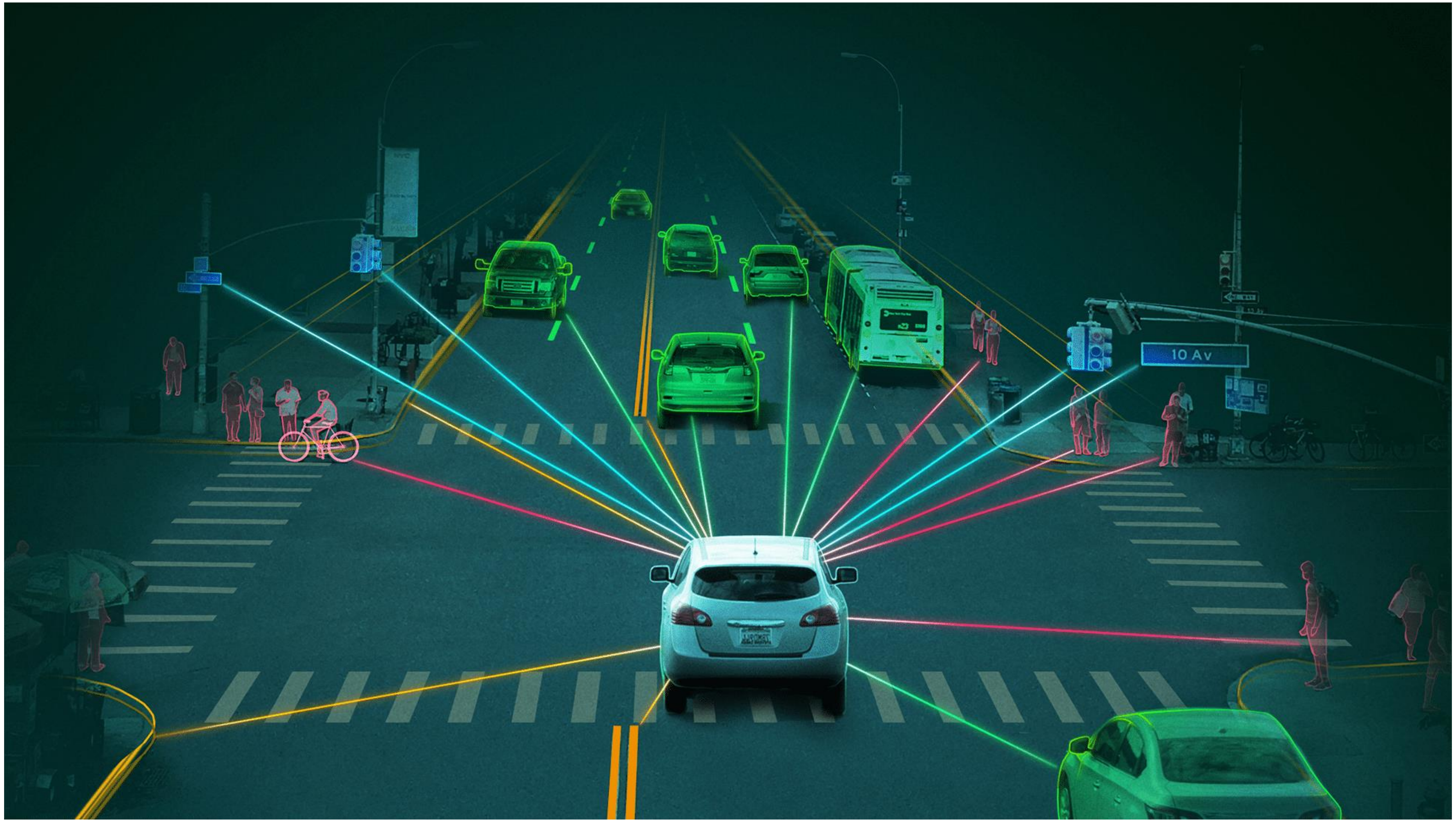
Institute of Automotive Technology
TUM School of Engineering and Design
Technical University of Munich

Methods for Sensor Failure Mitigation in 3D Object Detection

Alexander Fuchs, B.Sc.

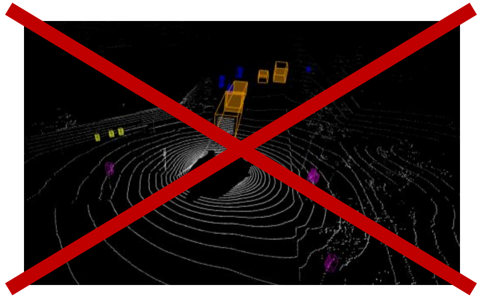
Garching, 27. March 2024

Sensing the Environment

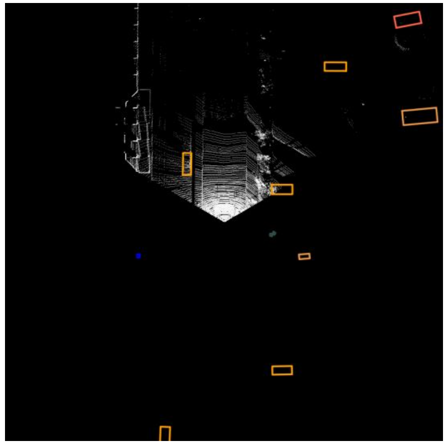


Real-World Conditions

Full Sensor Failure
(e.g. damaged sensor)



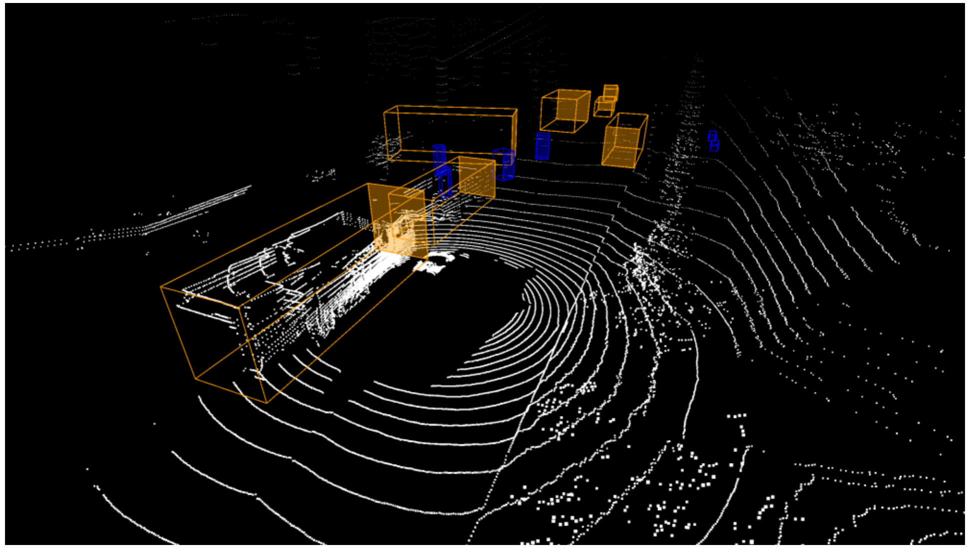
Partial Sensor Failure
(e.g. occlusion)



➔ Detection system should be **robust** against sensor failure

Input Modalities

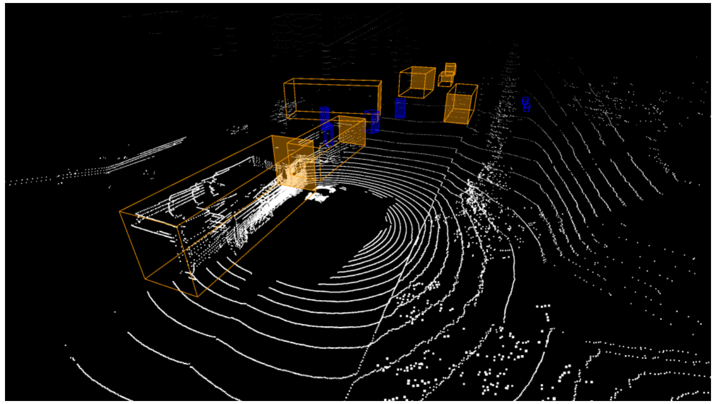
Camera

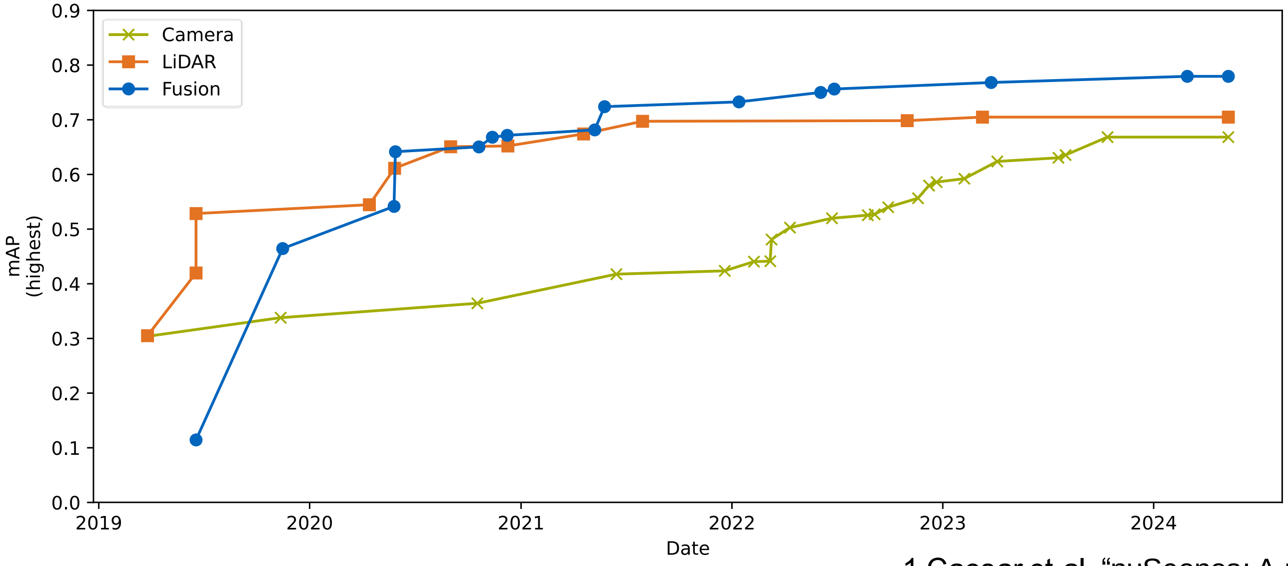


LiDAR

1 Caesar et al. “nuScenes: A multimodal dataset for autonomous driving“

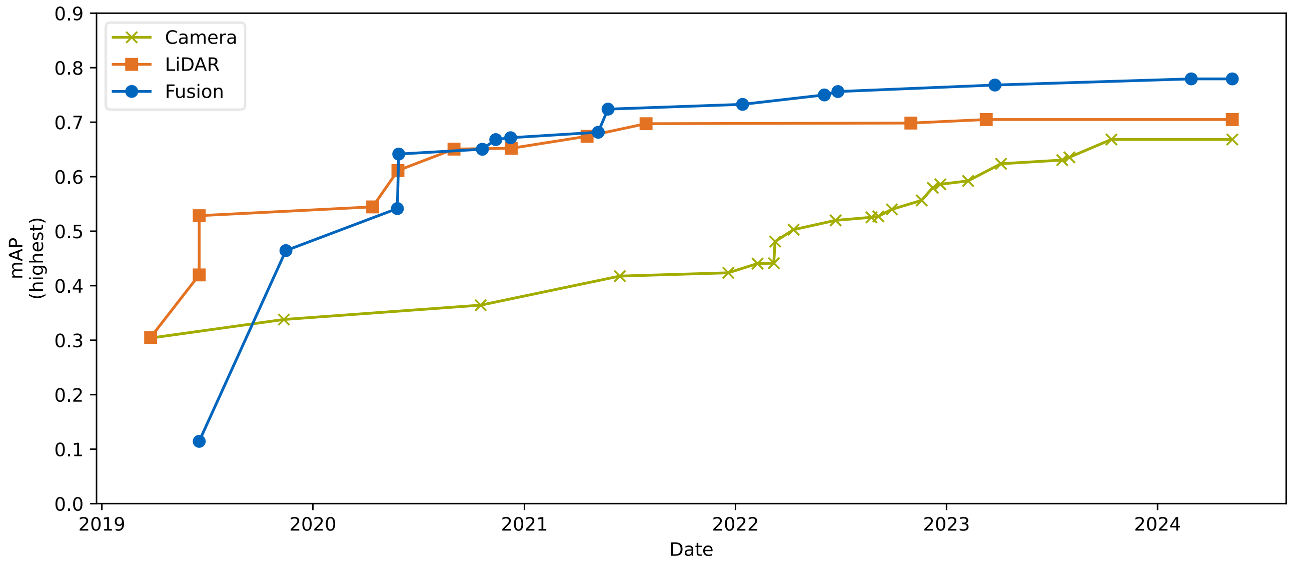
Input Modalities





1 Caesar et al. “nuScenes: A multimodal dataset for autonomous driving“

Input Modalities



LiDAR outperforms camera for 3D object detection

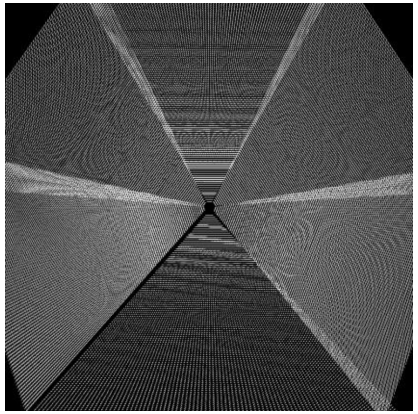
Fusion models have a higher reliance on LiDAR

1 Caesar et al. “nuScenes: A multimodal dataset for autonomous driving“

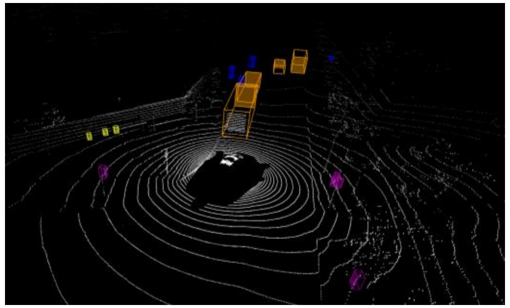
Sensor Fusion: Bird’s Eye View (BEV)

Camera

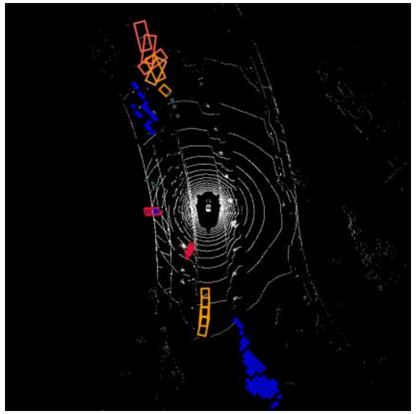
Explicit projection into 3D[2]



LiDAR



Flatten



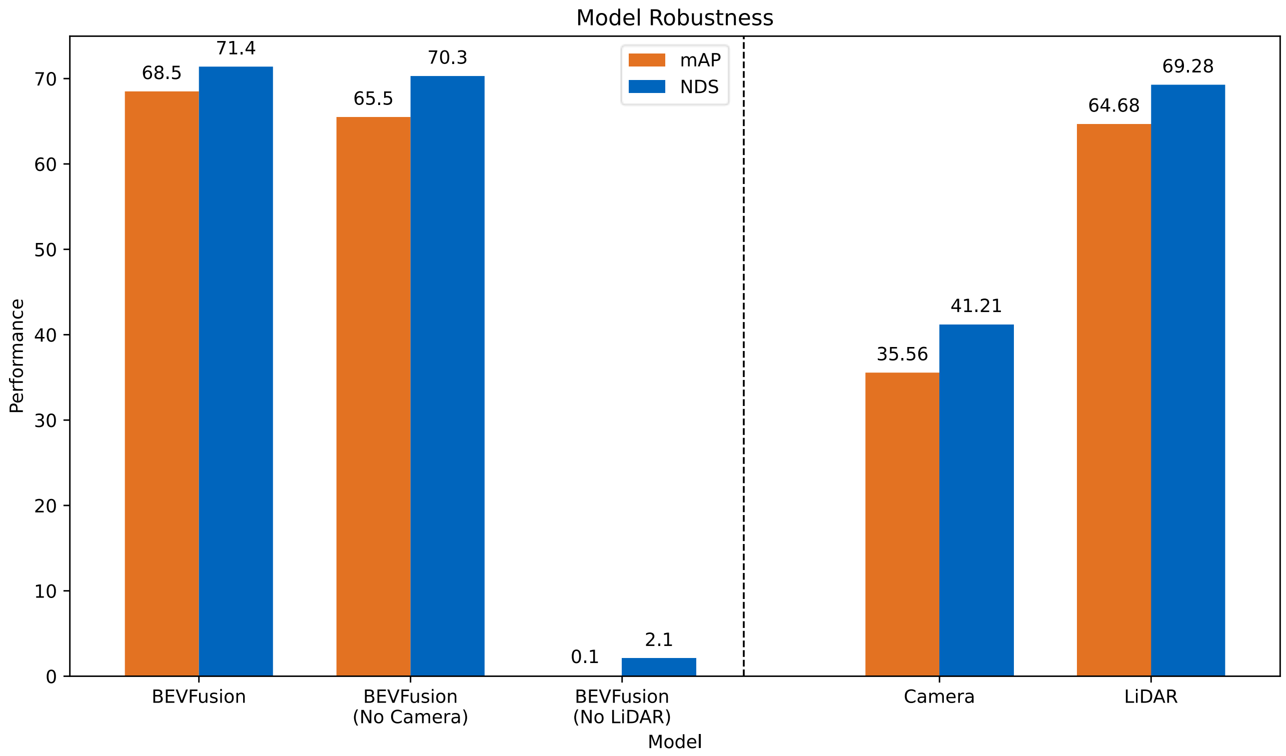
Fusion Model

2 J. Philion, S. Fidler “Lift, Splat, Shoot: Encoding Images from Arbitrary Camera Rigs by Implicitly Unprojecting to 3D”

Robustness of existing fusion methods

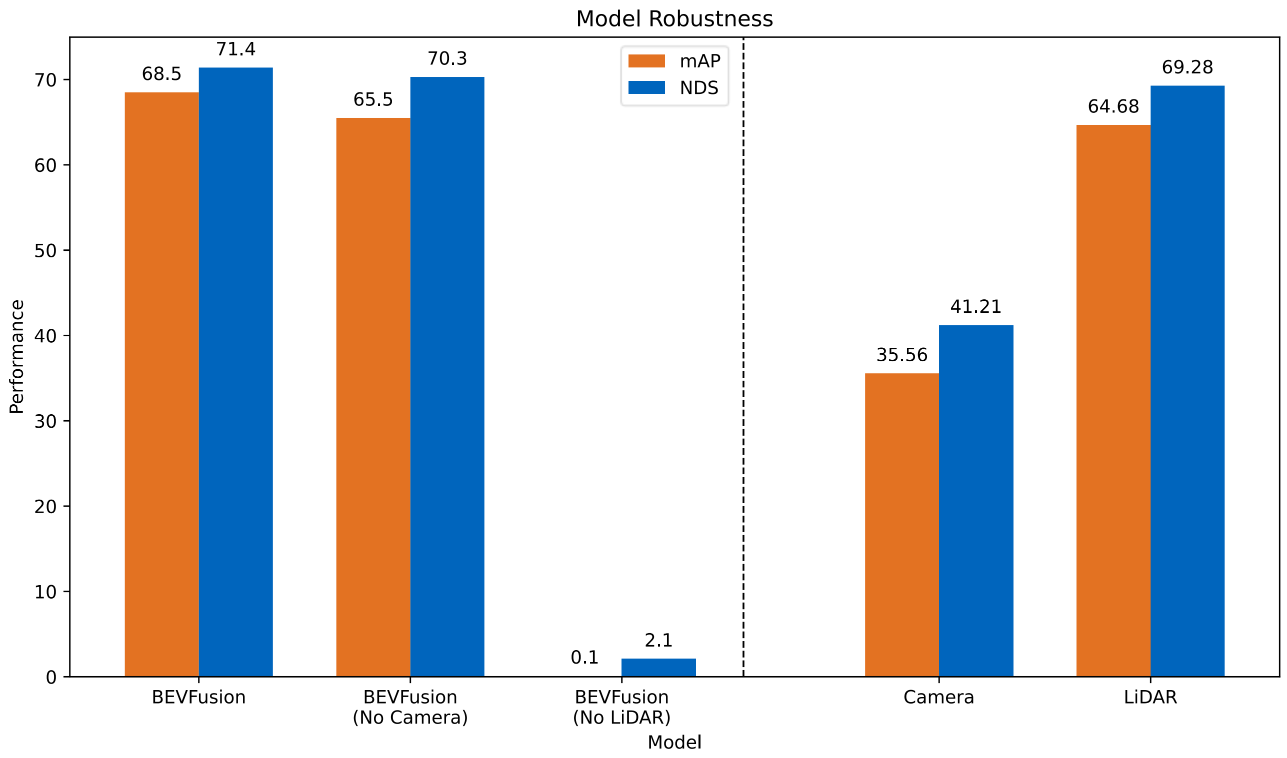
Performance in the event of **full sensor failure:**

Example: BEVFusion[3]



3 Liu et al. "BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation"

Robustness of existing fusion methods



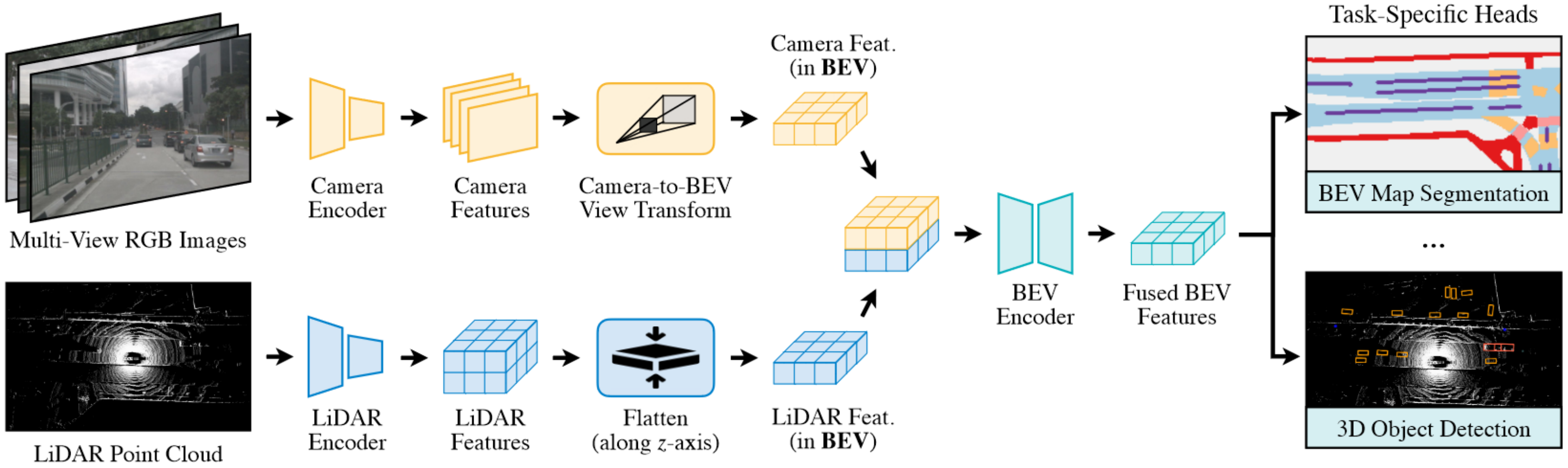
Existing sensor fusion methods **fail** when the LiDAR sensor fails

➔ **Goal:** Make sensor fusion robust against LiDAR failure

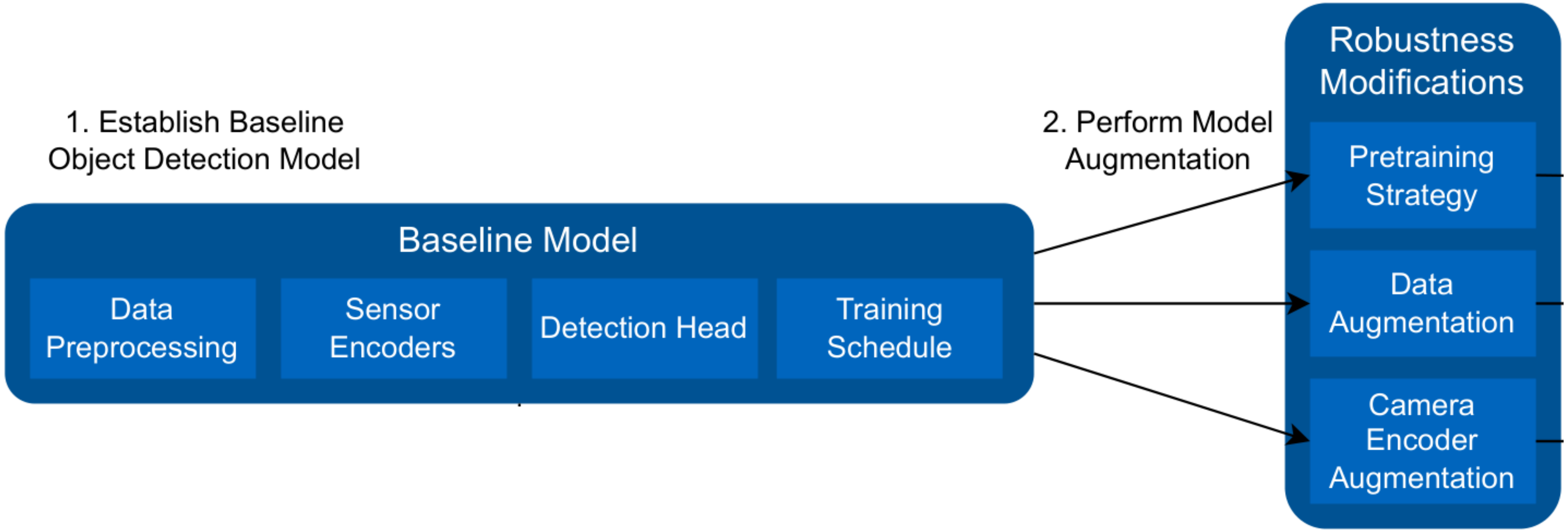
Approach: Baseline

Baseline Model

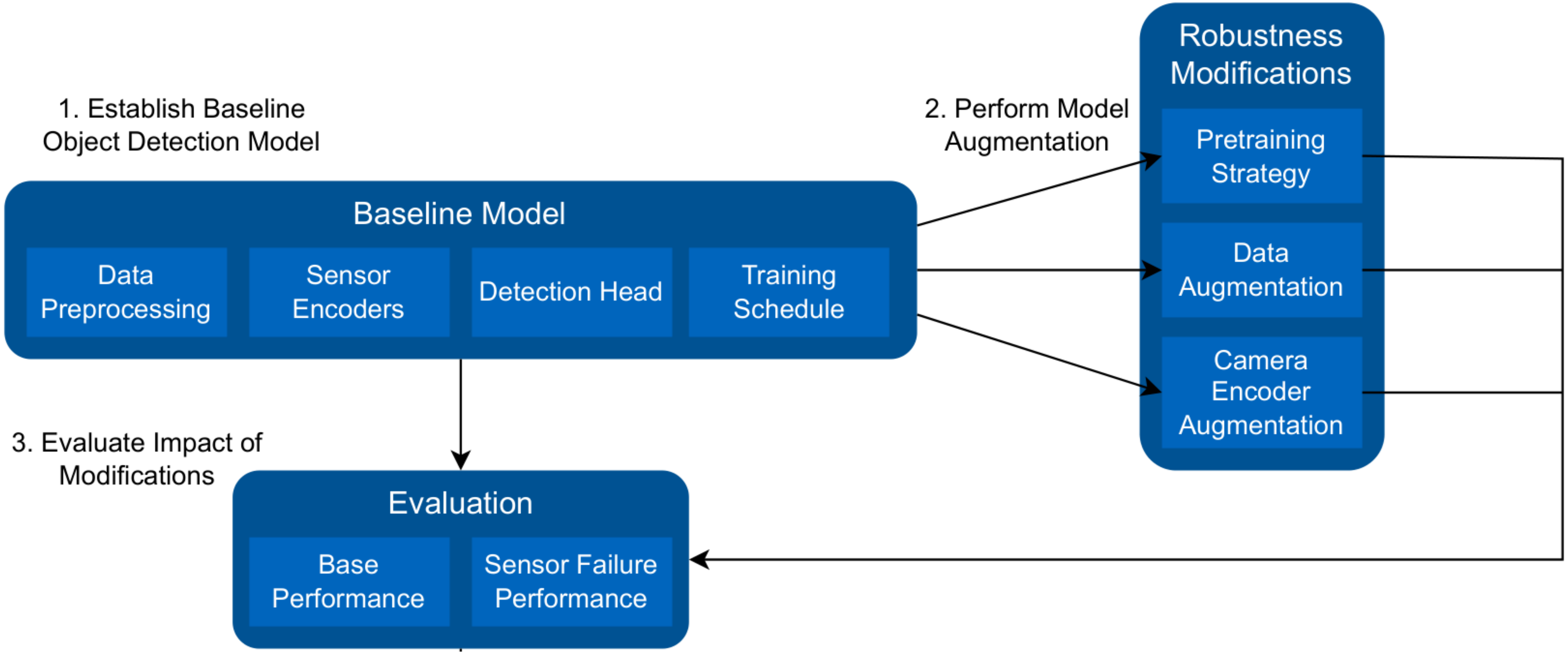
Data Preprocessing | Sensor Encoders | Detection Head | Training Schedule



Approach: Modifications

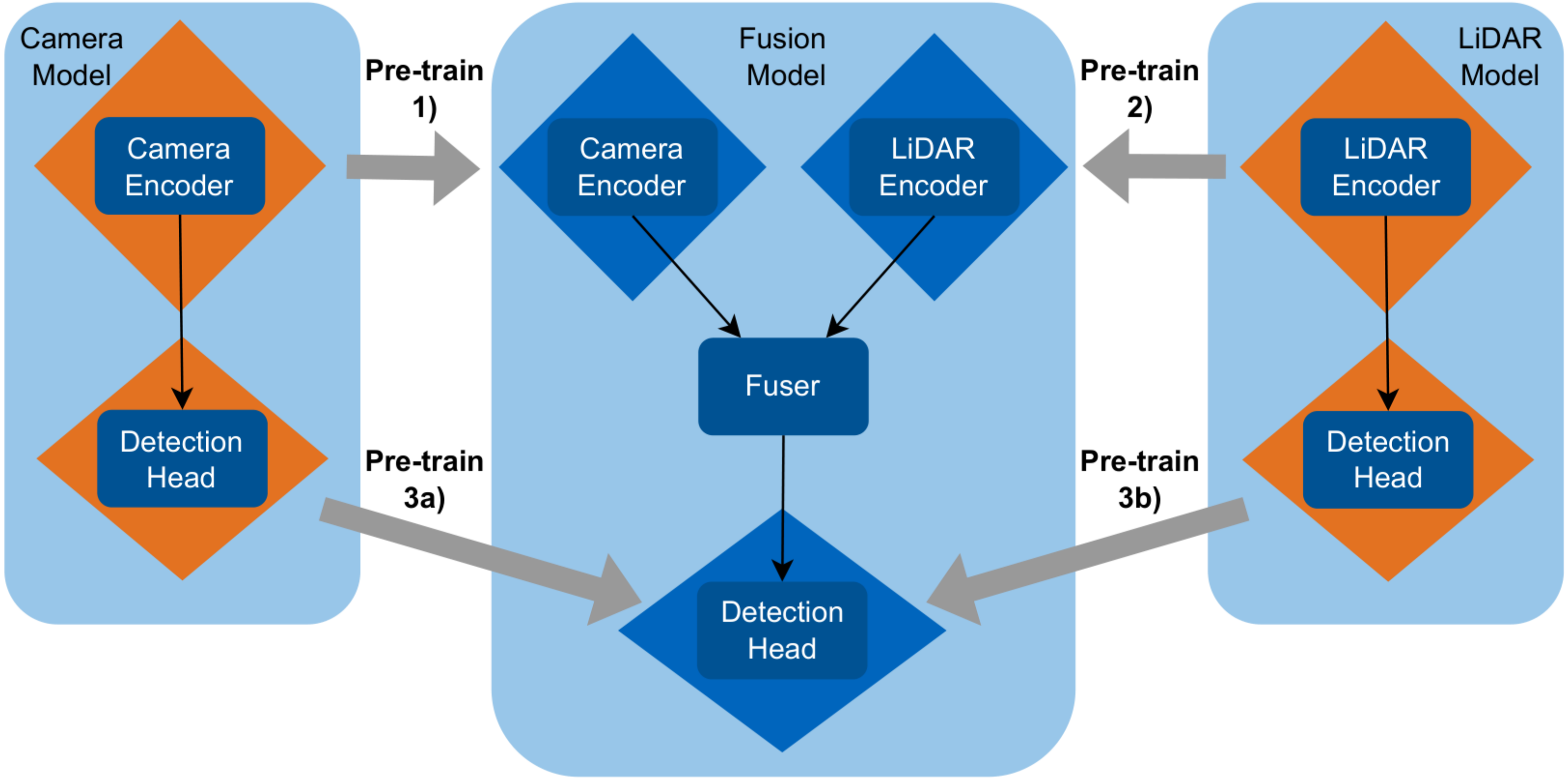


Approach: Evaluation



Method #1: Pretraining Strategy

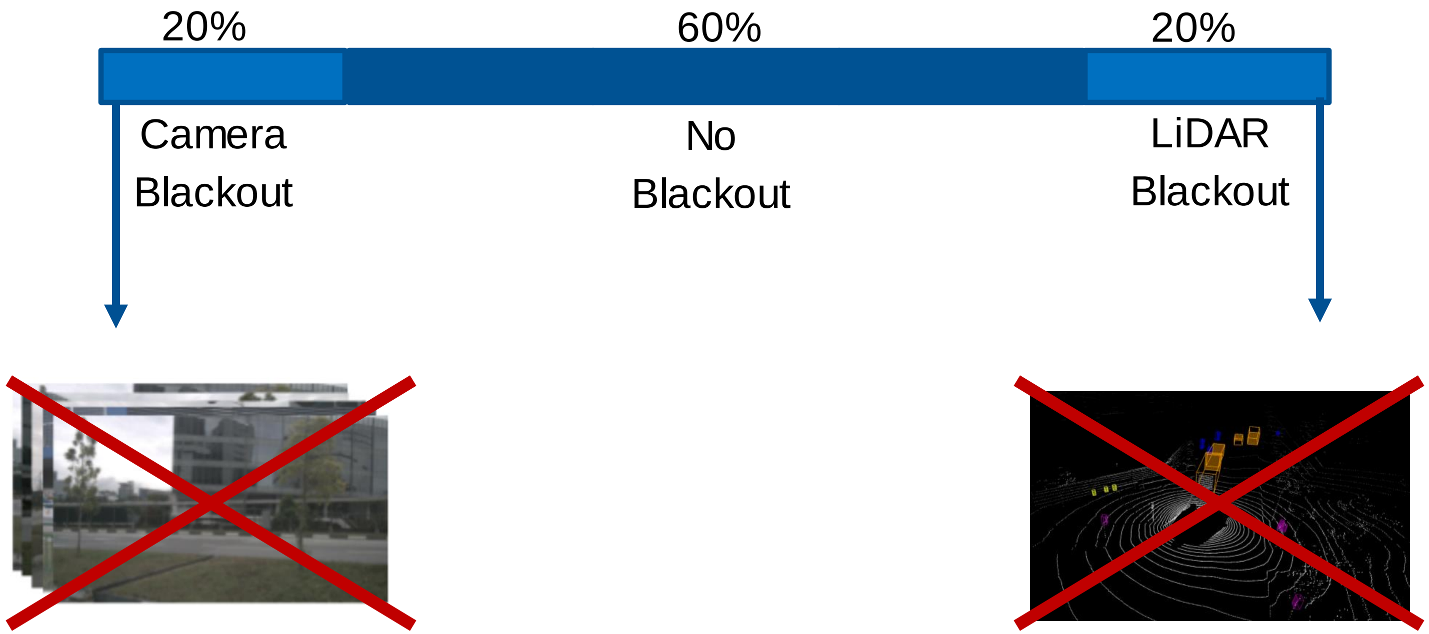
Original BEVFusion: **only pretrain LiDAR** model



➔ Try different variations

Method #2: Data Augmentation

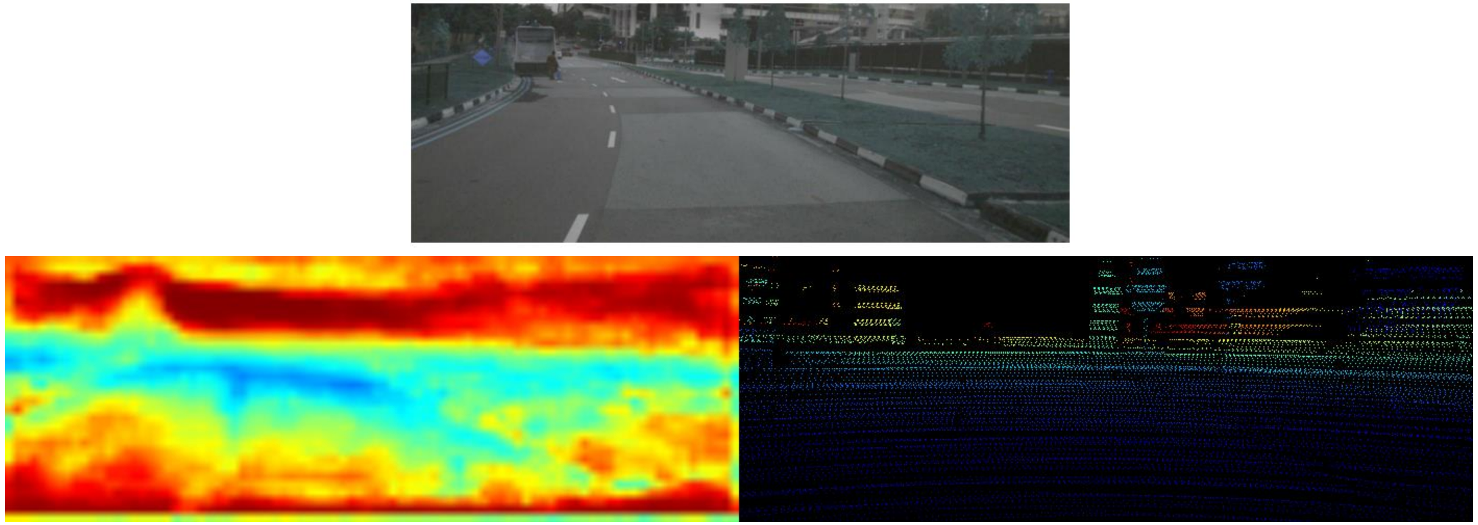
Simulate **total** sensor failure during training.



Blackout is applied randomly for each batch at 20% chance.

Method #3: Strengthen Camera Encoder

Strengthening the camera model could improve performance when the LiDAR sensor fails.



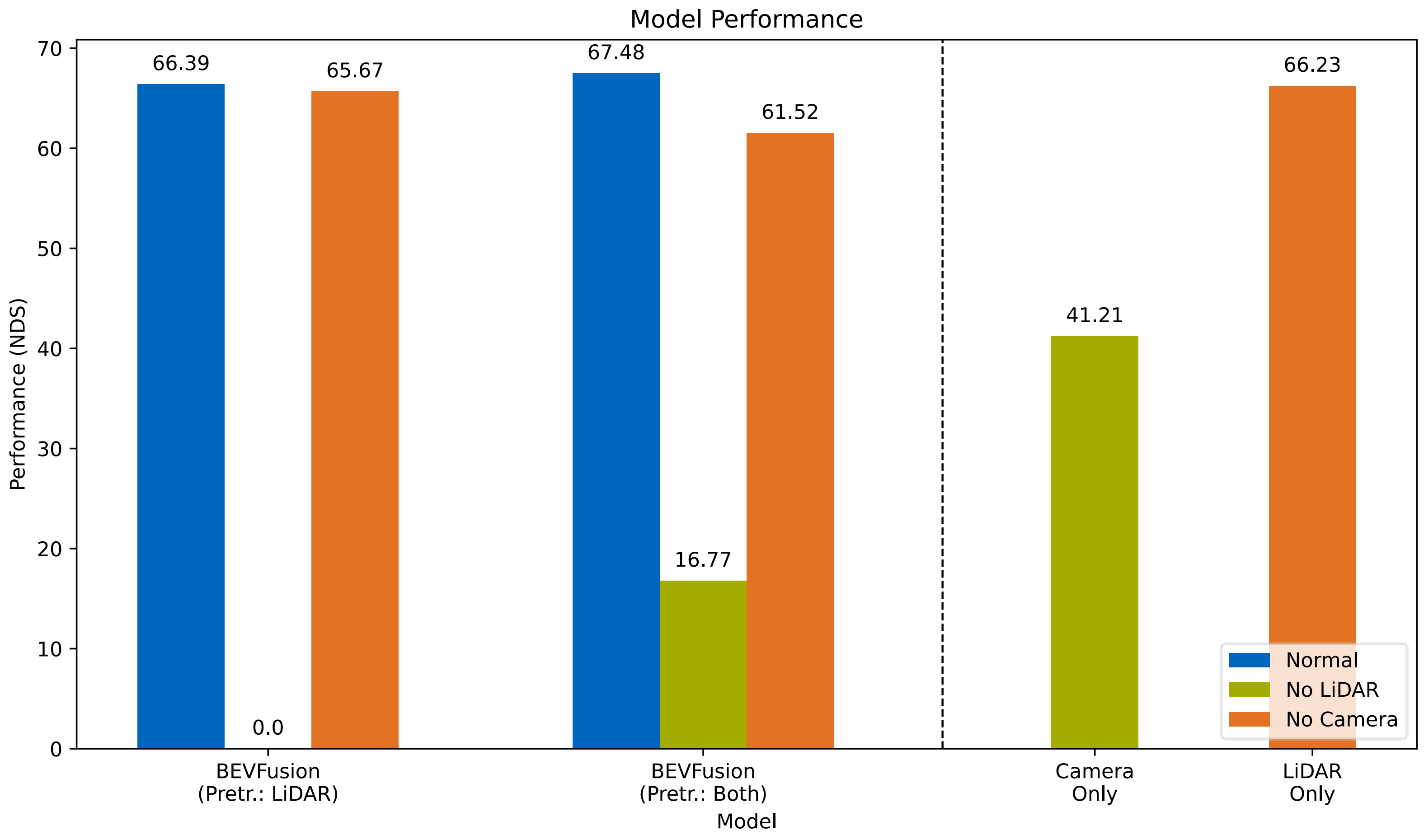
Camera depth prediction	True LiDAR-based depth

The camera model **explicitly predicts** the **depth** of each pixel.
This prediction is very inaccurate.

=> Improve prediction by using LiDAR-based depth as supervision during training

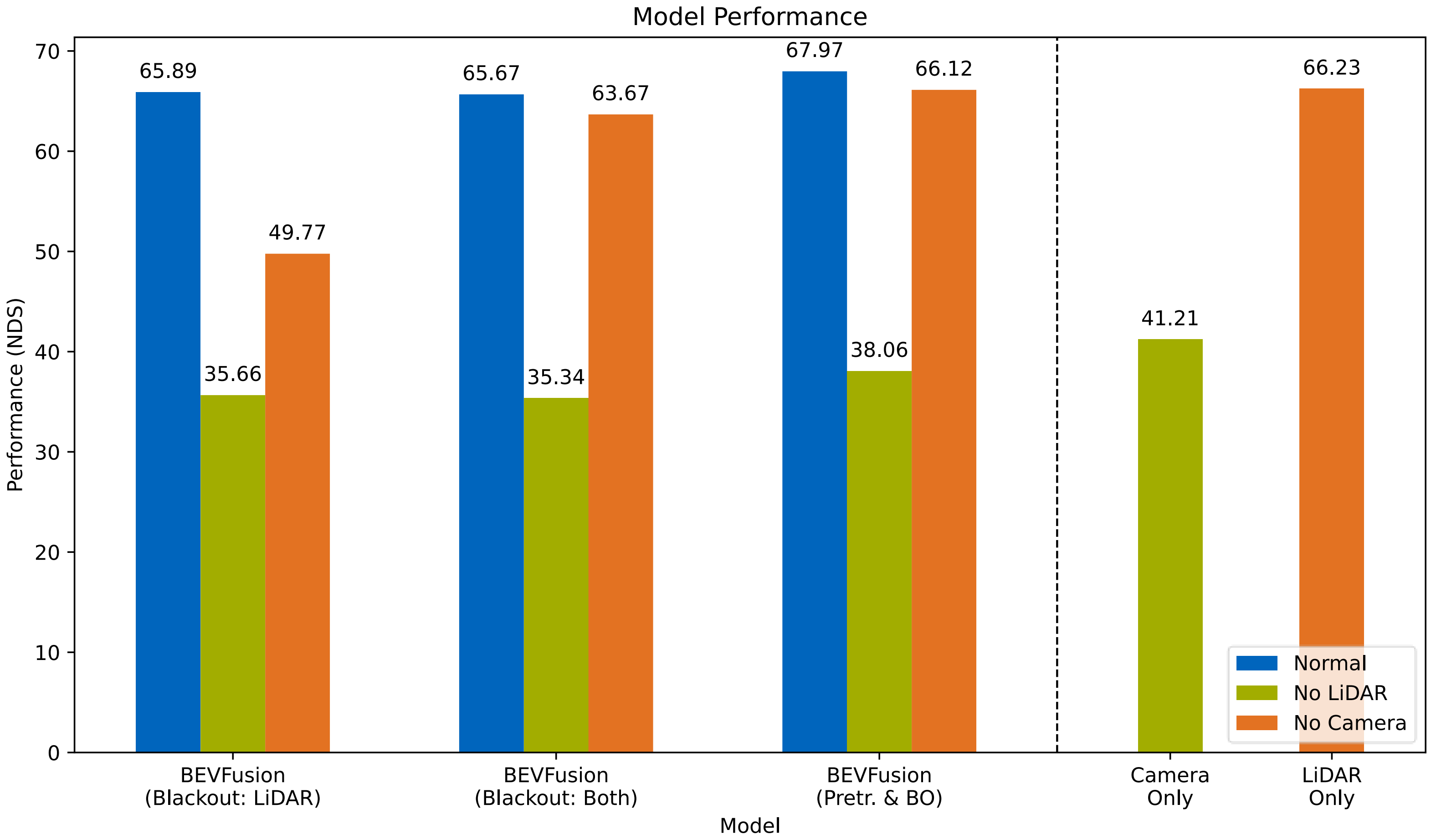
4 Li et al. “BEVDepth: Acquisition of Reliable Depth for Multi-view 3D Object Detection”

Results: Pretraining



=> Pretraining has a noticeable effect on robustness against sensor failure

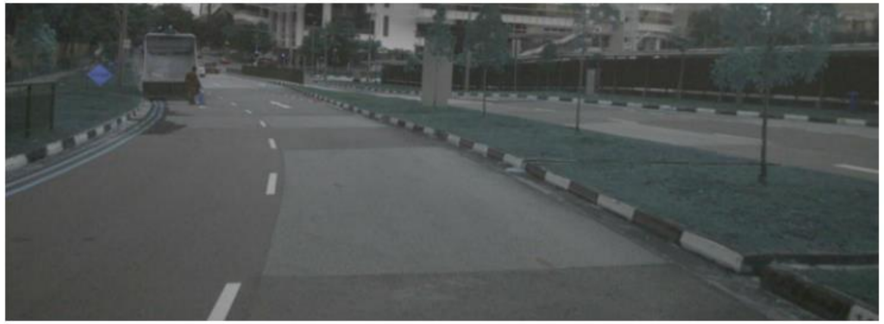
Results: Data Augmentation



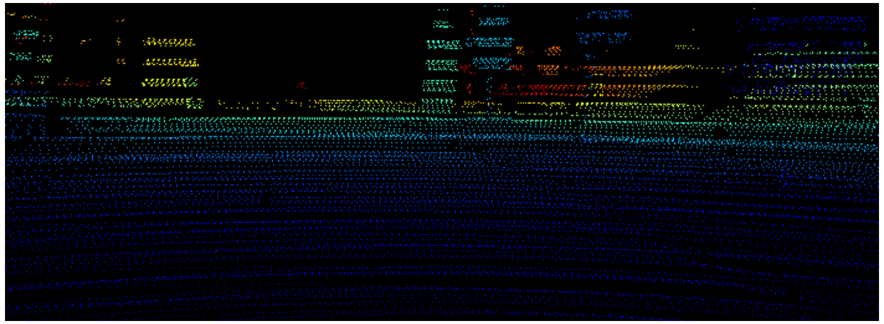
=> <u>Data augmentation has a substantial effect on robustness against sensor failure</u>

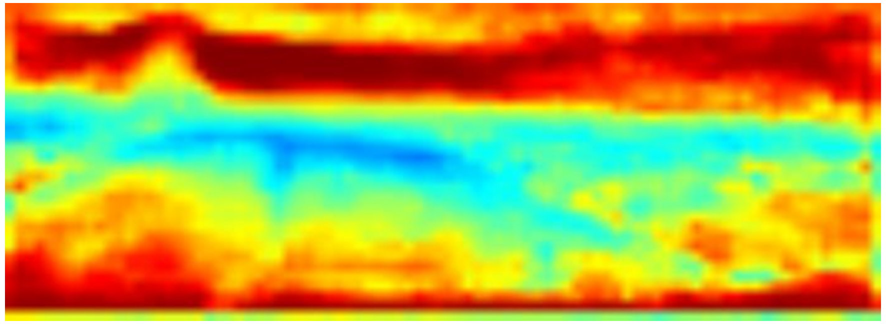
Results: Camera Encoder

Camera view

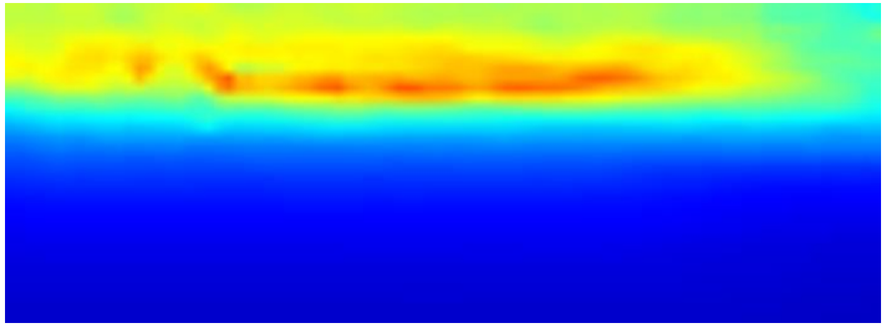


LiDAR-based depth





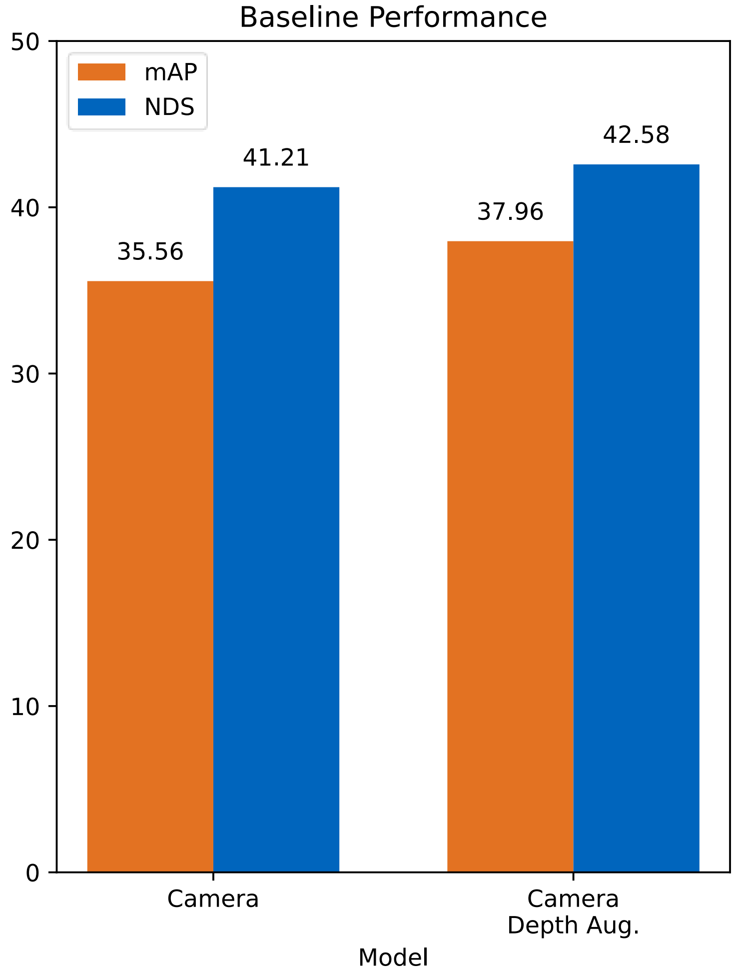
Original depth prediction



Improved depth prediction

Results: Camera Encoder

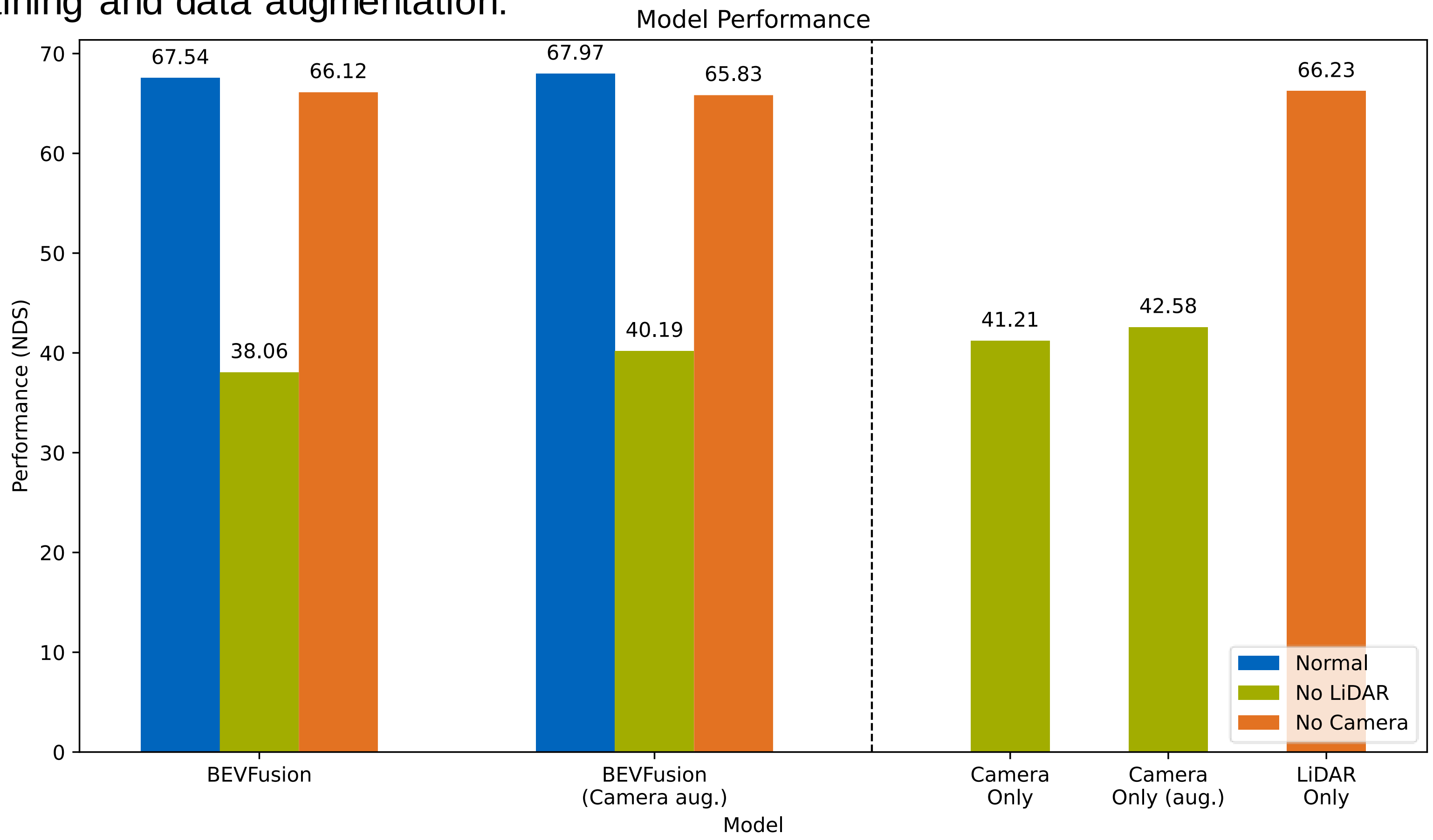
Noticeable increase in performance for the base camera model.



➔ Does this translate to the fusion model?

Results: Camera Encoder

Effect, including pretraining and data augmentation:

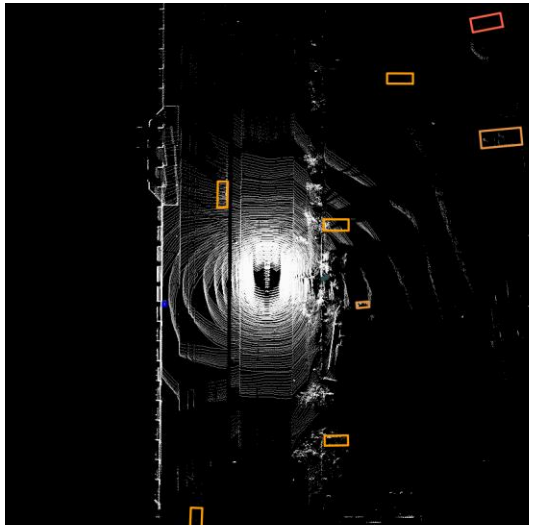


=> <u>Increased camera model performance directly translates to improved robustness during fusion</u>

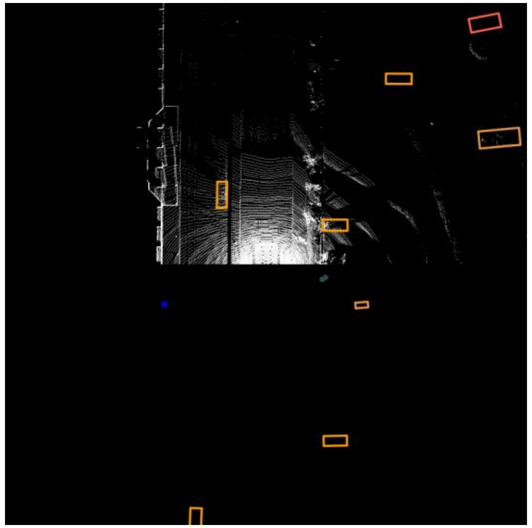
Generalization to partial sensor failure

So far, the focus has been only on **full** sensor failure (blackout)

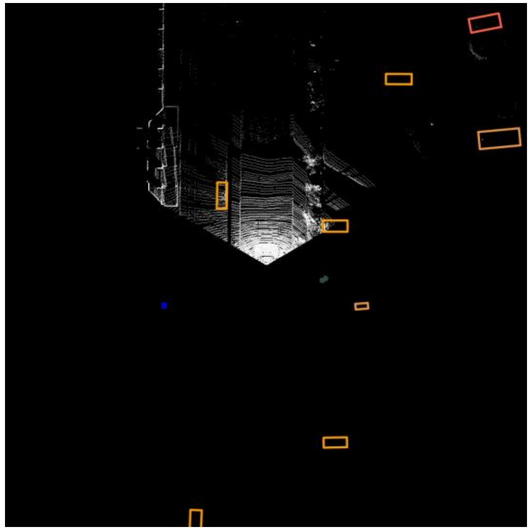
Do the presented methods improve performance in the event of **partial** sensor failure?



FoV 360°

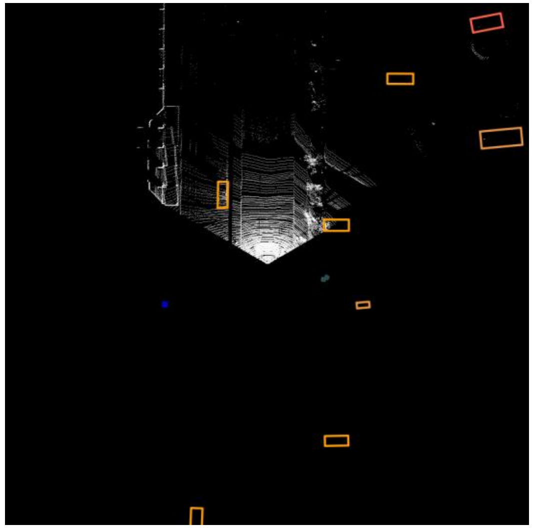


FoV 180°

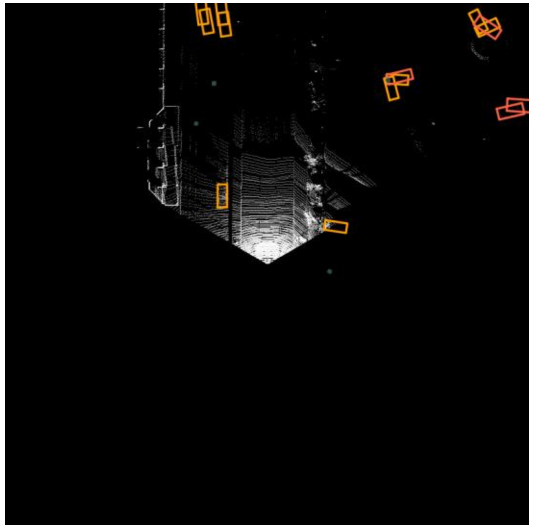


FoV 120°

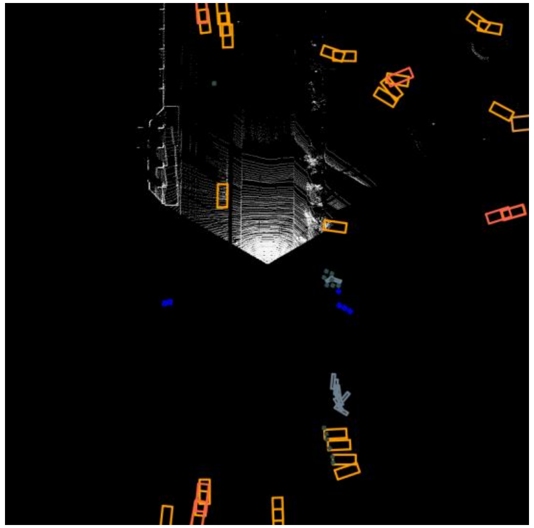
Generalization to partial sensor failure



Ground Truth



Baseline Model

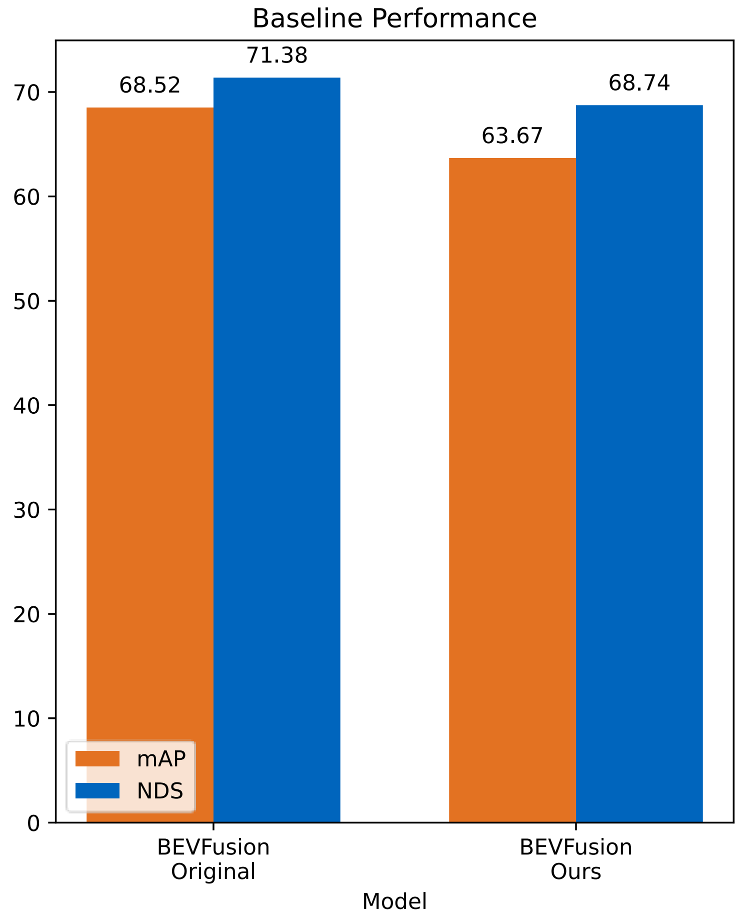


Augmented Model

=> Performance for **partial** sensor failure **increases significantly**

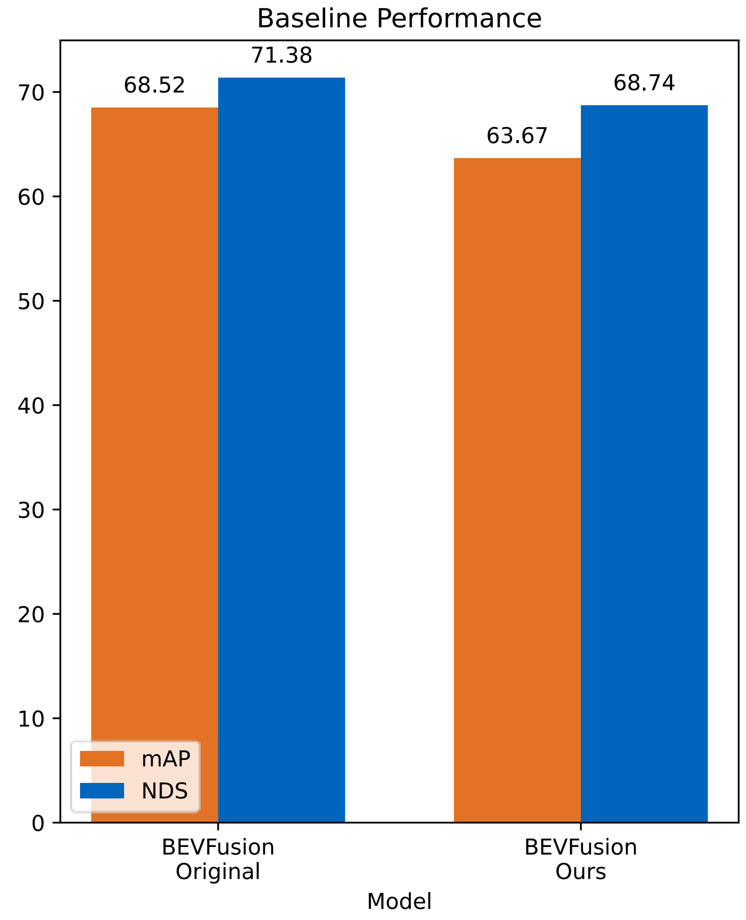
Limitations

- Reduced baseline performance vs. BEVFusion (resource and time constraints)
  - Higher performance might lead to trade-off with robustness



Limitations

- Reduced baseline performance vs. BEVFusion (resource and time constraints)
  - Higher performance might lead to trade-off with robustness
- Limited camera model performance
  - Possible trade-off between LiDAR and camera contribution
- Generalization to other sensor failure types unclear

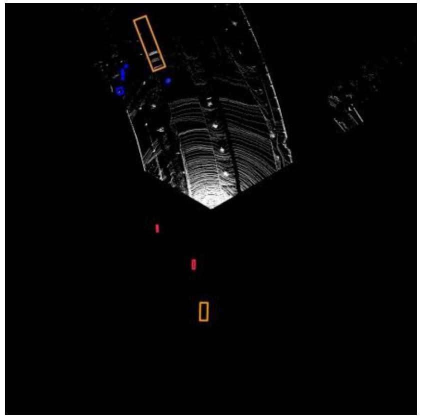


Summary

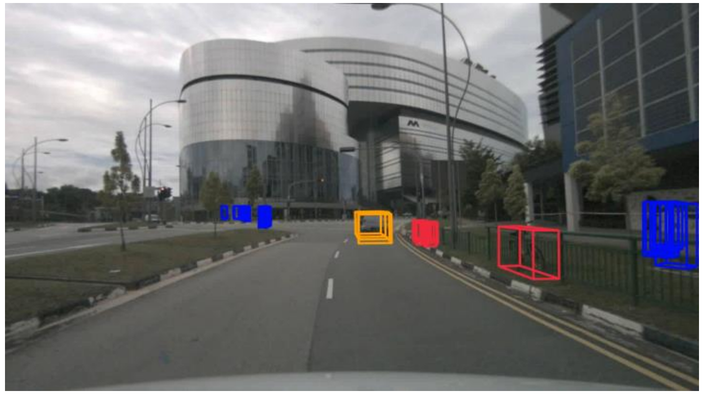
Visualization of LiDAR sensor failure

Before and after:

LiDAR view: Ground truth



Back view: Original model



Back view: Augmented model

Summary

Existing sensor fusion methods struggle in the event of sensor failure

- The presented methods **significantly** **increase robustness** against sensor failure
  - **Both** the camera failure and the LiDAR failure **cases** are **improved**
  - Strong results are achieved when using only **train-time** augmentations
- The **negative impact** on base performance is **negligible**
- Methods generalize well to partial sensor failure

Open Questions

- Do the presented methods also have a positive effect on **other** failure cases? (e.g. sensor miscalibration/misalignment)

- Are the presented methods also effective for **other** detection **tasks**? (e.g. map segmentation)

- When increasing base performance, is there a trade-off between a **robust** and a **high-performance** model?

Image Sources

https://www.eetindia.co.in/wp-content/uploads/sites/4/images/9d124bd9-84d5-4089-8e87-64831cb754d9.jpg