Predictive Coding Based Vision For Autonomous Cars

Roshan Prakash Rane[1], André Ofner[1], Shreyas Gite[2], Sebastian Stober[3]

[1]University of Potsdam, Germany
[2]Kopernikus Automotive GmbH, Germany
[3]Otto-von-Guericke-University Magdeburg, Germany

In recent decades, Predictive Coding has emerged as a unifying theory of human cognition. Related theories in cognitive neuroscience, such as Active Inference and Free Energy Minimization, have demonstrated that Predictive Coding can account for many aspects of human perception and action. However, little work has been done to explore the Predictive Coding framework in the practical domains like computer vision or robotics.

A popular implementation in the field of computer vision that is inspired by Predictive Coding is called the 'PredNet'. PredNet is trained on videos to perform future frame prediction. In a purely perceptual setup like this, Predictive Coding is defined as a hierarchical generative model that dynamically infers low-dimensional causes from high-dimensional perceptual stimuli. The architecture is trained at each level of it's hierarchy to learn low-dimensional causal factors from temporal visual data by actively generating top-down predictions or hypotheses and testing them against bottom-up incoming frames or sensory evidence. In our recent work, we inspected the PredNet architecture and found that it fails to emulate and therefore benefit from many core ideas of Predictive Coding. We will highlight these conceptual limitations of PredNet and present preliminary results from our improved Predictive Coding architecture.

Even though our architecture is inspired by PredNet, it differs from it in three main ways: (1) It is designed to perform semantic segmentation which is an important vision task for autonomous driving. The task is to classify pixels of an image as belonging to a semantic category like drivable road, pedestrian or car (2) The top-down predictions represent semantic class maps and not pixel values and (3) It performs not just short-term but also long-term predictions along its hierarchy.

Finally, we compare our architecture's performance against contemporary deep learning methods for the autonomous driving vision task. We access the semantic segmentation accuracy with an emphasis on the computational efficiency. This includes the model size, amount of training data it needs and the run-time. We also inspect the ability of the model to adjust to differing visual contexts like day time, night time and different weather conditions like rain or snow.