3.15 Consistent Joint Action and Explanation Prediction for Autonomous Driving

Leon Luithlen, Marsida Toska, Michelle Reckstadt, Daniel Kurta, Nikita Volkov, Erwin Rudi | Alexander Thamm GmbH

Formal Methods and Knowledge Representation

Background

Xu et al [1] demonstrated a novel approach for explainable decision making for autonomous driving decisions. The task is formulated as two multilabel classification problems, for which a joint classifier is trained to predict up to four (out of four) actions as „correct" and up to 21 preformulated explanations for those actions. This approach shows promising results [1] , however one drawback of the specific implementation is that it does not guarantee consistent action and reason prediction: reasons for actions that are predicted to **not** be correct can be predicted to be correct. This study investigates four methods of ensuring action-reason consistency.

Increasing Action-Reason Consistency

This study investigates three approaches to guarantee consistent action and explanation predictions.

1. On inference, output explanation logits, and use predicted actions to pick consistent explanations
   a. weak consistency: no explanation predicted if inconsistent with predicted action
   b. strong consistency: predict at least one explanation for each predicted action
2. Predict explanations only and derive actions consistent with the predicted explanations
3. Predict actions and use separate feedforward networks to predict explanations consistent with each predicted action

1) (1.) is a post-processing technique that could be applied to any architecture, but will be evaluated here on the original BDD-OIA architecture.

BDD-OIA-v2

The evaluation of the original architecture and dataset using Grad-CAM led us to question some aspects of the BDD-OIA dataset. Namely, (1) the frame selection from the 5 second clips (2) the presence of frames from the same 40 second scenes of the BDD dataset across training, testing and validation sets (3) incomplete and occasionally inconsistent action and reason labels.

To address each of these points, we developed a new dataset BDD-OIA-v2 that (1) uses the frame one second before the final frame of the 5 second clip, (2) has clean train, validation and test splits, (3) is exhaustively and consistently labelled. It is on this dataset that we evaluate the original architecture and novel variants.

Results

Evaluations show that imposing weak consistency leads to a small increase in precision and a small decrease in recall in explanation classification, while imposing strong consistency shows an increase in recall and a decrease in precision, and an increase in F1 score in explanation classification. The architectures (2.) and (3.) perform worse on all metrics than the strong consistency model, and worse than the base model on all but one metric (explanation recall).

	Action			Explanation		
Model	F1 Score	Recall	Precision	F1 Score	Recall	Precision
Base	**0.700**	**0.723**	**0.678**	0.479	0.376	0.663
Weak Consistency (1a)	**0.700**	**0.723**	**0.678**	0.479	0.374	**0.665**
Strong Consistency (1b)	**0.700**	**0.723**	**0.678**	**0.516**	**0.454**	0.597
Derive Action (2)	0.644	0.628	0.661	0.457	0.384	0.565
Explanation Heads (3)	0.634	0.649	0.620	0.277	0.180	0.605

Figure 2: F1 score, recall and precision for actions and reasons for the base model and the consistent model variants

References

[1] Xu, Yang, Gong et al.: Explainable Object-induced Action Decision for Autonomous Vehicles, 2020

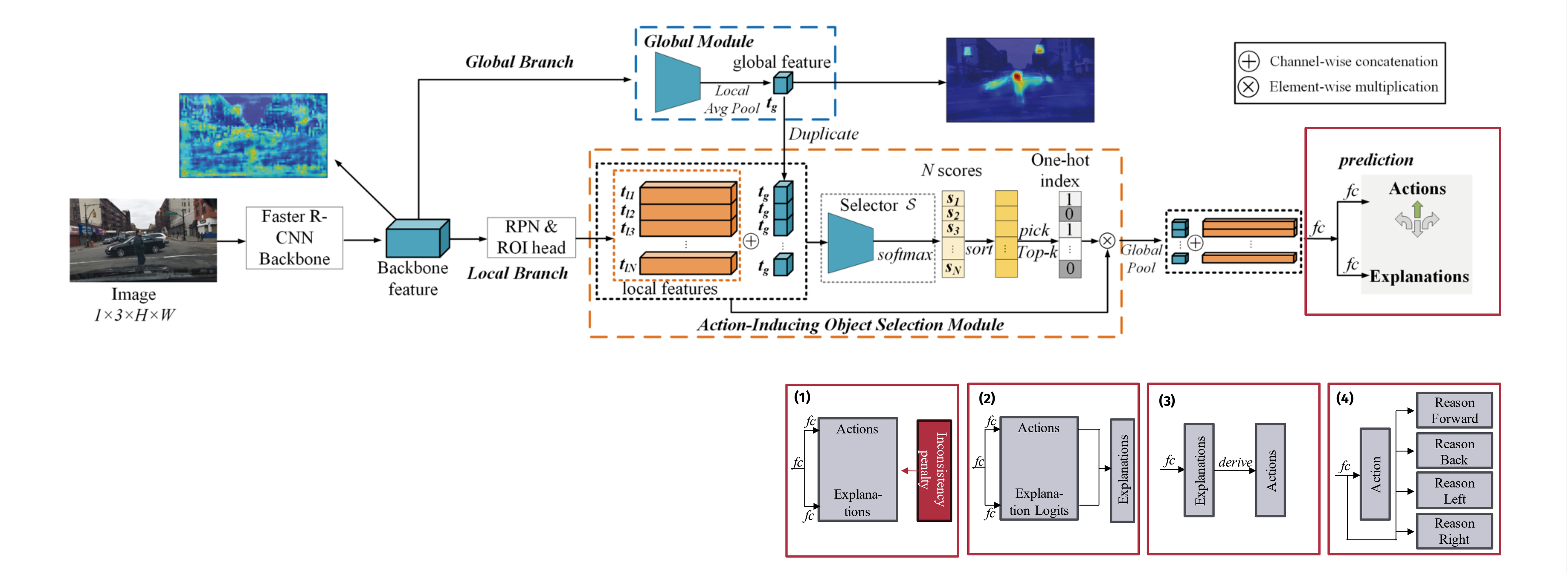

Figure 1: The architecture from BDD-OIA [1], with the final classification module highlighted. The four novel classification approaches are illustrated beneath. They consist in: (1) adding an inconsistency penalty, (2) deriving consistent explanations from the predicted actions and the explanation logits, (3) predicting explanations only and deriving the actions, (4) training a classifier for each set of explanations consistent with one action

Partners

External partners

For more information contact:

leon.luithlen@alexanderthamm.com
daniel.kaulbach@alexanderthamm.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

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