

**Распознавание трехмерных объектов дорожной сцены по данным  
бортовых камер и лидаров автомобиля**  
**И. И. Царин, Д. А. Юдин**

Московский физико-технический институт (национальный исследовательский  
университет)

Автоматизированные транспортные системы, включая беспилотные автомобили, становятся все более актуальными в современном мире, представляя собой перспективное направление развития транспортной отрасли. Для обеспечения безопасности и эффективности таких систем необходимо развивать и совершенствовать модели 3D детекции на лидарных облаках точек. В данной научной статье мы рассматриваем современные модели 3D детекции, сосредотачиваясь на модели VoxelNeXt, и исследуем их практическое применение в контексте построения беспилотного автомобиля.

В данном исследовании было проведено сравнение модели VoxelNeXt[1] с популярными алгоритмами 3D детекции, такими как PointPillars[2], SECOND и CenterPoint[3] на датасете nuScenes. Результаты экспериментов показали, что модель VoxelNeXt продемонстрировала самое высокое качество детекции объектов в трехмерном пространстве по сравнению с другими алгоритмами (рис. 1). Одновременно она обладает хорошей производительностью, достигая времени обработки 77 миллисекунд на графическом процессоре RTX3060Ti (рис. 2).

Рис. 1, сравнение качества работы алгоритмов детекции на датасете nuScenes

	mATE	mASE	mAOE	mAVE	mAAE	mAP	NDS
<a href="#">PointPillar-MultiHead</a>	33.87	26.00	32.07	28.74	20.15	44.63	58.23
<a href="#">SECOND-MultiHead (CBGS)</a>	31.15	25.51	26.64	26.26	20.46	50.59	62.29
<a href="#">CenterPoint-PointPillar</a>	31.13	26.04	42.92	23.90	19.14	50.03	60.70
<a href="#">CenterPoint (voxel_size=0.1)</a>	30.11	25.55	38.28	21.94	18.87	56.03	64.54
<a href="#">CenterPoint (voxel_size=0.075)</a>	28.80	25.43	37.27	21.55	18.24	59.22	66.48
<a href="#">VoxelNeXt (voxel_size=0.075)</a>	30.11	25.23	40.57	21.69	18.56	60.53	66.65

Рис. 2, сравнение скорости работы алгоритмов детекции на RTX3060Ti

checkpoint name	mean inference time	std inference time	основано на
<a href="#"><u>cbgs_voxel0075_voxelnext</u></a>	76.65	22.44	voxelnext
<a href="#"><u>cbgs_voxel0075_voxelnext_doubleflip</u></a>	254.25	50.96	voxelnext
<a href="#"><u>cbgs_dyn_pp_centerpoint</u></a>	34.88	12.76	pointpillar
<a href="#"><u>cbgs_pillar0075_res2d_centerpoint</u></a>	82.11	12.76	pillarnet
<a href="#"><u>cbgs_pp_multihead</u></a>	31.97	7.41	pointpillar
<a href="#"><u>cbgs_second_multihead</u></a>	50.32	12.71	second
<a href="#"><u>cbgs_voxel0075_res3d_centerpoint</u></a>	74.71	15.99	voxelnet + centerpoint?
<a href="#"><u>cbgs_voxel0075_voxelnext2d</u></a>	96.91	21.80	voxelnext

Кроме этого, проводились эксперименты по ускорению работы VoxelNeXt. Популярный в индустрии способ ускорения с помощью конвертации нейросети в TensorRT или ONNX тут оказался неприменим из-за использования submanifold sparse convolutions в архитектуре модели. Однако, удалось добиться ускорения скорости инференса, преобразовав модель в формат FP16 (77ms -> 66ms). Этот шаг позволил значительно увеличить скорость обработки данных, что сделало модель еще более эффективной для решения задач детекции объектов на лидарных облаках точек. Также был написан код для визуализации предсказаний, что позволяет наглядно отображать результаты работы модели.

### Список литературы

1. Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, Jiaya Jia VoxelNeXt: Fully Sparse VoxelNet for 3D Object Detection and Tracking  
arXiv:2303.11301v1
2. Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, Oscar Beijbom PointPillars: Fast Encoders for Object Detection from Point Clouds  
arXiv:1812.05784v2
3. Tianwei Yin, Xingyi Zhou, Philipp Krähenbühl Center-based 3D Object Detection and Tracking arXiv:2006.11275v2