Assessment of age estimation methods for forensic applications using non-occluded and synthetic occluded facial images

Elouan Derenee Jeuland[1][a], Aitor Del Río Ferreras[2,3][b], Deisy Chaves[2][c], Eduardo Fidalgo[2,3][d], Victor González-Castro[2,3][e] and Enrique Alegre[2,3][f]

[1]*Ecole Nationale Supérieure des Mines de Saint-Étienne, Saint-Étienne, France*

[2]*Department of Electrical, Systems and Automation, Universidad de León, León, Spain*

[3]*Researcher at INCIBE (Spanish National Cybersecurity Institute), León,Spain*

elouan.derenne@etu.emse.fr, {ariof, dchas, efidf, vgonc, enrique.alegre}@unileon.es

Keywords: Age estimation; Deep Learning; facial occlusion; CSEM

Abstract: Age estimation is a valuable forensic tool for criminal investigators since it helps to identify minors or possible offenders in Child Sexual Exploitation Materials (CSEM). Nowadays, Deep Learning methods are considered state-of-the-art for general age estimation. However, they have low performance in predicting the age of minors and older adults because of the few examples of these age groups in the existing datasets. Moreover, facial occlusion is used by offenders in certain CSEM, trying to hide the identity of the victims, which may also affect the performance of age estimators. In this work, we assess the performance of six deep-learning-based age estimators on non-occluded and occluded facial images. We selected FG-Net and APPA-REAL datasets to evaluate the models under non-occluded conditions. To assess the models under occluded conditions, we created synthetically occluded versions of the non-occluded datasets by drawing eye and mouth black masks to simulate the conditions observed in some CSEM images. Experimental results showed that the evaluated age estimators are affected more by eye occlusion than by mouth occlusion. Also, facial occlusion affects more the accuracy of the age estimation of minors and the elderly compared to other age groups. We expect that this study could become an initial benchmark for age estimation under non-occluded and occluded conditions, especially for forensic applications like victim profiling on CSEM where age estimation is essential.

1 INTRODUCTION

Age estimation has been an active research topic in Computer Vision due to its applications in the field of security and human-computer interaction [1] [2]. Nowadays, the COVID-19 pandemic has increased the use of the Internet as a communication tool by everybody, including children. However, they may be more exposed to online risks such as sexual abuse and exploitation. In the first months of the pandemic, Europol reported that the demand for Child Sexual Exploitation Material (CSEM) had increased up to 25% in some Member States[1]. Thus, new forensic tools have been developed to support Law Enforcement Agencies (LEAs) during the identification of victims from CSEM and speed-up the analysis process [3, 4]. These tools required an accurate age estimation to detect potential victims.

However, age estimation from facial images is still an open problem in Computer Vision due to external and internal factors, e.g., image quality, changes in expression, pose and illumination, and the ageing process itself [2, 5]. These are common issues seen in some CSEM images [6], which are also usually affected by face occlusion. Offenders tend to cover the face of victims using accessories or items present in the scene seeking to hide their identity [7], or they draw later, over the images, artificial black masks to cover the victim's eyes. Facial occlusion may affect

[a] https://orcid.org/0000-0001-6472-4081
[b] https://orcid.org/0000-0002-6993-6427
[c] https://orcid.org/0000-0002-7745-8111
[d] https://orcid.org/0000-0003-1202-5232
[e] https://orcid.org/0000-0001-8742-3775
[f] https://orcid.org/0000-0003-2081-774X

[1]https://portal.ieu-monitoring.com/editorial/eu-commission-proposes-new-rules-to-protect-children-against-sexual-abuse

https://doi.org/10.17979/spudc.9788497498418.0972

the performance of age estimators since it has been shown that regions such as the eyes and mouth corners are important features for predicting age.

Deep Learning methods has contributed to improving the performance in age estimation [8, 9, 10, 11, 12, 13, 14, 15, 16, 17], being now Deep Learning the state-of-the-art. However, Deep Learning-based approaches are trained mostly with unbalanced datasets composed of non-occluded facial images [5]. Besides, only a few works have studied the effect of facial occlusion during age estimation [18, 19, 20, 21], despite regions such as eyes and mouth corners are important features during the age prediction.

In this work, we assess the effect of occlusion during age estimation. Specifically, we present a comparison of the performance of six state-of-art age estimation models with non-occluded and occluded facial images. For this comparison, we choose two publicly available datasets for age estimation FG-Net [22], and APPA-REAL [23] datasets, reference for this task, that contain non-occluded images between 0 and 70+ years. Although people can have sunglasses in a few cases, we considered these datasets suitable for evaluating non-occluded conditions. Moreover, we created synthetic versions of FG-Net and APPA-REAL datasets by automatically adding eye and mouth occlusion using black masks. We choose eye occlusion to simulate the conditions present in some CSEM images and mouth occlusion to simulate the occlusion caused by face masks. Besides, we analyse the performance of the age estimators using the Mean Absolute Error (MAE) for eight age groups between 0 and 70+. The results of this study are expected to be used as an age estimation benchmark for forensic applications and help in the development of age estimators robust to eye and mouth facial occlusion.

2 RELATED WORKS

2.1 Age Estimation Models

Age estimation can be defined as the estimation from a facial image of a specific age or a range of ages. It is a challenging problem due to internal factors, such as gender and race, and external factors, such as lifestyle or environment. Nowadays, this problem is addressed mostly using Deep Learning-based models that extract features automatically and do not depend on handcrafted features [8, 9, 10, 11, 12, 13, 14, 15, 16, 17].

Rothe et al. [8] proposed Deep EXpectation of apparent age (DEX), a method to address the age estimation as a classification problem. DEX is built using the VGG16 architecture [24] pre-trained on the ImageNet dataset [25], and fine-tune with the IMBD dataset. DEX yield an MAE of 3.22 on the IMDB-Wiki dataset.

Niu et al. [9] have proposed an End-to-End learning approach for Ordinal Regression with Multiple Output CNN for Age Estimation (ORD). This approach transforms an ordinal regression problem into a series of binary classification problems that are solved using the outputs of the multiple CNNs. ORD obtained an MAE of 3.27 on the MORPH dataset [26].

Gao et al. [10] presented a Deep Label Distribution Learning With Label Ambiguity (DLDL) method. This method was successfully used to label ambiguity while learning the features and the classifier. This prevents the model from over-fit even with small datasets. DLDL was evaluated with MORPH and ChaLearn2015, achieving an MAE of 2.42 and 3.22, respectively. Later, Gao et al. [11] proposed an improvement of DLDL using a lightweight network architecture with a unified framework to jointly learn age distribution and regress age based on the VGG16 architecture. DLDLv2 was also assessed with MORPH and ChaLearn2015 datasets, obtaining an MAE of 3.54 and 1.97, respectively.

Pan et al. [12] proposed a new loss function, Mean-Variance (MV) loss which penalizes differences between the mean and the variance of the estimated age distribution and the true age. MV and softmax loss are embedded into a CNN for age estimation. This approach was evaluated with the datasets FG-Net (MAE of 2.68) and MORPH (MAE of 2.16).

Yang et al. [13] proposed a novel CNN for age estimation with a compact model called SSR-Net. The model estimates an initial age using multi-class classification and refines the predictions using regression to compute the expected age values. SSR-Net yielded an MAE of 3.16 on the MORPH2 dataset. Similarly, Serengil et al. [15] created a lightweight framework called Deepface for face recognition and facial attribute analysis, including age estimation. The age estimator is based on the VGG16 architecture and achieved an MAE of 4.65 on the IMDB-WIKI dataset.

Lin et al. [14] proposed a method called Leveraging Face Parsing Attention for Age Estimation (FP-Age), which incorporates facial semantics during age estimation. The model learn to correctly focus on the informative facial components from unaligned facial images regardless of the head pose and the non-rigid deformation. A clean version of the IMDB-WIKI dataset was created to train and test the model achieving an MAE of 4.68. Besides, FP-Age was tested on the MORPH obtaining an MAE of 1.90.

https://doi.org/10.17979/spudc.9788497498418.0972

Hiba et al. [16] proposed a dual image augmentation method based on attention to build an age estimator and a probabilistic hierarchical regressor framework to estimate the age labels. The data augmentation method allows the model to use multiple facial image augmentations whose embeddings were aggregated by a Transformer-Encoder. This proposed approach achieved an MAE of 2.53 on the MORPH2 dataset using as backbone VGG16.

Shin et al. [17] proposed a general regression algorithm called Moving Window Regression (MWR), and applied it to age estimation. First, MWR obtains an initial rank estimate of an input instance based on the nearest neighbour criterion. Then, it refines the estimated rank by selecting two reference instances to form a search window and estimating the relative rank (ρ-rank) within the search window, iteratively. The ρ-rank measures how much greater is an input compared to the first reference and how much smaller it is compared to the second one. Lastly, a local and a global regressor are used to manage the diverse characteristics in different rank groups. MWR was tested on the datasets MORPH2 (MAE of 2.13), ChaLearn2015 (MAE of 2.77), and UTK (MAE of 4.37).

Nevertheless, the reviewed age estimation methods were trained/tested using unbalanced datasets, i.e. datasets with a few samples for minors and elderly people compared to other ages. Hence, most of these methods may have a large error on the age prediction of the age of minors and the elderly. Besides, the training/testing datasets are composed mostly by non-occluded faces, and therefore, the accuracy of the age estimation may also drop during the analysis of occluded faces. Moreover, the methods have not been evaluated on a common dataset, which makes it difficult to compare their performance fairly, even only on non-occluded images. This work presents a comparison of several state-of-the are age estimators using the FG-Net and the APPA-REAL datasets, aiming to establish an initial benchmark for age estimation.

2.2 Age Estimation with occluded faces

Offenders sometimes cover facial regions, e.g. the eyes, of CSEM victims to avoid their recognition, which may also affect the performance of age estimators used to detect victims in this material. Besides, due to the COVID-19 pandemic, using face masks to cover the mouth is common, which may also hinder the performance of age estimation models. However, only a few works have studied the effect of facial occlusion in age estimation and focus on the creation of models robust against occlusion [18, 19, 20, 21].

Yadav et al. [18] investigated which facial cues are used by humans to estimate the age, the gender and the ethnic. They analysed different facial regions (i.e. the T-Region, binocular region, chin and mouth, and masked eye) and how they influenced on the estimation. As a result, they proposed an algorithm that integrated the knowledge learned during manual age estimation to enhance the facial recognition performance. The models were trained using the datasets MORPH, IIIT-Delhi and FG-Net.

Ye et al. [19] proposed a method to preserve the privacy of a specific age range which estimates the age of images with occluded eyes using a VGG architecture pre-trained on ImageNet and the IMDB-WIKI dataset. They proposed an attention mechanism to learn discriminative features from facial images. The age was estimated within the eight groups of the Adience dataset achieving an accuracy of 66.50% and a 1-off accuracy of 94.98%.

Chaves et al. [20] evaluated the SSR-Net model to estimate the age of minors and young adults using non-occluded and occluded facial images. They proposed a strategy to improve the age estimation using both types of images to fine-tune pre-trained models on IMDB and MORPH datasets. The model was trained using a combination of faces between 0-25 years from several publicly available datasets and obtained an average MAE of 4.07.

Cai et al. [21] proposed an Occlusion Contrast (OCCO) method for mouth occlusion that ignored the occluded region of the image and focused on non-occluded areas. To achieve this, they used self-supervised contrastive learning for learning non-occluded features. Also, they used a cost-sensitive strategy to constrain the learning of the classifier and coped with unbalance data. OCCO was tested with the ChaLearn2015 (MAE of 3.21) and MORPH (MAE of 2.29) datasets.

Although the reviewed methods considered facial occlusion to built robust age estimators, they focused on a specific occluded region, e.g. eye or mouth. Hence, similar to models developed considering mainly non-occluded faces, it is difficult to compare their performance. To cope with this, we create eye and mouth occluded versions of the FG-Net and the APPA-REAL datasets to assess the performance of state-of-the-art age estimators.

3 METHODOLOGY

We aim to assess the performance of age estimators from faces under non-occluded and occluded conditions to establish an initial age estimation benchmark for forensic applications, see Fig. 1.

 https://doi.org/10.17979/spudc.9788497498418.0972

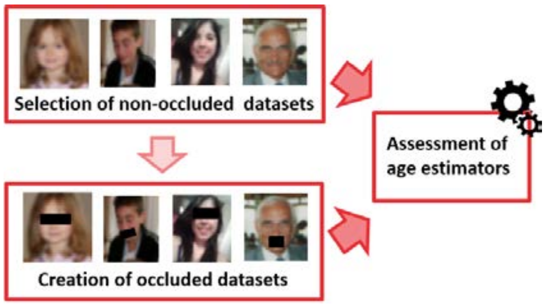

Figure 1: Evaluation methodology.

First, we selected two publicly available datasets for age estimation (FG-Net and APPA-REAL), composed by non-occluded images, and created occluded versions of them with eye and mouth occlusion drawing black masks automatically. We carried out the eye occlusion to reproduce the conditions present in some CSEM images, and mouth occlusion to simulate the occlusion caused by face masks.

Second, we chose six state-of-art age estimation models and evaluated them using (a) the non-occluded datasets and (b) the synthetically occluded ones in terms of MAE and analysed their performance considering eight age groups between 0 and 70+ years.

3.1 Non-Occluded Datasets

We selected the FG-Net and the APPA-REAL datasets to evaluate the age estimation methods under non-occluded conditions. Both datasets are publicly available and contain non-occluded facial images. Besides, we inspected the datasets manually and removed the images labelled with an incorrect age or with no human face. After cleaning the datasets, they contained 1002 facial images with the age range from 0 to 69 years old (FG-Net) and 6884 facial images with the age varying between 1 and 100 years old (APPA-REAL). Note that any image of the FG-Net dataset was not removed during the manual verification.

3.2 Creation of Occluded Datasets

In order to create synthetic occluded images, it is necessary to recognize the area of the face to be covered. Therefore, given a non-occluded facial image, first, the Multi-Task Cascade CNN (MTCNN) [27] method is used to identify the facial landmarks, i.e., the location of the right and the left eye, and the right and the left external points of the mouth. Second, the slope of the lines that connect these points is obtained and employed to determine the position and the dimensions of the rectangular mask to be drawn over the eyes and

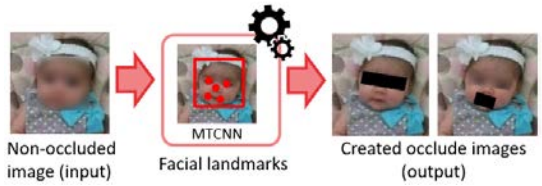

Figure 2: Creation of eye and mouth occluded images.

the mouth, respectively. The dimensions of the rectangle covering the eyes are the 25% of the height and the 95% of the width of the bounding box containing the face. While the dimensions of the rectangle to cover the mouth correspond to the 25% of the height and the 55% of the width of the bounding box containing the face, as illustrated in Fig. 2. See the paper [28] for more details.

3.3 Assessed Age Estimators

We chose six age estimators from the state-of-art considering the availability of public implementations and the reported performance (see Section 2.1). These methods are Deep EXpectation (DEX), Ordinal Regression with Multiple Output CNN (ORD), Deep Label Distribution Learning with Label ambiguity (DLDL), Expectation of Label Distribution Learning (DLDLv2), Mean-Variance Loss (MV), and Leveraging Face Parsing Attention (FP-Age). More details about the assessed methods are showed in Section 2.1.

In this work, we used the implementation of Yiming Lin[2] who re-implemented the selected age estimators (DEX, ORD, DLDL, DLDLv2, MV and FP-Age) from scratch using ResNet50 as a backbone. Input images were preprocessed using the RoI Tanh-polar transform [29] and models were trained using a clean version of the IMBD dataset [14], that contains 287683 images of people between 0 and 97 years, without facial occlusion. Even though Yiming Lin's implementation does not replicate the exact training conditions reported in the literature for the age estimators and the obtained results may vary from the ones reported in the state-of-art for non-occluded images, since the models were built using the same conditions, we considered that this implementation allows us to make a fair comparison of the age estimation performance using the non-occluded and occluded images.

3.4 Evaluation Metrics

We compared the performance of the models using the Mean Absolute Error (MAE), which corresponds

[2]https://github.com/ibug-group/fpage

to the average of the absolute errors between the predicted ages and the ground truth ones. Besides, similar to [30], we analysed the MAE considering eight age groups (i.e, 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, and 70+ years) to identify the age ranges where the models perform better. Finally, to measure the regularity of the age estimation over different age ranges, we computed the standard deviation of the MAE (σ) over the considered eight age groups.

4 EXPERIMENTAL RESULTS

Table 1, Table 2 and Table 3 condense, respectively, the MAE, the standard deviation of the MAE (σ) and the MAE for each of the eight age ranges between 0-97 years (see Section 3.4) obtained for the evaluated age estimators (DEX, ORD, DLDL, DLDLv2, MV and FP-Age) using non-occluded facial images from the FG-Net and the APPA-REAL datasets, as well as the performance of the models using the occluded versions of such datasets. Note that FG-Net does not have images of ages 70+, so no results are provided in that age group in those tables.

Results confirmed that all models perform poorly in both, i.e. non-occluded and occluded images, on pre-pubescents (0-9 years), pubescents (10-19 years) and elderly people (60-97 years), in comparison to the other age groups (20-59 years). In particular, we observed for the APPA-REAL dataset (i.e., non-occluded images), on average, an MAE of 15.58 for 0-9 years, 10.54 for 10-19 years, 5.83 for 20-59 years, and 11.65 for 60-97 years. This difference in the performance between the age groups is even higher on the eye occluded version of the APPA-REAL dataset where we obtained, in average, an MAE of 22.92 for 0-9 years, 16.10 for 10-19 years, 7.79 for 20-59 years, and 18.50 for 60-97 years. Presumably, this performance is caused by the unbalance of the existing datasets used for building the age estimators, which contain few examples in these age groups (pre-pubescents, pubescents and elderly people).

Age estimators drop their performance when they are used with occluded images, being slightly more affected by the eye occlusion than when the occlusion is on the mouth. This indicates that estimators rely more on features from the eye region. Moreover, the best performance on non-occluded images is achieved by MV on the FG-Net (MAE of 6.42) and on the APPA-REAL (MAE of 8.03) datasets. In the case of the occluded versions of the datasets, FP-Age was the model that obtained the best performance on the FG-Net (MAE of 9.90 and 9.66 for eye and mouth occluded images) and on the APPA-REAL (MAE of 11.97 and 10.71 for eye and mouth occluded images) datasets.

5 CONCLUSIONS

We present a comparison of the MAE of six age estimators based on Deep Learning (DEX, ORD, DLDL, DLDLv2, MV and FP-Age) by analysing non-occluded facial images from two datasets (FG-Net and APPA-REAL) and occluded facial images created from the non-occluded ones by drawing black masks over the eye and the mouth regions to simulate the occlusion seen in some CSEM images and the occlusion caused by wearing face masks.

Results show that age estimation models are slightly affected more by the eye occlusion than by the mouth one, indicating that estimators rely more on features obtained from the eye region. Besides, it is confirmed that models perform poorly in predicting the age of minors (0-19 years) and elderly people (60+ years), in both, non-occluded and occluded images, presumably because of the few examples of these age groups in the existing datasets used for training the age estimators. Moreover, MV and FP-Age achieved the best performance on non-occluded and occluded (i.e., eye and mouth) images, respectively.

This work aims to provide an initial benchmark for the assessment of Deep Learning-based age estimators and helps to understand the performance of age estimation models under non-occluded and occluded conditions, which are relevant for forensic applications that required age prediction, e.g. victim profiling on CSEM. Furthermore, future research lines will focus on evaluating novel architectures such as transformers to built age estimators robust against facial occlusion.

ACKNOWLEDGEMENTS

This research has been funded with support from the European Union's Horizon 2020 Research and Innovation Framework Programme, H2020 SU-FCT-2019 under the GRACE project with Grant Agreement 883341. This publication reflects the views only of the authors, and the European Union's Horizon 2020 Research and Innovation Framework Programme, H2020 SU-FCT-2019 cannot be held responsible for any use which may be made of the information contained therein.

 https://doi.org/10.17979/spudc.9788497498418.0972

Table 1: Evaluation of the DEX, ORD, DLDL, DLDLv2, MV and FP-Age age estimators on the FG-Net and, the APPA-REAL datasets (non-occluded images). The best MAE values are shown in bold.

Evaluation	FG-Net dataset						APPA-REAL dataset					
Metric	DEX	DLDL	DLDLv2	MV	ORD	FP-Age	DEX	DLDL	DLDLv2	MV	ORD	FP-Age
MAE 0-9 yr	10.90	22.71	10.97	10.00	**9.35**	17.50	13.76	26.03	13.82	**12.03**	12.21	15.62
MAE 10-19 yr	7.69	13.34	**7.20**	7.78	7.63	11.00	9.45	15.87	9.45	**9.02**	9.20	10.26
MAE 20-29 yr	5.73	6.50	**5.14**	5.46	**5.14**	5.72	6.22	7.80	6.31	6.02	5.86	**5.41**
MAE 30-39 yr	4.44	3.89	4.40	4.35	4.71	**3.29**	4.80	4.94	5.01	4.83	4.75	**4.09**
MAE 40-49 yr	**5.59**	6.17	6.07	6.06	6.38	5.76	5.60	6.21	5.84	5.67	5.56	**5.22**
MAE 50-59 yr	4.69	6.67	4.07	**3.88**	4.69	5.84	6.81	6.81	6.56	6.41	**6.32**	6.93
MAE 60-69 yr	10.83	11.33	10.08	**7.42**	9.92	8.64	7.23	9.00	7.81	7.25	**7.04**	8.68
MAE 70+ yr	—	—	—	—	—	—	13.01	21.96	13.60	**13.00**	14.02	17.28
MAE	7.13	10.09	6.85	**6.42**	6.83	8.25	8.36	12.33	8.55	**8.03**	8.12	9.19
σ	2.56	5.97	2.53	**1.98**	2.02	4.41	3.17	7.48	3.23	**2.85**	3.17	4.60

Table 2: Evaluation of the DEX, ORD, DLDL, DLDLv2, MV and FP-Age age estimators on the eye occluded versions of the FG-Net and, the APPA-REAL datasets . The best MAE values are shown in bold.

Evaluation	FG-Net dataset						APPA-REAL dataset					
Metric	DEX	DLDL	DLDLv2	MV	ORC	FP-Age	DEX	DLDL	DLDLv2	MV	ORC	FP-Age
MAE 0-9 yr	**19.20**	27.81	20.48	23.67	20.10	20.10	21.68	28.46	21.92	25.59	21.45	**18.41**
MAE 10-19 yr	**11.90**	19.15	12.96	16.01	13.94	14.74	14.69	19.57	15.38	17.71	15.47	**13.82**
MAE 20-29 yr	**6.77**	12.09	8.42	9.85	8.36	8.08	7.96	11.51	8.69	10.52	9.04	**7.60**
MAE 30-39 yr	**5.24**	6.74	5.31	7.34	5.50	3.83	5.55	6.18	5.75	6.00	5.75	**4.60**
MAE 40-49 yr	8.09	**6.30**	6.84	8.29	6.49	5.35	7.69	6.34	7.71	6.96	6.86	**6.00**
MAE 50-59 yr	12.50	8.55	7.88	**6.98**	8.10	7.95	10.57	**8.25**	9.97	9.04	9.09	9.39
MAE 60-69 yr	11.00	11.83	10.39	12.39	**9.56**	9.27	13.00	12.53	11.59	**10.88**	12.19	12.56
MAE 70+ yr	—	—	—	—	—	—	25.47	28.47	23.79	22.82	25.28	23.40
MAE	10.67	13.21	10.33	12.07	10.29	**9.90**	13.33	15.16	13.10	13.69	13.14	**11.97**
σ	**4.29**	7.20	4.73	5.58	4.72	5.24	6.60	8.66	6.24	6.94	6.61	**6.04**

Table 3: Evaluation of the DEX, ORD, DLDL, DLDLv2, MV and FP-Age age estimators on the mouth occluded versions of the FG-Net and, the APPA-REAL datasets. The best MAE values are shown in bold.

Evaluation	FG-Net dataset						APPA-REAL dataset					
Metric	DEX	DLDL	DLDLv2	MV	ORD	FP-Age	DEX	DLDL	DLDLv2	MV	ORC	FP-Age
MAE 0-9 yr	14.95	24.65	15.75	**14.82**	15.88	19.15	17.08	27.40	16.61	16.49	16.57	**16.21**
MAE 10-19 yr	10.21	15.39	10.18	**10.10**	10.96	11.92	11.54	18.36	11.77	11.94	11.79	**10.88**
MAE 20-29 yr	**4.98**	7.82	5.22	5.13	5.56	5.64	5.79	9.56	5.89	5.90	5.67	**5.22**
MAE 30-39 yr	4.92	**4.30**	5.33	4.77	4.50	3.67	5.24	5.27	5.29	4.89	4.98	**4.51**
MAE 40-49 yr	7.70	6.69	8.09	7.77	6.82	**6.69**	8.22	**6.19**	8.59	8.05	8.06	6.59
MAE 50-59 yr	10.98	9.45	11.10	9.33	10.31	**7.86**	11.62	**7.79**	11.64	11.33	11.31	8.49
MAE 60-69 yr	19.61	17.22	20.17	17.44	16.94	**12.66**	15.82	12.74	15.38	16.47	15.05	**11.09**
MAE 70+ yr	—	—	—	—	—	—	25.46	25.84	25.11	25.79	25.19	**22.65**
MAE	10.48	12.22	10.83	9.90	10.14	**9.66**	12.60	14.14	12.53	12.61	12.33	**10.71**
σ	4.97	6.65	5.09	**4.39**	4.53	4.90	6.30	8.18	6.09	6.43	6.21	**5.74**

https://doi.org/10.17979/spudc.9788497498418.0972

REFERENCES

[1] M. Alrjebi, N. Pathirage, W. Liu, and L. Li, “Face recognition against occlusions via colour fusion using 2d-mcf model and src,” *Pattern Recognition Letters*, vol. 95, pp. 1339–1351, 05 2017.

[2] R. Angulu, J. R. Tapamo, and A. O. Adewumi, “Age estimation via face images: a survey,” *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, p. 42, Jun 2018.

[3] A. Gangwar, V. GonzÃ¡lez-Castro, E. Alegre, and E. Fidalgo, “Attm-cnn: Attention and metric learning based cnn for pornography, age and child sexual abuse (csa) detection in images,” *Neurocomputing*, vol. 445, pp. 81–104, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523122100312X

[4] M. W. Al-Nabki., E. Fidalgo., E. Alegre., and R. Aláiz-Rodríguez., “File name classification approach to identify child sexual abuse,” in *Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods - ICPRAM,*, INSTICC. SciTePress, 2020, pp. 228–234.

[5] S. Silvia and S. H. Supangkat, “A literature review on facial age estimation researches,” in *2020 International Conference on ICT for Smart Society (ICISS)*, 2020, pp. 1–6.

[6] D. Chaves, E. Fidalgo, E. Alegre, F. Jáñez Martino, R. Alaiz-Rodríguez, and G. Azzopardi, “Assessment and estimation of face detection performance based on deep learning for forensic applications,” *Sensors*, vol. 20, no. 16, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/16/4491

[7] 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)*, vol. 1, 2019, pp. 344–345.

[8] 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.

[9] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, “Ordinal regression with multiple output cnn for age estimation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[10] B.-B. Gao, C. Xing, C.-W. Xie, J. Wu, and X. Geng, “Deep label distribution learning with label ambiguity,” *IEEE Transactions on Image Processing*, vol. 26, pp. 2825–2838, 04 2017.

[11] B.-B. Gao, H.-Y. Zhou, J. Wu, and X. Geng, “Age estimation using expectation of label distribution learning,” in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18.* International Joint Conferences on Artificial Intelligence Organization, 7 2018, pp. 712–718. [Online]. Available: https://doi.org/10.24963/ijcai.2018/99

[12] H. Pan, H. Han, S. Shan, and X. Chen, “Mean-variance loss for deep age estimation from a face,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5285–5294.

[13] 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.* International Joint Conferences on Artificial Intelligence Organization, 7 2018, pp. 1078–1084. [Online]. Available: https://doi.org/10.24963/ijcai.2018/150

[14] Y. Lin, J. Shen, Y. Wang, and M. Pantic, “Fp-age: Leveraging face parsing attention for facial age estimation in the wild,” 2021. [Online]. Available: https://arxiv.org/abs/2106.11145

[15] S. I. Serengil and A. Ozpinar, “Hyperextended lightface: A facial attribute analysis framework,” in *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, 2021, pp. 1–4.

[16] S. Hiba and Y. Keller, “Hierarchical attention-based age estimation and bias estimation,” 2021. [Online]. Available: https://arxiv.org/abs/2103.09882

[17] N.-H. Shin, S.-H. Lee, and C.-S. Kim, “Moving window regression: A novel approach to ordinal regression,” 2022. [Online]. Available: https://arxiv.org/abs/2203.13122

[18] D. Yadav, R. Singh, M. Vatsa, and A. Noore, “Recognizing age-separated face images: Humans and machines,” *PLoS ONE*, vol. 9, no. 12, pp. 1–22, 2014.

[19] L. Ye, B. Li, N. Mohammed, Y. Wang, and J. Liang, “Privacy-preserving age estimation for

 https://doi.org/10.17979/spudc.9788497498418.0972

content rating," in *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, Aug 2018, pp. 1–6.

[20] D. Chaves., E. Fidalgo., E. Alegre., F. Jáñez-Martino., and R. Biswas., "Improving age estimation in minors and young adults with occluded faces to fight against child sexual exploitation," in *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP,*, INSTICC. SciTePress, 2020, pp. 721–729.

[21] W. Cai and H. Liu, *Occlusion Contrasts for Self-Supervised Facial Age Estimation*. New York, NY, USA: Association for Computing Machinery, 2021, p. 1–7. [Online]. Available: https://doi.org/10.1145/3476098.3485052

[22] A. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, 2002.

[23] E. Agustsson, R. Timofte, S. Escalera, X. Baró, I. Guyon, and R. Rothe, "Apparent and real age estimation in still images with deep residual regressors on APPA-REAL database," in *FG 2017 - 12th IEEE International Conference on Automatic Face and Gesture Recognition*, May 2017, pp. 1–12. [Online]. Available: https://hal.inria.fr/hal-01677892

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[26] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," *FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, vol. 2006, pp. 341 – 345, 05 2006.

[27] 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.

[28] R. Biswas, V. González-Castro, E. Fidalgo, and E. Alegre, "A new perceptual hashing method for verification and identity classification of occluded faces," *Image and Vision Computing*, vol. 113, p. 104245, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885621001505

[29] Y. Lin, J. Shen, Y. Wang, and M. Pantic, "Roi tanh-polar transformer network for face parsing in the wild," *Image and Vision Computing*, vol. 112, p. 104190, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885621000950

[30] A. Greco, "Guess the age 2021: Age estimation from facial images with deep convolutional neural networks," in *Computer Analysis of Images and Patterns*, N. Tsapatsoulis, A. Panayides, T. Theocharides, A. Lanitis, C. Pattichis, and M. Vento, Eds. Cham: Springer International Publishing, 2021, pp. 265–274.

BY-NC-SA