Improving Age Estimation in Minors and Young Adults with Occluded Faces to Fight Against Child Sexual Exploitation

Deisy Chaves[1,2][a], Eduardo Fidalgo[1,2][b], Enrique Alegre[1,2][c], Francisco Jáñez-Martino[1,2][d] and Rubel Biswas[1,2][e]

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

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

Keywords: Age Estimation, Eye Occlusion, SSR-Net Model, CSEM, Forensic Images.

Abstract: Accurate and fast age estimation is crucial in systems for detecting possible victims in Child Sexual Exploitation Materials. Age estimation obtains state of the art results with deep learning. However, these models tend to perform poorly in minors and young adults, because they are trained with unbalanced data and few examples. Furthermore, some Child Sexual Exploitation images present eye occlusion to hide the identity of the victims, which may also affect the performance of age estimators. In this work, we evaluate the performance of Soft Stagewise Regression Network (SSR-Net), a compact size age estimator model, with non-occluded and occluded face images. We propose an approach to improve the age estimation in minors and young adults by using both types of facial images to create SSR-Net models. The proposed strategy builds robust age estimators that improve SSR-Net pre-trained models on IMBD and MORPH datasets, and a Deep EXpectation model, reducing the Mean Absolute Error (*MAE*) from 7.26, 6.81 and 6.5 respectively, to 4.07 with our proposal.

1 INTRODUCTION

Automatic age estimation from facial images has been extensively studied due to their applications in the field of security and human-computer interaction (Angulu et al., 2018). In forensic applications, during the analysis of Child Sexual Exploitation Materials (CSEM), accurate and fast age estimation is essential to detect possible victims. These systems aim to help investigators or Law Enforcement Agencies to speed-up the analysis of CSEM because the criminal's use of anonymization tools and private networks have increased significantly this kind of material (Gangwar et al., 2017; Anda et al., 2019; Al-Nabki et al., 2019). However, age estimation is still an open problem in computer vision as a result of several factors: image quality, variations in expression, pose and illumination, as well as the aging process itself. Aging is an inexorable process that affects at different rates the facial appearance of people of the same age (Angulu et al., 2018). These are common factors found in some CSEM images (Chaves et al., 2019), jointly with another concerning issue, as face occlusion. Criminals used accessories or items present in the scene to cover the face of victims in an attempt to hide their identity (Biswas et al., 2019) or they draw later, over the images, artificial glasses or black stripes covering the eyes, which may affect the performance of age estimators.

Deep learning methods have been developed to estimate age mainly in an interval between 0 and 60+ years (Rothe et al., 2015; Chen et al., 2017; Yang et al., 2018; Zhang et al., 2019). In general, a large amount of labeled facial images based on age label is required to create accurate age estimation models, but most of the available datasets used to build deep-learning-based estimators are highly unbalanced with few examples of minors and young adults, i.e. subjects between 0 and 25 years old. As a result, most of these approaches had a large error when are applied to minors and young adults (Anda et al., 2019). Notwithstanding, the problem of unbalanced data is not new in the literature with solutions including data augmentation (Carcagnì et al., 2015; Hase et al., 2019) and statistical methods (Galusha et al., 2019),

[a] https://orcid.org/0000-0002-7745-8111
[b] https://orcid.org/0000-0003-1202-5232
[c] https://orcid.org/0000-0003-2081-774X
[d] https://orcid.org/0000-0001-7665-6418
[e] https://orcid.org/0000-0003-1344-5968

Chaves, D., Fidalgo, E., Alegre, E., Jáñez-Martino, F. and Biswas, R.
Improving Age Estimation in Minors and Young Adults with Occluded Faces to Fight Against Child Sexual Exploitation.
DOI: 10.5220/0008945907210729
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 5: VISAPP*, pages 721-729
ISBN: 978-989-758-402-2; ISSN: 2184-4321

we considered this problem out of the scope of this research since we focus on the analysis of eye occlusion during age estimation.

In this paper, we propose a strategy to improve the age estimation in minors and young adults by combining non-occluded and artificially eye occluded face images during the training of Soft Stagewise Regression Network (SSR-Net) models (Yang et al., 2018). This allows creating compact size models that successfully estimate the age from non-occluded and occluded faces. This work is part of the European project Forensic Against Sexual Exploitation of Children (4NSEEK). Hence, the age estimation models resulted from this study will be integrated into the 4NSEEK tool for CSEM analysis.

The remaining of the paper is organized as follows: Section 2 describes relevant age estimation methods of the last years based on deep learning; Section 3 presents the methodology used to build age estimation models in minors and young adults; Section 4 the experimental set-up; Section 5 focuses on the experimental results and discussion; and Section 6 comprises the final remarks and future work.

2 RELATED WORKS

2.1 Age Estimation Models

The evolution of deep learning has improved the performance of automatic age estimators from facial images by using Convolutional Neural Network (CNN) architectures (Rothe et al., 2015; Yi et al., 2015; Chen et al., 2017; Yang et al., 2018; Zhang et al., 2019; Zhang et al., 2019). Deep EXpectation (DEX) method (Rothe et al., 2015) addressed the apparent age estimation as a deep classification problem using VGG-16 architecture through fine-tuning a pre-trained ImageNet[1] model with the IMBD dataset. This dataset was collected by the authors from the IMDB website. A multi-region CNN method was presented in (Yi et al., 2015) to estimate age employing features from eight sub-region of a facial image. A Ranking-CNN (Chen et al., 2017) used a deep ranking model for age estimation based on binary CNN outputs that adjust the age range until obtaining final age prediction. A method for fine-grained age estimation was developed in (Zhang et al., 2019) by combining the residual networks (ResNets) or the ResNets of RestNets (RoR) models with Attention Long Short-Term Memory (LSTM) to extract features of age-sensitive regions, the model also was pre-trained on ImageNet and fine-tuned on the IMDB dataset.

These works aimed to build robust and effective age estimation models generally based on bulky CNN architectures like VGG. However, some applications such as forensic analysis or surveillance, where a large number of images or videos are analyzed, require compact size and portable models that provide reliable age estimations in real-time. In this sense, an age estimation model called SSR-Net (Yang et al., 2018) was proposed based on DEX. This method mainly focused on reducing the size of models through classifying a small number of classes within the age group and refining them in each stage. Besides, achieving a similar the Mean Absolute Error (*MAE*) on MORPH-2 dataset (Ricanek and Tesafaye, 2006), the size of the SSR-Net model is more compact (0.32MB) in comparison to DEX model (500MB). Also, in (Zhang et al., 2019) a compact basic model was proposed using cascaded training and multi-scale context to tackle age estimation with small-scale facial images. These lightweight models allow age estimation regardless of hardware and memory capability, offering a more appropriate option for detecting possible CSEM victims than standard models.

Nevertheless, the reviewed age estimation approaches based on bulky or compact models were not specifically tested on minor-age facial images and used the age range from 0 to +66 years old. Hence, most of these methods tend to have a large error when working as minor-age estimators (Anda et al., 2019). To the best of our knowledge, few approaches focus on the age estimation of children (Antipov et al., 2016; Anda et al., 2019) and those methods were built on VGG architectures which are large size models.

2.2 Occluded Faces

In order to avoid the recognition of CSEM victims, criminals often cover the eyes of a victim, which may affect the performance of age estimators. Moreover, the occlusion is generally considered in other fields such as face recognition and face verification (Min et al., 2011; Zhao et al., 2016; Alrjebi et al., 2017; Cen and Wang, 2019; Biswas et al., 2019). Thus, only few works have studied the effect of eye occlusion in age estimation (Ye et al., 2018; Yadav et al., 2014).

In (Yadav et al., 2014) was improved the age estimation and face recognition, developing an algorithm inspired by human age estimation to determine the weight of facial features depending on the age group. They used facial images and partial face images, which contained areas of the face such as T-Region, binocular region, chin and mouth, and masked eyes. Ten age-groups are considered between 0 to +80 in-

[1] http://www.image-net.org/

Table 1: Description of age datasets used to create the training and the test sets. *Although the DiF dataset was created in 2016, IBM released it in 2019.

Dataset	Year	Age range	# of faces	# of images
IMDB-WIKI	2015	0–100	523051	105545
APPA-REAL	2017	0–95	7591	3115
AgeDB	2017	16–100	16488	1809
UTKFace	2017	0–116	20000	7941
DiF*	2019	0–60+	0.97M	335349

cluding four minor-age groups. They noticed that the chin area provides the most relevant features for age estimation in infants, i.e. subject between 0 to 5 years. Attention deep learning mechanism is introduced in (Ye et al., 2018) for age estimation with eye occluded face to remove recognizable areas of a face to preserve the privacy of a specific age group of audience, e.g. children, and to rank automatically content offered depending on the age. Age is estimated using the eight age groups, ranging from 0 to 100 years, of the Adience dataset.

In this work, we present an evaluation of SSR-Net models in minors, with and without eye occluded faces, and propose a training strategy to improve age estimation performance.

3 METHODOLOGY

We proposed a two-fold training strategy to improve the age estimation in minors and young adults with and without eye occlusion, see Figure 1.

First, given a set of non-occluded face images of minors and young adults, a set of occluded images is created artificially by covering the face eye area through a mask to simulate the observed conditions on CSEM. Second, both sets of images are combined into one and used it to build an age estimation model that is robust against eye occlusion.

3.1 Non-occluded Dataset

We collected images of minors and young adults from five different datasets: IMDB-WIKI (Rothe et al., 2015), APPA-REAL (Agustsson et al., 2017), AgeDB (Moschoglou et al., 2017), UTKFace (Zhang et al., 2017), Diversity in Faces, IBM (DiF) (Grd and Bača, 2016). Table 1 presents a summary of the content of each dataset and their distribution is shown in Figure 2. We manually inspected the datasets and removed images with an incorrect age label or without any human face. As a result, we gathered a balanced dataset with 130000 minor and young adult images —5000 images by age— for further training and test of age estimation models.

3.2 Creation of Eye Occluded Images

We created eye occluded face images of minors and young adults from existing non-occluded face dataset by adding a rectangular black mask over the face eyes area. Given a face image, first, the location of the right and the left eye is identified with the Multi-Task Cascade CNN (MTCNN) (Zhang et al., 2016) method. Second, the slope of the line that connects these points is computed and used to determine the position and the dimensions of the rectangular mask to be drawn. The rectangle height corresponds to the 25% of the height of the bounding box that contains the minor face. The rectangle width corresponds to the 95% of the width of the bounding box containing the minor or the young adult face.

3.3 Building of the Age Estimation Model

A training set is formed by non-occluded face images —selected from the dataset described previously— and their corresponding eye occluded version created artificially. Images are resized to 64×64 pixels and used to fine-tune a pre-trained SSR-Net model (Yang et al., 2018). We selected the SSR-Net method due to its age estimation performance and size compact models which can be used in any hardware regardless of their memory capability. SSR-Net models were trained considering an age interval of $[0,25]$ years at most for 90 epochs, i.e. the number of times the network sees the entire training set. The 80% of the training set was used to fine-tune the network and the remaining 20% was used to monitor overfitting (validation set). The model with the highest performance on the validation set was kept and used as age estimator.

4 EXPERIMENTAL SET-UP

We evaluated the performance of age estimation models using (i) non-occluded, (ii) eye occluded and (iii) a combination of both types of minor and young adult facial images. We assessed the impact of the size of the training set as well as the SSR-Net pre-trained models used to create the models by comparing the performance obtained with models trained using four datasets varying in size: 6500 images —250 images by age—, 13000 images —500 images by age—, 26000 —1000 images by age—, and 130000 images —5000 images by age—, and pre-trained models with IBMD and MORPH datasets. Note that IMDB and MORPH are unbalanced datasets that contain few minor examples. IMDB labels are very noisy while

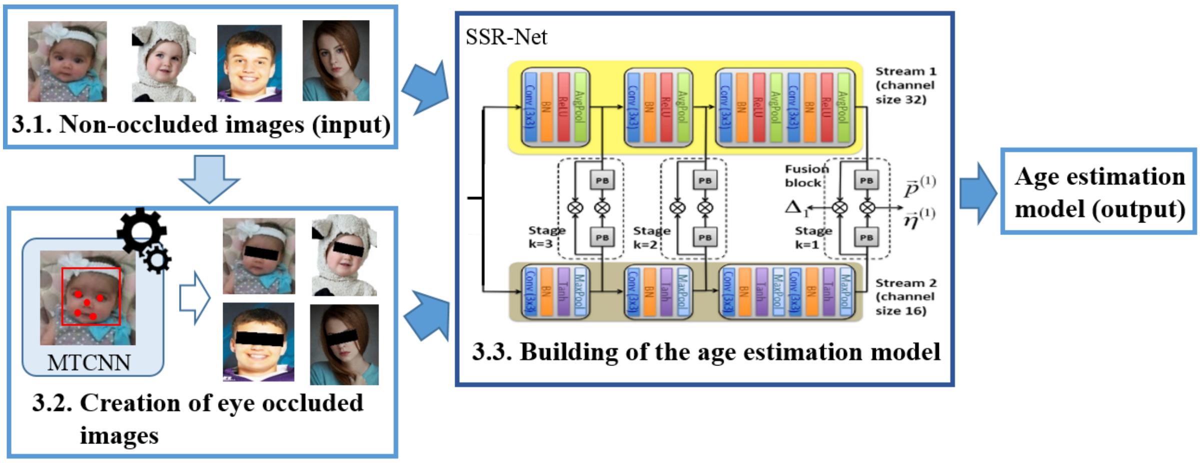


Figure 1: Proposed strategy to train age estimation models in minors and young adults with non-occluded and eye occluded images.

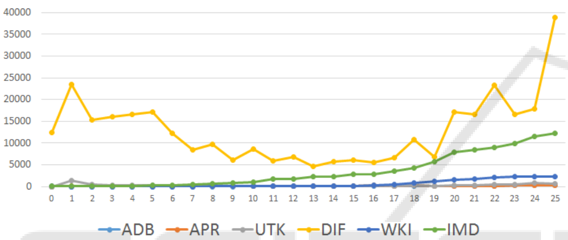


Figure 2: Minors and young adults age distribution per dataset.

MORPH only includes subjects eighteen years old. In addition, we compared the performance of our proposal against a DEX model, a bulky age estimator, trained considering an age interval of $[0,25]$ years. The training sets were randomly split into training and test set, containing 80% and 20% of the whole set, respectively.

Models were evaluated using the *MAE* and the Accuracy (*Acc*). The *MAE* corresponds to the average of the absolute errors between the predicted ages, *PredAge*, and the ground truth, *GtAge*. It is defined in Equation 1 as:

$$MAE = \sum_{i=1}^{n} \frac{|GtAge_i - PredAge_i|}{n} \tag{1}$$

The *Acc* is computed by considering five age groups: [0-5], [6-10], [11-15], [16-17], and [18-25], as the mean accuracy across all them. These groups were defined in (Anda et al., 2019) based on the report "Criminal networks involved in the trafficking and exploitation of underage victims in the European Union" of 2018.

Additionally, the improvement (*Impv*) of age estimation models in terms of *MAE* and *Acc* is analyzed. The improvement is defined as the relative error between the performance of an age estimator built using a baseline training conditions, *A*, and another one, *B*, as follows in Equation 2:

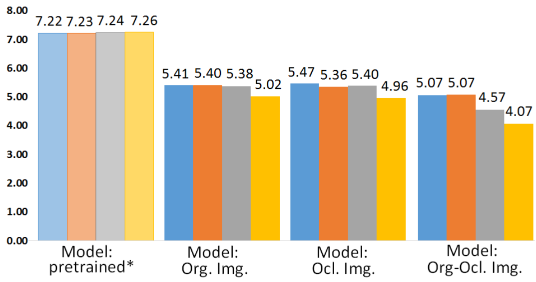
$$Impv = \left({}^{(A-B)}/_{A}\right) \times 100 \tag{2}$$

The interpretation of the improvement depends on the evaluation metric. Positive values of *Impv* in *MAE* indicate that B outperforms the baseline model, A. While, negative values of *Impv* in *Acc* imply that B performs better than A.

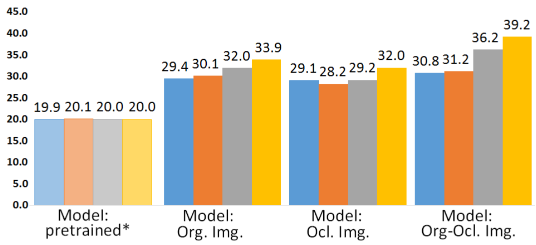
5 EXPERIMENTAL RESULTS

Figure 3 shows the average *MAE* and *Acc* values computed on test sets with non-occluded and eye occluded face images by age SSR-Net estimation models (fine-tuned) from pre-trained models with IMDB and MORPH datasets, respectively. In general, the use of large training sets improved the performance of age estimators. The best performance —*MAE* of 4.07 and *Acc* of 39.2%— is observed in models built with the larger dataset —130000 images—, including both non-occluded and eye occluded facial images, through fine-tuning of IMDB pre-trained models. This model outperformed the results obtained with the pre-trained SSR-Net models from IMDB and MORPH datasets. The pre-trained SSR-Net models from IMDB dataset achieved a *MAE* of 7.26 and an *Acc* of 20.0% while the pre-trained SSR-Net models from MORPH yielded a *MAE* of 6.81 and an *Acc* of 19.7%. Figure 4 illustrates the age predicted using the best age estimation model.

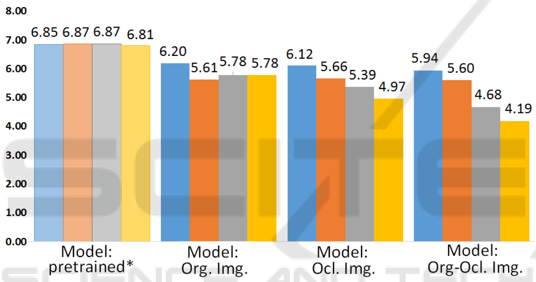
Afterward, we analyzed the performance of the SSR-Net models using non-occluded and eye occluded images, independently. Figure 5 presents the



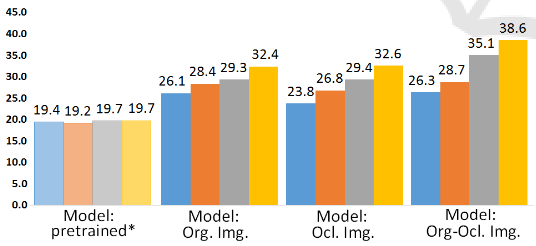
a) Avg. MAE values for IMDB dataset.



b) Avg. Acc values for IMDB dataset.



c) Avg. MAE values for MORPH dataset.



d) Avg. Acc values for MORPH dataset.

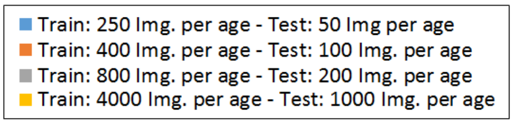


Figure 3: *MAE* and *Acc* values yielded on test sets with and without eye occluded images by SSR-Net age estimators fine-tuned from IMDB and MORPH dataset. *Pre-trained models reported in (Yang et al., 2018) built with unbalanced datasets.

Real age: 8 Predict age: 8 | Real age: 10 Predict age: 16 | Real age: 21 Predict age: 23

Real age: 8 Predict age: 8 | Real age: 10 Predict age: 17 | Real age: 21 Predict age: 22

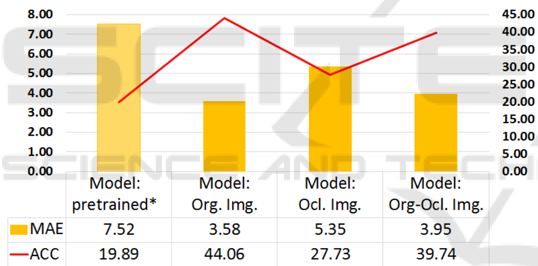
Figure 4: Illustration of ages estimated for non-occluded and eye occluded facial images with the best SSR-Net age estimation model. Images taken from UTKFace dataset (Zhang et al., 2017).

MAE and *Acc* values obtained with age estimators fine-tuned with 130000 images from pre-trained models with IMDB dataset. Models trained with these conditions yielded the best overall performance (see Figure 3). Furthermore, Table 2 depicts the *MAE* and the *Impv* of *MAE* values for the 12 SSR-Net models built using the training conditions described in Section 4, and Table 3 shows the Acc and the *Impv* of *Acc* values for those models. Similar to the reported in (Yadav et al., 2014), results showed that the use of eye occluded images, in most of the cases, do not affect negatively the performance of age estimators. Presumably, because eye information in facial images of minors and young adults does not provide the most significant information during the age estimation process.

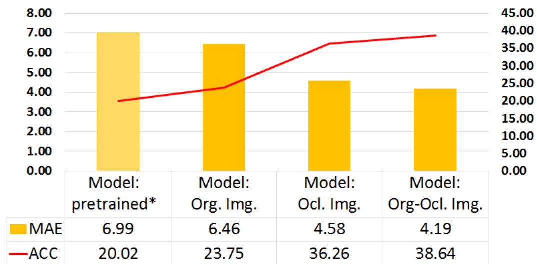
Besides, the use of balanced training sets improved the performance of age estimators in minors and young adults. The best *MAE* is obtained with age estimators built with non-occluded —*MAE* of 3.58— or occluded —*MAE* of 4.22— facial images by fine-tuning pre-trained models from IMDB and MORPH datasets, respectively. However, models trained only using non-occluded facial images perform poorly on eye occluded images —*MAE* of 7.93— despite the presence of some cases of eye occlusion, e.g. the use of glasses. Indicating that these models are not robust against artificial eye occlusion. Moreover, models built only using occluded face images performance better with non-occluded images —*MAE* of 6.46— but the error is higher in comparison to models trained with only non-occluded ones. Suggesting that the information provided by facial regions as nose or mouth may be more relevant than eye information during the

Table 2: *MAE* and *Impv* values for age estimation models fine-tuned from MORPH and IMDB dataset using training sets with images: non-occluded (Org), eye occluded (Ocl), and a combination of both types of facial images (Org-Ocl). Lower *MAE* values mean a better performance. Higher positive values of *Impv* indicate an improvement in *MAE* values regarding the baseline model. The best *MAE* and *Impv* in *MAE* values are highlighted in bold.

Model	Total img. per age		MORPH dataset				IMDB dataset			
			MAE MAE		*Impv. MAE (%)*		*MAE* Test		*Impv. MAE (%)*	
	Train	Test	Org.	Ocl.	Org.	Ocl.	Org.	Ocl.	Org.	Ocl.
Pre-train 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-tune Org. Img.	200	50	5.37	7.04	–	–	4.56	6.25	–	–
	400	100	4.40	6.82	17.99	3.10	4.27	6.53	6.48	-4.41
	800	200	4.24	7.32	20.94	-3.97	4.13	6.64	9.44	-6.10
	4000	1000	**3.63**	7.93	32.38	-12.65	**3.58**	6.46	**21.49**	-3.33
Fine-tune Ocl. Img.	200	50	6.57	5.67	–	–	5.73	5.22	–	–
	400	100	6.03	5.29	8.23	6.83	5.63	5.08	1.59	2.81
	800	200	5.80	4.97	11.66	12.44	5.72	5.07	0.02	2.84
	4000	1000	5.71	**4.22**	13.13	**25.54**	5.35	4.58	6.57	12.39
Fine-tune Org. - Ocl. Img.	200	50	6.08	5.81	–	–	5.00	5.13	–	–
	400	100	6.01	5.19	1.04	10.65	4.91	5.23	1.86	-1.92
	800	200	4.61	4.75	24.15	18.34	4.47	4.66	10.67	9.15
	4000	1000	3.93	4.44	**35.26**	23.57	3.95	**4.19**	21.02	**18.29**



a) Models tested on Org. images.



b) Models tested on Ocl. images.

Figure 5: *MAE* and *Acc* values obtained by age estimators fine-tuned from IMDB dataset using test sets of non-occluded (Org) and eye occluded images (Ocl). *Pre-trained models reported in (Yang et al., 2018) built with IMDB dataset.

age estimation of minors and young adults. Lastly, the models created using both, non-occluded and eye occluded images, are more stable and have similar performance for both evaluation conditions. In this case, the best *MAE* for non-occluded (3.95) and occluded (4.19) facial images is achieved with age estimators fine-tuned from pre-trained models with IMBD dataset.

Similar to the observed *MAE* values, the accuracy increases with large training sets, although this increase is not directly proportional to the number of training examples, see Table 3. The best accuracy for non-occluded —*Acc* of 44.06%— and eye occluded —*Acc* of 39.40%— images is attained with age estimators fine-tuned from pre-trained models using IMDB and MORPH datasets, respectively.

Finally, we compared the results obtained with the best SSR-Net model against a DEX model trained with 130000 images including non-occluded and eye occluded images. This dataset allowed to achieve the best overall performance for the built SSR-Net models (see Figure 3). Figure 6 presents the *MAE* and *Acc* values obtained with both age estimators. Results showed that the proposed age estimator based on the SSR-Net model outperformed the DEX model during the analysis of non-occluded and eye occluded facial images —*MAE* of 6.5 and *Acc* of 19.2— with an advantage in the size of the model. Our age estimator is very compact —with a size lower than 1 MB— in comparison to the DEX model based on VGG-16 architecture with a size larger than 500 MB. Hence, it can be used in any hardware despite their memory ca-

Table 3: *Acc* and *Impv* values for age estimation models fine-tuned from MORPH and IMDB dataset using training sets with images: non-occluded (Org), eye occluded (Ocl), and a combination of both types of facial images (Org-Ocl). Higher *Acc* values mean a better performance. Lower negative values of *Impv* indicate an improvement in *Acc* values against the baseline model. The best *Acc* and *Impv* in *Acc* values are highlighted in bold.

Model	Total img. per age		MORPH dataset				IMDB dataset			
			Acc Test (%)		Impv. *Acc* (%)		*Acc* Test (%)		Impv. *Acc* (%)	
	Train	Test	Org.	Ocl.	Org.	Ocl.	Org.	Ocl.	Org.	Ocl.
Pre-train MORPH	–	50	19.00	19.84	–	–	19.77	20.08	–	–
	–	100	18.39	19.99	–	–	19.81	20.44	–	–
	–	200	18.92	20.40	–	–	19.89	20.18	–	–
	–	1000	19.28	20.14	–	–	19.89	20.02	–	–
Fine-tune Org. Img.	200	50	30.21	22.06	–	–	35.73	23.15	–	–
	400	100	37.23	19.52	-23.25	11.50	38.07	22.19	-6.54	4.15
	800	200	38.44	20.25	-27.23	8.22	40.63	23.32	-13.70	-0.71
	4000	1000	**43.77**	20.95	-44.90	5.03	**44.06**	23.75	**-23.30**	-2.61
Fine-tune Ocl. Img.	200	50	20.71	26.81	–	–	26.40	31.72	–	–
	400	100	24.31	29.38	-17.35	-9.58	24.64	31.83	6.65	-0.32
	800	200	24.92	33.81	-20.33	-26.08	26.38	31.97	0.05	-0.78
	4000	1000	25.81	**39.40**	-24.63	**-46.96**	27.73	36.26	-5.04	-14.32
Fine-tune Org. - Ocl. Img.	200	50	26.37	26.28	–	–	32.30	29.25	–	–
	400	100	26.50	31.0	-0.49	-17.95	33.02	29.37	-2.23	-0.40
	800	200	35.62	34.54	-35.09	-31.41	37.85	34.60	-17.20	-18.30
	4000	1000	40.56	36.58	**-53.83**	-39.20	39.74	**38.64**	-23.06	**-32.09**

pability which is desirable in forensic applications as child detection on CSEM.

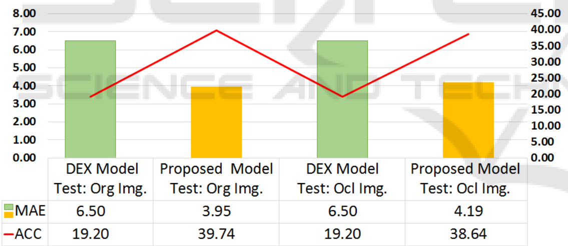


Figure 6: *MAE* and *Acc* values obtained from test sets of non-occluded (Org) and eye occluded images (Ocl) by the proposed SSR-Net model fine-tuned from IMDB dataset and the DEX model.

6 CONCLUSIONS

In this work, we presented a strategy to improve the estimation of age in minors and young adults with artificially eye occluded faces by fine-tuning SSR-Net models through a combination of non-occluded and occluded images. This kind of occlusion is frequent in CSEM to hide the identity of victims. Results showed that the proposed strategy allows building age estimation models in minors and young adults robust against eye occlusion —average *MAE* of 4.07 and *Acc* of 39.2% for non-occluded and occluded facial images— that outperformed models SSR-Net pre-trained with unbalanced set as MORPH —average *MAE* of 6.81 and *Acc* of 19.7% for non-occluded and eye occluded images—. Furthermore, our age estimator performance better than a DEX model trained using a dataset including non-occluded and occluded images —*MAE* of 6.5 and *Acc* of 19.2% for non-occluded and eye occluded images—. Finally, the SSR-Net based estimators are compact models —lower than 1 MB— in comparison to the DEX age estimator —more than 500 MB— allowing its use in any device without regarding their memory with a real-time performance, which is required in forensic applications.

As future work, an ensemble of classifiers will be used to reduce estimation errors by combining the best age estimators models trained with non-occluded or eye occluded facial images.

ACKNOWLEDGEMENTS

This work was supported by the framework agreement between the Universidad de León and INCIBE (Spanish National Cybersecurity Institute) under Addendum 01. We acknowledge NVIDIA Corporation with the donation of the TITAN Xp and Tesla K40 GPUs used for this research. 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.

REFERENCES

Agustsson, E., Timofte, R., Escalera, S., Baró, X., Guyon, I., and Rothe, R. (2017). 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*, pages 1–12.

Al-Nabki, M. W., Fidalgo, E., Alegre, E., and Fernández-Robles, L. (2019). Torank: Identifying the most influential suspicious domains in the tor network. *Expert Systems with Applications*, 123:212 – 226.

Alrjebi, M., Pathirage, N., Liu, W., and Li, L. (2017). Face recognition against occlusions via colour fusion using 2d-mcf model and src. *Pattern Recognition Letters*, 95:1339–1351.

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

Angulu, R., Tapamo, J. R., and Adewumi, A. O. (2018). Age estimation via face images: a survey. *EURASIP Journal on Image and Video Processing*, 2018(1):42.

Antipov, G., Baccouche, M., Berrani, S., and Dugelay, J. (2016). Apparent age estimation from face images combining general and children-specialized deep learning models. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 801–809.

Biswas, R., González-Castro, V., Fidalgo, E., and Chaves, D. (2019). Boosting child abuse victim identification in forensic tools with hashing techniques. In *V Jornadas Nacionales de Investigación en Ciberseguridad (JNIC)*, volume 1, pages 344–345.

Carcagnì, P., Coco, M. D., Cazzato, D., Leo, M., and Distante, C. (2015). A study on different experimental configurations for age, race, and gender estimation problems. *EURASIP Journal on Image and Video Processing*, 2015:1–22.

Cen, F. and Wang, G. (2019). Dictionary representation of deep features for occlusion-robust face recognition. *IEEE Access*, 7:26595 – 26605.

Chaves, D., Fidalgo, E., Alegre, E., and Blanco, P. (2019). Improving speed-accuracy trade-off in face detectors for forensic tools by image resizing. In *V Jornadas Nacionales de Investigación en Ciberseguridad (JNIC)*, pages 1–2.

Chen, S., Zhang, C., Dong, M., Le, J., and Rao, M. (2017). Using ranking-cnn for age estimation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 742–751.

Galusha, A., Dale, J., Keller, J. M., and Zare, A. (2019). Deep convolutional neural network target classification for underwater synthetic aperture sonar imagery. In Bishop, S. S. and Isaacs, J. C., editors, *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIV*, volume 11012, pages 18 – 28.

Gangwar, A., Fidalgo, E., Alegre, E., and González-Castro, V. (2017). 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)*, pages 37–42.

Grd, P. and Bača, M. (2016). Creating a face database for age estimation and classification. In *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1371–1374.

Hase, N., Ito, S., Kaneko, N., and Sumi, K. (2019). Data augmentation for intra-class imbalance with generative adversarial network. In *Fourteenth International Conference on Quality Control by Artificial Vision*, volume 11172, pages 34 – 41.

Min, R., Hadid, A., and Dugelay, J.-L. (2011). Improving the recognition of faces occluded by facial accessories. In *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*, pages 442 – 447.

Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., and Zafeiriou, S. (2017). Agedb: The first manually collected, in-the-wild age database. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1997–2005.

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

Rothe, R., Timofte, R., and Gool, L. V. (2015). Dex: Deep expectation of apparent age from a single image. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 10–15.

Yadav, D., Singh, R., Vatsa, M., and Noore, A. (2014). Recognizing age-separated face images: Humans and machines. *PLoS ONE*, 9(12):1–22.

Yang, T.-Y., Huang, Y.-H., Lin, Y.-Y., Hsiu, P.-C., and Chuang, Y.-Y. (2018). 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)*, pages 1–7.

Ye, L., Li, B., Mohammed, N., Wang, Y., and Liang, J. (2018). Privacy-preserving age estimation for content rating. In *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6.

Yi, D., Lei, Z., and Li, S. (2015). Age estimation by multi-scale convolutional network. In *Conference: Asian Conference on Computer Vision*, volume 9005, pages 144–158.

Zhang, C., Liu, S., Xu, X., and Zhu, C. (2019). C3AE: exploring the limits of compact model for age estimation. *CoRR*, abs/1904.05059:1–10.

Zhang, K., Liu, N., Yuan, X., Guo, X., Gao, C., Zhao, Z., and Ma, Z. (2019). Fine-grained age estimation in the wild with attention lstm networks. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–12.

Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503.

Zhang, Z., Song, Y., and Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4352–4360.

Zhao, Z.-Q., ming Cheung, Y., Hu, H., and Wu, X. (2016). Corrupted and occluded face recognition via cooperative sparse representation. *Pattern Recognition*, 56:77–87.