Proposal for Forensic Labeling

Sara Mousavi, Rosemary Dabbs, Tasmia Rahman, Zachary Randall

***Abstract*— The objective of the Forensic Labeling project is to improve the efficiency and accuracy of detecting what stage of decomposition a corpse is at in forensic images obtained from the Knoxville Body Farm by adding features to and identifying issues with the current model.**

I. Introduction and Motivation

In this project we are going to automate the labeling process of a forensic dataset including one million images taken from donated corpses to the body farm located in Knoxville Tennessee. To do so, we use a work-flow of web-based, image processing and deep learning approaches. Using web development techniques an online platform is build to facilitate the browsing and tagging process of the images. In the image processing stage, we do cleaning on the images including removing the redundant images and we do image augmentation to increase the size of the labeled data. We use deep learning to automate the feature detection in the unlabeled images and label them. In this proposal we plan on improving the current online platform, generate augmented images and use image processing approaches to clean the data and overcome insufficient labeled images and improve the accuracy of the current model built using Mask-RCNN. Later we would like to see how the weather impacts the decomposition process of the bodies by conducting statistical analytics.

II. Overview of the Dataset

III. Project Tasks

A. Refine Online Platform for Image Tagging and Browsing

The web development tasks for this project involve improving the current online platform that facilitates the manual labeling process for gathering training data. Some of the issues that need to be address include outdated links to directories where platform data is stored, troubleshooting the functionality of the search tags and list tags options on the tags dropdown menu, and manually identify missing labels in already labelled photos. In order to accomplish these tasks, the assigned team member will need to obtain access to the online platform from Dr. Mockus via superuser access and investigate the issues mentioned above. Once a solution has been found, the team member will suggest and/or add Javascript code to implement corrections to the online platform.

B. Analytic tasks : Tasmia

First, I will need to build a mapping of Weather associated to each image prior to the date(weather history for each image). Since I have the name of the images, I will parse the date from image name. Once I have that date, I will extract all the weather history prior to that date and associate the history to each image. Also, each image has several labels. For example, some images have scavenge label. I will extract the images with similar feature and with their weather history, I will try to find a pattern for each labels. With this pattern, we can make a sense out of the data which can tell us which weather features can cause which label. Once being done with that I will collect more weather data from agricultural campus that we can relate to this project. With the data, we may be able to find some patterns within different dataset. Finding this pattern might give us a more precise intuition how different features of weather can affect different labels.

C. Image processing tasks:

The goals of the image processing tasks include matching crime scene images from Google or other sources to an image within our existing collection as well as separating the background from the foreground of an image. The Google images that would be used for matching would be from real crime scenes that have been reviewed by forensic experts. The expert information can then be associated with the matching image in our collection. This will help with identifying photos without the need to bring in an expert as well as provide more data for training the algorithm.The purpose of separating the background and foreground is to make it easier for the program to identify the crucial details required to correctly label a forensic image. Both of these goals are important for success of the application.

D. Literature Review

The goal for the literature review is to have a better understanding of what the best practices are in the industry for improving deep learning environments for image classification. The assigned team member will explore what the different image segmentation methods are. He/she will also learn about the best cleaning, and pre-processing methods available for image-based datasets. Then, the assigned team member will compare image-based feature detections with Convolutional Neural Networks(CNNs). One approach to comparing the different algorithms is to use Big O notation to classify how the algorithms will scale. Some relevant sources have been listed under references at the end of this proposal.

E. Improving the model

The team member responsible for leading the improvement to the model tasks will work on programming with SIFT (Scale-Invariant Feature) in OpenCV to compare with the current CNN model. He/she will also work on increasing

the training data size by using image augmentation. Theano and Tensorflow are both deep learning frameworks. Currently, theano is used for the model. In order to improve the current model, the assigned team member will look at comparing these two frameworks.

IV. Outline of Team Member Responsibilities

Sara Mousavi will be the team lead for this project. She will be responsible for increasing the training data for the model with image augmentation. She will also be responsible for improving the current model. Rosemary Dabbs will be responsible for improving the web-application interface and reviewing literature. Tasmia Rahman will work on the analytic tasks to add weather data to the collection of images. Zachary Randall will work on improving image processing and look into separating the background from the foreground. Each member of the team will participate in the writing portions of the project and presentations.

V. Timeline and Milestones

A. Proposal: September 28, 2018

By September 28th, a written proposal for the project will be completed that outlines the objectives, dataset to be used, tasks of each team member, general approaches, and general timeline to be followed for the project.

B. Checkpoint One: October 12, 2018

Dive into the background reading and plan in detail how to approach assigned problem. Obtain permissions and access to current model and/or relevant material.

C. Checkpoint Two: October 26, 2018

Have code written to execute the tasks given. This is meant as a checkpoint to make sure no one is too stuck. The group can discuss their progress and offer any suggestions or help troubleshoot issues with current code.

D. Checkpoint Three: November 9, 2018

Troubleshoot code of individual members and report current standing results and progress. Outline what else needs to be accomplished before final report is due.

E. Checkpoint Four: November 23, 2018

Complete rough draft of written report and begin putting together a final presentation of results.

F. Checkpoint Five: November 30, 2018

Last checkpoint before the final project is due. Finalize written report and add any additions to the results. Have the PowerPoint slides finished so team members can prepare for the final presentation.

G. Due Date: TBD

VI. What are we expecting to get at the end

The end goal of this project is to produce a web-based application that uses image processing and deep learning methods to automatically label forensic images taken at the body farm located in Knoxville Tennessee. The cost, both money and time, of having a forensic expert review and label each image is too high due to the fact that the dataset includes one million images. The use of a tool to automate the labeling will greatly improve the process by allowing data to be quickly and cheaply reviewed with the same accuracy as a forensic expert. The body farm produces important scientific data on the decomposition of human bodies over time that can be used for many different applications. With the use of the web-based application proposed, the body farm will be able to generate more of the important data that it is designed for.

References

[1] https://www.quora.com/How-does-the-region-proposal-network-RPN-in-Faster-R-CNN-work

[2] Lin, T. Y., Dollr, P., Girshick, R. B., He, K., Hariharan, B., & Belongie, S. J. (2017, July). Feature Pyramid Networks for Object Detection. In CVPR (Vol. 1, No. 2, p. 3).

[3] https://blog.athelas.com/a-brief-history-of-cnns-in-image-segmentation-from-r-cnn-to-mask-r-cnn-34ea83205de4

[4] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).