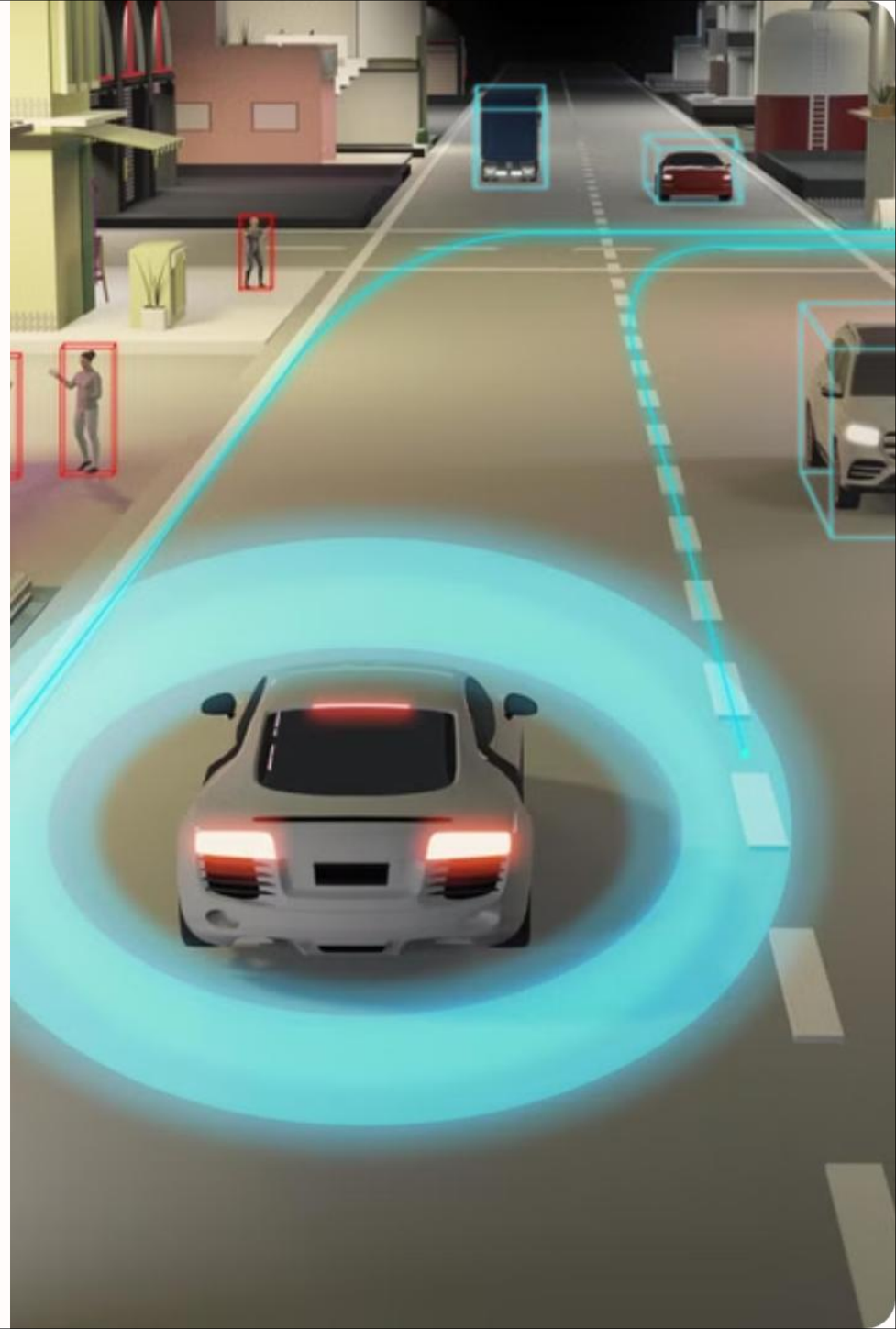
PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation

PointFusion is a generic 3D object detection method that leverages both image and 3D point cloud information. It processes raw sensor data and predicts 3D bounding boxes without dataset-specific assumptions.



Heterogeneous Network Architecture

Image Feature Extractor

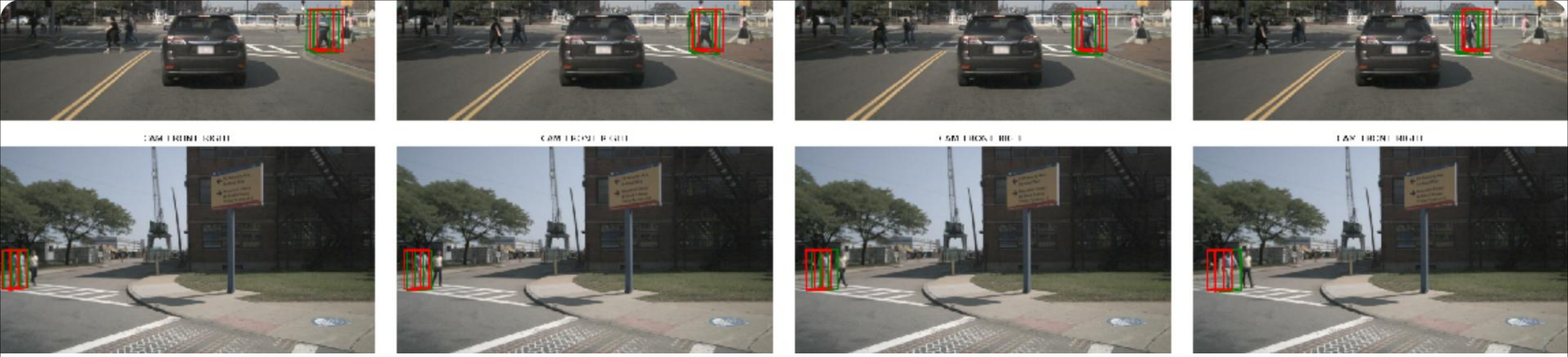
A CNN extracts appearance and geometry features from the input image.

Point Cloud Processor

A PointNet variant processes the raw 3D point cloud directly.

Fusion Network

A novel dense fusion network combines the image and point cloud features to predict 3D bounding boxes.



Spatial Anchor Prediction

1

2

3

Relative Offsets

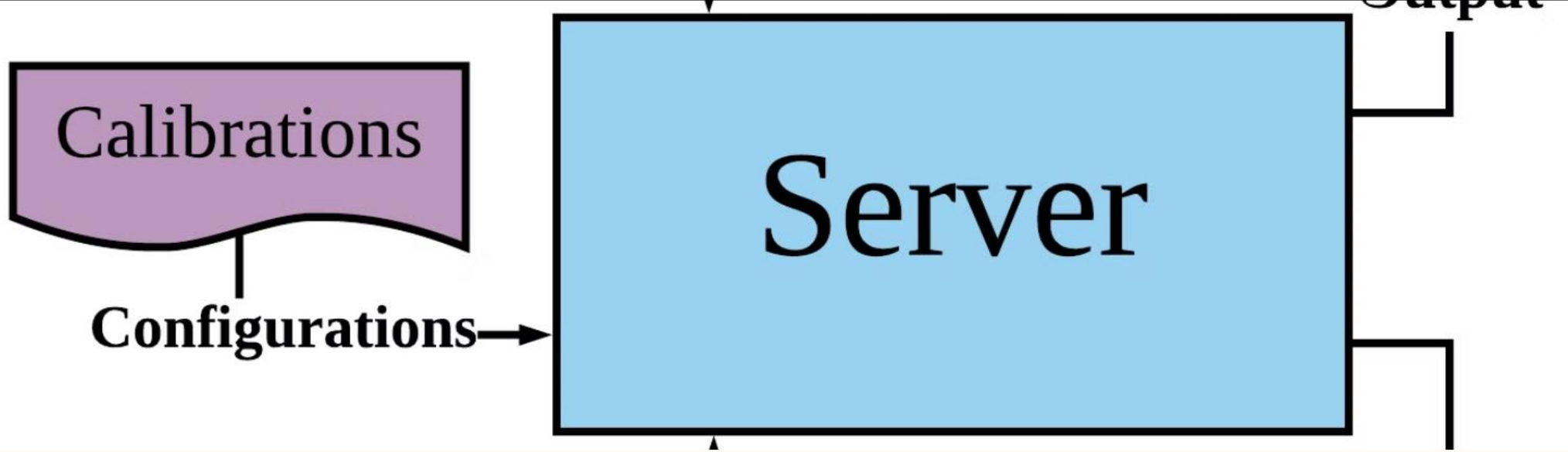
For each input 3D point, the network predicts the spatial

Scoring

The network learns to select the prediction with the highest

Robust Prediction

This spatial anchor approach reduces the variance in the



Evaluation on KITTI

1

Outperforms State-of-the-Art

PointFusion achieves better or on-par performance compared to specialized methods on the KITTI

2

Robust to Occlusion

The fusion model performs well even on heavily occluded objects, thanks to the complementary image and depth cues.

3

Generalizes Across Categories

A single PointFusion model works well for cars, pedestrians, and cyclists without category-specific

Evaluation on SUN-RGBD

Competitive Performance

PointFusion achieves results on-par with state-of-the-art methods on the diverse SUN-RGBD indoor dataset.

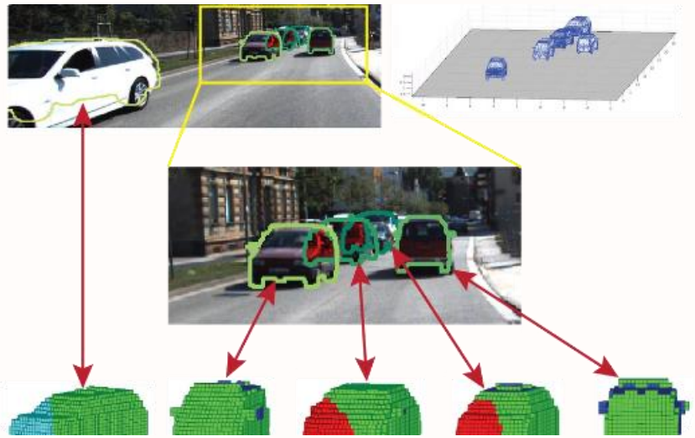
Faster Inference

PointFusion is much faster than specialized methods, enabling real-time 3D object detection.

Handles Varied Sensors

The generic architecture works well with both lidar and RGB-D camera inputs.

Qualitative Results

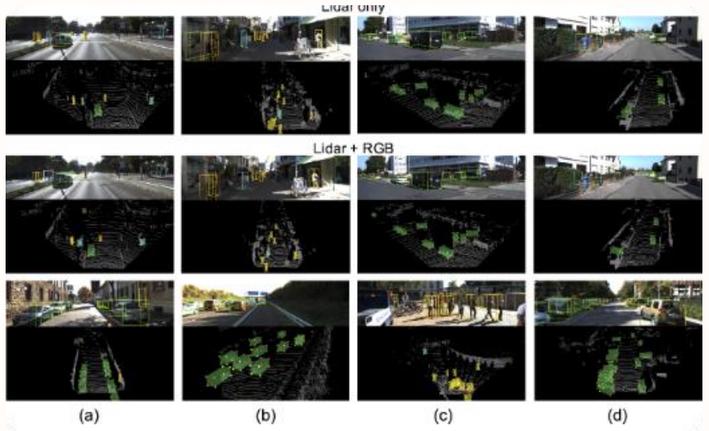


Accurate Car Detection

PointFusion accurately estimates the 3D bounding boxes of cars, including their dimensions and orientation.

Robust to Occlusion

The model performs well even on partially occluded objects like pedestrians and cyclists.



Handles Complex Scenes

PointFusion can detect objects in cluttered environments with multiple objects of different categories.

Ablation Studies

Point Cloud Input

Using raw point clouds is crucial for accurate 3D detection.

Image Fusion

Combining image and point cloud features significantly boosts performance.

Dense Prediction

The novel dense prediction architecture outperforms direct 3D box regression.

Unsupervised Scoring

The self-learned scoring function performs better than a supervised proxy objective.

Conclusion

1

Generic Architecture

PointFusion is a simple and generic sensor fusion method that works across diverse datasets and sensor configurations.

2

State-of-the-Art Performance

PointFusion achieves better or on-par results compared to specialized methods on both KITTI and SUN-RGBD.

3

Future Work

Potential future directions include end-to-end training and incorporating temporal information for joint detection and