KI WISSEN
Automotive AI Powered by Knowledge

KI Wissen Final Event | 21-22 March 2024

Interpretable Model-Agnostic Plausibility Verification for 2D Object Detectors Using Domain-Invariant Concept Bottleneck Models

Mert Keser | Continental

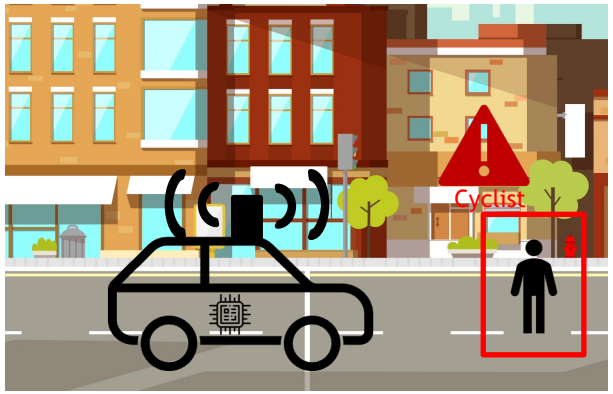
Motivation

Questions:

- How can we check the plausibility of DNN-based perception functions?

Requirements:

- Post-Hoc (Model-Agnostic)
- Operation-Time (Small & Cheap)
- Human-Interpretable
- Robust (In particular, domain invariant)

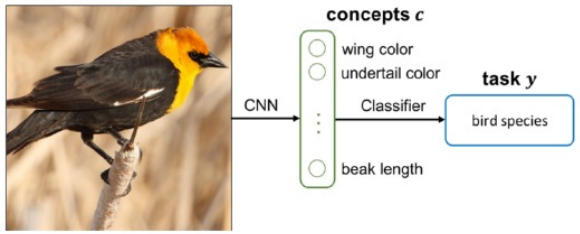


EU AI ACT:

- Under Title VIII „Post-Market Monitoring“ for high-risk AI
- Post-market monitoring requires **error (incident) monitoring** functionality. For this, one has to **identify incidents (implausible behavior) during operation.**

Concept-Bottleneck Models

x -> model - > y



(Koh, Concept Bottleneck Networks, 2020)

- **Challenges:**
  - Requires densely-labeled datasets
- **In Literature:**
  - Significant and extensively studied
    - Belem, Weekly Supervised Multi-Task Learning for Concept-Based Explainability, ICLR, 2021
    - Bento, ConceptDistil: Model Agnostic Distillation of Concept Explanations, ICLR, 2022
    - Sawada, Concept Bottleneck Model With Additional Unsupervised Concepts, IEEEAccess, 2022
    - Yuksekgonul, Poikarinen, Label-Free Concept Bottleneck Models, ICLR, 2023
    - Post-Hoc Concept Bottleneck Models, ICLR, 2023

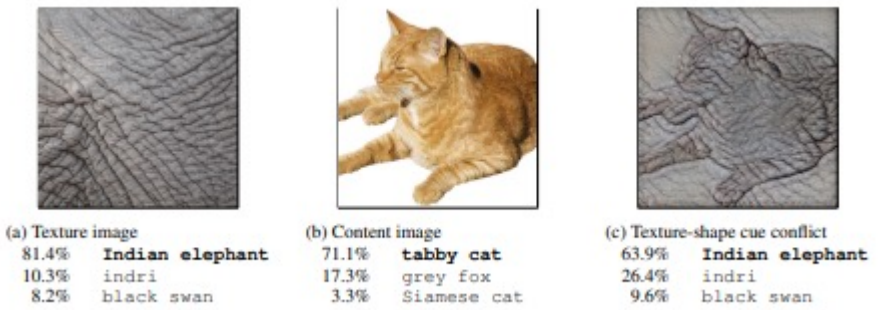
Concept-Bottleneck Models

- **Our Solution:**
  - Transfer Learning

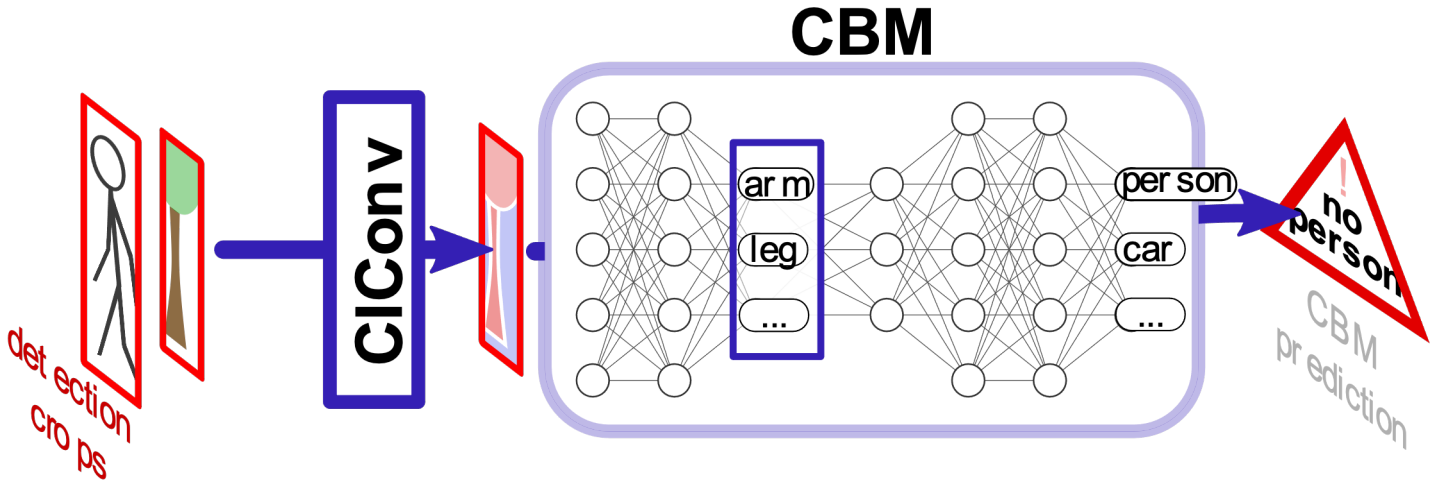
(Broden, ECCV 18)

- **Challenge:**
  - How can it perform well in diverse scenarios?

GIF courtesy of Waymo Open Dataset*

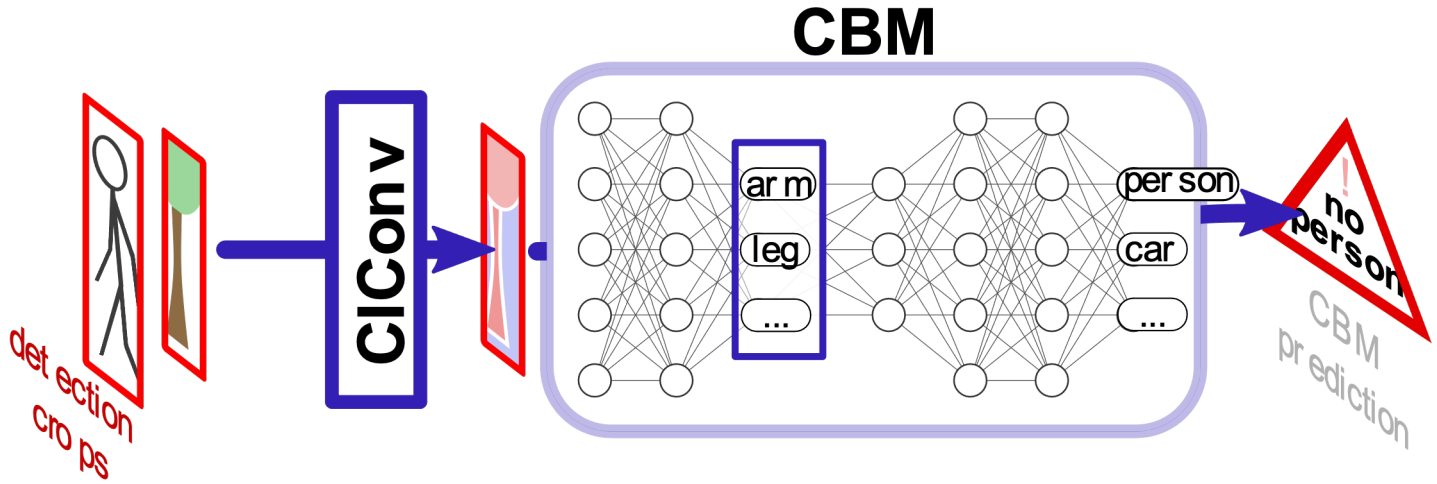


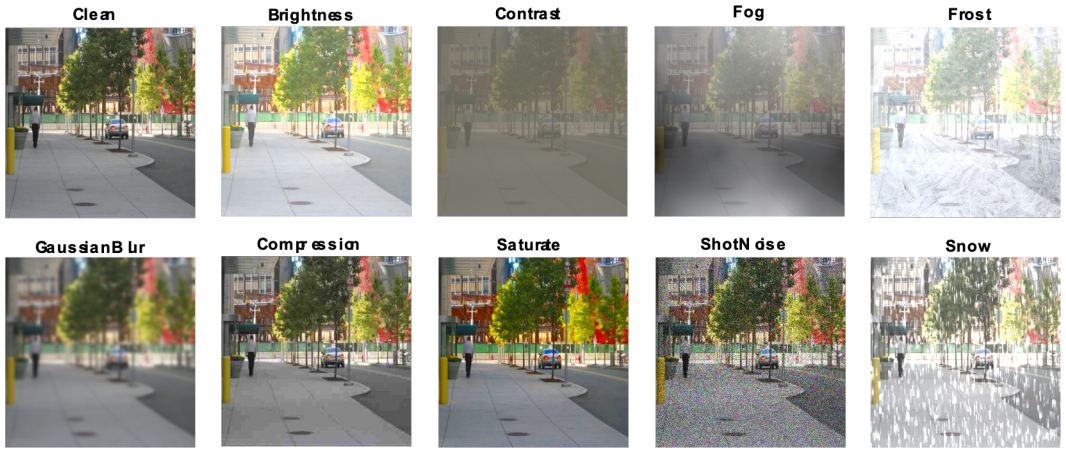
Robust, Generalizable Representations



- Lengyel et. Al., Zero-Shot Day- Night Domain Adaptation with a Physics Prior, ICCV, 2021
- Geirhos et. Al., ImageNet-Trained CNNs are Biased Towards Texture; Increasing Shape Bias Improves Accuracy and Robustness, ICLR, 2019

Learning Robust Concept Representations



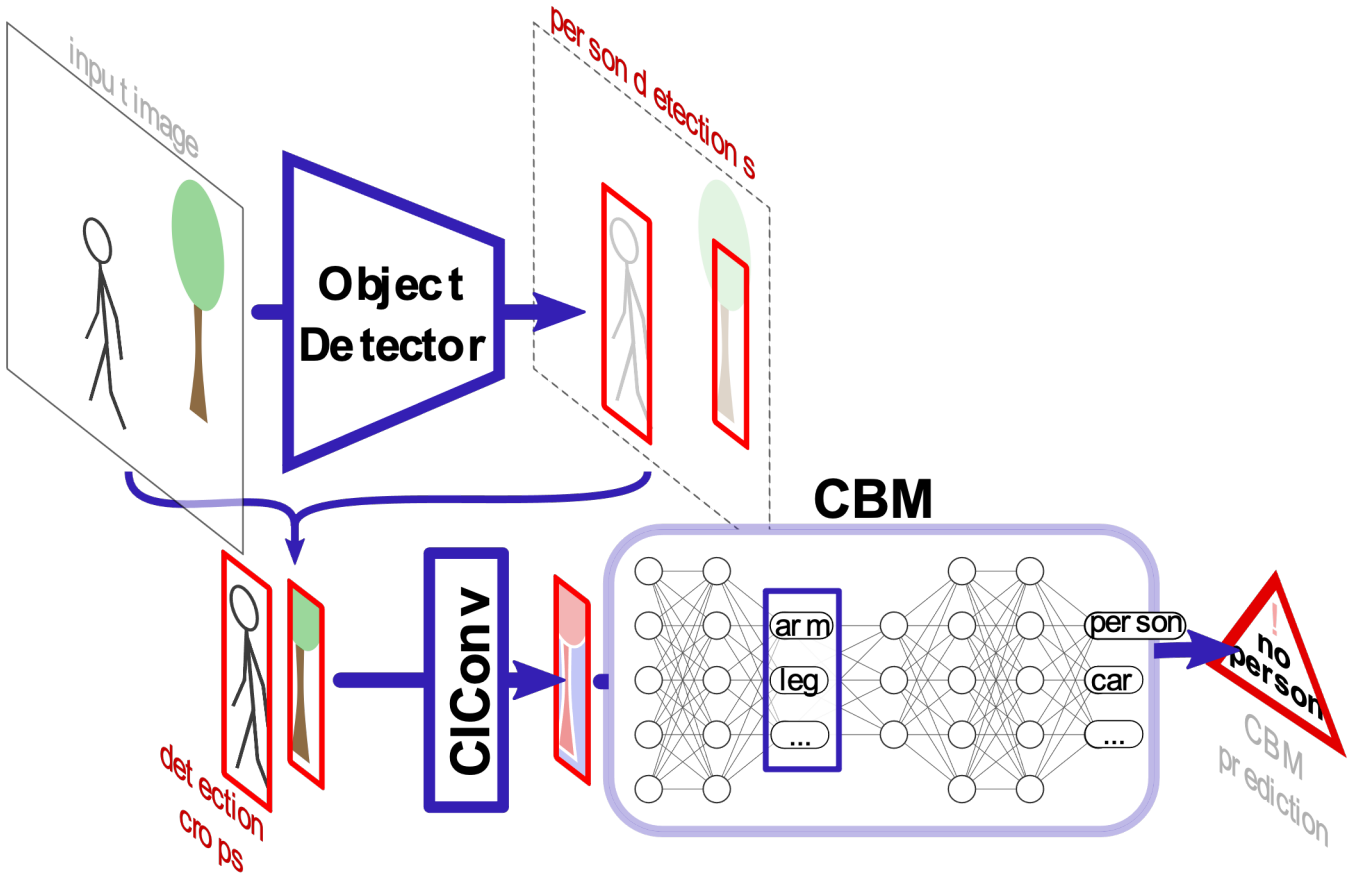


Perturbed Test Dataset of Broden

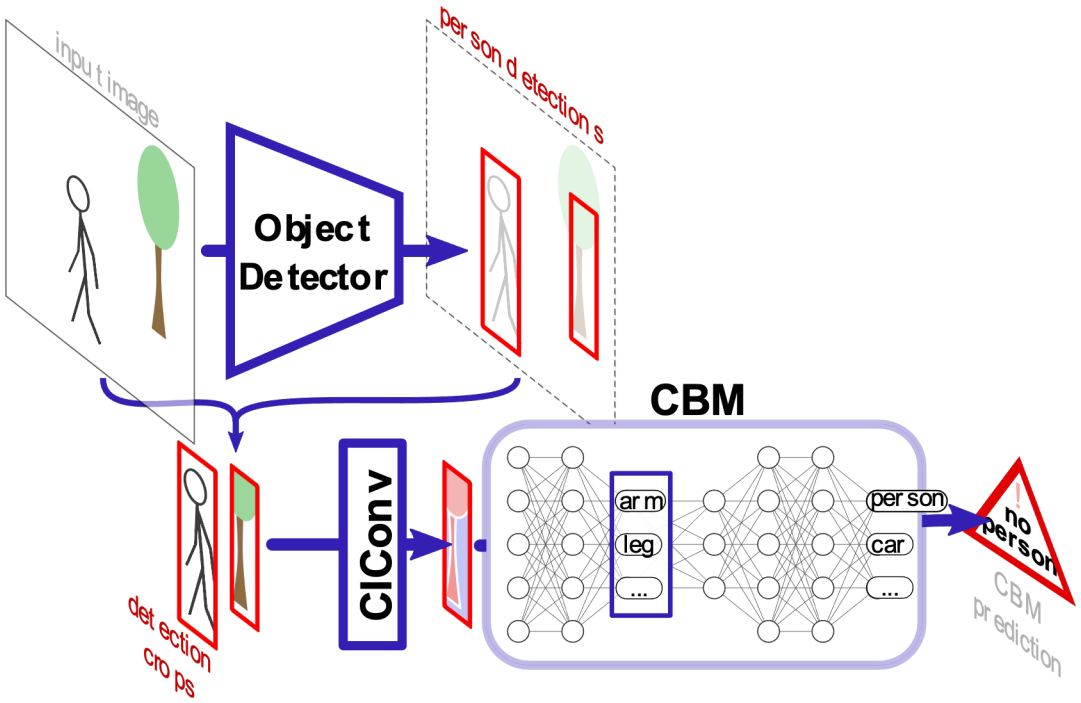
	Person		Car	
Corruptions	CBM	CIConv-CBM	CBM	CIConv-CBM
Clean	**89.3%**	87.4%	**91.4%**	90.3%
Brightness	33.88%	**85.69%**	64.92%	**88.77%**
Contrast	33.03%	**85.60%**	52.17%	**87.46%**
Fog	34.62%	**84.74%**	69.38%	**87.30%**
Frost	34.9%	**74.84%**	69.50%	**80.12%**
Gaussian Blur	35.16%	**75.19%**	70.04%	**81.73%**
Compression	35.06%	**83.84%**	69.74%	**87.55%**
Saturate	34.91%	**85.65%**	68.89%	**89.27%**
Shot Noise	34.27%	**65.53%**	42.13%	**74.55%**
Snow	35.27%	**65.37%**	68.33%	**77.90%**

Table 3. Object class prediction accuracy comparison between vanilla CBM and CBM with CIConv layer on Broden test data with applied corruptions (severity=3). Bold numbers highlight the best prediction performance for each class and corruption type.

FPs Monitoring with Color-Invariant CBMs



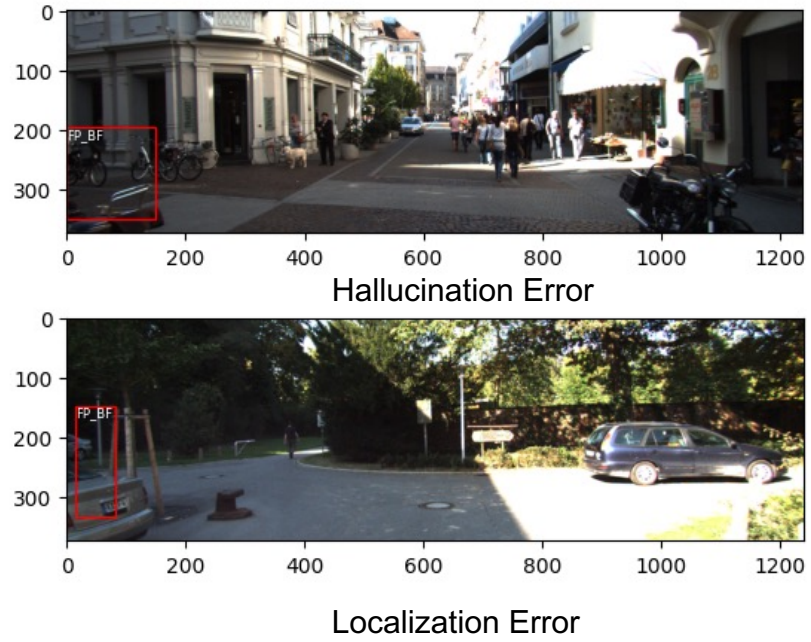
FPs Monitoring with Color-Invariant CBMs



Data, Model	Task	IoU	Precision	Recall
KITTI, SDet	Car	0.7	**0.96**	0.07
KITTI, SDet	(F) Car	0.7	0.81	**0.56**
KITTI, SDet	Ped	0.5	**0.83**	0.01
KITTI, SDet	(F) Ped	0.5	0.72	**0.95**

Table 4. Comparison of fine-tuning (FT) and zero-shot false positive monitoring for SqueezeDet (SDet) on KITTI easy

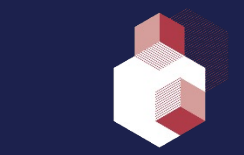
FPs



- Keser, Mert, et al. "Interpretable Model-Agnostic Plausibility Verification for 2D Object Detectors Using Domain-Invariant Concept Bottleneck Models." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

Conclusion & Future Work

- We proposed a novel method for model-agnostic, robust, flexible, and human interpretable operation-time plausibilisation of object detector detections.
- Exploring the potential of this monitoring approach in identifying and addressing other types of object detection errors, particularly focusing on false negatives.



Questions

KI WISSEN

Automotive AI Powered by Knowledge

Mert Keser| Continental AG | mert.keser@continental.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.

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

www.kiwissen.de @KI_Familie KI Familie