





# A Data Augmentation Strategy for Improving Age Estimation to Support CSEM Detection

Deisy Chaves<sup>1</sup><sup>a</sup>, Nancy Agarwal<sup>2</sup><sup>b</sup>, Eduardo Fidalgo<sup>1</sup><sup>c</sup>, and Enrique Alegre<sup>1</sup><sup>d</sup>

<sup>1</sup>*Department of Electrical, Systems and Automation, Universidad de León, León, Spain*

<sup>2</sup>*GNIOT, Greater Noida, Uttar Pradesh, India*

*nancy.iot@gniot.net.in, {dchas, efidf, enrique.alegre}@unileon.es*

**Keywords:** Age Estimation, Data Augmentation, Generative Adversarial Networks, Facial Occlusion, CSEM

**Abstract:** Leveraging image-based age estimation in preventing Child Sexual Exploitation Material (CSEM) content over the internet is not investigated thoroughly in the research community. While deep learning methods are considered state-of-the-art for general age estimation, they perform poorly in predicting the age group of minors and older adults due to the few examples of these age groups in the existing datasets. In this work, we present a data augmentation strategy to improve the performance of age estimators trained on imbalanced data based on synthetic image generation and artificial facial occlusion. Facial occlusion is focused on modelling as CSEM criminals tend to cover certain parts of the victim, such as the eyes, to hide their identity. The proposed strategy is evaluated using the Soft Stagewise Regression Network (SSR-Net), a compact size age estimator and three publicly available datasets composed mainly of non-occluded images. Therefore, we create the Synthetic Augmented with Occluded Faces (SAOF-15K) dataset to assess the performance of eye and mouth-occluded images. Results show that our strategy improves the performance of the evaluated age estimator.


## 1 INTRODUCTION


Facial age estimation is defined as the problem of automatically predicting the real age of a person (i.e. the number of years a person has been alive) or the age group (i.e. pre-pubescent, pubescent, or adult) from an image. In recent years, image-based age estimation has become an emerging research area due to its contributions to several real-world applications such as human-computer interaction (Angulu et al., 2018), (Osman and Yap, 2018), surveillance monitoring (Fu et al., 2010), and age-invariant face recognition (Angulu et al., 2018). Moreover, some studies showed the potential of age prediction to detect victims (pre-pubescent and pubescent) aid the analysis of CSEM (Anda et al., 2020), (Gangwar et al., 2021), (Grubl and Lallie, 2022). In (Gangwar et al., 2021), authors decompose the Child Sexual Abuse (CSA) detection problem into two modules: pornographic detection and age-group classification as minor or adult.


Deep learning algorithms are state-of-the-art for


designing the age estimators from facial images (Rothe et al., 2015), (Shen et al., 2018), (Yang et al., 2018), (Agbo-Ajala and Viriri, 2021), (Wang et al., 2022). However, their performance is hindered because the existing labelled datasets are limited in size. Constructing accurate estimators requires many labelled face images for every individual age or age group under consideration. Thus, the imbalance issue in these datasets leads to an overfitted model and prevents its generalisation capability from reaching its full potential. There are several strategies to handle small and imbalanced datasets, e.g. label distribution learning, data re-sampling, redefining class-balance losses, and transfer learning (Geng et al., 2014), (Kang et al., 2019), (Yan et al., 2022). These approaches mainly focus on designing the model with an equal probability of learning discriminative features of minority and majority class samples.

Instead of working at the algorithm level, another widely adopted strategy to deal with imbalanced dataset issues is adding more samples of minority classes to the training (Liu et al., 2020), (Zhong et al., 2020), (Zhang and Bao, 2022). However, finding the expected number of real images labelled with chronological ages is challenging and time-consuming for the age estimation problem. Therefore, in this work,

<sup>a</sup> <https://orcid.org/0000-0002-7745-8111>

<sup>b</sup> <https://orcid.org/0000-0003-4392-0520>

<sup>c</sup> <https://orcid.org/0000-0003-1202-5232>

<sup>d</sup> <https://orcid.org/0000-0003-2081-774X>

we presented a data augmentation approach to generate target samples from the existing facial images artificially. To the best of our knowledge, StyleGAN is a state-of-the-art method for synthesising high-resolution images (Karras et al., 2019), (Karras et al., 2020). In this work, StyleGANv2 and High-Resolution Age Editing Face (HREAF) (Yao et al., 2021) have been adopted.

Since we are focused on designing an age estimation system that could assist in CSEM investigations, occlusion plays an essential role as the criminals tend to cover certain parts of the victim, such as eyes and mouth, to hide their identity (Gangwar et al., 2017). However, only a few works have studied the effect of facial occlusion during age estimation (Yadav et al., 2014), (Ye et al., 2018), (Cai and Liu, 2021), despite regions such as eyes and mouth corners are important features during the age prediction. In our previous work, we explored the use of eye occlusion and improved the estimation of subjects between 0 and 25 years old. In this work, we use artificial eye and mouth occluded images with different levels of transparency for designing an age estimator focused on minor and adult age groups that are robust to face occlusion. Moreover, the Soft Stagewise Regression Network (SSR-Net) (Yang et al., 2018) architecture is applied to build a compact size age estimator. Furthermore, for evaluation purposes, we generate two artificial occluded versions of the APP-Real (Agustsson et al., 2017) and the FG-Net datasets (Fu et al., 2014). Hence, the main contributions of this work are summarised as follows:

- A data augmentation strategy for age estimation using facial occlusion and synthetic images aid to support CSEM detection.
- The Synthetic Augmented with Occluded Faces (SAOF-15K)<sup>1</sup>, created for evaluating purposes, which is composed of the eye and mouth-occluded facial versions of the FG-Net and the APP-Real dataset.
- Study of the effect of facial occlusion during age estimation.
- The outcome of this study will be applied to the European project Global Response Against Child Exploitation (GRACE).

---

<sup>1</sup><https://gvis.unileon.es/dataset/synthetic-augmented-with-occluded-faces-saof/>

## 2 RELATED WORKS

### 2.1 Age Estimation Models

Age estimation is mainly addressed as a regression problem to learn a non-linear function for mapping the facial features of the images to their chronological age. However, learning a global ageing mapping is challenging due to the inhomogeneous nature of faces observed as intra-class facial appearance variations (i.e. significant changes in the face attributes of different persons at the same age) (Shen et al., 2018). As a solution, some studies (Shen et al., 2018), (Shen et al., 2019) presented modelling of multiple local regressors to learn the variations in the ageing pattern, e.g. random forest-based regression (Montillo and Ling, 2009) and CNN-based tree regression (Shen et al., 2018). In some studies, the age estimation is treated as a label distribution learning (LDL) problem by modelling the correlation pattern between the neighbouring ages based on the assumption that facial images of a person at adjacent ages appear similar (Geng et al., 2014), (Gao et al., 2018). Shen et al. (Shen et al., 2019) combine the concept of multiple local functions and LDL to introduce the Deep Label Distribution Learning Forests (DLDLFs) algorithm for age estimation. The DLDLF model consists of an ensemble of LDL trees employing the VGG-16 convolutional neural network architecture (Simonyan and Zisserman, 2014).

Besides, age regression can also be formulated as a classification problem by partitioning the ages into a set of discrete classes. The authors (Rothe et al., 2015) leverage the softmax function at the outer layer of the deep neural network to classify the predicted age into labels between 0 to 100 years. The final age is predicted by combining the output probabilities of the neurons via the expected value function. Inspired by the above work, the authors (Yang et al., 2018) perform multi-stage classification, where each stage corresponds to a set of age groups and refines the decision of the previous step with a finer granularity. The advantage of the multi-stage framework is that the number of output neurons is small at each stage, leading to a more compact model than (Rothe et al., 2015) without sacrificing performance. Recently, in (Shin et al., 2022) proposed a general regression algorithm called Moving Window Regression (MWR) and applied it to age estimation. MWR obtains an initial rank estimate of an input instance based on the nearest neighbour criterion. It is refined by selecting two reference instances to form a search window and estimating the relative rank within the search window iteratively. The diverse characteristics in different rank

groups are managed using a local and a global regressor.

In our work, we applied a similar stagewise methodology (Yang et al., 2018), so we built a compact size age estimator useful for machines with limited memory and computation resources.

## 2.2 Age Estimation with occluded faces

The majority of age estimation works have used fully-visible facial images as input data for prediction and, therefore, do not generalise well when certain parts of the faces are hidden, or occluded (Chaves et al., 2020). In forensic applications such as CSEM detection, age estimation from occluded images becomes crucial as the criminals tend to cover the eyes of the victim to hide their identity (Gangwar et al., 2017). Moreover, the recent global COVID-19 pandemic, where people should wear a face mask, has further urged the need to consider occlusion in modelling age estimators.

Face occlusion is commonly studied in domains like face recognition and face verification (Min et al., 2011), (Kortli et al., 2020), (Zhao et al., 2016). However, only a few studies have considered occluded faces during the age estimation (Yadav et al., 2014), (Cai and Liu, 2021), (Ye et al., 2018), (Chaves et al., 2020). Yadav et al. (Yadav et al., 2014) conducted an experiment where some regions of facial images, such as T-region, binocular region, chin portion, faces with masked eyes, and masked T-region, were shown to the participants to understand which facial section contains helpful information for age prediction. It is seen that the chin area provides sufficient clues for the age group 0 – 5, and the faces with obfuscated T-region are recognisable in the age group 6 – 10. In (Ye et al., 2018), the authors mask eye regions to build a model for age estimation with more discriminate features around the mouth and nose. Furthermore, the work (Cai and Liu, 2021) focused on occlusion caused by wearing a mask and employed a self-supervised contrastive learning framework to model the relationship between the fully-visible and masked face.

Our work is similar to (Chaves et al., 2020) where artificially eye occluded images are included during training to design an age model to support the recognition of CSEM victims showed that the model built using both occluded and non-occluded images is more stable, robust, and efficient. As an extension to this approach, we considered several levels of transparency during the occlusion and synthesised mouth occluded images, which may assist the age estimation of people wearing facial masks.

## 2.3 Data augmentation

One of the solutions to deal with the imbalanced data issue is to increase the number of minority samples in the dataset. However, finding the expected number of images labelled with chronological age is challenging and time-consuming for the age estimation problem. Data augmentation overcomes this problem by defining a variety of simple operations such as rotation, translation, scaling, flipping, and cropping (Liu et al., 2020), (Krizhevsky et al., 2012) to artificially generate different versions of images from the existing ones. In addition to geometric image transformations, synthetic instances can also be obtained by altering the brightness of the image, infusing some noise to it, or erasing its certain regions via masking technique (Liu et al., 2020), (Zhong et al., 2020).

Besides these simple transformation principles, data augmentation also consists of complex transformations such as the work of Oliveira et al. (de Pontes Oliveira et al., 2016) where a high-level deformation function is proposed to induce variance in specific facial features (e.g., chin, nose, and jaw) by detecting fiducial points (de Pontes Oliveira et al., 2016). Nowadays, the use of generative models such as GAN (Generative Adversarial Network) is one promising approach in data augmentation that relies on neural networks to create artificial data (Wang et al., 2018). The primary motivation for using these models in age estimation is to generate a modified version of an existing facial image with a younger or older appearance (Golubović and Risojević, 2021). In (Georgopoulos et al., 2020), the authors integrated the style transfer optimising technique with GAN, allowing them to modify the existing image based on the features of another image. The framework assists them in creating a variety of faces of extreme ages (i.e. very old/young).

StyleGAN is nowadays the state-of-the-art method for synthesising high-resolution images (Karras et al., 2019), (Karras et al., 2020). Conventional GAN modelling struggles with the feature entanglement issue, where a small change in latent input space affects multiple features and their capability to produce images with finer details. In contrast, styleGAN allows better control and offers us to modify the details from high-level features such as face shape to minute alterations (e.g., hair colour, wrinkles) in the image by comprising the methodology that learns progressively from low-level to high-resolution details. Since GAN-based models are complex and known for producing artefacts, the study (Yao et al., 2021) presented the encoder-decoder-based system to synthesise images for

performing ageing/de-ageing with fewer artefacts and high resolution, i.e., High-Resolution Face Age Editing (HRFAE). We employ HRFAE for creating the older images and styleGANv2 (Karras et al., 2020) for obtaining the images of the minor group. styleGANv2 also solves the artefacts problem in images caused by the styleGAN.

### 3 METHODOLOGY

We proposed a two-step training strategy to improve the performance of age estimation models built with imbalanced data on facial images with and without occlusion; see Figure 1. First, given a set of non-occluded facial images, data is augmented by creating a set of occluded eye and mouth images and adding masks to simulate the observed conditions on CSEM. A collection of synthetic images of minors and elders are generated using GAN-based methods to increase the samples with low frequency on the imbalance datasets. Second, artificial occluded images and synthetic generated ones are combined into one set to build an age estimation model robust against facial occlusion.

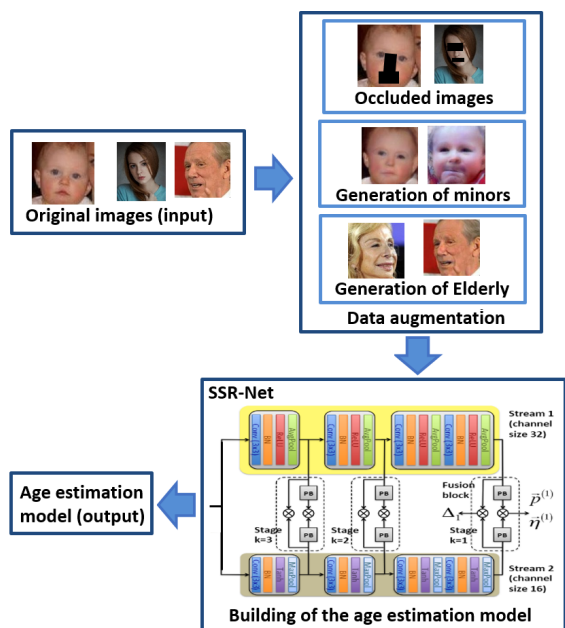


Figure 1: Strategy for training age estimation models with imbalanced datasets.

#### 3.1 Non-occluded dataset as input

We use facial images of the Mivia Age Dataset provided last year for the Guess The Age contest (Greco,

2021) as input to build the age estimators. This dataset includes 575.073 images of subjects aged between 1 and 81 years old. Apart from being one of the largest publicly available datasets with age annotations, the Mivia Age Dataset is highly imbalanced on minors and the elderly making it ideal to evaluate our proposed data augmentation strategy.

#### 3.2 Augmentation of facial images

We increase the number of images in the training dataset using three strategies: (i) creation of facial occluded images by adding a rectangular mask over the eye and the mouth regions to simulate what offenders do to difficult victims identification, (ii) generation of synthetic images of minors using transfer style (StyleGANv2 (Karras et al., 2020)), and (iii) generation of elder faces using an ageing method (HREAF (Yao et al., 2021)).

##### 3.2.1 Creation of facial occluded images

Given a non-occluded facial image, first, the Multi-Task Cascade CNN (MTCNN) (Zhang et al., 2016) method is used to identify 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 connects these points is obtained and used to determine the position and the dimensions of the rectangular mask to be drawn over the eyes and 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.

Finally, several degrees of transparency were considered to draw the black rectangles, as illustrated in Figure 2.

Age	Original Images	Occluded Images			
		50% transparency		0% of transparency	
2					
14					
45					

Figure 2: Eye and mouth occluded facial images created artificially with different levels of transparency.

### 3.2.2 Generation of synthetic minor images

We use StyleGANv2 to generate minor images between 1 and 17 years old. StyleGANv2 is an extension of StyleGAN to reduce the blob-shaped artefacts in the generated images. StyleGAN and StyleGANv2 create images through the incremental expansion of discriminator and generator models from small to large images during the training process.

To generate a new minor image, first, we select two facial images of minors and generate latent vectors using a StyleGANv2 trained from scratch using 2000 images per age group of the Mivia Age Dataset – when available. We choose images of minors of the same age and gender to keep the facial features representing a particular age. Second, we combine the latent vectors of two minor images using linear interpolation. Third, we generated the new minor image from the latent vector using a truncation parameter set experimentally to 0.7. We generate around 4000 images per age group; see Figure 3.

















Age	Original Image		Generated Image	
3				
9				
12				
15				

Figure 3: Minor facial images generated with StyleGANv2 per age group.

### 3.2.3 Generation of synthetic elder images

We use HRFAE to generate elder images between 68 and 81 years old because the number of images on the Mivia Age Dataset for this age group were not enough to successfully train a StyleGANv2 model. HRFAE is an encoder-decoder architecture for face age editing that uses a latent space representation containing the face identity, a feature modulation layer and its age.

We selected an adult image to generate a new elder image and aged it using the HRFAE model trained from scratch. As a result, we generate around 4000 images per age; see Figure 4.













Ageing Years	Original Images		Generated Images	
68 - 73				
69 - 76				
69 - 81				

Figure 4: Generation of aged facial images using HRFAE per age group.

## 3.3 Building of the age estimation model

We split the Mivia Age dataset into training and test subsets, containing the 80% and 20% of all images, respectively. We augmented the training dataset by combining the non-occluded faces of the Mivia Age dataset, their corresponding eye and mouth occluded version created artificially, and the generated images of minors and the elderly using StyleGANv2 and HRFAE (see section 3.2).

Images on the augmented training dataset are resized to  $64 \times 64$  pixels and used to build SSR-Net models (Yang et al., 2018) from scratch. We select a low image resolution ( $64 \times 64$  pixels) and the SSR-Net model for age estimation because it allows us to construct compact-size models which can be used in any hardware including mobile devices regardless of their memory capacity. SSR-Net primarily focuses on reducing models' size by classifying a small number of classes within the age group and refining them at each stage.

We built the SSR-Net models on a 12GB Nvidia TITAN Xp using a loss function focus on the Mean Absolute Error (MAE) value and the Adam optimiser with a decay learning rate varying between 30 and 60. Models were trained until a maximum number of 180 epochs using an early stop strategy with a patience parameter of 10 epochs.

## 4 EXPERIMENTAL EVALUATION

We evaluated the performance of the SSR-Net models using the MAE computed on the entire test set, the MAE values per age intervals obtained considering eight range intervals between 1 and 81 years, and the standard deviation ( $\sigma$ ) of these MAE per age intervals.

First, we assess the effect of using eye and mouth-occluded images during the training of SSR-Net mod-

els. For this, we compared the performance of models built from scratch following the procedure described in Section 3.3 with several training sets composed of artificially occluded facial images using black masks with several degrees of transparency: (a) original images (Org. Img.); (b) original images and eye occluded images (Org. Img & Eye Ocl. Img.) with a degree of transparency of 90%, 75%, 50%, 25% and 0% (solid mask); (c) original images and mouth occluded images (Org. Img & Mouth Ocl. Img.) with a degree of transparency of 90%, 75%, and 0% (solid mask); (d) original images with eye and mouth occluded images (Org. & Ocl. Img.); (e) original images with eye and mouth occluded images and synthetic images generated using GANs models (Org. & Ocl. & GANs Img).

Table 1 condenses the general MAE, the MAE per age interval, and the MAE standard deviation obtained for the five evaluated configurations. The data augmentation using eye and mouth-occluded images effectively improves the performance during the age estimation of minors and the elderly (MAE of 3,30 and standard deviation,  $\sigma$ , of 4,79). Moreover, the use of synthetic images allows the building of more stable models with the lowest MAE standard deviation ( $\sigma$  of 3,71) in comparison to the model trained only with original images ( $\sigma$  of 5,64) and occluded ones ( $\sigma$  of 4,79). Thus, the best SSR-Net model for estimating the age is built using the combination of original, occluded and synthetic images; it achieves an MAE of 3,31 and  $\sigma$  of 3,71 on the Mivia Age Dataset.

Besides, we assess the performance of the best age estimation model on occluded conditions using the testing set of the Mivia Age dataset (20 % of the data) and the reference datasets, APP-Real (Agustsson et al., 2017), and FG-Net (Fu et al., 2014). FG-Net is composed of 1002 facial images of subjects between 0-66 years, and APP-Real comprises 7591 facial images of subjects between 0-95 years. Note that after a manual inspection of the APP-Real dataset, we use for evaluation 6884 images from this dataset. Moreover, we created the SAOF-15K dataset, which comprises 15772 images corresponding to two versions of the APP-Real and the FG-Net datasets (artificially occluded eyes and mouth).

Table 2 shows the general MAE, the MAE per age interval, and the MAE standard deviation. Although eye and mouth occluded images were used to train the age estimation models, results showed that the performance decreased on this type of images, especially on eye occluded ones (MAE of 6,46), in comparison to the prediction on non-occluded or original images (MAE of 3,31). This indicates that the information in the eye region is more important for

predicting age than the information in the mouth region. The same behaviour was observed on the evaluated datasets. Therefore, more robust strategies are required to improve age prediction on occluded faces.

## 5 CONCLUSIONS

In this study, we presented a data augmentation strategy to improve the estimation of age on imbalanced datasets and support the detection of CSEM. The proposal is based on the facial occlusion usually found on CSEM and synthetic image generation to increase the number of faces with few samples in the datasets.

We evaluated SSR-Net models built using the Mivia Age and the SAOF-15K datasets. We create the SAOF-15K dataset to assess the age estimation with occluded faces from the APP-Real and the FG-Net datasets by occluding the eye, and the mouth with black masks since these datasets contain mainly non-occluded faces. Results show that the best SSR-Net model for estimating the image is built using the combination of original, occluded and synthetic images. Moreover, the data augmentation strategy allows for improving the performance of non-occluded images. However, it is not robust enough to accurately predict age on the eye and mouth-occluded images. In future work, we will evaluate more robust architectures as the backbone for age estimation and an ensemble of classifiers trained only with non-occluded or occluded facial images. Also, we will assess in-painting techniques to reconstruct the images with occluded faces.

## 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.

## REFERENCES

- Agbo-Ajala, O. and Viriri, S. (2021). Deep learning approach for facial age classification: a survey of the state-of-the-art. *Artificial Intelligence Review*, 54(1):179–213.

Table 1: Effect of the data augmentation strategy on the performance of the SSR-Net models. The best values are shown in bold.

Evaluation Metric	Org. Img.	Org. Img. & Eye Ocl. Img.					Org. Img. & Mouth Ocl. Img.			Org. & Ocl. Img.	Org. & Ocl. & GANs Img.
		90%	75%	50%	25%	0%	90%	75%	0%		
MAE <sup>1</sup> , 1-10 yrs	18,11	<b>16,28</b>	17,99	17,5	17,26	20,54	17,77	19	17,61	16,28	<b>13,00</b>
MAE <sup>2</sup> , 11-20 yrs	5,01	<b>4,17</b>	5,04	4,70	4,40	6,73	4,62	5,78	4,75	<b>4,17</b>	4,40
MAE <sup>3</sup> , 21-30 yrs	2,77	2,74	3,37	2,90	<b>2,56</b>	4,04	2,86	3,44	2,76	2,74	2,90
MAE <sup>4</sup> , 31-40 yrs	3,47	3,17	3,44	3,24	3,75	4,10	4,05	3,55	<b>3,11</b>	<b>3,17</b>	3,23
MAE <sup>5</sup> , 41-50 yrs	3,84	3,49	3,64	3,42	5,18	3,98	5,11	3,54	<b>3,36</b>	3,49	<b>3,43</b>
MAE <sup>6</sup> , 51-60 yrs	3,61	<b>3,26</b>	3,91	3,27	5,55	3,83	5,62	3,55	3,30	3,26	<b>3,13</b>
MAE <sup>7</sup> , 61-70 yrs	5,36	3,99	5,15	<b>3,68</b>	6,81	3,89	7,68	4,78	4,28	3,99	<b>3,68</b>
MAE <sup>8</sup> , +70 yrs	9,90	6,92	8,84	5,71	10,65	<b>4,92</b>	12,75	8,26	7,18	<b>6,92</b>	7,12
MAE	3,64	<b>3,30</b>	3,78	3,35	4,30	4,20	4,51	3,77	3,31	<b>3,30</b>	3,31
$\sigma$	5,64	<b>4,79</b>	5,38	5,10	5,25	5,86	5,68	5,68	5,28	4,79	<b>3,71</b>

Table 2: Assessment of the age estimator with eye and mouth-occluded images on the Mivia Age, the APP-Real, and the FG-Net datasets. The best values are shown in bold.

Evaluation Metric	org. Img.	Mivia Age Dataset			APP-Real dataset			FG-Net dataset		
		org.	eye ocl.	mouth ocl.	org.	eye ocl.	mouth ocl.	org.	eye ocl.	mouth ocl.
MAE <sup>1</sup> , 1-10 yrs	18,11	<b>13,00</b>	28,18	18,87	31,40	32,10	<b>31,05</b>	<b>15,36</b>	31,74	25,08
MAE <sup>2</sup> , 11-20 yrs	5,01	<b>4,40</b>	12,64	8,52	<b>17,52</b>	18,75	18,55	<b>9,80</b>	21,41	17,84
MAE <sup>3</sup> , 21-30 yrs	2,77	<b>2,90</b>	7,89	4,74	<b>9,68</b>	9,84	10,38	<b>6,49</b>	12,63	11,07
MAE <sup>4</sup> , 31-40 yrs	3,47	<b>3,23</b>	4,39	4,14	5,46	<b>4,44</b>	4,99	5,94	<b>5,35</b>	5,39
MAE <sup>5</sup> , 41-50 yrs	3,84	<b>3,43</b>	4,03	5,41	<b>6,85</b>	8,41	6,93	6,56	6,45	<b>5,37</b>
MAE <sup>6</sup> , 51-60 yrs	3,61	<b>3,13</b>	6,84	6,14	<b>11,90</b>	16,01	12,63	<b>7,79</b>	9,57	9,07
MAE <sup>7</sup> , 61-70 yrs	5,36	<b>3,68</b>	11,69	7,53	<b>18,48</b>	24,92	20,85	<b>11,57</b>	19,43	14,43
MAE <sup>8</sup> , +70 yrs	9,90	7,12	18,41	13,19	<b>26,42</b>	34,58	28,95	—	—	—
MAE	3,64	3,31	6,46	5,34	12,70	13,58	13,06	11,01	22,08	18,03
$\sigma$	5,64	3,71	9,30	5,72	9,29	11,56	9,85	3,48	10,43	8,00

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.

Anda, F., Le-Khac, N.-A., and Scanlon, M. (2020). Deep-uage: improving underage age estimation accuracy to aid csem investigation. *Forensic Science International: Digital Investigation*, 32:300921.

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):1–35.

Cai, W. and Liu, H. (2021). Occlusion contrasts for self-supervised facial age estimation. In *Multimedia Understanding with Less Labeling on Multimedia Understanding with Less Labeling*, pages 1–7. Association for Computing Machinery.

Chaves, D., Fidalgo, E., Alegre, E., Jánéz-Martino, F., and Biswas, R. (2020). Improving age estimation in minors and young adults with occluded faces to fight against child sexual exploitation. In *VISIGRAPP (5: VISAPP)*, pages 721–729.

de Pontes Oliveira, Í., Medeiros, J. L. P., de Sousa, V. F., Júnior, A. G. T., Pereira, E. T., and Gomes, H. M. (2016). A data augmentation methodology to improve age estimation using convolutional neural networks. In *2016 29th SIBGRAP Conference on Graphics, Patterns and Images (SIBGRAP)*, pages 88–95. IEEE.

Fu, Y., Guo, G., and Huang, T. S. (2010). Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 32(11):1955–1976.

Fu, Y., Hospedales, T. M., Xiang, T., Yao, Y., and Gong, S. (2014). Interestingness prediction by robust learning to rank. In *ECCV*, pages 488–503.

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 2017)*, pages 37–42.

Gangwar, A., González-Castro, V., Alegre, E., and Fidalgo, E. (2021). Attm-cnn: Attention and metric learning based cnn for pornography, age and child sexual abuse (csa) detection in images. *Neurocomputing*, 445:81–104.

Gao, B.-B., Zhou, H.-Y., Wu, J., and Geng, X. (2018). Age estimation using expectation of label distribution learning. In *IJCAI*, pages 712–718.

Geng, X., Wang, Q., and Xia, Y. (2014). Facial age estimation by adaptive label distribution learning. In *2014 22nd International Conference on Pattern Recognition*, pages 4465–4470. IEEE.

- Georgopoulos, M., Oldfield, J., Nicolaou, M. A., Panagakis, Y., and Pantic, M. (2020). Enhancing facial data diversity with style-based face aging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 14–15.
- Golubović, A. and Risojević, V. (2021). Impact of data augmentation on age estimation algorithms. In *2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pages 1–6. IEEE.
- Greco, A. (2021). Guess the age 2021: Age estimation from facial images with deep convolutional neural networks. In Tsapatsoulis, N., Panayides, A., Theodoridis, T., Lanitis, A., Pattichis, C., and Vento, M., editors, *Computer Analysis of Images and Patterns*, pages 265–274.
- Grubl, T. and Lallie, H. S. (2022). Applying artificial intelligence for age estimation in digital forensic investigations. *arXiv preprint arXiv:2201.03045*.
- Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J., and Kalantidis, Y. (2019). Decoupling representation and classifier for long-tailed recognition. *arXiv preprint arXiv:1910.09217*.
- Karras, T., Laine, S., and Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410.
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119.
- Kortli, Y., Jridi, M., Al Falou, A., and Atri, M. (2020). Face recognition systems: A survey. *Sensors*, 20(2):342.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Liu, X., Zou, Y., Kuang, H., and Ma, X. (2020). Face image age estimation based on data augmentation and lightweight convolutional neural network. *Symmetry*, 12(1):146.
- 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 & Gesture Recognition (FG)*, pages 442–447. IEEE.
- Montillo, A. and Ling, H. (2009). Age regression from faces using random forests. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 2465–2468. IEEE.
- Osman, O. F. and Yap, M. H. (2018). Computational intelligence in automatic face age estimation: A survey. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 3(3):271–285.
- Rothe, R., Timofte, R., and Van Gool, L. (2015). Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 10–15.
- Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B., and Yuille, A. (2019). Deep differentiable random forests for age estimation. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):404–419.
- Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B., and Yuille, A. L. (2018). Deep regression forests for age estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2304–2313.
- Shin, N.-H., Lee, S.-H., and Kim, C.-S. (2022). Moving window regression: A novel approach to ordinal regression.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Wang, H., Sanchez, V., and Li, C.-T. (2022). Improving face-based age estimation with attention-based dynamic patch fusion. *IEEE Transactions on Image Processing*, 31:1084–1096.
- Wang, Z., Tang, X., Luo, W., and Gao, S. (2018). Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7939–7947.
- Yadav, D., Singh, R., Vatsa, M., and Noore, A. (2014). Recognizing age-separated face images: Humans and machines. *PLoS one*, 9(12):e112234.
- Yan, C., Meng, L., Li, L., Zhang, J., Wang, Z., Yin, J., Zhang, J., Sun, Y., and Zheng, B. (2022). Age-invariant face recognition by multi-feature fusion and decomposition with self-attention. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 18(1s):1–18.
- Yang, T.-Y., Huang, Y.-H., Lin, Y.-Y., Hsiu, P.-C., and Chuang, Y.-Y. (2018). Ssr-net: A compact soft stage-wise regression network for age estimation. In *Proceedings of IJCAI-18*, pages 1078–1084. IJCAI.
- Yao, X., Puy, G., Newson, A., Gousseau, Y., and Hellier, P. (2021). High resolution face age editing. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 8624–8631. IEEE.
- 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. IEEE.
- Zhang, B. and Bao, Y. (2022). Cross-dataset learning for age estimation. *IEEE Access*, 10:24048–24055.
- 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.
- Zhao, Z.-Q., Cheung, Y.-m., Hu, H., and Wu, X. (2016). Corrupted and occluded face recognition via cooperative sparse representation. *Pattern Recognition*, 56:77–87.
- Zhong, Z., Zheng, L., Kang, G., Li, S., and Yang, Y. (2020). Random erasing data augmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13001–13008.