# Viterbi Internship - Final Work Report

Arka Sadhu
Supervised by: Prof. Ram Nevatia

July 7, 2017

## Contents

# 1 Abstract

# 2 Introduction

The work is done as a part of the MediFor Project. The MediFor project aims at pushing the state of the art research in the field of media forensics which in broad sense deals with the tampering of the media (image, video or audio) and its detection. This work only deals with image forensics. For each manipulated image the MediFor project demands the actual image on which manipulation is done (this is called the baseline image), the kind of manipulation, and in case of splice manipulation where one image is spliced onto another image it also demands the donor image. This work focuses only on the first part, where the aim is to find the baseline image. It is assumed that the world set contains the true baseline image. All experiments are done on Nimble Dataset which is publicly available for use.

# 3 Implementation Details

## 3.1 Datasets Used

The datasets used for this project are Nimble Datasets

| Dataset version | # Probe Images | # World Images | # Provenance Images |
|---|---|---|---|
| NC2016 | 1124 | 874 | - |
| NC2017 Dev1 Beta4 | 515 | 1631 | 65 |
| NC2017 Dev3 Beta1 | 2261 | 4098 | 2157 |
| Self-Generated | 1000 | 4098 | 1000 |

The neural nets used for evaluations are

| Neural Net Used | Dataset Trained on |
|---|---|
| AlexNet | Places365 |
| AlexNet | ImageNet |

## 3.2 Baseline Detection

All experiments have been done on the Nimble Datasets. For the neural nets, the corresponding caffe models are used. All code is written in Python.

The baseline detection problem is essentially finding the base image of the corresponding manipulated image. There can be different types of manipulation. The Nimble 2016 dataset has the following while the Nimble 2017 dataset has the following

For the purpose of baseline detection, two models are used. First is AlexNet trained on Places365 and second is AlexNet trained on ImageNet. The reason for using AlexNet over others like VGG or ResNet is primarily that it requires low computation memory and time. The models used are pre-trained caffemodels found from respective websites. As such a comparison between the Places365 and ImageNet dataset is also shown for the purpose of baseline detection.

- The first step was to understand which layer of the Net should be used.The final layer which is after the Softmax layer is good when the purpose is classification. But it didn't turn out to be as good when used for the purpose of image matching.
- Image matching problem is basically trying to understand wheather or not the two images are actually the same or not. By same we mean wheather the underlying scene is the same or not. This is where the Places365 kicks in because it is scene-centric database. So in cases where a completely new object is placed on the new original image, the ImageNet gives more weight for the new object, while the Places365 doesn't change too much. This is the intuition behind using the Places365 dataset.

- The metrics used are SSD (sum of square distance), SAD (sum of absolute distance), inner product , NCC (pearson's correlation). Of all the metrics the pearson's correlation coefficient turned out to be the most consistent, and hence all further computations are done using this metric only.
- Now onto which layer to choose. This is dictated by emperical results on the datasets (NC2016). The output of the last layer (Softmax layer) is good for classification purposes, but not for image matching. This is especially seen in this image
- On further investigation it is found that the fc8 layer performs the best out of all, i.e. the output of the layer just before the Softmax layer. This is intuitive in the sense that, the deeper down the layers one is, the more semantic features are observed. Also after going through the Softmax layer, it weighs the scenery so that it falls into one of the categories of the dataset. But there is always the chance that a part of the image heavily influences the scene to which it should be categorised, and the Softmax layer gives it a exponential boost in some sense, which is detrimental for the purpose of image matching. From experimentation as well as by intuition it is seen that considering the output of the fc8 layer which is a 365 x 1 vector is the most suitable.
- Interestingly it is also found that for small changes (manipulations less than 10%), the AlexNet trained on ImageNet outperforms AlexNet trained on Places365, but for significant changes (more than 25%) the AlexNet trained on Places365 does significantly better.

## 4 Results

## 5 Discussion