CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday - January 28, 2022

Unsupervised multimodal data-assimilation for AI-accelerated high-throughput experimentation

Nathaniel Albert Trask, Sandia National Laboratories

The past year has seen a surge of interest in multimodal machine learning, whereby data of different types describing a given event is assimilated in a manner exploiting synergistic information. Typically, this is some combination of video, audio, and text data pertaining to computer vision applications. However, in high-throughput scientific experimentation where robotics are used to perform massive numbers of scientific experiments, typically there is a different type of multimodal data: high-fidelity microscopy measurements are often accompanied by a plethora of unutilized 1D signals characterizing e.g. acoustics, spectroscopy, or electrical properties. We present a variational inference framework allowing identification of a latent space shared across all modalities. A novel variational autoencoder enables discovery of conditional probabilities embedding the multimodal data into a Gaussian mixture. Sampling of this model provides a means of generative modeling, so that one may synthesize high-fidelity measurements from simple signals amenable to high-throughput testing; one may generatively sample from the distribution of a high-cost experiment conditioned on a low-cost measurement. This framework provides exciting possibilities to incorporate physics-informed machine learning into experimentation; we share an overview of a new SNL project applying these tools to additive manufacturing processes.