CS231N Project Proposal

Cedrick Argueta
Department of Computer Science
Stanford University
Stanford, CA 94305
cedrick@cs.stanford.edu

Kevin Wang
Department of Physics
Stanford University
Stanford, CA 94305
kwang98@stanford.edu

Abstract

1 What is the problem that you will be investigating? Why is it interesting?

We are investigating unpaired image-to-image translation with generative adversarial networks. This work was demonstrated in [4]. Work done in the same lab [2] demonstrates how we can use paired images in training to create an image-to-image mapping from one domain to another. The specific image-to-image task that we're aiming to implement is this, which demonstrates a couple examples of sketch-to-photo translations trained using the pix2pix model. We aim to do something similar, using CycleGAN to instead train on unpaired images. The relaxation that unpaired training gives us allows for easier creation of datasets and combinations of domains – we can simply swap a whole domain and retrain rather than find image-to-image pairs for that specific translation task. Our project differs from [**?**] in that we aren't doing multi-modal image translation, just from one domain to another. We are essentially implementing the demo from [2] with an unpaired dataset and architecture matching that of [4]

2 What reading will you examine to provide context and background?

GAN
CycleGAN
BicycleGAN
pix2pix
pix2pix demo
SketchyGAN
SPADE
Google Drawing Guessing Game

3 What method or algorithm are you proposing? If there are existing implementations, will you use them and how? How do you plan to improve or modify such implementations? You don't have to have an exact answer at this point, but you should have a general sense of how you will approach the problem you are working on.

We will use CycleGAN to generate the image-to-image translations. The idea is to get a dataset of sketches of objects (bananas for example) and a dataset of photos of that object. Then we'd train the GAN to be able to produce a photo corresponding to sketch input. We'd use the sketch as input, and then the GAN would produce an image that it believes is in the target domain. The discriminator judges this output, deciding whether the generated image was real or fake. When a threshold has been reached, we have a GAN capable of producing fake images that belong to the target domain that corresponds to the input image. We imagine that much of the difficulty in the project will come from cleaning and putting together a decently sized dataset (either through using existing ones of even creating sketches from photos as in [2]) and also from tuning the model to produce reasonable outputs. Again, we are planning on implementing one of the applications suggested in the CycleGAN paper, where we use unpaired images of sketches and the real object for image translation. As we go through implementations of CycleGAN and learn more about GANs in general, we will try to experiment with our implementation and see if we can improve on it.

Several libraries exist for training GANs and specifically CycleGANs, such as the reference implementation used in the paper, code used in a class at University of Toronto, and code from other researchers.
CycleGAN from Yunjey Choi
UToronto Code
Code from paper

4 What data will you use? If you are collecting new data, how will you do it?

We will try to combine data from the TU-Berlin sketch dataset [1] and ImageNet [3] to create our domains. Because CycleGANs don't require paired images, we think it will be easier to create our dataset than with traditional GANs. We won't need to label the data ourselves, we'll simply pull data that belongs to our domains and use them directly in training. If need be, we can follow the approach of [2] and generate edges from photos to augment our dataset.

5 How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g. plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics or statistical tests)?

As for qualitative evaluation, it will be simple to look at the output for a test image and decide whether it is

- photorealistic
- a faithful representation of the original sketch
- actually belongs to the target domain

Because we are doing unsupervised learning, it is difficult to quantify how well our generator works. However, a few metrics exist:

- Inception Score, comparing conditional label distribution to marginal label distribution for generated samples
- Fréchet Inception Distance, comparing activation distributions of generated samples vs. real samples
- MTurk classification, determining whether real people can distinguish real images vs. fake images
- FCN score, using semantic classifiers to determine whether what the classifier sees is what we wish to generate

References

[1] Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):44:1–44:10, 2012.

[2] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.

[3] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Fei-Fei Li. Imagenet large scale visual recognition challenge. *CoRR*, abs/1409.0575, 2014.

[4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *CoRR*, abs/1703.10593, 2017.