XAI – Autonomous Driving

Motivation

- Internship at automotive company
- Family in industry
- Popularity

Abbreviations

Abbreviation	Meaning
AD	Autonomous Driving
AV	Autonomous Vehicle
DL	Deep Learning
DLM	Deep Learning Model

Overview

- Study content
- Results
- Critique

Explanations in AD: A Survey

- IEEE
- November 2021
- 272 citiations

Explanations in AD: A Survey - Content

- Survey for behaviour of AVs
- Need for XAI for AVs
- Examination of AV regulations and standards
- Explainable AD operations
- Framework for AV explainability
- Fundamental knowledge for researchers

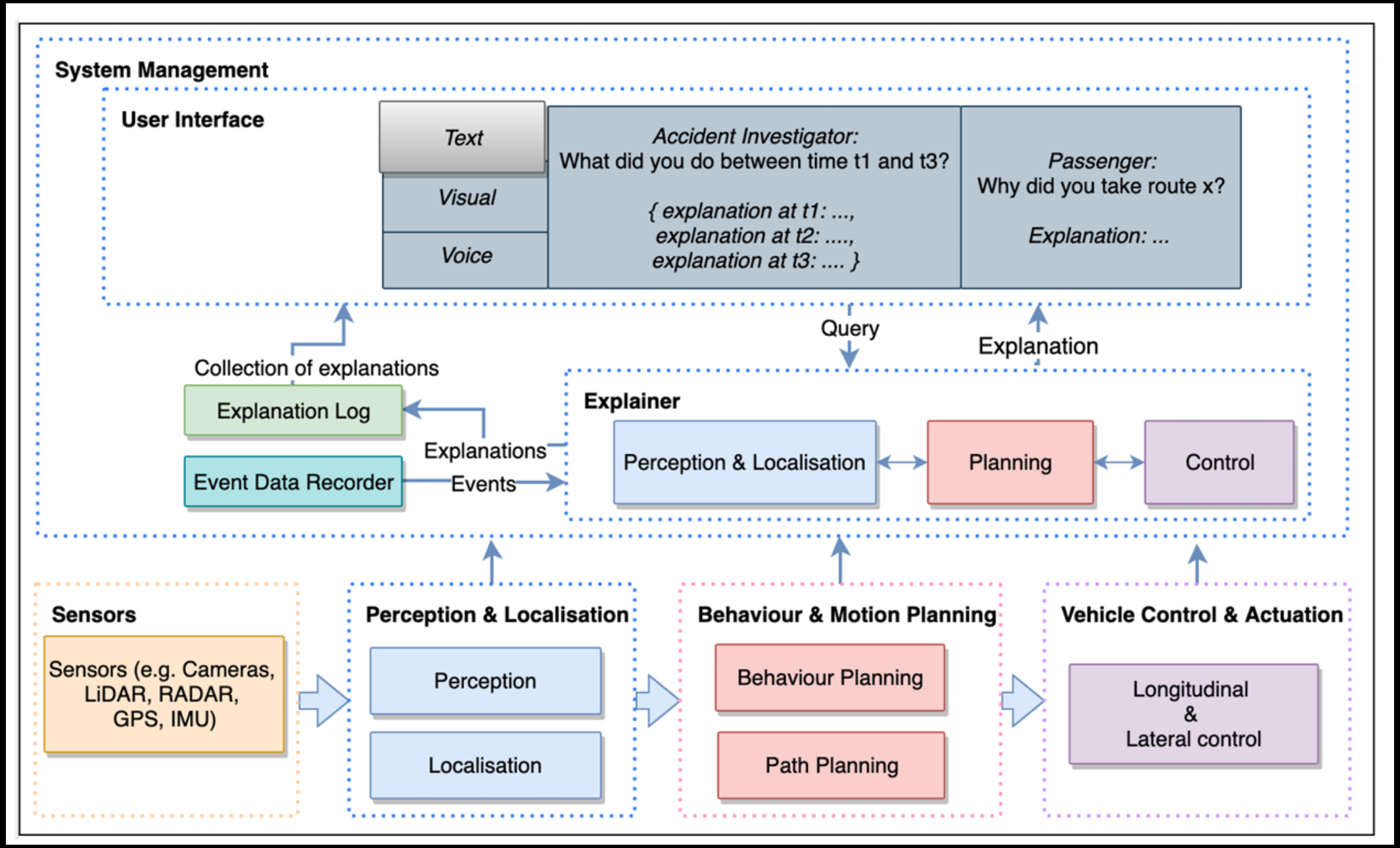
Need for XAI for AVs

- Trust
- Safety
- Future
- Decisions

Aim	Standard & Description	Stakeholder
Human Safety	**ISO 19237:2017** Pedestrian detection and collision mitigation systems	**Class B and C** AV Developers, Regulators, System Auditors, Accident Investigators, Insurer
	ISO 22078:2020 Bicyclist detection and collision mitigation systems	
	ISO 26262:2011: Road vehicles – Functional safety. An international standard for functional safety of electrical and/or electronic (E/E) systems in production automobiles (2011). It addresses possible hazards caused by the malfunctioning behaviour of E/E safety-related systems, including the interaction of these systems.	
	ISO 21448:2019: Safety Of The Intended Functionality (SOTIF). Provides guidance on design, verification and validation measures. Guidelines on data collection (e.g. time of day, vehicle speed, weather conditions) (2019). (complementary to ISO 26262).	
	UL 4600: Standard for Safety for Evaluation of Autonomous Products. a safety case approach to ensuring autonomous product safety in general, and self-driving cars in particular.	
	SaFAD: Safety First for Automated Driving. White paper by eleven companies from the automotive industry and automated driving sector about frameworks for development, testing and validation of safe automated passenger vehicles (SAE Level 3/4).	
	RSS (Intel) / SFF (NVIDIA): Formal Models & Methods to evaluate safety of AV on top of ISO 26262 and ISO 21448 (proposed by companies).	
	IEEE Initiatives: "Reliable, Safe, Secure, and Time-Deterministic Intelligent Systems (2019)"; "A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems" (2019); "Assessment of standardization gaps for safe autonomous driving (2019)".	
	The Autonomous: Global safety reference, created by the community leading automotive industry players, which facilitates the adoption of autonomous mobility on a grand scale (2019).	
Information/Data Exchange	**ISO/TR 21707:2008: Integrated transport information, management, and control—Data quality in intelligent transport systems (ITS)**. "specifies a set of standard terminology for defining the quality of data being exchanged between data suppliers and data consumers in the ITS domain" (2018).	**Class A and C** Passengers, Auxiliary Drivers, Pedestrians, Regulators, System Auditors, Accident Investigators Insurers
	ISO 13111-1:2017: The use of personal ITS station to support ITS service provision for travellers. "Defines the general information and use cases of the applications based on the personal ITS station to provide and maintain ITS services to travellers including drivers, passengers, and pedestrians" (2017).	
	ISO 15075:2003: In-vehicle navigation systems—Communications message set requirements. "Specifies message content and format utilized by in-vehicle navigation systems" (2003).	
	ISO/TR 20545:2017: Vehicle/roadway warning and control systems. "Provides the results of consideration on potential areas and items of standardization for automated driving systems" (2017).	
	ISO 17361:2017: Lane departure warning.	
	ISO/DIS 23150: Data communication between sensors and data fusion unit for automated driving functions.	

Explainable Autonomous Driving Operations

- Perception
- Localisation
- Planning
- Vehicle Control


System Management
User Interface
Text
Visual
Voice
Accident Investigator:
What did you do between time t1 and t3?
{ explanation at t1: ...,
explanation at t2: ....,
explanation at t3: .... }
Passenger:
Why did you take route x?
Explanation: ...
Query
Explanation
Collection of explanations
Explanation Log
Explanations
Event Data Recorder
Events
Explainer
Perception & Localisation
Planning
Control
Sensors
Sensors (e.g. Cameras, LiDAR, RADAR, GPS, IMU)
Perception & Localisation
Perception
Localisation
Behaviour & Motion Planning
Behaviour Planning
Path Planning
Vehicle Control & Actuation
Longitudinal & Lateral control

Critique

- Good if no previous knowledge
- „irrelevant“ sections
- regulations
- Future research directions

OD-XAI: Explainable AI-Based Semantic Object Detection for Autonomous Vehicles

- Appl. Sc. Journal
- May 2022
- 49 citations

OD-XAI: Explainable AI-Based Semantic Object Detection for Autonomous Vehicles - Content

- Detection of road using DLM
- XAI to understand predictions and decisions
- XAI integrated AV system to improve explainability of black-box models

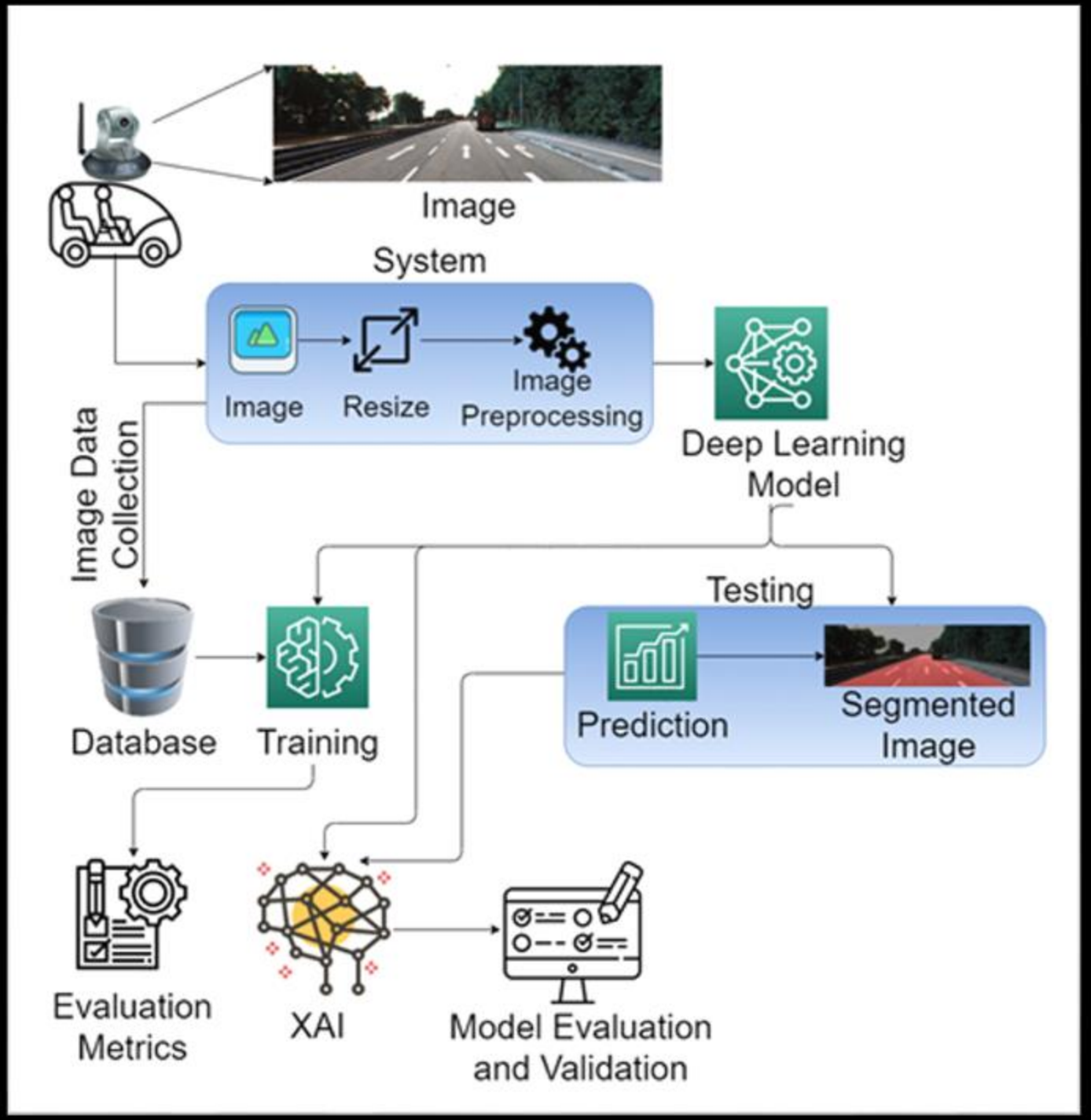


Image
System
Image
Resize
Image Preprocessing
Deep Learning Model
Image Data Collection
Testing
Database
Training
Prediction
Segmented Image
Evaluation Metrics
XAI
Model Evaluation and Validation

Approach

Model	Layers	Model Parameters
ResNet-18	111	17.271.149
ResNet-50	215	39.030.637
SegNet	65	73.613.762

XAI Integration

- Grad-CAM
- Saliency Map
- Observational Explanations

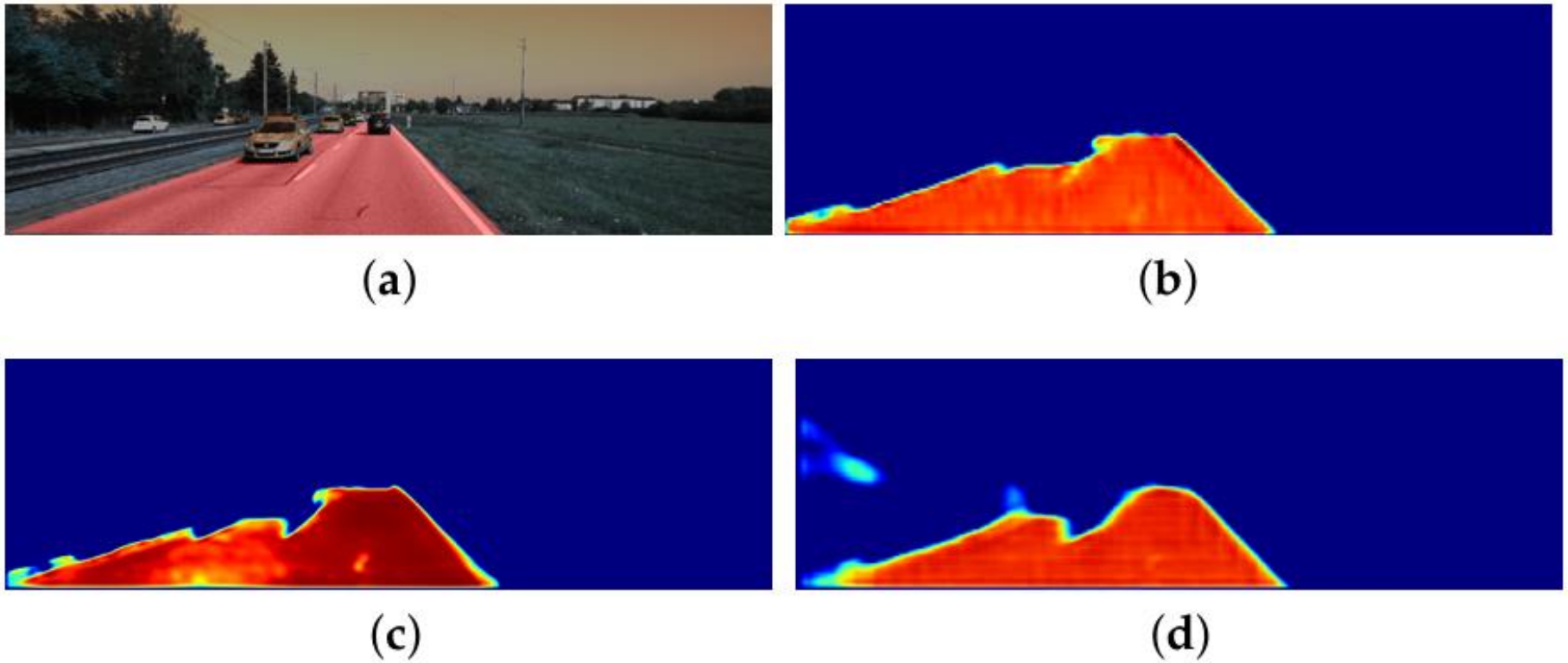


Figure 6. (**a**) Original road image from the KITTI dataset with an overlapped semantic map. (**b**) Grad-CAM results of ResNet-18 model. (**c**) Grad-CAM results of the SegNet model. (**d**) Grad-CAM results of the ResNet-50 model.

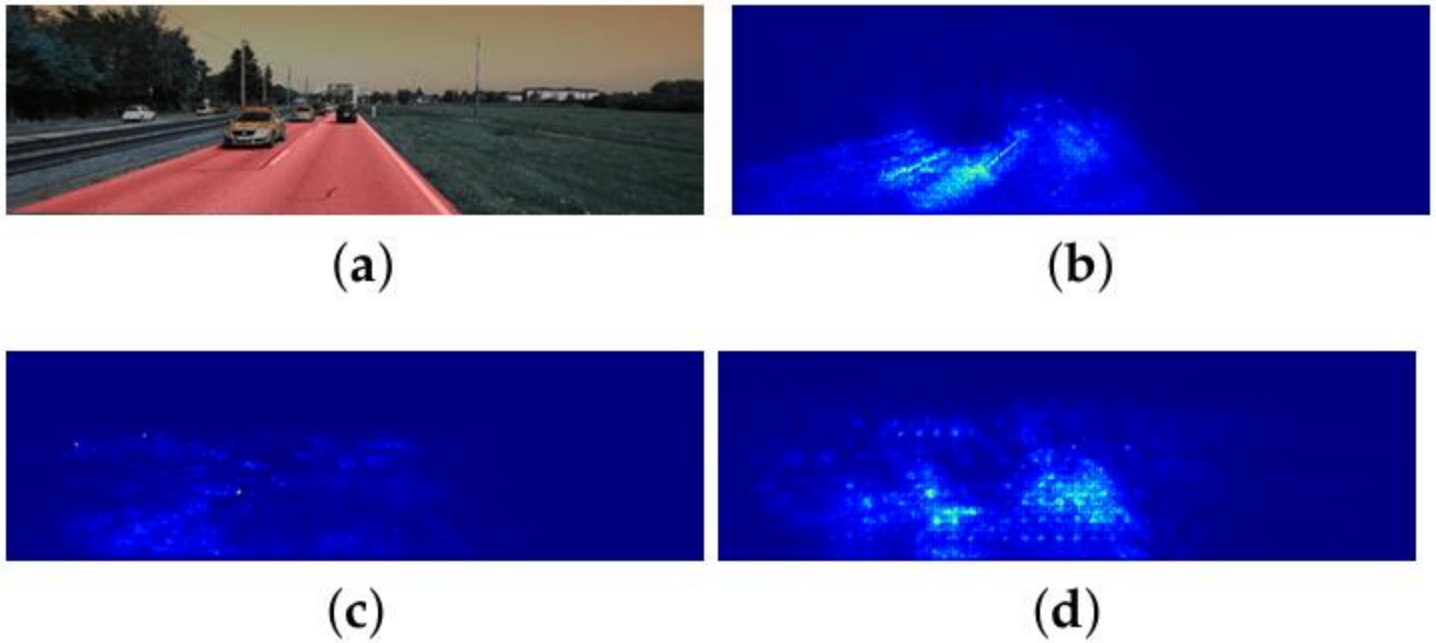


Figure 8. (**a**) Original road image from the KITTI dataset. (**b**) Saliency Map results of the ResNet-18 model. (**c**) Saliency Map results of the SegNet model. (**d**) Saliency Map results of the ResNet-50 model.

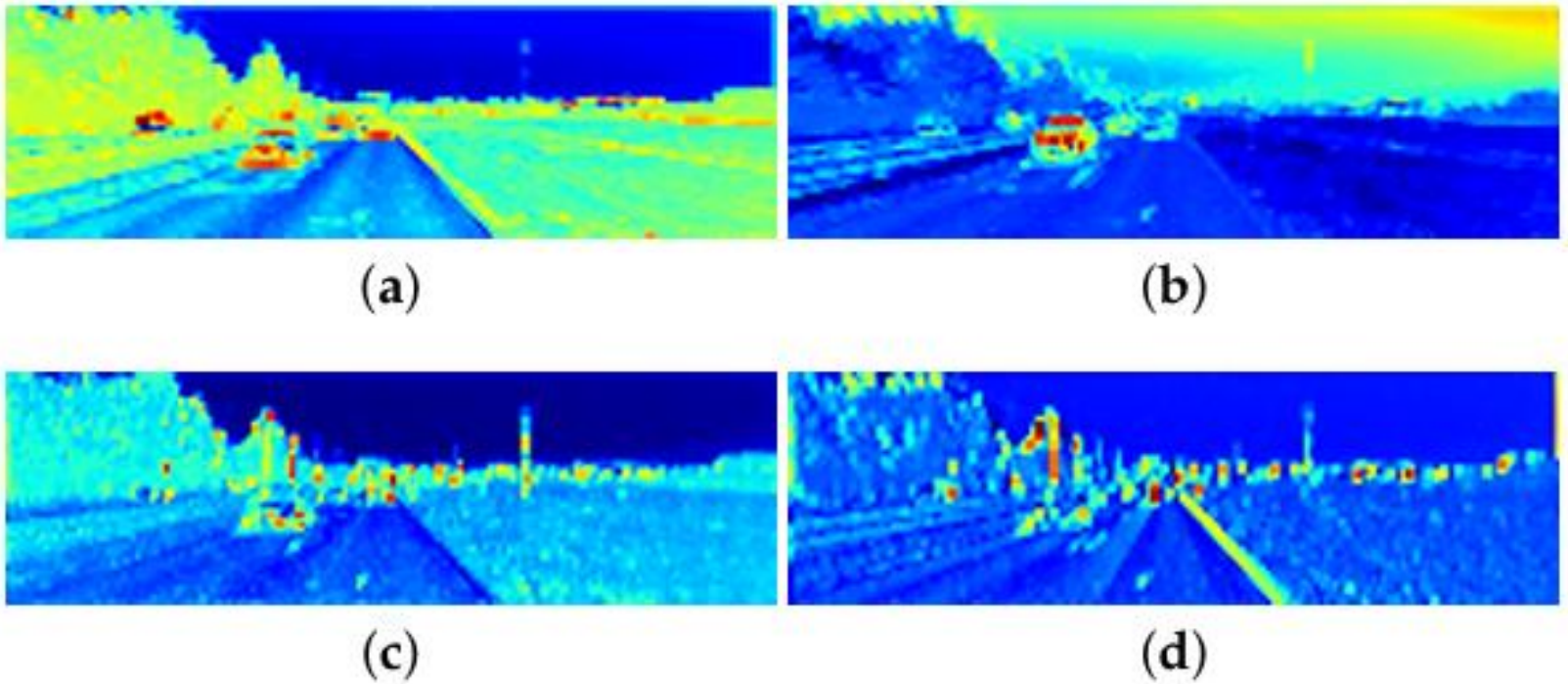


Figure 9. Model—ResNet18, layer 7, total channels—64, layer type—Max Pooling. (**a**) Output of channel 0. (**b**) Output of channel 5. (**c**) Output of channel 33. (**d**) Output of channel 63.

OD-XAI - Results

	ResNet-18	ResNet-50	SegNet
Train accuracy	0.9761	0.9281	0.9655
Test accuracy	0.9786	0.9318	0.9630

Critique

- „Irrelevant“ content
- Detailed description of approach and setup
- Different XAI integrations

XAI-ADS: An Explainable Aritifical Intelligence Framework for Enhancing Anomaly Detection in Autonomous Driving Systems

- IEEE
- April 2024
- 8 citations

XAI-ADS: An Explainable Aritifical Intelligence Framework for Enhancing Anomaly Detection in Autonomous Driving Systems - Content

- End-to-end XAI framework for anomaly detection
- Introduction of two novel feature selection methods

Framework

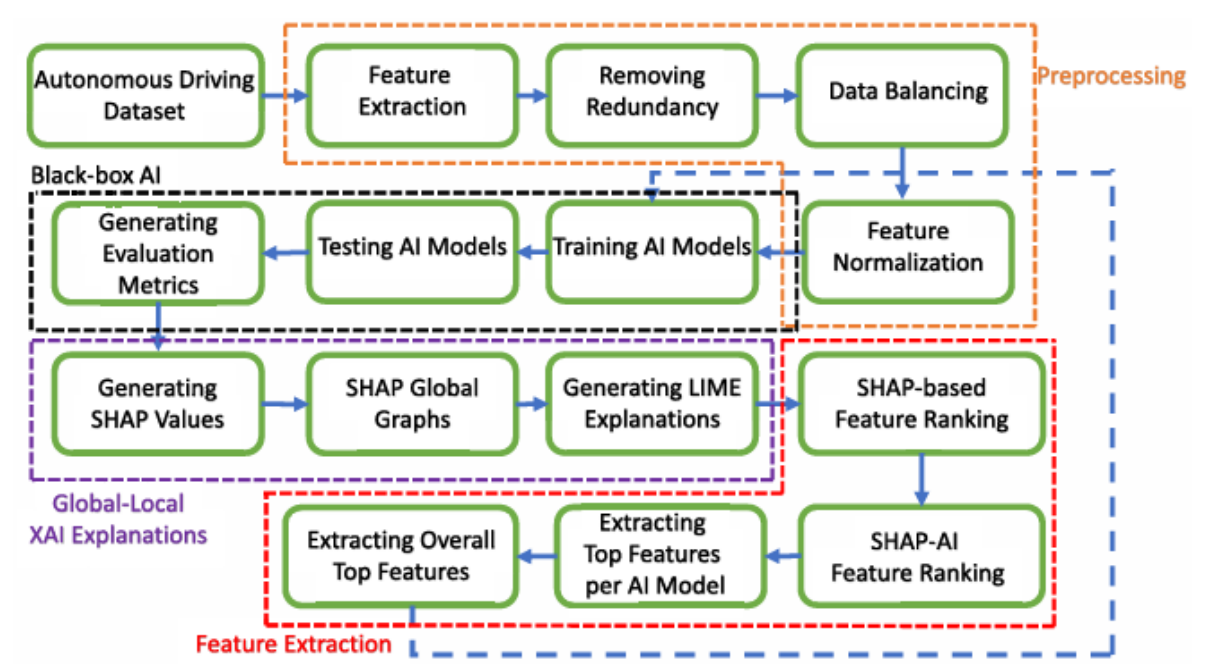


FIGURE 1. An overview of different components in our XAI framework for AV status classification.

- feature extraction to understand attacks
- Remove invaluable data
- No feature dominates model analysis
- AI: 70% training, 30% evaluation
- XAI: SHAP, LIME

Novel Feature Selection Methods

- Select features
- Generate SHAP values
- SHAP-based: values as performance contribution
- SHAP-AI: F_{i} as sum of product of SHAP values
- Rank features
- Extract top k ranked features

Novel Feature Selection Methods

	Evaluation	DT		RF		DNN		KNN		SVM		AdaBoost	
		SHAP-based	SHAP-AI	SHAP-based	SHAP-AI	SHAP-based	SHAP-AI	SHAP-based	SHAP-AI	SHAP-based	SHAP-AI	SHAP-based	SHAP-AI
k=1	Accuracy	0.75	**0.76**	0.78	0.78	0.53	0.34	0.77	0.77	0.53	0.53	**0.79**	0.79
	Precision	0.77	0.77	0.80	0.79	0.67	**0.68**	**0.82**	0.82	0.53	0.53	0.83	0.82
	Recall	**0.91**	0.91	0.89	**0.90**	0.59	0.03	**0.85**	0.84	0.54	0.51	0.86	0.87
	F-1 score	0.83	0.83	0.84	0.84	0.63	0.06	0.83	0.83	0.54	0.52	**0.85**	0.84
k=2	Accuracy	0.77	0.78	0.79	**0.80**	0.42	0.57	0.77	0.77	0.55	0.54	0.79	0.79
	Precision	0.79	0.81	0.83	**0.83**	0.65	0.67	0.82	0.82	**0.54**	0.53	0.82	0.83
	Recall	0.89	0.88	0.87	0.88	0.28	0.72	0.85	0.85	0.67	0.61	**0.87**	0.87
	F-1 score	0.84	**0.84**	0.85	**0.86**	0.39	0.69	0.83	**0.84**	0.59	0.57	0.85	0.85
k=3	Accuracy	0.76	0.78	0.80	0.80	0.46	0.63	0.77	0.77	**0.55**	0.55	0.79	0.79
	Precision	0.78	**0.82**	0.83	0.83	0.67	0.66	0.82	0.82	0.55	0.54	0.83	0.83
	Recall	0.90	0.87	0.89	0.88	0.37	**0.91**	0.85	0.84	0.61	0.66	0.86	0.86
	F-1 score	0.83	0.84	0.86	0.86	0.47	0.77	0.83	0.83	0.58	**0.61**	0.85	0.85
k=4	Accuracy	0.76	0.79	0.80	0.80	**0.66**	0.41	**0.78**	0.78	0.55	0.55	0.79	**0.80**
	Precision	0.78	0.82	0.83	0.83	0.67	0.67	0.82	0.82	0.54	0.54	0.83	**0.84**
	Recall	0.90	0.87	0.88	0.88	0.98	0.24	0.85	0.85	0.69	**0.69**	0.86	0.86
	F-1 score	0.83	0.84	0.85	0.85	**0.79**	0.36	0.84	0.84	0.61	0.61	0.85	0.85
k=5	Accuracy	0.79	**0.79**	0.80	0.80	0.66	0.34	0.78	0.77	0.55	0.55	0.79	0.79
	Precision	0.82	0.82	0.83	0.83	0.67	0.64	0.82	0.82	0.54	0.54	0.84	0.83
	Recall	0.87	0.87	0.88	0.88	0.96	0.01	0.85	0.85	0.66	0.66	0.85	0.86
	F-1 score	0.84	0.84	0.86	0.85	0.79	0.03	0.84	0.83	0.60	0.60	0.85	0.85
k=6	Accuracy	0.79	0.79	0.80	0.80	0.66	0.66	0.79	0.79	0.66	0.66	0.80	0.80
	Precision	0.82	0.82	0.73	0.73	0.67	0.67	0.82	0.82	0.69	0.69	0.84	0.84
	Recall	0.87	0.87	0.64	0.64	0.99	0.99	0.85	0.85	0.92	0.92	0.85	0.85
	F-1 score	0.84	0.84	0.68	0.68	0.80	0.80	0.84	0.84	0.85	0.85	0.85	0.84

Critique

- Testing features
- Lots of graphics

Paper for long presentation

OD-XAI: Explainable AI-Based Semantic Object Detection for Autonomous Vehicles

Harsh Mankodiya [1], Dhairya Jadav [1], Rajesh Gupta [1], Sudeep Tanwar [1,*], Wei-Chiang Hong [2,*] and Ravi Sharma [3]

Sources

- AV - https://innovationatwork.ieee.org/autonomous-vehicles-for-today-and-for-the-future/
- Molly Problem - https://www.itu.int/en/ITU-T/focusgroups/ai4ad/Pages/MollyProblem.aspx
- Molly Problem picture - https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281
- Explanations in Autonomous Driving: A Survey – https://ieeexplore.ieee.org/abstract/document/9616449
- OD-XAI - https://www.mdpi.com/2076-3417/12/11/5310
- XAI-ADS - https://ieeexplore.ieee.org/abstract/document/10486915