Video generation using deep learning

Deep learning is a subcategory of machine learning, which automates feature extraction using large amounts of data to train complex deep neural networks. In recent years, deep learning has quickly become one of the most sought-after areas in the field of computer science and, due to its effectiveness, has helped to achieve breakthrough results in many different research areas, e.g. computer vision, satellite and medical image analysis, sound analysis, speech recognition, natural language processing, decision making by autonomous systems, and others.

Generative Adversarial Networks (GANs) are powerful deep learning models capable of generating realistic, but synthetic data. GANs have witnessed increasing attention due to their ability to model complex data distributions, which allows them to generate realistic images, as well as to translate images. While realistic video generation is the natural sequel, it is substantially more challenging w.r.t. complexity and computation, associated with the simultaneous modelling of appearance, as well as motion.

Specifically, in inferring and modelling the distribution of human videos, generative models face three main challenges: (a) generating uncertain motion, (b) retaining human appearance throughout the generated video, as well as (c) modelling spatio-temporal consistency. Finding suitable representation learning methods, which are able to address these challenges, is critical to the final visual quality and plausibility of the rendered novel video sequences.

This project proposal focuses on video generation and consists of 3 main research topics. Firstly, we are going to focus on the highly intricate problem of video generation without prior knowledge w.r.t. either appearance or motion, and on learning a disentangled representation of the video generative factors appearance and motion, allowing for manipulation of both. A disentangled representation has been defined here as one, where single latent units are sensitive to changes in single generative factors, while being relatively invariant to changes in other factors. Secondly, we are going to work on conditional video generation to allow for manipulation of generated video content. Finally, we are going to focus on probabilistic video generation, which is particularly important for video forecasting.