Pix2Pix Paper

Dr. Alireza Aghamohammadi

Introduction to Pix2Pix

In this lecture, we discuss the paper titled "Image-to-Image Translation with Conditional Adversarial Networks".
The primary contribution of this paper is to demonstrate that conditional GANs produce reasonable results across a wide variety of problems.

- Synthesizing photos from label maps
- Reconstructing objects from edge maps
- Colorizing images

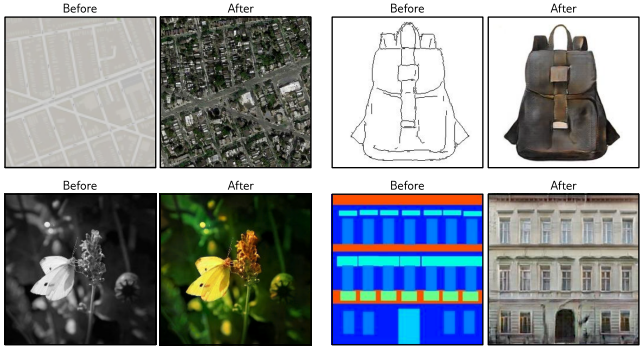

Image credit: Understanding Deep Learning

Pix2Pix Model Architecture

The Pix2Pix model maps one image x to a different style image y. This GAN can be interpreted as supervised learning, which is unconventional.

$$\min_G \max_D V(G,D) = \mathbb{E}_{(x,y)}[\log D(x,y)] + \mathbb{E}_x[\log(1 - D(x, G(x)))] + \lambda \cdot \mathbb{E}_{(x,y)}\left[\|y - G(x)\|_1\right]$$

- The generator aims to fool the discriminator and be close to the ground truth output in an L_1 sense.
- The generator uses a U-Net based architecture.
- The discriminator uses a PatchGAN classifier, classifying each $N \times N$ patch in an image as real or fake.
- The discriminator runs convolutionally across the image, averaging responses to provide the final output of D.
- At the last layer of PatchGAN, each hidden unit indicates whether the region within its receptive field is real or synthesized.
- N can be much smaller than the full image size and still produce high-quality results, making PatchGAN efficient with fewer parameters, faster runtime, and applicability to large images.