Tutorial 11: Deep Learning For Image and Video Processing

The goal of this tutorial is twofold: i) to briefly review the fundamentals of popular deep network architectures and approaches for training them, and ii) introduce successful deep learning methods in image and video processing, mainly in image/video restoration, superresolution and compression. Background in signal processing concepts is sufficient to follow the tutorial. Some familiarity with neural networks is beneficial but not necessary. This tutorial is organized into three parts:

PART 1: Fundamentals of Deep Learning
The first session will introduce the fundamentals of deep networks and learning, including

- The universal approximation theorem – function approximation using single- or multi-layer neural networks
- Convolutional neural networks (CNN/ConvNet) and the classic backpropagation+gradient descent method for training them via supervised learning
- Recurrent neural networks (RNN) and variations (e.g., LSTM and GRU) for modeling temporal dynamics, and back-propagation in-time method for training them
- Multi-scale nets and auto-encoder architecture
- Adversarial nets and applications, how do we compute adversarial loss
- Popular deep learning frameworks, including PyTorch and TensorFlow

PART 2: Learned Image Restoration and Single-Image Superresolution (SISR)
Traditional human-engineered filters perform inversion of a linear model of a blur/down-sampling function using an assumed signal model, such as sparsity, to compute a regularized solution to an ill-posed inverse problem. Using deep learning, we now learn both the degradation (possibly nonlinear) and signal model directly from a large number of "original and degraded image pairs," without the standard linear modelling assumptions. Deep learned models are more general and more accurate than traditional blind restoration methods that estimate a *linear* degradation model from a "*single*" blurred image. Furthermore, results show that deep image restoration methods produce artefact-free inverse solutions.

- Comparison of performance of traditional optimal filtering methods and learned filtering methods – artefact free inverse solutions
- What is adversarial loss and what good is it?
- Comparison of training with mean square loss only vs. combined mean square and adversarial loss
- Discussion of the results of CVPR NTIRE 2017 and 2018 competitions
- Should you prefer CNN or RNN architecture for video processing?

PART 3: Learned Image and Video Compression
Deep learning methods may replace motion-compensation with learned frame prediction and interpolation; transforms such as discrete cosine transform and discrete wavelet transform with auto-encoders that are trained to learn image-dependent transforms or generative codecs that are trained to synthesize decoded images without artefacts. Learned methods are also very effective as pre- and/or post-processing filters. Results show that learned transforms and generative models can be highly effective for image/video compression.

- Can learned frame prediction compete with motion-compensation?
- Should we employ mean square loss only training or combined mean square and adversarial loss training in image/video compression?
- Discussion of the results of the CVPR Challenge on Learned Image Compression 2018

Speaker:

A. Murat Tekalp
Koç University, Istanbul, Turkey