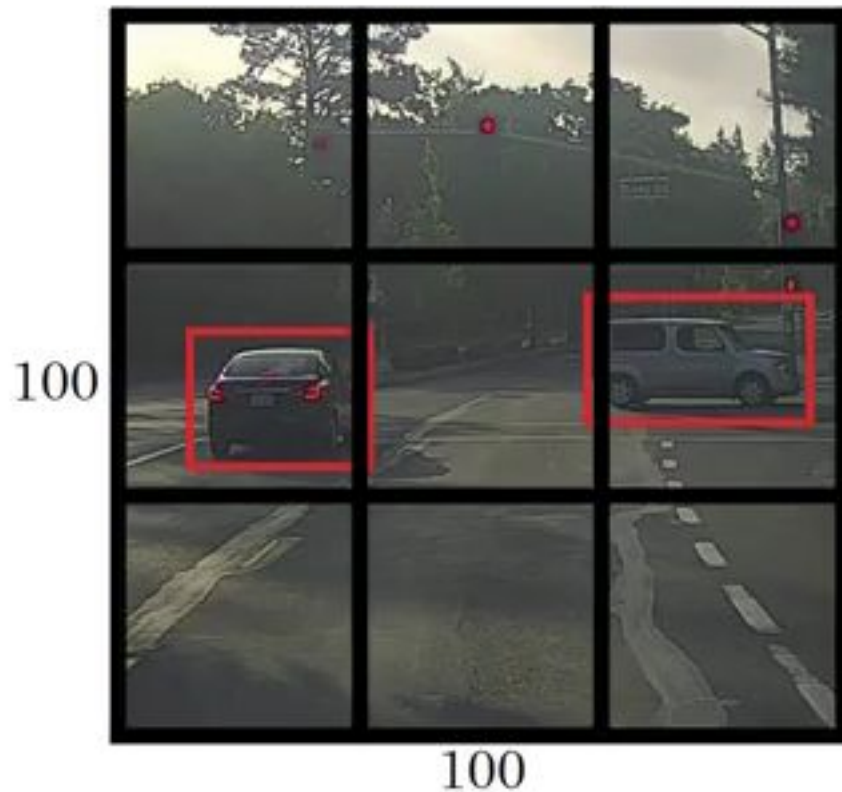
PSPNet-ResNet



YOLO

Data captured by camera (2D)

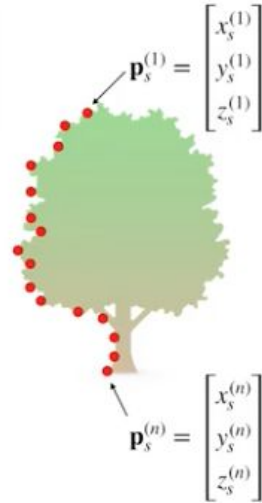
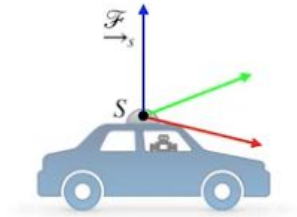
Fast with high accuracy but 2D

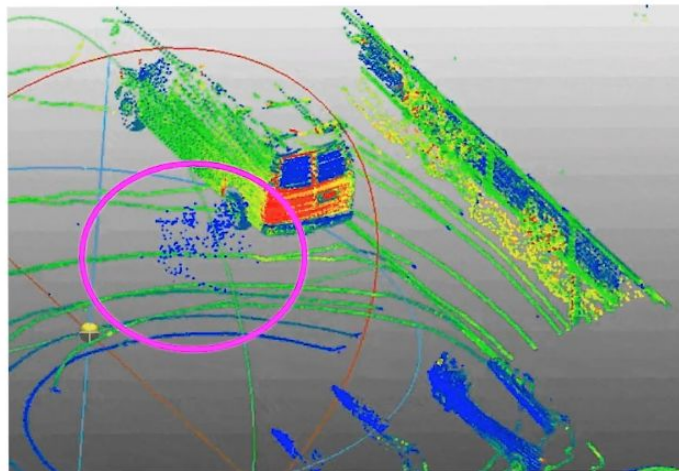


Point Clouds

Captured by Lidar

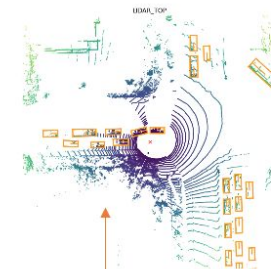
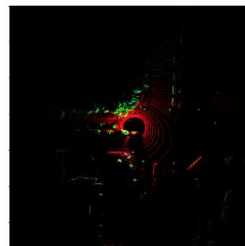
$$\mathbf{P}_s = [\mathbf{p}_s^{(1)} \quad \mathbf{p}_s^{(2)} \quad \dots \quad \mathbf{p}_s^{(n)}] = \begin{bmatrix} x_s^{(1)} & x_s^{(2)} & \dots & x_s^{(n)} \\ y_s^{(1)} & y_s^{(2)} & \dots & y_s^{(n)} \\ z_s^{(1)} & z_s^{(2)} & \dots & z_s^{(n)} \end{bmatrix}$$





U-Net

host	scene_name	date	scene_token	first_sample_token
0	host-a007-a0a0-129048631150282196-153325303087	2019-01-27	473283b48a76b78265a36245851e192866a74361	c77f6a32c0c588334a754633087a875a6d87196d...
1	host-a007-a0a0-129048631150282196-153325303087	2019-01-24	65218935a80026c035a3a30a15199147f5210c1e1...	07899693171746223040400c1c32a180a0e1240c...
2	host-a007-a0a0-129048631150282196-153325303087	2019-01-07	6944b7646219174463b18229b1e600320e10ed12a0d...	83ce1a2b4a077a08819202199152a0890a072070d4...
3	host-a007-a0a0-129048631150282196-153325303087	2019-01-09	8a52a7961a61a4c5c3f70e772288f60a24a4b6e511...	c1d5d342700c0266c4f5a6d55044a4b733a6a5105...
4	host-a008-a0a0-123120001193312000-125150087	2019-01-11	a25a48966c49980a115a78222374c79a6f930c2f...	690ba6e1a1464047a16a76c332b1a611a1c2a19a69...



```

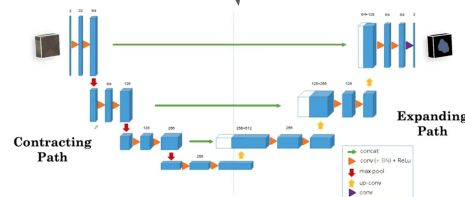
scene_token = '2290184a4016e9803a62ef6a6e1a4211f14e029f4532890001e734114d',
scene_name = '129048631150282196-153325303087',
scene_date = '2019-01-27',
scene_token = '473283b48a76b78265a36245851e192866a74361',
first_sample_token = 'c77f6a32c0c588334a754633087a875a6d87196d...',
scene_token = '65218935a80026c035a3a30a15199147f5210c1e1...',
scene_name = '129048631150282196-153325303087',
scene_date = '2019-01-24',
scene_token = '6944b7646219174463b18229b1e600320e10ed12a0d...',
first_sample_token = '83ce1a2b4a077a08819202199152a0890a072070d4...',
scene_token = '8a52a7961a61a4c5c3f70e772288f60a24a4b6e511...',
scene_name = '129048631150282196-153325303087',
scene_date = '2019-01-09',
scene_token = 'a25a48966c49980a115a78222374c79a6f930c2f...',
first_sample_token = '690ba6e1a1464047a16a76c332b1a611a1c2a19a69...'

```

```

array([[ -97.33626,  -95.891556,  -95.581436,  ...,  -4.3027453,
         -4.3097425,  -4.3089889,  ...,
         [  8.9488838,  9.516727,  10.167154,  ...,  0.17457172,
         0.19416472,  0.21448715],
         [ 11.900117,  11.759933,  11.736123,  ...,  0.15442245,
         0.15207818,  0.15207745],
         [100.,  100.,  100.,  ...,  100.,
          100.,  100.,  100.]], dtype=float32)

```



PSPNet-ResNet Model

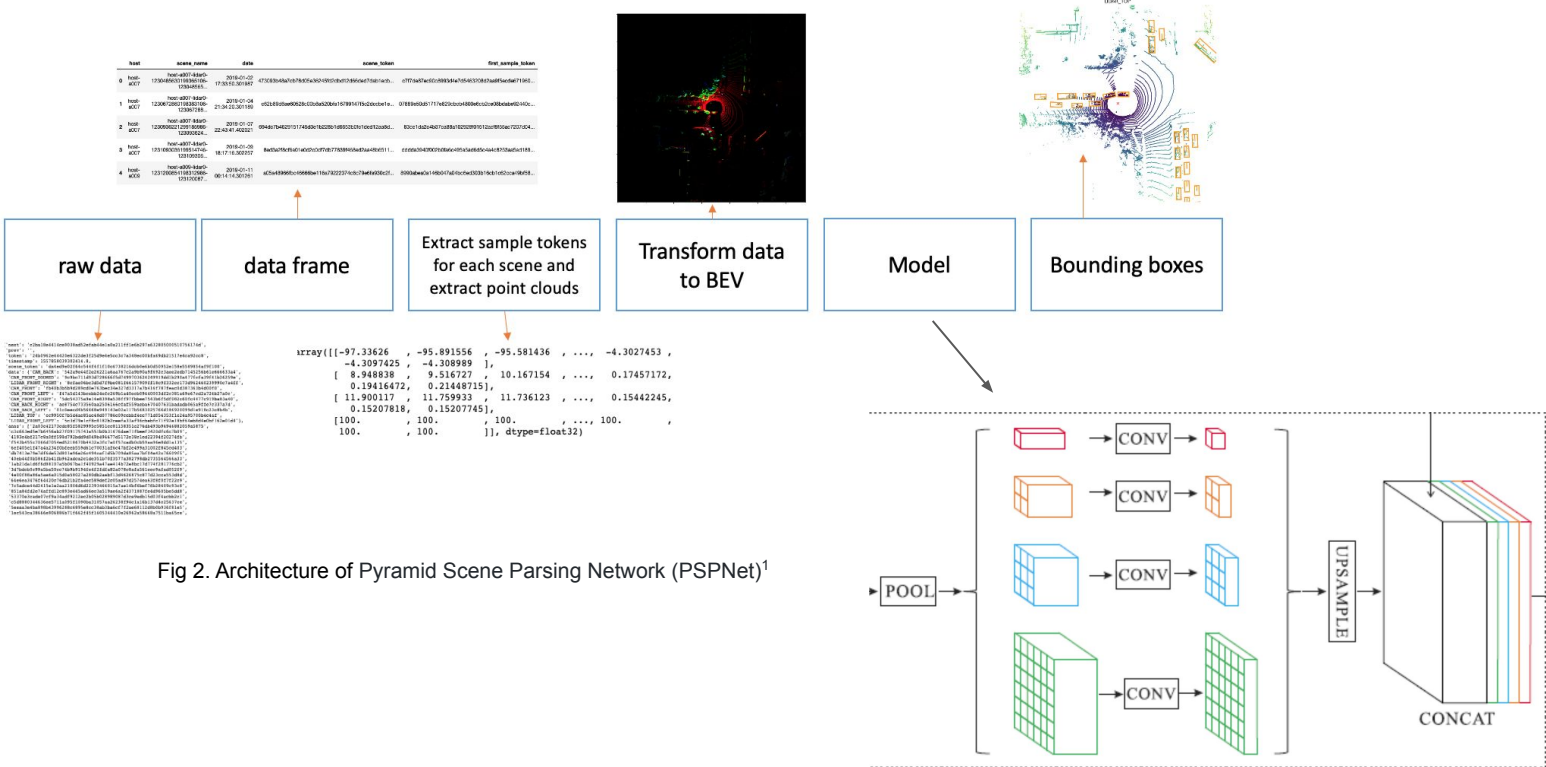
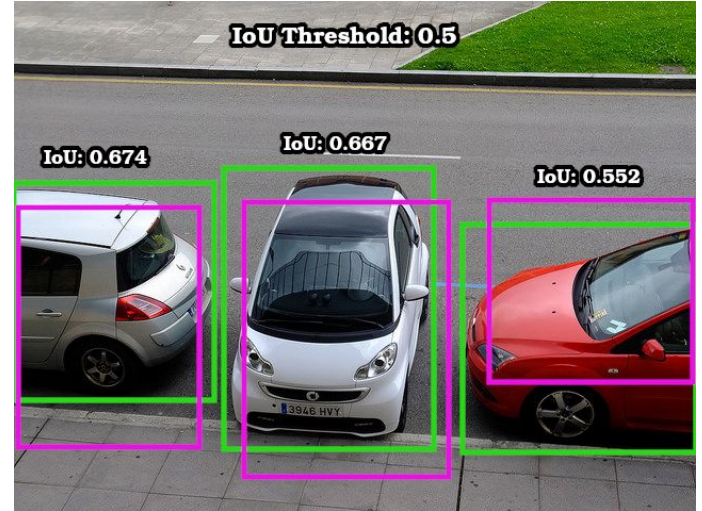


Fig 2. Architecture of Pyramid Scene Parsing Network (PSPNet)¹

¹ Pyramid Scene Parsing Network by Zhao et al, <https://arxiv.org/abs/1612.01105>

Evaluation and Findings

Model	Number of Epochs	MAP	MAP - car	MAP - pedestrian	MAP - animal
U-Net	1	0.09	0.46	0.00021	0.0
U-Net	15	0.01	0.50	0.0027	0.0
PSPNet-ResNet	15	0.09	0.71	0.0134	0.0



Conclusions

Cloud points gathered from Lidar sensor are very important in AV 3D object detection:

- They do not get affected by adverse weather conditions
- They do not rely on light for object detection

But:

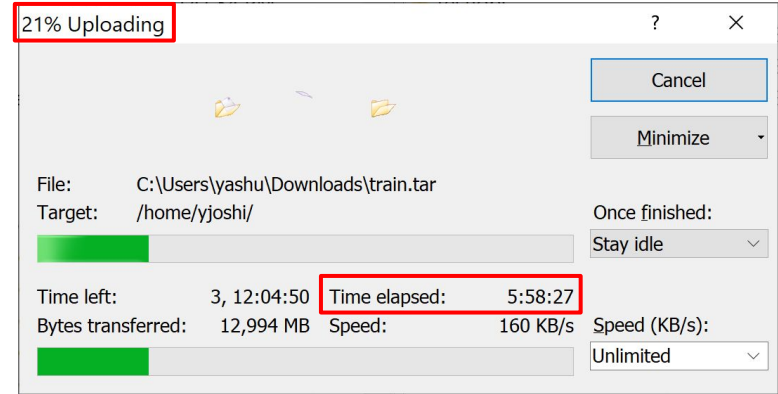
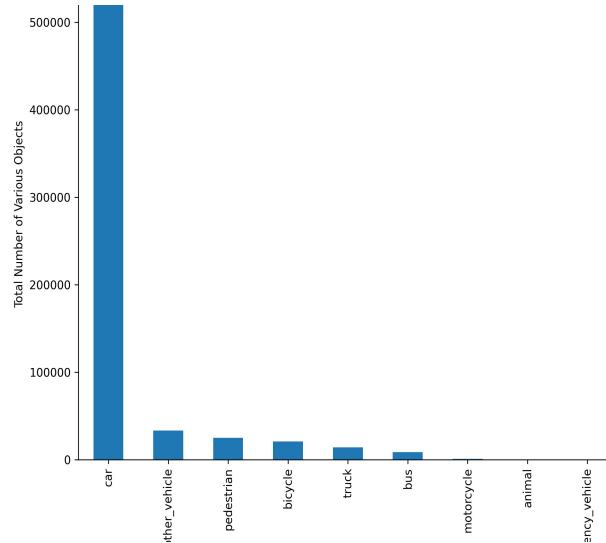
They are not reliable for detecting small objects such as animals, or any smaller class in the training set.

Possible Solutions and Future Works

- Using both images and point clouds to detect objects (They are complementary)
- Having separate pretrained model for each class such as animals, pedestrians,...and fine tune model with autonomous vehicle data
- Training model with more epochs
- Using ensemble models (combination of various trained model) for prediction

Problems

- Related to data:
 - Downloading Dataset
 - Understanding the data
 - Transferring data to GL
 - Installing some of the packages
 - The dataset we have is highly imbalance.



First two scenes of the data:

```
[{'log_token': 'da4ed9e02f64c544f4f1f10c6738216dcb0e6b0d50952e158e5589854af9f100',  
  'first_sample_token': '24b0962e44420e6322de3f25d9e4e5cc3c7a348ec00bfa69db21517e4ca92cc8',  
  'name': 'host-a101-lidar0-1241893239199111666-1241893264098084346',  
  'description': '',  
  'last_sample_token': '2346756c83f6ae8c4dladec62b4d0d31b62116d2e1819e96e9512667d15e7cec',  
  'nbr_samples': 126,  
  'token': 'da4ed9e02f64c544f4f1f10c6738216dcb0e6b0d50952e158e5589854af9f100'},  
{ 'description': '',  
  'first_sample_token': '3b30673b9d944ec6058ef5a8debb4c0a6fe075bca7076e06cf42a2bc10dc446e',  
  'log_token': '0a6839d6ee6804113bb5591ed99cc70ad883d0cff396e3aec5e76e718771b30e',  
  'name': 'host-a006-lidar0-1236037883198113706-1236037908098879296',  
  'token': '0a6839d6ee6804113bb5591ed99cc70ad883d0cff396e3aec5e76e718771b30e',  
  'last_sample_token': '5ec01e4634ca91751311eaafb45a9196ba8616bf05edc85593aff158db653a34',  
  'nbr_samples': 126}]
```

Information inside one sample

```
{'next': 'c2ba18e4414ce9038ad52efab44e1a0a211ff1e6b297a632805000510756174d',
'prev': '',
'token': '24b0962e44420e6322de3f25d9e4e5cc3c7a348ec00bfa69db21517e4ca92cc8',
'timestamp': 1557858039302414.8,
'scene_token': 'da4ed9e02f64c544f4f1f10c6738216dcb0e6b0d50952e158e5589854af9f100',
'data': {'CAM_BACK': '542a9e44f2e26221a6aa767c2a9b90a9f692c3aee2edb7145256b61e666633a4',
'CAM_FRONT_ZOOMED': '9c9bc711d93d728666f5d7499703624249919dd1b290a477fcfa39f41b26259e',
'LIDAR_FRONT_RIGHT': '8cfae06bc3d5d7f9be081f66157909ff18c9f332cc173d962460239990c7a4ff',
'CAM_FRONT': 'fb40b3b5b9d289cd0e763bec34e327d3317a7b416f787feac0d387363b4d00f0',
'CAM_FRONT_LEFT': 'f47a5d143bcebb24efc269b1a40ecb09440003df2c381a69e67cd2a726b27a0c',
'CAM_FRONT_RIGHT': '5dc54375a9e14e8398a538ff97fbbee7543b6f5df082c60fc4477c919ba83a40',
'CAM_BACK_RIGHT': 'ae8754c733560aa2506166cfaf559aeba670407631badadb065a9ffe7c337a7d',
'CAM_BACK_LEFT': '01c0eecd4b56668e949143e02a117b5683025766d186920099d1e918c23c8b4b',
'LIDAR_TOP': 'ec9950f7b5d4ae85ae48d07786e09cebbf4ee771d054353f1e24a95700b4c4af',
'LIDAR_FRONT_LEFT': '5c3d79e1cf8c8182b2ceefa33af96cbebf71f92e18bf64eb8d4e0bf162e01d4'},
'anns': ['2a03c42173cde85f5829995c5851cc81158351e276db493b96946882059a5875',
'c3c663ed5e7b6456ab27f09175743a551b0b31676dae71fbee73420dfc6c7b09',
'4193e4bf217c8a0ff598d792bdd9d049b496677d5172e38c1ed22394f20274fb',
'f543b455c7066d7054ed5218670b4432a3fc7a0f57cadb0cb59aa96e8dd1a135',
'6ef405e1f47a4a234f0bfeeb559d61c70031af6c47bf2e499a31002f845cd403',
'db7413e79e7df6de53d801e96e26c494caf1d5b709da85aa7bf04e42c76609f5',
'40eb44f0b586f2b41fb962adca2c1de351b70f3577a382798db2735564566a33',
'1ab21dal6f6d88107a5b067bal40929a47ae414b72e8bc17d774f281776cb2',
'3d7bdc0c99a5ba50cc74b9b9194fe4f2fdfa82a078c8afa561eec9afad052f9',
'4e00f88a86a5ae6a015d0a8027a280db2aebf13d4626875c877d23cca553d8d',
'64e6ea3476f64420c76db21b2fa4ec589def2c05ad97d2574ea63f8f0f7f22c9',
'7c5adce44d2415e1e2aa21006d6d22393466015a7ae14bf6bef76b28409c93c8',
'851a84fd2e74affd12c893e445ad66ec3a519ae6a2f4371887fe4d9605be5dd8',
'53370e3cade07cf9a34adf9212ec2b05b028989087d3ce9edb15d03f4acbb2c1',
'c5d8880344636ee5711a095f1090ba31057aa26238f94c1a16b137d4c25637ce',
'5eaa3e4ba898b43996288c6895e8cc38ab3ba6cf7f2ae68112d8b0b936f81e5',
'1ec543ca38666e906886b71f662f45f1605344410e26962a58668a7511ba65ce',
'b46fe64538510e926250a400f261c85d32eae75ef8b54fac9e089185362140ed',
'd924771fb6bc5ab4adb00053e9f0ea3396df9caebd24b3c25ec9223a90ebe4a2',
'f1ed482b393935fa0e7d7b4ad3facb731d9f37bd38df500a134049cba776d4ae',
'20d79c70f6b730dc04416e8b8b1e75c38ee1c10c57a8ca2081a33858ec08944a',
'000e252002f62f01e566e14c347e80f02722e57ba13e62a1805eb6e5eb30e23'
```

Problems

- Related to models:
 - VoxelNet² Implementation
 - Models failed to predict the small objects, for example pedestrians.

Average per class mean average precision = 0.09416772164775028
(`'animal'`, 0.0)
(`'bicycle'`, 0.0012601587242598114)
(`'bus'`, 0.0537961585349724)
(`'car'`, 0.5057703547292053)
(`'motorcycle'`, 0.0010937927262783702)
(`'other_vehicle'`, 0.18122473782388235)
(`'pedestrian'`, 0.002728630097923848)
(`'truck'`, 0.007467940545480116)

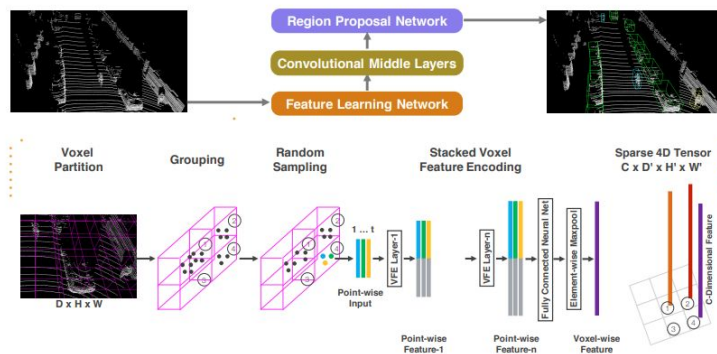


Fig 2. Architecture of VoxelNet²

² VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection by Zhou et al, <https://arxiv.org/abs/1711.06396>

What we learned Our Journey...

Object detection using torchvision

Greg Teichert
CSCAR Consultant

Explore MRI data

Image segmentation

- Video: MRI Data and Image Registration 3 min
- Video: Segmentation 3 min
- Lab: Get a sub-section 1h
- Reading: Convolutional Neural networks 10 min
- Video: 2D U-Net and 3D U-Net 2 min
- Reading: More about U-Net (Optional) 10 min
- Lab: Implement U-Net 1h
- Video: Data augmentation for segmentation

2D U-Net and 3D U-Net

U-Net

Segmentation Architecture

Perform Real-Time Object Detection with YOLOv3

by Coursera Project Network

Congratulations!

You've successfully completed **Perform Real-Time Object Detection with YOLOv3!**

Convolutional Neural Networks > Week 3 > YOLO Algorithm

Pre

Detection algorithms

- Video: Object Localization 11 min
- Video: Landmark Detection 5 min
- Video: Object Detection 5 min
- Video: Convolutional Implementation of Sliding Windows 11 min
- Reading: Convolutional Implementation of Sliding Windows "CORRECTION" 1 min
- Video: Bounding Box Prediction 14 min
- Video: Intersection Over Union 4 min

Object Detection

Putting it together:
YOLO algorithm

deeplearning.ai

Congratulations!

You've successfully completed Course 4 of 5: **Convolutional Neural Networks!**

[Read course reviews](#)

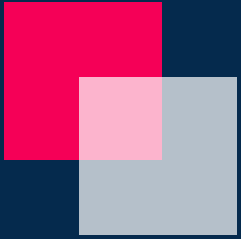
☆☆☆☆

[View your certificate](#)

Keep going!

Convolutional Neural Networks

by DeepLearning.AI



Questions?