The ability to forecast behaviors is essential for systems, such as self-driving cars, designed to interact dynamically with humans. A self-driving car must understand how other agents in its field of view will: 1) act, and 2) react to the car’s own movements. It is insufficient to recognize an action or behavior as it is occurring; the behavior must also be anticipated. For example, consider a car driving through a neighborhood. Several meters ahead, a child plays basketball in her driveway. The rebound of her shot hits the front of the rim and launches toward the street. Without hesitating, she leaps out into the street to grab her ball.

A system which is only capable of recognizing a child playing or running could not understand the potential danger of the situation. Whereas, an experienced driver would see a child playing with a ball in a driveway, and know that, among the possible futures, the child and/or the ball may end up in the street directly in the path of the vehicle.

Based on additional, personalized context, the driver may even adjust their prior for how the child, the ball, and the street may interact. An older child may be more likely to throw the ball harder, increasing the likelihood of the ball landing in the street. But, the older child may be more likely to know better than to run into the street. Whereas, a younger child may be more likely to have a supervising parent, who may stop the child from chasing the ball into the street.

Learning to forecast directly from visual input drives this research theme. Over the past year, we have focused on developing intuitions within the sports domain, a more constrained setting than the driving domain. Sports has a pre-defined number of agents who interact for a specified amount of time in a relatively small space. Whereas, the number of agents, time horizon, and spatial extent of a drive is less well-defined. In recent work, we learn to predict ball passing behavior with performance almost comparable to non-expert humans. The abstract and link to the paper are below:

Abstract: A large number of very popular team sports involve the act of one team trying to score a goal against the other. During this game play, defending players constantly try to predict the next move of the attackers to prevent them from scoring, whereas attackers constantly try to predict the next move of the defenders in order to defy them and score. Such behavior is a prime example of the general human faculty to make predictions about the future and is an important facet of human intelligence. An algorithmic solution to learning a model of the external world from sensory inputs in order to make forecasts is an important unsolved problem. In this work we develop a generic framework for forecasting future events in team sports videos directly from visual inputs. We introduce water polo and basketball datasets towards this end and compare the predictions of the proposed methods against expert and non-expert humans.

Also in progress is work on personalizing the predictions of behaviors, where we try to understand how actor identity and context can contribute to more accurate predictions. Preliminary experiments show noticeable performance gains for predicting multiple agent’s trajectories when we include information about an agent’s identity (personalization) and information about where other agents will go in the future (context).

Over this next year, we plan to continue developing intuitions for understanding how to forecast behaviors. And we plan to apply the techniques we have developed (and continue to develop) to the more general setting of the driving domain, where we focus on predicting lane changing behaviors of other drivers directly from visual input.