MSAN 631 - Deep Learning

1. Team

Harrison Mamin

2. Project Title

Generative Adversarial Networks for Photo and Sketch Generation

3. Background and Motivation

I've been curious about GANs for quite a while but haven't gotten a chance to try building them yet. There are a few reasons for my interest. For one thing, they seem to take a very different approach to the goal of learning than standard neural networks. I'm curious about other ways to utilize multiple models (for example, I wonder if we could use some sort of collaborative "teacher" and "student" networks, or perhaps an ensemble of models that are incentivized to learn different weights from each other). Experimenting with novel architectures sounds much more challenging to jump into immediately, so I think a good start would be trying to implement existing two-model approaches like GANs. The second motivation is simply that I find the task these models take on to be more intriguing. Classification and regression are interesting too, but the idea of teaching a computer to create something new just feels more exciting.

4. Project Objectives

My goal is to implement multiple GAN architectures in PyTorch and try variations of the training process or the models themselves. I hope to become much more comfortable experimenting in PyTorch, rather than just sticking with the standard workflow of funneling data into the same CNN or RNN. I found a dataset of photos and sketches which I believe presents several interesting opportunities. To start with, I will take on the two tasks separately (generating photos from random noise and generating sketches from random noise). The next step would be to see if the models can learn to convert between the two modes. I'm hoping to try out both directions: "teaching the model to draw" (in the case of converting photos to sketches) and rendering photo-realistic images (in the reverse case) both seem interesting. While not the most immediately practical project, I could see this having applications in various design- or media-related contexts. Companies like Autodesk have explored using generative models to augment workflows of human industrial designers, and photo reconstruction from sketches could make for a fun mobile app or even a tool for police sketch artists.

5. Data

The dataset comes from Georgia Tech in the form of the Sketchy Database. The owners selected 125 different categories with 100 photos each, then asked people to sketch each image. There are roughly 6 sketches for each photo in the dataset, making for a total of 12,500 photos and 75,471 sketches. The photos are stored as jpegs and the sketches are stored as png files. I think the quantity should be sufficient (I've certainly worked with smaller datasets in the past, though of course not with GANs), but I've also found a few similar datasets that could be used in addition if necessary. Links to all datasets are below, with the first link being the primary dataset:

http://sketchy.eye.gatech.edu/
http://cybertron.cg.tu-berlin.de/eitz/tvcg_benchmark/index.html
https://www.kaggle.com/vikramtiwari/pix2pix-dataset

6. Techniques Overview

To start with, I will build a deep convolutional GAN (DCGAN) in PyTorch and use that on the two datasets (photos and sketches) separately. I plan to implement at least one other GAN variation as well (hopefully more), which will allow me to compare the results of different approaches. I'm still researching options but cycle GAN will likely be the second approach. Of course, GANs rely on both a generator and a discriminator, so each approach involves building multiple models.

7. Optional Outcomes

There are many different types of GANs, from Wasserstein GANs to Conditional GANs to InfoGANs. Implementing one of these methods (or a similar one) in addition to DCGAN and cycle GAN would be ideal. A less likely but more exciting outcome would be to develop a variation that improves on one of these established models. Given my limited experience and the timeframe, I'm not expecting anything here but hope to try some experiments nonetheless.

As a fun extension, I would also like to try the models on a few different datasets once I have them working. I have some other paired datasets (outlines and photos of shoes), as well as unpaired ones (Leonardo Da Vinci's sketches) which I think would be interesting to train on. These may or may not be included in the final report, however.

8. Evaluation

Evaluation of GANs is not a completely solved problem, so this may be a challenge. We can certainly look at the images and judge them as humans. A simple metric would be to calculate human level performance at discriminating between real and fake images. On a more rigorous level, there are measures such as inception score, MS-SSIM, and precision/recall variants.

Based on the GAN output I've seen from other projects and some of my preliminary results, I suspect the majority of generated samples will be easily distinguishable from the real samples. Ultimately, though, I view this as more of an experimental/research-based project, so my real goal is to better understand these models, develop an ability to build them, and start to gain intuition for training them effectively. I think that will be extremely beneficial going forward, and should help me to explore my own ideas rather than implementing others'.

9. References

The idea for a project involving sketch to image or image to sketch generation is something I've been thinking about since the start of the program. (Actually, my initial hope was to attempt a sketch to 3D model generator, but this project already seemed challenging enough at the moment.) Some sources that I've started to investigate are linked below: the papers for cycle GAN and pix2pix models, a Google blog post on sketch generation (though their methods are somewhat different), a distill.pub roundup of several interesting challenges related to GANs, and a github repo with various models implemented in TensorFlow.

https://arxiv.org/pdf/1703.10593.pdf
https://arxiv.org/pdf/1611.07004.pdf
https://ai.googleblog.com/2017/04/teaching-machines-to-draw.html
https://distill.pub/2019/gan-open-problems/#eval
https://github.com/hwalsuklee/tensorflow-generative-model-collections

10. Schedule

With a 1 person team, delegation of responsibilities will not be an issue. By the end of week 3, I should have all the data downloaded and the initial DCGAN models built. By the end of week 4, I want to have the cycle GAN models done. At that point, I will re-evaluate depending on my progress. The final week can be used for hyperparameter tuning or experimenting with variations in training or architectures. If time allows, week 5 may also provide a chance to try a third GAN variant. I will try to write parts of the paper along the way, but most of it will likely need to wait until the final week to get all my results in.