Integrity Monitoring of 3D Object Detection in Automated Driving Systems using Raw Activation Patterns and Spatial Filtering

Hakan Yatbaz, Mehrdad Dianati, Konstantinos Koufos, Roger Woodman

Presented by Alireza Ahrabian from HITACHI Europe

25.09.24

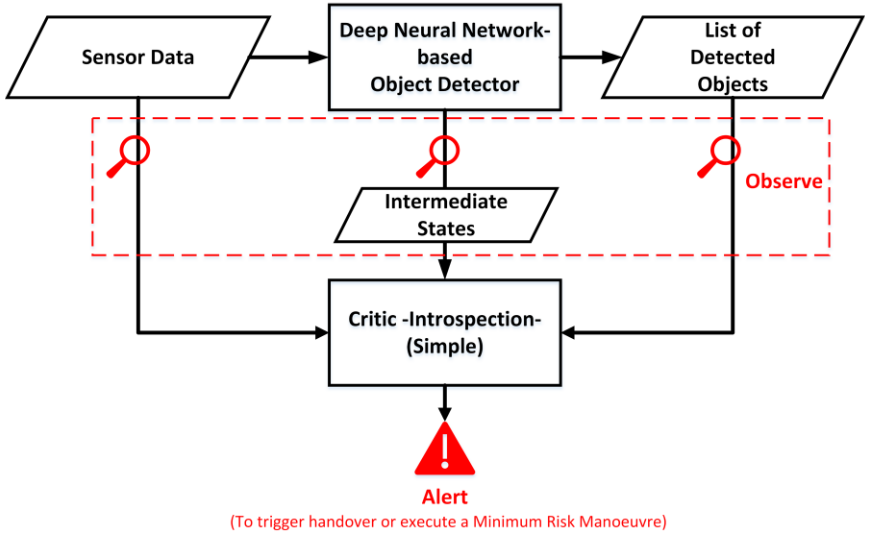
WMG
THE UNIVERSITY OF WARWICK
EVENTS

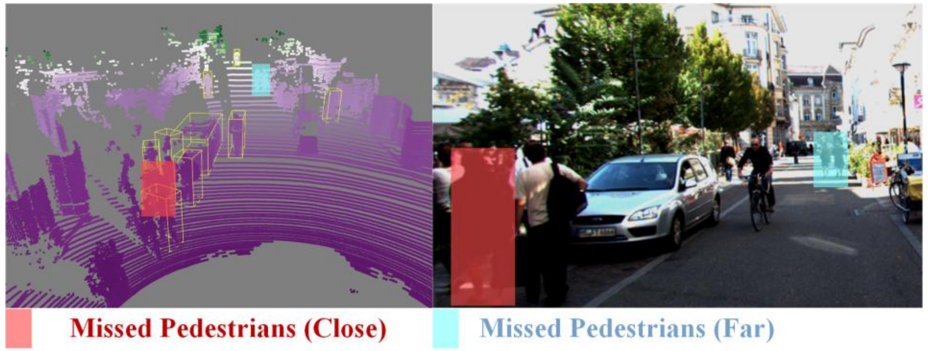
Outline

- Motivation for Using Perception Integrity Monitoring
- Proposed Integrity monitoring Mechanism for 3D LiDAR-based Object Detection
- Performance Evaluations
- Conclusions

Motivation

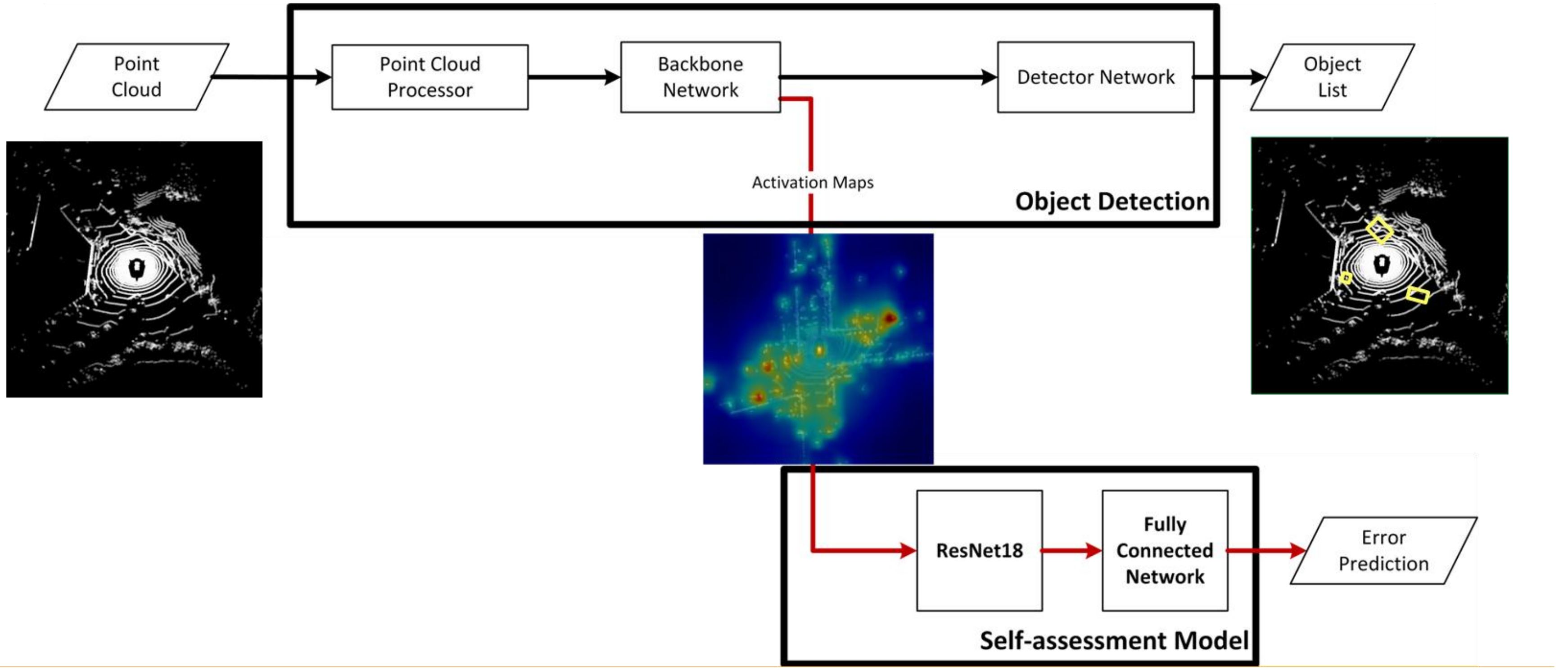
- DNN-based object detectors for perception are vulnerable to faults and failure for a variety of reasons.
- In automated driving systems (ADS), run-time perception monitoring mechanisms can mitigate the impact of these failures.
- We adopt an actor-critic architecture where the critic is a secondary system that continuously assesses the performance of the actor (perception system).
- Within a given scene, not all objects are equal for safety criticality– hence some filtering mechanism is recommended.





Activation Maps

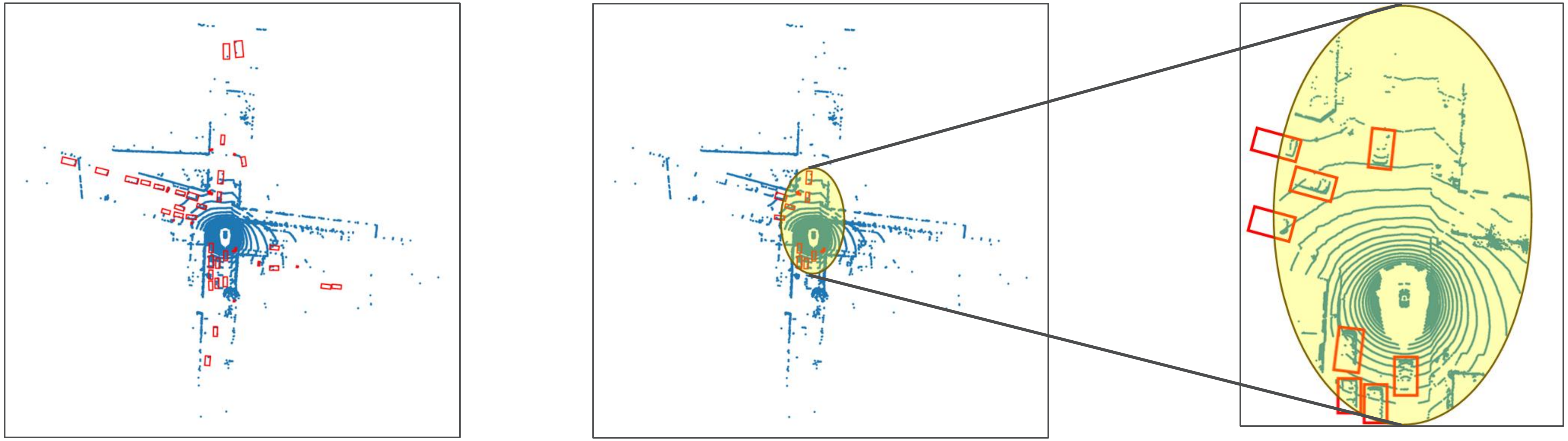
- For perception integrity monitoring (or self-assessment), we propose to use activations (features) extracted for 3D object detection.
- The aim is to associate the activations with the objects that might be missed.



Spatial Filtering

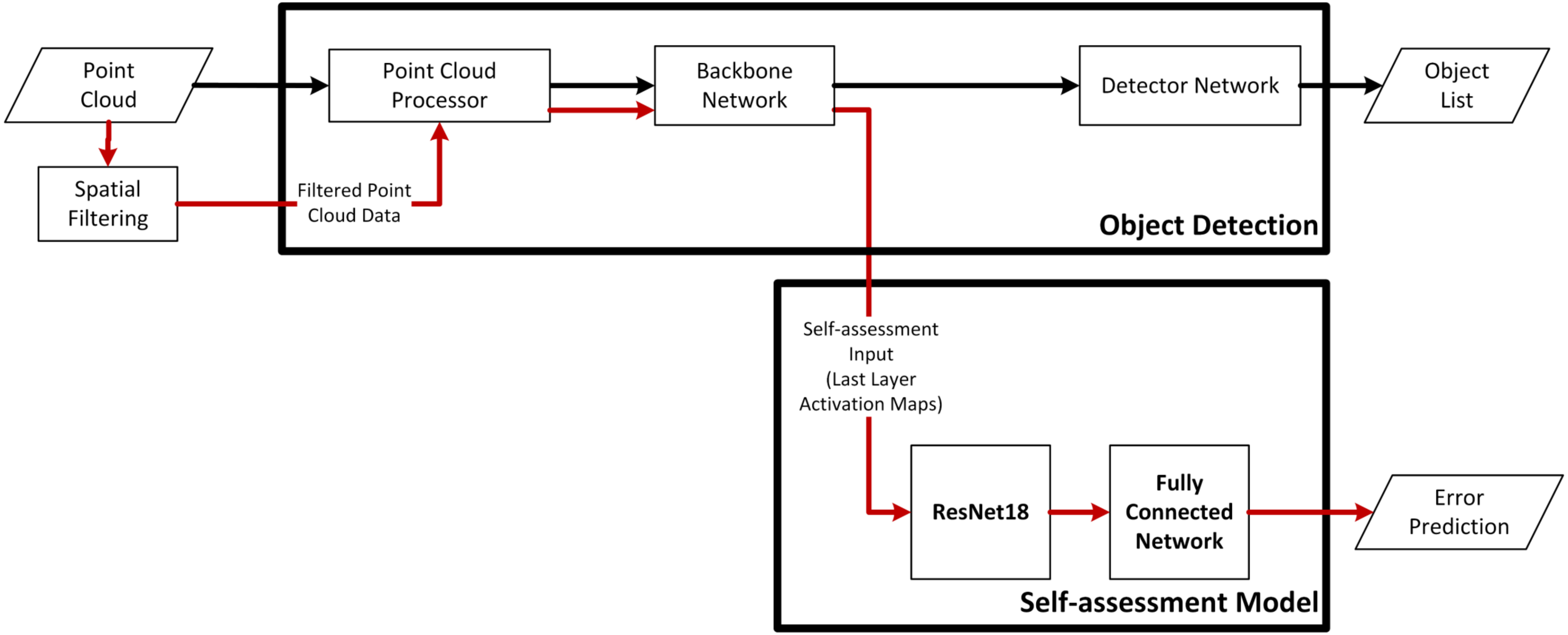
- To focus on the immediate danger zone within the local surroundings of the Ego vehicle, we propose spatial filtering of the 3D point cloud, removing any content outside an ellipse-shaped filter.
- **Filter Range:** [-20m,30m] longitudinal, [-15m,+15m] lateral.

Full Scene with All Objects | Full Scene with Only Objects Filtered | Spatially Filtered Scene



Activation maps and Spatial Filtering

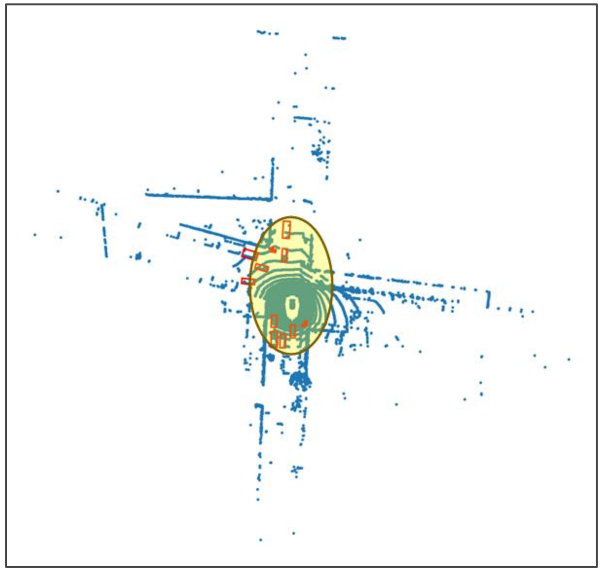
- To identify erroneous scenes while operating with spatial filtering the following architecture is used.



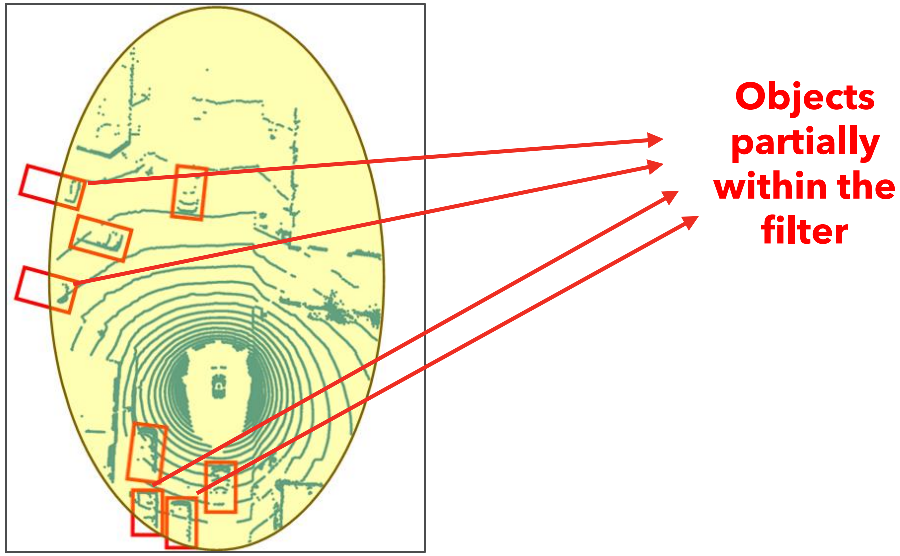
Proposed Mechanism

- We may be losing information if an object lies partially within the spatial filter.
- We also investigate error detection with full scene point cloud data where only the labels filtered.

Full Scene with Only Labels Filtered



Spatially Filtered Scene



Performance Evaluations

EVENTS WMG THE UNIVERSITY OF WARWICK

Datasets

KITTI Dataset:
- ~7000 Samples.
- Only the front of the ego vehicle is annotated.
- Car, pedestrian and cyclist classes are used.

NuScenes Dataset:
- ~34k samples (~850 scenes).
- 360° Annotation.
- Vehicles and pedestrian classes are used.

Object Detectors

Point Pillars:
- Common baseline model.
- Processes point cloud into 2D xy plane as pillars.
- Provided in common frameworks, and Autoware Auto's software stack.

Centerpoint:
- Better performing baseline model for ADS.
- Provided in the latest Autoware Auto software stack and currently used in WMG's prototype vehicle.

Metrics

AUROC:
- An indicator of how well a classifier distinguishes between the positive ('error') and negative ('no-error') classes.

Recall:
- Measures the classifier's ability to correctly identify the class.
- Both positive and negative class recall is calculated.

Implementation Details

Libraries and Frameworks

- PyTorch
- Torchvision
- Torchmetrics
- OpenMMDetection3D
- Grad-CAM

Training Parameters & Hyperparameter Tuning

Loss Function: Focal Loss

$$L(q) = -\sum \alpha_i \ (1 - q_i)^{\gamma} \log(q_i) \qquad i \in \{0, 1\}$$

where:

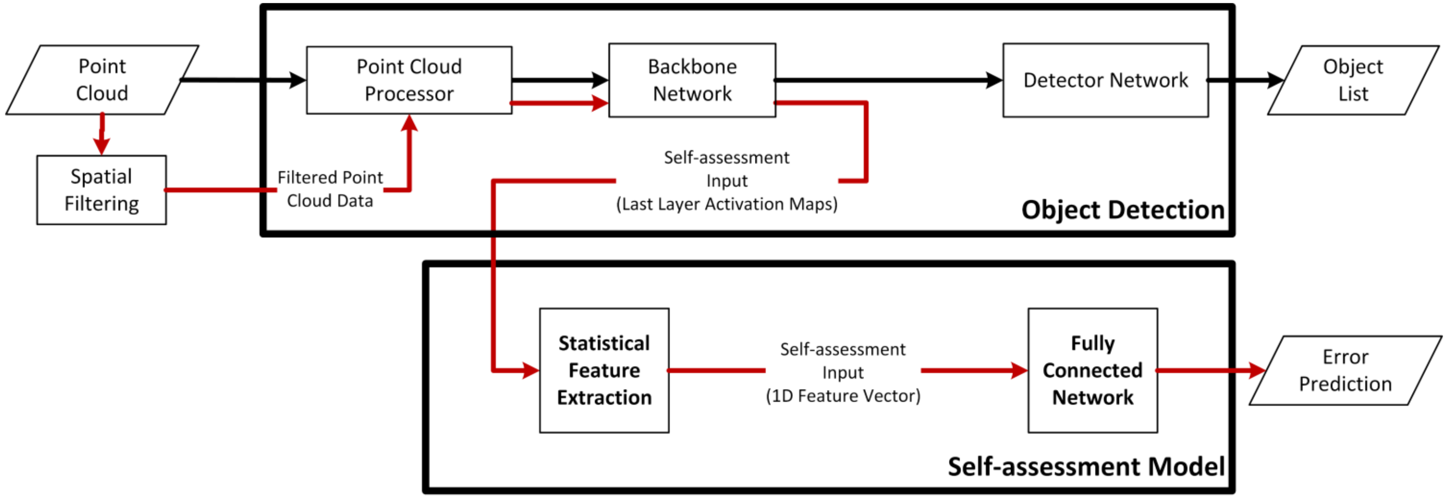
- log is the natural logarithm
- q is the predicted probability vector for the,
- α_i is a scaling factor (class weights)
- $\gamma \in \{0,2,4,5\}$ is a parameter that down-weights easy examples and emphasises hard examples.

Optimizers: **Stochastic Gradient Descent (SGD)**, Adam
Batch Sizes: 32,**64**,128
Learning Rates: 0.1, **0.01**, 0.001, 0.0005

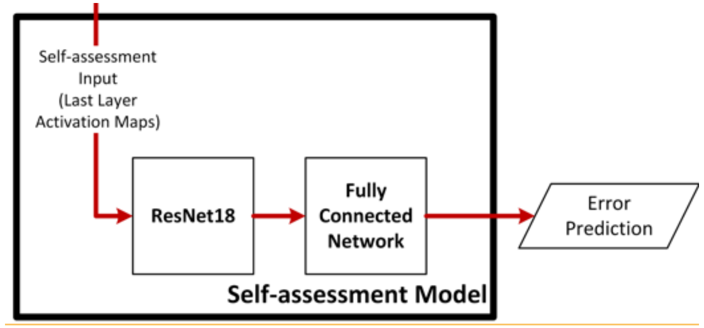
Best performing model is obtained with the hyperparameters shown in **red**.

SOTA Mechanism

- The state-of-the-art (SOTA) mechanism adapted from literature for comparison is called Statistical Features [1].
- It performs global pooling of the last layer activation maps with max, mean and standard deviation functions to create a 1D feature vector.



Statistical Features



Proposed

[1] Q. M. Rahman, N. Sunderhauf, and F. Dayoub, "Per-frame map prediction for continuous performance monitoring of object detection during deployment," in 2021 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW), 2021, pp. 152–160

Performance Evaluations

EVENTS WMG THE UNIVERSITY OF WARWICK

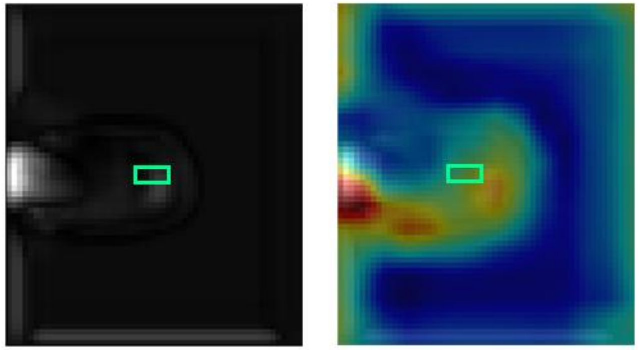
Dataset / Method	Input	Rec(-)	Rec(+)	AUROC
KITTI / PointPillars	Statistical Features	0.9649	0.0625	0.6193
	Filtered Activations	**0.6368**	**0.7091**	**0.7384**
NuScenes / Centerpoint	Statistical Features	0.4819	0.8067	0.7317
	Filtered Activations	**0.5330**	**0.8646**	**0.8092**

- For both datasets, the proposed mechanism performs better than statistical features regarding AUROC and error class recall.
- Spatial Filtering has reduced the number of error frames on both datasets when compared with previous work [2], highlighting that significantly less critical frames are labelled as errors without filtering.

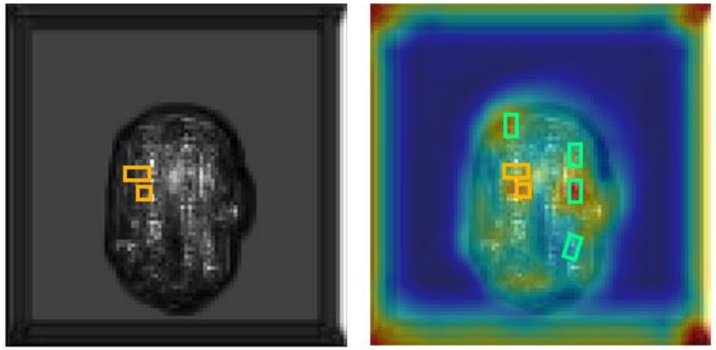
[2] H. Y. Yatbaz, M. Dianati, K. Koufos, R. Woodman, "Run-time Monitoring of 3D Object Detection in Automated Driving Systems Using Early Layer Neural Activation Patterns," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2024, pp. 3522-3531

Performance Evaluations

- Channel-wise maximum of activation maps (left images).
- Activation visualization with EigenCAM Heatmap of model attention (right images). (**red**: high, **blue**: low)
- Diving direction is from left to right for KITTI, and from bottom to top for NuScenes.



KITTI



NuScenes

Correctly Detected Object

Missed Object

- The self-assessment mechanism directly focuses on the area of interest with spatial filtering, and the missed objects have got higher attention.

Conclusion

- Our mechanism outperforms the state-of-the-art mechanism (statistical features) by around 7% in overall with AUROC and 6% in identifying error class.

- The reduction in the number of error frames with filtering highlights that a significant number of errors in the datasets are caused by point cloud sparsity for far away objects, which are less safety-critical.

Funded by the European Union

EVENTS project has received funding under grant agreement No 101069614. It is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Thank you.

hakan.yatbaz@warwick.ac.uk

hakanyektayatbaz

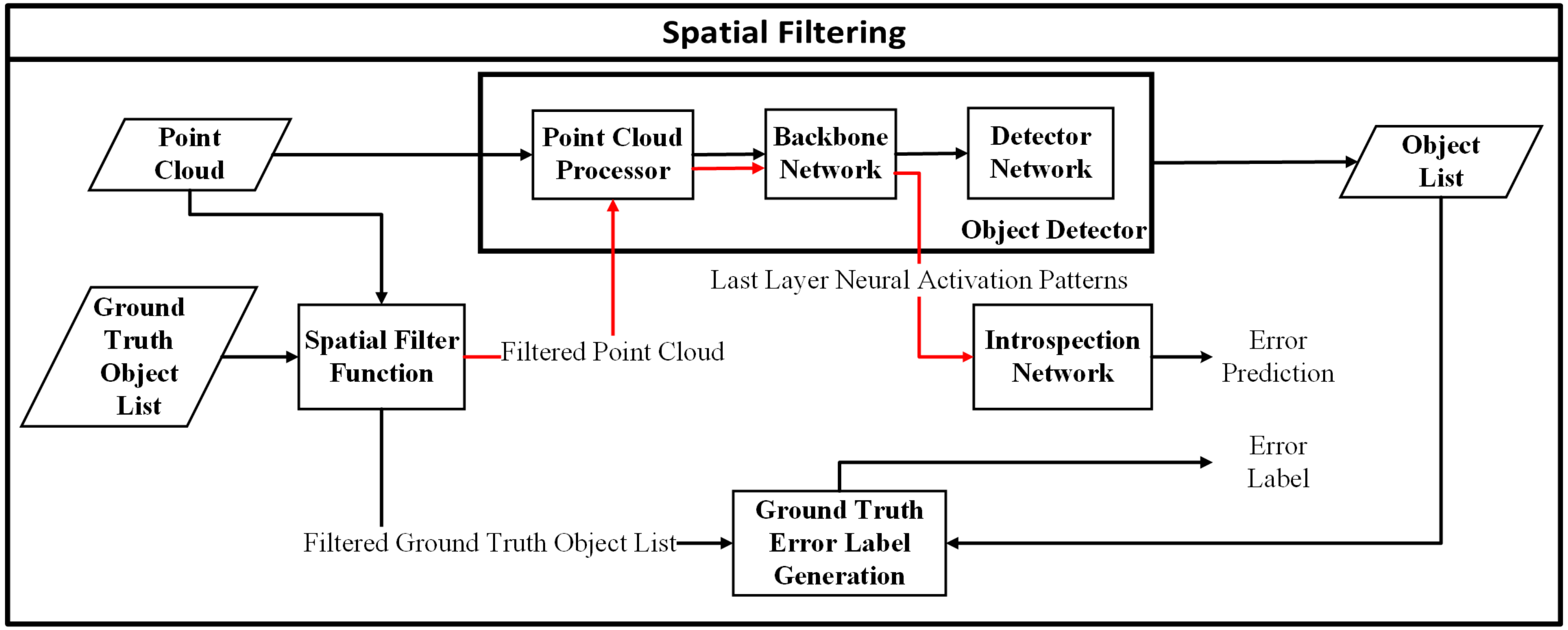
WMG
THE UNIVERSITY OF WARWICK

Q&A


AVL F-FEM-DDC SNAP-IN
AVL F-FEM-AIT SNAP-IN
25 TO 36
AVL PRESSURE MEASUREMENT SNAP-IN
AVL SNAP-IN
AVL

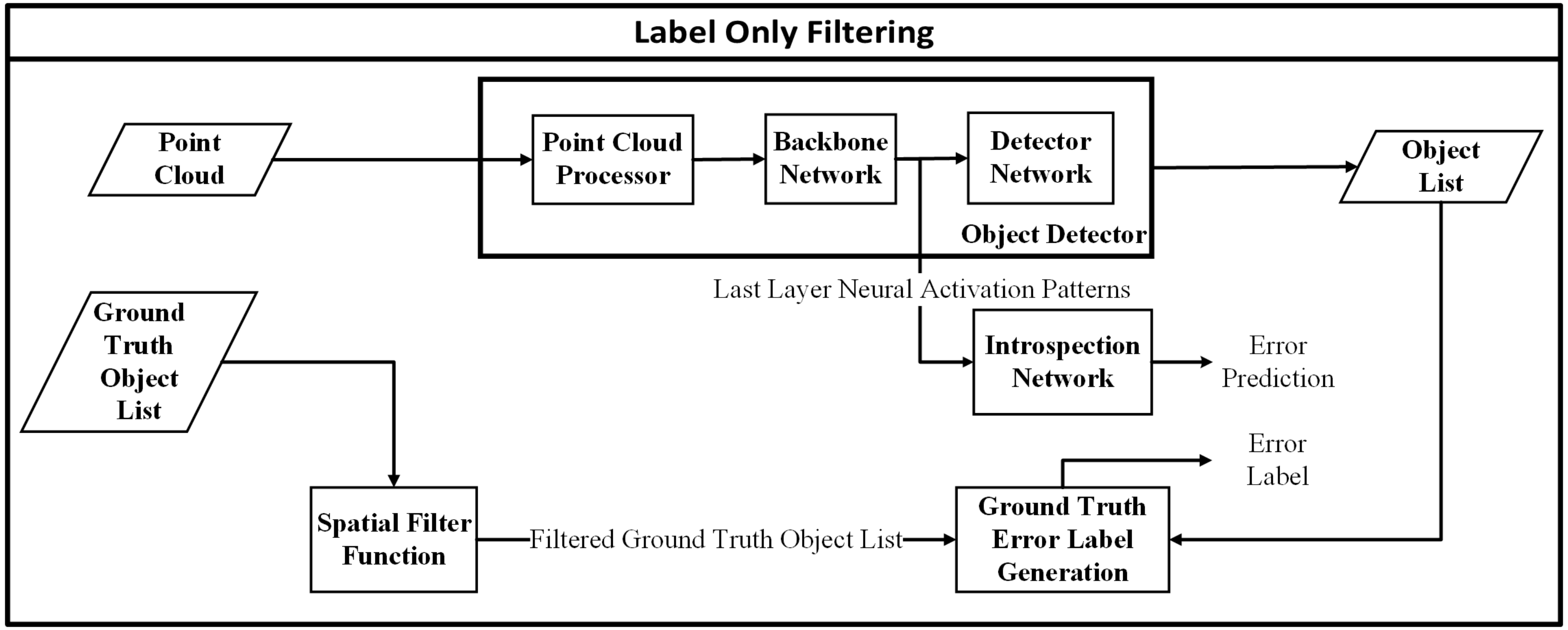
Proposed Mechanism

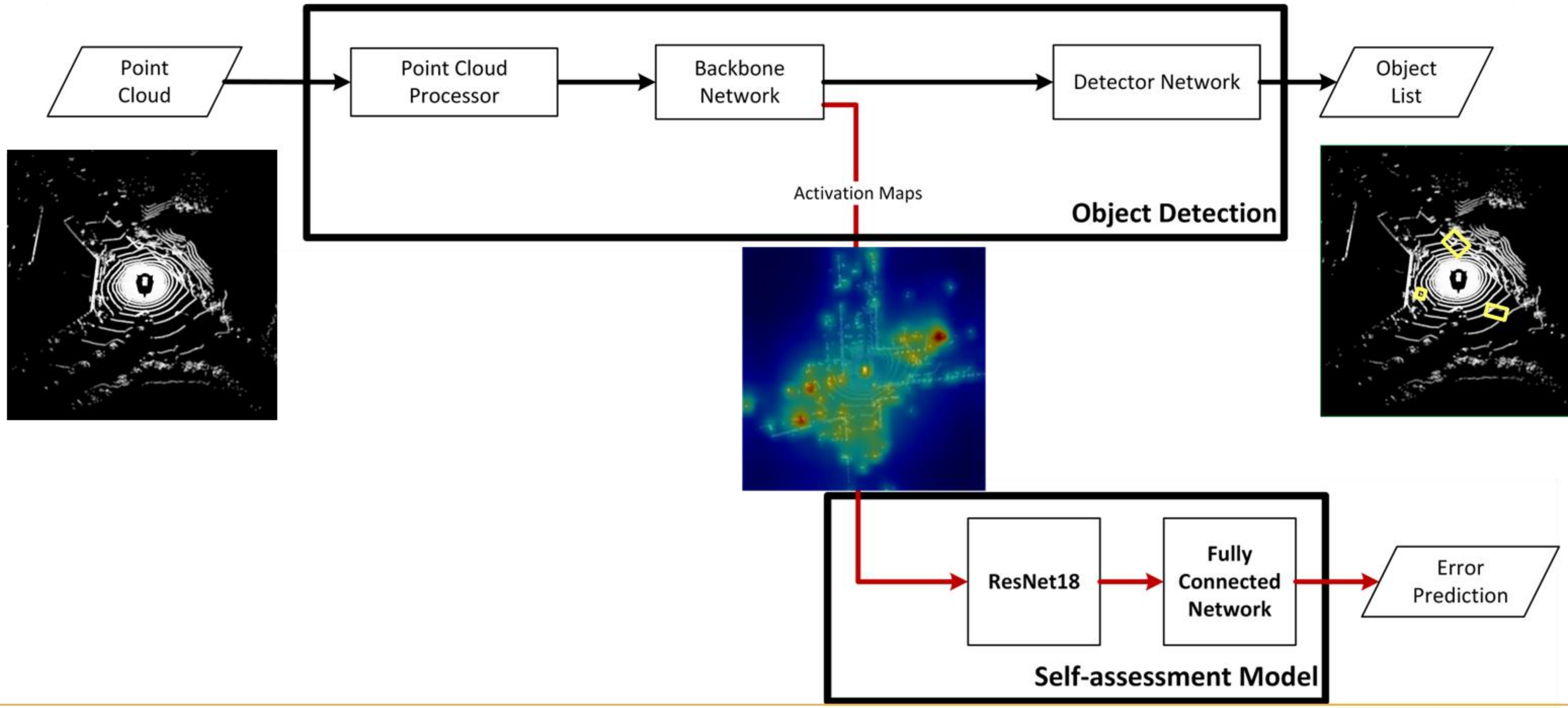
- To identify erroneous scenes while operating with spatial filtering the following architecture is used.



Proposed Mechanism

- To identify erroneous scenes, with label only filtering:





Point Cloud
Point Cloud Processor
Backbone Network
Detector Network
Object List
Activation Maps
Object Detection
ResNet18
Fully Connected Network
Error Prediction
Self-assessment Model