Workshop#4

Automatic Image Captioning using Transformers and Deep Multimodal Learning

Dr. Ali Nadian Ghomsheh, Ms. Maryam Nazari

Generating automatic captions for the image is a challenging problem in the field of machine vision and a major step towards solving the image understanding problem. In this case, to produce a text description for the image, text, and image need to be processed simultaneously. Such problems where two modes of data are used jointly for learning are considered as multimedia issues. Thanks to high-end graphics processors, it is now possible to train deep neural networks, and with the help of these networks, complex multimodal problems such as image captioning can be handled. In this workshop, we will first introduce the problem of image captioning with the help of deep neural networks. Then we will move to the concept of attention and self-attention mechanisms for improving the image captioning problem. We also look at how transformers can be used in the neural image captioning problem.

Content:
An overview of neural image captioning   - 15 mins
Feature extraction from text and images - 10 mins
A base-line image captioner using Tensorflow and Keras - 30 mins
Attention Mechanism - 10 mins
Self-Attention Mechanism - 15 mins
Image captioning based on Transformer Architectures 20 mins

Ali Nadian-Ghomsheh received the BS degree from Shahid Chamran University in 2005, MSc degree from Shahid Beheshti University Iran in 2007, and later the Ph.D. degree from the same university in 2012. He is currently an assistant prof. in the department of Cyber Space research inst. At Shahid Beheshti University. His current research interests include assistive technologies, vision based human centric systems, and virtual reality.

Maryam Nazari is a master's student at Shahid Beheshti University. She is interested in computer vision and machine learning. Currently, she is completing her thesis on deep neural networks for image captioning. She has received her bachelor's degree in Information technology from Semnan University in 2017.

Skyroom Link: https://www.skyroom.online/ch/mvip2022/workshop

Time and Date (In Iran): 3PM-5PM {22 February 2022 (3th of Esfand 1400)}