

READING ASSIGNMENT 6

16785-Integrated intelligence in robotics



1. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs

This paper focuses on synthesizing high resolution photo-realistic images from semantic label maps using conditional Generative Adversarial Networks. It is also successful in allowing object manipulation and category changing. It provides an interface for users to interactively edit the appearance of objects.

This approach builds on the pix2pix framework which was discovered to have unstable training and produce images of poor quality. To improve on photo realism and resolution, coarse to fine generators, multi-scale discriminators and modifications in adversarial loss were included. In order to generate diverse images and allow instance-level control, low-dimensional feature channels were added in addition as the input to the generator.

This approach can be extended to several applications in the field of medicine and biology as the author suggests in conclusion. For example, high quality images are often needed in cases such as endotracheal intubation, laparoscopy and endoscopy. Another improvement on existing framework could be to generate images for training in projects that have sparse data or where the data-set is extremely inadequate. Human system interactions can be greatly improved by building on the framework suggested for interactive editing.

2. Progressive growing of GANS for improved quality, stability and variation

This paper describes a ground-breaking approach in the domain of generative adversarial networks which greatly reduces training time whilst also stabilizing it. Very high quality images are produced. NVIDIA proposed the growth of both the generator and discriminator progressively. New layers will be added later in training to model finer features.

The input of the system is a facial image. The model learns its distribution and applies the learned characteristics to a novel synthesized image. Prior work could not control the specific features a GAN wished to regenerate. However this framework allows controlling effects of individual features without varying the other features simultaneously. It demonstrates extraordinary success with regards to facial features of a human being.

This paper suggests that true photo realism whilst now attainable is still a long way off. A better semantic perception and contextual relationship would be a great improvement in rendering images. Introducing interactive editing may also help especially in scenarios which are data-set specific. The author also suggests that it is imperfect when it comes to microstructures of images and instances of curved roads