Reinforcement of age estimation in forensic tools to detect Child Sexual Exploitation Material

Rubel Biswas*†, Deisy Chaves*†, Francisco Jáñez-Martino*†, Pablo Blanco-Medina*†, Eduardo Fidalgo*†, Carlos García-Olalla†, George Azzopardi‡

*Department of Electrical, Systems and Automation, Universidad de León, León, ES
†Researcher at INCIBE (Spanish National Cybersecurity Institute), León, ES
‡Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, NL
Email:{rubel.biswas, deisy.chaves, francisco.janez, pablo.blanco, eduardo.fidalgo}@unileon.es, carlos.olalla@incibe.es, g.azzopardi@rug.nl

***Abstract*—Several image-based approaches for estimating the age of a person are available in computer vision literature. However, most of them perform poorly on minors and young adults, especially when the eyes are occluded. This type of occlusion is common in Child Sexual Exploitation Materials (CSEM), in order to hide the identity of victims. We introduce an approach that builds Soft Stagewise Regression Network (SSR-Net) models with natural and eye-occluded facial images, to estimate the age of minors and young adults. Our proposal reduces the Mean Absolute Error from 7.26 to 6.5, and 6.81 to 4.07 for SSR-Net pre-trained models on the IMDB and MORPH datasets, respectively.**

***Index Terms*—Age estimation, Occlusion, SSR-Net model, CSEM, Forensic images**

Type of contribution: *Research already published*

I. Introduction

In forensic applications, accurate and fast age estimation solutions enhance the detection of victims in Child Sexual Exploitation Materials (CSEM) [1]. Forensic tools may also support Law Enforcement Agencies (LEAs) in identifying criminals through enhanced image analysis [2].

Age estimation is a challenging problem due to factors such as pose and illumination variation, which are commonly found in CSEM images [3]. It is also common for offenders to use accessories or black stripes to hide the face or eyes of the victims [4], which presents further challenges to the performance of age estimators.

An increasing number of deep-learning-based age estimators have been proposed during the last years. However, most of these approaches are designed for the age interval between 0 and $60+$ years, and are trained with unbalanced data [5], [6]. Thus, many of them do not perform well for minors and young adults, aged between 0 and 25 years old.

To address this problem, we present an improved solution for the age estimation of minors and young adults by training Soft Stagewise Regression Network (SSR-Net) models [5] using natural face images and faces with occluded eyes.

II. Related Work

Due to the advancement of deep learning architectures, the performance of age estimators has improved significantly in recent years [5], [7]. Despite this, to our knowledge, there are very few approaches that estimate the age of minor/young adults [8], [9] or eye-occluded facial images [10].

Zhang et al. [7] introduced an accurate, fine-grained age estimation model by combining Long Short-Term Memory (LSTM) networks with residual networks (ResNets) to extract face features from age-sensitive regions. The resulting model is complex and computationally intensive, which does not make it suitable for (near) real-time analysis. To address this requirement, Yang et al. proposed an age estimation model, called SSR-Net [5], based on the Deep EXpectation (DEX) method [11]. They reduced the model size by classifying a small number of classes within the age group. In contrast, Zhang et al. [6] introduced a compact model using cascaded training and multi-scale context to estimate the age with small-scale facial images. These compact models are preferable for real-time tasks due to a reduced computational cost.

III. Methodology

We introduce a two-fold solution for age estimation of minors and young adults, as presented in Fig. 1.

First, we created a balanced dataset with natural face images of minors and young adults and their corresponding eye-occluded versions. The natural facial images, in the range $[0, 25]$ years, were collected from five different well-known datasets, namely IMDB-WIKI, APPA-REAL, AgeDB, UTK-Face, and Diversity in Faces, IBM (DiF). We gathered a total of 130000 minor and young adult images by inspecting these datasets manually, removing images with an incorrect age label or without any human face. Afterwards, we created the occluded version of these images by locating the eye region

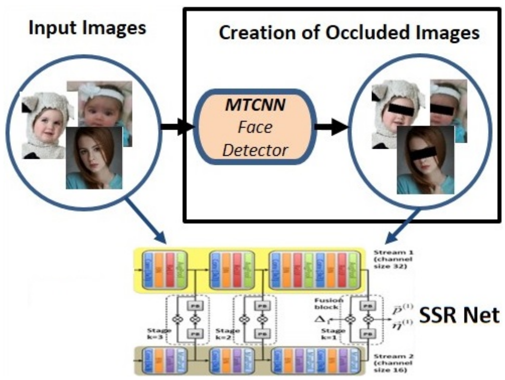

Figure 1. Steps to train age estimation models in minors and young adults.

using the Multi-Task Cascade Convolutional Neural Network (MTCNN) [12] and then masking it in order to simulate the referred conditions on CSEM. Finally, both image sets were merged into one.

Using these images, we implemented a lightweight pre-trained SSR-Net age estimator [5] to build new, fine-tuned age estimation models focused on minor and young adults. Our images were resized to 64×64 pixels to fine-tune the model. Lastly, we split the dataset into a training (80%) and a test (20%) set using stratified random sampling.

IV. Experimental results

We evaluated the age estimation performance using the Mean Absolute Error (MAE) of the SSR-Net pre-trained models that have been trained with face images considering the age range $[0, 25]$ years from four balanced datasets varying in size, [6500-130000], and with two unbalanced datasets, namely MORPH and IMDB.

Then, we measured the MAE's performance enhancement of fine-tuned age estimators using our non-occluded (Org.), eye-occluded (Ocl.), and a combination of both types (Org. - Ocl.) of minor and young adult facial images. Our results are presented in Table I.

We noticed that the age estimation performance was more stable in SSR-Net models —pre-trained on the IMDB dataset— fine-tuned with our merged dataset. These models achieved the best MAE of 3.58 and 4.19 for non-eye-occluded and eye-occluded images, respectively.

Furthermore, we compared our results with the best SSR-Net model against a state-of-art approach, VGG16-based DEX model, trained with our merged dataset. The proposed models outperformed the DEX model with MAE of 6.5 for non-eye-occluded and eye-occluded facial images. In addition, the size of the SSR-Net-based age estimators was much lower than the DEX age estimator, with sizes of $< 1MB$ and $500MB$ respectively.

Finally, we have successfully integrated our proposal, i.e. fine-tuned SSR-Net age estimation model, into the 4NSEEK[1] tool to support the detection of minors on CSEM.

Table I
MAE values of SSR-Net age estimation models. The best MAE values are highlighted in bold.

Model	# images per age		MORPH dataset MAE Test		IMDB dataset MAE Test	
	Train	Test	Org.	Ocl.	Org.	Ocl.
Pre-trained MORPH	–	50	7.16	6.53	7.51	6.93
	–	100	7.19	6.56	7.52	6.93
	–	200	7.17	6.56	7.54	6.94
	–	1000	7.06	6.55	7.52	6.99
Fine-tuned Org. Img.	200	50	5.37	7.04	4.56	6.25
	400	100	4.40	6.82	4.27	6.53
	800	200	4.24	7.32	4.13	6.64
	4000	1000	**3.63**	7.93	**3.58**	6.46
Fine-tuned Ocl. Img.	200	50	6.57	5.67	5.73	5.22
	400	100	6.03	5.29	5.63	5.08
	800	200	5.80	4.97	5.72	5.07
	4000	1000	5.71	**4.22**	5.35	4.58
Fine-tuned Org. - Ocl. Img.	200	50	6.08	5.81	5.00	5.13
	400	100	6.01	5.19	4.91	5.23
	800	200	4.61	4.75	4.47	4.66
	4000	1000	3.93	4.44	3.95	**4.19**

[1] https://www.incibe.es/en/european-projects/4nseek

V. Conclusions

We present an improved age estimator focused on minors and young adults with SSR-Net models, fine-tuned using natural and eye-occluded face images. Results show that our solution performs better in minors and young adults (MAE of 4.07) in comparison to the DEX model (MAE of 6.5), being more robust against eye occlusion.

Moreover, our SSR-Net-based estimators are compact models and suitable for any hardware despite memory capability, as well as forensic applications of child detection on CSEM.

Acknowledgements

This work was supported by the framework agreement between the Universidad de León and INCIBE (Spanish National Cybersecurity Institute) under Addendum 01. Also, this research has been funded with support from the European Commission under the 4NSEEK project with Grant Agreement 821966. This publication reflects the views only of the authors, and the European Commission cannot be held responsible for any use which may be made of the information contained therein. Finally, we acknowledge NVIDIA Corporation with the donation of the TITAN Xp and Tesla K40 GPUs used for this research.

References

[1] D. Chaves, E. Fidalgo, E. Alegre, F. Jánez-Martino, and R. Biswas, "Improving age estimation in minors and young adults with occluded faces to fight against child sexual exploitation," in *VISIGRAPP*, 2020, pp. 721–729.

[2] A. Gangwar, E. Fidalgo, E. Alegre, and V. González-Castro, "Pornography and child sexual abuse detection in image and video: A comparative evaluation," in *8th International Conference on Imaging for Crime Detection and Prevention (ICDP)*, 2017, pp. 37–42.

[3] D. Chaves, E. Fidalgo, E. Alegre, and P. Blanco, "Improving speed-accuracy trade-off in face detectors for forensic tools by image resizing," in *V Jornadas Nacionales de Investigación en Ciberseguridad (JNIC)*, 2019, pp. 222–223.

[4] R. Biswas, V. González-Castro, E. Fidalgo, and D. Chaves, "Boosting child abuse victim identification in forensic tools with hashing techniques," in *V Jornadas Nacionales de Investigación en Ciberseguridad (JNIC)*, 2019, pp. 344–345.

[5] T.-Y. Yang, Y.-H. Huang, Y.-Y. Lin, P.-C. Hsiu, and Y.-Y. Chuang, "Ssr-net: A compact soft stagewise regression network for age estimation," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, 2018, pp. 1–7.

[6] C. Zhang, S. Liu, X. Xu, and C. Zhu, "C3AE: Exploring the limits of compact model for age estimation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 579–12 588.

[7] K. Zhang, N. Liu, X. Yuan, X. Guo, C. Gao, Z. Zhao, and Z. Ma, "Fine-grained age estimation in the wild with attention lstm networks," *IEEE Trans. Circuits and Systems for Video Technology*, pp. 1–12, 2019.

[8] F. Anda, D. Lillis, A. Kanta, B. A. Becker, E. Bou-Harb, N.-A. Le-Khac, and M. Scanlon, "Improving borderline adulthood facial age estimation through ensemble learning," in *14th International Conference on Availability, Reliability and Security (ARES '19)*, 2019, pp. 1–8.

[9] L. M. Demajo, K. Guillaumier, and G. Azzopardi, "Age group recognition from face images using a fusion of cnn- and cosfire-based features," in *Proceedings of the 2nd International Conference on Applications of Intelligent Systems*, ser. APPIS '19, 2019.

[10] L. Ye, B. Li, N. Mohammed, Y. Wang, and J. Liang, "Privacy-preserving age estimation for content rating," in *IEEE 20th International Workshop on Multimedia Signal Processing*, Aug 2018, pp. 1–6.

[11] R. Rothe, R. Timofte, and L. V. Gool, "Dex: Deep expectation of apparent age from a single image," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, December 2015, pp. 10–15.

[12] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.