Improving the Accuracy of Automated Facial Age Estimation to Aid CSEM Investigation

Felix Anda, David Lillis, Aikaterini Kanta, Brett Becker, Elias Bou-Harb, Nhien An Le Khac, and Mark Scanlon

Forensics and Security Research Group, University College Dublin, Ireland.

felix.andabasabe@ucdconnect.ie, {mark.scanlon, an.lekhac}@ucd.ie

ABSTRACT

The investigation of child sexual exploitation material (CSEM) is one of the more commonly encountered criminal investigation types throughout the world. While hash lists of known illegal files are used to identify previously encountered material on new devices, new previously unencountered material requires expert, manual analysis and categorisation. The discovery and analysis of these digital images and videos has the potential to be greatly expedited with the use of automated artificial intelligence (AI) based techniques. Intelligent, automated human-in-the-loop evidence processing and evidence prioritisation could also help alleviate some of the digital evidence backlogs that have become commonplace worldwide. In order for AI aided CSEM investigations to be beneficial, the fundamental question when analysing multimedia content becomes "how old is each subject encountered?''

Figure 1. Example of Borderline Adulthood Identification Challenge (Photo is CC0/Public Domain)

PROBLEM STATEMENT

Accurate facial age estimation has been an arduous task for both humans and machines (Figure 1 depicts a sample challenge for both humans and machines encountering borderline cases). Moreover, the influence of factors such as environment, health habits, lifestyle, makeup, emotions, uncontrolled lightning, and partial facial occlusion hinder the age prediction process and thus impact the final result . Nevertheless, humans are quite accurate at estimating the age of other humans. The mean absolute error (MAE) rate has been measured to vary from 2.07 to 8.62 years depending on the age of the assessor and the studied subject, and the difference between the two [1]. Conversely, machines have reached MAE figures of approximately 4.1 [2] and vary according to the datasets used in the implementation. The use of automation in DF has been criticised due to the constant need of a human-in-the loop approach and the lack of accuracy that models achieve for predicting age.

METHODOLOGY

Our work presents the evaluation of existing cloud-based and offline age estimation services, outlines our deep learning age estimation model, DS13K (built with a VGG-16 Deep Convolutional Neural Network architecture) and uses an ensemble technique to improve the accuracy of underage subject age estimation. The contributing services consisted of Amazon Rekognition, Microsoft Azure Cognitive Services, How-Old.net, and Deep Expectation (DEX). It was found that for the borderline adulthood age range of 16 to 17 years old, our deep learning model performed best with a accuracy performance rate of 68%. A comparative examination of the obtained results allowed us to identify performance trends and issues inherent to each services/tool and develop ensemble techniques to improve the accuracy of automated adulthood determination.

PRELIMINARY RESULTS

While there are a number of methods that can contribute to easing the digital evidence backlog [3][4], accuracy of automated age inference in images and videos is one important technique needed to streamline CSEM cases. This study evaluated four different age prediction services from both online and offline sources. The four services evaluated were Amazon Rekognition (AWS), Microsoft Azure, Deep Expectation (DEX), and How-Old.net. Initial evaluation results on an age range from 0 to 25, indicated that AWS had the overall lowest error rate, followed by How-Old.net. However, the ages that surround the borderline between minority and adulthood (considered to be 18 for this study) were found to follow a different pattern, where DEX surpassed the performance of AWS and Azure. Following these observations, a separate dataset was curated featuring a higher number of sample images in the age range of 1 to 25 inclusive.

Experiments on this dataset indicated that ensemble approaches based on regression substantially outperformed the four systems used for this test. Gradient Boosting and Bagging Regressor approaches not only outperformed the best individual system for the borderline range (16-17) by over 40 %, they were also superior to a hypothetical perfect system that chooses the best prediction from those available each time. Furthermore, the DS13K VGG-16 based model trained for this work had the best figures for the borderline age range (16-17) and a competing third place for the age range (0-5), as depicted in Figure 2. The overall conclusion of the study is that even off-the-shelf regression techniques have been demonstrated to improve upon the performance of commercial offerings, by combining their outputs effectively. This offers a clear motivation for further work on bringing AI based techniques to bear on this and associated DF challenges.

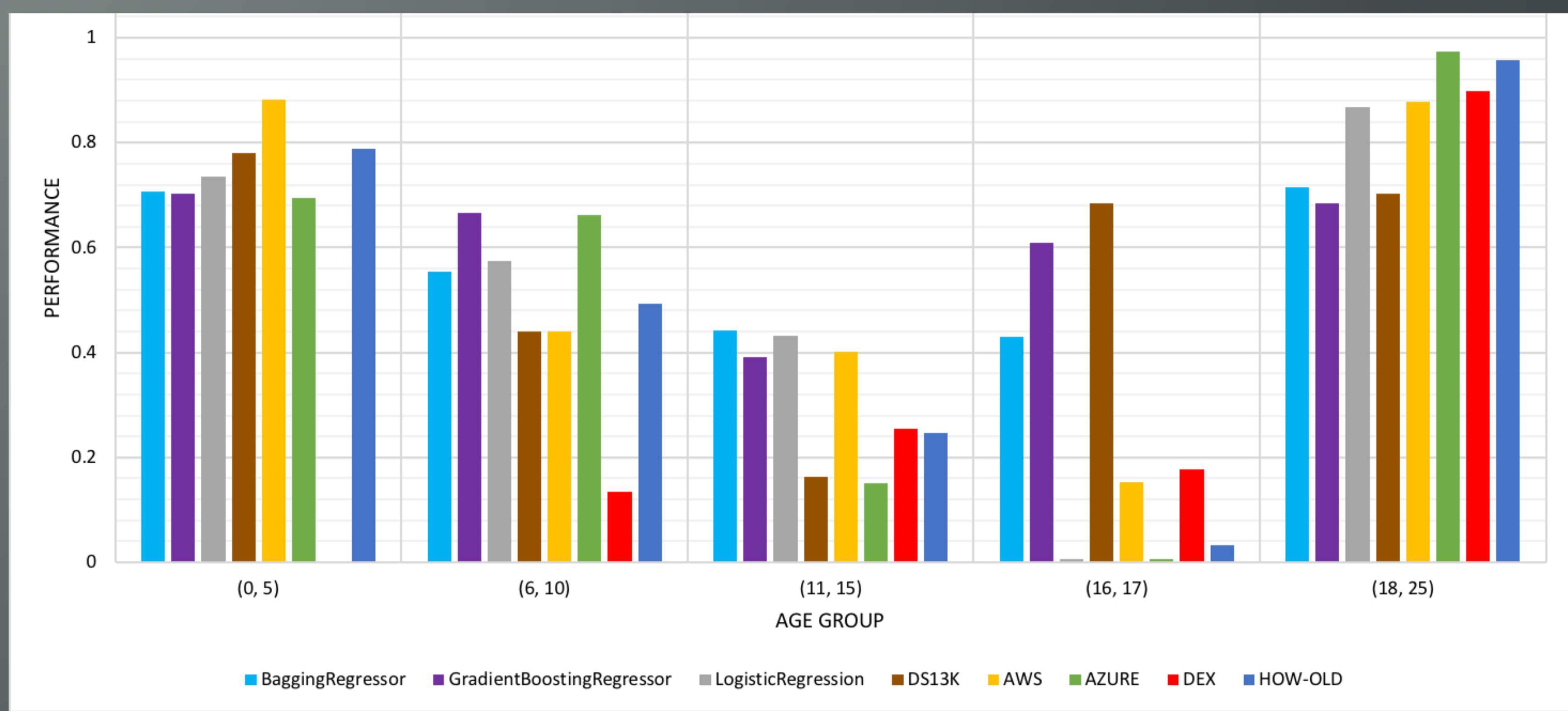

Figure 2. Age Estimator Performance per Age Group

Further information is contained in our published paper on Age Estimation entitled "Evaluating Automated Facial Age Estimation Techniques for Digital Forensics":

Scan me

REFERENCES

1. E. Moyse and S. Brédart. An own-age bias in age estimation of faces. Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology, 62(1):3 – 7, 2012.
2. Khoa Luu, Keshav Seshadri, Marios Savvides, Tien D Bui, and Ching Y Suen. Contourlet appearance model for facial age estimation. In Biometrics (IJCB), 2011 International Joint Conference on, pages 1–8. IEEE, 2011.
3. Anda, F., Lillis, D., Le-Khac, N-A., & Scanlon, M. (2018, May). Evaluating Automated Facial Age Estimation Techniques for Digital Forensics. In 12th International Workshop on Systematic Approaches to Digital Forensics Engineering (SADFE), IEEE Security & Privacy Workshops. IEEE.
4. James, J. I., & Gladyshev, P. (2013). Challenges with Automation in Digital Forensic Investigations. Retrieved from http://arxiv.org/abs/1303.4498