KI WISSEN
Automotive AI Powered by Knowledge

3.18 Local Concept-Based Explanations for Object Detectors

Franz Motzkus, Christian Hellert | Continental AG

Formal Methods and Knowledge Representation

Problems with xAI for Object Detection

- Explainability methods are optimized for *classification only*
- *Implementations*/best practices are missing for object detection models
- Pixel-level information is *hard to interpret*

Goals of this Work

- Locally explain object detection models using LRP
- Provide class-wise and instance-wise explanations
- Local-to-global concept and attribution analysis

Locally Explaining Detections

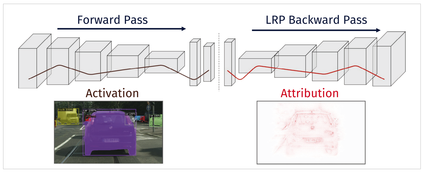

Figure 1: Model forward pass (left) and LRP backward pass (right)

- Local attribution maps show pixel-wise importance scores w.r.t a target
- Layer-Wise Relevance Propagation (LRP)[1] as rule-based backward pass of attribution
- Focus on „what has been used“ by the model w.r.t the target
- Model specific configuration (especially for object detection) required

Figure 2: The multi-dimensional output of object detection models requires for a profound choice on what to explain and how to initialize LRP.

Local-to-Global Approaches

1. Concept-based decomposition with
   a) partitioning into most attributed concepts [2]
   b) testing for a pre-defined semantic concept

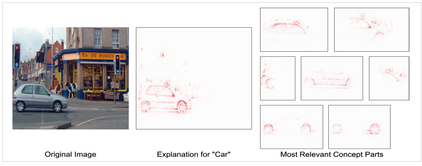

Figure 3: LRP explanation and concept decomposition for the detection of a car by an SSD object detector (© 1. COCO dataset)

Figure 4: Local attribution for globally encoded concept

2. Attribution-based clusterings for extracting detection strategies (Figure 6).
   - Usage of contextual features
   - Usage of false correlations

Application to Automotive

We successfully scaled the methods to a real-world application with examples of the Cityscapes dataset on a MaskRCNN (Figure 5).

References

[1] Bach, et al.: On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation, 2015.
[2] Achtibat, et al.: From "Where" to "What": Towards Human-Understandable Explanations through Concept Relevance Propagation, 2022.
[3] Anders, et al.: Software for Dataset-wide XAI: From Local Explanations to Global Insights with Zennit, CoRelAy, and ViRelAy, 2021.
[4] Lin, et al.: Microsoft COCO: Common Objects in Context, 2014.
[5] Cordts, et al.: The Cityscapes Dataset for Semantic Urban Scene Understanding, 2016.
[6] Liu, et al.: SSD: Single Shot Multibox Detector, 2016.
[7] He, et al.: Mask R-CNN, 2017.

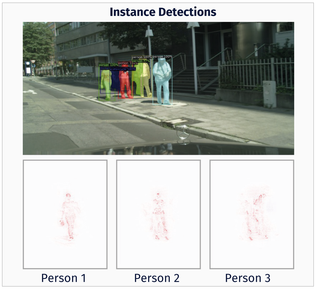

Figure 5: LRP explanations for detections of a MaskRCNN model on the Cityscapes dataset. (© 1. Cityscapes dataset)

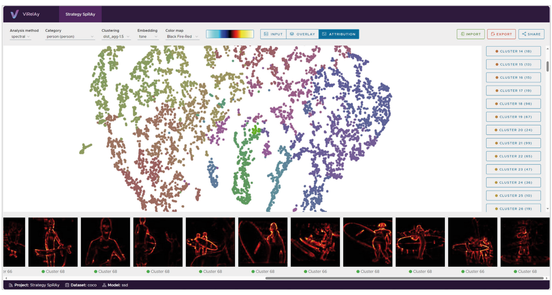

Figure 6: Cluster visualization for detections of class „Person“ in the COCO dataset using the SSD model (depicted in Virelay[3]). Multiple clusters with highly attributed contextual information like „surfboard“ can be found.

Partners

External partners

For more information contact:
franz.walter.motzkus@continental-corporation.com
christian.hellert@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.

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Funded by the European Union NextGenerationEU

www.kiwissen.de @KI_Familie KI Familie