ABSTRACT: Forecasting long-term human motion is a challenging task due to the non-linearity, multi-modality and inherent uncertainty in future trajectories. The underlying scene and past motion of agents can provide useful cues to predict their future motion. However, the heterogeneity of the two inputs poses a challenge for learning a joint representation of the scene and past trajectories. To address this challenge, we propose a model based on grid representations to forecast agent trajectories. We represent the past trajectories of agents using binary 2-D grids, and the underlying scene as a RGB birds-eye view (BEV) image, with an agent-centric frame of reference. We encode the scene and past trajectories using convolutional layers and generate trajectory forecasts using a Convolutional LSTM (ConvLSTM) decoder. Results on the publicly available Stanford Drone Dataset (SDD) show that our model outperforms prior approaches and outputs realistic future trajectories that comply with scene structure and past motion.

REFERENCES: