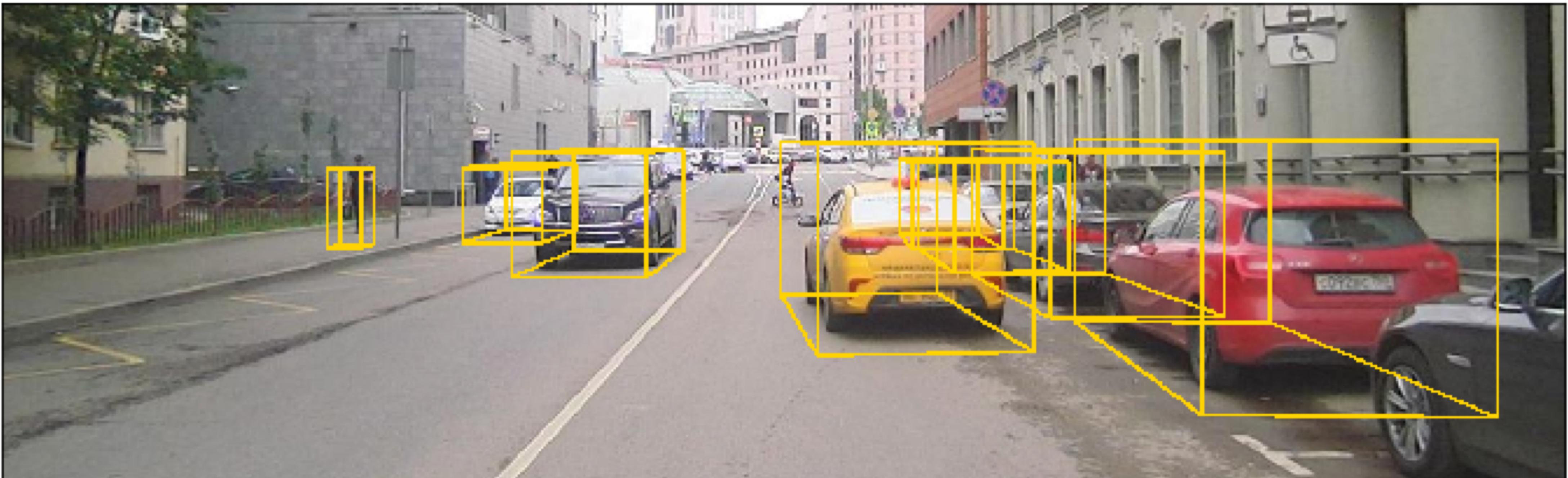


Обзор 3D детекторов

Артём Филатов
Yandex Self-Driving Team

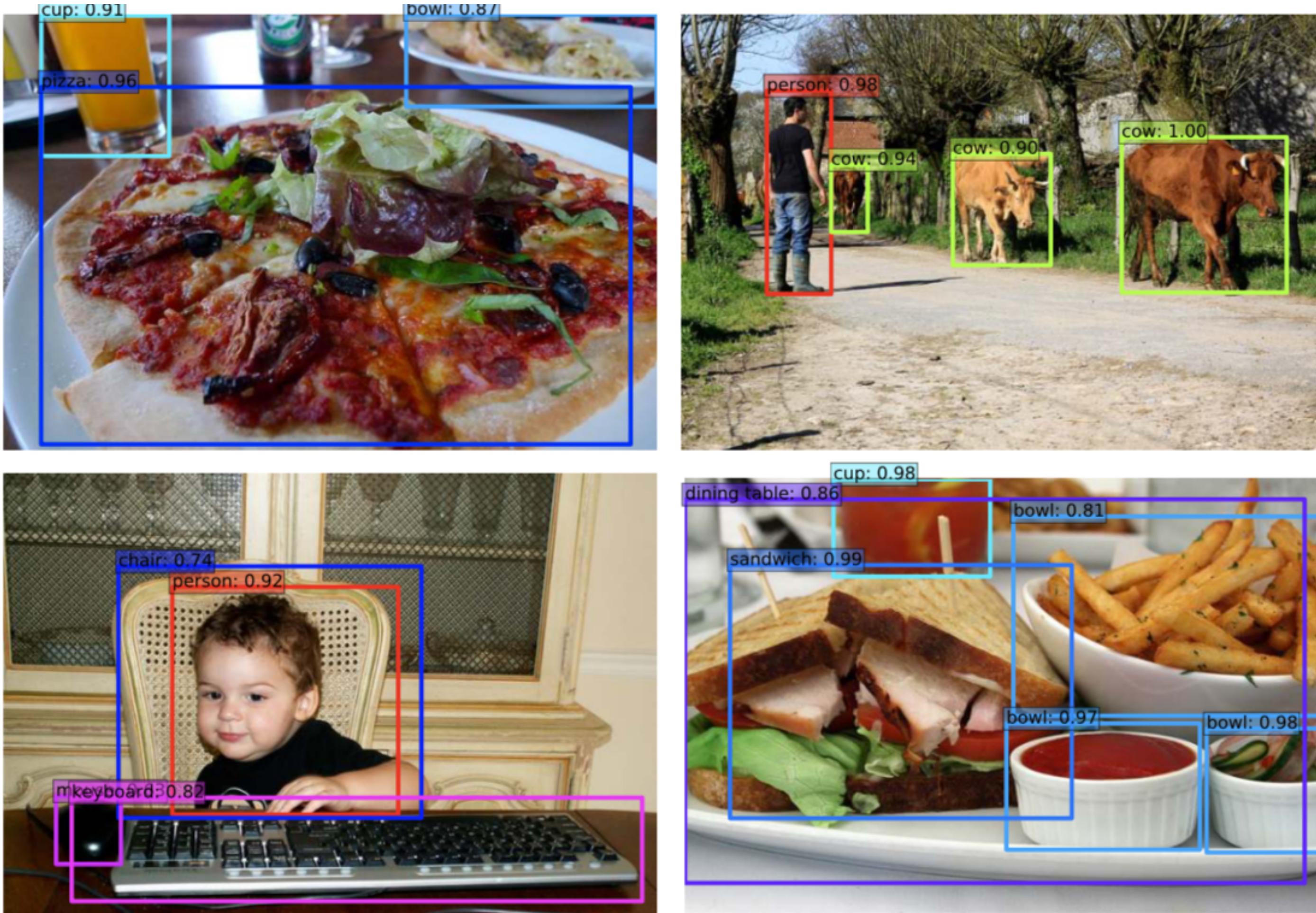
Цель

Хотим поставить коробку на каждого участника движения



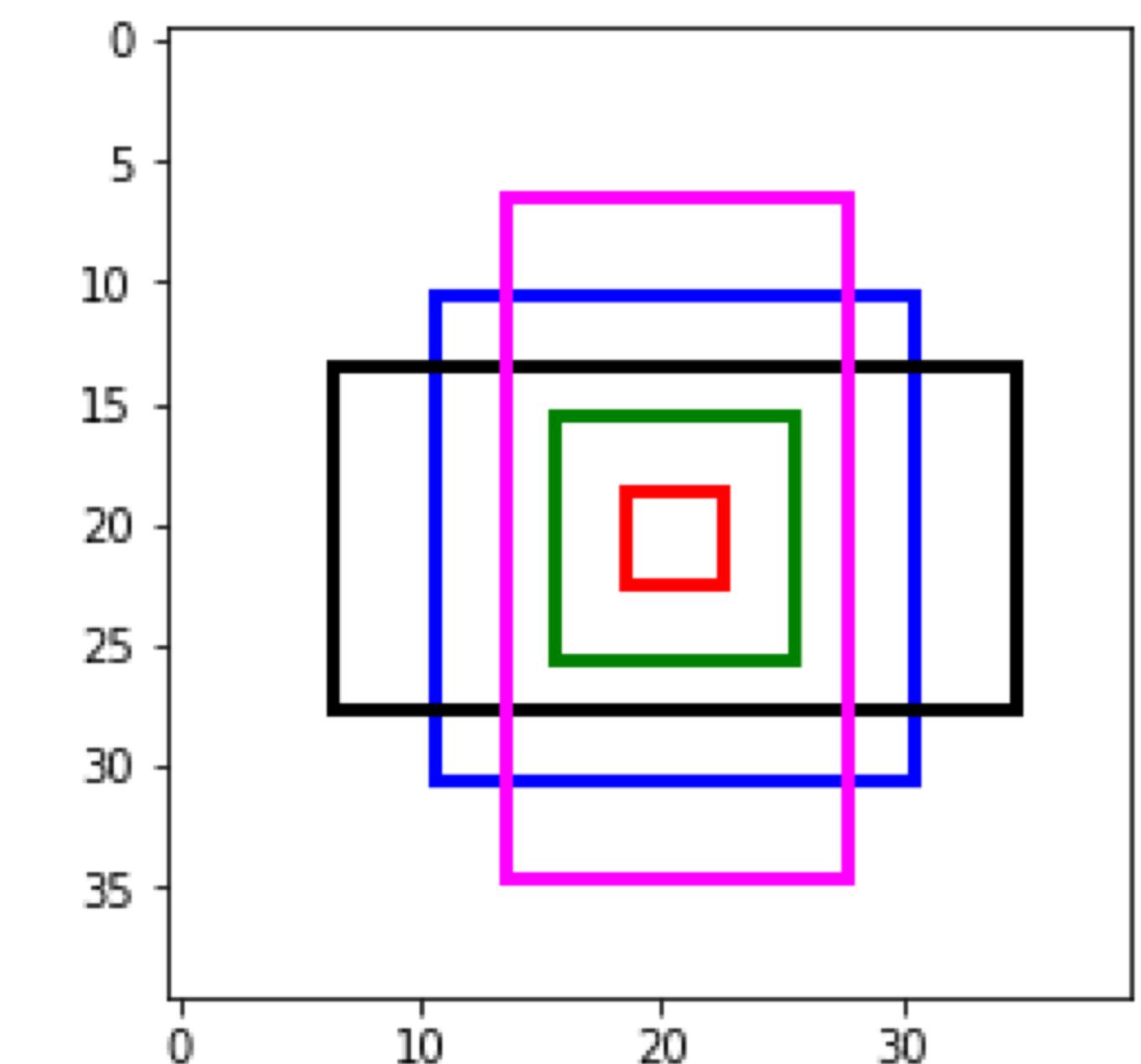
Но как? Обратимся к 2D!

SSD



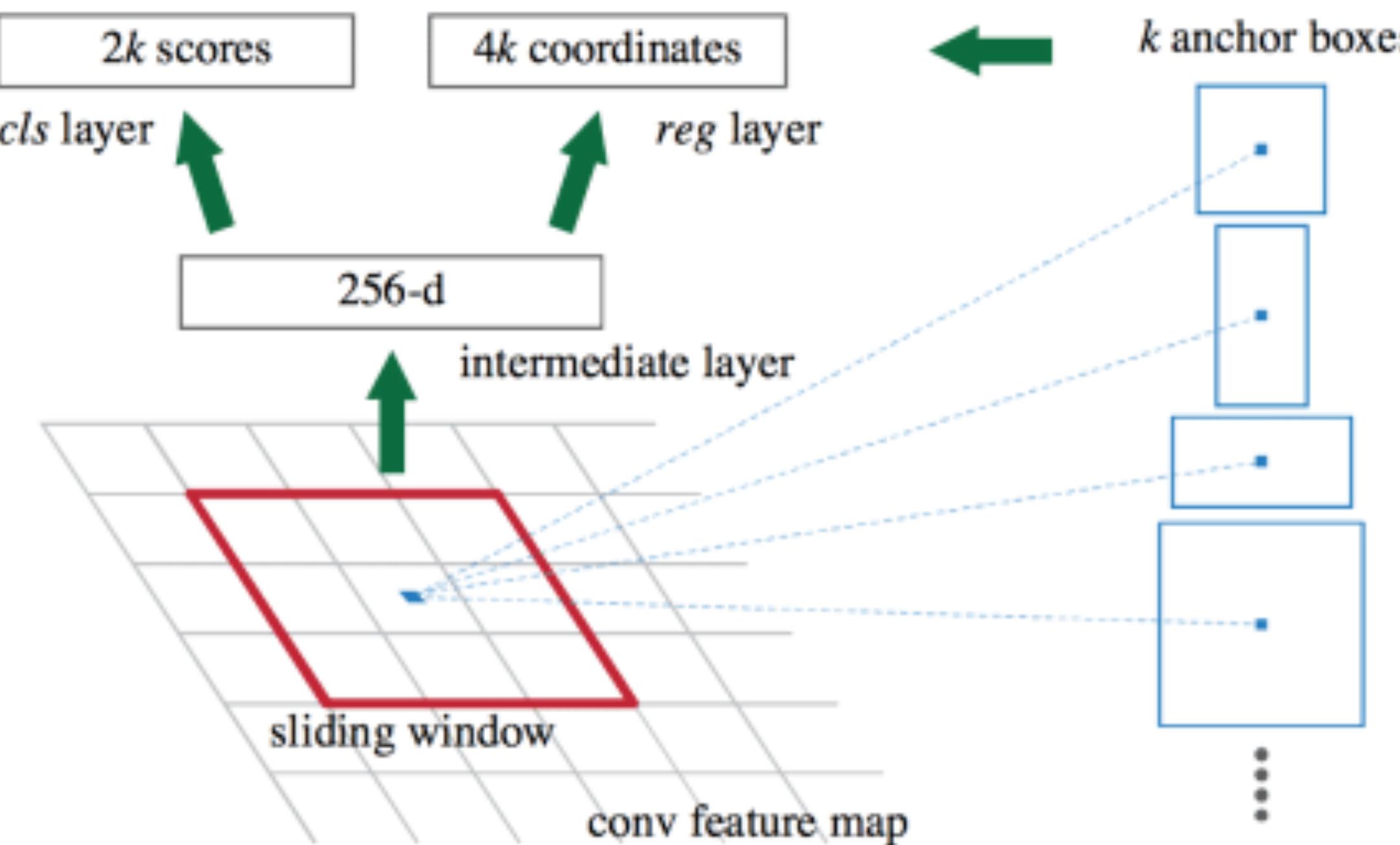
Anchor Boxes

- › Каким образом мы будем предсказывать коробки?
- › Давайте зафиксируем конечный набор коробок на картинке
- › Для каждой такой коробки (anchor box) мы будем предсказывать сдвиг в сторону ground truth
- › Для построения обучающей выборки для каждой anchor box вычисляется сдвиг в сторону ближайшей ground truth box



RPN: Region Proposal Network

- › Предскажем для каждой коробки вероятность того, что она принадлежит ground truth и сдвиг для более точной локализации

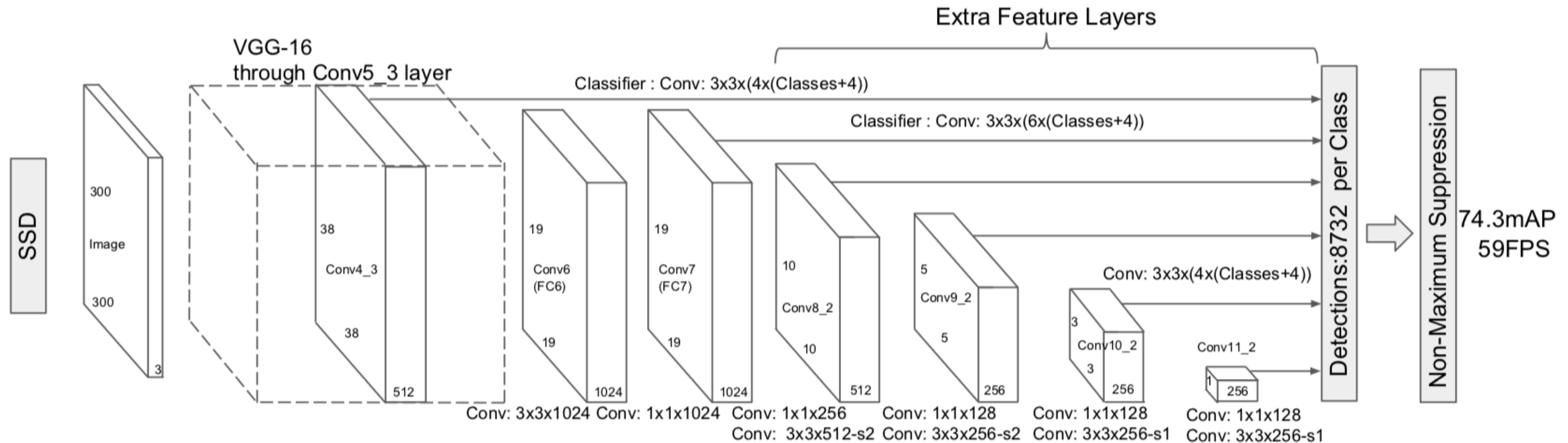


NMS

- › К одному ground truth объекту могло быть прибито несколько якорей
- › Следовательно, от RPN мы можем ожидать аналогичного выхода
- › Что делать?



SSD



Results

| Method | data | Avg. Precision, IoU: 0.5:0.95 0.5 0.75 | | | Avg. Precision, Area: S M L | | | Avg. Recall, #Dets: 1 10 100 | | | Avg. Recall, Area: S M L | | |
|-------------|-------------|---|-------------|-------------|--------------------------------|-------------|-------------|---------------------------------|-------------|-------------|-----------------------------|-------------|-------------|
| | | - | - | - | - | - | - | - | - | - | - | - | - |
| Fast [6] | train | 19.7 | 35.9 | - | - | - | - | - | - | - | - | - | - |
| Fast [24] | train | 20.5 | 39.9 | 19.4 | 4.1 | 20.0 | 35.8 | 21.3 | 29.5 | 30.1 | 7.3 | 32.1 | 52.0 |
| Faster [2] | trainval | 21.9 | 42.7 | - | - | - | - | - | - | - | - | - | - |
| ION [24] | train | 23.6 | 43.2 | 23.6 | 6.4 | 24.1 | 38.3 | 23.2 | 32.7 | 33.5 | 10.1 | 37.7 | 53.6 |
| Faster [25] | trainval | 24.2 | 45.3 | 23.5 | 7.7 | 26.4 | 37.1 | 23.8 | 34.0 | 34.6 | 12.0 | 38.5 | 54.4 |
| SSD300 | trainval35k | 23.2 | 41.2 | 23.4 | 5.3 | 23.2 | 39.6 | 22.5 | 33.2 | 35.3 | 9.6 | 37.6 | 56.5 |
| SSD512 | trainval35k | 26.8 | 46.5 | 27.8 | 9.0 | 28.9 | 41.9 | 24.8 | 37.5 | 39.8 | 14.0 | 43.5 | 59.0 |

Table 5: COCO test-dev2015 detection results.

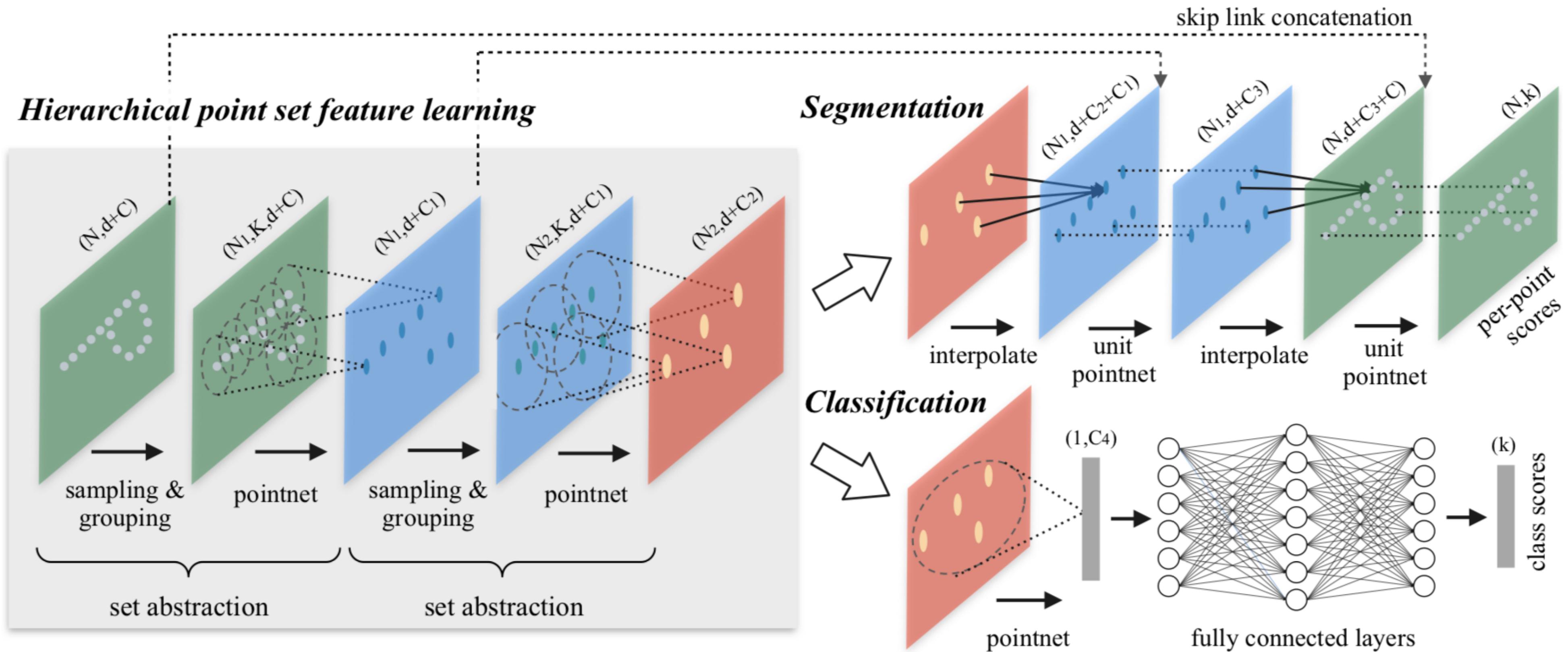
Перейдем в 3D!

- › Воспользуемся идеями SSD
- › Но для начала нам нужно научится вытаскивать фичи из облака точек

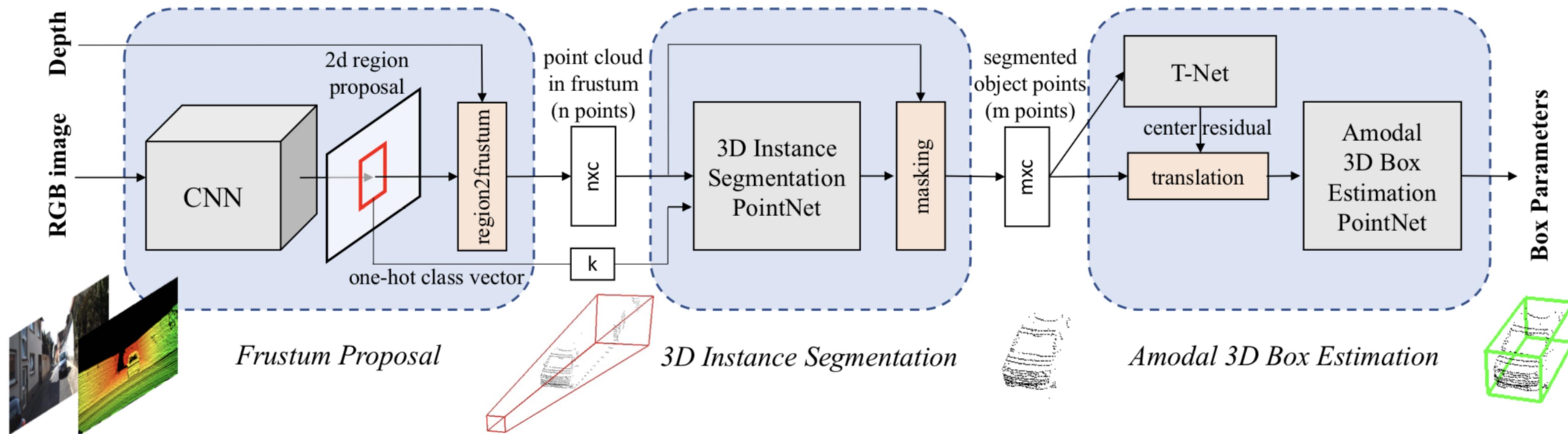
Облако точек



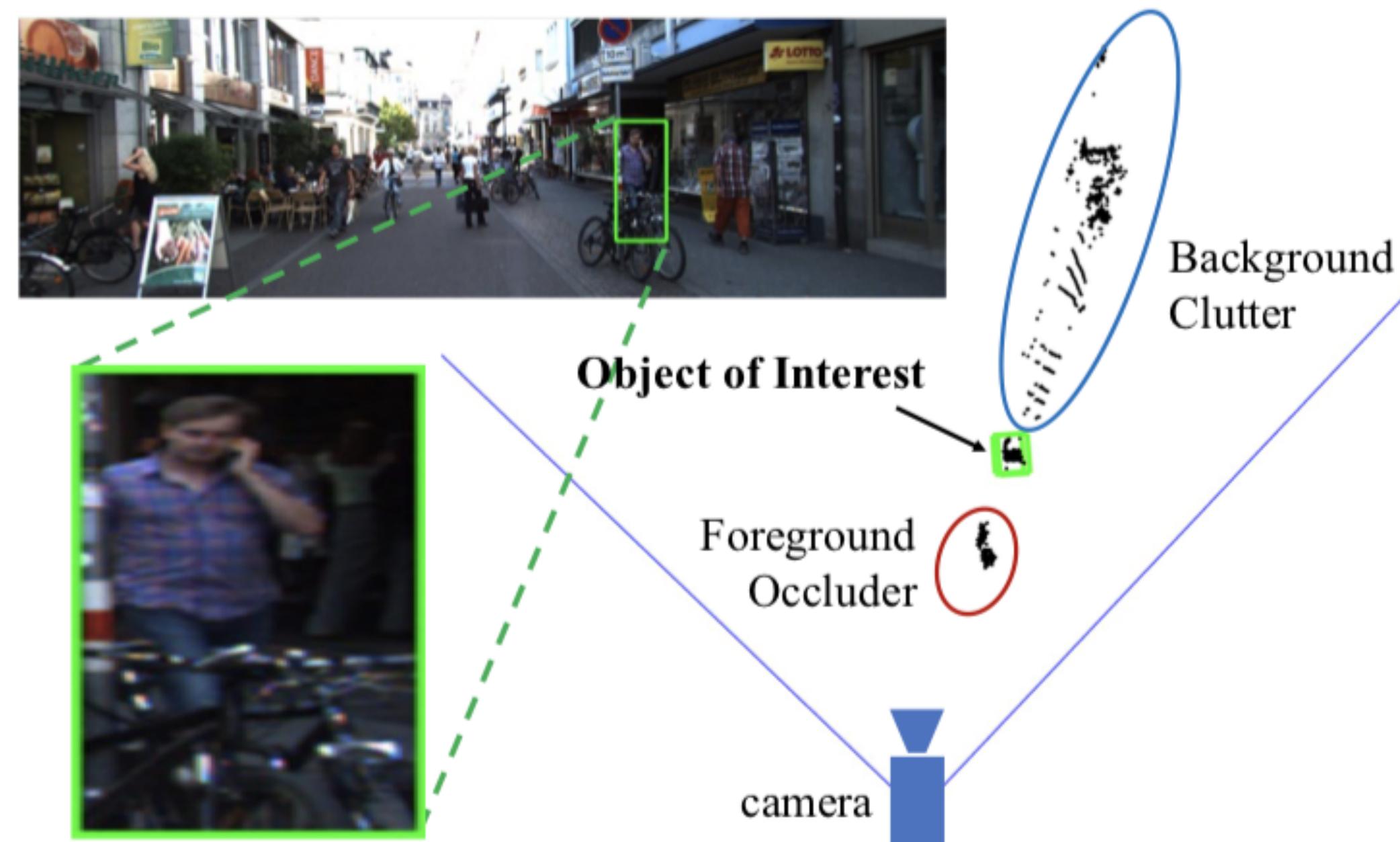
PointNet++

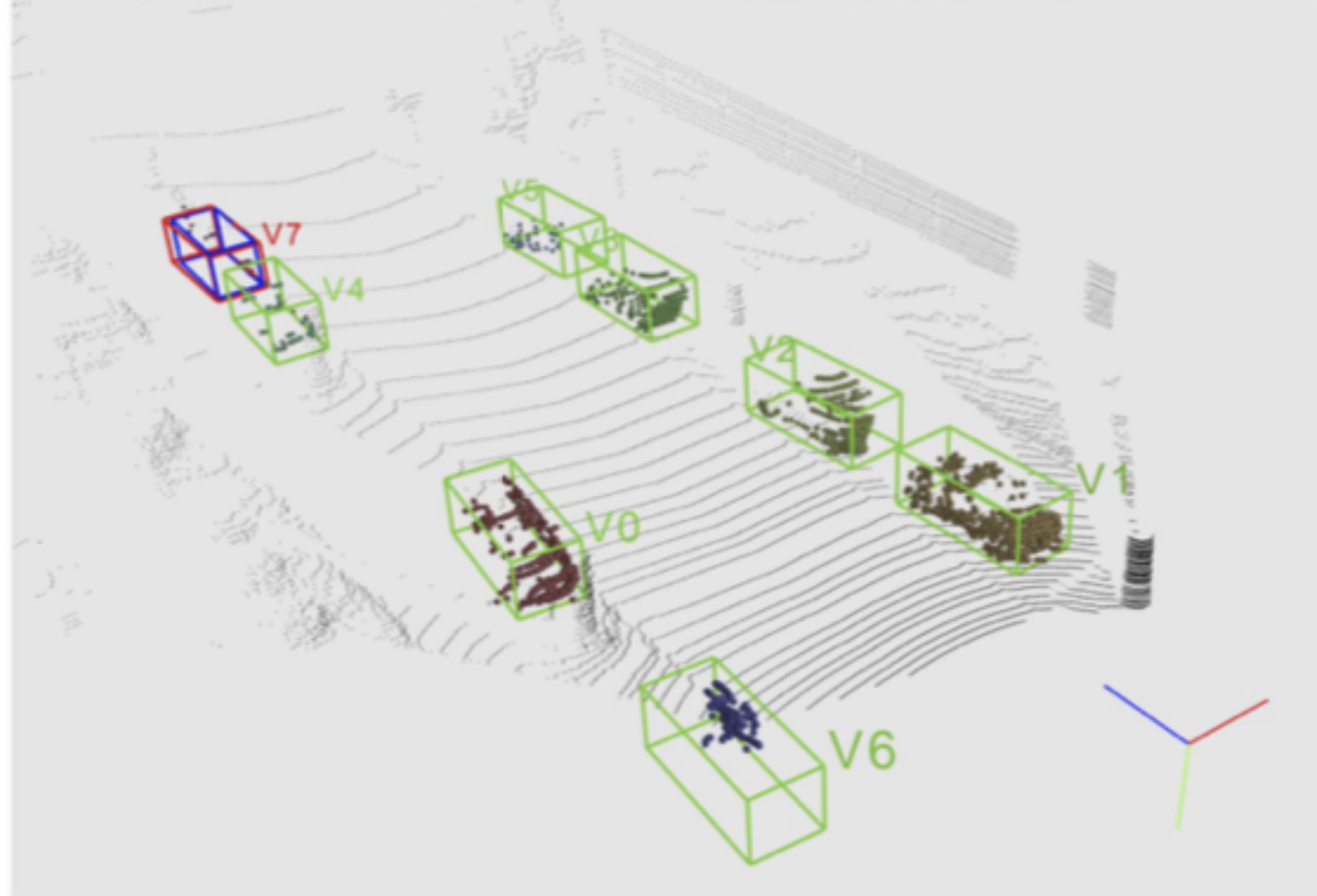
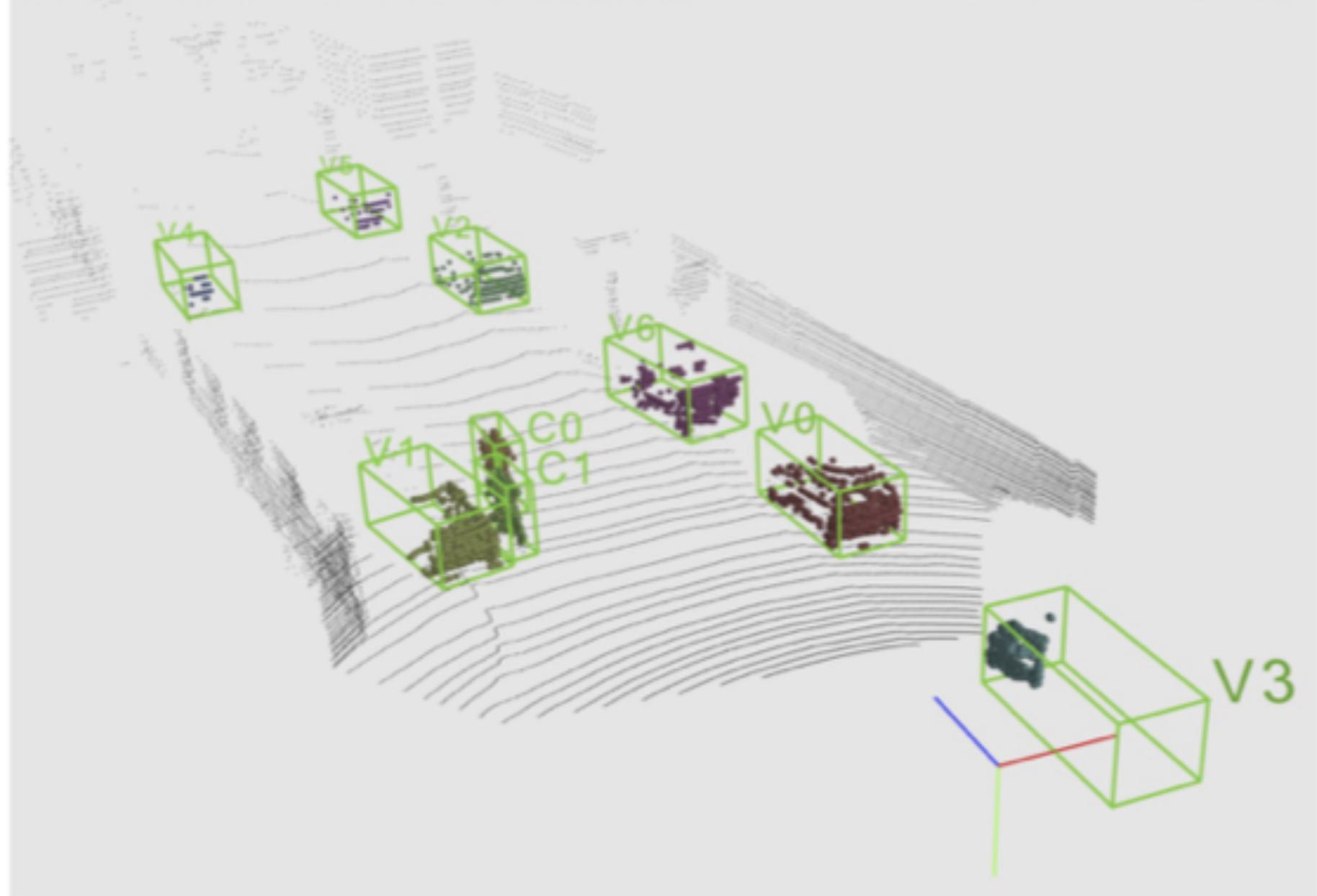
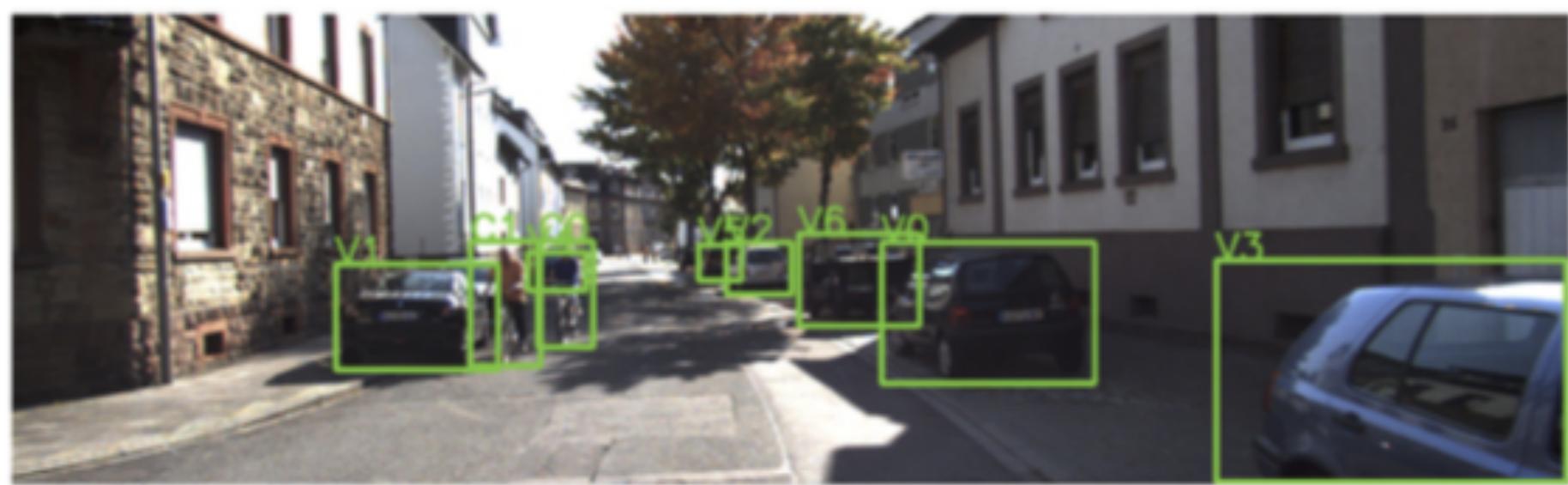


Frustum

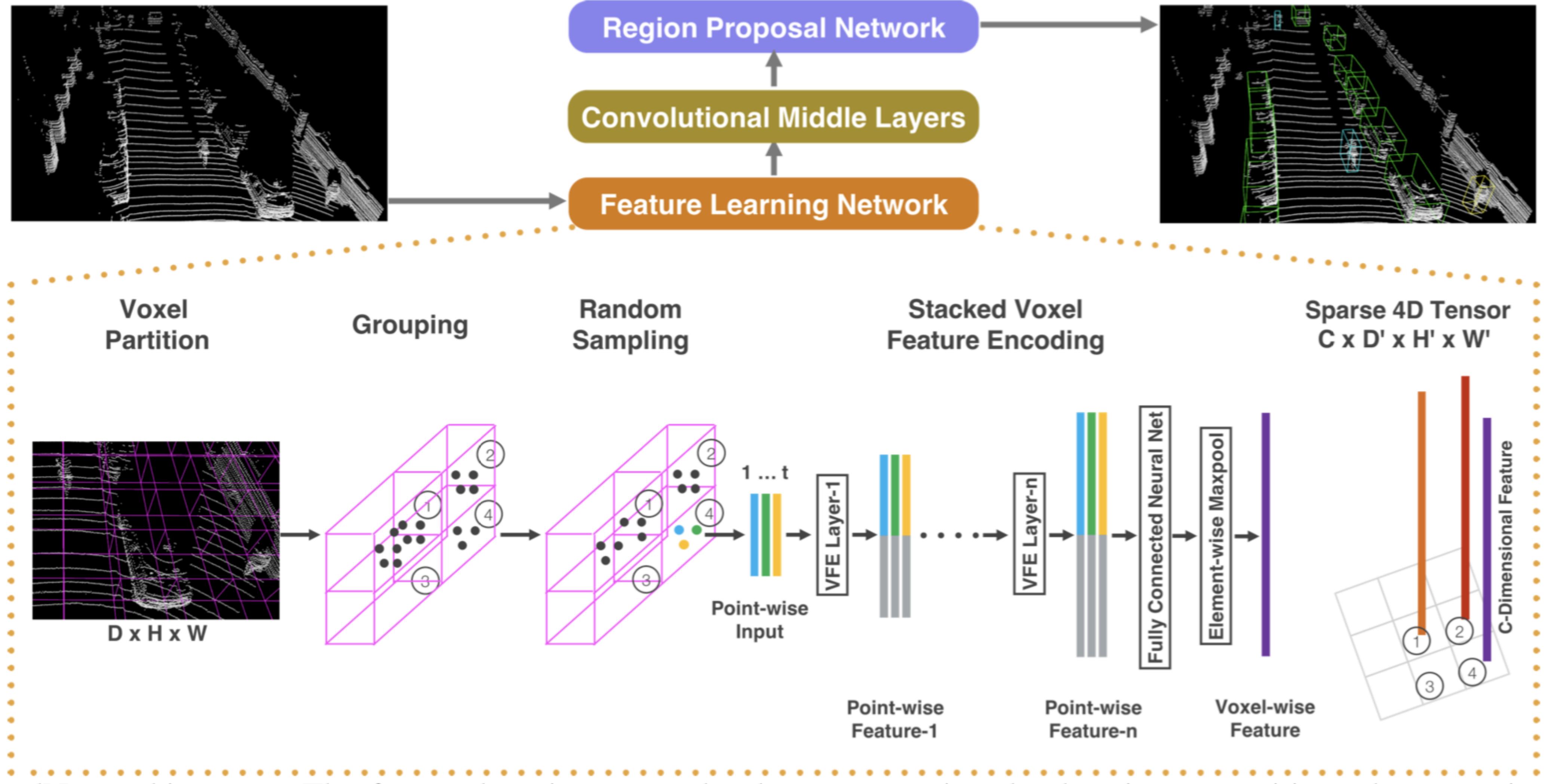


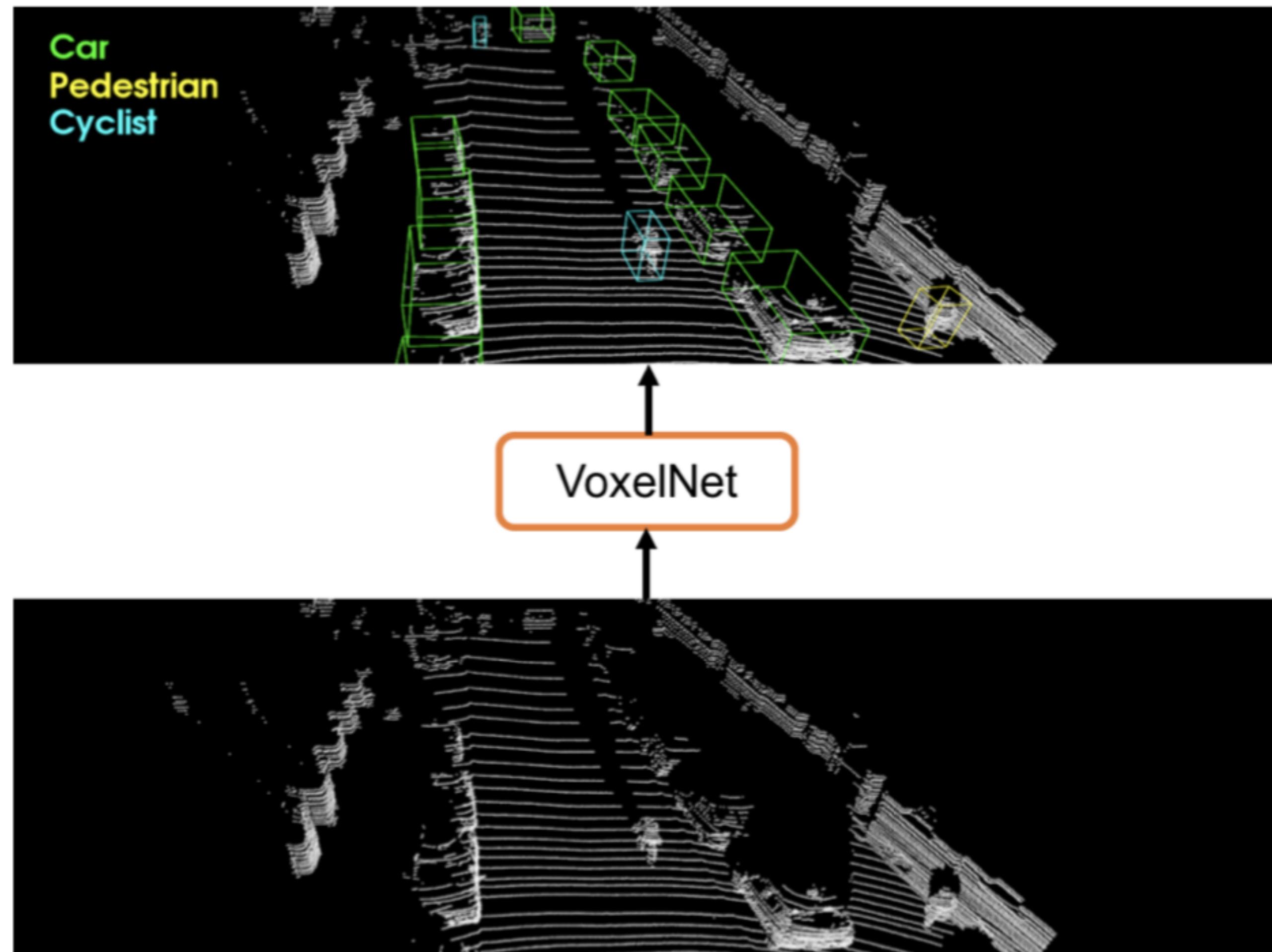
Challenges





VoxelNet





Results

| Method | Modality | Car | | | Pedestrian | | | Cyclist | | |
|--------------------|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | Easy | Moderate | Hard | Easy | Moderate | Hard | Easy | Moderate | Hard |
| Mono3D [3] | Mono | 5.22 | 5.19 | 4.13 | N/A | N/A | N/A | N/A | N/A | N/A |
| 3DOP [4] | Stereo | 12.63 | 9.49 | 7.59 | N/A | N/A | N/A | N/A | N/A | N/A |
| VeloFCN [22] | LiDAR | 40.14 | 32.08 | 30.47 | N/A | N/A | N/A | N/A | N/A | N/A |
| MV (BV+FV) [5] | LiDAR | 86.18 | 77.32 | 76.33 | N/A | N/A | N/A | N/A | N/A | N/A |
| MV (BV+FV+RGB) [5] | LiDAR+Mono | 86.55 | 78.10 | 76.67 | N/A | N/A | N/A | N/A | N/A | N/A |
| HC-baseline | LiDAR | 88.26 | 78.42 | 77.66 | 58.96 | 53.79 | 51.47 | 63.63 | 42.75 | 41.06 |
| VoxelNet | LiDAR | 89.60 | 84.81 | 78.57 | 65.95 | 61.05 | 56.98 | 74.41 | 52.18 | 50.49 |

Table 1. Performance comparison in bird's eye view detection: average precision (in %) on KITTI validation set.