COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

MSc Defence

Friday May 3, 2024, at 1PM, online via Zoom (Remote)

Arslan Kazmi

Text to Image Synthesis from Scene Descriptions with a Focus on Relative Positioning

Chair: Dr. Stacey Scott
Advisor: Dr. Andrew Hamilton-Wright
Advisory: Dr. Fei Song
Non-Advisory: Dr. Stefan Kremer

Abstract:

Text-to-Image Synthesis (also known as text to image generation) refers to the process of generating images given some input description or caption. While traditional scene generation tools used databases of objects and programmed rules for placement, modern deep-learning systems can infer details about an image from images and their captions only. Generative Adversarial Network (GAN) deep learning architectures have demonstrated the ability to generate scenes in the form of two-dimensional bitmap images after learning from a set of similar scenes. Models capable of generating detailed images conditioned on complex scene structure within linguistic descriptions but struggle to internalize inter-object relative position information such as “to the right and above” and “close to.”

We attempt to demonstrate that a conditional GAN can be trained to generate 2D bitmap representations of simple 2D scenes with a focus on relative positioning relationships matching those in the input linguistic description.