KI WISSEN
Automotive AI Powered by Knowledge

3.4 Knowledge-Based Comparison of CNNs

Georgii Mikriukov, Christian Hellert | Continental AG

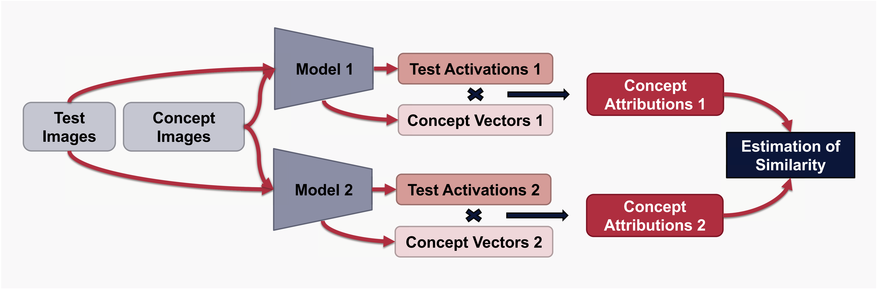

Figure 1: Indirect feature space comparison via semantic concepts and sample attributions

Motivation

- Informed model selection for AI safety.
- Selection of the best model (among several) based on their knowledge.

Problem

- Existing methods foremostly utilize performance or error-estimation metrics.
- Different networks have different architectures and may learn different internal representations.

Solution

- Knowledge representation via *semantic concepts* which correspond to natural language concepts (e.g., head, tyre).
- Abstraction of concepts grows with the depth.
- Concept vector in the feature space is *Concept* Activation Vector (CAV) [1].
- Concept can be available *a priori* [1] or *mined* [2].
- CAVs enable the measurement of *concept attribution* in samples.
- By *measuring concept attributions* in different networks for the same samples, we compare these networks.
- Knowledge may be compared in *pair of layers* or in *whole networks*.
- *Saliency-based* and ranking-based knowledge comparison are proposed.

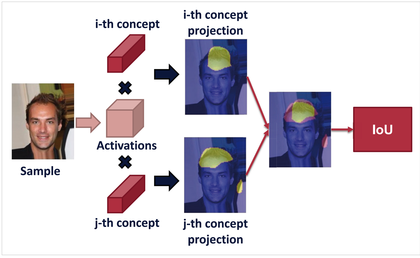

Figure 2: Saliency-based concept similarity estimation

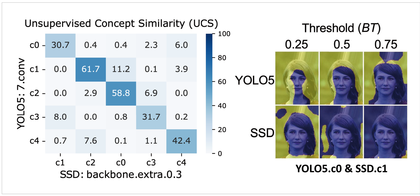

Figure 3: Comparison of mined concepts in layer 7.conv of YOLOv5 and layer backbone.extra.0.3 of SSD

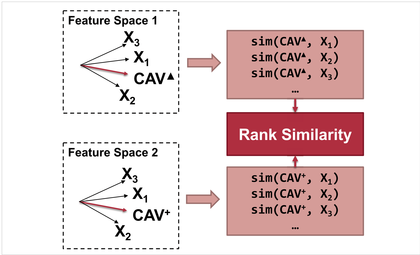

Figure 4: Ranking-based concept similarity estimation

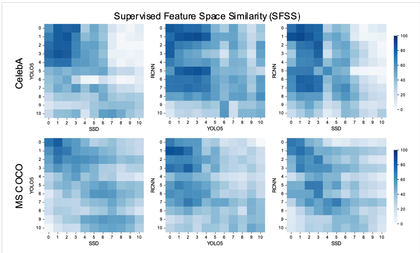

Figure 5: Layer-wise comparison of YOLOv5, RCNN and SSD around a set of concepts related to human body parts

References:

[1] Kim, et al., "Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV)", PMLR, 2018.

[2] Zhang, Ruihan, et al. "Invertible concept-based explanations for cnn models with non-negative concept activation vectors." AAAI, 2021.

Formal Methods and Knowledge Representation

Partners

External partners

For more information contact:
georgii.mikriukov@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