Disentangling Stochastic PDE Dynamics for Unsupervised Video Prediction

Xinheng Wu, Jie Lu, Zheng Yan and Guangquan Zhang

Abstract: Unsupervised video prediction aims to predict future outcomes based on the observed video frames, thus removing the need for supervisory annotations. This research task has been argued as a key component of intelligent decision-making systems, as it presents the potential capacities of modeling the underlying patterns of videos. Essentially, the challenge of video prediction is to effectively model the complex spatiotemporal and often uncertain dynamics of high-dimensional video data. In this context, an appealing way of modeling spatiotemporal dynamics is to explore prior physical knowledge, such as partial differential equations (PDEs). In this article, considering real-world video data as a partly observed stochastic environment, we introduce a new stochastic PDE predictor (SPDE-predictor), which models the spatiotemporal dynamics by approximating a generalized form of PDEs while dealing with the stochasticity. A second contribution is that we disentangle the high-dimensional video prediction into low-level dimensional factors of variations: time-varying stochastic PDE dynamics and time-invariant content factors. Extensive experiments on four various video datasets show that SPDE video prediction model (SPDE-VP) outperforms both deterministic and stochastic state-of-the-art methods. Ablation studies highlight our superiority driven by both PDE dynamics modeling and disentangled representation learning and their relevance in long-term video prediction.

All authors are with the Decision Systems and e-Service Intelligence Laboratory (DESI), Australian Artificial Intelligence Institute (AAII)