Fashion Forward-Propagation: Machine Learning Tackles Fashion Curation

Hailey LeClair

supervised by
Dr. Andrew GODBOUT

October 31, 2017

1   Introduction

When we see someone wearing a really nice pair of shoes, wouldn't it be nice if we could just take a photo or their shoes, and have some third party tell us where these shoes came from? My project will attempt to do this by using convolutional neural networks trained on specific clothing items that will try to find a match to the photo that is input. Over the past twenty years or so, the fields of computer vision and machine learning have been able to come up with innovative ways to detect and classify many different kinds of objects in images. Most rigid objects like boxes, pens, and even shoes can be detected in an image and a neural network can be trained classify a box as a box, or a shoe as a shoe. This kind of image recognition and classification seems seems to work very well for rigid objects, but what about deformable objects? If I take a photo of a dress on a woman, and want to find an image of the same dress online, what if the women wearing the dresses have completely different body shapes? What if in one photo the dress is blowing in the wind and one is not? This paper will look at these questions, and see if and how they can be answered.

2   Objects in Images

Rigid, articulated, and deformable objects all need to be considered when doing object detection, recognition, or classification in images. A rigid object (or body) is one that where, in terms of physics, "The distance between any two given points on a rigid body remains constant in time regardless of external forces exerted on it "[**?**]. In terms of images, this means that the rigid object will always have the same dimensions, or some translation of the same dimensions in any photo taken of this object. Articulated objects are ones that may have rigid parts, but like human body parts, are connected at a joint which can move[**?**]. Multiple images of the same articulated object may show the object in different formations or positions. Although parts of these objects are rigid, possible movements from the joints, mean that there are many different positions each rigid part of an articulated object could take on. A deformable object is one that "changes its shape and/or volume while being acted upon by any kind of external force"[**?**]. Depending on the degree to which an object deforms, or the amount of force applied, many images of an object could show it as having completely different dimensions and shape.

3   PUT A FIGURE OF ARTICULATED RIGID AND DEFORMABLE OBJECTS HERE

The idea here is to take an image, classify it and find its' match amongst a set of similar photos, which means we need to be able to recognize the object in another image. There exist many methods to recognize rigid objects in an image, like SIFT (Scale Invariant Feature Transform). SIFT can recognize any object that doesn't vary in images based on image rotation, scaling, or translation. It also does well with different illumination and 3D pro-

jection [**?**]. Since rigid objects often have the same shape or dimensions in any image, edge detection methods can also be a good starting point to recognizing an object amongst noise in an image. It is great that that we can use methods like SIFT or edge detection to recognize a rigid object in an image, but what about articulated and deformable objects?

With a few exceptions(shoes, jewelry, accessories, etc.), most clothing items are deformable to a degree. For instance, a pair of skinny jeans may have a general shape on a hanger, but will have different dimensions and possibly even a different shape on different people. Like jeans, many clothing items take on the shape of the person wearing them. We cannot rely on something like SIFT[**?**] to recognize these sorts of items in an image. Usually, in images, clothing items are being worn by someone. This simple fact adds an extra dimension to recognizing the clothing item itself because the human body is an articulated body. Different articulations of the joints in the human body (especially our knees, shoulders, and elbows) can dramatically effect the shape and dimensions of a clothing item in an image. A loose denim jacket, which has a similar shape on anyone who wears it can be easily deformed by placing the arms in a different position. These obstacles specific to clothing items add a layer to object recognition for clothing items that is not present in recognizing many other types of objects in images.

4   PUT A FIGURE OF A CLOTHING ITEM ON TWO DIFFERENT PEOPLE IN DIFFERENT POSES

Along with the problem of objects being deformable and on an articulated body, we also have to consider other parts of the image. An image containing a clothing item could be convoluted with other objects or people in it, or could have an obvious plain background. A method that is able to recognize and classify these objects must account for these degrees of freedom that come with this task. For instance, if we have the same object in multiple images with a white background, a computer may then recognize the white background as a defining feature of this image. When introduced to another image of the same object with a noisy background a computer may then think this is a different object because of the dramatic different in backgrounds amongst this photo and the previous one(s).

5   PUT A PICTURE OF A VERY CONVOLUTED IMAGE HERE

6   Neural Networks

- What is a Neural Network?

- Why are they used

- layers, weights, and activation functions

- Training and testing neural networks

- regularization techniques(not sure if this should go here)

7 Back Propogation

- What is back propogation?

- How does back propogation help to train a Neural network?

- Hyper parameters(Tuneable Parameters) and tuning them

8 TensorFlow

- What is tensorflow

- what is a tensor

- Using Tensorflow to build and train a neural network

9 Convolutions

- What is a convolution (or filter) ?

- How are convolutions useful for image classification and computer vision?

10 Convolutional Neural Networks

- What happens in Biology when we look at things and try to identify them(Local receptive fields, etc.)

- What is a convolutional layer?

- What is a pooling layer/max pooling layer?

- Building and training Convolutional Neural Networks with TensorFlow

11 Image Classification Using Convolutional Neural Networks

- How is Image classification done using convolutional neural networks?

- storing and accessing images (database)

- Using a pretrained CNN to classify an image as a label

- Using the label to get a subset of photos in a database and (possibly) image to image classification using another convolutional neural network

12   Implementation

- What I did/used to solve this problem

13   Results

- What happened when I used, built and trained the convolutional neural networks to classify clothing items

- How did my project do at classifying clothing items

- Problems I encountered, how and why they occurred

- Possible solutions to the problems I encountered

14   Conclusions

- Conclusions made about image classification for clothing using convolutional neural networks