AVSS'19 Tutorial
Deep Learning for Video Compression and Understanding

1. Format: Half-day

2. Title: Deep Learning for Video Compression and Understanding
   Abstract:
   This tutorial aims at reviewing the recent progress in the deep learning based video compression and video understanding. In the past years, deep learning techniques have been successfully applied to a lot vision task. However, for the image/video compression task, traditional non-learning approaches (DCT, Block based motion estimation and motion compensation, etc) are still employed for the mainstream image and video codec. Considering the powerful representation capability, it is possible to improve the image/video compression efficiency further by employing the latest progress in the deep learning. Recently, deep leaning based image and video compression have received significant attention from both academia and industry in the field of computer vision and image/video compression.
   In addition, based on the deep learning techniques, video understanding achieves amazing performance in the past years. In this tutorial, we will introduce the related deep learning techniques (including backbone networks and task specific network architectures) and the corresponding vision tasks (such as video detection or tracking).

3. Motivation and Focus:
   Motivation:
   AVSS is the leading conference for the underlying theory, methods, systems, and applications of video and signal based surveillance. It is obvious that video compression is one of the most important topics in this conference. In fact, video compression plays an important role in the surveillance system. The content in our tutorial covers the latest progress in learning based video compression and may provide the insight for designing efficient video compression system for surveillance systems.
   In addition, video understanding is widely used in the surveillance system (such as detection). The content in our tutorial will review the latest progress in video understanding and introduce the state-of-the-art algorithms for video understanding.
   In summary, our tutorial covers two fundamental techniques in video surveillance: video compression and video understanding and can provide the insights for a lot of attendees.
   Focus:
   In this tutorial, we will introduce the fundamental deep learning techniques, learning based image/video compression and the fundamentals of conventional image and video compression. First, the deep learning techniques have developed rapidly in the past decades. A lot of classic backbone network such as VGG, Resnet, Densenet have been proposed for vision tasks. In addition, some critical techniques such as dropout, BN, residual learning, attention scheme have been proved to be effective for both high level (classification, detection, etc) and low level task (image restoration). These knowledges comprise the basic networks and techniques for image and video compression. Second, for the learning based

image/video compression, how to reduce the redundancy and optimize the network through Rate-distortion trade-off are two critical issues. We will give the several feasible approaches and analyse the advantages of different pipelines. In addition, for the learning based video compression task, how to reduce the temporal redundancy and compress the motion information in an end-to-end manner are critical. In this tutorial, we will introduce the latest work on video compression and provide the more in-depth analysis of the architecture of learning based video compression.

Lecture1 : Fundamental techniques of deep learning, by Wanli Ouyang
In this lecture, we will cover the recent progress in the network design. In order to improve the representation ability, a lot of backbone networks have been proposed to solve the problem. In this lecture, we will analyse the classic network architecture and provide the insight for network design. In addition, we will also analyse the baseline architecture for different task.
In addition, for the video understanding task, we will introduce several benchmark algorithms and analyse the corresponding network architectures. More importantly, we will provide the successful learning based application for video understanding.

Lecture2: Learning based video compression, by Guo Lu
In this lecture, we will discuss the recent works for image compression. First, we will introduce several baseline network architectures for image compression (including both CNN based and RNN based approach). Then we will analyse the techniques that utilized in the learning based image compression (including learning based entropy model, priming and spatial adaptive bitrate allocation, rate-distortion optimization, multiscale image decomposition, adversarial training, importance map and intra prediction). In the end, we will introduce the feasible research topics by combining image compression with other vision tasks.
In addition, we will introduce the learning based video compression. First, we introduce the feasible approach for motion generation and motion compression. Then we will analyse the architecture of learning based method and traditional architecture. In the end, we will analyse the computational complexity of video compression and give some solutions for reducing the model complexity.

Lecture3 : The application of deep learning for video understanding, by Wanli Ouyang
We will introduce several benchmark algorithms and analyse the corresponding network architectures. More importantly, we will introduce the successful learning based application for video understanding.

4. Presenters:

Wanli Ouyang
Senior Lecturer
University of Sydney
wanli.ouyang@sydney.edu.au
https://wlouyang.github.io/

Dr. Wanli Ouyang received the PhD degree in the Department of Electronic Engineering, the Chinese University of Hong Kong. He is now a senior lecturer at the University of Sydney. His research interests include image processing, computer

vision and pattern recognition. He is the first author of 7 papers on TPAMI and IJCV, and has published 70+ papers on top tier conferences like CVPR, ICCV and NIPS. He received the best reviewer award of ICCV. He is a senior member of the IEEE.

Guo Lu
Ph.D candidate
Shanghai Jiao Tong University
luguo2014@sjtu.edu.cn

Guo Lu received his B.S. degree from Ocean University of China, China, in 2014. He is currently pursuing the Ph.D. degree from the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include video compression and processing. He has published several top tier conference and journal papers like CVPR, ECCV and TIP.