



Contents lists available at ScienceDirect

# Digital Investigation

journal homepage: [www.elsevier.com/locate/diin](http://www.elsevier.com/locate/diin)

## Improving the accuracy of automated facial age estimation to aid CSEM investigations



Felix Anda<sup>a</sup>, David Lillis<sup>a</sup>, Aikaterini Kanta<sup>a</sup>, Brett Becker<sup>a</sup>, Elias Bou-Harb<sup>b</sup>, Nhien An Le Khac<sup>a</sup>, Mark Scanlon<sup>a</sup>. <sup>a</sup>*Forensics and Security Research Group, University College Dublin, Ireland*; <sup>b</sup>*Cyber Threat Intelligence Laboratory, College of Engineering and Computer Science, Florida Atlantic University, USA*

E-mail addresses: [felix.andabasabe@ucdconnect.ie](mailto:felix.andabasabe@ucdconnect.ie) (F. Anda), [david.lillis@ucd.ie](mailto:david.lillis@ucd.ie) (D. Lillis), [aikaterini.kanta@ucdconnect.ie](mailto:aikaterini.kanta@ucdconnect.ie) (A. Kanta), [brett.becker@ucd.ie](mailto:brett.becker@ucd.ie) (B. Becker), [ebouharb@fau.edu](mailto:ebouharb@fau.edu) (N.A. Bou-Harb), [an.lekhac@ucd.ie](mailto:an.lekhac@ucd.ie) (N.A. Le Khac), [mark.scanlon@ucd.ie](mailto:mark.scanlon@ucd.ie) (M. Scanlon).

**Keywords:** Facial Age Estimation, CSEM Investigation, Artificial Intelligence, Digital Forensics.

### Abstract

The investigation of violent crimes against individuals, such as 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 CSEM content are commonly used to identify previously encountered material on suspects' devices, previously unencountered material requires expert, manual analysis and categorisation. The discovery, analysis, and categorisation of these digital images and videos has the potential to be significantly expedited with the use of

automated artificial intelligence (AI) based techniques.

Intelligent, automated evidence processing and prioritisation has the potential to aid investigators in alleviating 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?". Our work presents the evaluation of existing cloud-based and offline age estimation services, introduces our deep learning model, DS13K, which was created with a VGG-16 Deep Convolutional Neural Network (CNN) architecture, and develops an ensemble technique that improves the accuracy of underage facial age estimation.

In addition to our model, a number of existing services including Amazon Rekognition, Microsoft Azure Cognitive Services, How-Old.net, and Deep Expectation (DEX) were used to create an ensemble learning technique. It was found that for the borderline adulthood age range (i.e., 16–17 years old), our DS13K model substantially outperformed existing services, achieving a performance accuracy of 68%. A comparative examination of the obtained results allowed us to identify performance trends and issues inherent to each service/tool and develop ensemble techniques to improve the accuracy of automated adulthood determination.