Preliminary Research Topic and Motivation

Sanket Mehrotra (832449213) and Rachit Dalal (832835557)

Research Idea #1

Application: Understanding classification failure in Machine Learning particularly for Road Signs, using the CNNs and GANs.

Models used: Convolutional Neural Networks, Deconvolutional Neural Networks and Deep Convolutional Generative Adversarial Networks (DCGANs).

Implementation

Understanding how and why neural networks can be fooled into misclassifying images is an interesting topic that has been widely explored. This topic has gained much attention of the researchers as the problem is very critical and as we are in the world of self-driving car this kind of failures causes hurdles to bring technologies to the mainstream society. Some papers [5] show that one-pixel change is enough to force a misclassification. This can be done using CNNs or GANs [5], and even uses evolutionary algorithms in some cases [6]. We want to try to this for ourselves and explore some possible reasons to explain this phenomenon. We plan on trying to work with GANs, something outside both of our machine learning experience.

Datasets

1. German Traffic Sign Recognition Benchmark Dataset [2]
   Published by researchers at the Ruhr-Universität Bochum, Germany in 2011 for the International Joint Conference on Neural Networks (IJCNN). The dataset has the following properties:
   - Single-image, multi-class classification problem
   - More than 40 classes
   - More than 50,000 images in total

   The images in this dataset are ~ 32x32 images of road signs. Below are a few samples:

Possible Experiments

1. CNNs:
    a. Train CNN model to high accuracy on part of our dataset
    b. Manually write a program to add random noise to our test dataset
    c. Write a program to add directed/targeted forms of noise on the images: (colored patches, blurring, black rows/cols of pixels.)
    d. ompare the classification accuracy of the classifier with the above types of noise.
2. GANs:
    a. Write a simple Deep Convolutional GAN with the aim of generating images similar to class A, but different enough to fool the discriminator network.

    We are new to GANs, but we will try to experiment with how the inputs to the generator can be more than just random input.

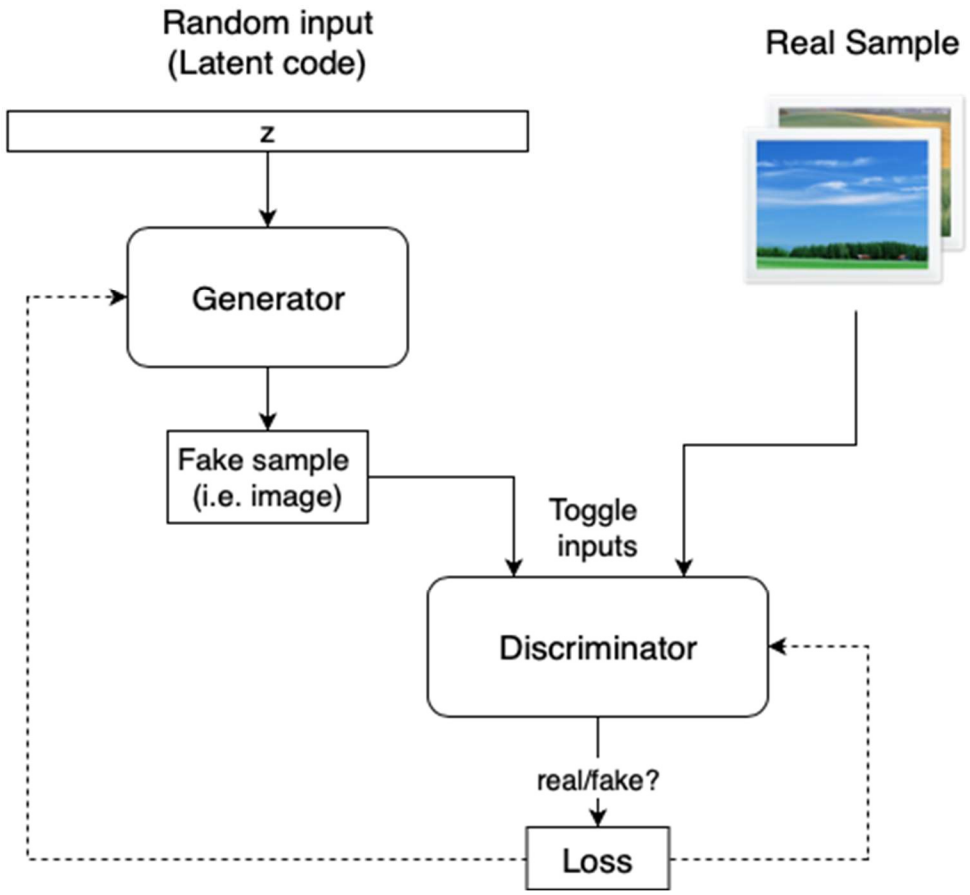

Figure 1: GAN Description: Ref[8]

Qualitative Metrics:

True Positive Rate, False Positive Rate, Precision, Recall, F1 score, SKLearn's Classification report, saliency prediction

Things to focus on:

1. Using this new type of network for the first time (implemented using the tf/keras API).
2. Short and simple implementations of the networks.
3. Hyper parameter tweaking and comparison experiments.
4. Large dataset means that the train-test split must be carefully calculated. It may train the network to perform better but will take a long time and may be prone to overfitting issues.
5. Effect of optimizations to get a small yet well performing tflite model.
6. Well documented code.

Related Work

1. Reference CNN architecture for road sign classification - navoshta/traffic-signs: Building a CNN based traffic signs classifier. (github.com)
2. German Traffic Sign Recognition Benchmark (GTSRB) – Institut fur Neuroinformatik, Ruhr-Universität Bochum, published dataset for IJCNN 2011 Competition. Available at German Traffic Sign Benchmarks (rub.de)
3. Metrics for Multi-Class Classification: An Overview - Margherita Grandini, Enrico Bagli, Giorgio Visani
4. Zeiler, M. D., Krishnan, D., Taylor, G. W., & Fergus, R. (2010). Deconvolutional networks. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2010 (pp. 2528-2535). [5539957] (Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition). https://doi.org/10.1109/CVPR.2010.5539957
5. J. Su, D. V. Vargas and K. Sakurai, "One Pixel Attack for Fooling Deep Neural Networks," in *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 5, pp. 828-841, Oct. 2019, doi: 10.1109/TEVC.2019.2890858.
6. https://github.com/Hyperparticle/one-pixel-attack-keras - Reference code for paper [5]
7. Tensorflow Documentation for DCGANs - https://www.tensorflow.org/tutorials/generative/dcgan
8. https://www.lyrn.ai/2018/12/26/a-style-based-generator-architecture-for-generative-adversarial-networks/

Research Idea #2

Application: Using StyleGANs to try to generate realistic images of room interior design using AI/Deep Learning. For this we've found the large Princeton LSUN dataset used for understanding scenes. This dataset is more recent and was used in the CVPR 2015 and 2016 conferences image classification challenge.

We're interested in trying out new types of AI search and applications in fields we may not have explored yet in our graduate studies so far. NLP and Recommendation Systems are of great interest to us as well and are within the scope of possible project topics.