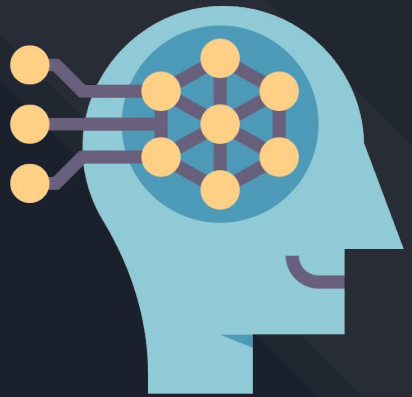
Takshak'21

Team ML_Enthusiasts Presentation (ISHANA):



Domain- ML/AI (Computer vision/ Perception)

Title- Optimal visual path planning for autonomous vehicles

Team Members: 1. Kshitij M Bhat (B.Tech 2nd year, IIT Indore)
2. Bhavya Dalal (B.Tech 2nd year, IIT Indore)
3. Tanishq Selot (B.Tech 2nd year, IIT Indore)

Basic Tools

Language - Python 3 (Jupyter Notebook)

Image processing - OpenCV, NumPy

Machine Learning - PyTorch

Path Planning - NetworkX, Scikit-Learn

Visualization - Matplotlib

Basic Outline

The task was to basically help the car reach its destination in the shortest time possible.

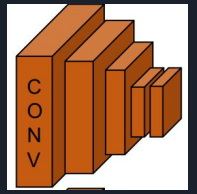
- **We divided our task into two sections -**
  1. Binary Semantic Segmentation (Road Extraction- labelling each pixel as road/non-road)
  2. Finding the shortest path between the coordinates given.

- **To build a model for road extraction, we referred to the paper Zhou *et. al.*, CVPR Workshops 2018[1]**

- **We referred to the DeepGlobe '18 dataset- which contains the ground truth mask data for 6.25k aerial images (1024x1024 px)[2].**

Road Extraction using D-linkNet[1]

Road extraction as a task has various challenges which are elegantly tackled by D-linkNet:



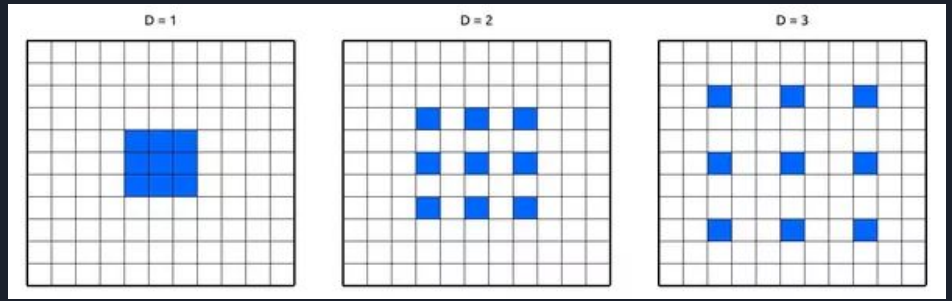
- As D-linkNet is an improved version of LinkNet, it has an additional part B in its architecture which helps in enlarging the receptive field and preserving the the detailed spatial information.

- The encoder ResNet34 follows 5 layers of downsampling (using max pooling) to yield a 32x32 px image.

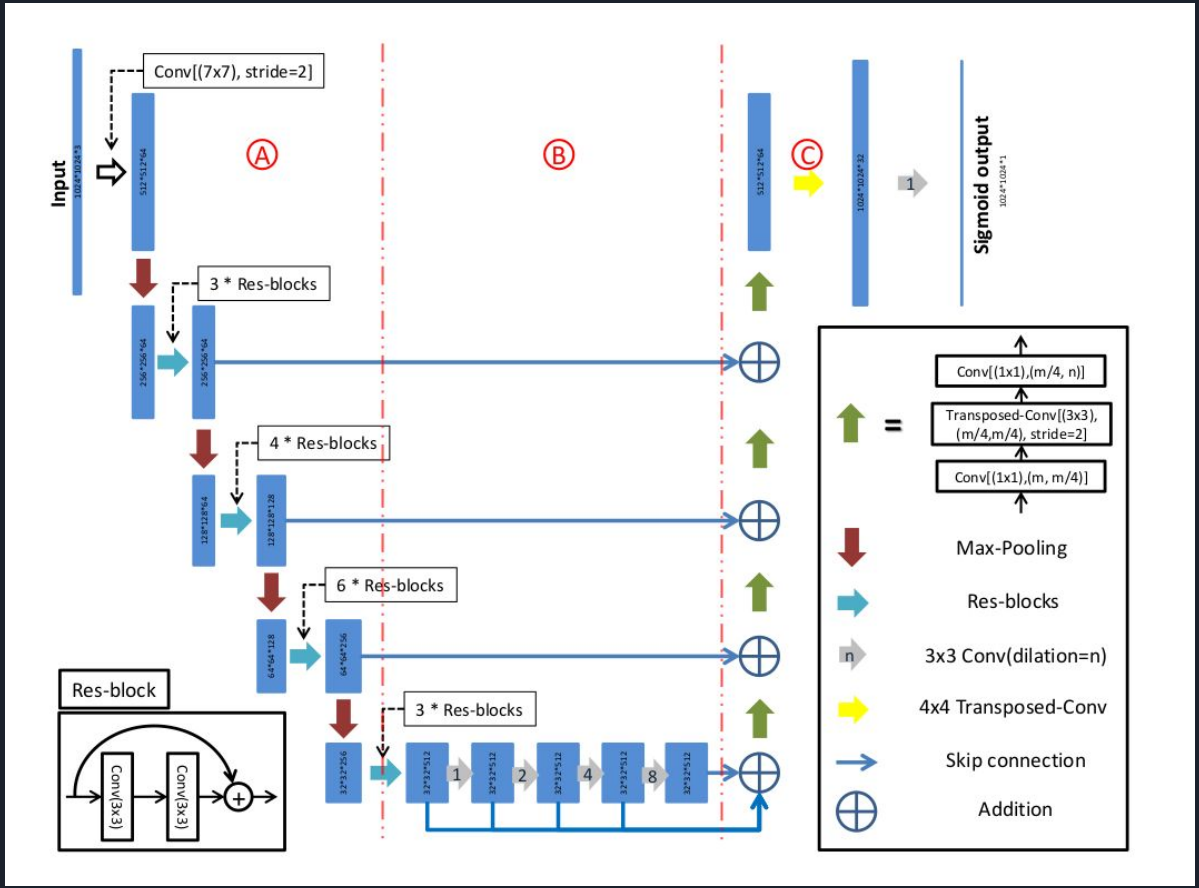
- Decoder same as in LinkNet was used for upsampling.

D-LinkNet34 (Centre part - The X-Factor)[1]

- **Dilated convolutional neural network** : The role of the ConvNet is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction.
- Using **pooling layers** could multiply increase the receptive field of feature points, but may reduce the resolution of center feature maps and drop spatial information. Hence, **DCNN** are a great alternative.
- They do this by **introducing holes** in the kernel thereby increasing the receptive field.



D-LinkNet Architecture[1]



Our Approach

Improving the output mask

The primary objective of this step is to avoid any pixel outside the road

Solution: A road narrowization function was implemented.

Path Planning

Optimal path generation using the binary occupancy grid

- Each of the road pixels were assigned a node in a weighted graph.
- The A* search algorithm was used for determining the shortest path between the specified start node and the end node.
- The k-D Tree Algorithm was used to find the nearest node to the start and end coordinates.

References

[1] L. Zhou, C. Zhang, and M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2018.

[2] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. arXiv preprint arXiv:1805.06561, 2018.