Announcing the Final Examination of Amir Mazaheri for the degree of Doctor of Philosophy

Time & Location: April 3, 2020 at 9:00 AM in Virtual Defense Link https://meet.ucf.edu/ctran/FQRPC6KK
Title: VIDEO CONTENT UNDERSTANDING USING TEXT

The rise of the social media and video streaming industry provided us a plethora of videos and their corresponding descriptive information in the form of concepts (words) and textual video captions. Due to the mass amount of available videos and the textual data, today is the best time ever to study the Computer Vision and Machine Learning problems related to videos and text. In this dissertation, we tackle multiple problems associated with the joint understanding of videos and text. We first address the task of multi-concept video retrieval, where the input is a set of words as concepts, and the output is a ranked list of full-length videos. This approach deals with multi-concept input and prolonged length of videos by incorporating multi-latent variables to tie the information within each shot (short clip of a full-video) and across shots. Secondly, we address the problem of video question answering, in which, the task is to answer a question, in the form of Fill-In-the-Blank (FIB), given a video. Answering a question is a task of retrieving a word from a dictionary (all possible words suitable for an answer) based on the input question and video. Following the FIB problem, we introduce a new problem, called Visual Text Correction (VTC), i.e., detecting and replacing an inaccurate word in the textual description of a video. We propose a deep network that can simultaneously detect an inaccuracy in a sentence while benefiting 1D-CNNs/LSTMs to encode short/long term dependencies and fix it by replacing the inaccurate word(s). Finally, as the last part of the dissertation, we propose to tackle the problem of video generation using user input natural language sentences. Our proposed video generation method constructs two distributions out of the input text, corresponding to the first and last frames latent representations. We generate high-fidelity videos by interpolating latent representations and a sequence of CNN based up-pooling blocks.

Major: Computer Science

Educational Career:
Bachelor's of Computer Engineering - Information Technology, BS, 2013, Sharif University of Technology
Master's of Computer Science, MS, 2016, University of Central Florida

Committee in Charge:
Mubarak Shah, Chair, Computer Science
Peter Hancock, University of Central Florida
Lotzi Boloni, University of Central Florida
Fei Liu, University of Central Florida

Approved for distribution by Mubarak Shah, Committee Chair, on January 27, 2020.

The public is welcome to attend.