Photoshopped Image Detection with Deep Neural Networks

Ozi Anyanwu, Thapelo Sebolai
{ozijesus, tsebolai}@stanford.edu

Introduction

Just under 30 years ago, Adobe introduced the first analog photograph into the digital space, changing the way we interact with imagery. This sparked a chain of events to form a world where digital media is so prominent. Today, anybody can interact and edit images and videos for reasons ranging from artistic expression to financial gain. However this increase in the sophistication of image manipulation technology introduced new methods of using technology for malicious purposes.

Consequently, we propose a novel approach of using three methods to detect image manipulation.

Data- PS Battles Dataset

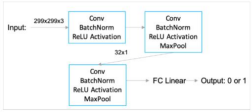
Figure 1: Photoshopped to Original Pair

The PS-Battles dataset, (by Heller et. al) consists of 102,028 images with RGB channels, preprocessed as follows:

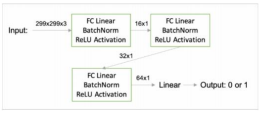
- ~10000 original to photoshopped pairs.
- Center crop to 299x299 as input(224x224 for ResNeT)
- Demeaned according to ImageNet specifics
- Random horizontal flip for data aug. purposes

Models:

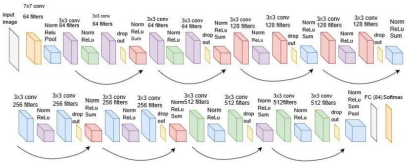
We employed four main models to solve this problem:



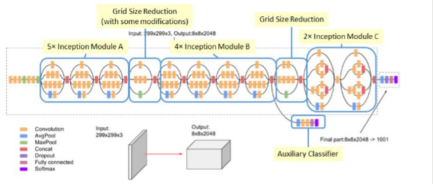
Fully-connected NN trained end-to-end



Convolutional NN trained end-to-end



ResNet18 pre-trained on ImageNet

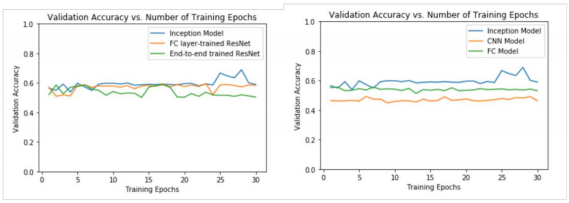


InceptionV3 network pre-trained on ImageNet

$$L_i = -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) \quad \text{or equivalently} \quad L_i = -f_{y_i} + \log\sum_j e^{f_j}$$

Loss Function: Cross Entropy Loss

Results



Model	Accuracy
FC NN	0.4916
CNN	0.5555
ResNeT End-to-End	0.5859
ResNeT Last-Layer Trained	0.5951
InceptionV3	0.6848

The results showcase that the InceptionV3 network performed best. Unfortunately, however it didn't reach the benchmark of 0.7 accuracy. We hypothesize that a possible reason for this was that freezing all the convolutional layers might not have worked in our benefit. However, we discovered that we tended to overfit for our ResNeT model, when trained end-to-end. We did tune on multiple values for the optimizer like step size for the learning rate schedule and learning rate and evaluated our test set on the best performing models.

Future Work

One key aspect that we failed to work on was actively combating over-fitting and under-fitting using methods beyond solutions like dropout.

We would work on devising new ways to extract features from each image, ones that captured the most salient aspects would be a huge step. This could follow a similar paradigm to previous work we researched:
Interpolating two streams, one that extracts low-level features like noise and another that created high level features like RGB values, and passing them through as a feature would have enabled our classifiers to function better.