Automated Image Timestamp Inference Using Convolutional Neural Networks

Prafull Sharma
prafull7@stanford.edu

Michel Schoemaker
michel92@stanford.edu

David Pan
napdivad@stanford.edu

Stanford University

Abstract

With the rise in amateur and professional photography, metadata associated to images could be valuable for both users and companies. In particular, the prediction of the time a photograph has been taken is not currently an active area of public research, and vast amounts of accurately labeled data is not available. In this paper, we propose methods to classify the time taken of a picture in two ways: using user submitted tags, namely "morning", "afternoon", "evening" and "night" and four time buckets (i.e. 12 AM to 6 AM, 6 AM to 12 PM, etc.). Among the prediction models used were vanilla SVMs and their variants, along with Convolutional Neural Networks ranging from three layer architectures to deeper networks, namely, AlexNet and VGGNet. The best performing models were the vanilla SVM and the three layer AlexNet (50 and 60 percent accuracy, respectively) suggesting deeper networks that are better equipped to deal with complex features do not necessarily perform better in this particular task.

1. Introduction

Amateur photography has become a novel and popular interest around the world with the emergence of social applications like Flickr, Instagram and Snapchat. These applications allow users to take and upload photographs taken on the spot along with the tags that describe the image. On the other hand, professional photography involves the usage of high-end cameras that record all the information involving location, settings and time taken is also rising in popularity. This information is called the EXIF data which includes ISO, shutter speed, aperture among many other information categories. Perhaps the most insightful and necessary information one can gather from EXIF data is the time of the day when a photograph was taken, as it can allow users to search through their photographs more efficiently as well as provide invaluable data to users and image hosting companies. Most photographs on the internet either don't have the time information or are tagged with the incorrect time.

In this paper, we apply Convolutional Neural Networks to predict when a given input image was taken during the day both using both time windows from EXIF data as well as user submitted time tags. Convolutional Neural Networks have been proven to be efficient in recognizing vision features such as edges, curvatures, corners, etc. We will explore how Convolutional Neural Networks perform in identifying brightness and contrast, among other features, for this task. Specifically, we interpret time in two ways for the purpose of the paper: 1) by tags like morning, afternoon, evening and night and 2) by time bucketing like 00:00 to 6:00, 6:00 to 12:00 etc. The input to our algorithm will be an image which will be classified into a time tag or bucket using SVM and several variants of convolutional neural networks.

2. Previous Work

The premise of inferring time windows or time stamps from images is fairly novel, as such, there is no previous public work to be found that closely aligns to our project. There is however a myriad of related work, such as geospatial location detection. In a collaborative effort between Google and RWTH Aachen University, the publication PlaNet - Photo Geolocation with Convolutional Neural Networks attempted to determine the location where a photo was taken merely from its pixels [1]. The task is straightforward when the image contains a famous landmark or recognizable patterns, but it becomes more complex as abstract features related to weather, markings, architectural details among others need to be inferred. Similarly, we presumed our task would face similar issues, since time detection relies on a variety of factors as well. Their classification approach (they subdivided the earth's surface into thousands of cells) achieved accuracies ranging from $\sim$15% by street ($\sim$1km) to $\sim$42% by country and $\sim$62% by continent. The results improved significantly by introducing LSTMs to solve the problem. Google, however, has access to millions of pictures with extremely accurate location tags, whereas time tags are rarer to find and not very reliable.