

The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: DeepPatrol: Finding Illicit Videos for Law Enforcement

Author(s): Marco Alvarez Vega

Document Number: 254636

Date Received: April 2020

Award Number: 2016-MU-CX-K015

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

FINAL SUMMARY REPORT

Federal Agency and Organization Element to Which Report is Submitted:

National Institute for Justice, U.S. Department of Justice

Cooperative Agreement Number: 2016-MU-CX-K015

Project Title: DeepPatrol: Finding Illicit Videos for Law Enforcement

PI Name, Title, and Contact Information:

Marco Alvarez Vega, Assistant Professor

Department of Computer Science and Statistics

University of Rhode Island

9 Greenhouse Road, Suite 2, Kingston, RI 02881

malvarez@cs.uri.edu

Submission Date:

6/28/2019

Recipient Organization:

University of Rhode Island

70 Lower College Road

Kingston, RI 02881

Project/Grant Period (Start Date, End Date):

01/01/2017 to 06/30/2019

2. INTRODUCTION:

This project addresses the problem of automatically detecting child pornography in media files such as images and videos. Through the development of research in machine intelligence/vision this project results in the development of DeepPatrol, an innovative software tool to assist law enforcement agencies in investigating child pornography cases. DeepPatrol will be made public as an open-source project, providing direct benefits to law enforcement organizations.

The proliferation of inexpensive storage devices and mass digital media production has led to the most frenzied growth of child pornography in history. As of 2017, Child Victim Identification Program from the National Center for Missing and Exploited Children (NCMEC) has sent more than 209,000 notifications to service providers regarding publicly accessible websites (URLs) containing suspected images or videos of children being sexually abused.

During the course of an investigation of suspected child pornography, a computer forensics specialist typically spends many hours looking at hundreds of thousands of images and videos. The seizure and further analysis of a suspect's computer and data is tedious, error prone, invades the privacy of the suspect, wears on the investigator, and demands time that the investigator could be using to address the backlog of cases he/she likely faces. Automating the process of searching images and videos on seized media would drastically reduce the amount of time that investigators have to spend looking at suspected files, and would allow investigators to concentrate on other aspects of the case.

In this project we use recent computer vision and machine learning advances to automate the process of identifying Sexually Exploitative Imagery of Children (SEIC) in order to significantly decrease the amount of time law enforcement agents spend on child pornography investigations. Traditional machine learning methods for this task rely on manual feature engineering and tend to generate many false positives, serving only as a coarse filter for suspected material. When a large number of files are present on the suspect's hard drive, the agent reviewing the case may become overwhelmed with imagery falsely flagged as being SEIC, especially when many pornographic images are present. This is because pornographic content is difficult to distinguish from SEIC using traditional techniques. Recently, researchers have found greater success at nudity detection, but these models still struggle at differentiating between pornography and SEIC. To address this issue, we fuse the predictions of these

more accurate deep learning models for nudity detection with apparent age estimation to explicitly identify SEIC material.

This novel approach results in a framework which may be as fine or coarse a filter as the agent specifies, in a way previous approaches cannot. It can distinguish between challenging examples of pornographic and SEIC videos with 89% accuracy using the default thresholds. Since our approach relies on automatic representation learning through the use of convolutional neural networks, those involved in this work never had to be directly exposed to pornographic or SEIC content. Additionally, in contrast to most other works on SEIC content detection, we rigorously evaluate our models on a series of challenging datasets to analyze their performance before presenting results on data collected from real law enforcement cases.

3. APPARENT AGE ESTIMATION:

The development of computational methods for age estimation from face images has been one of the most challenging problems within the field of facial analysis. In addition to common difficulties in facial analysis, such as pose and illumination, age estimation proves a significant challenge due to the subjective nature of the problem. The process of aging is unique to every individual and is influenced by genetics, diet, occupation, and hobbies. This implies that two individuals with the same biological age can have quite different appearances.

Age estimation may be considered from either of two different representations: biological age estimation or apparent age estimation. In biological age estimation, the actual age of a human subject is predicted while in apparent age estimation the label is the aggregation of a group of guesses made by human labelers. This aggregation is usually the arithmetic mean of the collection of guesses. Apparent age estimation is a recent topic, which has received increasing attention, as a result of the deep learning revolution and two apparent age estimation competitions run by ChaLearn in 2015 and 2016. State-of-the-art results are already beating the human reference.

By focusing the attention on the apparent age of individuals, the hopes are to alleviate the subjectivity underlying biological age estimation tasks, since human guesses are expected to agree more on how old a subject looks like.

We developed research to approach the apparent age estimation problem under the framework of label distribution learning (LDL). Unlike classic single-label or multi-label classification, in which instances are assigned to a single or multiple labels, the aim of LDL is to assign instances to label distributions, i.e., vectors containing the probabilities of the instance having each label. Our motivation is essentially to find better ways to model the label ambiguity underlying the apparent age estimation problem.

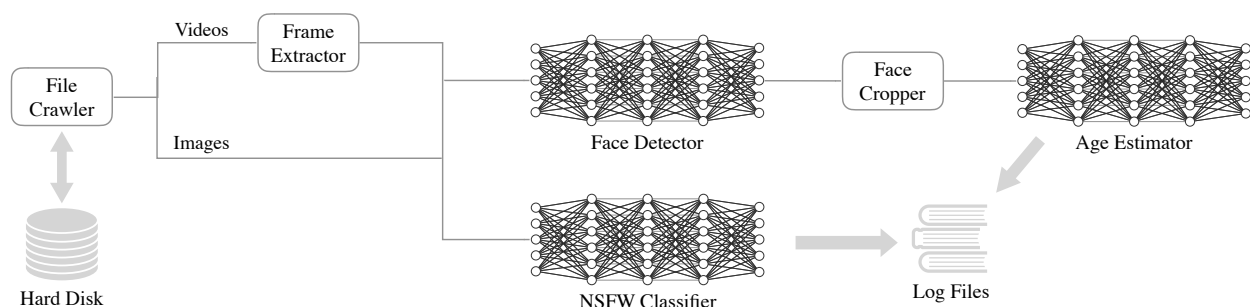
We were able to increase the state-of-the-art for apparent age estimation by proposing an end-to-end framework, based on convolutional neural networks, to learn distributions of apparent age labels. Given an input image of a human face, we produce a discretized probability distribution vector where each value k represents the probability of being k years old.

We published a conference paper (IJCNN 2018) highlighting the following contributions:

- A novel end-to-end framework, based on learning label distributions, that leverages the availability of human guesses in the APPA-Real dataset for modeling the apparent age estimation problem;
- Better performance than state-of-the-art methods on apparent age estimation using the APPA-Real dataset. We improve the mean absolute error to 3.688 years;
- Empirical evidence that pre-training using label distributions yields higher performing models regardless of the target task.

4. DEEPPATROL ARCHITECTURE

DeepPatrol has been developed with the goal of a unified data pipeline in mind. A basic diagram of the approach is provided in the figure below. Components that use deep learning are highlighted in gray. The entire pipeline was efficiently implemented by loading all of the modules in memory and passing batches of images through each step. By moving batches of images through the pipeline, the application is able to process media in real-time, and work with live video.



Processing of image batches starts with the file crawler, which given a starting base directory, identifies every image and video file under that directory. Next, videos are sent to the frame extractor to separate each video into a series of discrete frames to be processed as images. Currently each video is sampled at an arbitrary rate of 1 frame per second and saved to the filesystem as a set of images. We explored more intelligent frame extraction techniques which perform multiple passes on a video to collect more frames from difficult to analyze portions of a video, but we could achieve similar performance with a simpler and more efficient strategy.

Yahoo's OpenNSFW model serves as the base of the pornography detection module. This model is a pre-trained Convolutional Neural Network (CNN) publicly available which achieves satisfactory performance at identifying pornography in images containing SEIC without any fine-tuning. It accepts images as input and outputs the probability of the image containing NSFW content. Preprocessing for this CNN includes converting to an RGB color format, and resizing the image to be 256x256 using a bilinear interpolation method.

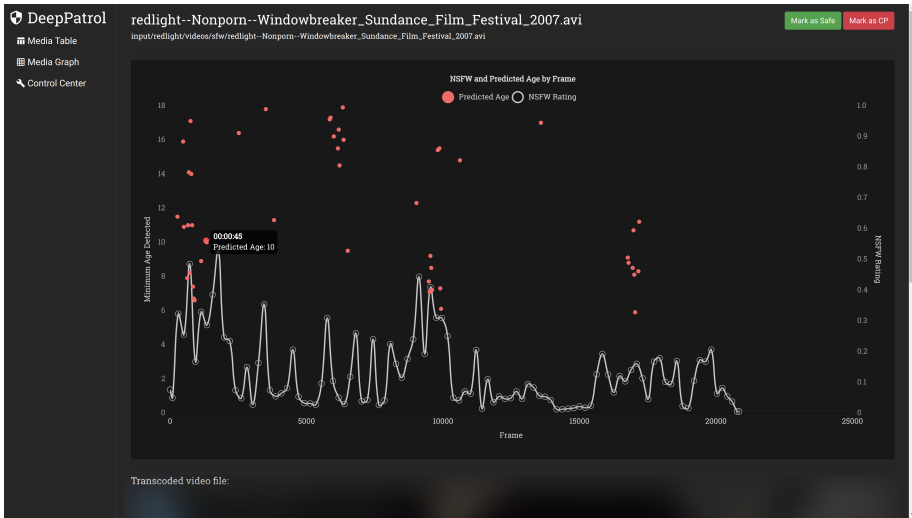
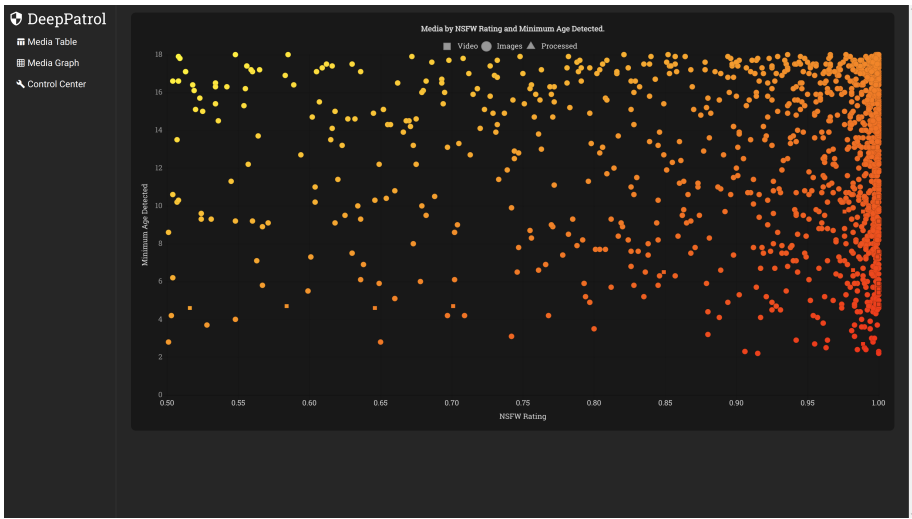
For face detection, the publicly available S3FD: Single Shot Scale-invariant Face Detector is used. The performance of this face detector is impressive, yielding few false positives and tight, high quality bounding boxes around human faces of all sizes and resolutions. After detecting all the human faces in an image, each face is cropped and saved to file with an extended margin of 40% of the height and width of the face. This gives the age estimator “context” in the region around the face. Only faces with a confidence score of 0.80 or higher and a width and height greater than 25 are sent to the age estimator in the hopes of filtering out smaller and lower quality faces. Input images for the S3FD Face Detector are converted to an RGB color format, resized to 640x640 using a bilinear interpolation method, and each color channel has the mean color value of image net subtracted from itself.

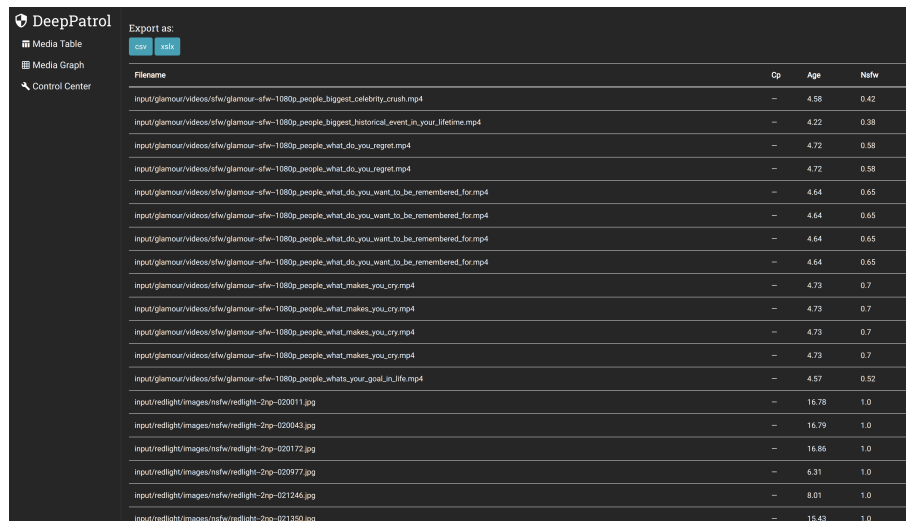
The age estimation module is based on our research. As input it accepts human facial images, and outputs a probability distribution across ages. In order to determine if an image contains a minor, we simply take the expected value of the distribution and check to see if the predicted value is less than the threshold for adulthood. To prepare input images, they are converted to an RGB color format, resized to 256x256 with a bilinear interpolation, center cropped to a size of 224x224, and each channel is standardized to the mean and standard deviation of the ImageNet database.

When performing inference on images, DeepPatrol will simply combine the predictions of both the age estimator, and the NSFW detector. If a minor is detected and there is NSFW content in an image, then

it is flagged as suspected child pornography. The ages of each person detected in the image and the NSFW score are logged to local files. For videos, inference is performed on each extracted frame. Results for each frame are then aggregated for the final video classification. A video is not flagged as illicit material if only a single frame contains contraband because the application may produce many false positives. As the number of frames sampled in a video grows, so would the likelihood of a false positive using this criteria.

The interface created for DeepPatrol is based on web technologies. The following screenshots depict the look-and-feel of DeepPatrol.





DeepPatrol

Media Table
Media Graph
Control Center

Export as:
CSV XLSX

Filename	Cp	Age	Nsfw
input/glamour/videos/sfw/glamour-sfw-1080p_people_biggest_celebrity_crush.mp4	-	4.58	0.42
input/glamour/videos/sfw/glamour-sfw-1080p_people_biggest_historical_event_in_your_lifetime.mp4	-	4.22	0.38
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_do_you_regret.mp4	-	4.72	0.58
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_do_you_regret.mp4	-	4.72	0.58
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_do_you_want_to_be_remembered_for.mp4	-	4.64	0.65
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_do_you_want_to_be_remembered_for.mp4	-	4.64	0.65
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_do_you_want_to_be_remembered_for.mp4	-	4.64	0.65
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_do_you_want_to_be_remembered_for.mp4	-	4.64	0.65
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_makes_you_cry.mp4	-	4.73	0.7
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_makes_you_cry.mp4	-	4.73	0.7
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_makes_you_cry.mp4	-	4.73	0.7
input/glamour/videos/sfw/glamour-sfw-1080p_people_what_makes_you_cry.mp4	-	4.73	0.7
input/glamour/videos/sfw/glamour-sfw-1080p_people_whats_your_goal_in_life.mp4	-	4.57	0.52
input/redlight/images/nsfw/redlight-2rp-020011.jpg	-	16.78	1.0
input/redlight/images/nsfw/redlight-2rp-020043.jpg	-	16.79	1.0
input/redlight/images/nsfw/redlight-2rp-020172.jpg	-	16.86	1.0
input/redlight/images/nsfw/redlight-2rp-020977.jpg	-	6.31	1.0
input/redlight/images/nsfw/redlight-2rp-021246.jpg	-	8.01	1.0
input/redlight/images/nsfw/redlight-2rp-021330.jpg	-	16.43	1.0

In summary, the contributions of our work developing DeepPatrol can be stated as (further details are provided in Jared Rondeau’s Master Dissertation and the Journal Paper):

- A novel framework and software for the automatic detection of child pornography in videos and images, freely available to law enforcement agencies;
- A rigorous analysis of the performance of nudity detection and age estimation models on ethnically diverse and challenging pornographic and non-pornographic images and videos;
- Empirical evidence that treating video classification as a per-frame image classification task with prediction aggregation achieves competitive results at pornography detection;
- An evaluation of our framework for child pornography detection on a real world dataset collected by local law enforcement agents from about 20 real world cases.

5. DATASETS

We used publicly available datasets for training our age estimator and for evaluating the performance of DeepPatrol.

1. The IMDB-Wiki dataset is the largest publicly available labeled aging database, with 523,051 images of 20,284 individuals. It consists of face images of actors and actresses collected from the popular Internet Movie Database and Wikipedia. Each image is labeled by its corresponding biological age. In all experiments, we use only a subset of IMDB-Wiki. At the time of our experiments, the creators of the dataset reported issues with images from the Wikipedia portion of the dataset. All Wikipedia images were excluded from our local copy of the dataset. Images

containing multiple faces were also excluded since the identity of the human associated with the age is not easily inferred. Likewise, all images where the face of the subject could not be located were also removed. This pre-processing was done with the meta-information provided by the maintainers of the dataset. We used 90% of the remaining images for training and 10% for validation, that is, 165,970 and 18,442 images respectively.

2. The APPA-Real dataset was introduced with the objective of gathering a large and robust set of human apparent age guesses on images of people “in the wild” (referring to pictures taken in uncontrolled conditions). The authors relied on crowd-sourcing to label each of the 7,591 images. The pictures collected are of about 7,000 different individuals taken in varying conditions of lighting and image quality, which makes the dataset more representative of real world photo capture. Each image is labeled with both biological and an average of 38 apparent age guesses;
3. The RedLight dataset contains 12,180 images, manually separated into several categories describing the type of pornography. It also contains 15,401 images which are non-pornographic to help evaluate a model's performance. This dataset was collected by the Digital Forensics and Cyber Security Center at the University of Rhode Island in 2010 to assist in the development of the RedLight pornography scanner. One issue prevalent in skin tone based pornography detection tools is the inability to generalize to all races. RedLight provides several hundred NSFW images categorized into various ethnicities, which we view as important to report results on;
4. NPDI is another pornographic dataset containing only videos, separated into three groups: non-porn easy, non-porn difficult, and porn. These contain 200, 200, and 400 videos respectively. The non-porn hard subset of the dataset was intended to be extremely challenging and contains many tricky videos, such as those containing people at the beach, wrestling, swimming, and even infants breastfeeding. The non-porn easy videos were randomly selected from YouTube. The group of pornography videos were selected from websites which exclusively host content of pornographic nature, and includes examples of animated pornography. In total, there is 77 hours of footage in the NPDI dataset. The pornographic video group is also ethnically diverse, as 46% of the videos contain Caucasians, 16% Asians, 14% Africans, and 24% are multi-ethnic;
5. A private dataset which we didn't have access to, containing 1,109 videos and 84,619 images, was collected over a multi-year period, from approximately 20 real-world SEIC cases, by our local law-enforcement partner. We were able to test our software on their data through coordination with our

liaison. Although each video is a known instance of child pornography, the true label of each image in the dataset is somewhat ambiguous, as an image did not have to contain both nudity and a minor for it to be included in the set. Each image comes from a larger collection of identified child pornography involving the given victims for a case, but is not clearly indicated whether or not actual nudity, or the face of a child, is present in each image. Therefore, as of now, we can only approximate the generalization of the NSFW detection and age estimation model on the images in the set.

6. COMPUTING INFRASTRUCTURE AND TRAINING DETAILS

Models were trained and evaluated locally on an Ubuntu 14.04 server configured with 4x NVidia Titan X Pascals, 2x Intel Xeon CPU E5-2620v4, and 64GB of memory. All CNNs were trained using the GPUs. OpenNSFW and the S3FD face detector used the Caffe deep learning framework while our age estimation model runs on PyTorch. Yahoo!'s OpenNSFW nudity detection model was initialized with ImageNet weights and then trained on a private dataset of both NSFW and SFW images.

Our age estimation model uses a DenseNet-161 architecture pre-trained on ImageNet, fine-tuned onto a collection of 130,000 images of actors and actresses crawled from the publicly available IMDB-Wiki dataset for the task of label distribution learning. Finally, our model is fine-tuned onto APPA-REAL for the task of apparent age estimation using normal label distributions parameterized by the mean and standard deviation of the per-image human guesses. For further details, see our conference paper (IJCNN 2018).

Analysis of the runtime of the models deployed at the local law enforcement agency is also provided. This workstation is less powerful than the one the models were trained on, and only has 1 NVidia Quadro P4000. In total, 1,956 videos were split into 1,044,577 frames. 123,200 images were detected. It took the file crawler 11 minutes to identify all files (single thread), 3 hours and 4 minutes for the frame extractor to extract the frames at a rate of 1 per second (using a single thread), 25 hours to detect the faces in the resulting 1,167,777 files (batch size of 1, single GPU), 56 minutes to crop the detected faces into 599,232 files (single thread), 1 hour and 13 minutes to run the age estimation model on each detected face (batch size of 64, single GPU), and 8 hours and 56 minutes to run the nudity detection model on the 1,167,777 total images/frames (batch size of 1, single GPU) for a total runtime of 39 hours.

7. SUPPORTED STUDENTS

During the execution of this project, a total of 2 graduate (Thomas Howard III and Jared Rondeau) and 7 undergraduate students (Douglas Deslauriers, Davin Bernier, Jeremy Peacock, Jonathan Ibanez, Peter Pinto, Luke Watkins, and Jake Ward) were directly involved in research and software development tasks. They were supported by funds from this grant.

8. DELIVERABLES

1. Conference Paper for our research on Apparent Age Estimation (International Joint Conference on Neural Networks 2018)
2. Masters in Computer Science Dissertation (Jared Rondeau)
3. Journal Paper for our child pornography approach (In preparation)
4. Conference Paper for DeepPatrol (In preparation)
5. Github repository for DeepPatrol (source code, installation manual, user manual)
6. Presentations (posters and talks)

9. IMPACTS

This project combined research in age estimation and machine learning together with backend/frontend software development. The development of DeepPatrol will have a positive impact on the state-of-the-art for child pornography detection, and the state-of-the-art on automatic age estimation from face images. We believe that by providing the general public, and specifically law-enforcement agencies, with open-source cutting-edge technology will strongly support the fight against child exploitation crimes.