EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR OBJECT DETECTION IN THE AUTOMOTIVE SECTOR

MARIOS SIGANOS (K3Y LTD)

AUTHORS: M. SIGANOS, P. RADOGLOU-GRAMMATIKIS, T. LAGKAS, V. ARGYRIOU, S. GOUDOS, K. E. PSANNIS, K-F KOLLIAS, G. F. FRAGULIS, AND P. SARIGIANNIDIS

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070214 (TRUSTEE).

TRUSTEE

INTRODUCTION

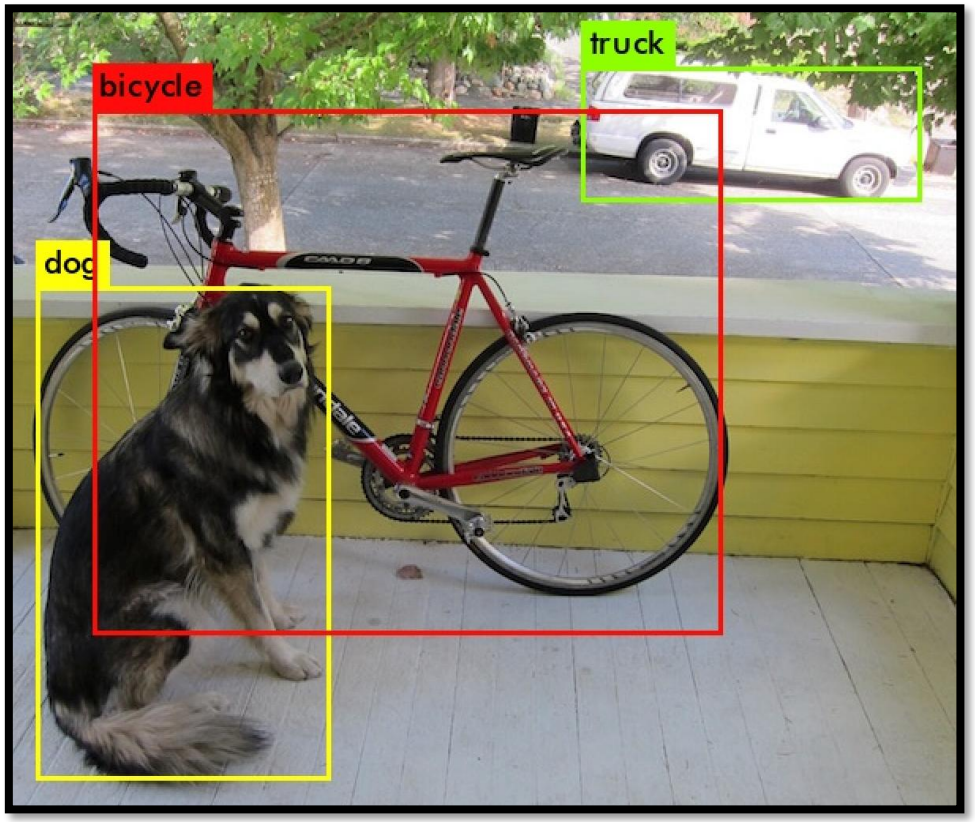
- Object detection is vital in automotive applications, enhancing safety and autonomy by identifying and classifying objects in the vehicle's surroundings, such as pedestrians, vehicles, cyclists, and road signs.
- Explainable Artificial Intelligence (XAI) plays a crucial role in making AI systems more transparent by providing insights into their decision-making processes and explaining the reasoning behind detections.
- In the automotive industry, the stakes are exceptionally high, especially in autonomous driving, where transparency, accountability, and trust are non-negotiable. Regulators and users demand systems that are not only accurate but also interpretable and reliable.
- By offering interpretable insights into object detection algorithms, XAI addresses these concerns, fostering confidence in the deployment of autonomous systems and ensuring compliance with ethical and safety standards.

BACKGROUND

- Computer Vision and Object detection
- Explainable AI (XAI) and Explainability
- XAI for Images and Explainable Object Detection

BACKGROUND
COMPUTER VISION AND OBJECT DETECTION

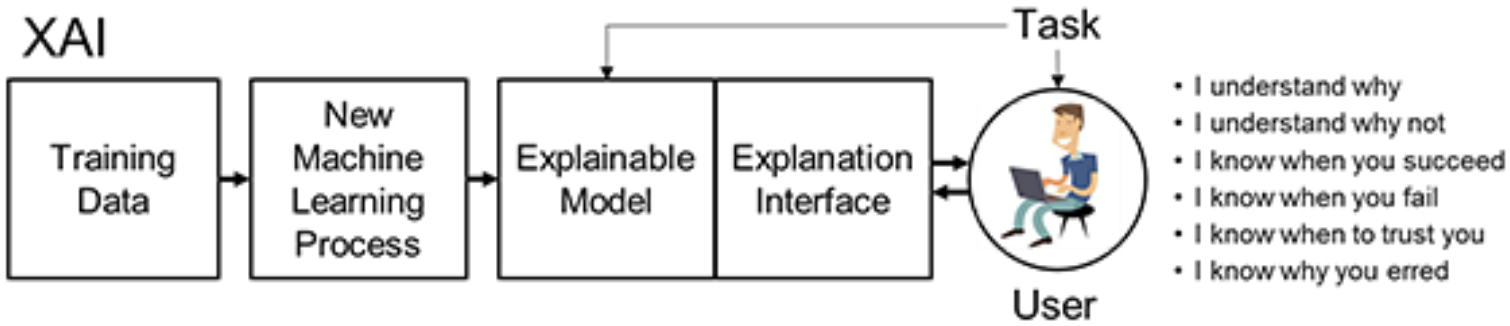
- **Computer Vision**: A branch of AI enabling computers to interpret and understand visual information.
  - *Key tasks: a) Image classification, b) Object detection, c) Image segmentation and d) Image generation*
- **Object Detection**: Identifies and localizes objects in images/videos by predicting bounding boxes and class labels.
  - *Applications: a) Autonomous vehicles, b) Surveillance systems, c) Medical imaging and d) Augmented reality*
- **Approaches to Object Detection**: a) One-stage detection algorithms (e.g., YOLO, SSD) and b) Two-stage detection algorithms (e.g., Faster R-CNN)
  - Comparison: One-stage algorithms prioritize speed and efficiency while two-stage focus on accuracy, often at the cost of slower performance.



Source: https://pjreddie.com/darknet/yolo

BACKGROUND
EXPLAINABLE AI (XAI) AND EXPLAINABILITY

- **What is XAI?** XAI creates systems that provide **clear, understandable explanations** for their decisions. It bridges the gap between **complex** ML models and the need for **transparency, accountability, and trust**.
- **What is Explainability?** The ability of an AI system to explain: a) **Why** specific decisions were made, b) **How** decisions were arrived at, and c) **What** factors influenced the output.
- **Approaches to Explainability**: a) **Interpretable models** – simple, inherently understandable models (e.g., decision trees, linear regression), b) **Transparent algorithms** – algorithms designed with built-in explainability (e.g., rule-based systems) and c) **Post-hoc explanations** – explanations for black-box models using techniques like feature importance scores, attention maps and sensitivity analysis.



Source: https://www.aporia.com/learn/explainable-ai/explainable-ai/

BACKGROUND
XAI FOR IMAGES AND EXPLAINABLE OBJECT DETECTION

- **XAI for images**: Explains how DL models process visual data and **identifies which image regions influenced the model's decisions** using techniques like feature visualization, saliency mapping, and attention mechanisms.
- **Class Activation Maps**: CAMs a) **visualize discriminative regions** in an image that contribute to a model's predictions, b) aid in **interpreting decisions** of DL models, particularly CNNs used in object detection and c) **produce heatmaps** that highlight key areas of focus, making model behavior more transparent.
- **Methods for generating CAMs**: a) **Gradient-based methods** (e.g., Grad-CAM) that compute gradients of the target class score relative to the last convolutional layer and highlight regions of importance using weighted feature maps, and b) **Gradient-free methods** (e.g., EigenCAM) that use principal components from learned representations to create visual explanations.

Source: Eigen-CAM: Class Activation Map using Principal Components

RELATED WORK

- **Topic**: XAI applications in the automotive sector
- **Focus areas**: object detection, dangerous vehicle behavior prediction, pedestrian intention recognition, traffic light identification, steering angle estimation, and traffic sign/vehicle classification
- **Techniques used**: Saliency maps, Grad-CAM, sensitivity analysis, and visual attention mechanisms
- **Challenges**: a) Black-box nature of DL models remains a limitation despite visualization techniques and b) limited scalability and real-time applicability of XAI methods in complex automotive scenarios.

RELATED WORK

Mankodiya et al. (2022)	Li et al. (2020)	Kim and Joe (2022)	Cultrera et al. (2020)
Semantic object detection for autonomous vehicles based on XAI. ❑ Training ResNet-18, ResNet-50, and SegNet on the KITTI road dataset for road detection ❑ Using GradCAM and saliency maps to generate heatmaps for explainability.	Vision-based framework tailored for autonomous driving. ❑ Focusing on object detection, pedestrian intention recognition, dangerous vehicle estimation, and signal recognition. ❑ Applying RISE to generate saliency maps for interpretable predictions.	A novel XAI method tailored for CNN-based models utilized in self-driving cars. ❑ Performing sensitivity analysis for CNNs, tested on traffic signs and vehicle/non-vehicle datasets. ❑ Outperforming SHAP, LIME, Grad-CAM and XCNN in delivering effective explanations.	Framework for explaining decisions made in autonomous driving scenarios through visual attention mechanisms. ❑ Using visual attention with a convolutional network to explain steering angle predictions. ❑ Utilising data from the CARLA urban driving simulator.

CONTRIBUTIONS

- **XAI in Automotive Object Detection**: Focuses on applying XAI techniques to object detection systems in the automotive domain, enhancing transparency and trust.

- **Explainability Module**: Presents a software component that integrates XAI methods into vehicle object detection systems, functioning as an explainability layer within a broader framework.

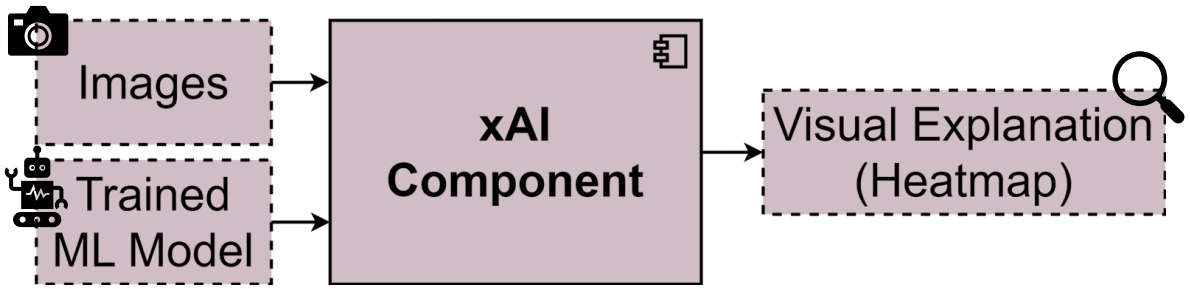
- **Object Detection with Visual Explanations**: Utilizes XAI techniques to generate heatmaps, clarifying which image features influenced object detections and the model's confidence in its predictions.

ARCHITECTURE

Steps:

1. Load an input image for processing.
2. Load a pre-trained YOLO model for object detection.
3. Pass the image through the YOLO model for object detection.
4. Identify and localize objects within the image, generating bounding boxes and class probabilities for each detected object.
5. Extract feature maps from various layers of the YOLO model that correspond to the detected objects. These maps capture the network's activations and responses to different image regions.
6. Apply EigenCAM to the extracted feature maps to create heatmaps. These highlight the image regions most influential to the model's decision for each detected object class.

High-level architecture of the XAI component



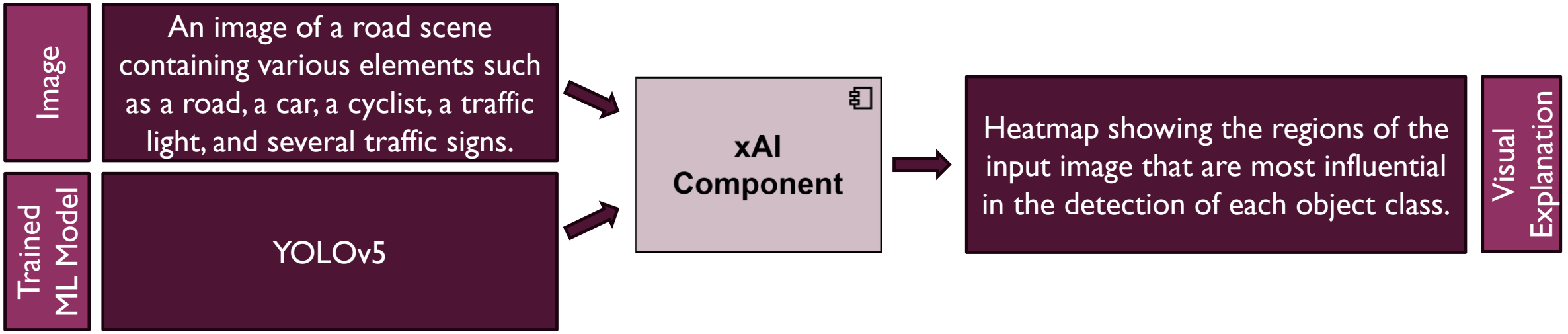
ARCHITECTURE
WHY EIGEN-CAM?

- **Seamless integration**: Eigen-CAM works with any CNN model without requiring changes to the architecture or retraining of the model.

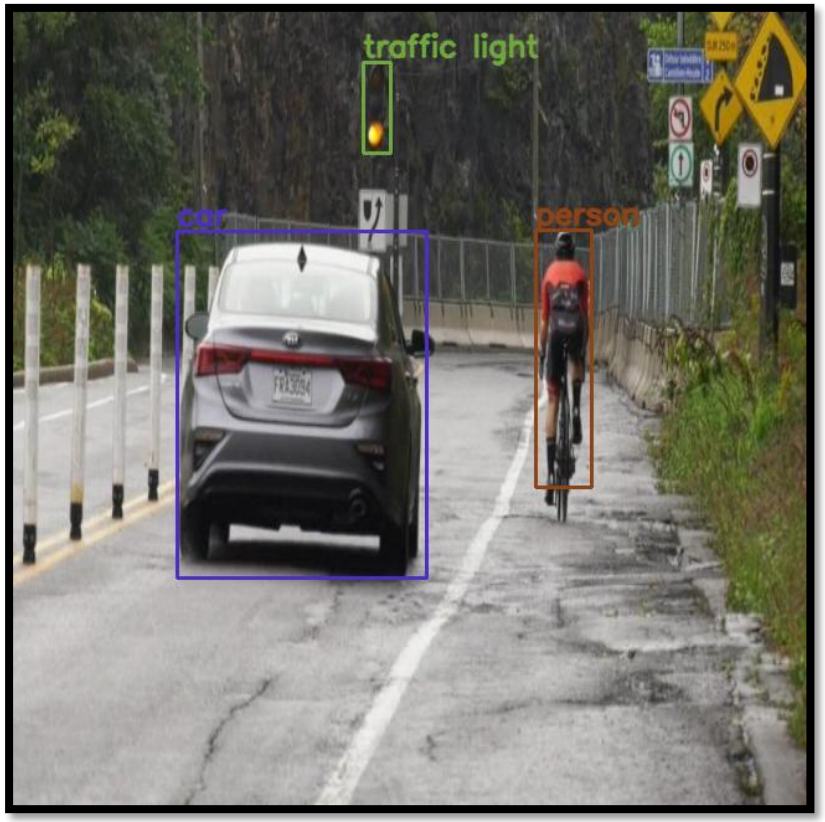
- **Multi-object explanations**: Eigen-CAM is capable of generating visual explanations for multiple objects in a single image, making it suitable for complex scenes.

- **Gradient-free operation**: Eigen-CAM operates independently of gradient back-propagation, improving its versatility and computational efficiency.

EXPERIMENTAL FINDINGS



EXPERIMENTAL FINDINGS

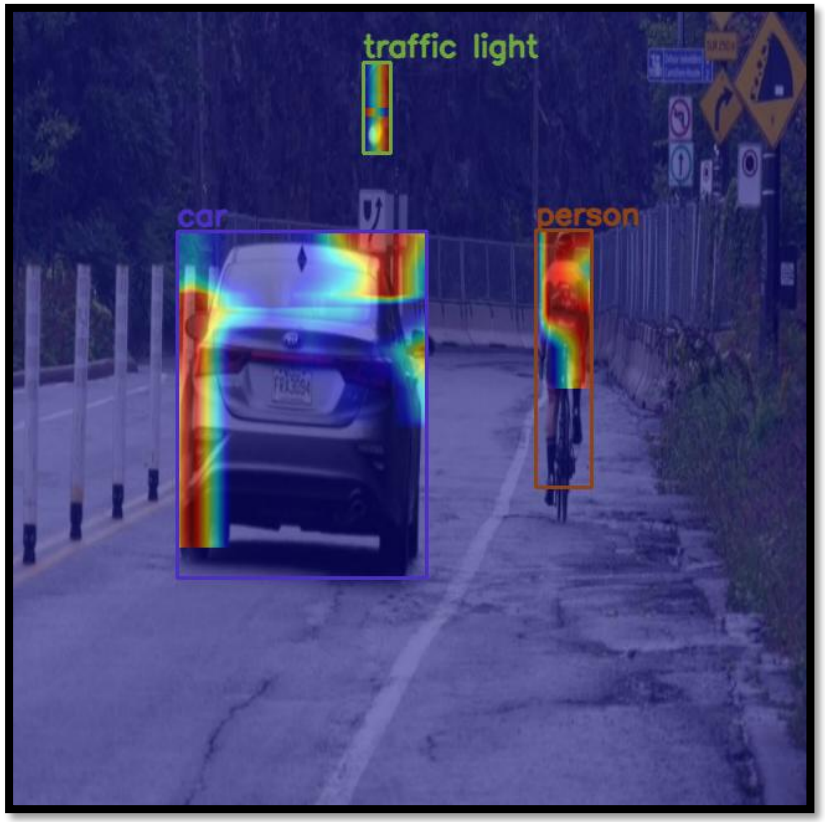


- The original image including the bounding boxes of the detected objects.

EXPERIMENTAL FINDINGS

- The regions with the most impact are highlighted.

EXPERIMENTAL FINDINGS



- Areas that have the most influence on each individually detected object are highlighted.

EXPERIMENTAL FINDINGS

CONCLUSIONS

- This work introduced a software component designed to enhance the explainability of pre-trained object detection models in the automotive domain.
- The proposed XAI component was evaluated using a YOLOv5 detection model, generating visual explanations for automotive-related images to demonstrate its capabilities.

Next steps and future directions:

- Incorporating additional CAM-based XAI methods for broader evaluation and improved interpretability.
- Assessing the XAI component with more state-of-the-art object detection models like Faster R-CNN and SSD.
- Measuring the the effectiveness of the XAI component through user satisfaction surveys and quantitative metrics.

THANK YOU!

EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR OBJECT DETECTION IN THE AUTOMOTIVE SECTOR

M. Siganos et al.

K3Y LTD

msiganos@k3y.bg

K3Y
R&D AND CYBER SECURITY

ETLTC-ICETM 2025
ETLTC CONFERENCE SERIES