Real-time de-identification of minors

Filip Karlo Došilović, Antonio Gauta, Lara Lokin,
Domagoj Pavlović, Ana Puljčan

Students at *Faculty of Electrical Engineering and Computing, University of Zagreb*

January 2021

Abstract

Machine learning models are becoming ubiquitous by the day. From simple applications, such as movie and recommendations, to more serious applications, such as recommending medical diagnosis and treatment, assessing a suspect's likelihood of re-offending, and fraud detection. Interesting applications in computer vision systems include object detection and recognition, optical flow estimation, semantic segmentation, and others. One such application is age estimation, or adulthood classification. This work investigates the use of ensemble methods for classifying people into adults and minors for, possibly, real-time applications of such system in, for example, the de-identification of minors in a video.

1 Introduction

Computer vision models are now used in various situations and pipelines. From identifying customers in a shopping warehouse to self-driving cars. One application is age estimation and adulthood classification. All this technology is now available thanks to deep learning, and, specifically, convolutional neural networks. We can detect and identify different objects, segment foreground from background and estimate age from images. This project investigates whether a simpler model is capable of reliable adulthood classification with possible use in real-time systems.

This project report is organized as follows: *Related Work* section provides a short summary of similar work on age estimation and adulthood classification. Section *Adulthood classification using an ensemble* describes an approach to adulthood classification, namely, using an ensemble of regularized, shallow multilayer perceptrons (MLP). *Experiments* section provides details on hyperparameter search and experimental results. We finish with with a conclusion and proposals for future work.

2 Related Work

Earlier work on age estimation usually extract facial features (landmarks), and use those for classification using a multilayer perceptron [1] [2].

Recent advancements in architecture of convolutional neural networks allowed for even more accurate age estimation and adulthood classification. Instead of using traditional feature extraction methods, recent work uses convolutional neural networks. [3] use class specific features, learned using an autoencoder, for classifying face images into minors and adults, while in [4] apply ResNet and Xception architectures to estimate age and gender.

Our work is similar to [1] and [2], but, instead of extracting facial landmarks, we perform linear dimensionality reduction and on the reduced representation apply an ensemble of shallow multilayer perceptrons.

3 Adulthood classification using an ensemble

Our pipeline is split into several steps. First, face detection algorithm is run on an input image. Detected faces are then pre-processed before feeding them to the classifier. In second to last step we classify images into minors or adults. Last step is, simply, blurring parts of the original, input image that correspond to the faces of minors.

In the following, each step is described in detail.

3.1 Face detection

First step in our pipeline is detecting faces in an input image. Popular *face_recognition*[1] [5] package is used for face detection. Package allows two options for face detection: detection using histogram of oriented gradients (HOG) and detection using convolutional neural network (CNN). We opted for HOG.

After receiving bounding boxes by face detection algorithm, we extracted faces from the original image, and continued with the next step.

Before feeding an image to face detector, input image was resized to either 800x600 or 640x480 RGB image.

3.2 Preprocessing

Extracted face images are then resized, pixel values normalized, and, finally, dimensionality is reduced using principal component analysis (PCA).

First, image was resized to either 32x32, 48x48 or 64x64 pixel (RGB) image. Next, pixel values are normalized to $[0, 1]$ range. Final pre-processing step is dimensionality reduction using principal component analysis (PCA). PCA was set to capture 99% of variance in the dataset. to

[1] `https://github.com/ageitgey/face_recognition`

3.3 Adulthood classification

After pre-processing, face images are feed to a classifier.

Since the used dataset, UTKFaces [6], contains imbalanced classes, various methods of over- and under- sampling were used to minimize the bias present in the dataset. The following models were trained and tested:

3.3.1 Bagging and multilayer perceptron

First model is a bagging model, an ensemble method, combined with a shallow multilayer perceptron (MLP) as a base, underlying estimator. In order to handle the imbalanced classes present in the dataset, ADASYN [7] was used for over-sampling the minority class.

3.3.2 Balanced bagging and multilayer perceptron

Second classificator is also a bagging model combined with a shallow MLP, buta random under sampling is used to handle the imbalanced dataset.

Since the goal of an ensemble method is to use a collection of weak models to build a strong model, only shallow, no more than 4 computational layers, MLPs were used. Furthermore, various regularization approaches were employed to avoid over-fitting. Namely, standard L_2 regularization was applied, classifiers were trained on a subset of dataset and subset of features.

Final step in our pipeline is, simply, blurring the faces of minors.

Our implementation[2] is based on models and algorithms available in *scikit-learn* [8] and *imbalanced-learn* [9] libraries.

4 Experiments

4.1 Experiment setup and hyper-parameter search

All of the hyper-parameters were tuned using a random search [10]. As it is recommended in [10], 60 random hyper-parameter configurations (out of 486) were checked using 5-fold cross-validation on a subset of the training set. Best 3 configurations were then trained on a whole training set and their generalization tested on the test set.

The metrics used for hyper-parameter search for the two models were different. The model that over-samples minority class using ADASYN method uses accuracy as a metric, while balanced accuracy was used for for the models which (randomly) under-samples majority class.

We refer the reader to the implementation, specifically configuration files, for more information on the hyperparameter search space.

[2]Source code, configuration files and trained models can be found here `https://github.com/fkdosilovic/minor-adult-classification-ensemble`.

4.2 Results

Two sets of the results are presented. Each set contains two tables, one with the results per image size and the other table contains confusion matrices that correspond to the best results. Rows in confusion matrix represent true lables, while columns represent predicted labels. Since the dataset is imbalanced, we use balanced accuracy as a final metric for model comparison.

As mentioned above, the 3 best configurations were trained and tested for each image size. Full results are available in the Appendix.

Results for ADASYN + Bagging with MLP are available in Tables 1 and 2. We see from the Table 1 that the best results achieveable with ADASYN + Bagging with MLP model uses 64x64x3 face images.

Table 1: Results for ADASYN + Bagging with MLP

Image size	PCA features	Accuracy	Balanced accuracy
32x32x3	379	0.95	0.8896
48x48x3	597	0.955	0.8972
64x64x3	769	0.953	**0.897**

Table 2: Confusion matrices for best ADASYN + Bagging with MLP models

(a) 32x32x3 images

	Minor	Adult
Minor	989	253
Adult	100	5745

(b) 48x48x3 images

	Minor	Adult
Minor	1007	235
Adult	96	5749

(c) 64x64x3 images

	Minor	Adult
Minor	1001	241
Adult	75	5770

Results for Balanced Bagging with MLP are available in Tables 3 and 4. We see from the Table 3 that the best results achieveable with Balanced Bagging with MLP model uses 48x48x3 face images.

Table 3: Results for Balanced Bagging with MLP

Image size	PCA features	Accuracy	Balanced accuracy
32x32x3	379	0.936	0.906
48x48x3	597	0.942	**0.909**
64x64x3	769	0.939	0.908

We should be careful when interpreting the aforementioned results. Even though the balanced accuracy is much better suited for datasets with imbalanced classes than accuracy alone, we can see from the confusion matrix that, for instance, the best ADASYN + Bagging model is the one that uses 48x48x3 images, and not the one that uses 64x64x3 images.

Table 4: Confusion matrices for best Balanced Bagging with MLP models

(a) 32x32x3 images

	Minor	Adult
Minor	1063	179
Adult	252	5593

(b) 48x48x3 images

	Minor	Adult
Minor	1069	173
Adult	238	5607

(c) 64x64x3 images

	Minor	Adult
Minor	1068	181
Adult	236	5609

5 Conclusion

Determining age and adulthood of individuals in images and videos is an important problem. New laws protect the individuals, especially minors, from various exploitations that are even more present with easy availability of *plug-n-play* machine learning models. As those treats continue to proliferate, models and methods that counteract them should also develop, thereby making it harder to exploit the large availability of data. One such system is real-time de-identification (of minors).

Our results show that imbalanced datasets can be problematic, but managable. Future work should focus on improving the misclassification rate of minors as adults, as that is more problematic than classifying adults as minors. One could employ asymmetric loss function, instead of balanced accuracy, as a metric. Also, using images of size 96x96x3 and 128x128x3 should improve classification accuracy.

References

[1] Nabil Hewahi. Age estimation based on neural networks using face features. *Journal of Emerging Trends in Computing and Information Sciences*, 1, 1 2010.

[2] Sarah Kohail. Using artificial neural network for human age estimation based on facial images. pages 215–219, 03 2012.

[3] Maneet Singh, Shruti Nagpal, Mayank Vatsa, and Richa Singh. Are you eligible? predicting adulthood from face images via class specific mean autoencoder. *Pattern Recognition Letters*, 119, 03 2018.

[4] Dmitry Yudin, Maksim Shchendrygin, and Alexandr Dolzhenko. *Age and Gender Recognition on Imbalanced Dataset of Face Images with Deep Learning*, pages 30–40. 06 2020.

[5] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[6] Zhifei Zhang, Yang Song, , and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[7] Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *IN: IEEE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE), IJCNN 2008*, pages 1322–1328, 2008.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[9] Guillaume Lemaitre, Fernando Nogueira, and Christos K. Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *CoRR*, 2016.

[10] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(10):281–305, 2012.

Appendix

Table 5: Confusion matrices and metrics for variations of ADASYN + bagging with MLP for 32x32x3 images

(a)

	Minor	Adult
Minor	990	252
Adult	114	5731

(b)

	Minor	Adult
Minor	976	266
Adult	89	5756

(c)

	Minor	Adult
Minor	989	253
Adult	100	5745

(d)

	Accuracy	Balanced accuracy
(a)	0.948	0.889
(b)	0.95	0.885
(c)	**0.95**	**0.8896**

Table 6: Confusion matrices and metrics for variations of ADASYN + bagging with MLP for 48x48x3 images

(a)

	Minor	Adult
Minor	1005	237
Adult	92	5753

(b)

	Minor	Adult
Minor	1003	239
Adult	82	5763

(c)

	Minor	Adult
Minor	1007	235
Adult	96	5749

(d)

	Accuracy	Balanced accuracy
(a)	0.954	0.8967
(b)	**0.955**	0.8968
(c)	0.953	**0.8972**

Table 7: Confusion matrices and metrics for variations of ADASYN + bagging with MLP for 64x64x3 images

(a)

	Minor	Adult
Minor	994	248
Adult	64	5781

(b)

	Minor	Adult
Minor	993	249
Adult	75	5770

(c)

	Minor	Adult
Minor	1001	241
Adult	75	5770

(d)

	Accuracy	Balanced accuracy
(a)	**0.956**	0.895
(b)	0.954	0.893
(c)	0.955	**0.897**

Table 8: Confusion matrices and metrics for variations of balanced bagging with MLP for 32x32x3 images

(a)

	Minor	Adult
Minor	1079	163
Adult	340	5505

(b)

	Minor	Adult
Minor	1069	173
Adult	294	5551

(c)

	Minor	Adult
Minor	1063	179
Adult	252	5593

(d)

	Accuracy	Balanced accuracy
(a)	0.929	0.905
(b)	0.934	0.905
(c)	**0.939**	**0.906**

Table 9: Confusion matrices and metrics for variations of balanced bagging with MLP for 48x48x3 images

(a)

	Minor	Adult
Minor	1066	176
Adult	234	5611

(b)

	Minor	Adult
Minor	1066	176
Adult	273	5572

(c)

	Minor	Adult
Minor	1069	173
Adult	238	5607

(d)

	Accuracy	Balanced accuracy
(a)	**0.942**	**0.909**
(b)	0.937	0.906
(c)	**0.942**	**0.909**

Table 10: Confusion matrices and metrics for variations of balanced bagging with MLP for 64x64x3 images

(a)

	Minor	Adult
Minor	1068	174
Adult	260	5585

(b)

	Minor	Adult
Minor	1059	183
Adult	223	5622

(c)

	Minor	Adult
Minor	1061	181
Adult	236	5609

(d)

	Accuracy	Balanced accuracy
(a)	0.939	**0.908**
(b)	**0.943**	0.907
(c)	0.941	0.907