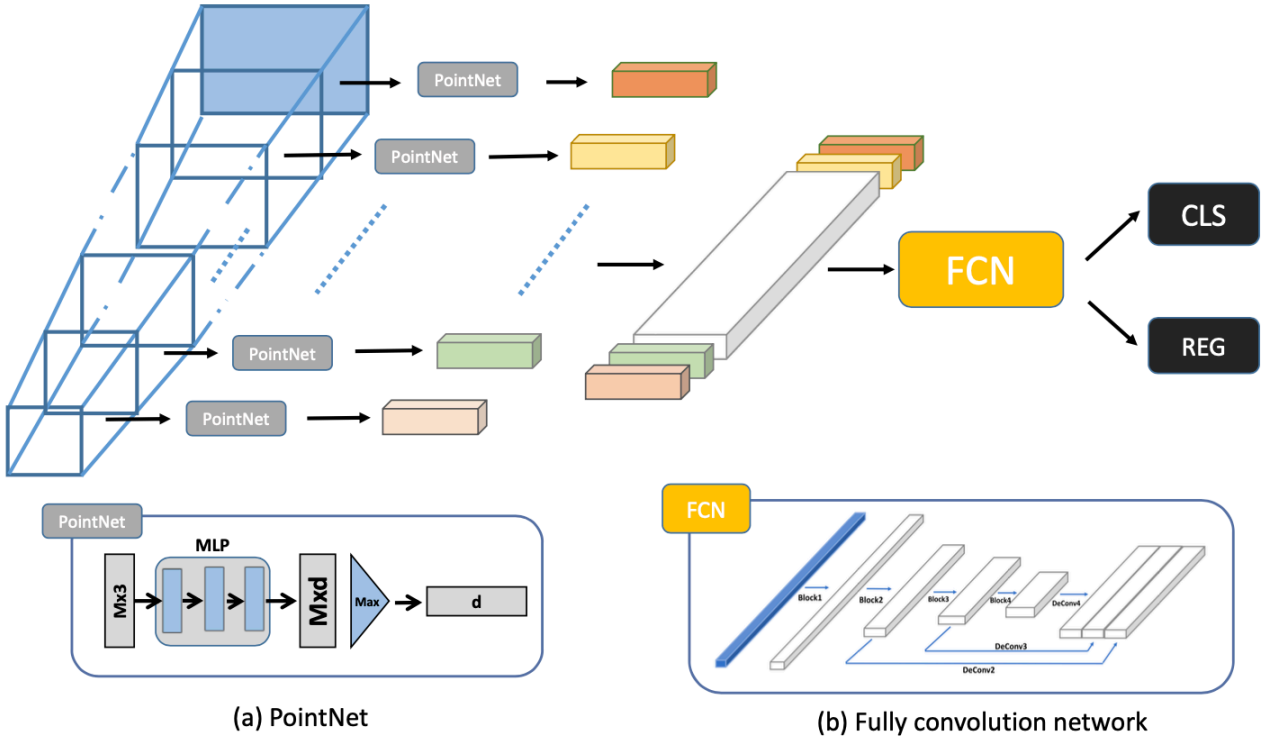
Lyft 3D Object Detection for Autonomous Vehicles

TEAM V

Model Selection

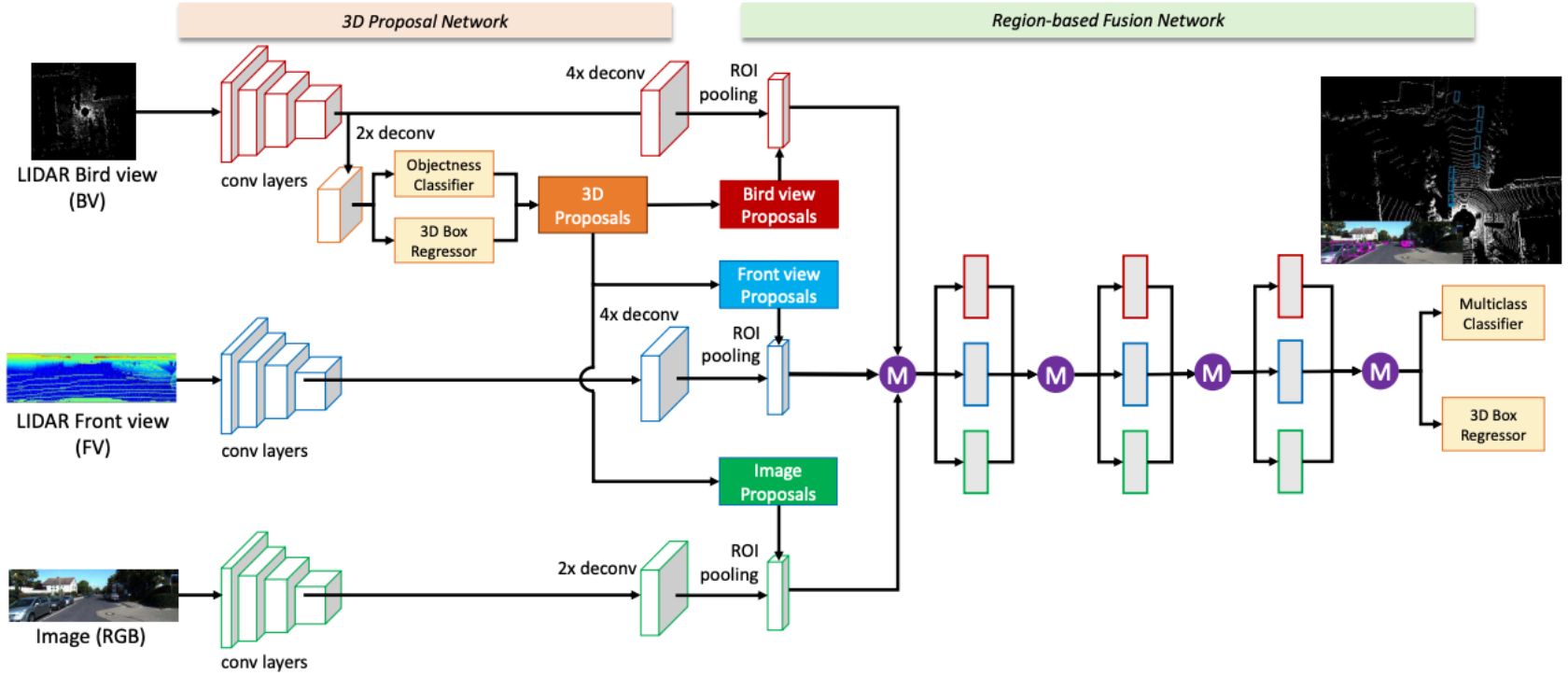
F-ConvNet (Frustum ConvNet)



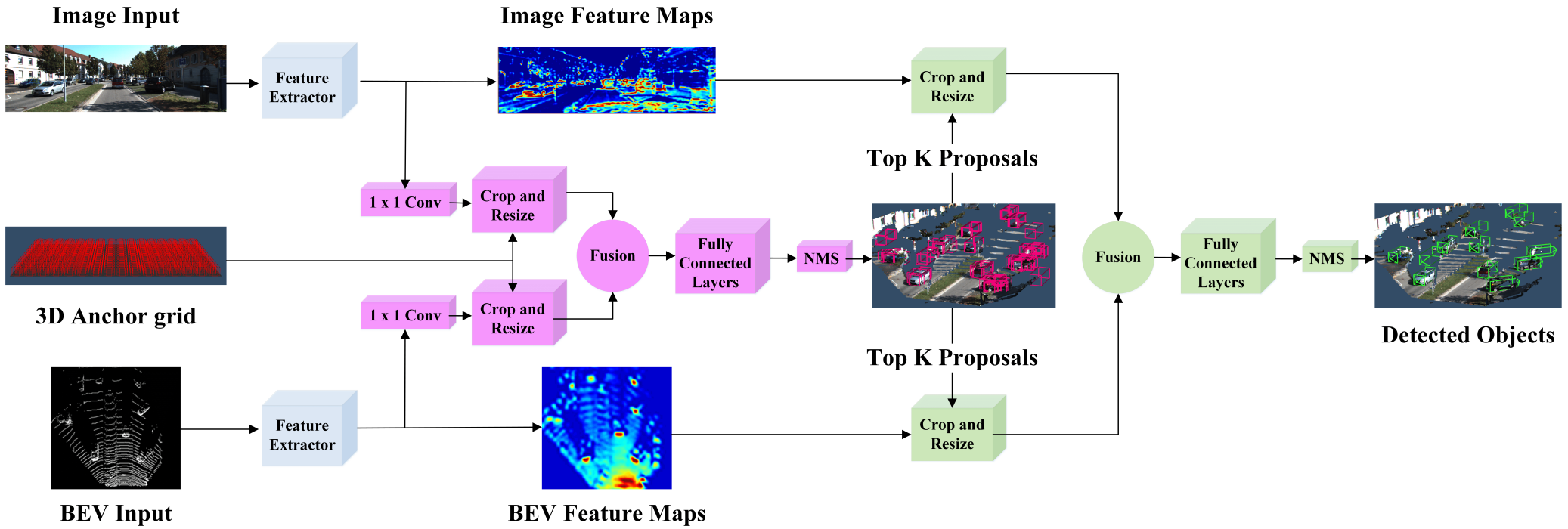
(a) PointNet

(b) Fully convolution network

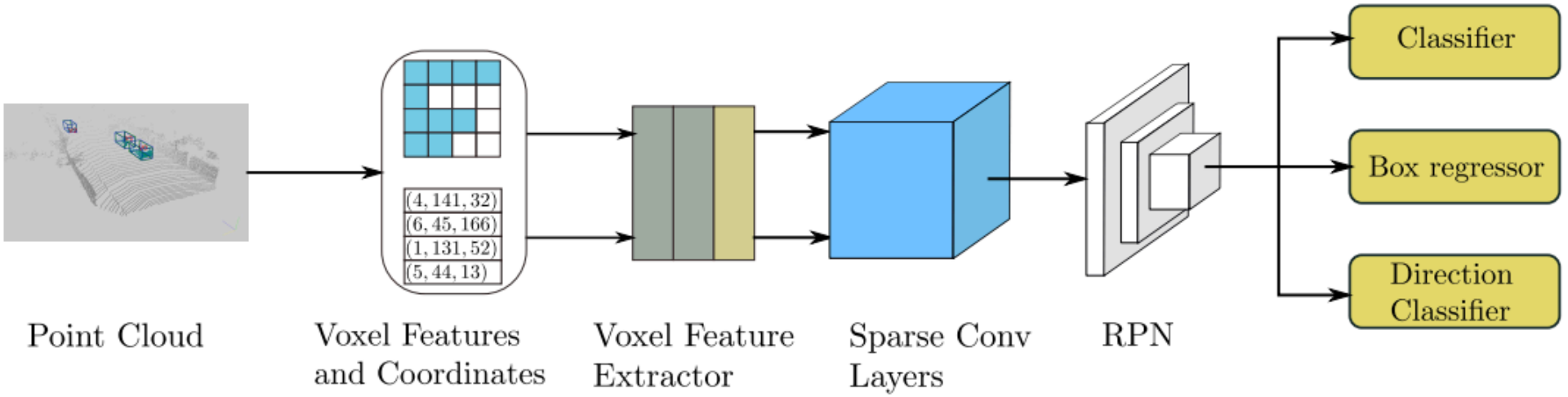
MV3D (Multi-View 3D Object Detection Network)



AVOD (Aggregate View Object Detection Network)



SECOND (Sparsely Embedded CONvolutional Detection)



Model Comparison

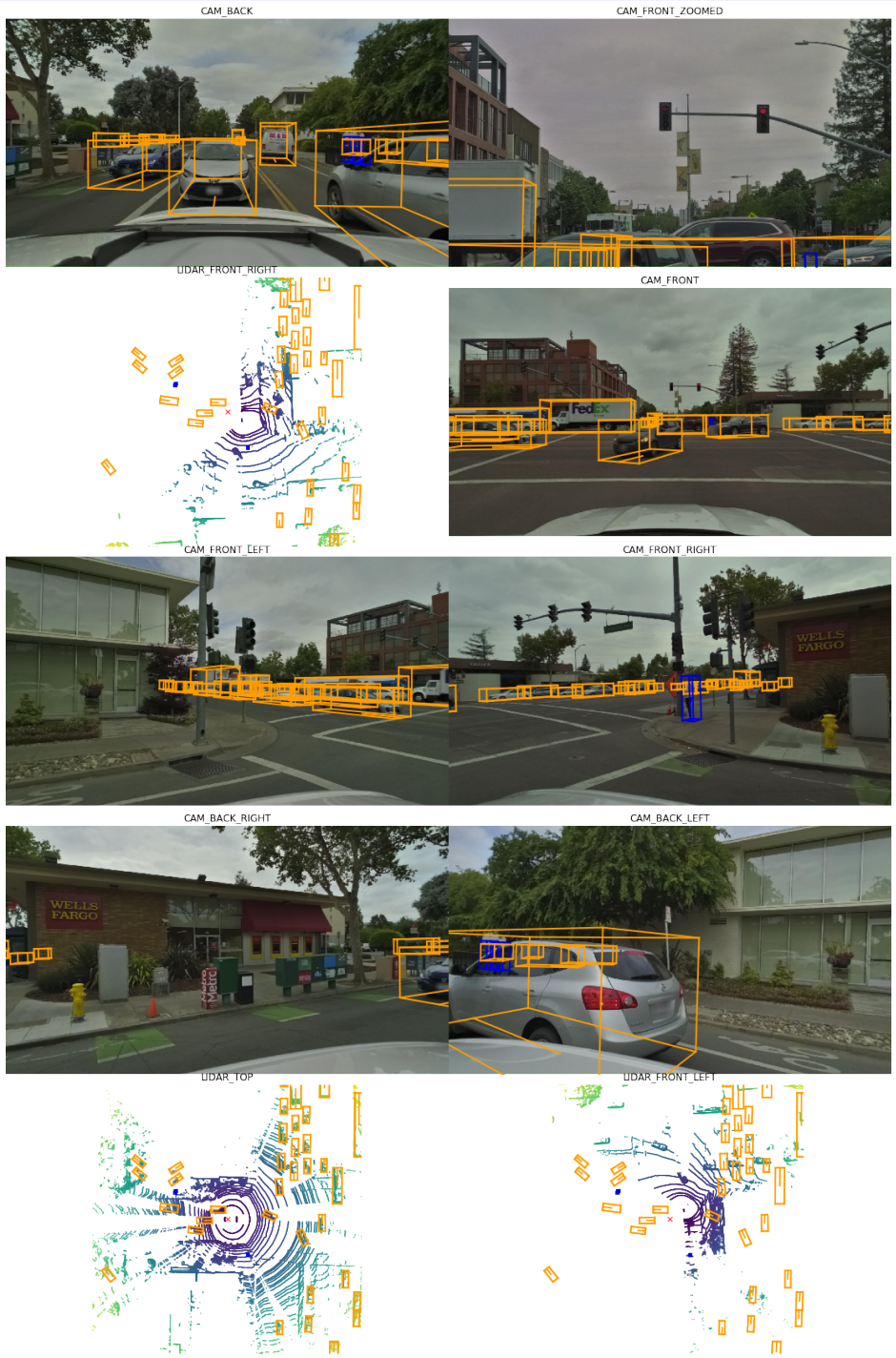
Model	Objects	Easy	Moderate	Hard	Runtime
F-ConvNet	Car	87.36 %	76.39 %	66.69 %	470ms
	Pedestrian	52.16 %	43.38 %	38.80 %	
	Cyclist	81.98 %	65.07 %	56.54 %	
MV3D	Car	74.97 %	63.63 %	54.00 %	360ms
AVOD	Car	76.39 %	66.47 %	60.23 %	80ms
	Pedestrian	36.10 %	27.86 %	25.76 %	
	Cyclist	57.19 %	42.08 %	38.29 %	
SECOND	Car	83.34 %	72.55 %	65.82 %	38ms
	Pedestrian	48.96 %	38.78 %	34.91 %	
	Cyclist	71.33 %	52.08 %	45.83 %	

Note:

- Results for object detection are given in terms of average precision (AP)
- Results from *http://www.cvlibs.net/datasets/kitti/eval_object.php*

Data Visualization

- 7 cameras
- 3 lidars



Modification

- **Step 1 – Multi-view mixing**

  - Compose 7 camera images in each scene to one as input image

  - Extend the input image and lidar to 7 images and 3 lidars

- **Step 2 – Dataset emerge**

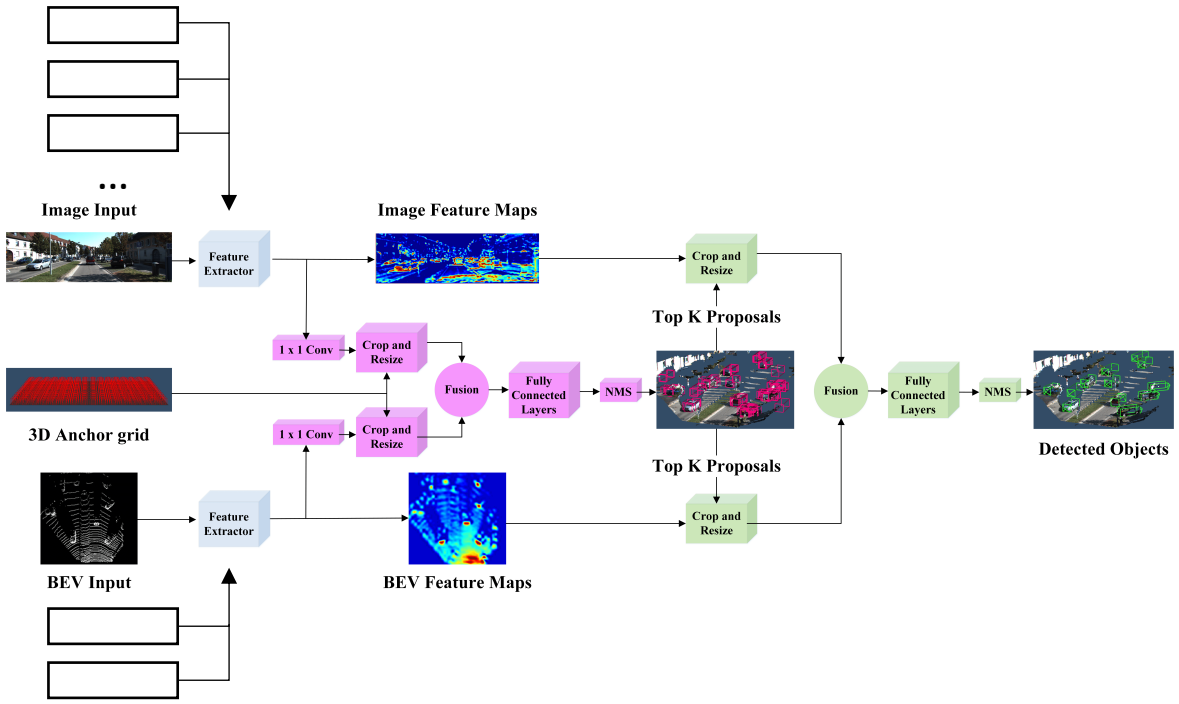
  - Emerge nuScenes dataset and Lyft dataset

- **Step 3 – Data volume selection**

  - Maybe select a part of training and test data from the whole dataset (100GB) to train and test the model

Multi-view mixing

- Extend input one image and lidar to 7 images and 3 lidars



Multi-view mixing

- Compose 7 camera images in each scene to one as input image