Advanced 6D Radar for Software-Defined Vehicles

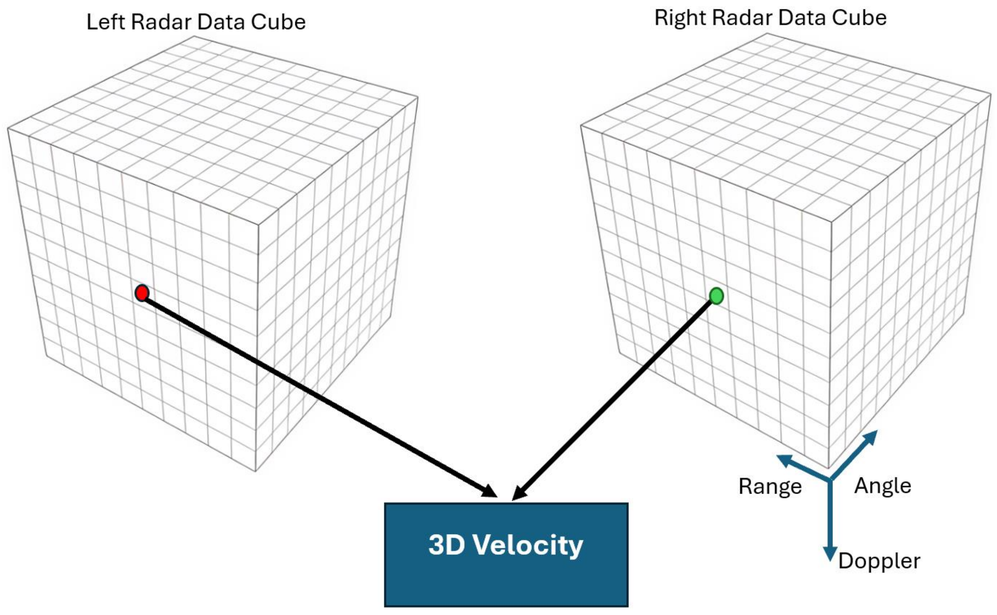
Oxford RF is the Market Leader in Solid-State 360^{O} and 6D Automotive Radar Sensing for Current and Future Sensing Requirements

3 Dims for Image Formation (Point Cloud)
+3 Dims for Target Kinematics, Image Correction & Enhancement (3D SAR)

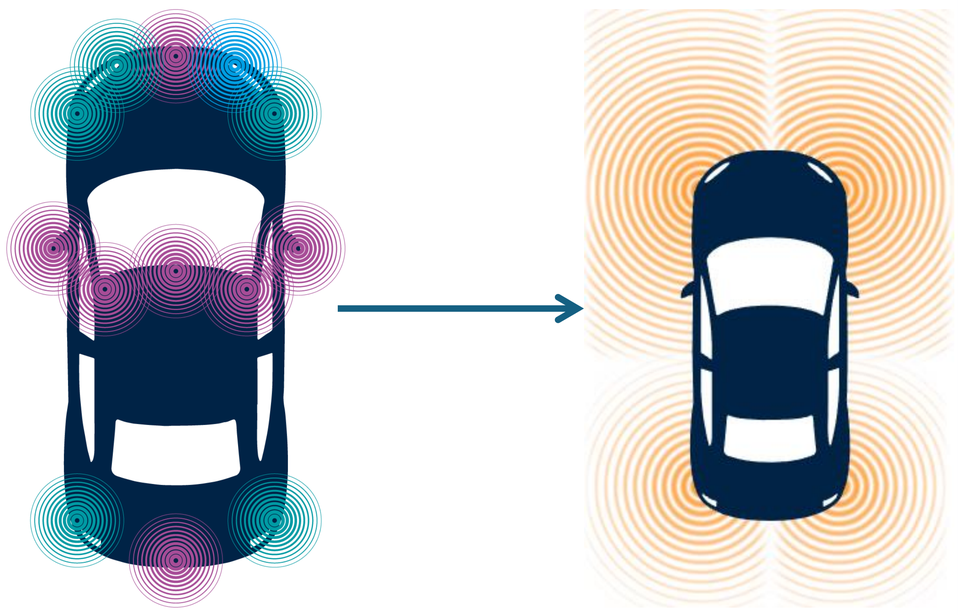
Field-of-View
1080-degree field of view = 270^{o} x 4

Multi-Aspect Networked Sensor System
Higher Diversity Gain
Tristatic Coverage in Almost All the Space

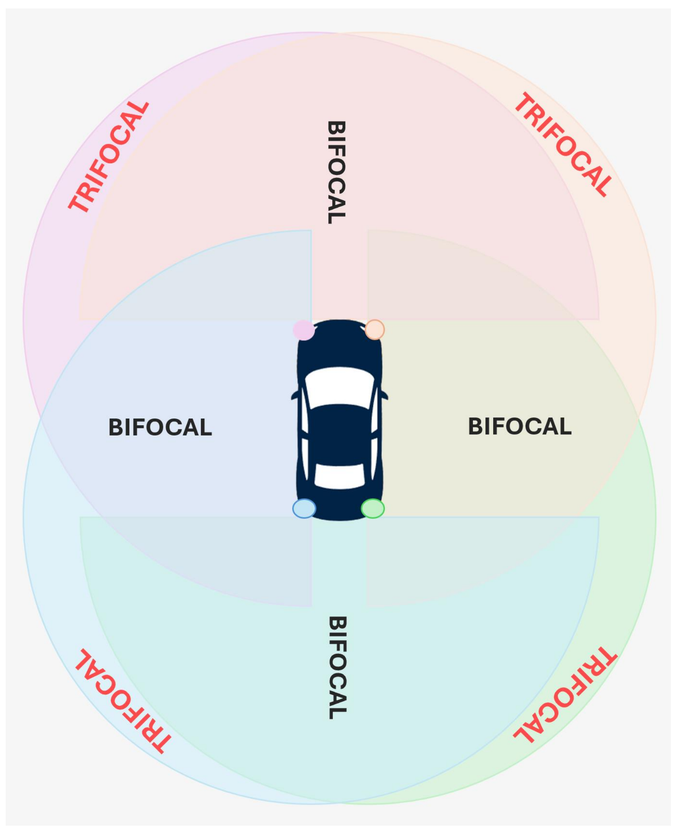
3D Velocity Computation



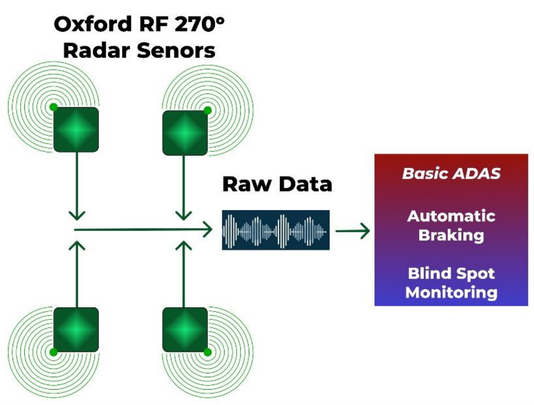
Fewer Wide FoV Sensors



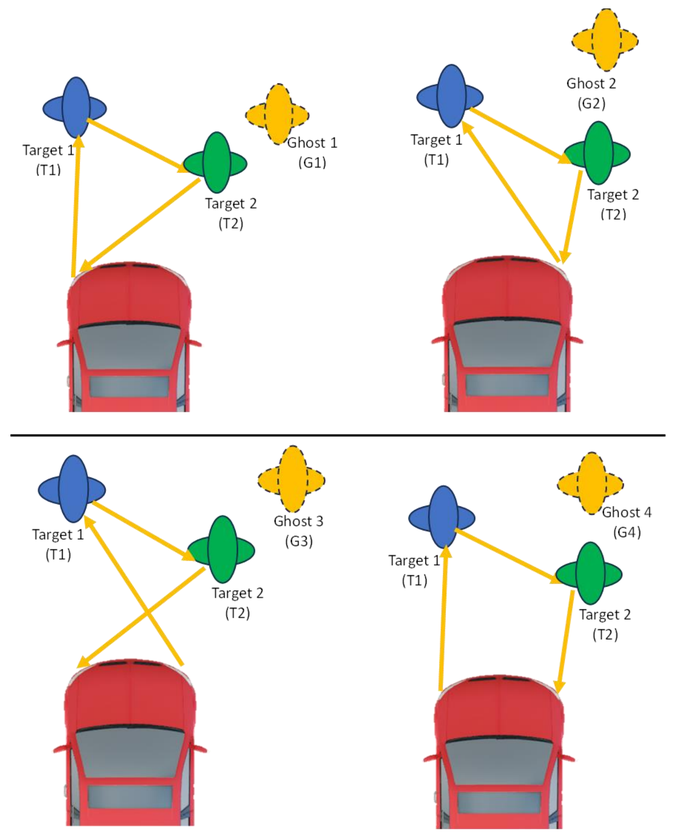
Enhanced Multi-Focal/Multistatic Sensing

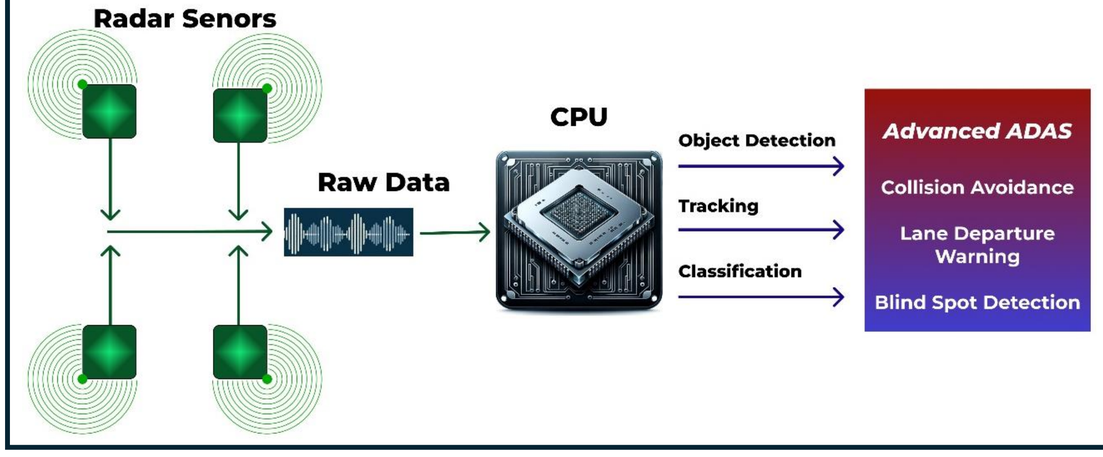


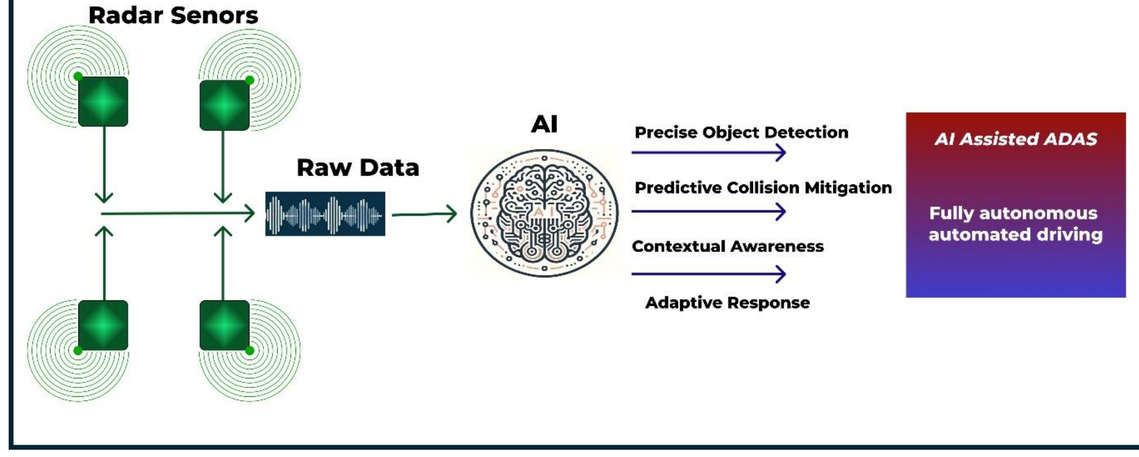
- Higher precision localisation of targets all around the vehicle with fewer sensor
- Enhanced Target Kinematics and Attitude Info
- Better cross-range 3D SAR imaging
- Redundancy/Augmentation
- Signal security
- Enhanced Accuracy & Resolution
- Better immunity to Noise and Weather Phenomena
- Improved Reliability!



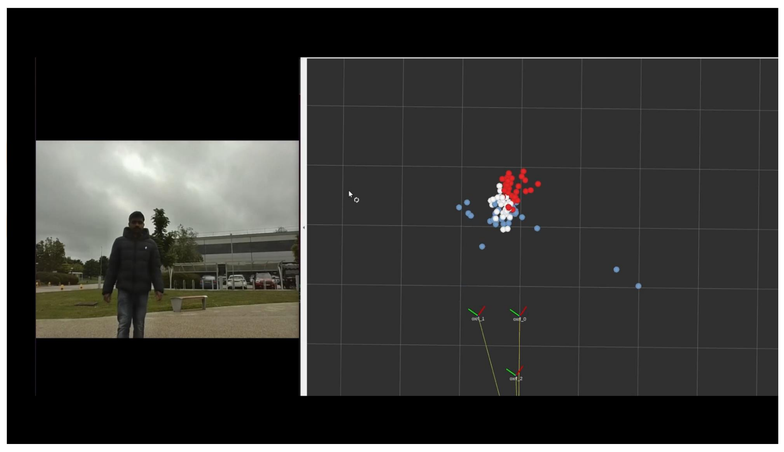
Multipath Ghost Rejection

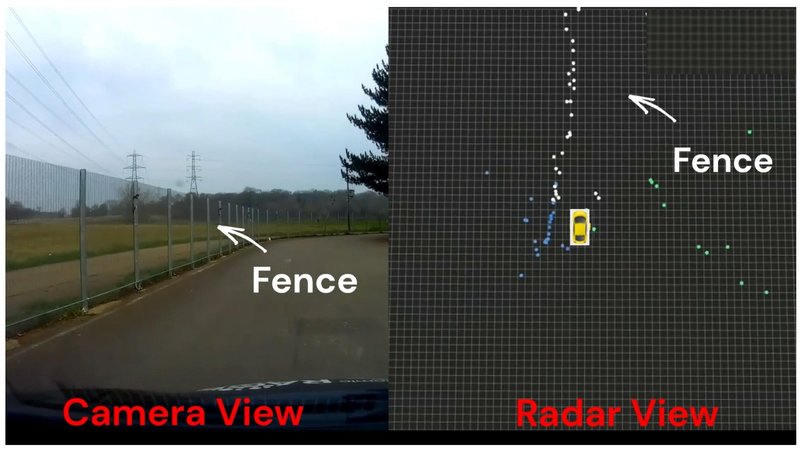




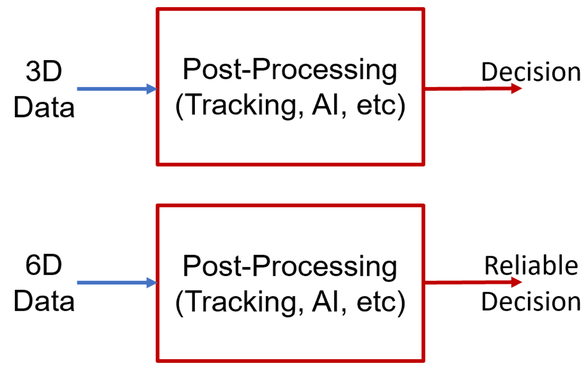


Trifocal Radar Trials and Data





Physics-Based Data => Better Driving Decision



OXFORD RF

Contact: Dr Kashif Siddiq & Mr Jomin Joy / Oxford RF Solutions Ltd
e-mails: Kashif.Siddiq@oxrf.co.uk / info@oxrf.co.uk

Sim4CAMSens

BEVSeg2GTA: Joint Vehicle Segmentation and GNNs for Ego Vehicle Trajectory Prediction in BEV

Sushil Sharma , Ciaran Eising and Mark Halton

Department Electronic and Computer Engineering, University of Limerick, Limerick, V94 T9PX Ireland

Abstract

Predicting the ego vehicle's trajectory is crucial for autonomous driving, yet real-world complexities challenge comprehensive capture within defined traffic regulations. Our paper introduces *BEVSeg2GTA*, which integrates perception and trajectory prediction into a unified system. By leveraging a deep network with EfficientNet and a Graph Neural Network (GNN), we improve accuracy in perceiving and predicting ego vehicle trajectories. Our proposed framework, *BEVSeg2GTA*, has been thoroughly evaluated on the nuScenes dataset.

Introduction

The development of autonomous driving technologies emphasizes the importance of safety, which relies on continuous coordination between perception, prediction, planning, and control systems, as depicted in the Figure below

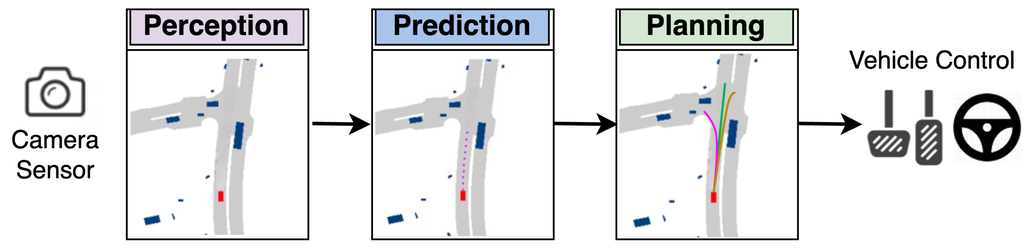


Figure 1: Self-driving systems use standard components for various tasks.

Our Proposed Methodology

Our proposed BEVSeg2GTA architecture: Our method for jointly segmenting vehicles and predicting the trajectory of the ego vehicle consists of several stages. Initially, we extract image features across different scales and integrate a camera-aware positional embedding to address perspective distortion. Following this, we employ map-view positional embedding and cross-attention layers to gather contextual insights from various viewpoints and enhance the segmentation accuracy of the ego vehicle. The segmented results are then passed through a graph neural network to generate embeddings representing the surrounding environment. These embeddings are further utilized as input to a probabilistic layer for trajectory prediction, leveraging contextual information from the surrounding scene.

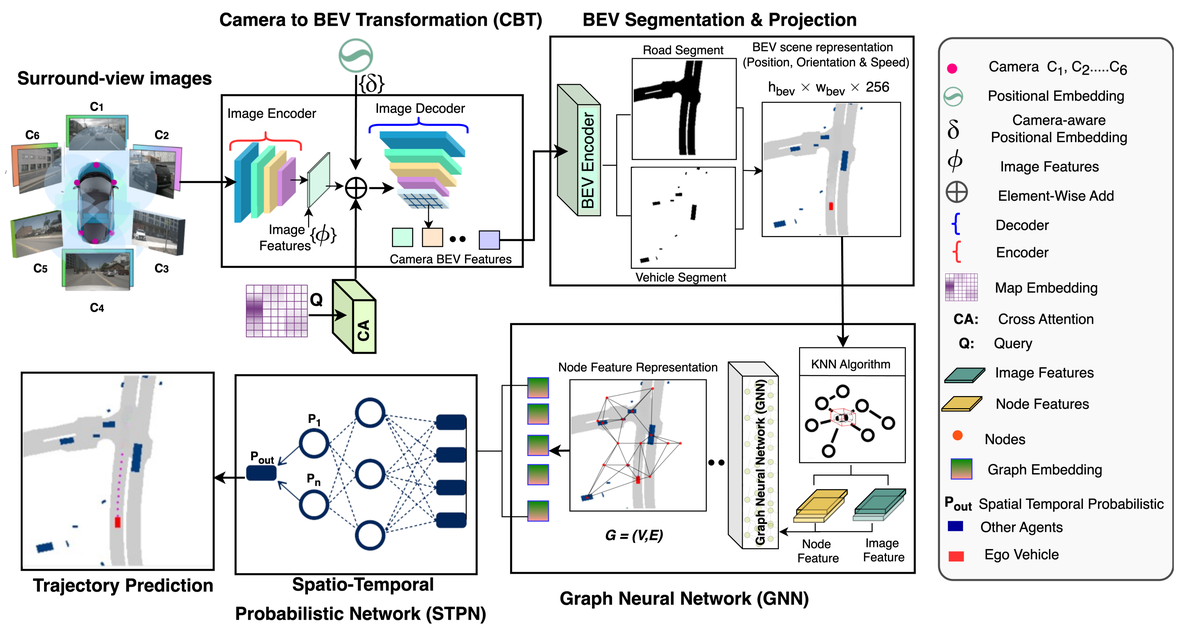


Figure 2: . Our proposed **BEVSeg2GTA** architecture

Graph Neural Network

Within the domain of ego vehicle trajectory prediction, the integration of GNNs represents a significant enhancement. This approach builds upon prior research efforts that have explored similar methodologies.

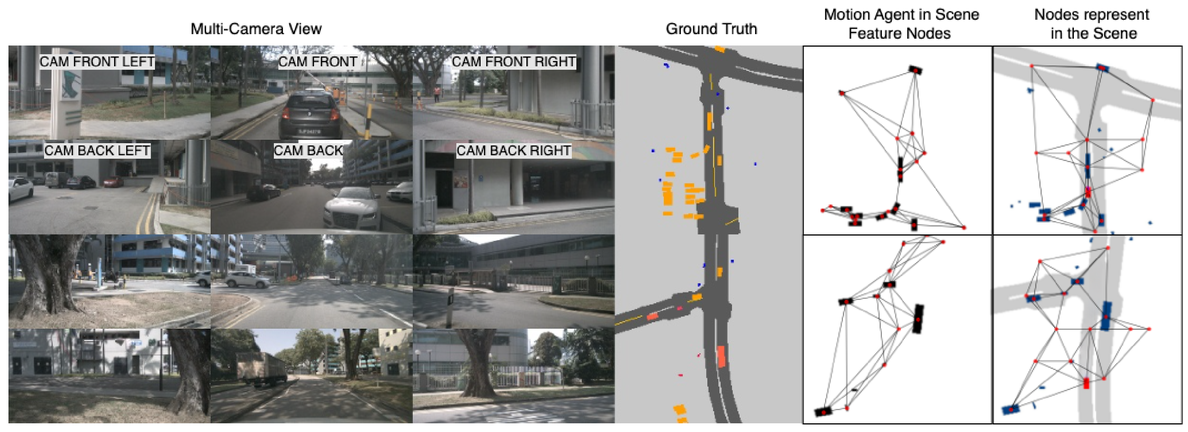


Figure 3: Motion agents with feature nodes and road representations.

Results

Qualitative outcomes of BEVSeg2GTA the combined vehicle segmentation and ego vehicle trajectory prediction. The six camera perspectives of nuScenes surrounding the vehicle are shown, with the top three facing forward and the bottom three facing backwards. Ground truth segmentation is displayed on the right. Our trajectory prediction approach integrates improved map-view segmentation with ego vehicle trajectory (second from the right), and it is compared to the LSS method and the CVT method (third and fourth from the right)

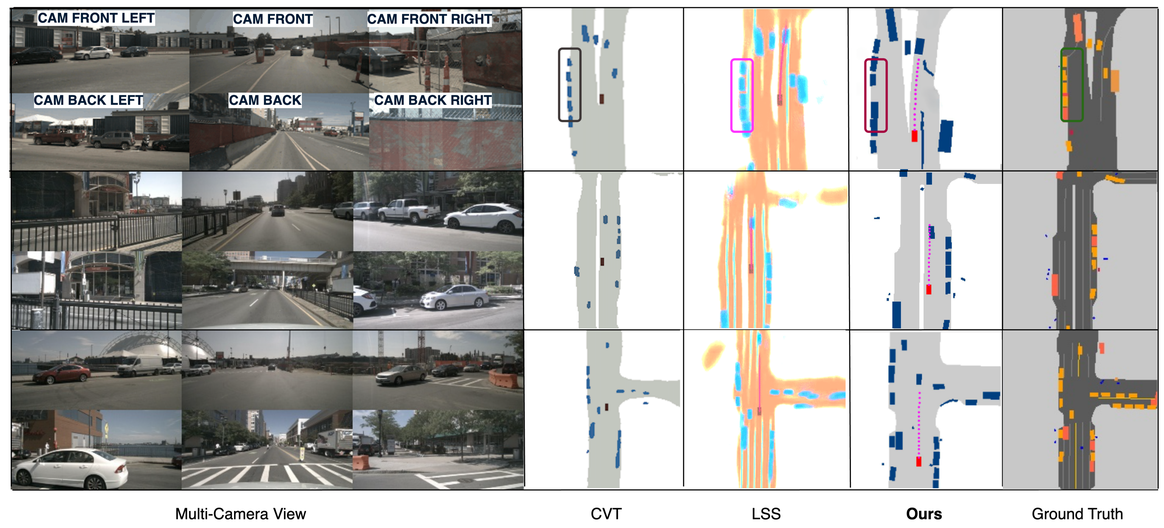


Figure 4: Qualitative outcomes of BEVSeg2GTA the combined vehicle segmentation and ego vehicle trajectory prediction

Method	$MinADE_5 \downarrow$	$MinADE_{10} \downarrow$	$MinADE_{15} \downarrow$	$MinFDE_5 \downarrow$	$MinFDE_{10} \downarrow$	$MinFDE_{15} \downarrow$	$MissRate_{5,2} \downarrow$	$MissRate_{10,2} \downarrow$
Const. Vel and Yaw	4.61	4.61	4.61	11.21	11.21	11.21	0.91	0.91
Physics oracle	3.69	3.69	3.69	9.06	9.06	9.06	0.88	0.88
CoverNet	2.62	1.92	1.63	11.36	-	-	0.76	0.64
Trajectron++	1.88	1.51	-	-	-	-	0.70	0.64
MTP	2.22	1.74	1.55	4.83	3.54	3.05	0.74	0.67
MultiPath	*1.78*	1.55	1.52	**3.62**	2.93	2.89	0.78	0.76
MHA-JAM	1.85	*1.24*	**1.03**	3.72	*2.23*	*1.67*	*0.60*	**0.46**
Ours	**1.63**	**1.19**	*1.06*	*3.63*	**2.13**	**1.65**	**0.56**	*0.51*

Table 1: Evaluation of competing methods on the nuScenes Dataset

Conclusion

In this study, we introduce BEVSeg2GTA, Initially, an encoder-decoder transformer processes images to create BEV representations. These representations are then used in a GNN to identify spatial relationships among agents. The K-Nearest Neighbors (KNN) algorithm generates a graph embedding, which is integrated with a Spatio-Temporal Probabilistic Network (STPN) to predict ego vehicle trajectory

Acknowledgment

This publication has emanated from research supported in part by a grant from Science Foundation Ireland under Grant number 18/CRT/6049.

Deformable Convolution Based Road Scene Semantic Segmentation of Fisheye Images in Autonomous Driving

Anam Manzoor, Aryan Singh, Ganesh Sistu, Reenu Mohandas, Eoin Grua, Anthony Scanlan, Ciarán Eising

1 Introduction

- ***Semantic Segmentation:*** is the process of labeling each pixel in an image with a semantic category, providing a detailed understanding of the scene's content.
- ***Fisheye Problem:*** Fisheye images, common in autonomous driving scenarios, pose challenges due to their wide field of view and geometric distortions, making it challenging to extract accurate spatial and geometric information.
- ***Why Deformable Convolution?:*** Deformable convolutions are used to address the challenges presented by fisheye imagery by allowing neural networks to adapt and learn from geometric distortions. This helps improve the accuracy of semantic segmentation by capturing intricate spatial relationships more effectively.

Research Goals

- Evaluate the effectiveness of Deformable Convolutions in segmenting fisheye images using U-Net.
- Explore multi-view scene processing to assess model versatility.
- Investigate how incorporating images from various scenes can improve segmentation accuracy.
- Highlight the role of Deformable Convolutions in facilitating view-agnostic learning and their intrinsic advantages for fisheye image.

Figure 1. Four fisheye cameras mounted around the vehicle to provide complete 360-degree coverage.

2 Methodology

- ***Objective:*** Our study aims to explore the effectiveness of Deformable Convolutions in semantic segmentation of fisheye images using U-Net architectures.
- ***Approach:*** We leverage U-Net and Residual U-Net architectures, incorporating Deformable convolutions to enhance spatial understanding in fisheye images.
- ***Model Architecture:*** We modify traditional U-Net and Residual U-Net models by replacing convolutional blocks with Deformable convolutions, as discussed in previous literature.
- ***Dataset:*** Utilizing the **WoodScape** dataset, comprising 10,000 annotated fisheye images from four view angles, we evaluate model performance across diverse real-world scenarios.
- ***Experimental Setup*:** Implementing vanilla and Residual U-Net models with and without Deformable convolutions, we explore variants such as **V_U-Net, V_DeU-Net, R_U-Net**, and **R_DeU-Net**. The models are trained from scratch using PyTorch on NVIDIA **GeForce RTX 3080 GPU**.
- ***Evaluation:*** Model performance is evaluated based on standard segmentation metrics including accuracy and IoU score, with training continuing for 50 epochs using Adam optimizer and a batch size of 1 for smoother training.

Model Workflow Diagram

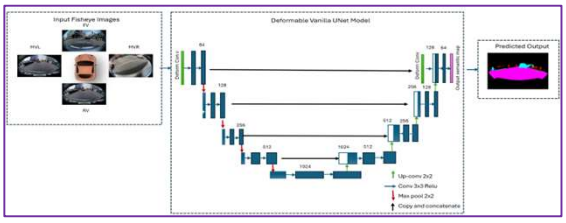


Figure 2. Baseline Vanilla DeU-Net model where Deformable Convolution block injected into the first layer of the encoder and last layer of decoder path to better account the spatial and geometric characteristics of fisheye images during training.

3 Results and Analysis

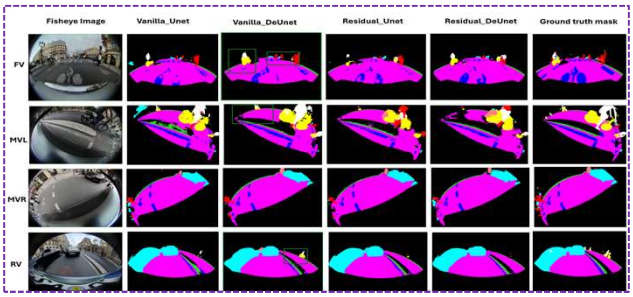


Figure 3. Visualizations of results on Woodscape fisheye images and corresponding ground truth masks are presented across baseline models including Vanilla_U-Net, Residual_U-Net, Deformable_U-Net, and Deformable_Residual model. Notably, the visualization performance with Vanilla_DeU-Net surpasses the other models, as indicated with green boxes compared to ground truth masks, with particular emphasis on distorted edges.

Sr. #	Categories	V_Unet_ce		V_deUnet_ce		V_deUnet_mwf		V_deUnet_wf		R_Unet_ce		R_deUnet_ce		R_deUnet_wf		R_deUnet_EnAll_wf	
		Acc ↑	IoU ↑	Acc ↑	IoU ↑	Acc ↑	IoU ↑	Acc ↑	IoU ↑	Acc ↑	IoU ↑	Acc ↑	IoU ↑	Acc ↑	IoU ↑	Acc ↑	IoU ↑
1	Background	0.98	0.96	0.99	0.97	0.99	0.96	0.98	0.95	0.98	0.95	0.98	0.96	0.97	0.94	0.98	0.93
2	Road	0.97	0.92	0.97	0.93	0.97	0.92	0.96	0.88	0.96	0.89	0.97	0.91	0.96	0.87	0.94	0.86
3	Lanemark	0.64	0.61	0.54	0.53	0.55	0.53	0.40	0.39	0.39	0.38	0.53	0.51	0.32	0.31	0.27	0.27
4	Curb	0.57	0.49	0.56	0.50	0.55	0.48	0.53	0.48	0.38	0.36	0.46	0.42	0.47	0.43	0.41	0.38
5	Person	0.28	0.23	0.37	0.32	0.34	0.27	0.29	0.23	0.04	0.04	0.15	0.12	0.27	0.19	0.17	0.15
6	Rider	0.44	0.41	0.54	0.46	0.42	0.36	0.51	0.44	0.18	0.16	0.17	0.15	0.10	0.10	0.42	0.33
7	Vehicles	0.91	0.84	0.90	0.85	0.90	0.85	0.87	0.76	0.89	0.73	0.88	0.81	0.84	0.67	0.71	0.58
8	Bicycle	0.60	0.52	0.74	0.53	0.49	0.43	0.47	0.38	0.33	0.28	0.45	0.37	0.07	0.06	0.19	0.18
9	Motorcycle	0.46	0.37	0.73	0.47	0.54	0.42	0.56	0.44	0.14	0.10	0.29	0.22	0.44	0.31	0.29	0.22
10	Traffic Sign	0.09	0.09	0.11	0.12	0.10	0.10	0.39	0.36	0.02	0.01	0.02	0.02	0.18	0.17	0.09	0.09
Average mIOU			0.93		0.93		0.91		0.88		0.89		0.89		0.85		0.84
Average Accuracy		0.99		0.99		0.99		0.98		0.94		0.94		0.93		0.98	

Table 1. The table presents class-specific accuracy and IoU scores for various configurations of Vanilla and Residual U-Net models, including their deformable versions, trained with different loss functions: cross-entropy (ce), standard focal loss (nwf), and weighted focal loss (wf). The highest IoU score for each class is highlighted in green for improved clarity.

4 Summary

- The study investigated the effectiveness of integrating Deformable convolutions for semantic segmentation of fisheye images in the automotive domain.
- Four models were explored as a baseline: **Vanilla_U-Net, Residual_U-Net, Deformable_U-Net**, and **Deformable_Residual_U-Net**.
- ***Findings:*** Through experiments, it was observed that integrating Deformable convolutional blocks allows for more refined and efficient modeling of fisheye images.
- ***Future Directions***: could explore incorporating these blocks into alternative backbone architectures of Transformers or multitask networks to enhance semantic segmentation for synthetic and real-world datasets.

References

- [Dai et al., 2017] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., and Wei, Y. (2017). Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).
- [[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, pages 234–241 Springer.
- [El Jurdi et al., 2023] El Jurdi, R., Sekkat, A. R., Dupuis, Y., Vasseur, P., and Honeine, P. (2023). Fully residual unet-based semantic segmentation of automotive fisheye images: a comparison of rectangular and deformable convolutions. Multimedia Tools and Applications, pages 1–23.

Acknowledgement

*This work was supported, in part, by the Science Foundation Ireland grant **13/RC/2094 P2** and co-funded under the European Regional Development Fund through the Southern \& Eastern Regional Operational Programme to Lero - the Science Foundation Ireland Research Centre for Software **(www.lero.ie)***

HOST INSTITUTION

PARTNER INSTITUTIONS

FUNDED BY:

Camera-Radar Fusion in Autonomous Vehicles For Perception Tasks

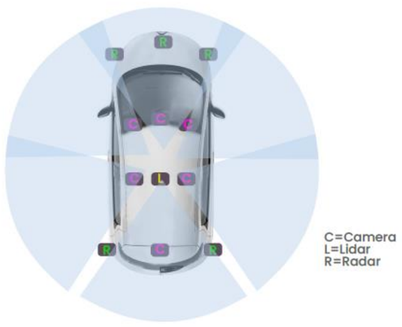
Authors: Sanjay Kumar **Supervisor:** Ciaran Eising, University of Limerick

1

INTRODUCTION:

In the field of autonomous driving, developing robust and accurate perception systems is essential for ensuring safety and efficiency. These systems integrate multiple sensors: cameras for high-resolution visuals, LiDAR for depth information though less effective in adverse weather and radar for reliable long-range detection in challenging conditions [1]. The combination of these technologies enables comprehensive environmental perception, vital for complex driving tasks like motion prediction, path planning, and automated control.

However, integrating these diverse inputs involves complex challenges in sensor selection and data fusion at various levels (early, intermediate, and late), which are key research areas to advance autonomous vehicle capabilities [2].



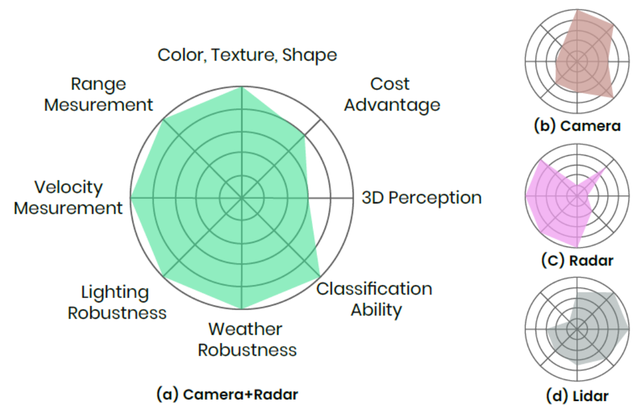
Vehicle with Multiple Sensors

2

Objective:

The primary objective of our research are:

- ❖ To assess the individual and combined capabilities of cameras, and radar in detecting and interpreting complex driving environments.
- ❖ To explore intermediate and late fusion techniques to optimize the integration of camera and radar data for enhanced accuracy in environmental perception.
- ❖ To evaluate the performance of the integrated system in diverse weather conditions and driving scenarios to ensure consistent functionality.

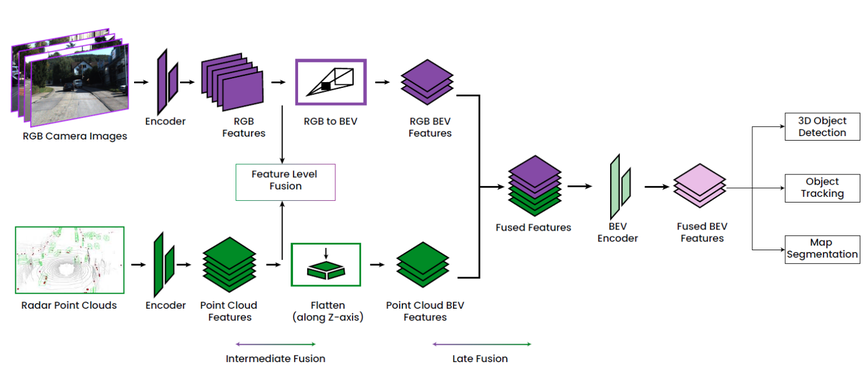


Characteristics of Sensors [2]

3

Proposed Methodology:

- ❖ Our approach uses radar point clouds and RGB camera images, encoding these data sources to extract and convert features into a Bird's Eye View (BEV) format. This involves specifically transforming RGB features to BEV and flattening radar data into BEV, optimizing them for fusion.
- ❖ The prepared BEV features from both radar and RGB data are fused at either an intermediate or late stage, ensuring thorough integration. This process uses a BEV encoder to align and optimize the features for accurate data synthesis.
- ❖ The resulting fused BEV features are crucial for performing advanced functions such as tracking, map segmentation, and 3D object detection, enhancing the system's capabilities in dynamic and complex environments.



Proposed Workflow [3]

4

Conclusion:

In conclusion, our research integrates advanced sensors and fusion techniques to enhance autonomous vehicle perception systems. By merging camera and radar data, we enhance accuracy and adaptability under various conditions. This leads to more reliable and efficient autonomous driving in complex environments.

References:

[1] Wu, Zizhang, Guilian Chen, Yuanzhu Gan, Lei Wang, and Jian Pu. "Mvfusion: Multi-view 3d object detection with semantic-aligned radar and camera fusion." In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2766-2773. IEEE, 2023.

[2] Yao, Shanliang, Runwei Guan, Xiaoyu Huang, Zhuoxiao Li, Xiangyu Sha, Yong Yue, Eng Gee Lim et al. "Radar-camera fusion for object detection and semantic segmentation in autonomous driving: A comprehensive review." *IEEE Transactions on Intelligent Vehicles* (2023).

[3] Liu, Zhijian, Haotian Tang, Alexander Amini, Xinyu Yang, Huizi Mao, Daniela L. Rus, and Song Han. "Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation." In *2023 IEEE international conference on robotics and automation (ICRA)*, pp. 2774-2781. IEEE, 2023.

Acknowledgements:

This proposed work is supported, in part, by the Science Foundation Ireland grant 13/RC/2094 P2 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero.

HOST INSTITUTION

PARTNER INSTITUTIONS

FUNDED BY:

MEASUREMENT TECHNIQUES

AUTOMATED MEASUREMENT OF VISIBLE WARNINGS ON VEHICLE INSTRUMENT CLUSTER

TEAM MEMBERS: William Dunnion, Joshua Haywood, Daniel Rodrigues, Réka Kassai, George Giles, George Maynard

PROJECT MENTOR: Dr. Gian Matteo Bianchi

An **Advanced Driver-Assistance System (ADAS)** feature validation improvement project

1 PROBLEM

Test engineers record a video of a test vehicle's instrument cluster to see when ADAS features are being activated during a test

The test video is manually processed by an analyst, who identifies when ADAS features are being activated

The current process is not automated and relies on hours of a data analyst's time to process

The current process for acquisition of AAD cluster data is not time synced to the rest of the AAD test data

All the above means the process of validation and defect analysis is very complex, monotonous and time consuming

2 OBJECTIVES

Research & Planning	To research and investigate existing image processing projects to explore and potentially re-use previous work proof of concepts conducted by JLR
Design & Development	To develop: 1. A recording system for image capture with GigE PTP camera 2. An image processing model for single feature detection for one specific car line
Integration	To integrate into JLR's current MT/GT data collection and processing systems
Testing	To analyse: 1. Icon detection error margin 2. Accuracy of time sync 3. Robustness testing for camera position

3 METHODS

RESEARCH & PLANNING

Lane Keep Assist (LKA)

Feature only operates within the specified speed range

When the vehicle is projected to breach the lane marking, LKA initiates an intervention

A visual warning must be displayed

Machine Learning

Investigation of existing JLR solutions

YOLO Algorithm:
- Open source
- Availability of documentation & community support
- Supports python

DESIGN & DEVELOPMENT – RECORDING SYSTEM

- Script developed to connect to a GigE camera
- Upload a camera configuration and record frame by frame
- Images saved to a named location in a .png format
- Images named using the session date and timestamp collected from PTP calibrated system time

Connect to camera → Start recording frame by frame → Collect PTP timestamp from system → Save to chosen directory (png)

DESIGN & DEVELOPMENT – IMAGE PROCESSING PIPELINE

Image Preprocessing → Custom YoloV5 Object Detection model (ultralytics YOLOv5) → Output results to log file

- Images are processed through a custom YoloV5 model trained on a 5000+ image dataset
- Pipeline can be executed in one click by the test engineer
- Script developed which generates augmented images and labels to enhance training dataset
- Results output into a log file to be used by test engineers

INTEGRATION

- Integrated with the current Ground truth (GT) and Measurement Technique (MT) boot kits
- Vector CANape recording system

3 METHODS CONTINUED

- The IEEE1588 PTP protocol will be provided using an OXTS INS system

GigE camera → MT Core running CANape ← OXTS INS for PTP

VECTOR

TESTING

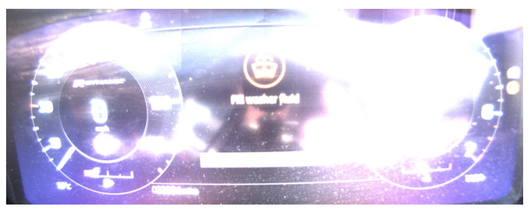
Vehicle or ECU rig → Image capture → Image processing pipeline → Timestamped detection file output

Rig Testing Pipeline
- Bench test rig setup for calibration
- Uses Vector CANalyzer to drive requisite signals to the display
- Uses in-house recording script to capture frames

Vehicle Testing Pipeline
- Camera mounted in vehicle
- Recording software for vehicle clusters implemented

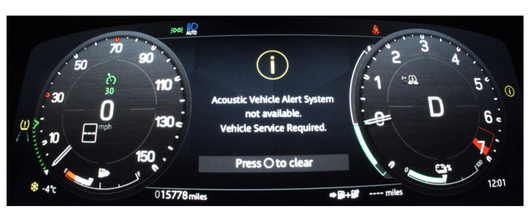
4 CHALLENGES

Glare (environmental factors)



- Sunny weather
- Region of interest (ROI) not specified
- Camera on auto brightness and exposure

- Sunny weather
- Region of interest (ROI) specified
- Brightness and exposure on set values



- Sunny weather
- Region of interest (ROI) not specified
- Brightness and exposure on set values
- Cluster protective plastic shield removed (not feasible in all vehicles)

Experimentation with polarising filters and different camera configurations, potential cabin adjustments to reduce glare.

5 RESULTS

- Trained model is currently able to detect the LKA icon in the 'two red lines' state as a Proof of Concept
- Labelling and training is ongoing for a model to identify the other 9 states
- The model can identify the LKA icon, but with reduced confidence, in images with high amounts of environmental factors (glare, vibration)
- The training dataset of rig images will be enriched with in-vehicle images with environmental factors present to improve performance

6 CONCLUSION

Research & Planning	Research undertaken around chosen ADAS feature and existing JLR projects YOLO v5 algorithm chosen for image recognition
Design & Development	Camera script for image capture developed Image processing pipeline developed
Integration	Recording function can be implemented as a part of MT Core, shrinking overall software footprint and integrating with existing MT data collection
Testing	Algorithm has successfully detected LKA icon on images obtained from both vehicle and rig testing

7 FUTURE WORK

1. Improve the accuracy of the model
2. Train the model on different ADAS icons
3. Enable integration with multiple vehicle lines

VELOCITY DRIVEN VISION: ASYNCHRONOUS SENSOR FUSION BIRDS EYE VIEW MODELS FOR AUTONOMOUS VEHICLES

Seamie Hayes, Ciarán Eising
University of Limerick

hayes.seamie@ul.ie, ciaran.eising@ul.ie

Abstract

Fusing different sensor modalities can be a difficult task, particularly if they are asynchronous. Difficulties arise due to incorrect spatial and temporal alignment. Our approach to resolving the issue of asynchrony of sensors yields promising results. For one instance of the asynchronous case, when utilising velocity information, we narrow the gap between camera+radar (C+R) to camera+LiDAR (C+L) from a difference of 5.1 IOU to 2.7. This is a major leap forward for the utilisation of the often-neglected radar sensor modality, which is less favoured than LiDAR for autonomous driving purposes.

Methodology

Construction of Synchronous and Asynchronous Datasets

To push the limits of asynchrony, we modify the nuScenes dataset [1] to construct two different datasets: *asynchronours* dataset, *synchronous* dataset. The synchronous dataset is constructed by removing the first data point, and then one more data point for every frame-lag. The asynchronous dataset is constructed very similarly except that every frame is given the radar data of the previous frame (this is for frame-lag 1). Additionally, we translate this data from its old ego frame to the new ego frame. We can now examine the performance of these datasets by utilising the Simple-BEV [2] architecture.

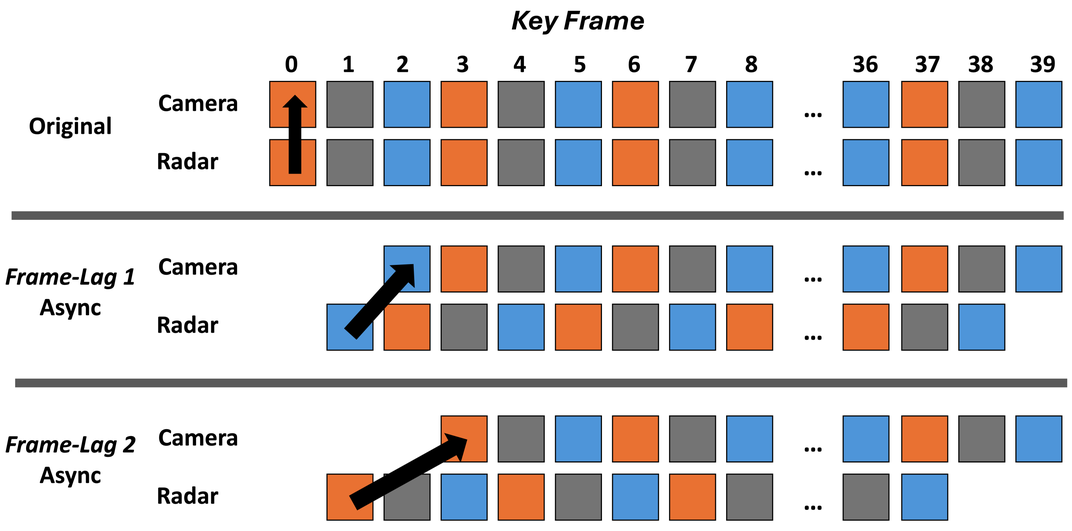


Figure 1: Dataset Visualisation: Original vs Asynchronous for frame-lags 1 and 2

Inferring Future Radar Point Clouds

We can enhance the radar data using velocity information, notably the asynchronous data as it is lagging behind. This is not possible for LiDAR, as their sensors cannot capture velocity information. We update the position of the current radar points, $P_{current}$, by inferring what their future position should be, P_{future}, with the xy velocity information, $P_{velocity}$. To finalise the calculation, we require the time passes which is simply the time CAM_FRONT was captured (t_{cam}) minus the time our radar sensor was captured t_{radar}. We use the following formula to predict future position:

$$P_{future} = P_{current} + P_{velocity} \times [t_{cam} - t_{radar}]$$

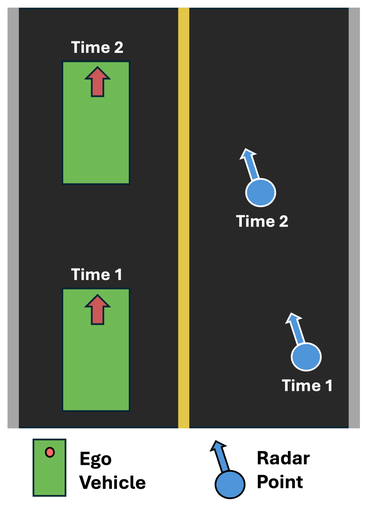


Figure 2: Inferring future radar point for *Time 2*

Results

Table 1: Comparison of ablation studies across C+R and C+L datasets: Asynchronous and Synchronous, with various frame-lags (1, 2, 3), and radar inference on future positions. Green highlights synchronous, red highlights asynchronous, with the best performer in each asynchronous metric **bolded**

Datasets				Metrics	
Sync/Async	*Modality*	*Frame-Lag*	*Infer Position*	*IOU*	*Total Loss*
Synchronous	C+R	1	✓	50.9	2.4
Synchronous	C+R	1	✗	50.3	2.5
Synchronous	C+L	1	N/A	58.9	2.0
Asynchronous	C+R	1	✓	48.4	2.6
Asynchronous	C+R	1	✗	47.0	2.6
Asynchronous	C+L	1	N/A	**53.1**	**2.3**
Synchronous	C+R	2	✓	49.0	2.9
Synchronous	C+R	2	✗	49.5	2.9
Synchronous	C+L	2	N/A	58.2	2.4
Asynchronous	C+R	2	✓	45.6	**2.3**
Asynchronous	C+R	2	✗	43.2	3.3
Asynchronous	C+L	2	N/A	**48.3**	2.9
Synchronous	C+R	3	✓	49.0	2.6
Synchronous	C+R	3	✗	49.3	2.7
Synchronous	C+L	3	N/A	59.0	1.9
Asynchronous	C+R	3	✓	43.8	2.9
Asynchronous	C+R	3	✗	43.1	2.9
Asynchronous	C+L	3	N/A	**48.7**	**2.6**

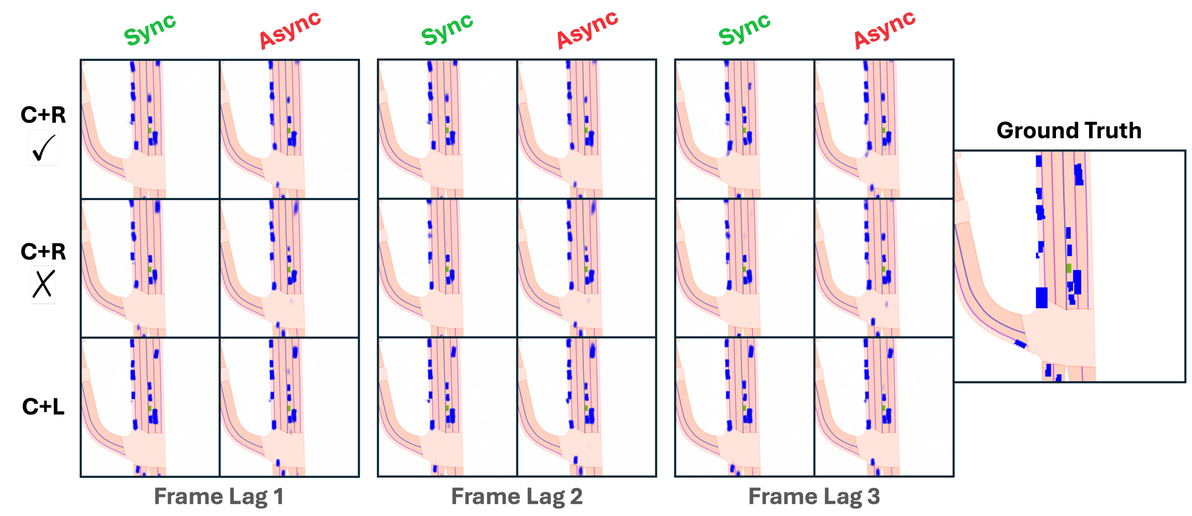


Figure 3: BEV image comparison for all experiments. C+R tick refers to inferring future position. C+R cross refers to not inferring future position

Analysing our results, it is very apparent from both the table and our predicted BEV images, that asynchrony introduces severe degradation, especially for C+L models where for frame-lag 3 we experience a 10.3 drop in IOU compared to the synchronous model. C+R receives a 5.2 drop in IOU for frame-lag 3.

Additionally, inferring future positions for C+R synchronous datasets yields inconsistent improvements. However, for the asynchronous datasets, we see a noteworthy boost, namely for frame-lag 2 where there is a 2.4 increase in IOU compared to the non-inferred dataset. This is an advantage of utilising radar for asynchronous datasets, as LiDAR lacks the ability to capture velocity information.

For both asynchronous and synchronous datasets, C+L outperforms C+R with respect to IOU, simply due to the sheer density of LiDAR point cloud data. However, for the asynchronous inferring frame-lag 2 datasets, C+R greatly outperforms C+L with respect to loss.

Conclusion

- Asynchrony **severely degrades** models. C+L datasets take a larger hit in performance than C+R datasets.
- C+R Synchronous datasets **cannot** be consistently improved using radar velocity information. However, C+R Asynchronous datasets **can** be consistently improved.
- Performance of the C+R Asynchronous datasets **approaches** that of the C+L Asynchronous datasets, namely when utilising **velocity** information.

References

[1] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. nuscenes: A multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027*, 2019.

[2] A. W. Harley, Z. Fang, J. Li, R. Ambrus, and K. Fragkiadaki. Simple-bev: What really matters for multi-sensor bev perception?, 2022.

Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6049. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

ARDÁN: Automated Reference-free Defocus characterization for Automotive Near-field Cameras

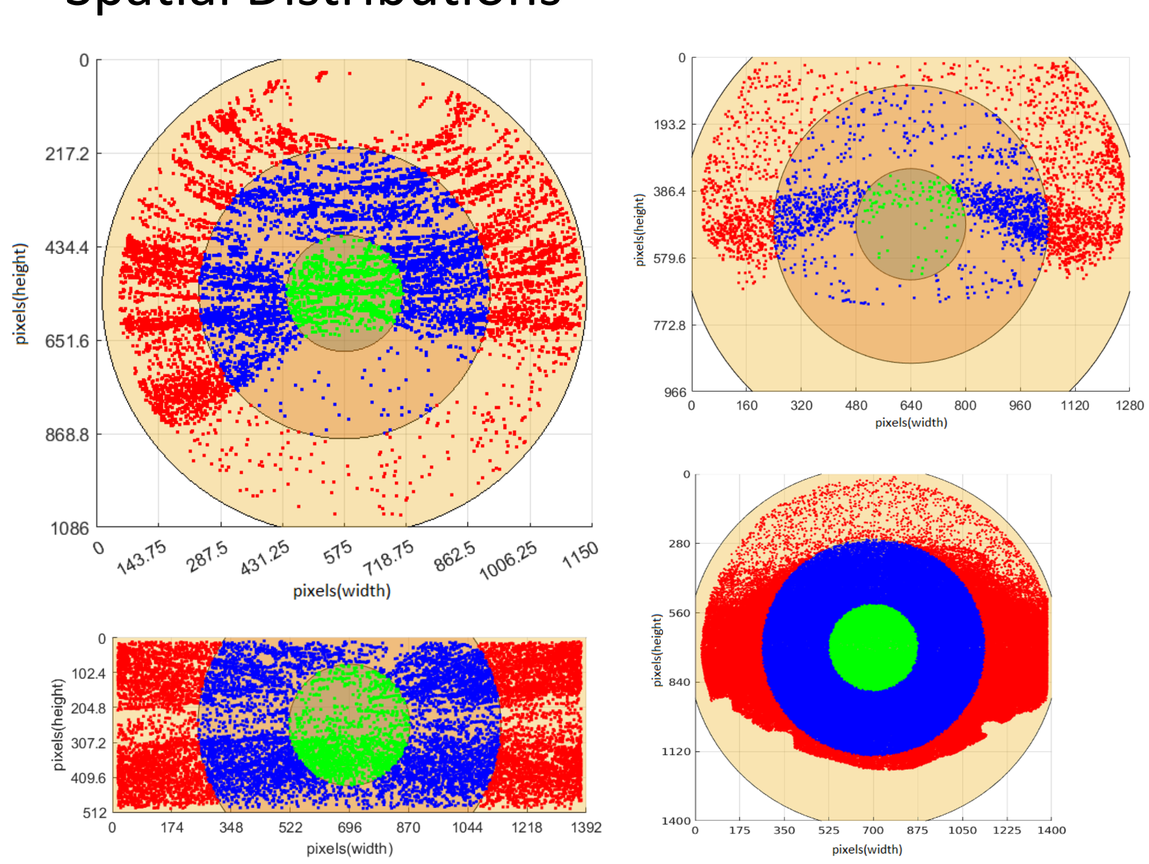
Daniel Jakab (University of Limerick)
Dr. Brian Deegan (University of Galway)
Dr. Anthony Scanlan (University of Limerick)
Dr. Ciarán Eising (University of Limerick)

1 Introduction

Measuring **optical quality in camera lenses** is a crucial step in evaluating cameras, especially for safety-critical visual perception tasks in automotive driving. While ground-truth labels and annotations are provided in **publicly available automotive datasets** for computer vision tasks, there is a lack of information on the **image quality of camera lenses** used for data collection. To compensate for this, we propose an **A**utomated **R**eference-free **D**efocus characterization for **A**utomotive **N**ear-field cameras ***(ARDÁN)*** to evaluate **Horizontal Slanted Edges for ISO12233:2023** in four publicly available automotive datasets using a Region of Interest (ROI) selection system in natural scenes. We use the mean of 50% of the Modulation Transfer Function **(MTF50) to measure optical quality** in three **Camera Radii (CaRa)** segments. We use **Regional Mask to Lens Alignment (RMLA)** to remove ego-vehicle occlusion and vignetting from cameras.
See below, for further information about the four datasets used in the experiments:

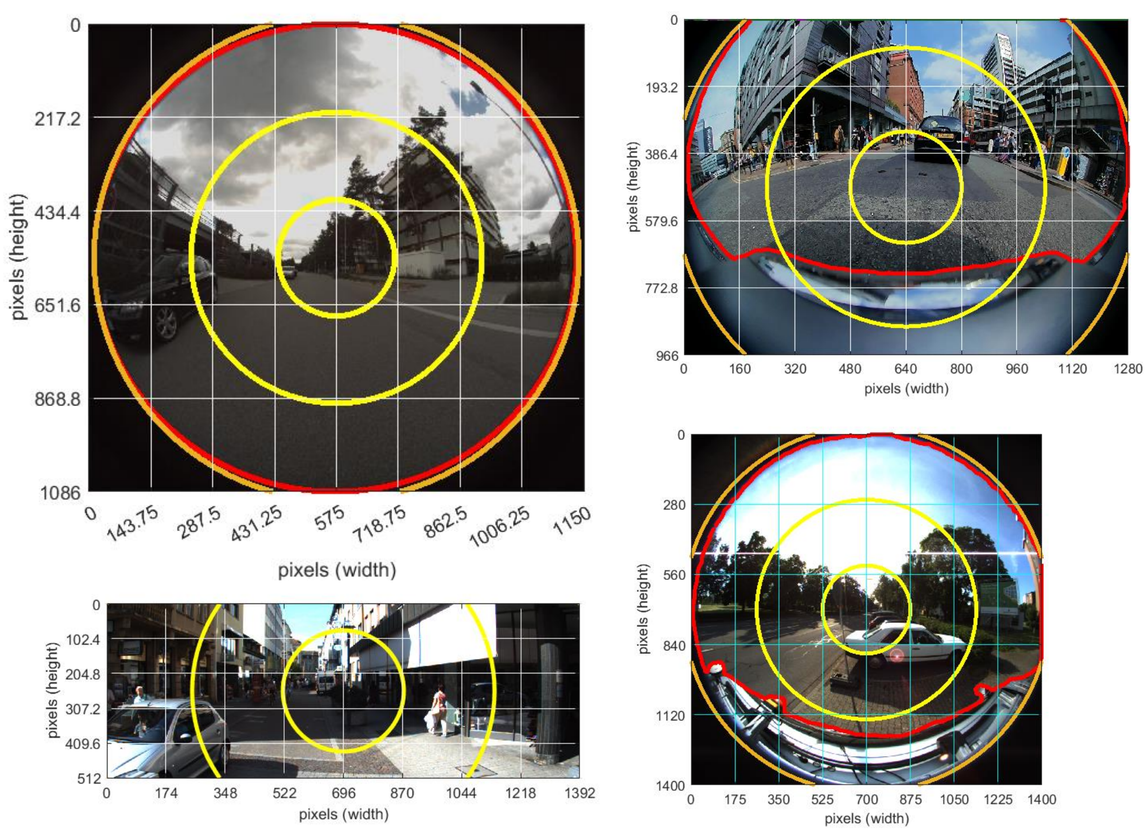
Datasets	No. Images
KITTI (4th Camera RGB)	1065
KITTI-360 (Left Camera)	11518
Woodscape (Front View)	1514
LMS (Rear Camera)	1251

2 Spatial Distributions



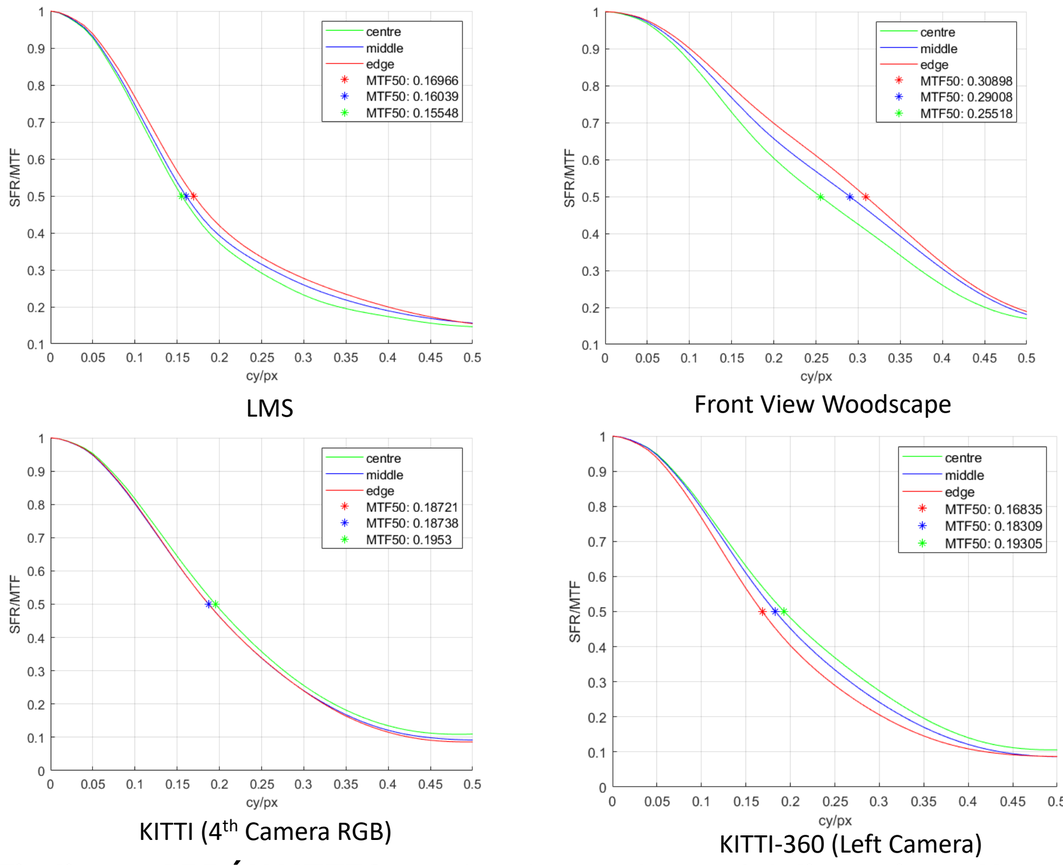
Spatial Distributions of MTF50 measurements for **KITTI, KITTI-360, LMS and Front View Woodscape** datasets. Notice, the clusters of data points for each dataset are unique where no two datasets give the same distribution of measurements. Most notably, the pattern of results follows the geometry of the scenes where the outline of the road in each dataset can be seen from the clustered points.

3 Camera Radii (CaRa)



To measure the optical quality of public datasets, we propose **Camera Radii (CaRA)** to segment the spatial domain of cameras. The outer CaRa circle(orange) represents the periphery of the camera aperture, the two inner CaRa circles (yellow) and finally, the red outlines are the **Regional Mask to Lens Alignment (RMLA)**.

4 Experimental Results



LMS — Front View Woodscape

KITTI (4th Camera RGB) — KITTI-360 (Left Camera)

Findings: *ARDÁN* MTF50 results show substantial **distortion effects** on the image quality of fisheye datasets. **Distortion** may have an **unforeseen impact on *Computer Vision*** performance.

Acknowledgements
This work was supported, in part, by the Science Foundation Ireland grant 13/RC/2094 P2 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero.

HOST INSTITUTION

PARTNER INSTITUTIONS

FUNDED BY:

Influence of AVC and HEVC Compression on Detection of Vehicles Through Faster R-CNN

Authors: Pak Hung Chan[1], Anthony Huggett[2], Georgina Souvalioti[1], Paul Jennings[1], Valentina Donzella[1] *Affiliation*: [1]WMG, University of Warwick, [2]onsemi

Abstract

IEEE Xplore (Published)

Due to the considerable data amount produced by the vehicle's perception sensors, there is the need to investigate techniques to reduce the datarate, e.g. for camera well established **lossy compression** techniques can be explored and evaluated. These techniques must be analysed in combination with the consumer of the data, which will most likely be a perception algorithm based on deep neural networks (DNNs). This work shows that **compression tuned DNNs** have enhanced performance with respect to traditionally trained DNNs, and the performance is higher when evaluating not only compressed data, but also uncompressed data. Overall, the DNN performance is steady when transmitting data with increasing lossy compression rate **(up to~130:1)**, but above this value there is a performance decrease. The results presented in this work demonstrate that **compression can be used in automotive sensors**, particularly leveraging the hereby proposed and optimised compression-tuned DNNs. The full paper is available at DOI: 10.1109/TITS.2023.3308344.

Pre-Print (Open-Access)

1 Motivation

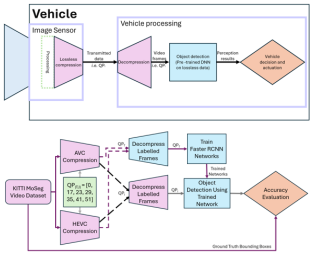
- Generated raw data by a perception suite is tremendous (can exceed 40 Gbit/s) and the majority this data is produced by camera (from 500 Mbit/s to several Gbit/s) [1].
- Automotive ethernet networks are providing higher bandwidth (from 100 Mbps to 10 Gbps) for camera video transmission than traditional wired networks, but will still be below what is required [2].
- Automotive systems employ lossless compression through expensive network cables per camera to ensure that the video information has not been altered.
- **Lossless** compression encodes the video to reduce datarate. When decoded, the video will be an exact match of the original video.
- **Lossy** compression encodes the video but selectively remove information which are less noticeable to the human visual system to reduce the datarate. The decoded video will not match the original video, but higher compression ratios can be achieved.
- For assisted and automated vehicles, the consumer of the camera data is machine learning based perception algorithms. This work investigates whether lossy compression can be used without impacting perception.

Key Contributions

- Proposal of a robust methodology to evaluate the effects of data compression on perception step.
- Demonstration that high level of lossy compression (up to ~130:1) can be applied without degrading the performance of Deep Neural Network (DNN) based vehicle detection.
- Established that re-training DNN with lossy compressed data is beneficial to the training performance. This is true even when different compression standard/rates are used for transmission.
- Proposal of a process to optimise the compression ratio of data used in training the perception algorithm when the compression ratio of the transmitted data is known.

2 Methodology

- **Dataset:** The KITTI MoSeg collated and provides sequential frames from the KITTI dataset [3].
- **Compression:** Compression schemes *Advanced Video Coding* (AVC) and *High Efficiency Video Coding* (HEVC) have been selected to compress the dataset sequences using the library *ffmpeg* with the *libx264* and *libx265* codecs respectively.
- **Rate Control:** Constant Quantisation Parameter (CQP) was selected as the rate control for compression. This ensures that the lossy nature of the data across the dataset is computationally consistent, as opposed to CRF which can vary the Quantisation Parameter (QP) based on the human perception system.
- **Quantisation Parameter:** Seven values of QP for each of the 2 compression schemes are used to compress the dataset, resulting in 14 compressed datasets, ranging from *losslessly* compressed dataset (0 QP) to *visually lossless* (23 QP) and the *highest compression* possible in these schemes (51 QP).



Trained Networks (QP_T):

14

AVC 7 | HEVC 7

Inference Datasets (QP_I):

14

AVC 7 | HEVC 7

Total Number of Inference:

196

(T)AVC-(I)AVC 49 | (T)HEVC-(I)HEVC 49 | (T)AVC-(I)HEVC 49 | (T)HEVC-(I)AVC 49

Homogeneous | Heterogeneous

Object Detection (Perception)

- Transfer learning with a DNN network was performed using each of the created compressed dataset, creating 14 trained networks.
- Each of the trained network is used to detect vehicles in every inference dataset and evaluated the performance using average precision.

3 Compression Ratios

- Compression ratio exponentially increases, for every increase in 6 QP the compression ratio doubles.
- Different compression schemes introduces different losses in the information. This translate to different visible artefacts at higher compression.

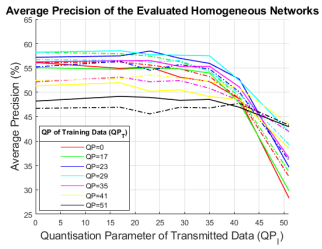
QP	Total Compressed Dataset File Size (Bytes)		Compression Ratio (based on the raw dataset size of 2,024,616,250 Bytes)	
	AVC	HEVC	AVC	HEVC
0	780,118,312	879,129,975	1:3	1:2
17	146,862,722	206,085,465	1:14	1:10
23	64,705,207	101,412,155	1:31	1:20
29	28,299,339	43,227,152	1:72	1:47
35	12,616,765	15,586,079	1:160	1:130
41	5,740,979	5,152,473	1:353	1:393
51	2,015,010	1,585,235	1:1005	1:1277

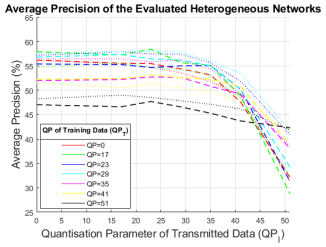
Blurring and loss of high frequency information

Loss of contrast causing vehicles to blend together

Distortion of features

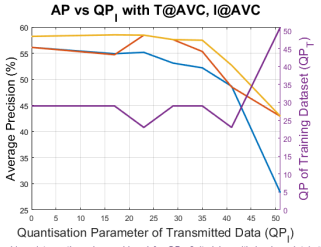
4 Results



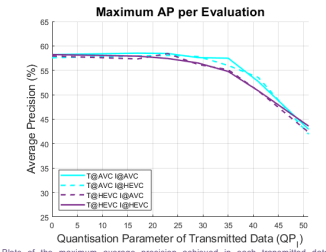


Solid line represents using AVC – AVC compression for training and transmitted datasets; dash-dotted line represents using HEVC – HEVC; dashed line represents HEVC – AVC; dotted line represents AVC – HEVC

- DNN-based object detection has a small variation in performance inferring from datasets compressed with up to a QP_I of 35, demonstrated by the relatively flat average precision performance for all the trained networks.
- From QP_I of 35, the trained networks all have a sharp drop in performance >25% in the worse case (except the network trained with QP_T of 51). Compression past this point is detrimental and cannot be deployed on assisted and automated vehicles.



The blue plot are the values achieved for QP_T=0 (training with lossless data), the orange are for QP_T=QP_I (training with the same level of compression of the transmitted data), yellow are for the maximum achievable AP as a function of QP_I, and the purple plot (right axis) show the related QP_T for each point in the yellow plot.



Plots of the maximum average precision achieved in each transmitted dataset produced. The line colour represent the QPT training compression scheme and the line style represent the QPI transmitted compression scheme.

- In all cases, training with lossless compressed data (i.e. the blue plot, QP_T = 0) did not perform the best.
- Compared to the baseline (blue) average precision performance improves by 2% to 15% when retraining with compressed data.
- Compressing with a QP of 29 for training is demonstrated to be optimal in most cases using AVC scheme.
- Performance gap between QP_I 29 and 41 when training with different compression schemes. The DNNs could learn particularities in artefacts generated by compression.

5 Conclusions

- **Aim**: Investigate the effects of compression on a downstream perception task (object detection) in assisted and automated vehicles *i.e.* Object Detection.
- **Focus:** Evaluating DNN based object detection algorithm using lossy compression for both transfer learning and inference.
- **Results:** Compression of sensor data on vehicles up to 160:1 before the performance degraded. Additionally, networks trained with compressed data improved performance (up to around 15%).
- **Performance:** Lossy compression on the training data of the DNN has demonstrated an improvement in detection performance when using lossless and lossy data for transmission. With careful choice of compression, it can be used as a pseudo hyperparameter.
- **Significance:** Lossy compression can significantly reduce the amount of data to transmit from camera sensors and reduce the complexity and amount of wiring harness and transmitting protocols to lower both cost and weight of the vehicle.
- **Generalisation:** The results are similar in AVC, HEVC and Motion-JPEG compression schemes [4]. However, training with different compression schemes affects the performance more than the compression of the inference dataset.
- **Future Research:** Automotive use have strict requirements such as latency and no corruption/artefacts of the data to ensure safety. *Ad hoc* compression schemes will need to be optimised to meet these requirements [5].

References:

[1] S. Heinrich, "Flash memory in the emerging age of autonomy," *in Proc. Flash Memory Summit*, 2017.
[2] L. Bello *et al.*, "A Perspective on Ethernet in Automotive Communications – Current Status and Future Trends," *Applied Sciences*, 2023.
[3] M. Siam *et al.* "MODNet: Motion and Appearance based Moving Object Detection Network for Autonomous Driving," in *2018 ITSC, 2018.*
[4] P. H. Chan *et al.* "The data conundrum: compression of automotive imaging data and deep neural network based perception," in *LIM*, 2021.
[5] Y. Wang *et al.* "Semantic-Aware Video Compression for Automotive Cameras," IEEE Transactions on Intelligent Vehicles, 2023.

Robust downsampling for LiDAR Point Clouds in Assisted and Automated Driving

Authors: Jiangyin Sun, Boda Li, Pak Hung Chan, Kurt Debattista, Valentina Donzella

Affiliation: [1]WMG, University of Warwick

Abstract

LiDAR, used to support Assisted and Automated Driving functions produces a significant amount of data that needs to be transmitted to the vehicle processing units. This work proposes a novel processing algorithm for LiDAR pointclouds that enables a significant reduction of the size of data to be transmitted without losing essential information and ensuring robustness against some of the potential noise factors that can degrade the point cloud. This proposed algorithm combines removal of ground points with Region of Interest/Non-ROI Separation and downsampling (deploying and comparing uniform and non uniform techniques). Noisy and 'ideal' downsampled pointclouds are evaluated using machine learning based object detection, i.e. the Complex YOLO v4 deep neural network. The experimental results show that the detection performance are stable, the average precision and average similarity are calculated and remained robust compared to the original baseline around 90% and a trend has been found from multiple combinations of parameters. For the rain noise model adding on, the detection results are robust to approximately same average precision untill 1mm/h of rain rate, and decreasing follows the increasing of rain rates till 8mm/h.

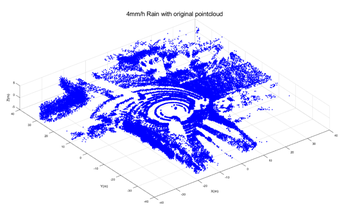
1 Motivation

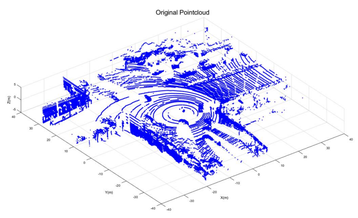
- Challenges with large data amount produced by perception sensors
- Traditional Networking/Real-time processing. [1]
- Compression challenges (data loss/artifacts/robustness). [2]
- Preservation of vital information. [3]

- Challenges with noise factors and sensor data degradation
- Prediction and elimination of influence cased by noise factors.
- Prediction of influence under extreme weather conditions. [5] [6] [7]

2 Methodology

Panda Dataset selected [8]

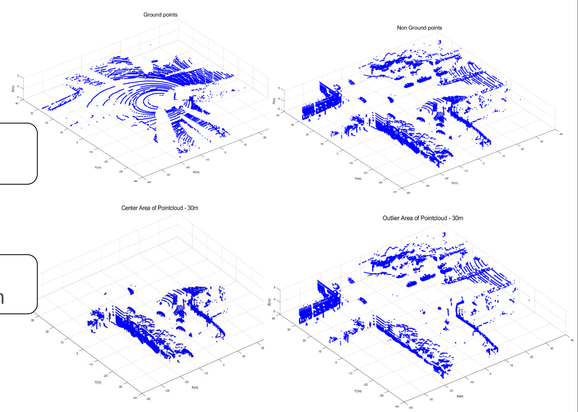




Original Point Cloud

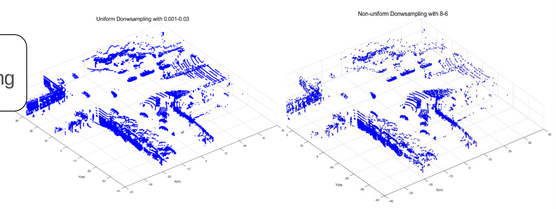
Rain Noise model [7] [9]

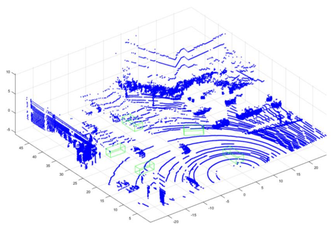
Pre-processing



Region of Interest/
Non-ROI Separation

Uniform/
Non-Uniform Downsampling





DNN - Object Detection

The DNN object detection network was trained from a dataset with 6 different rain rates combined from 0 mm/h to 8 mm/h.

Evaluation & Comparison

3 Downsampling Comparison

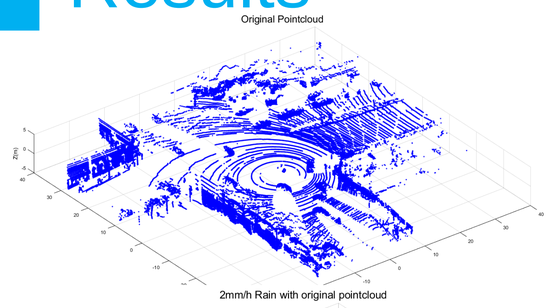
1. Random Downsampling
With the parameter of percentage %

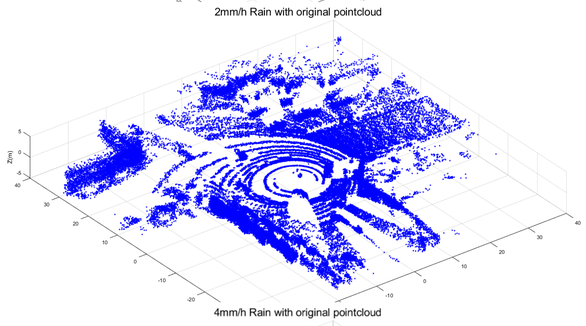
2. Uniform Downsampling
With the parameter of grid-size to determine how large grids are and take the average value of all points in a single grid

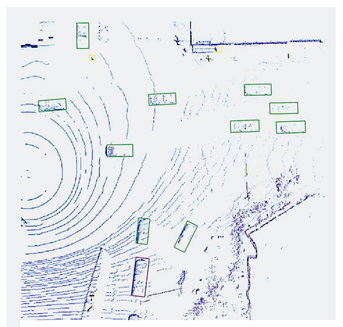
3. Non-uniform Downsampling
With the parameter of Maximum points can be included in each slot and randomly select one of them.

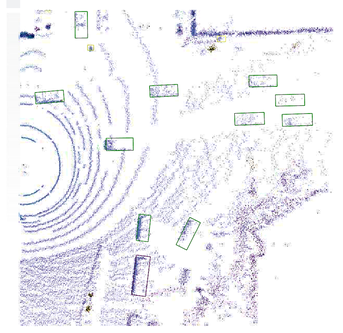
All 3 methods with results to around 10% points remained

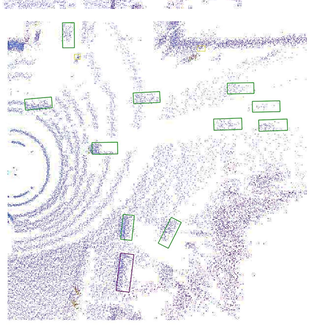
4 Results



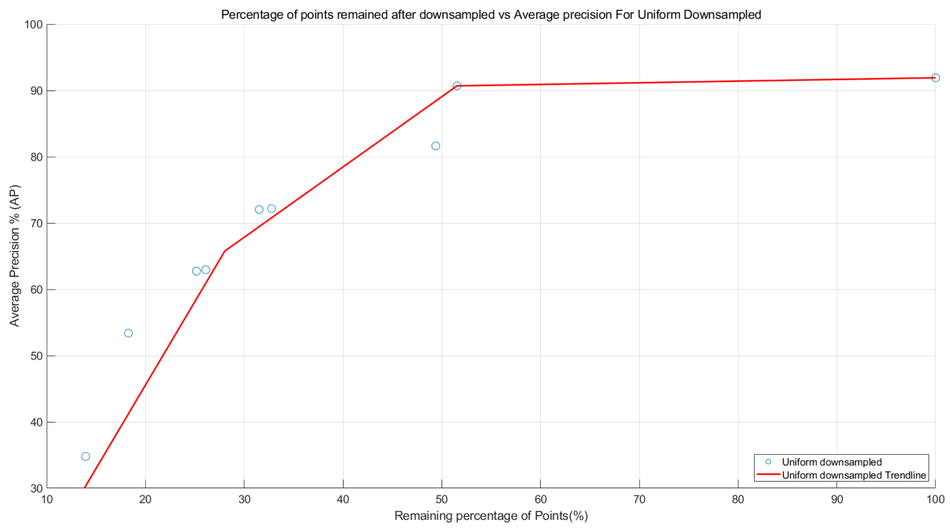








The results showing the difference for the same frame of point cloud under 2mm/h and 4mm/h rain rates in Bird-eye view



- The detection results are robust with 50% of reduction of the pointcloud size.
- A trend has been identified with AP decreasing following the pointcloud size reduction.
- Robust to approximately same average precision untill 1mm/h of rain rate, and decreasing follows the increasing of rain rates till 8mm/h.
- The performance of detection results can be improved more by using machine-learning base downsampling methods for the next step

5 Conclusions

- **Research gap** : The combined effect on LiDAR performance with noise models and data reduction techniques.
- **Innovation of this project**: Proposing a noise robust method which is able to significantly reduce the point cloud size (e.g. 50% size reduction) and to preserve key information
- **Future Research:** Experiment and contrast different noise models, adapt various object detection network as well as downsampling methods.

[1] B. Anand, V. Barsaiyan, M. Senapati, and P. Rajalakshmi, 'Real time LiDAR point cloud compression and transmission for intelligent transportation system', IEEE Veh. Technol. Conf., vol. 2019-April, pp. 1–5, 2019, doi: 10.1109/VTCSpring.2019.8746417.
[2] C. Tu, E. Takeuchi, A. Carballo, and K. Takeda, 'Point cloud compression for 3d lidar sensor using recurrent neural network with residual blocks', Proc. - IEEE Int. Conf. Robot. Autom., vol. 2019-May, no. May, pp. 3274–3280, 2019, doi: 10.1109/ICRA.2019.8794264.
[3] H. M. Sayed, S. A. Taie, and R. A. El-Khoribi, 'An improved technique for LIDAR data reduction', Int. Conf. Electron. Devices, Syst. Appl., pp. 3–6, 2017, doi: 10.1109/ICEDSA.2016.7818479.
[4] M. Sualeh and G. W. Kim, 'Dynamic Multi-LiDAR based multiple object detection and tracking', Sensors (Switzerland), vol. 19, no. 6, 2019, doi: 10.3390/s19061474.
[5] T. Fersch, A. Buhmann, A. Koelpin, and R. Weigel, 'The influence of rain on small aperture LIDAR sensors', GeMiC 2016 - 2016 Ger. Microw. Conf., pp. 84–87, 2016, doi: 10.1109/GEMIC.2016.7461562.
[6] P. D. Kumar and T. Ramashri, Noise Reduction in Lidar Signal Based on Sparse Difference Method, vol. 1040. Springer Singapore, 2020
[7] P. H. Chan, G. Dhadyalla, and V. Donzella, 'A Framework to Analyze Noise Factors of Automotive Perception Sensors', IEEE Sensors Lett., vol. 4, no. 6, pp. 4–7, 2020, doi: 10.1109/LSENS.2020.2996428.
[8] Scale AI, Inc. and Hesai Photonics Technology Co. (2020). PandaSet by Hesai and Scale AI High-quality open-source dataset for autonomous driving, https://pandaset.org/
[9] Goodin, C., Carruth, D., Doude, M., & Hudson, C. (2019). Predicting the Influence of Rain on LIDAR in ADAS. Electronics, 8(1), 89.

Synthetic Bayer Dataset with Headlight Flare

Authors: Boda Li[1], Darryl Perks[2], Yiting Wang[1], Pak Hung Chan[1], Valentina Donzella[1]

Affiliation: [1]WMG, University of Warwick, [2]onsemi

1 Abstract

Automotive cameras are widely deployed on cars, however the data generated by cameras can be affected by various noise factors. Flare, also known as straylight, is a common noise factor especially during the night. The automotive headlight can be dazzling for human drivers and might alter camera data use for Assisted and Automated Driving (AAD) functions. To enable higher levels of driving automation, investigating and testing this noise factor can be key to achieve AAD in challenging lighting conditions. Therefore, accurate automotive camera flare models need to be thoroughly investigated and developed. This work develops, describes, and validates an automotive specific parameterised method for modelling flare induced by automotive headlights. By this method, the model can be validated using real camera data. Additionally, this work introduces a method to seamlessly integrate the modelling results into images generated by a widely used simulation software, the CARLA simulator. Using the newly proposed model, paired automotive datasets with and without the realistic headlight flare can be generated. Overall, this work can pave the way for more accurate automotive camera modelling, which is key to speeding up the simulation and testing of assisted and automated driving functions.

2 Motivations

- Flare is a camera artefact occurring when light arrives at the image sensor through an unintentional path; this effect can be due to various reasons, e.g. surface scattering by sensor components, aperture diffraction, lens coating reflection, etc. These unpredictable events make flare inevitable in camera systems, causing image contrast reduction and undesirable imaging [1][2].
- Scenarios with flare from oncoming traffic shall be considered yet to be included into (a) virtual/physical AAD testing process, (b) open-source driving datasets representing realistic camera data.

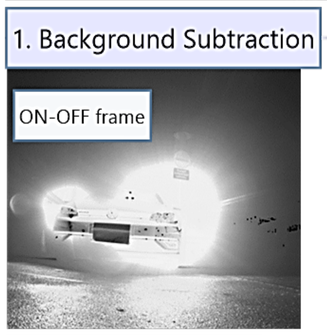
3 Objectives

- Proposing a method to model and validate headlight flare by using real automotive camera data.
- Development of an **empirical model** of headlight flare, adaptable to any real automotive camera.
- A workflow to integrate the realistic flare mask into simulated data.

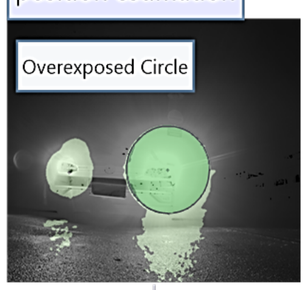
4 Methodology

- Processing flow from collected data to flare model parameter extraction with examples of processed images:

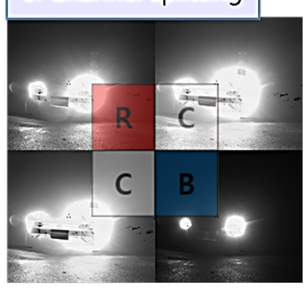
1. Background Subtraction



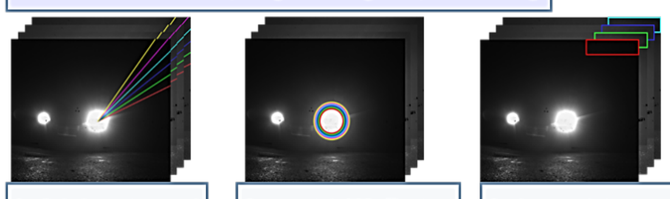
2. Headlight centre position estimation



3. Channel splitting



4. GSF feature sampling and signal processing

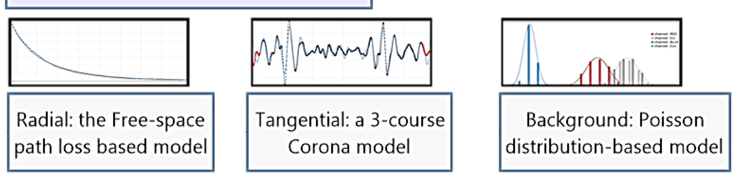


Radial: Cropping, and normalisation

Tangential: FFT, filtering, and cropping

Background: Area sensitivity analysis and outlier checking

5. Physical model identification



Radial: the Free-space path loss based model

Tangential: a 3-course Corona model

Background: Poisson distribution-based model

6. Output modelled flare image mask

- Workflow to integrate the modelled headlight flare into synthetic images generated in CARLA simulator:

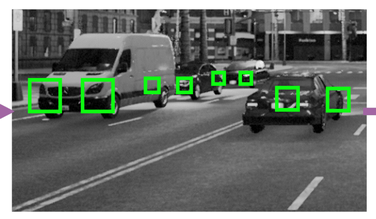
Traffic locations in the sim world

Headlight locations estimation with the help of traffic 3D bounding boxes

Headlight rough locations in the sim world

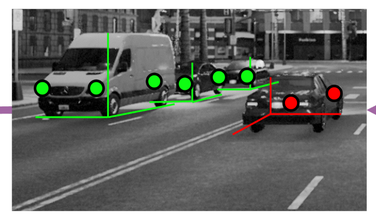
$$k_{2\times3} \times w2c_{3\times3} \times \begin{pmatrix} x_{loc} \\ y_{loc} \\ z_{loc} \end{pmatrix} = \begin{pmatrix} col_{pos} \\ row_{pos} \end{pmatrix}$$

Positions of headlight proposal areas in the synth image

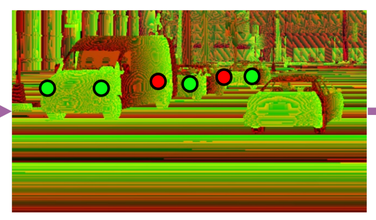


Pick the maximum possible location of headlight in each proposal area by checking pixel values

Proposal screening by traffic and ego vehicle's orientations



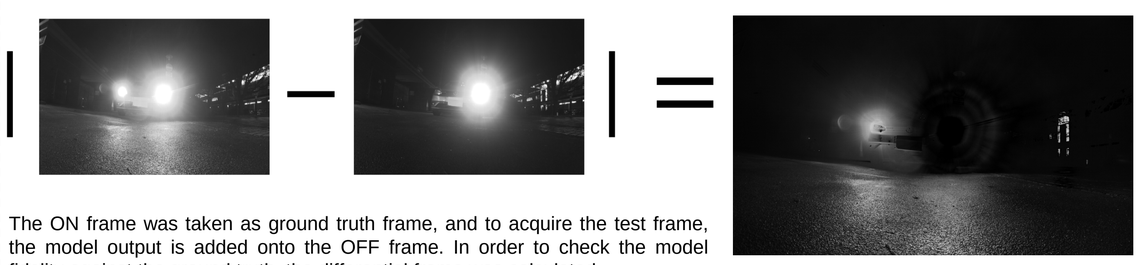
Proposal screening by checking against value from depth images generated by depth camera model from CALRA



Headlight flare patch integration parameterised by the light source distance and the strength based on the pixel values in the proposal area

Save the image pair with and without the headlight flare in raw Bayer format together with ground truth

- Headlight flare model validation by real sensor data:

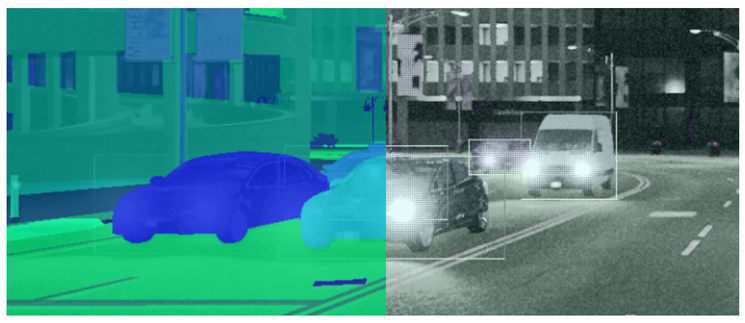


The ON frame was taken as ground truth frame, and to acquire the test frame, the model output is added onto the OFF frame. In order to check the model fidelity against the ground truth, the differential frame was calculated.

5 Results

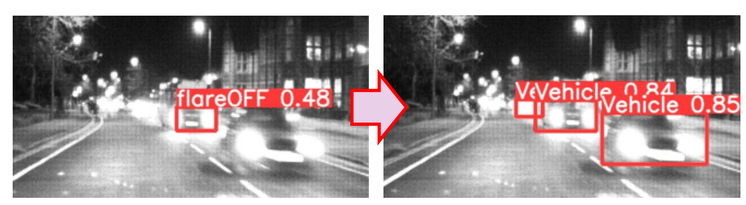
- Frames in pair output by the proposed methodology:
- They are 1920x1080 in single channel Bayer format. The CFA layout can be generated to mimic the specific CFA pattern of any selected camera.

- Ground truth automation in the form of 2d bounding boxes (shown on right) and full frame instance segmentation (shown on the left):



6 Conclusions

- **Aim:** Investigate headlight flare in automotive camera data by modelling and validating with real automotive camera data.
- **Focus:** A framework to model flare and a workflow to generate it onto synthetic images by CARLA simulator were proposed.
- **Results**: 3842 frames of synthetic Bayer data sets in pairs with and without the modelled headlight flare are generated.
- **Significance:** A validated model to enable faster AAD testing using sensor data emulation and augmentation.
- **Generalisation:** The proposed model can be used to enhance any real or synthetic data. The generated dataset can be used for bench-marking flare removal algorithms or other AAD functions or features.
- **Future Research:** Investigation of flare model in combination with downstream perception tasks. e.g. vehicle detector performance can be potentially improved by retraining with the proposed dataset:



Acknowledgments: Boda Li was supported by the State Scholarship Fund by China Scholarship Council (CSC) (award no. 202108060176). This work is partially funded by the European Union (grant no. 101069576). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. The UK participant (WMG) in this project is supported by Innovate UK (contract no. 10045139). Prof. Donzella acknowledges the Industrial Fellowships scheme of the Royal Academy of Engineering. The Authors thank HVM CATAPULT for the support, and Darryl Perks and Dr. Anthony Huggett from onsemi, for fruitful discussions.

References: [1] IEEE P2020 Working Group, "Image quality for computer vision," *IEEE P2020 Automotive Imaging White Paper*, pp. 14–22, 2018. [2] B. Li, P. H. Chan, G. Baris, M. D. Higgins, and V. Donzella, "Analysis of automotive camera sensor noise factors and impact on object detection," *IEEE Sensors Journal*, vol. 22, no. 22, pp. 22 210–22 219, 2022.

A New Approach for Bayer Adaption Techniques in Compression.

Authors: Hetian Wang[1], Anthony Huggett[2], Valentina Donzella[1]

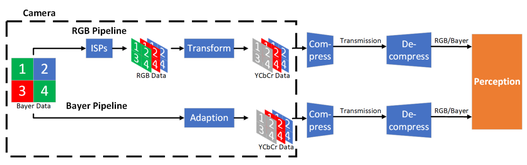
Affiliation: [1]WMG, University of Warwick, [2]onsemi

Abstract

Existing wired communication vehicle network technology lacks the bandwidth required to support the data rates produced by the sensor suite for assisted and automated driving functions in the next generation of vehicles. Video from image sensors demands a very high bandwidth and cameras are continually being developed giving better resolution, bit-depth (dynamic range) and frame rate. To overcome this challenge, compression is a potential solution to reduce the required bandwidth. As modern automotive cameras produce Bayer images, it might be more effective to compress Bayer images directly instead of compression for red-green-blue (RGB) images, which has been traditionally implemented in the most widely deployed compression schemes. By using Bayer, we avoid to triple the memory storage and preserve data fidelity by bypassing demosaicing and other non-reversible process in the Image Signal Processing (ISP) pipeline. Bayer adaption techniques indicate methods to convert Bayer data to be used in traditional compression pipelines. In this research, we propose two novel Colour Space Transform (CST) techniques and implement them in combination with the widely used H.264 codec. Our results show that CST techniques are superior to other techniques when paired with object detection, specifically when the bit rate is reduced between approximately 700 to 1250 kb/fr.

1 Motivation

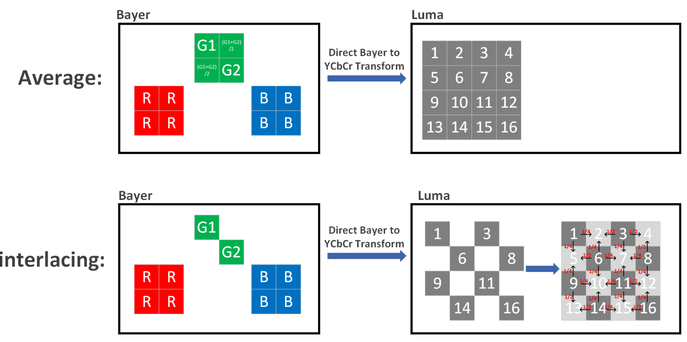
- Existing wired communication vehicle network technology lacks the bandwidth required to support the data rates produced by perception cameras in the next generation of vehicles
- We investigate Bayer compression in automotive for two reasons: firstly, compression reduces the amount of data used to represent the original image/video to meet a specific bit rate requirement [2]. It is promising in automotive but should also address concerns around safety associated with data loss, information distortion, artefacts, error propagation, latency, etc [1,3,4]. Secondly, as RGB frames are mainly for human consumption, there is an interest in using Bayer frames directly in automotive perception tasks [1]. Therefore, the possibility of adapting and using Bayes frames in current compression pipelines is an important aspect that needs to be investigated.



An example of a processing Bayer pipeline VS. a RGB pipeline.

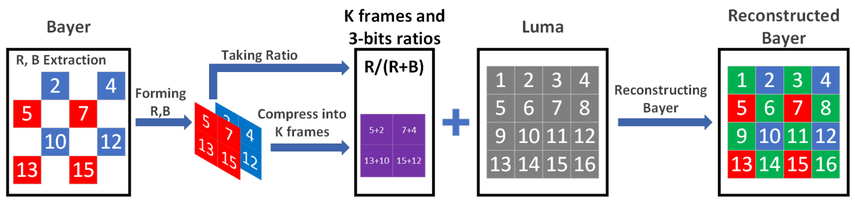
2 Colour Space Transform Techniques

- The principle of Bayer adaption techniques is essentially trying to break the chessboard pattern possessed by Bayer data. Based on the principles of Bayer adaption techniques, we hereby propose and apply CST techniques to use Bayer data and convert them into YCbCr format. We also design ***K frames*** and ***3-bit ratios*** mechanisms to accompany CST techniques for Bayer image reconstruction.



The newly proposed direct Bayer to YCbCr Transforms: From top to bottom, Average technique, and De-interlacing technique.

- **Direct Bayer to YCbCr Transforms:**
- We use direct Bayer to YCbCr transform to eliminate the intermediate RGB colour step which would triple the data size. Also, in the CST, only Luma data is preserved in YCbCr colour space.
- **Average:** In every 2x2 Bayer pattern, we have two true green pixels, one true red pixel and one true blue pixel. We duplicate true blue and red values assuming they are the same in all four pixels. And while we keep green values of two true green pixels, we assume other two pixels both have green values that is the average of two true green pixels. In this way, we have an estimated RGB data in every pixel and acquire Luma data of the image.
- **De-interlacing:** De-interlacing method is identical to the Average method in assuming red and blue values are the same in every 2x2 Bayer pattern but only calculates Luma data in the true green pixels. Later, it will fill the void pixels in Luma colour space by taking portions from their neighbours.



Schematic representation of how Bayer Data are reconstructed using K frames and 3-bit ratios.

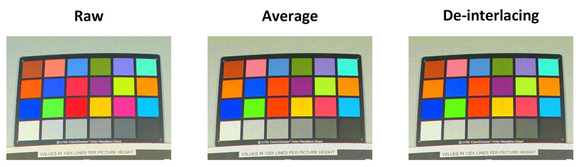
- **Reconstructing Bayer Data:**
- **K Frames:** Since only Luma channel is calculated in the proposed CST techniques, it brings an imminent problem which is there is no way to reconstruct the Bayer pattern from the Luma channel. In `traditional' YCbCr conversion, these components can be determined by Cb and Cr channels. Here, we develop a lightweight solution, based on using K frames.
- We determine K frames as compressed Red and Blue frames: $K = 0.257 \times R + 0.098 \times B$
- Hence, Luma channel can be determined by: $Y = 0.504 \times G + K + 16$
- **3-bit Ratios:** Here, we use 3-bit ratios to approximate red and blue data. Namely, in every 2x2 Bayer pattern, we determine the ratio of red pixel value out of the sum value of both red and blue and encode this ratio according to the 3-bit coding scheme. Given the K frames, the ratio information helps approximate the actual red and blue values.

3-bit ratios								
Code	000	001	010	011	100	101	110	111
Symbol	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7

3-bit Ratios: Codes and Symbols

3 Back-to-back Analysis

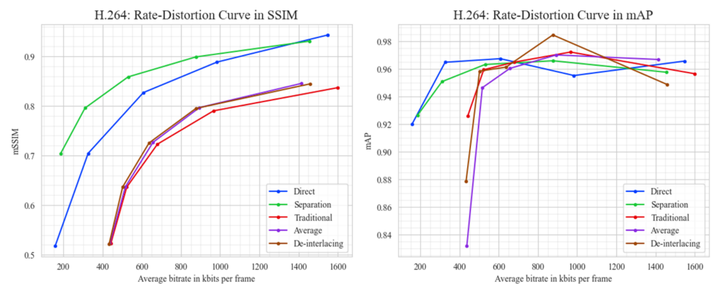
- The main purpose of our back-to-back analysis is to analyse the pixel-level effects of the proposed transformations. In our back-to-back analysis, we used an 8-bit Bayer testing image captured in an imaging lab. It contains a colour test chart for colour error analysis. The testing image is processed by the 2 CST techniques, one at a time. Then, we demosaiced both our Raw (original) testing image and CST processed images for comparison. By observing the colour checker, some colour blocks show limited colour shifting but in general error is limited.
- With our 8-bit testing image, both CST techniques show that they have deviations to Raw of approximately 0.98 in Mean Saturation and 0.7432 in Delta-E00, which are limited errors in the colour information.



Resultant Images for Back-to-back Analysis: From left to right are images processed by, Raw (Original), Average, and De-interlacing techniques.

4 Results

- We implemented our CST techniques in combination with the H.264 compression codec (Quantisation parameter (QP) set to QP={15, 23, 29, 35, 41}) with 300 frames from RobotCar Dataset (8 bits, cropped to 1280x320) and compared the compression performance between CST techniques and other techniques [6].
- For comparison, we included the Direct technique which directly compresses Bayer data with H.264 codec as a benchmark. We also included the Traditional technique which uses traditional demosaicing on Bayer data to acquire Luma channel. We finally included the Separation technique as a comparison with other approaches in designing Bayer adaption techniques, details refer to [5].We evaluated image quality of compressed images via a traditional image quality assessment (IQA) metric, namely SSIM, for pixel-level accuracy quantification, and via a Faster-RCNN based object detector for an assessment oriented towards applications in automotive perception [7].



Evaluation Results of H.264 Experiment: From left to right, in SSIM, and mAP for Object Detection.

- With direct Bayer to YCbCr methods in place, our proposed Average and De-interlacing techniques have improved compression performance than the baseline Traditional technique. However, when comparing our CST techniques with Direct technique and Separation technique, it is clear that they perform slightly better in terms of SSIM. In object detection evaluation, although Direct and Separation techniques still have better performance at lower bit rates, when bit rates are between approximately 700 to 1250 kb/fr, CST demonstrates better performance, with higher detection accuracy than other techniques.

5 Conclusions

- **Aim**: Develop novel Bayer adaption techniques to transmit the oversized images produced by automotive camera in next-generation automotive
- **Focus:** Evaluate performances of proposed CST techniques for automotive applications.
- **Limitation:** CST techniques have limited colour errors, Within acceptable range, CST techniques are able to represent most of the colour information with less required storage, even without compression.
- **Results:** In general, although CST techniques have performance below other conversion techniques in terms of IQA and specifically SSIM, they can achieve higher performance when combined with perception.
- **Significance:** The improvement of using CST techniques compressed images in object detection is specifically related to object detection performance at specific bit rates, which can pave the way for their use in automotive.
- **Future Research:** Our techniques enrich research in Bayer adaption techniques and in a broader view, in designing novel Bayer compression.

Acknowledgments: The work was partially supported by onsemi and High Value Manufacturing CATAPULT. This research is partially sponsored by the Centre for Doctoral Training to Advance the Deployment of Future Mobility Technologies (CDT FMT) at the University of Warwick. Co-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 Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. Project grant no. 101069576.

UK participants in this project are co-funded by Innovate UK under contract no. 10045139.Swiss participants in this project are co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract no. 22.00123.

References:

[1] P. H. Chan, A. Huggett, G. Souvalioti, P. Jennings and V. Donzella, "Influence of AVC and HEVC Compression on Detection of Vehicles Through Faster R-CNN," in IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 1, pp. 203-213, Jan. 2024, doi: 10.1109/TITS.2023.3308344.

[2] D. R. Bull and F. Zhang, Intelligent Image and Video Compression: Communicating Pictures. 2021. doi: 10.1016/B978-0-12-820353-8.00002-5.

[3] H. Hsiang, K. -C. Chen, P. -Y. Li and Y. -Y. Chen, "Analysis of the Effect of Automotive Ethernet Camera Image Quality on Object Detection Models," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 2020, pp. 021-026, doi: 10.1109/ICAIIC48513.2020.9065232.

[4] P. Krömer, M. Prauzek, M. Stankuš and J. Konečný, "Adaptive Fuzzy Video Compression Control for Advanced Driver Assistance Systems," 2018 26th International Conference on Systems Engineering (ICSEng), Sydney, NSW, Australia, 2018, pp. 1-9, doi: 10.1109/ICSENG.2018.8638184.

[5] Ning Zhang and Xiaolin Wu, "Lossless compression of color mosaic images," in IEEE Transactions on Image Processing, vol. 15, no. 6, pp. 1379-1388, June 2006, doi: 10.1109/TIP.2005.871116.

[6] Maddern W, Pascoe G, Linegar C, Newman P. 1 year, 1000 km: The Oxford RobotCar dataset. The International Journal of Robotics Research. 2017;36(1):3-15. doi:10.1177/0278364916679498.

[7] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

Investigating the Mutual Interference of State-of-the-Art Automotive LiDAR Sensors

Authors: Milan Lovric[1], Jess Smith[2], Gurdit Bhakar[1], Joel Rapley[2], Hsun Yang[2], Imran Mohamed[2], Valentina Donzella[1]

Affiliation: [1]WMG, University of Warwick, [2]The National Physical Laboratory (NPL)

Abstract

The increasing deployment of LiDAR-equipped vehicles along with their expected interactions in close proximity raise concerns about potential LiDAR-LiDAR interference. However, research into the mutual interference of state-of-the-art automotive LiDARs is still scarce. Earlier WMG/NPL experiments focused on measuring LiDAR mutual interference in an enclosed laboratory setting. In this study, the aim was to replicate those laboratory experiments in an outdoor environment – to reduce potential LiDAR beam multipath – as well as to conduct a preliminary analysis of potential interference mitigation by changing LiDAR (victim/offender) inclination angles. The results suggest that possible interference effects can be detected in the victim's point cloud as changes (mostly increases) in the percentage of outliers and the values of reflectivity, when the offender LiDAR is turned on. There is currently insufficient evidence to support changing LiDAR inclination angles as an effective interference mitigation strategy. Further research is also needed to disentangle interference effects from other noise factors.

1 Background

- As LiDAR-equipped vehicles are being increasingly used on the road, the likelihood of their LiDAR beams interacting in close proximity also increases (Fig.1), which raises concerns about potential LiDAR-LiDAR interference and the implications it might have on road safety.
- Two distinct types of LiDAR mutual interference have been identified in the literature as: (I) *direct interference* – when one LiDAR's beams are directly received by another LiDAR's sensor; and (II) *indirect interference* – when LiDAR beams are first reflected off other objects [1]. Receiving foreign laser pulses can give rise to issues such as ghost targets and diminished signal-to-noise ratio [2].
- Multiple approaches to attempt *interference mitigation* have been proposed using pre-processing, post-processing and hardware-based techniques [3].

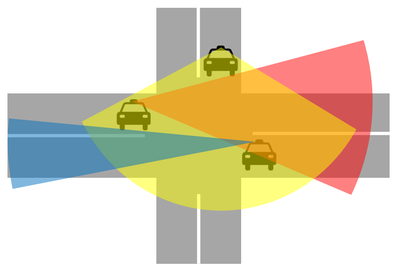


Fig.1: An illustration of possible close encounters of LiDAR-equipped vehicles at a road junction. LiDAR devices may have different FoV and different ranges.

- Earlier WMG/NPL experiments [4] investigated LiDAR mutual interference in an enclosed laboratory setting (Fig.2) at various distances (Fig.3). Possible interference effects were detected as an increased percentage of outliers in the victim LiDAR's point cloud, most prominently at (closer) distances of 1 m and 2 m.

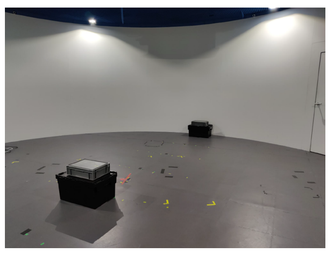


Fig.2: Image of the experimental chamber where measurements were taken in previous study [4].

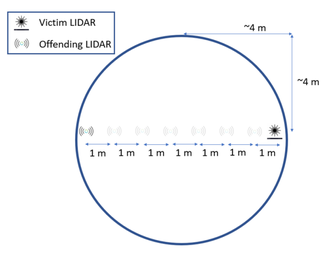


Fig.3: Schematic top-view diagram of LiDAR positions within the walls of the experimental chamber used in [4].

2 Objectives

- Previous study on LiDAR mutual interference [4] was potentially affected by LiDAR beam multipath due to the wall confines of the experimental chamber. Hence, the objective of this study was to replicate those experiments in an outdoor environment.
- Another objective was to conduct a preliminary investigation into the possibility of interference mitigation by changing LiDAR (victim/offender) inclination angles.
- The last objective was to understand the limitations of the current methodology to inform future experiments.

3 Methodology

- LiDAR positions:
  - Offender and victim LiDARs were oriented towards each other with their sensors aligned.
  - The two LiDARs were placed at the following distances: 1 m, 2 m, 3 m, 5 m, 7 m, 15 m
  - LiDAR inclination angles were used at 1 m - 3 m:
    - Victim LiDAR angles: 0°, -8°, +8°
    - Offender LiDAR angles: 0°, -8°
  - Offender LiDAR was turned OFF (for baseline dataset) and ON (for interference dataset).

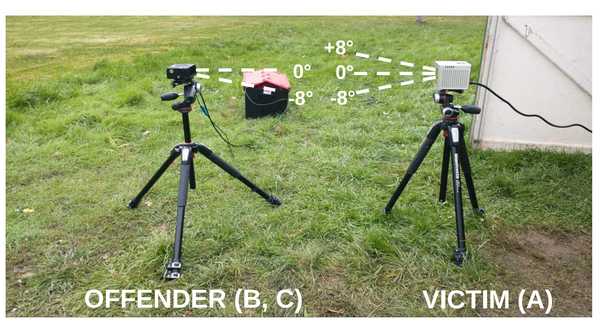


Fig.4: The position of the offender/victim LiDAR at a 1 m distance and 0° angles.

- Data collection:
  - Each point cloud was recorded for 80 seconds (at a frame rate of 10 fps).
  - All measurements were taken twice, with LiDAR angles repositioned before each measurement.
  - In total, 240 point clouds (62 GB) were recorded.
  - Notes were taken for special events that may have impacted measurements (e.g., rain droplets on the sensor, sun hitting the sensor, the lawn being mowed, birds in the background etc.)

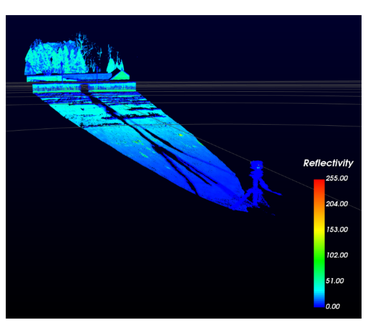


Fig.5: Visualisation of the victim LiDAR's point cloud.

4 Results

- Metrics for the point cloud analysis of the victim A (MEMS LiDAR):
  - **Outliers:**
    - Near outliers: range < 7.44 m (for consistency with previous study [4])
    - Far outliers: range > 7.64 m (for consistency with previous study [4])
    - Very Far Outliers: range > 73 m (typical stopping distance from 60mph, the UK Highway Code)
  - **Reflectivity (0 to 255):**
    - Values 0 to 150 (0% to 100% in the Lambertian reflection model)
    - Values 151 to 255 (reflectivity of target objects with retroflection properties)
- Results for the offender B (mechanical spinning LiDAR):

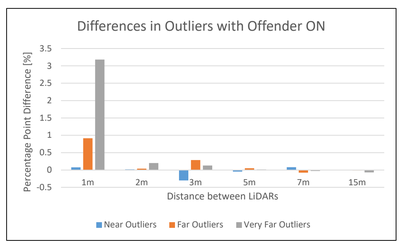


Fig.6: Percentage point differences in outliers at various distances with offender ON (both LiDARs at 0°).

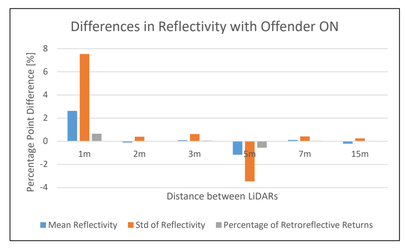


Fig.7: Percentage point differences in reflectivity at various distances with offender ON (both LiDARs at 0°).

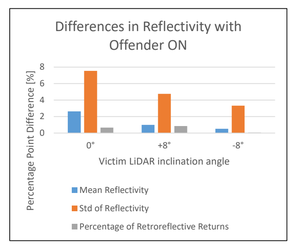


Fig.8: Differences in reflectivity at 1 m distance and victim inclination angles.

- At a 1 m distance, there were notable increases in the percentage of far and very far outliers with offender turned ON.
- At a 1 m distance, there were notable increases in mean reflectivity, st.dev. of reflectivity, and retroreflective returns.
- The victim inclination angles of +8° and -8° resulted in smaller changes in reflectivity.

- Results for the offender C (solid-state LiDAR):

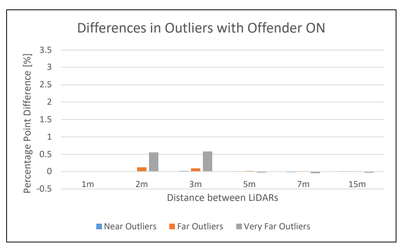


Fig.9: Percentage point differences in outliers at various distances with offender ON (both LiDARs at 0°).

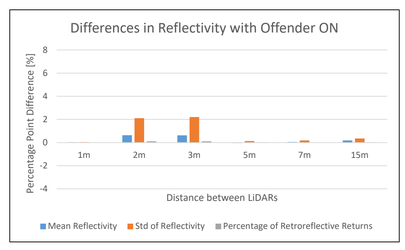


Fig.10: Percentage point differences in reflectivity at various distances with offender ON (both LiDARs at 0°).

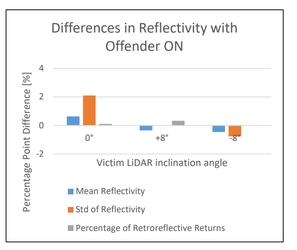


Fig.11: Differences in reflectivity at 2 m distance and victim inclination angles.

- At 2 m and 3 m distances, there were some increases in the percentage of far and very far outliers with offender turned ON.
- At 2 m and 3 m distances, there were some increases in mean reflectivity, st.dev. of reflectivity, and retroreflective returns.
- The victim inclination angles of 8° and -8° impacted the differences in reflectivity.

5 Conclusions

- **Aim**: Investigating LiDAR mutual interference in an outdoor setting.
- **Focus:** Specific LiDAR positions (distances and angles) were considered, using a narrow FoV MEMS LiDAR as the victim and different LiDARs as the offender.
- **Results:**
  - When the offender LiDAR was turned ON, possible interference effects were observed as changes (mostly increases) in the percentage of far and very far outliers in the victim LiDAR's point cloud. Possible interference effects were also registered as points with higher reflectivity values (including in the retroreflective range, as per the value definitions of the victim LiDAR).
  - These changes were most prominent at a 1 m distance for LiDAR B, and at 2 m and 3 m for LiDAR C. The size (and sign) of these changes were somewhat affected by inclination angles.
- **Significance:** One of the few studies on LiDAR interference in an outdoor setting.
- **Practical Applicability:** Based on the current results, there is insufficient support for changing LiDAR inclination angles as an effective interference mitigation strategy. More sophisticated mitigation techniques are likely needed.
- **Generalisation**: Given the magnitude of changes in the metrics, it appears that mechanical spinning LiDARs are stronger offenders than solid-state LiDARs, but more research is required using other LiDARs with different scanning patterns.
- **Future Research:** Accounting for environmental noise factors (in outside experiments) and wall scattering (in laboratory experiments). Incorporating calibrated LiDAR targets to isolate a smaller area of the point cloud for the analysis.

Acknowledgments: This research is part of the WMG-NPL Project on LiDAR Mutual Interference funded by the CCAV. This research is also partially supported by the High Value Manufacturing Catapult and Sim4CAMSens. The Authors would like to thank all the NPL staff that supported data collection at NPL, Teddington, 11-15 March 2024.

References:

[1] G. Kim, J. Eom, S. Park and Y. Park, "Occurrence and characteristics of mutual interference between LIDAR scanners," in SPIE Optics + Optoelectronics, Prague, 2015.
[2] I.-P. Hwang and C.-H. Lee, "Mutual Interferences of a True-Random LiDAR With Other LiDAR Signals," IEEE Access, vol. 8, pp. 2169-3536, 2020.
[3] D1.2 Interference Mitigation Report, Deliverable of the WMG-NPL project on LiDAR Mutual Interference. 2024.
[4] J. Robinson, A. Venturi, R. Dudley, M. Bevilacqua and V. Donzella, "The presence and effect of mutual interference in state-of-the-art LiDAR sensors," [In preparation].

Correlating Image Quality Metrics and DNN-Based Detection for Automotive Camera Data

Authors: Daniel Gummadi[1], Pak Hung Chan[1], Hetian Wang[1], Qingyang Wang[1], Valentina Donzella[1]

Affiliation: [1]WMG, University of Warwick

Abstract

Assisted and automated driving (AAD) systems rely on perception sensors like cameras. This study finds that traditional image quality assessment (IQA) metrics (e.g. SSIM, and a retrained BRISQUE), strongly correlate with deep neural network (DNN) object detection performance against image degradation due to compression. Metrics such as IW-SSIM and DSS showed very high positive correlations over 0.9, while retrained BRISQUE achieved almost perfect correlations over 0.97 against perception performance degradation under increasing levels of compression noise. These insights enable better prediction of perception degradation, crucial for enhancing AAD system development. The full paper is available at DOI:10.36227/techrxiv.24566371.v1

1 Background

AAD systems rely on camera images for perception tasks like object detection. But degradations like adverse weather or compression can degrade image quality, thus affecting system performance. Traditional methods (IQA Metrics) evaluate image quality in terms of realism and satisfaction for human consumers but fail to address automotive and AAD functions' needs, therefore there is a need to:

- Evaluate and quantify image quality (and more generally perception sensor data quality) for perception.
- Understand and quantify the level of degradation within images that can be tolerated by perception algorithms with minimal/acceptable variation of the perception performance.

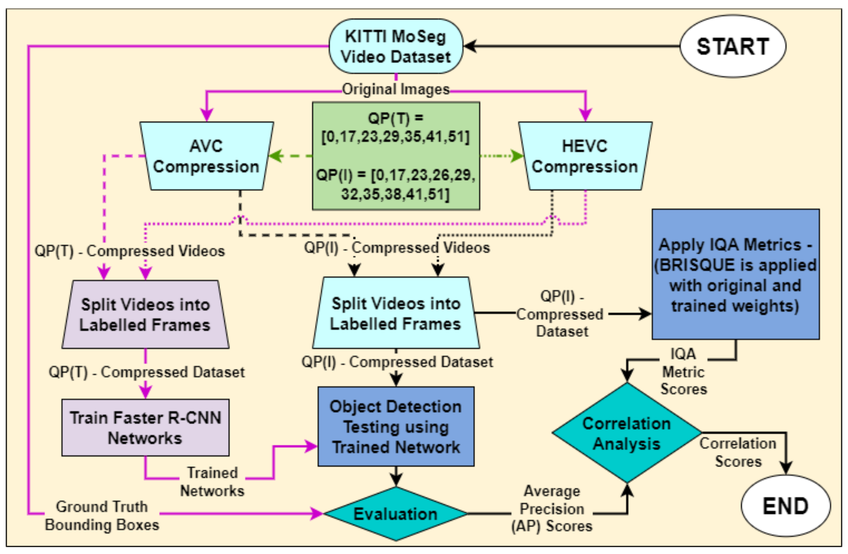
Our study focuses on evaluating image quality degradation caused by compression and its impact on DNN-based object detector performance.

- We prioritise compression because it is vital for efficient data transmission in automotive systems. As camera data transmission increases, cost-effective compression becomes necessary to handle limited bandwidth. Linking image quality degradation caused by compression to perception performance can help determine compression levels needed to maintain acceptable perception performance within bandwidth limits, using camera data alone.

2 Objectives & Contributions

- Explore 17 widely used image quality assessment (IQA) metrics and their relationship with DNN-based algorithms as data degradation increases – demonstrating high correlations between IQA Metrics and DNN performance.
- Investigate why certain metrics show lower correlations with DNN performance based on their features.
- Show that some IQA metrics can be retrained to achieve high correlation scores with deep neural network (DNN) performance – thereby predicting DNN performance from only the input data.
- Create an IQA metrics coding library with IQA metrics used in this work. Find the library at:
  https://github.com/WMG-IV-Sensors/IQ_Metrics

3 Methodology



Dataset Compression:

- Used KITTI MoSeg Dataset – compressing images using AVC (H.264) and HEVC (H.265) compression.
- Compressed training dataset at 7 QP levels (excluding 26, 32, 38 QP Levels) – resulting in 7 training datasets
- Compressed testing dataset at 10 QP levels – resulting in 10 testing datasets
- Created 14 compressed training datasets and 20 compressed testing datasets.
- This resulted in 140 train and test combinations (70 for AVC, 70 for HEVC).

Faster R-CNN Training, Testing and Evaluation:

- Selected Faster R-CNN model for Object Detection.
- Used stochastic gradient descent with batch size 1, learning rate 1e-3, and 10 epochs for training.
- Trained 14 detectors: 7 on AVC, 7 on HEVC.
- Evaluated detectors on corresponding testing datasets.
- Measured performance using average precision (AP) across 140 scores: 70 for AVC and 70 for HEVC. This resulted in 14 sets of AP scores (7 AVC, 7 HEVC), each set contained 10 AP scores corresponding to the 10 testing datasets compressed at 10 QP levels.

IQA Metrics Evaluation:

- Applied 17 IQA metrics to 20 compressed testing datasets (10 AVC, 10 HEVC).
- Calculated sets of mean, median, max, and min values for each IQA metric. For each IQA metric and averaging operation (mean, median, max, min) per compression method (AVC, HEVC), this resulted in each set containing 10 IQA scores corresponding to the 10 testing datasets compressed at 10 QP levels.

Correlation Analysis:

- Compared each set of IQA scores with each set of AP scores from the same testing dataset using Pearson's and Spearman's correlation coefficients
- Analysed correlation scores between IQA metrics and AP scores

4 Results

- Tables I and II contain the Pearson correlation scores which represent the linear correlation between the mean scores of each IQA metric and the evaluated DNN average precision (AP) for every object detector trained at 7 QP levels. Table I shows the mean of all correlation scores across all 7 QP levels. For both tables, green cells indicate a positive correlation, while red cells indicate a negative correlation, darker shades in colour represent stronger correlation. As seen in Tables I and II, most IQA metrics have very high correlations with DNN performance

Table I: Correlation Scores Between Average Precision And Mean IQA Metric Scores (Red Background = Negative Correlation, Green Background = Positive Correlation)

PLCC - AVC Training, AVC Testing											
Positive Correlation											
	PSNR	SSIM	CW-SSIM	DSS	HAAR-PSI	VIF	VSI	MS-SSIM	SR-SIM	IW-SSIM	FSIM
Mean correlation	0.367	0.934	0.888	0.950	0.874	0.700	0.894	0.963	0.936	0.978	0.885
Negative Correlation											
	GMSD		MS-GMSD		MDSI	BRISQUE		NIQE		PIQUE	
Mean Correlation	-0.900		-0.898		-0.468	-0.803		-0.948		-0.779	
PLCC - HEVC Training, HEVC Testing											
Positive Correlation											
	PSNR	SSIM	CW-SSIM	DSS	HAAR-PSI	VIF	VSI	MS-SSIM	SR-SIM	IW-SSIM	FSIM
Mean Correlation	0.363	0.928	0.895	0.945	0.866	0.713	0.900	0.903	0.940	0.963	0.892
Negative Correlation											
	GMSD		MS-GMSD		MDSI	BRISQUE		NIQE		PIQUE	
Mean Correlation	-0.905		-0.900		-0.515	-0.930		-0.920		-0.851	

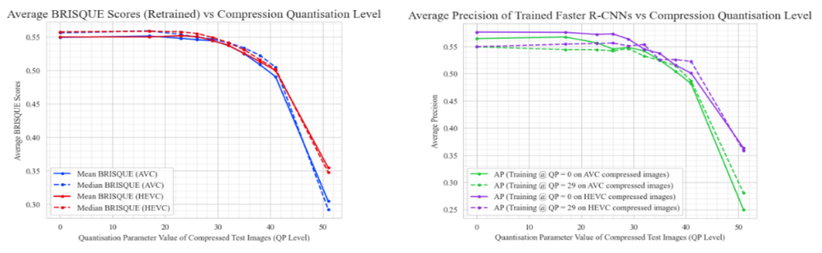
Table II: Correlation Scores Between Average Precision And Mean IQA Metric Scores (Original BRISQUE Vs Re-trained BRISQUE) (Red Background = Negative Correlation, Green Background = Positive Correlation)

PLCC - Mean BRISQUE scores					
AVC Training and AVC Testing			HEVC Training and HEVC Testing		
QP Training Value	BRISQUE (Original)	BRISQUE (Retrained)	QP Training Value	BRISQUE (Original)	BRISQUE (Retrained)
AP @ QP = 0	-0.858	0.999	AP @ QP = 0	-0.954	0.992
AP @ QP = 17	-0.853	0.999	AP @ QP = 17	-0.956	0.996
AP @ QP = 23	-0.830	0.998	AP @ QP = 23	-0.956	0.990
AP @ QP = 29	-0.821	0.998	AP @ QP = 29	-0.938	0.992
AP @ QP = 35	-0.810	0.994	AP @ QP = 35	-0.948	0.995
AP @ QP = 41	-0.698	0.969	AP @ QP = 41	-0.880	0.925
AP @ QP = 51	-0.748	0.965	AP @ QP = 51	-0.878	0.955
Mean Correlation	-0.803	0.989	Mean Correlation	-0.930	0.978

- Figures 1(a) and 1(b) show the decrease in average scores of one IQA metric (BRISQUE (Re-trained)) and the AP scores of two trained object detectors when applied to/tested on compressed images with increasing QP levels (increasing compression). As seen, the IQA metric trend matches the performance decrease of the detectors as compression levels (QP) increases – reinforcing the high correlation scores seen in Tables I and II above.

Figure 1(a): Average Retrained BRISQUE scores vs QP (Blue – Mean Score vs QP, Red = Median Score vs QP, Solid-Line = AVC, Dashed-Line = HEVC).

Figure 1(b): Faster-RCNN Average Precision vs QP (Purple = AP vs QP (HEVC Trained Detectors), Green = AP vs QP (AVC Trained Detectors), Solid-Line = Detectors Trained @ 0 QP, Dashed-Line = Detectors Trained @ 29 QP.



(a) (b)

- Table III shows the time taken to compute each IQA metric from the images in the KITTI MoSeg Dataset in milliseconds. This provides some context as to which IQA metrics can be used to estimate ideal compression levels (in terms of QP level) in real-time.

Table III: Time Taken To Compute A Quality Score For Each IQA Metric Per Input – For FR-IQA Metrics The Input Is A Single Pair Of Images And For NR-IQA Metrics The Input Is A Single Image (Image Size = (1242 by 375) Pixels)

Metrics With Positive Correlation											
	PSNR	SSIM	CW-SSIM	DSS	HAAR-PSI	VIF	VSI	MS-SSIM	SR-SIM	IW-SSIM	FSIM
Computing Time (milliseconds)	66.7	246	982	196	196	589	368	540	515	1350	3320
Metrics With Negative Correlation											
	GMSD		MS-GMSD		MDSI	BRISQUE		NIQE		PIQUE	
Computing Time (milliseconds)	98.2		221		418	982		295		1840	

5 Conclusions

- **Aim:** Create a relationship between traditional image quality analysis and machine learning-based perception for driving automation.
- **Focus:** Performance degradation of a DNN-based object detector using compressed camera data in assisted and automated driving functions.
- **Results**: Strong correlation between traditional IQA metrics and object detector performance, notably IW-SSIM and retrained BRISQUE showing extremely high positive correlations.
- **Significance**: The paper highlights importance of visual salience features and ineffectiveness of chrominance components in assessing object detector performance.
- **Practical Applicability:** Most metrics exhibit high correlations and computational efficiency, enhancing their practicality for evaluating/estimating object detector performance solely by assessing inputted data.
- **Generalisation:** Correlations made can be inferred to different types of DNN-based object detectors – not just Faster R-CNN.
- **Future Research:** Extend findings to explore use of IQA metrics for assessing a wider range of object detector models, other computer vision tasks or on other forms of noisy data to enhance understanding of broader utility in automotive applications.

6 Acknowledgments

Co-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 Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. Project grant no. 101069576. UK participants in this project are co-funded by Innovate UK under contract no.10045139. This work was also partially funded by Innovate UK (contract no. 10064608) as part of the Sim4CAMSens project.

References:
P. H. Chan, A. Huggett, G. Souvalioti, P. Jennings, and V. Donzella,"Influence of AVC and HEVC compression on detection of vehicles through faster r-cnn," IEEE Transactions on Intelligent Transportation Systems,
pp. 1–11, 2023.
B. Li, P. H. Chan, G. Baris, M. D. Higgins, and V. Donzella, "Analysis of automotive camera sensor noise factors and impact on object detection," IEEE Sensors Journal, vol. 22, no. 22, pp. 22 210–22 219, 2022.
Y. Wang, P. H. Chan, and V. Donzella, "Semantic-aware video compression for automotive cameras," IEEE Transactions on Intelligent Vehicles, 2023
P. H. Chan, G. Souvalioti, A. Huggett, G. Kirsch, and V. Donzella, "The data conundrum: compression of automotive imaging data and deep neural network based perception," London Imaging Meeting, vol. 2, no. 1, pp. 78–78, 2021. [Online]. Available: https://library.imaging.org/lim/articles/2/1/art00015

Benchmarking the Robustness of Panoptic Segmentation in Automated Driving

Authors: Yiting Wang[1], Haonan Zhao[1], Daniel Gummadi[1], Mehrdad Dianati[1], Kurt Debattista[1], Valentina Donzella[1]

Affiliation: [1]WMG, University of Warwick

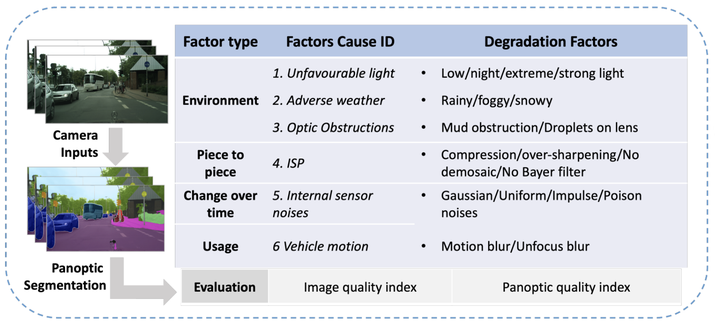
Abstract

Precise situational awareness is required for the safe decision-making of assisted and automated driving (AAD) functions. Panoptic segmentation is a promising perception technique to identify and categorise objects, impending hazards, and driveable space at a pixel level. While segmentation quality is generally associated with the quality of the camera data, a comprehensive understanding and modelling of this relationship are paramount for AAD system designers. Motivated by such a need, this work proposes a unifying pipeline to assess the robustness of panoptic segmentation models for AAD, correlating it with traditional image quality. The first step of the proposed pipeline involves generating degraded camera data that reflects real-world noise factors. To this end, 19 noise factors have been identified and implemented with 3 severity levels. Of these factors, this work proposes novel models for unfavourable light and snow. After applying the degradation models, three state-of-the-art CNN- and vision transformers (ViT)-based panoptic segmentation networks are used to analyse their robustness. The variations of the segmentation performance are then correlated to 8 selected image quality metrics. This research reveals This research reveals some interesting findings (see conclusion).

1 Motivation

- The **panoptic segmentation** aims to predict the semantic meaning for object instances and backgrounds represented with different colours.
- **Panoptic segmentation** is a promising perception technique to identify and categorise objects, impending hazards, and driveable space at the pixel level.
- **There are lots of noises in the real world,** which proposes the highest impact on the scene understanding task. and **which segmentation model** is most robust in facing noise factors?
- Can we **predict the perception degradation** simply using the image quality index, and which index to use?

2 Methodology



Factor type	Factors Cause ID	Degradation Factors
Environment	*1. Unfavourable light*	• Low/night/extreme/strong light
	2. Adverse weather	• Rainy/foggy/snowy
	3. Optic Obstructions	• Mud obstruction/Droplets on lens
Piece to piece	*4. ISP*	• Compression/over-sharpening/No demosaic/No Bayer filter
Change over time	*5. Internal sensor noises*	• Gaussian/Uniform/Impulse/Poison noises
Usage	*6 Vehicle motion*	• Motion blur/Unfocus blur
Evaluation	Image quality index	Panoptic quality index

*(a) The **19** identified noise factors.*

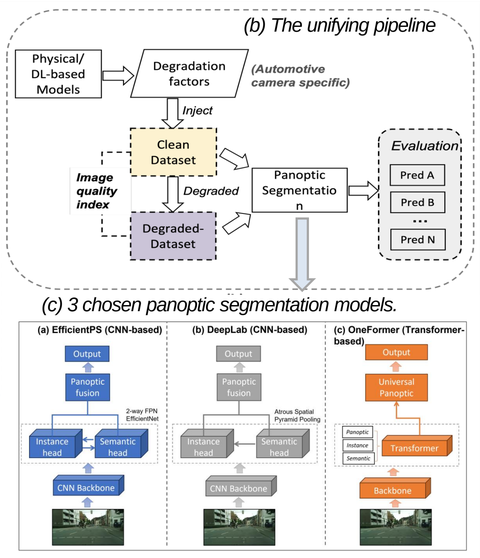


Fig. 1 The unifying pipeline to assess the robustness of panoptic segmentation.

- **Step 1**- Synthetic degraded driving dataset generation, with **19** noise factors under 3 different severity levels.
- **Step 2** - Panoptic segmentation models implementation. Using both the CNN and VIT architectures.
- **Step 3 –** Evaluation of the relationship between the perception models performance (PQ) using 8 backbones and 8 image quality index.

3 Noise Models

- The proposed snow model with the veiling effect is closer to the observation of the real-world.

Fig. 3 More vivid adverse environmental images under different severity levels, compared with other robustness research.

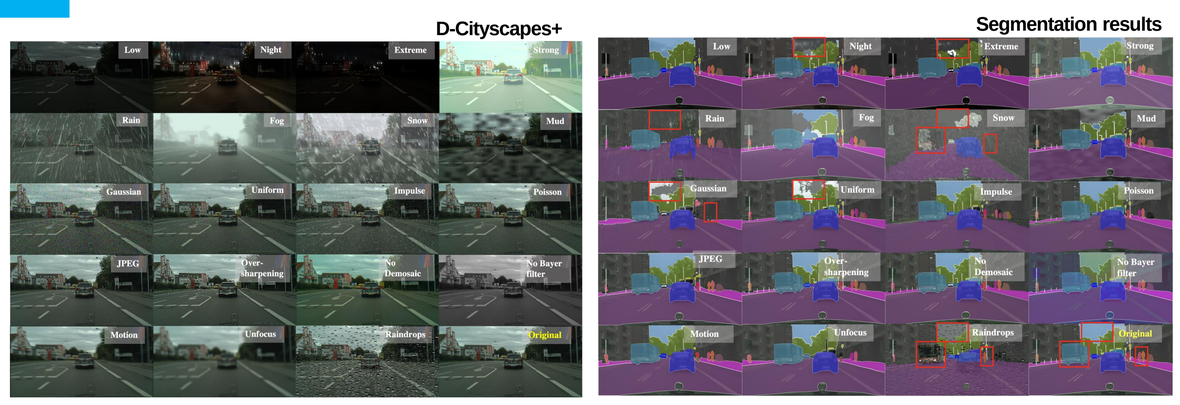
- The new definition of unfavourable light conditions.

low light	night light	extreme light
Entails uniform darkening of images.	Replicates AAD scenarios in urban areas characterized by traffic lights, headlights, and vibrant streetlights with a glare effect.	Represents even darker illumination in comparison to night light.

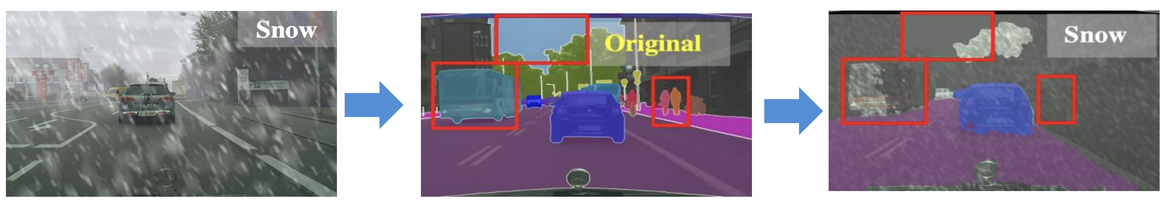
- Deep-learning-based methods for the unfavourable light model.

EC-Zero-DCE *CycleGAN* *Compound model*

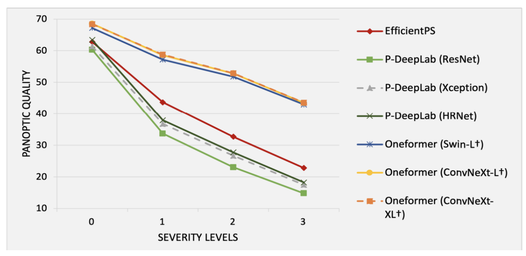
4 Results



- The panoptic segmentation results on D-Cityscapes+.



- Under extreme weather conditions, we find: **(1)** the sky cannot be accurately segmented. **(2)** pedestrians and bus cannot be identified under **extreme snow, raindrops** on the camera lens, even using the most SOTA panoptic segmentation model, posing safety risks.



- The average panoptic segmentation results using 8 backbones.
- **Oneformer** has the best PQ.
- EfficientPS is the fastest model.

- The panoptic quality are correlated to 8 selected image quality metrics.

Model	Index	PSNR↑	SSIM↑	FID↓	LPIPS↓	NIQE↓	CW-SSIM↑	FSIM↑	BRISQUE↓
EfficientPS	**PLCC**	0.807	0.948	-0.923	-0.948	-0.513	0.954	0.921	-0.519
EfficientPS	**SRCC**	0.800	0.933	-0.967	-0.867	-0.533	0.967	0.933	-0.6
Oneformer(ConNeXt-L)	**PLCC**	0.813	0.961	-0.947	-0.979	-0.569	0.975	0.966	-0.661
Oneformer(ConNeXt-L)	**SRCC**	0.800	0.933	-0.967	-0.933	-0.533	0.967	0.933	-0.6

5 Conclusions

- **Aim**: Benchmarking the Robustness of Panoptic Segmentation in Automated Driving under camera data degradation
- **Focus**: 19 noise factors, especially the adverse environmental conditions
- **Results:** ***(1)*** **c**ertain specific noise factors produce the highest **impact** on panoptic segmentation, i.e. droplets on lens and Gaussian***; (2)*** The **ViT-based** panoptic segmentation backbones show better robustness to the considered noise factors ***(3)*** Some image quality metrics (i.e. LPIPS and CW-SSIM) correlate strongly with panoptic segmentation performance
- **Significance:** The first robustness research in panoptic segmentation, benefiting the community when facing data degradation.
- **Practical Applicability:** ***(1)*** when considering the corner cases, pay more attention to the higher impact degradation conditions to improve the safety, robust and resilient of the AAD functions ***(2)*** when designing a more robust panoptic model, consider the use of transformer-based backbones ***(3)*** *consider the image quality index to indicate the perception degradation*
- **Feature Research:** ***(1)*** indicate the superiority of the proposed new synthetic model using real-world datasets ***(2)*** to design a more robust panoptic segmentation model in AAD

Acknowledgements: Funded by the European Union (grant no. 101069576). This research is partially sponsored by the Centre for Doctoral Training to Advance the Deployment of Future Mobility Technologies (CDT FMT) at the University of Warwick, High Value Manufacturing CATAPULT, Innovate UK (contract no. 10045139) and the Swiss State Secretariat for Education, Research and Innovation (contract no. 22.00123).

Darwick: A Paired Dataset in Low-Light Driving Scenarios for Advanced Perceptual Enhancement and Benchmarking Assessment

Authors: Zixiang Wei[1], Yunruo Jiang[1], Pak Hung Chan[1], Yiting Wang[1], Yuri Poledna[2] and Valentina Donzella[1]

Affiiliation: [1]WMG, University of Warwick; [2]CARISSMA , Technische Hochschule Ingolstad

Abstract

To effectively address the perception challenges present in low-light vehicular environments, we introduce the Darwick dataset, a high-quality resource specifically designed to bolster the visual perception capabilities of assisted and automated driving functions under low-light conditions. Darwick dataset stands out as the first of its kind, focusing on pixel-level paired low-light imaging tailored to diverse driving scenarios. Furthermore, we provide a benchmark analysis of various low-light enhancement techniques, demonstrating the practical applicability and superiority of our dataset in enhancing the performance of vision based systems in low-light conditions.

1 Background

- Existing strategies for camera-based low-light driving perception enhancement is divided into two categories: direct image manipulation and high-level enhancement.
- The former encompasses methodologies aimed at ameliorating image quality on a pixel-level basis. On the other end, indirect high-level enhancement eschews direct alterations to the image, encompassing domain adaptation and direct learning on low-light images.

2 Contributions

- **Uniquely Designed for Low-Light Driving:** The Darwick dataset is the **first** paired dataset tailored for low-light imaging across varied driving scenarios.
- **Benchmark of Low-Light Enhancement Methods:** Our studies benchmark the effectiveness of low-light enhancement techniques through image quality assessment and downstream perception tasks.

3 Related Works Comparison

- Comparison between Darwick and other low-light datasets

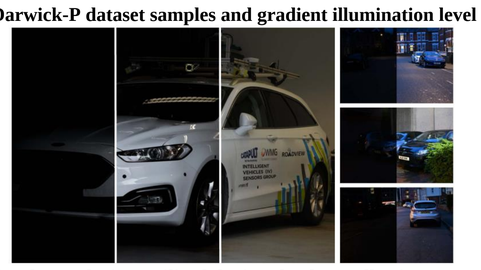
Datasets	Scale	Real?	Anno.	Pair?	Drive?
RENOIR	120	R	×	✓	×
LOL	500	R+Syn	×	✓	×
VE-LOL	13400	R+Syn	D	✓	×
SID	5,094	R	×	✓	✓
SICE	4,413	R	×	✓	×
ExDARK	7,363	R	D	×	✓
CRIE	2,450	R	×	×	✓
A*3D	11,700	R	D	×	✓
BDD	110,000	R	D+S	×	✓
KAIST	53,000	R	D	×	✓
TDND	115,072	R	D	×	✓
NightCity	4297	R	S	×	✓
DarkZurich	8377	R	S	×	✓
Nighttime	35,000	R	S	×	✓
Darwick (ours)	5,300	R	D+S	✓	✓

*R in the table indicates real-world and Syn indicates synthetic; D indicates object detection annotation and S indicates semantic segmentation annotation; (✓) indicates partially conforming.

4 Dataset Overview

Section	Biref Description	Pair?	Purpose
Darwick-P	2500 normal-light and 2500 low-light.	Yes	Training, Quantitative Evaluation
Darwick-UP	300 unpaired low-light captured using an in-vehicle camera.	No	Testing
Darwick-V	20 low-light videos, each one of 3-5 minutes in 60 fps.	No	Testing, Real Driving Processing Evaluation

Darwick-P dataset samples and gradient illumination level



Hardware devices utilized during the data collection process

(a) Camera Set (b) Spectroradiometer (c) Color Board

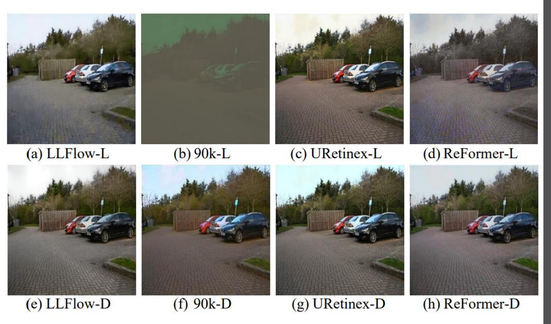
5 Experiment

- Quantitative results of SOTA low-light image enhancement baselines and following perception tasks based on Darwick.

Benchmarking IQA, perception accuracy metric on Darwick Test Set

Methods	Training Set	PSNR	SSIM	mAP	mIoU
Input	N/A	11.34	0.32	0.41	23.57
CLAHE	N/A	17.29	0.72	0.49	24.01
MSRCR	N/A	13.03	0.55	0.43	17.49
Wavelet	N/A	14.81	0.57	0.43	13.25
Gamma	N/A	11.47	0.33	0.41	14.68
KinD	VE-LOL	16.59	0.61	0.50	30.79
	Darwick-P	20.05	0.77	0.54	37.54
ReFormer	VE-LOL	16.93	0.75	0.52	30.62
	Darwick-P	21.42	0.83	0.56	46.92
URetinex	VE-LOL	13.46	0.53	0.45	15.65
	Darwick-P	18.74	0.67	0.49	28.51
LLFlow	VE-LOL	18.55	0.68	0.53	33.47
	Darwick-P	20.75	0.89	0.58	42.28
SMG-LLIE	VE-LOL	17.78	0.74	0.47	35.46
	Darwick-P	20.34	0.87	0.51	38.38
90k	VE-LOL	12.40	0.47	0.44	26.13
	Darwick-P	18.48	0.65	0.48	37.10

Visual results obtained from testing same image enhancement baselines after training on the Darwick and VE-LOL dataset.

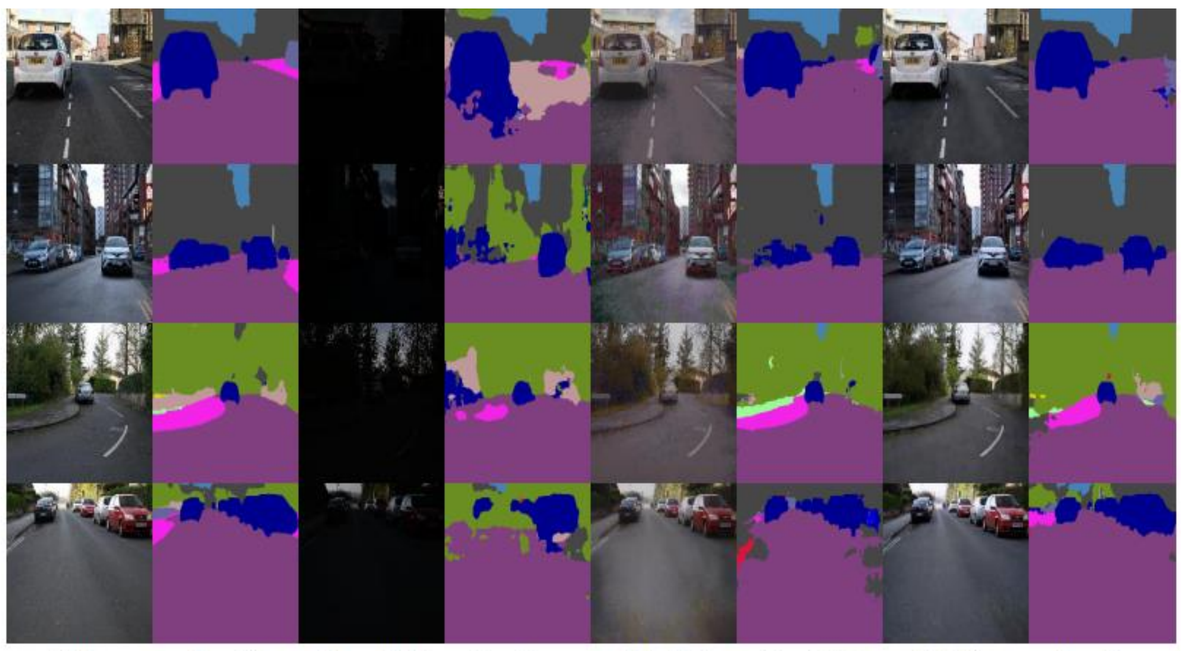


(a) LLFlow-L (b) 90k-L (c) URetinex-L (d) ReFormer-L
(e) LLFlow-D (f) 90k-D (g) URetinex-D (h) ReFormer-D

Visual results by testing RetinexFormer on DarkZurich dataset after training on the VE-LOL and Darwick dataset.

(a) Input frame 474 (b) Trained-LOL (c) Trained-Darwick
(d) Input frame 715 (e) Trained-LOL (f) Trained-Darwick

- Sample images from the Darwick dataset along with their corresponding enhancements and segmentation results.



GT Seg-GT Low-light Seg-Low Enh-LOL Seg-LOL Enh-Ours Seg-Ours

6 Conclusion

- **Higher Data and Annotation Quality**: Compared to similar datasets, Darwick has the best image resolution, lighting gradient reference, and perceptual artificial label quality.
- **Better Learning Resource:** The beneficial effects of using Darwick have been shown in terms of IQA of restored frames, and also in terms of downstream perception tacks (i.e. detection and segmentation). The learning methods using Darwick outperform using SOTA datasets.
- **Future Work:** Future updates to the dataset will incorporate weather variability and motion blur. Motion blur will also be added to simulate dynamic effects.

Acknowledgement: *Co-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 Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. Project grant no. 101069576. UK participants in this project are co-funded by Innovate UK under contract no.10045139.*