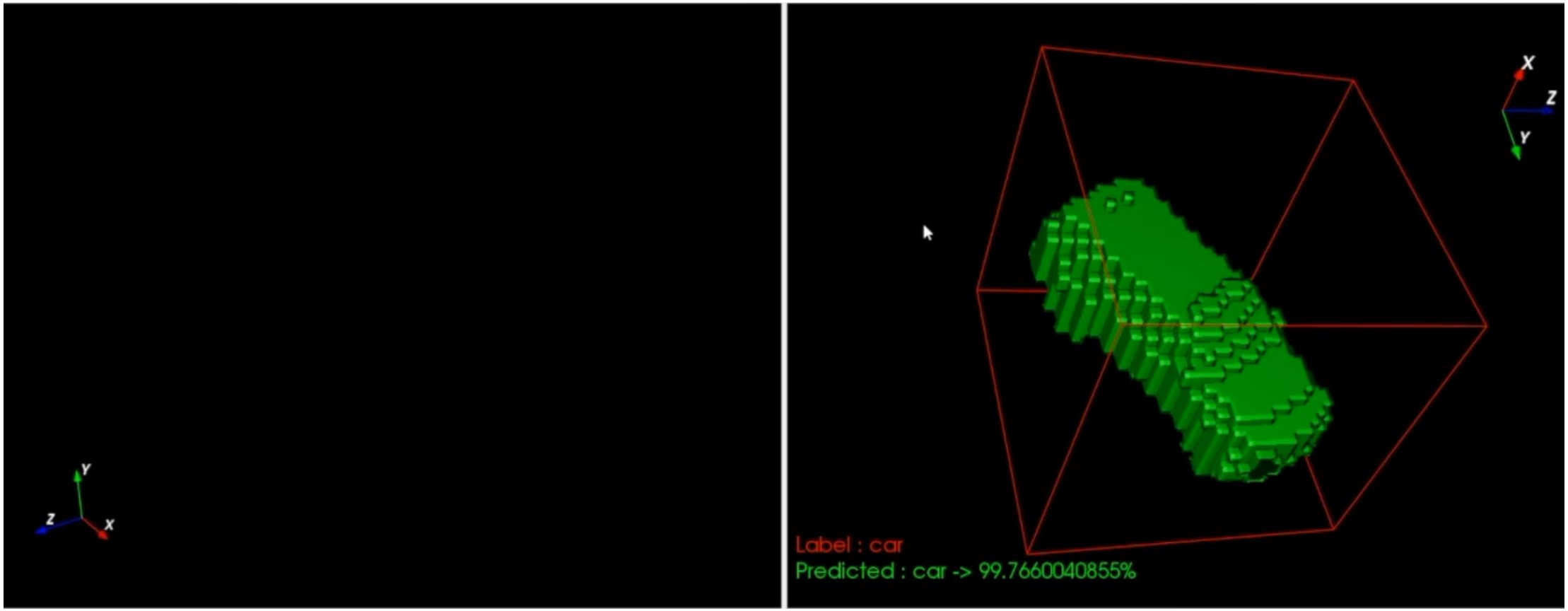
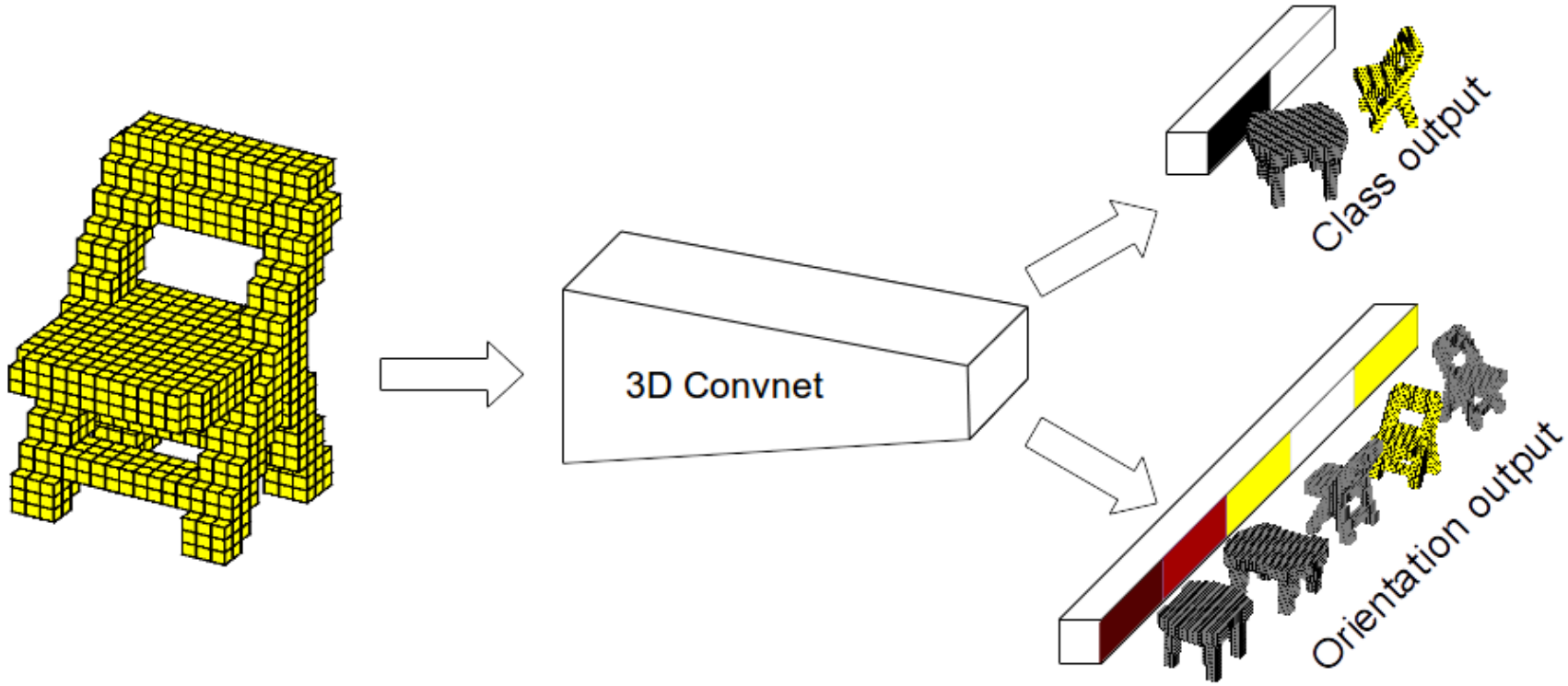
Orientation-boosted Voxel Nets for 3D Object Recognition

Object identification and 3D CNN



category-level classification task as a multi-task problem



Method

- ORIentation-boosted vOxel Net- ORION
  - Z-axis
- Sensors:
  - LiDAR (3D Laser scanning)
  - RGBD cameras (Depth as a fourth input)
  - Synthetic/CAD
- Experiments:
  - Detection
  - Classification

ORION

- https://github.com/lmb-freiburg/orion

- Language: Mathlab
- Built on the Voxnet architecture

Datasett

- Sydney Urban Objects - LiDAR/Pointcloud
- NYUv2 - Kinect/RGBD
- ModelNet - Synthetic/CAD
- KITTI - LiDAR/Pointcloud

Sydney Urban Objects - LiDAR/Pointcloud

- LiDAR scans of 631 objects in 26 categories
- Incomplete point-clouds
  - Single viewpoint
- Converted to voxel-grids of size 32x32x32

NYUv2 - Kinect/RGBD

- 2808 RGBD images, 10 object classes
- Converted to voxel-grids of size 32x32x32
- Orientation annotation from SUN-RGBD benchmark

ModelNet - Synthetic/CAD

- synthetic CAD models
- uniformly aligned objects of the same classes
- voxel grids of size 28x28x28
- Manual annotation and automated annotations

KITTI - LiDAR/Pointcloud

- Object detection: 7481 training images and 7518 test images
- Only used for the detection experiment

Classification results

Method↓				Dataset		
		# Conv	# param	Sydney	NYUv2	ModelNet10
Hand-crafted feat.	Recursive D [31]	-	-	-	37.6	-
	Recursive D+C [31]	-	-	-	44.8	-
	Triangle+SVM [7]	-	-	67.1	-	-
	GFH+SVM [4]	-	-	71.0	-	-
Deep Network	FusionNet [13]		118M	-	-	93.1
	VRN† [2]	43	18M	-	-	93.6
Shallow Network	ShapeNet [38]	3	-	-	57.9	83.5
	DeepPano [29]	4	-	-	-	85.5
	VoxNet [21] (baseline)	2	890K	72	71	92
	ORION (Ours)	2	910K	**77.8**	**75.4**	**93.8**
		4	4M	77.5	**75.5**	**93.9**

Classification results

- Results are just over the state of the art alternatives
  - Far shallower architecture than previous state of the art
    - 2-conv layers vs 43
- Slightly deeper networks => overfitting (Too small dataset)

Results: Sydney Urban Objects - LiDAR/Pointcloud

- Average F1 score
  - Dataset unbalanced but taken into account
  - 4 folds/subset used for cross validation (3 folds training, 1 testing)
    - Small dataset => ran three times with different seeds

Results: ModelNet - Synthetic/CAD

Method	Conv. Layers	Batch Norm.	Accuracy (%) No Alignment	Accuracy (%) Rough, Automatic Alignment	Accuracy (%) Perfect, Manual Alignment
VoxNet [21] (baseline)	2	×	83	-	-
ORION (Ours)	2	×	-	88.1	87.5
	2	✓	-	88.6	88.2
	4	✓	-	89.4	**89.7**

Classification examples

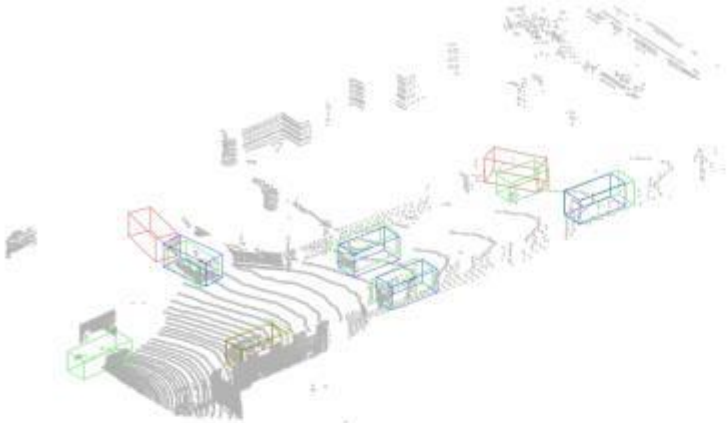
Ground Truth	4wd	building	bus	desk	monitor
Baseline	car	bus	car	sofa	chair
Ours	4wd	building	bus	desk	monitor
Ground Truth	table	chair	ute	toilet	bathtub
Baseline	nite-stnd	table	ute	toilet	table
Ours	table	chair	truck	chair	bed

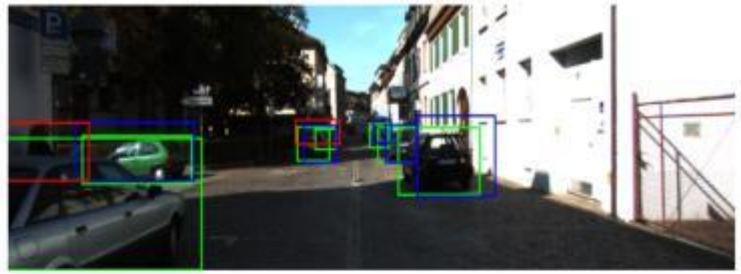
Figure 4: Some exemplar classification results. We show examples on which the outputs of the two networks differ.

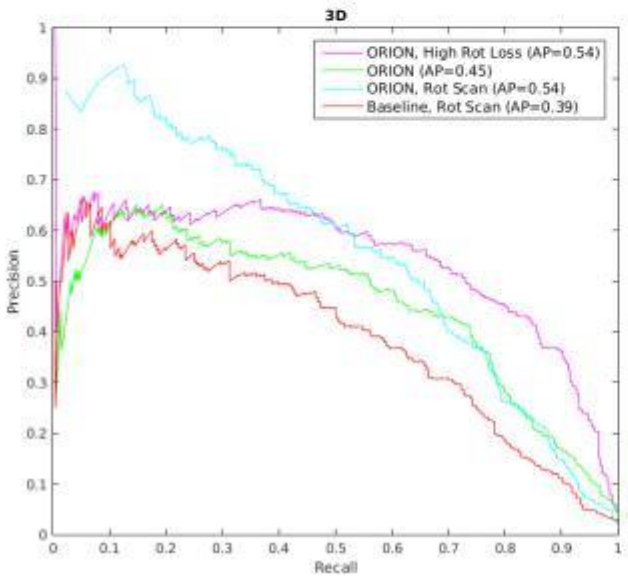
Results: Detection (KITTI)

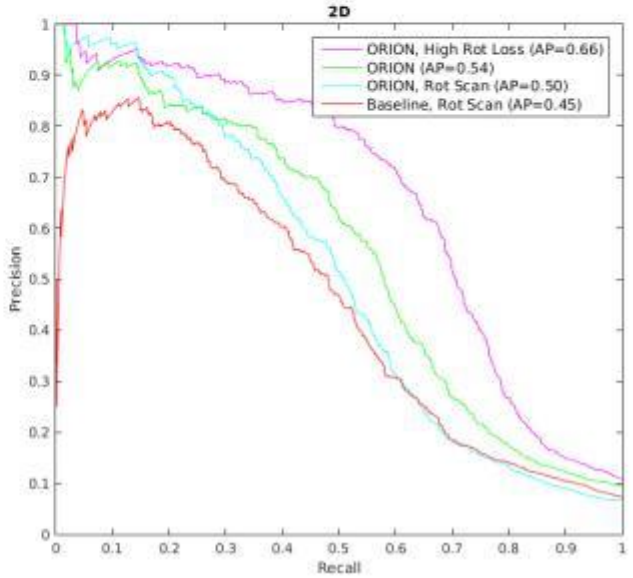
- 3D detector to detect cars
- Orientation sensitive network => binary object classifier
- 3D point cloud (No RGB)

Results: Detection (KITTI)









Conclusion

Object orientation during training => better classification results at test time.

Source

- https://www.youtube.com/watch?v=a246GAffWZk modelnet40 trained example