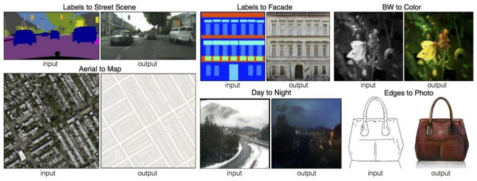
Wasserstein GANs for Image-to-Image Translation

Noah Makow

CS 230, Spring 2018

Introduction

A large number of tasks in computer vision can be represented as the translation from an input to an output image.



Recent research has formalized the task of *image-to-image translation*, or translating from one scene representation to another given sufficient training data. As data and compute has become increasingly available, deep learning approaches are now seeing competitive performance on many of these tasks. In this project we explore the effectiveness of Wasserstein GANs, a recently proposed modification designed to stabilize GANs and improve performance, to the image translation problem.

From GANs to wGANs

We use as our baseline the Pix2Pix architecture which makes use of the vanilla conditional GAN loss functions (with L1 loss):

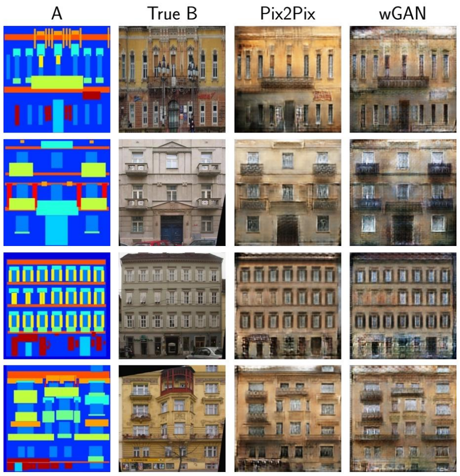
$$\begin{aligned}L_{cGAN}(G,D) =&\mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)]+\\&\mathbb{E}_{x\sim p_{data}(x),z\sim p_z(z)}[\log(1-D(x,G(x,z))]\\L_{L1}(G) =&\mathbb{E}_{x,y\sim p_{data}(x,y),z\sim p_z(z)}[||y-G(x,z)||_1]\\L_{total}(G,D) =&L_{cGAN}(G,D)+\lambda L_{L1}(G)\end{aligned}$$

In contrast, the Wassterstein GAN loss is formulated as (omitting L1 loss below):

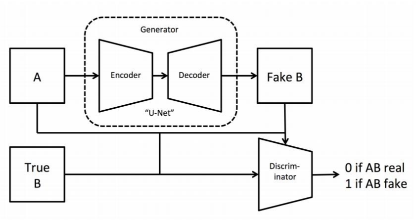
$$\begin{aligned}L_{wGAN}(G) &= -\mathbb{E}_{x\sim p_g(x)}[D(x)]\\L_{wGAN}(D) &= \mathbb{E}_{x\sim p_g(x)}[D(x)]-\mathbb{E}_{x\sim p_r(x)}[D(x)]\\L_{wGAN}(G,D) &= L_{wGAN}(G)+L_{wGAN}(D)\end{aligned}$$

Building Facade Generation

Below we present results after training both the Pix2Pix and wGAN models for 100 epochs:



Model Architecture



Results

Quantitatively measuring the accuracy of generated images is an open challenge in research. We report two metrics, L2 distance (lower is better) and cosine similarity in the VGG16 feature space (higher is better). We compare our models to a naive baseline of randomly generated images and feature vectors:

Model	L2 Distance	VGG Cos-Sim
Random	40990.1	0.15610
Pix2Pix	**25702.1**	**0.46944**
wGAN	25889.5	0.43776

Discussion

- Pix2Pix has the slight edge quantitatively, although the generated samples are not necessarily observably better than wGAN
- wGAN took 3x as long to reach 100 epochs as there is roughly a 5:1 discriminator/generator update ratio compared to a 1:2 ratio in Pix2Pix
- Neither Pix2Pix nor wGAN is particularly effective in generating *diverse* images, other models like BicycleGAN or cVAE-GAN are better
- Empirically the benefits of the wGAN formulation do not seem to be worth the extended training time, ultimately producing samples very similar to Pix2Pix in quality and appearance

Future Work

- Further training of these models would continue to result in higher quality images, however it may be worth exploring other architectures like those mentioned above
- Exploration of how to accurately assess the quality of generated images both in the loss function and in evaluation metrics