# Run-time Monitoring of 3D Object Detection in Automated Driving Systems Using Early Layer Neural Activation Patterns

Hakan Yekta Yatbaz[1], Mehrdad Dianati[1,2], Konstantinos Koufos[1], and Roger Woodman[1]
University of Warwick[1], Queen's University Belfast[2]

## Overview

**Problem:**

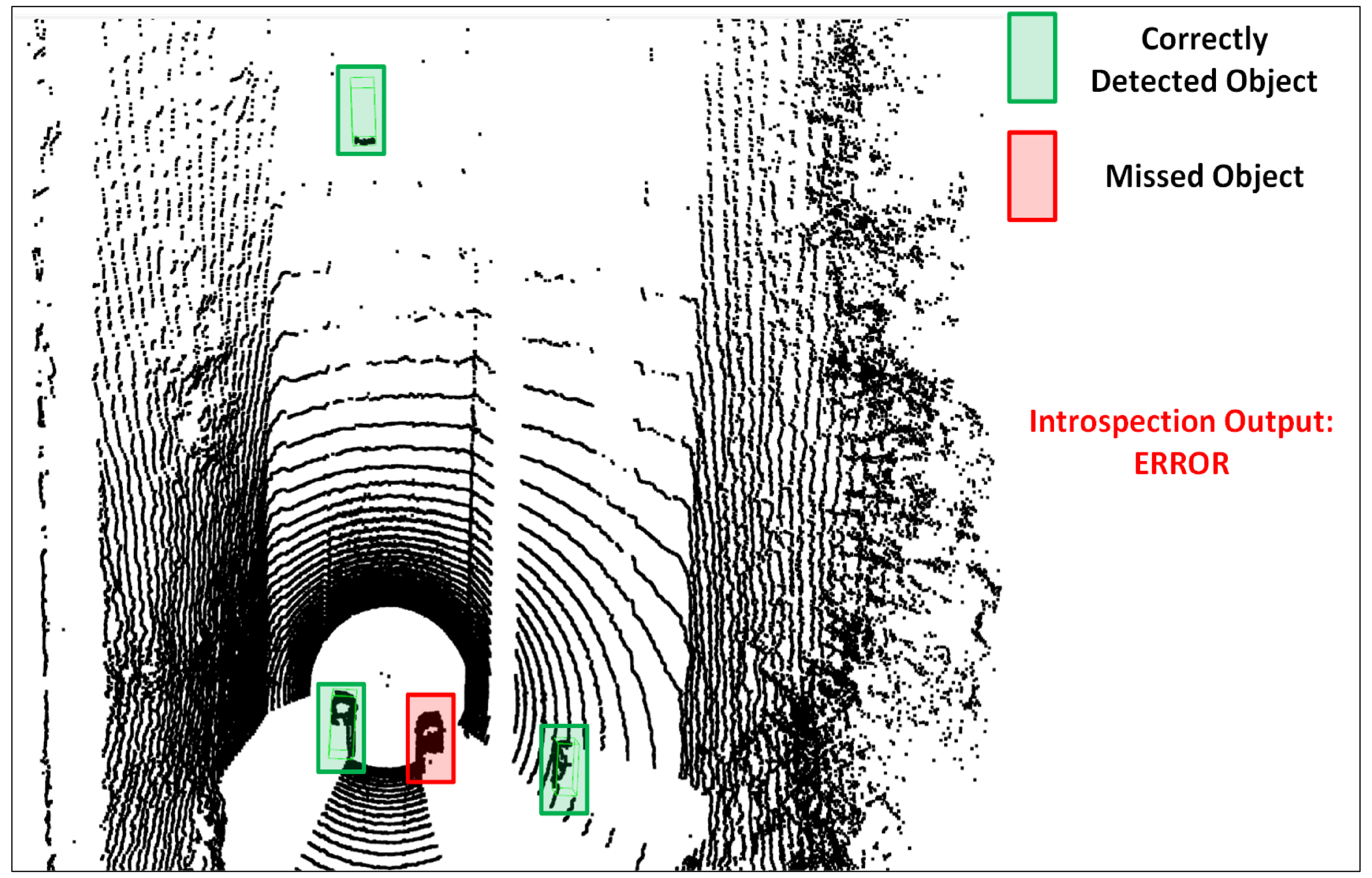
- Detecting objects in the scene is crucial for automated driving systems (ADS). However, widely used deep neural network-based object detectors are susceptible to errors.
- Developing and deploying a run-time monitoring mechanism to identify erroneous cases is essential for safety.
- Existing studies primarily focus on 2D object detection and use final layer activations which may not be sufficient for 3D detection.

**Contributions:**

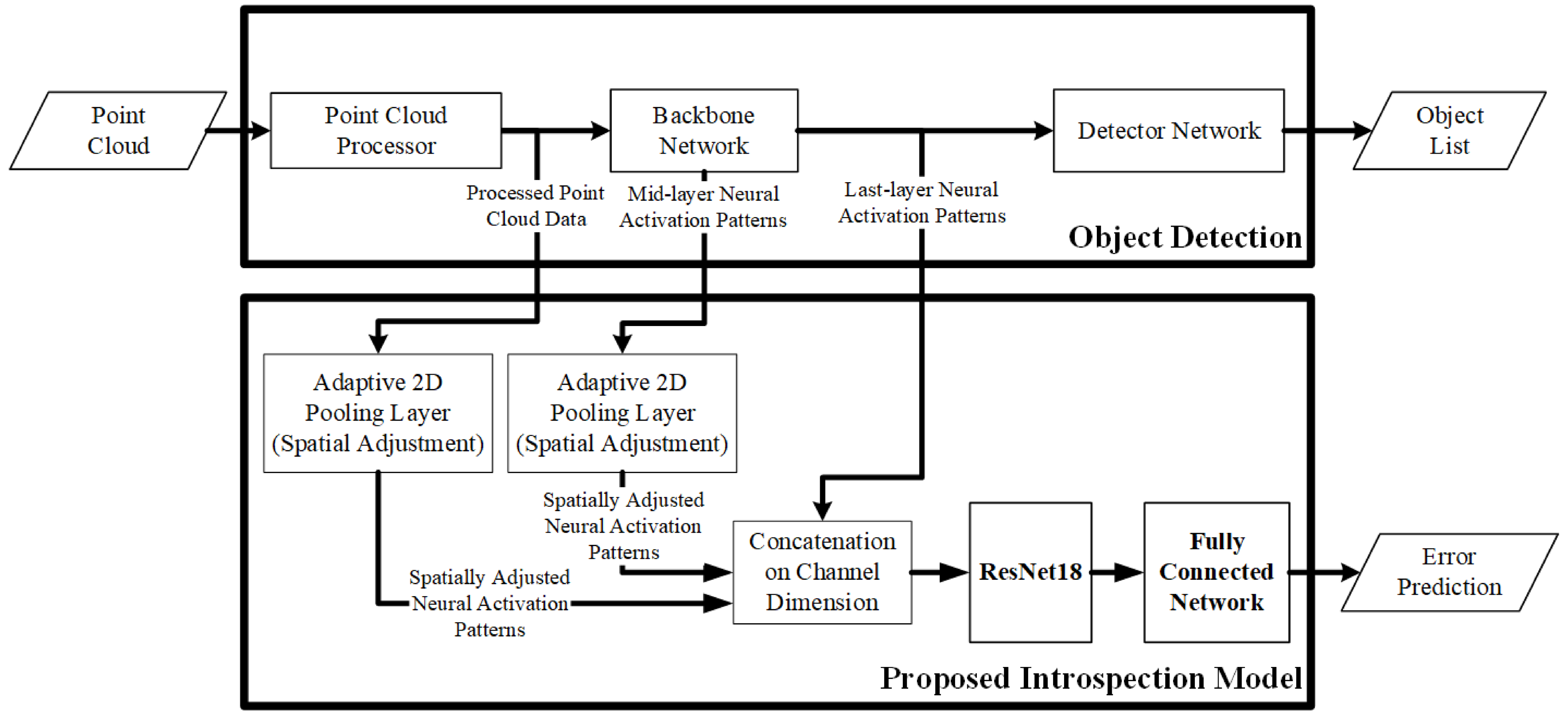
- Focusing on run-time monitoring of 3D object detection, which is not widely investigated.
- Investigating the use of earlier activation layers for error detection.
- Proposing a concatenation-based mechanism to combine activation patterns from multiple layers for better error detection.

**Findings:**

- Early layer activations provide better error detection capabilities at the cost of increased computational complexity.
- Proposed method offers a balanced performance in terms of accuracy and computational requirements.




## Proposed Mechanism




## Experimental Settings

**Datasets:** KITTI and NuScenes.

**Object Detectors:** PointPillars (for KITTI), CenterPoint (for NuScenes).

**Metrics:** AUROC, $\text{Recall}_{\text{Error (+)}}$, $\text{Recall}_{\text{No-Error (-)}}$.

## Performance Evaluations

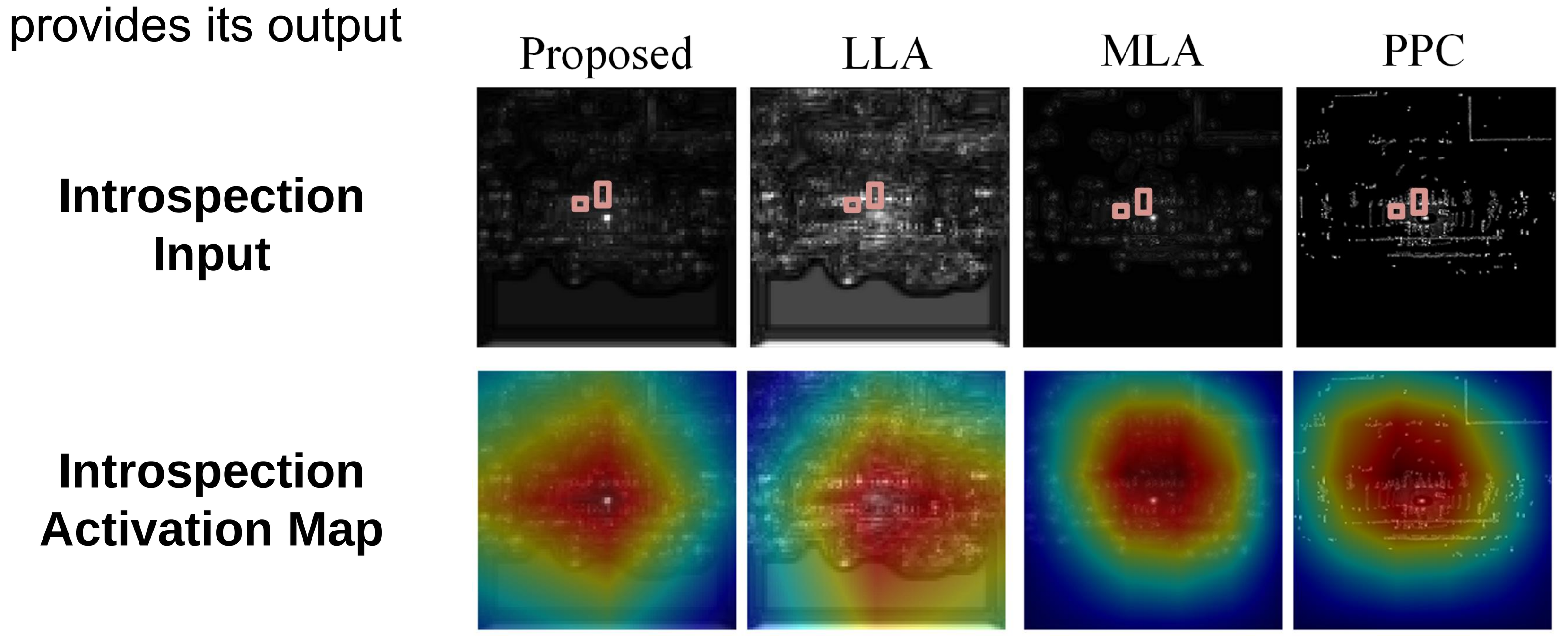
| Dataset / Model | Input | Rec.(-) | Rec.(+) | AUROC |
|---|---|---|---|---|
| Kitti / PointPillars | SF | 0.1479 | 0.9408 | 0.6000 |
| | PPC | **0.7764** | **0.7524** | **0.8420** |
| | MLA | <u>0.7500</u> | <u>0.7460</u> | <u>0.8368</u> |
| | LLA | 0.6268 | 0.8104 | 0.8036 |
| | Proposed | 0.7077 | 0.7858 | 0.8309 |
| NuScenes / CenterPoint | SF | 0.2607 | 0.9217 | 0.7322 |
| | PPC | 0.7945 | 0.8995 | 0.9198 |
| | MLA | **0.7945** | **0.9060** | **0.9330** |
| | LLA | 0.7123 | 0.8581 | 0.8919 |
| | Proposed | <u>0.8650</u> | <u>0.8630</u> | <u>0.9288</u> |

- Statistical Features (SF).
- Processed Point Cloud (PPC).
- Middle-Layer Activations (MLA).
- Last-Layer Activations (LLA).
- Proposed concatenation.

**Best result is in bold.**
<u>Second best is underlined.</u>

| Method | CPU Time (ms) | GPU Time (ms) | FLOPs (G) |
|---|---|---|---|
| PPC | 54.32 (9.54) | 11.47 (1.21) | 36.32 |
| MLA | 9.43 (3.26) | 2.01 (0.10) | 3.68 |
| LLA | **5.01 (0.47)** | **1.80 (0.06)** | **1.60** |
| Proposed | <u>4.94 (0.32)</u> | <u>1.95 (0.07)</u> | <u>2.60</u> |

- The statistics are calculated based on 1000 iterations excluding initial warm-up (700-800 ms), on an Intel(R) Core(TM) i9-10980XE CPU and NVIDIA RTX 3090 GPU.
- The time-lapse is measured from the point where the backbone network outputs all activation patterns till the point where the introspection model provides its output




Driving direction is from left to right in the example.

**Example Error Scene from NuScenes**

- Boxes show missed object locations.
- Attention of the introspection network shown with a heatmap (**red** :high, **dark blue**: low)
- All introspection models identified the error in the scene.
- High activation areas in PPC, MLA, and proposed methods correspond to the locations of missed objects.
- The proposed mechanism is attending the drivable area with a high focus on the missed objects.

EVENTS

Contact | Paper | Q&A

This research has been conducted as part of the EVENTS project, which is funded by the European Union, under grant agreement No 101069614. Views and opinions expressed are however thoseof 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.