Traffic Risk Assessment: A Two-Stream Approach Using Dynamic-Attention

Corcoran Gary (Patrick), Clark James
Department of Electrical and Computer Engineering
McGill University
Montreal, Quebec
gary.corcoran@mail.mcgill.ca, clark@cim.mcgill.ca

***Abstract*—The problem being addressed in this research is performing traffic risk assessment on visual scenes captured via outward-facing dashcam videos. To perform risk assessment, a two-stream dynamic-attention recurrent convolutional neural architecture is used to provide a categorical risk level for each frame in a given input video sequence. The two-stream approach consists of a spatial stream, which analyses individual video frames and computes high-level appearance features and a temporal stream, which analyses optical flow between adjacent frames and computes high-level motion features. Both spatial and temporal streams are then fed into their respective recurrent neural networks (RNNs) that explicitly models the sequence of features in time. A dynamic-attention mechanism which allows the network to learn to focus on relevant objects in the visual scene is added. These objects are detected by a state-of-the-art object detector and correspond to vehicles, pedestrians, traffic signs, *etc*. The dynamic-attention mechanism not only improves classification performance, but it also provides a method to visualise what the network "sees" when predicting a risk level. This mechanism allows the network to implicitly learn to focus on hazardous objects in the visual scene. Additionally, this research introduces an offline and online model that differ slightly in their implementations. The offline model analyses the complete video sequence and scores a classification accuracy of 84.89%. The online model deals with an infinite stream of data and produces results in near real-time (7 frames-per-second); however, it suffers from a slight decrease in classification accuracy (79.90%).**

***Keywords*-Traffic risk assessment, dynamic-attention, recurrent neural network, two-stream, spatial stream, temporal stream, optical flow**

I. Introduction

Driving any vehicle can be a difficult task even for the most experienced drivers. This difficulty is shown by the high number of collisions in Canada alone. In 2015, there were $118,404$ collisions that were either fatal or involved a personal injury [2]. In about 84% of these accidents the cause could be traced back to driver error [7]. Although the majority of vehicles are currently equipped with passive safety systems, *i.e.* systems to help reduce the outcome of an accident such as seat belts, airbags, *etc*., there are still a high number of serious incidents. Newer intelligent car models are becoming equipped with active safety systems that utilize an understanding of the vehicle's state to avoid and minimize the effects of a crash. Some of these systems include collision warning, adaptive cruise control, automatic braking, *etc*. Research into these active safety systems have expanded into applications that work with or for the driver. This new generation of advanced driver-assistance systems go beyond automated control systems by attempting to work in combination with the driver. These advanced safety systems include predicting driver intent [22], warning drivers of lane departures [15], *etc*. Additionally, one of the most recent trends in the automotive industry is the emergence of autonomous vehicles. In North America, especially in Canada, development and production of autonomous vehicles are growing at a rapid pace. These vehicles are integrated with active safety systems to enhance vehicle safety and reduce road accidents.

Active safety systems have many benefits; however, they are often difficult to implement as they require knowledge about the driver, the vehicle, and the environment. To address this problem, various research papers have attempted to gather information by utilizing multiple cameras, sensors, GPS locations, vehicle trajectories, and the list continues [10, 28, 18, 21]. Although many of these systems can provide a broad understanding of the driver and their surroundings, they require a difficult installation and calibration process. To this end, the proposed research aims to gather knowledge related to driver safety from a single outward-facing dashcam. To accomplish this, dashcam videos are processed by a two-stream dyanmic-attention recurrent convolutional architecture to produce a label corresponding to the perceived risk level in each visual scene. The risk levels are divided into four categories:

1) Low risk: visual scenes which do not include any hazards, resulting in a little-to-no probability of an incident (ideal driving situation).
2) Moderate risk: visual scenes which include hazards with a low-to-medium probability to cause an incident (normal driving situations).
3) High risk: visual scenes which include hazards with a high probability to cause an incident (unsafe driving situations).
4) Critical risk: visual scenes which include hazards with an almost certain probability to cause an incident (impending disaster).

Additionally, because of their novelty, self-driving vehi-