Communication Papers of the of the 17th Conference on Computer Science and Intelligence Systems pp. 39–43
DOI: 10.15439/2022F98
ISSN 2300-5963 ACSIS, Vol. 32

Face and silhouette based age estimation for child detection system.

Tomasz Lehmann[*], Andrzej Pacut[†] and Piotr Paziewski[‡]
Research and Academic Computer Network (NASK), Warsaw, Poland
Email: [*]tomasz.lehmann@nask.pl, [†]andrzej.pacut@nask.pl, [‡]piotr.paziewski@nask.pl

***Abstract*—The problem of age estimation based on facial images is a well-known computer vision task that is widely applied in identification systems. With help of the special *Dyzurnet.pl* unit detecting Internet content, related to sexual abuse of children we slightly redefined a problem. Our *Convolutional Neural Network (CNN)* solution is focused on infants and prepubescents recognition and in the particular age ranges can be considered as *the-state-of-the-art* in children detection.**

Silhouette-based age estimation is often concentrated on the human gait or body proportions analysis. Single image age estimations on the dressed (fully or partly) body are not typically researched because of a lack of properly labeled data. In our work, we present the method used to train image preparation and the final effectiveness of age estimation of that kind.

The proposed solution is a part of the system for responding to threats to children's safety in cyberspace with particular emphasis on child pornography

I. Introduction

CHILD pornography is a form of child sexual exploitation. European law defines child pornography as any visual depiction of sexually explicit conduct involving a minor (persons younger than an age adopted in the resolution of the particular country). The advancement of artificial intelligence technology, particularly in image classification and object detection, could be applied to overcome the current flaws of child sexual abuse material. The problem of the age estimation of people detected on the single image is considered one of the most important tasks in this project. In the following paper, we show methods for face-based and silhouette-based age estimation. The author's main contributions to the age estimation problem are a novel child categorization based on their sexual maturity and the untypical methodology of database creation.

The presented approach is based on convolutional neural network *CNN* classification. *CNNs* are regularized versions of multilayer perceptions where they take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in their filters. This kind of artificial neural network has become dominant in various computer vision tasks.

To implement face age estimation in the wild, age estimation must be combined with face detection. Figure 1 illustrates face age estimation combined with face detection. In this paper, we analyze the usefulness of publicly available datasets. Also, we address the issue of the head pose in the context of face-based age estimation.

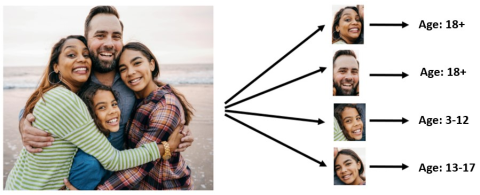

Fig. 1: Face age estimation vizualization.

A. Related Works

Age estimation based on face images is an object of many scientific papers and projects [1], [2], [3]. Most often, it is defined as an estimation of accurate age (with 1-year precision). Exception of this rule is the *oui-adience* project [4], where authors defined 8 age ranges. In our paper, we define age estimation as classification into one of 4 age ranges: 0-2, 3-12, 13-17, and 18+. These ranges were proposed by the *Dyżurnet.pl* team. The team responds to anonymous reports received from Internet users about potentially illegal material, mainly related to the sexual abuse of children. Age ranges have been established according to sexual maturity stages.

Silhouette-based age estimation is much less popular in scientific papers. Age prediction from human body images was presented in [5] where authors used their training dataset which contained just 1500 elements. In the context of child detection, the method based on the ratio between head and body heights was also proposed in [6]. The greatest number of researches touched an area of gait-based human age classification ([7], [8], [9]) but to our newest knowledge, there is any application for a single image age classification based on still body images. In this paper, we present our method of database collecting and our results achieved on the test dataset.

II. Experimental setup

A. Datasets

1) Face based age estimation: As there are many papers and projects addressing face age estimation, there are also many datasets for this task. However, the distribution of age in the majority of available datasets does not fit our purposes. Table I shows the number of samples for each class for different datasets. Publicly available datasets differ not only in sample quantity and age distribution but also in label quality. *Imdb* [1], *wiki* [1], *magaage* [14], and *adience* [4] datasets have relatively many samples but labels quality is low while *appa* [10] and *fgnet* [11] datasets have relatively few samples but labels quality is high. *Appa* dataset is often used as a benchmark for face age estimation solutions. For every image, it contains 2 labels. One label corresponds to real age and the second label corresponds to apparent age that is estimated by a group of experts. In our research, we decided to use the apparent labels as they had fewer outliers than real-age labels. For research purposes, we also created an in-house dataset based on part of images from *AFW* dataset [12] and *HELEN* dataset [13].

TABLE I: Number of samples for each class for different datasets.

Dataset	**Number of samples**			
	0-2	**3-12**	**13-17**	**18+**
imdb	14	1552	4215	177235
wiki	6	107	738	43373
megaage	0	4986	9239	27326
adience	1817	3174	1122	6151
appa	108	533	284	5947
fgnet	70	372	159	362
in-house	82	207	89	407

One of the most important characteristics of face age estimation datasets is head pose distribution. To examine head pose in publicly available datasets, we used *WHENet* estimator [15]. Figure 2 shows head poses distribution on publicly available datasets. Results show that the great majority of faces in publicly available datasets have a frontal pose. This is a serious problem if the trained model is supposed to be used on *in-the-wild* photos. To address this issue, we propose the usage of the dataset used in one of the face alignment problem projects [16] (later referred as face alignment dataset). In the referring dataset, for every face image, there are several images with artificially rotated heads. An example of an original image with several artificially rotated heads is shown in Figure 3. As images from face alignment dataset come among others from *AFW* and *HELEN* datasets, we were able to easily match images from our in-house dataset (which also contains images *AFW* and *HELEN* datasets) with images from *face_alignment* dataset.

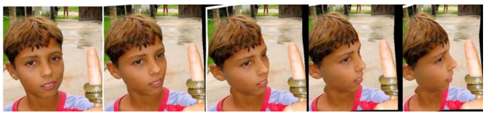

Fig. 3: Examples of artificial head rotations.

2) Silhouette based age estimation: Because of the lack of databases dedicated to silhouettes based age estimation as a base for our research, we used *imdb* database [1]. Age distribution and image diversity were not satisfying and we decided to exploit a web crawler that downloaded extra images of people mostly younger than 15 years old (photos were examined by downloaded content). Single silhouette images were obtained with *YOLOv3* algorithm [17]. We used *Dlib-ml* [18] software for face detection and we removed images that contained none or more than a single one. Detected faces were used to label our non-label part of the database with the face-based age estimation model which predicted an accurate age on a scale 0-99 [3]. In this way, we received a database containing 794567 labeled images.

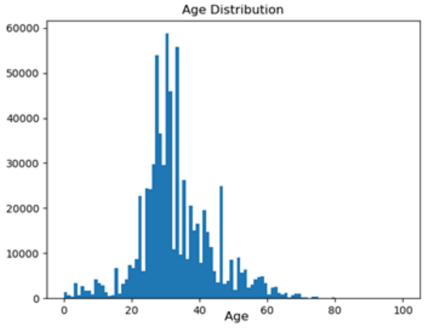

Fig. 4: Age distribution of received silhouette based database.

Figure 5 depicts few examples of images included to the final database.

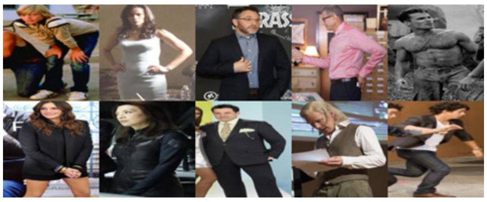

Fig. 5: Silhouette-based age estimation database examples.

B. Proposed methods

1) Face based age estimation: For face based age estimation we decided to use *resnet* [21] architecture with 4 neurons in the output layer. It is a convolutional neural network that can be utilized as a *state-of-the-art* image classification model. During tests, we used *resnet* networks with different depths (*resnet50*, *resnet101*) in order to keep optimal model size compared to training data size. Similar implementation techniques were presented in [20]. For the loss function, we used cross-entropy presented in Figure 6. As a metric, we used *Balanced Mean Absolute Error (BMAE)* and for the training algorithm, we used ADAM [24]. To level the distribution of labels during training, we decided to use an undersampling technique. Undersampling is a popular method used for balancing uneven datasets by keeping all of the data in the minority class and decreasing the size of the majority class. For data augmentation we used *autoAgment* [19] transformations. During tests, every training configuration was run 5 times with different data split for training, validation, and test sets. It is a popular technique for assessing how the results of a statistical analysis will generalize to an independent data set.

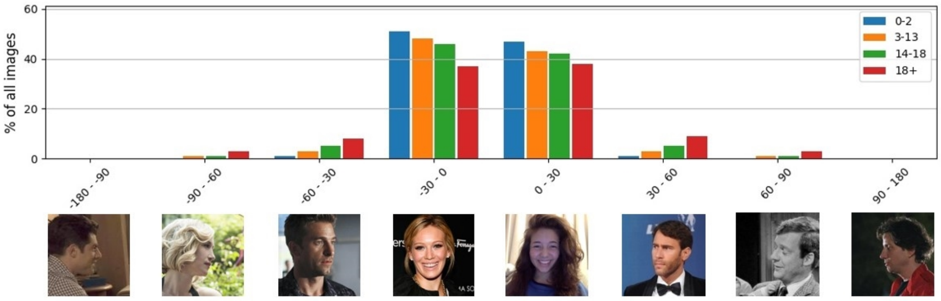

Fig. 2: Head poses distribution on publicly available datasets.

$$Loss = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$$

Mnumber of classes

y binary indicator (0 or 1)

if class label c is the correct classification for observation o

p predicted probability observation o is of class c

Fig. 6: Cross-Entropy loss function.

2) Silhouette based age estimation: For silhouette based age estimation we proposed pretrained on *ImageNet* [22] dataset *senet154* [23] architecture with added *perceptron* to in the final layer. The mentioned fully connected neural network changes size of output to 100 elements which corresponds to age classes. *SENet* architectures generalise extremely well across challenging datasets which made results slightly better than using *resnet*. As a loss function we chose *Mean Squared Error (MSE)* defined as:

$$MSE = \frac{1}{n} \sum_{n} (x_n - y_n)^2$$

n number of tensors

x_n tensor ground truth

y_n tensor prediction

Fig. 7: MSE formula.

ADAM [24] was chosen as the optimizer.

We decided to split the dataset into three parts: train, test and validation subsets, with proportions 80:15:5. Validation process was supposed to determine optimal training parameters.

C. Results

1) Face based age estimation:

a) Dataset selection: Due to the high number of available datasets and their different characteristics, datasets selection is not obvious. To analyze which datasets increase the accuracy (*BMAE*) of the model, we run multiple tests with the usage of different datasets. We started using only *appa* dataset and with every next test, we expanded the used datasets. For every configuration, we tested accuracy on in-house dataset for *resnet50* and *resnet101* architectures. Table II shows results of described tests. Even though some of the datasets had low-quality labels, they still improved the accuracy of the model. Usage of deeper networks did not improve model performance.

b) Model optimization for wide range of head poses: To prepare the model for in-the-wild images with a wide range of head poses, we produced an in-house dataset with artificially rotated heads as described in the ”datasets” section. For research purposes, we split images from in-house datasets to train, valid, and test. Then, for every image from the in-house dataset, we took 3 corresponding images from *face_alignment* dataset with different rotations: R0 (head without rotation), R1 (heads with 45 degrees rotation), and R2 (heads with 90 degrees rotation). Finally, we compared the performance of the model with and without rotated head samples in train/valid sets. We tested the performance of models on images with different head poses (R0, R1, R2) and on *appa* images to make sure that additional data do not decrease model performance on *appa* dataset. Results of the described experiment are shown in Table III. Results show that age estimation accuracy decreases with head rotation. That is a highly not desirable situation that could be problematic during age identification on the real sexual abuse images. The addition of images from the face_alignment dataset has significantly increased model accuracy on images with head rotation.

c) Model performance on appa dataset: Appa dataset is widely used as benchmark for age estimation problem. In order to obtain optimal performance on refereed dataset, we split training into two phases. Inspiration for this technique comes from [1]. In first phase, model is trained on *imdb*, *appa*, *adience*, *megaage* and *fgnet* dataset. In second phase, model is trained only on *appa* dataset. Model performance on *appa* test data (10% of appa dataset) is shown on table IV and V.

TABLE II: BMAE for different datasets sets and different network depths.

appa	fgnet	adience	megaage	imdb	wiki	BMAE	
						resnet50	**resnet101**
✓	X	X	X	X	X	0.374 +/- 0.042	0.403 +/- 0.046
✓	✓	X	X	X	X	0.358 +/- 0.024	0.381 +/- 0.034
✓	✓	✓	X	X	X	0.337 +/- 0.024	0.366 +/- 0.043
✓	✓	✓	✓	X	X	0.302 +/- 0.047	0.303 +/- 0.017
✓	✓	✓	✓	✓	✓	0.221 +/- 0.021	0.232 +/- 0.027

TABLE III: BMAE for different models and different test data.

usage of face_alignment dataset during training	BMAE			
	R0	**R1**	**R2**	**appa**
yes	0.244 +/- 0.054	0.285 +/- 0.03	0.616 +/- 0.094	0.256 +/- 0.016
no	0.221 +/- 0.023	0.242 +/- 0.03	0.383 +/- 0.039	0.255 +/- 0.028

TABLE IV: MAE score on *appa* for age range classification for different age ranges for face based age estimation.

Age range	MAE
0-2	0.134 (+/-0.055)
3-12	0.175 (+/-0.048)
13-17	0.217 (+/-0.1)
18+	0.401 (+/-0.079)

TABLE V: Confusion matrix for age range classification for different age ranges for face based age estimation.

Label age range	Predicted age range			
	0-2	**3-12**	**13-17**	**18+**
0-2	86%	14%	<1%	<1%
3-12	10%	81%	8%	1%
13-17	<1%	8%	77%	15%
18+	<1%	1%	14%	85%

Depicted results should be interpreted as promising, however, because of untypical age ranges and labeling procedures they cannot be compared with other *state-of-the-art* solutions. The innovation of the method makes it incomparable.

2) Silhouette based age estimation: We assessed results by analyzing two main criteria: *Mean Absolute Error (MAE)* in the given age range and confusion matrix for the processed output where age was grouped in the same way as it was during face-based age estimation.

TABLE VI: MAE score for different age ranges.

Age range	MAE
0-2	3.42 +/- 0.15
3-12	3.68 +/- 0.17
13-17	2.17 +/- 0.10
18+	7.77 +/- 1.20

TABLE VII: Confusion matrix for output processed data.

Confusion matrix	Age range predicitons			
Age range labels	**0-2**	**3-12**	**13-17**	**18+**
0-2	65%	25%	8%	3%
3-12	<1%	68%	20%	11%
13-17	<1%	9%	65%	26%
18+	<1%	3%	9%	88%

We can observe that *MAE* scores are the best for categories that include images of children and teenagers. It is what we have expected because of the tightest age ranges. From the point of an ongoing project that is an important feature. Otherwise, the confusion matrix shows us that adults' predictions are more accurate than in the other age categories. It might be an effect of unbalanced data or human body biological attributes which indicates sexual maturity. Presented results are much less correct than in face-based age estimation. This is the result of the worse quality of images and accumulating approximation which is the consequence of database collecting and labeling procedures.

III. Conclusions

Achieved results will be evaluated on the database specially constructed for *the APAKT* project. The database will include real images of child sexual exploitation as well as adult pornography. Face-based age estimation results can be considered the best solution dedicated to the detection of infants, prepubescent and pubescent children. Untypical age ranges and labeling procedures make our models incomparable with other *state-of-the-art* solutions. Silhouette-based age estimation delivers relatively worse results but requires the construction of a new database. It is a new perspective on the case of age estimation. The next part of the research will be optimizing results for a given database which will also contain illegal and sensitive content with minor nudity.

IV. Acknowledgement

Presented work is a part of the *System for responding to threats to children's safety in cyberspace with partic-*

ular emphasis on child pornography (APAKT) CYBERSECIDENT/455132/III/NCBR/2020 project financed by *Polish National Center for Research and Development*.

References

[1] R. Rothe, R. Timofte, L. Van Gool. 2015. ”DEX: Deep EXpectation of Apparent Age from a Single Image”. 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 252-257. doi: 10.1109/ICCVW.2015.41.

[2] H. Pan, H. Han, S. Shan, X. Chen. 2018. ”Mean-Variance Loss for Deep Age Estimation from a Face”. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5285-5294. doi: 10.1109/CVPR.2018.00554.

[3] Y. Tingting, W. Junqian, W. Lintai, X. Yong. 2019. ”Three-stage network for age estimation”. 2019. CAAI Transactions on Intelligence Technology. doi: 10.1049/trit.2019.0017

[4] G. Levi, T. Hassner. 2015. ”Age and Gender Classification Using Convolutional Neural Networks”. IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG) at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109/cvprw.2015.7301352

[5] Y. Ge, J. Lu, W. Fan and D. Yang. 2013. ”Age estimation from human body images”. 2013 IEEE Conference on Acoustics. Seech and Signal Processing. doi: 10.1109/ICASSP.2013.6638072

[6] O. F. Ince, J. Park, J. Song, B. Yoon. 2014. ”Child and Adult Classification Using Ratio of Head and Body Heights in Images”. International Journal of Computer and Communication Engineering. doi: 10.7763/IJCCE.2014.V3.3042

[7] O. F. Ince, J. Park, J. Song, B. Yoon. 2021. ”Real-Time Gait-Based Age Estimation and Gender Classification from a Single Image”. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). doi: 10.1109/WACV48630.2021.00350

[8] N. Mansouri, M. Aouled Issa, Y. B. Ben Jemaa. 2017. ”Gait-based Human Age Classification Using a Silhouette Model”. IET Biometrics. doi: 10.1049/iet-bmt.2016.0176

[9] H. Zhu, Y. Zhang, G. Li, J. Zhang, H. Shan. 2019. ”Ordinal Distribution Regression for Gait-based Age Estimation”. SCIENCE CHINA Information Sciences. doi: 10.1007/s11432-019-2733-4

[10] E. Agustsson, R. Timofte, S. Escalera, X. Baro, I. Guyon, R. Rothe, 2017. ”Apparent and Real Age Estimation in Still Images with Deep Residual Regressors on Appa-Real Database”. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 87-94. doi: 10.1109/FG.2017.20.

[11] Y. Fu, T. M. Hospedales, T. Xiang, J. Xiong, S. Gong, Y. Wang, Y. Yao. 2016. ”Robust Subjective Visual Property Prediction from Crowdsourced Pairwise Labels”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(3). 563–577. doi: 10.1109/tpami.2015.2456887

[12] X. Zhu, D. Ramanan. 2012. ”Face detection, pose estimation, and landmark localization in the wild”. 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2879-2886. doi: 10.1109/CVPR.2012.6248014.

[13] V. Le, J. Brandt, Z. Lin, L. Bourdev, T. S. Huang. 2012. ”Interactive Facial Feature Localization”. Lecture Notes in Computer Science. 679–692. doi:10.1007/978-3-642-33712-3_49

[14] Y. Zhang, L. Liu, C. Li, C.-C. Loy, Chen Change. 2017. ”Quantifying Facial Age by Posterior of Age Comparisons”. Proceedings of the British Machine Vision Conference (BMVC). doi: 10.5244/C.31.108

[15] Zhou, Yijun, Gregson, James. 2020. ”WHENet: Real-time Fine-Grained Estimation for Wide Range Head Pose”. arXiv. doi: 10.48550/arXiv.2005.10353

[16] X. Zhu, Z. Lei, X. Liu, H. Shi, S. Li. 2016. ”Face Alignment Across Large Poses: A 3D Solution”. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 146-155. doi: 10.1109/CVPR.2016.23

[17] J. Redmon, A. Farhadi. 2018. ”YOLOv3: An Incremental Improvement”. arXiv:1804.02767. doi: 10.48550/arXiv.1804.02767

[18] D. King, A. Farhadi. 2009. ”Dlib-ml: A machine learning toolkit”. Journal of Machine Learning Research

[19] Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., Le, Q. V. 2019. ”AutoAugment: Learning Augmentation Strategies From Data”. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2019.00020

[20] M. Leppioja, P. Luuka, C. Lohrmann. 2021. ”Image based classification of shipments using transfer learning”. Recent Advances in Business Analytics. Selected papers of the 2021 KNOWCON-NSAIS workshop on Business Analytics. ACSIS, Vol. 29. pages 37–44 (2021), doi: 10.15439/2021B4

[21] K. He, X. Zhang, S. Ren, J. Sun. 2016. ”Deep Residual Learning for Image Recognition”. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770-778. doi: 10.1109/CVPR.2016.90.

[22] J. Deng, W. Dong, Socher, R., Li-Jia Li, Kai Li, Li Fei-Fei. 2009. ”ImageNet: A large-scale hierarchical image database”. 2009 IEEE Conference on Computer Vision and Pattern Recognition. doi:10.1109/cvprw.2009.5206848

[23] J. HU, L. Shen, G. sun. 2018. ”Squeeze-and-Excitation Networks”. IEEE Conference on Computer Vision and Pattern Recognition. doi: 10.1109/CVPR.2018.00745

[24] D. P. Kingma, J. Ba. 2014. ”Adam: A Method for Stochastic Optimization”. Proceedings of the 3rd International Conference on Learning Representations. doi: 10.48550/arXiv.1412.6980