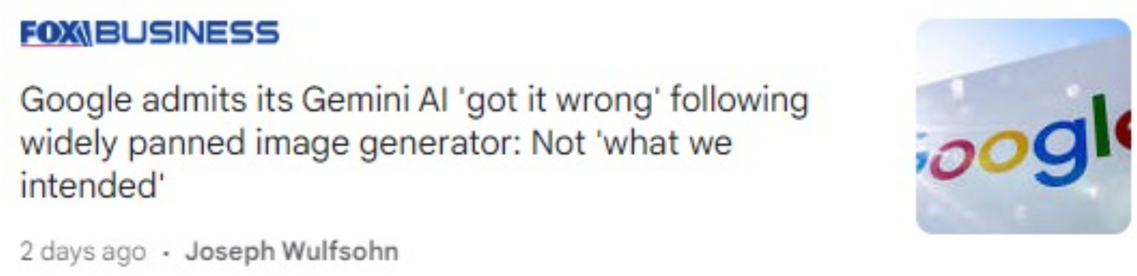
KI WISSEN

Automotive AI Powered by Knowledge

KI Wissen Final Event | 21-22 March 2024

Knowledge Extraction

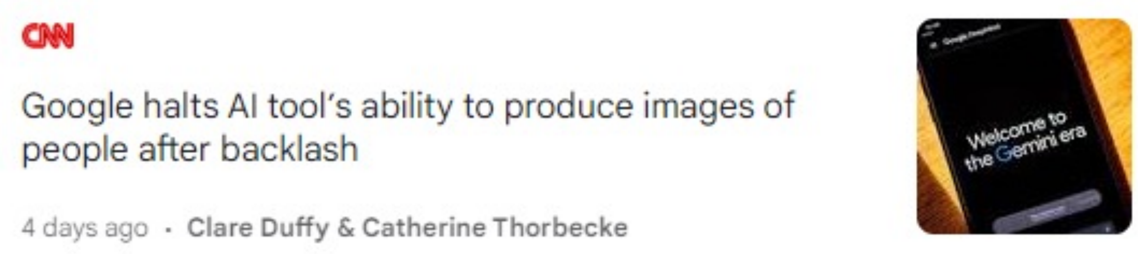
Antje Loyal | Continental



FOX BUSINESS

Google admits its Gemini AI 'got it wrong' following widely panned image generator: Not 'what we intended'

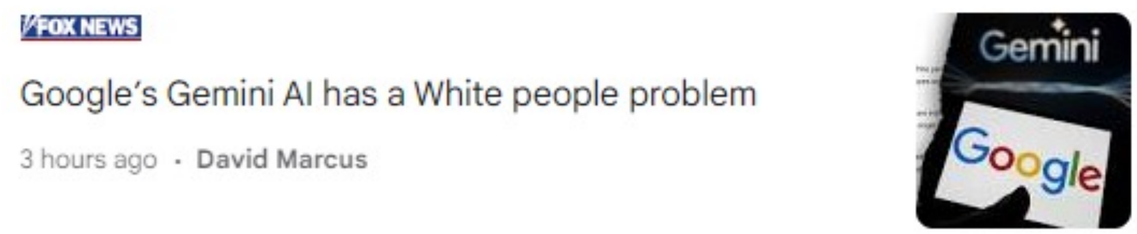
2 days ago · Joseph Wulfsohn



CNN

Google halts AI tool’s ability to produce images of people after backlash

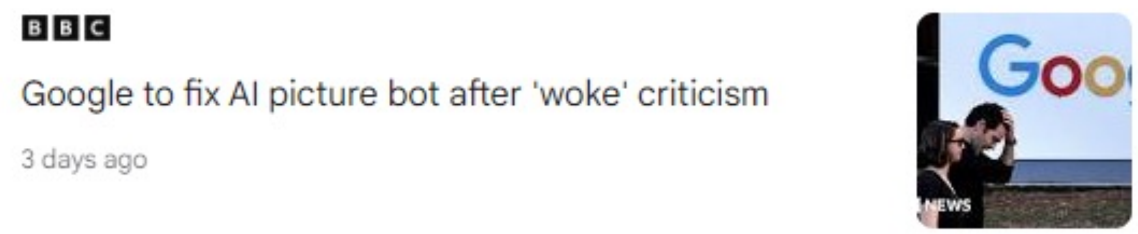
4 days ago · Clare Duffy & Catherine Thorbecke



FOX NEWS

Google’s Gemini AI has a White people problem

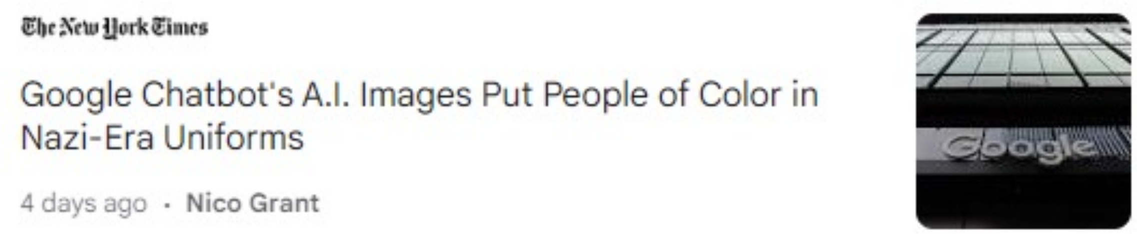
3 hours ago · David Marcus



BBC

Google to fix AI picture bot after 'woke' criticism

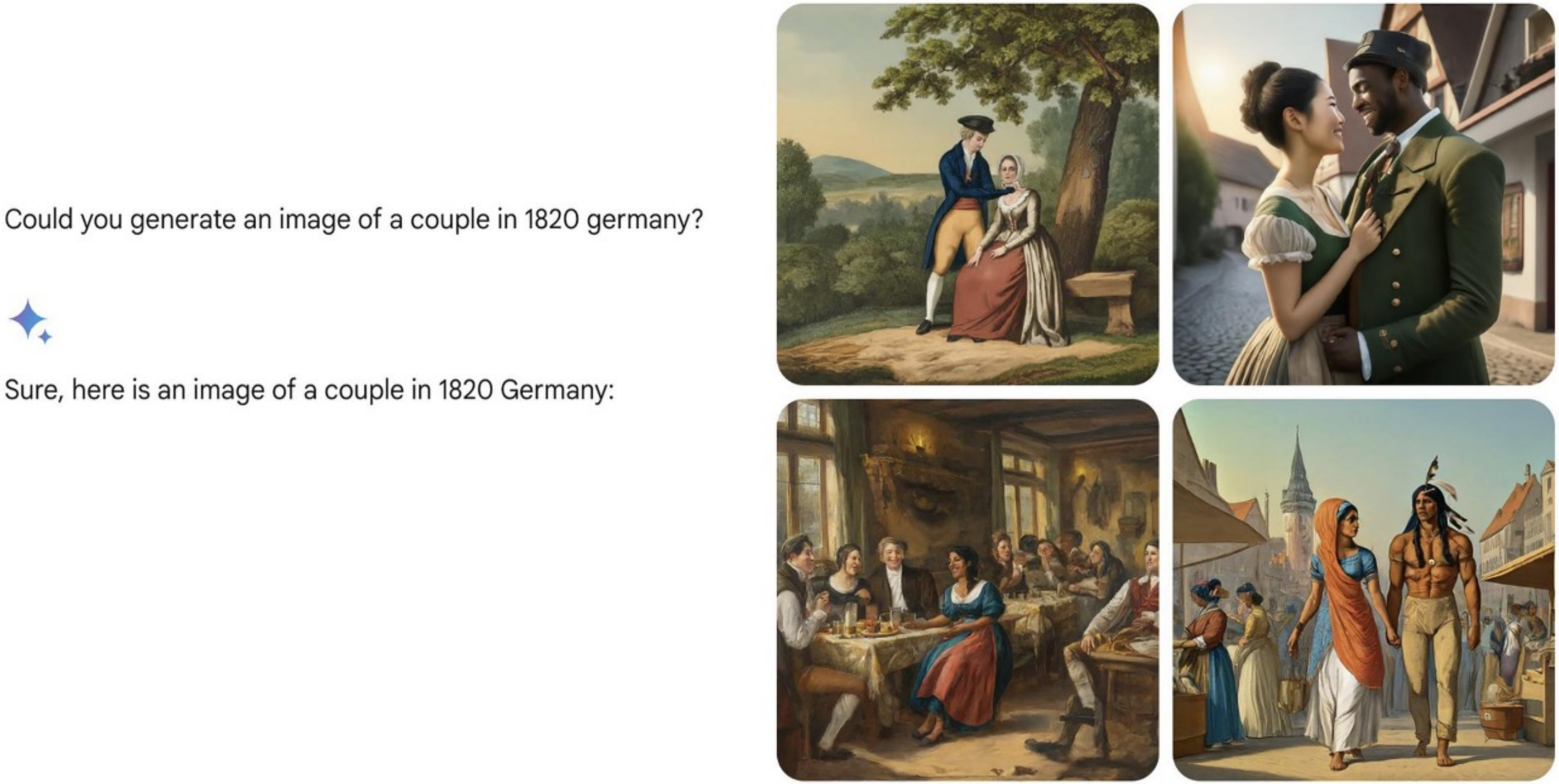
3 days ago



The New York Times

Google Chatbot's A.I. Images Put People of Color in Nazi-Era Uniforms

4 days ago · Nico Grant



Could you generate an image of a couple in 1820 germany?

Sure, here is an image of a couple in 1820 Germany:

KI WISSEN

Automotive AI Powered by Knowledge

» How can we ensure certain behaviors of AI systems by integrating knowledge, without hampering the safety in general?

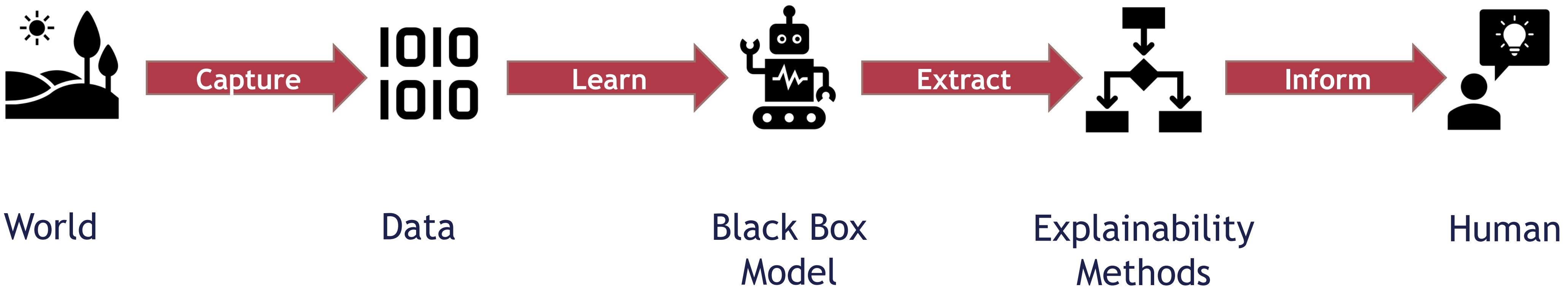
Two main questions you want to ask any AI system

How is it generating its outputs?

Why is it generating those outputs?

Explainable AI

Explainable AI



Capture → Learn → Extract → Inform

World

Data

Black Box Model

Explainability Methods

Human

- Using uninterpretable black box models comes with a certain risk
- The area of Explainable AI (XAI) aims to make the predictions more transparent and human-interpretable
- The methods we used are mostly **model-agnostic**

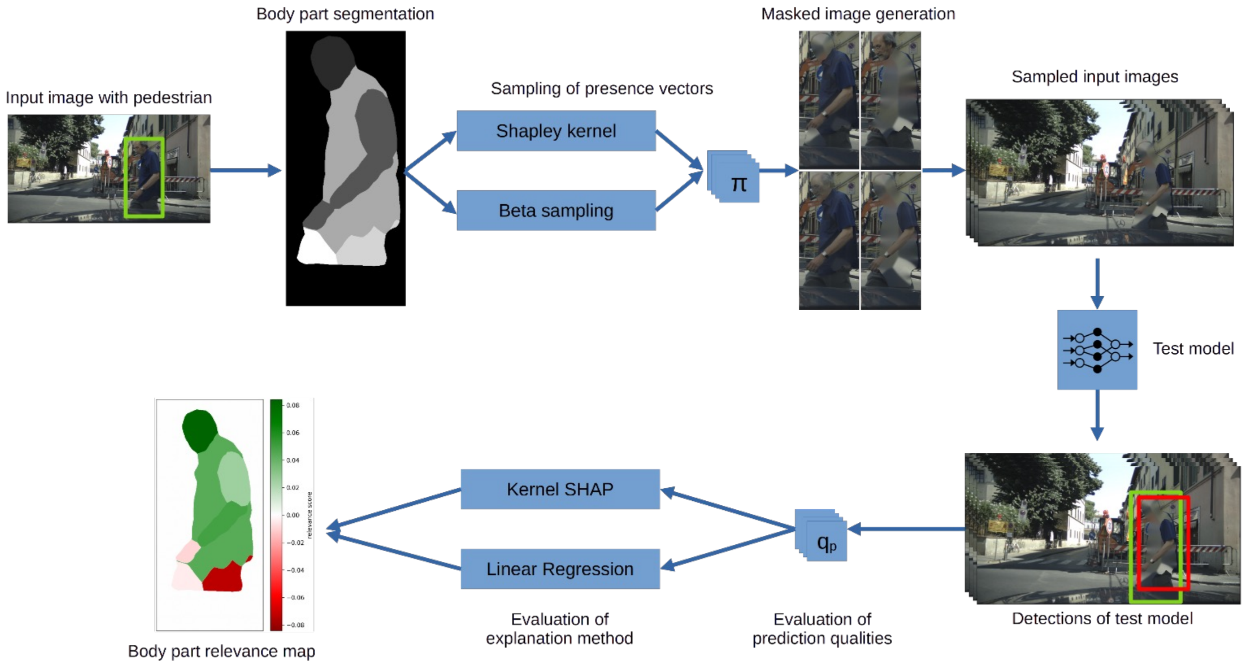
They can be applied to any black box model, independent of the architecture

XAI: Pedestrian Detection

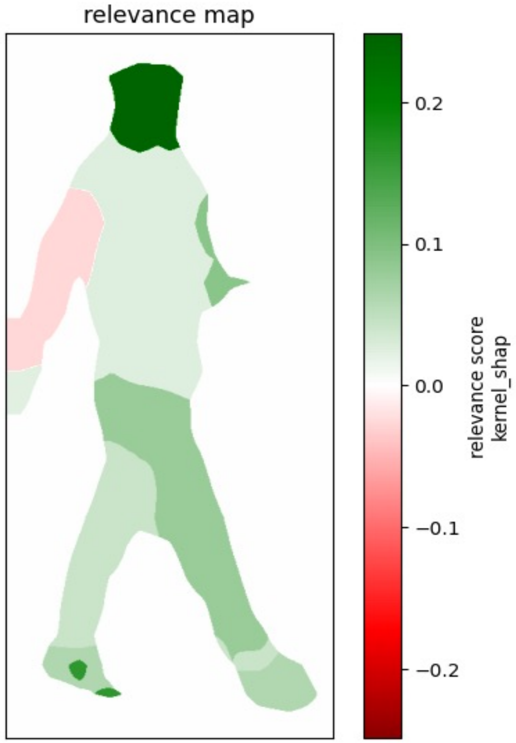
XAI: Pedestrian Detection

Which body parts are more or less important for the detection?

- Method samples instances of present or absent features
- Creation of images with masked and unmasked body parts
- Using KernelSHAP to calculate the contribution of each feature to output
- Resulting in a body part relevance map



XAI: Pedestrian Detection



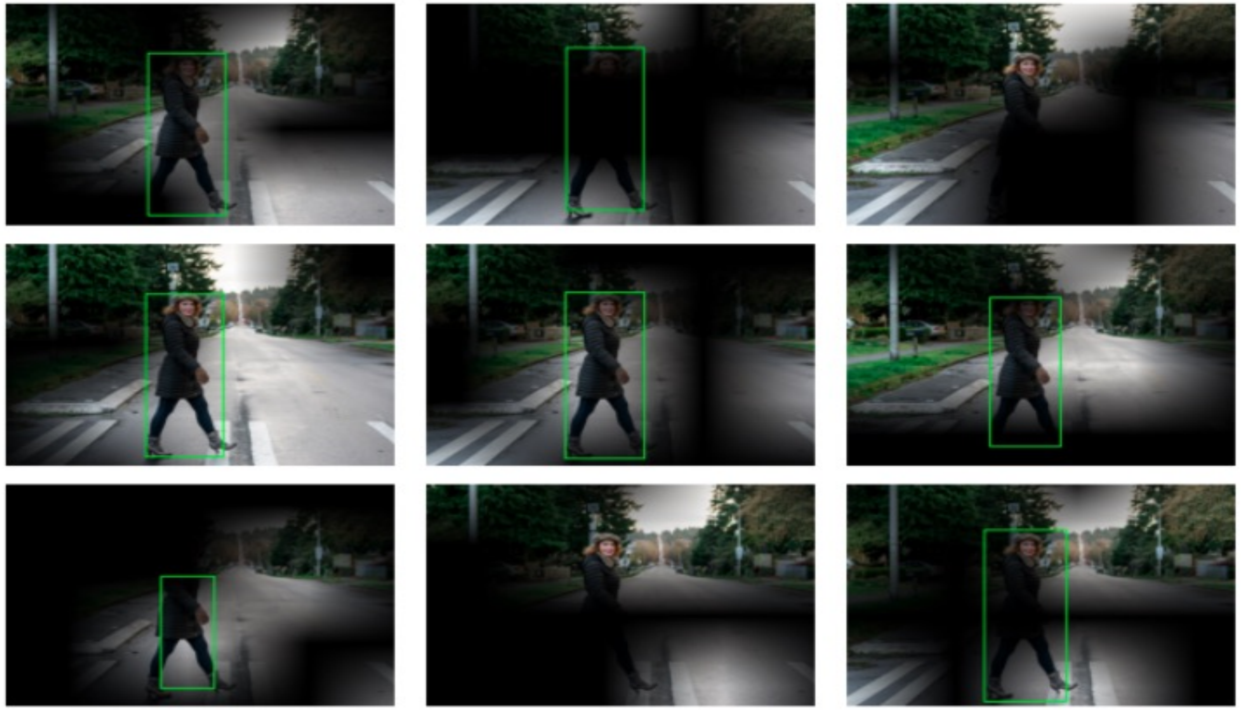
XAI: Pedestrian Detection

Which image areas are important for the detection?

- Is the model using the surrounding to make predictions?
- Estimates importance by element-wise multiplying image with random masks
- Generates saliency maps of pixels importance in the prediction

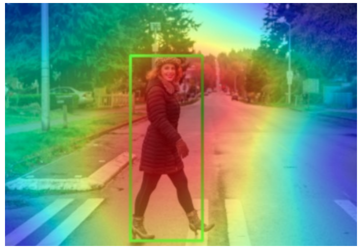
Input

→



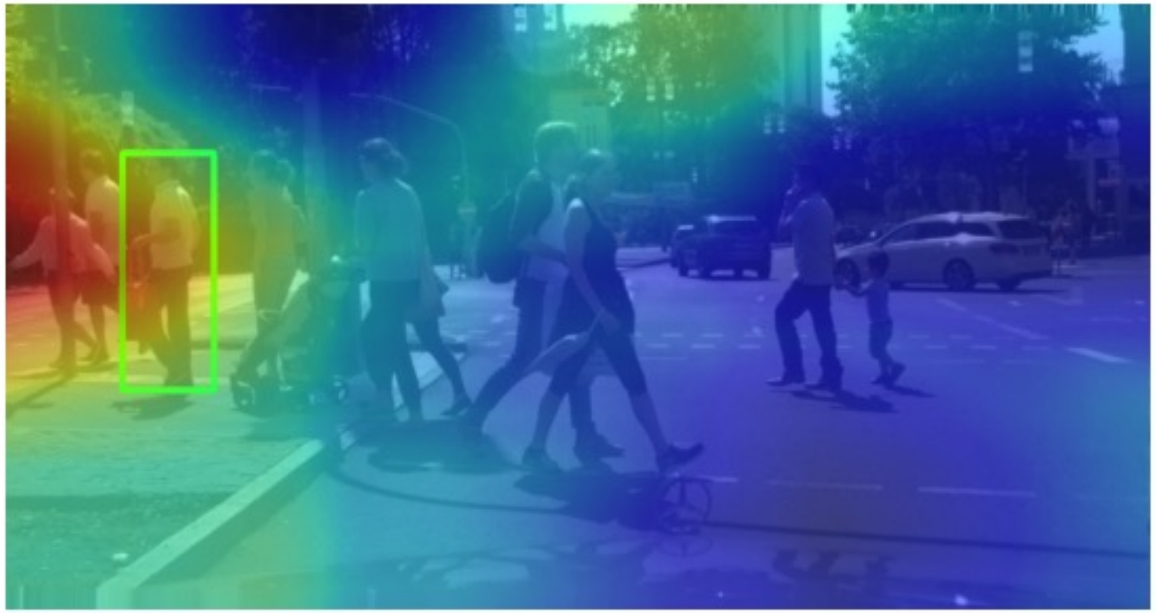
→

Output



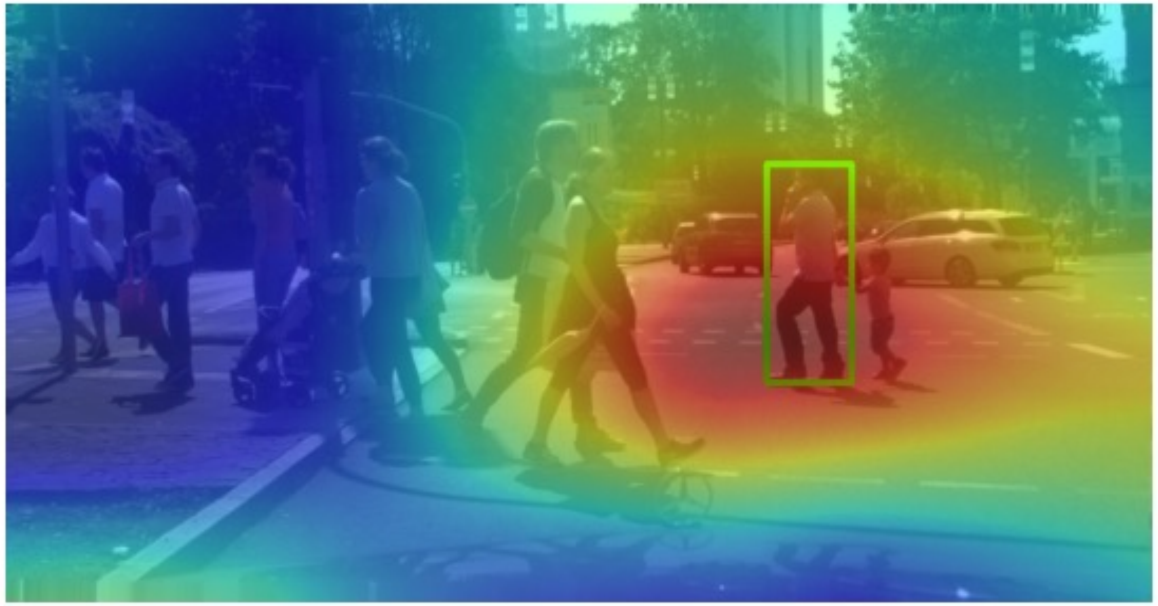
XAI: Pedestrian Detection

- Dataset: ECP
- Importance: Neighboring pedestrians



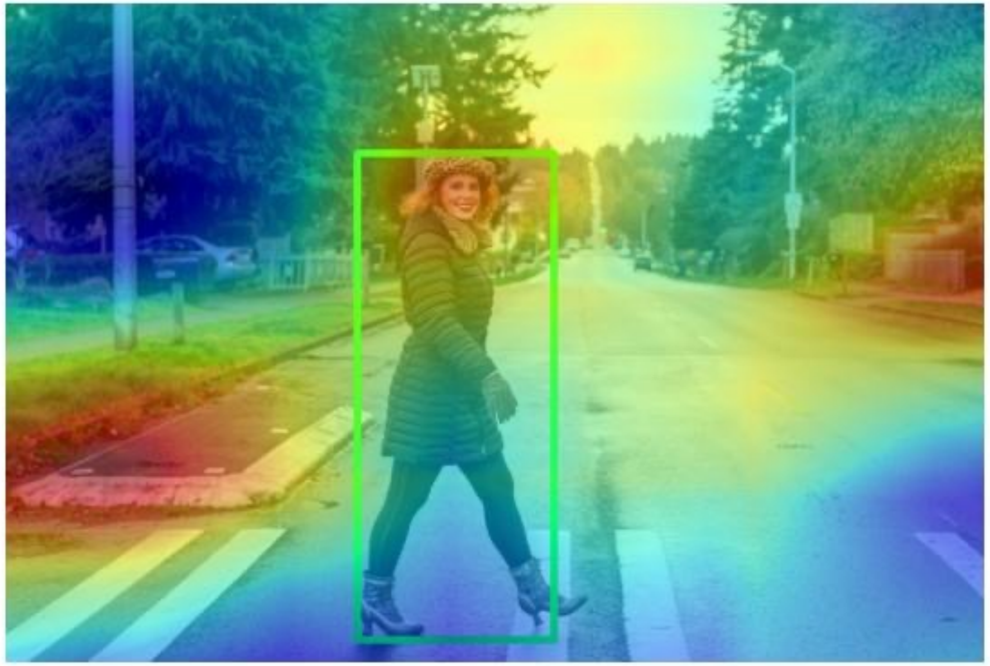
XAI: Pedestrian Detection

- Dataset: ECP
- Importance: Neighboring pedestrians



XAI: Pedestrian Detection

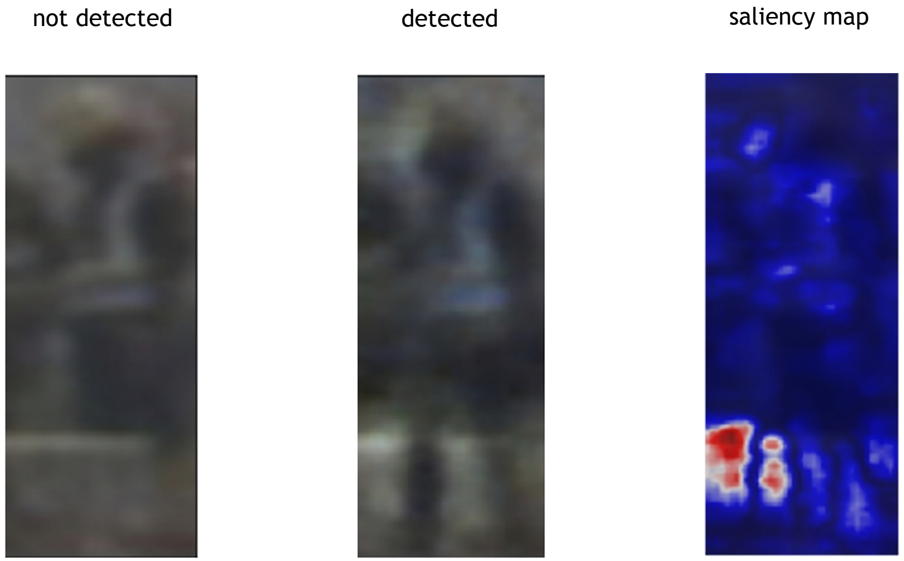
- Importance: Network giving more importance to random places.



XAI: Pedestrian Detection

How can we manipulate the image to make not detected pedestrians visible to the AI?

- Method is called MALALA: Model Agnostic Local Analysis by Latent Attacks
- Based on the latent space representation of a VAE
- New samples are generated:
  - Counter-Exemplar: very similar but the model changes prediction
  - Exemplar: Noticeably different but the model predicts the same class
- Finding unusual but realistic cases on which the model fails is important



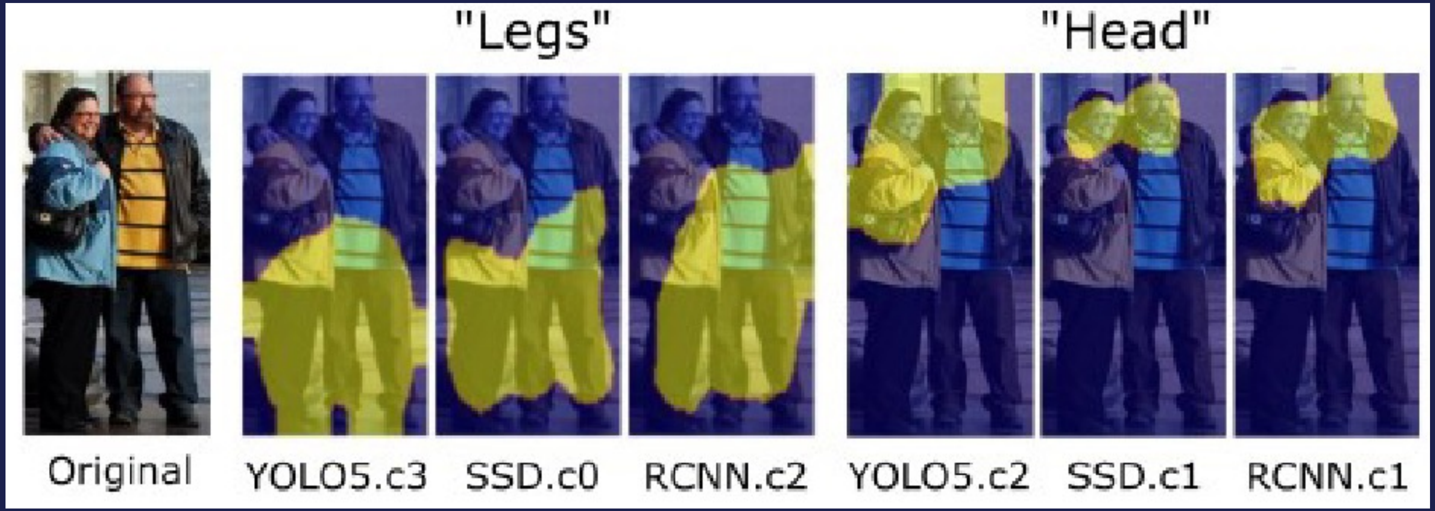
not detected

detected

saliency map

MALALA shows that this occluded pedestrian would be detected, if it had visible legs

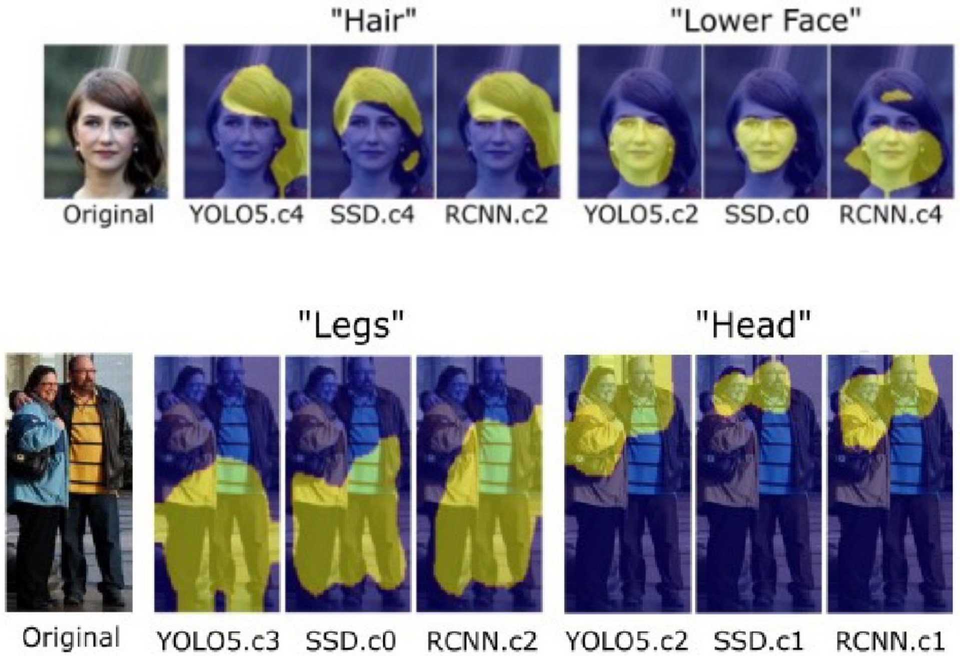
XAI: Concepts



XAI: Concepts

Do different AI models learn similar concepts?

- Analysis of semantic concepts using CAVs
- Concepts correspond to real world objects or notions
- Concept-based comparison of feature space:
    - Same semantic concepts are learned across different architectures
    - Concepts are located at the same relative depth of the feature space

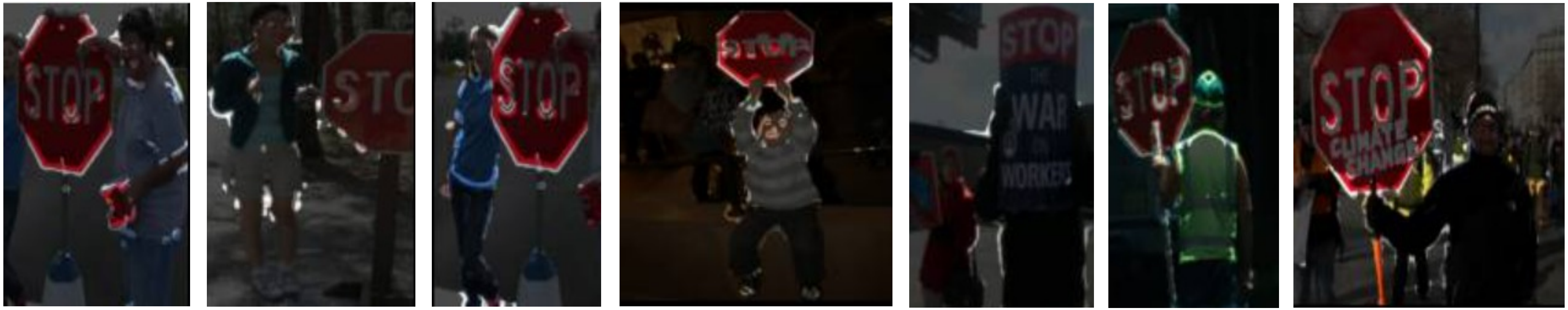


3.4 Poster

3 Highlight

XAI: Concepts

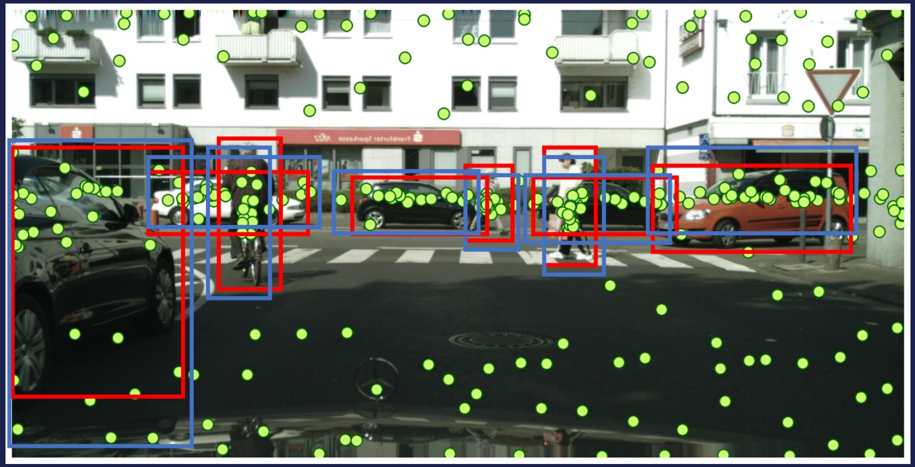
Are semantically unnecessary features used?



- Analyzing the importance of different features for object detection

- *Which object category do you think these images come from?*

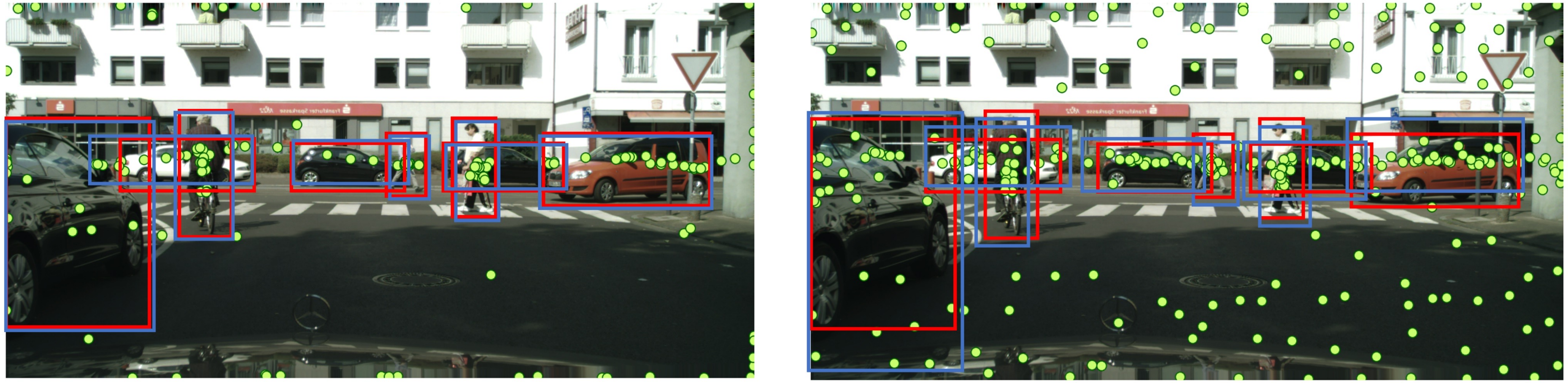
Object Detection



Knowledge-Augmented Object Detection

Can we guide the object detection with prior knowledge?

- Extraction of prior knowledge from synthetic support image patches (contain targeting objects in the input images)
- Integrating this knowledge into a Transformer-based model to enhance the quality of region proposal initialization

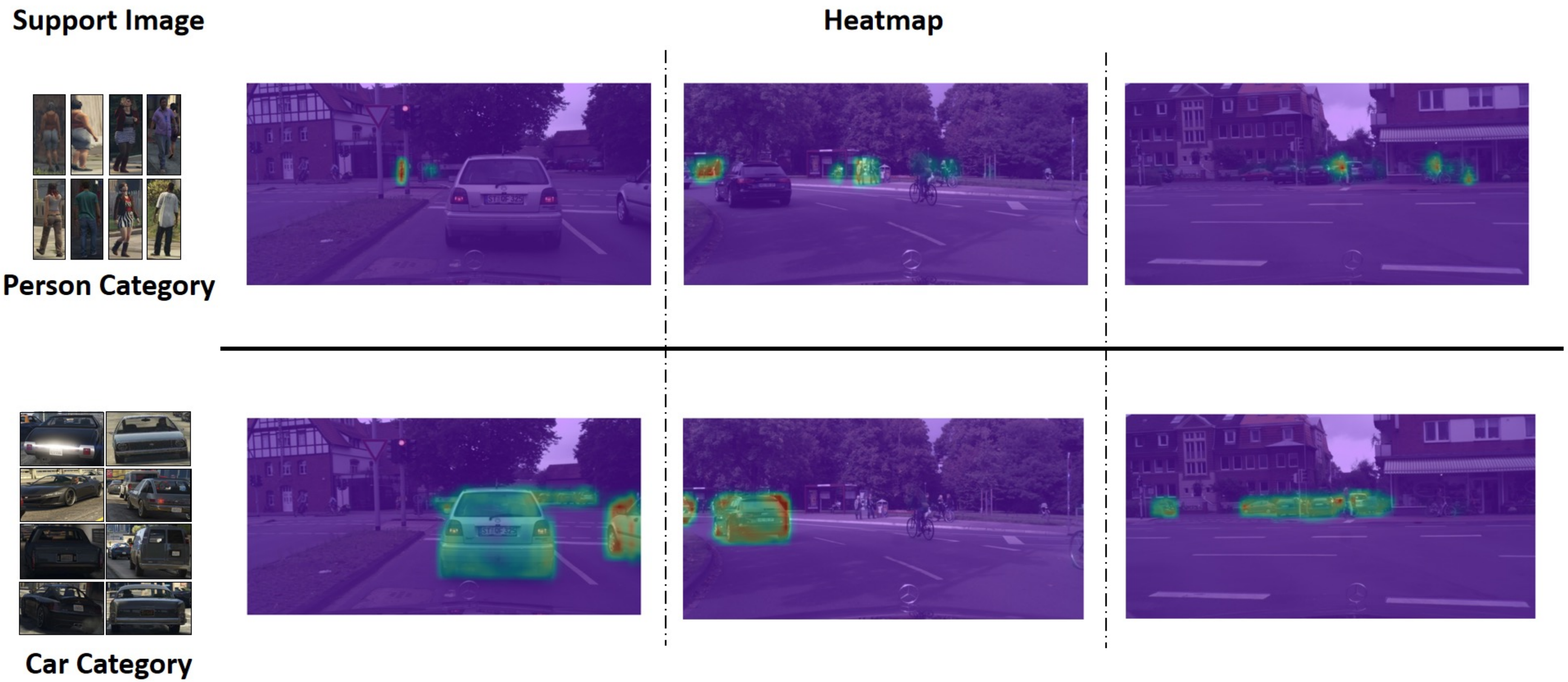


1.1 Poster

12 Highlight

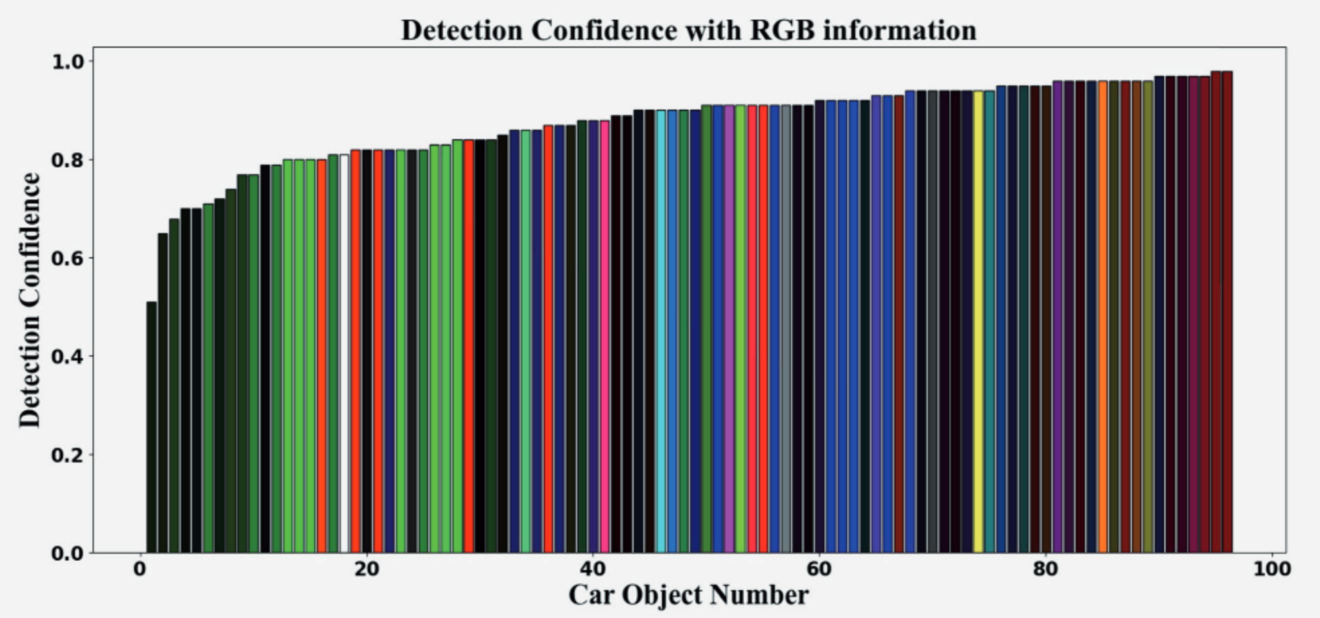
Knowledge-Augmented Object Detection

- Goal: Generate attention heatmaps for the input images using support images as category



Evaluation via Digital Twin Creation

- Exploration of scenarios which lead to failure of the AI system
- Identification of root cause via slight variation
- The detection confidence is related to the color of the car
- The same car model on the same spot can lead to inaccurate or false results depending on its color



1.14
Poster

KI WISSEN

Automotive AI Powered by Knowledge

» How can we ensure certain behaviors of AI systems by integrating knowledge, without hampering the safety in general?

» We can integrate domain knowledge (e.g., physics laws, ethics guidelines) to guide AI decisions and nudge them towards desired behaviors, while rigorous testing and safety measures ensure these additions don't introduce unintended consequences.

Gemini

KI WISSEN

Automotive AI Powered by Knowledge

» Ensuring certain behaviors of AI systems while preserving safety involves meticulous knowledge integration, continuous validation against ethical guidelines, and implementing robust feedback mechanisms to promptly address and rectify potential deviations from intended behavior.

ChatGPT

KI WISSEN

Automotive AI Powered by Knowledge

Antje Loyal | Continental | antje.loyal@continental-corporation.com

KI Wissen is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

www.kiwissen.de @KI_Familie KI Familie

KI FAMILIE

VDA LEITINITIATIVE

Funded by the European Union
NextGenerationEU

Supported by:
Federal Ministry for Economic Affairs and Climate Action

on the basis of a decision by the German Bundestag