

inaFaceAnalyzer: a Python toolbox for large-scale face-based description of gender representation in media with limited gender, racial and age biases

David Doukhan¹ and Thomas Petit¹²

¹ French National Institute of Audiovisual (INA) ² Univ Lyon, INSA Lyon, LIRIS (UMR 5202 CNRS)

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Arfon Smith](#) ↗

Reviewers:

- [@lukassnoek](#)
- [@vimalthilak](#)

Submitted: 22 February 2022

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

In partnership with



This article and software are linked with research article DOI [10.3847/xxxxx<-update this with the DOI from AAS once you know it](https://doi.org/10.3847/xxxxx<-update this with the DOI from AAS once you know it) published in the Astrophysical Journal <- The name of the AAS journal..

Summary

inaFaceAnalyzer is a Python toolbox designed for large-scale analysis of faces in image or video streams. It provides a modular processing pipeline allowing to predict age and gender from faces. Results can be exported to tables, augmented video streams, or rich ASS subtitles. inaFaceAnalyzer is designed with speed in mind to perform large-scale media monitoring campaigns. The trained age and gender classification model provided is based on a ResNet50 architecture and evaluated on 4 face databases. Evaluation results are highly competitive with respect to the current state-of-the-art, and appear to reduce gender, age and racial biases.

Statement of need

Automatic facial attribute analysis allows to extract information from images or video streams containing faces such as gender, age, emotion, identity or *race*. This information can be used in a wide range of applications including biometrics, human-computer interaction, multimedia indexation, digital humanities and media monitoring - but also surveillance and security.

inaFaceAnalyzer is a Python framework aimed at extracting facial attribute information from massive video and image streams. It was realized to meet the needs of French National Audiovisual Institute (INA), in charge of archiving and providing access to more than 22 million hours of TV and radio programs. The emergence of computational digital humanities and data journalism has increased the need of INA's users to access meta-data obtained from automatic information extraction methods.

Since 2018, INA has realized several large-scale studies (up to 1 million hours of program analyzed) in the context of Gender Equality Monitor project, which aims at describing men and women representation differences in media. INA's automatic softwares aimed at estimating men and women speech time ([Doukhan et al., 2018](#)) or decoding TV text incrustations ([Doukhan et al., 2020](#)) were used in digital humanity studies as well as in French public reports ([Calvez, 2020](#); [CSA, 2021](#)). Earlier inaFaceAnalyzer prototypes were used to detect gender from face in TV programs, and compare these estimates to speech-time ([Baudry et al., 2020](#); [Doukhan et al., 2019](#); [Rezgui, 2019](#)).

inaFaceAnalyzer has a modular design allowing easy integration of new components. With respect to the high social impact associated to the studies using this software, it should provide high accuracy prediction models. Being aimed at describing the representation of under-represented categories of people in media, it should minimize gender, age or racial biases that are known to also affect machine learning datasets and softwares. It is highly configurable, allowing to define trade-offs between accuracy and processing time depending on the scale of the analyses to be performed and on the available computational resources.

Image and Video processing pipeline

The face analysis pipeline is composed of several modules :

- **stream decoding** : image and video frames are converted to 8-bits RGB arrays. Trade-offs between processing time and result robustness can be defined by restricting analyses to a given amount of frames per second of video content (FPS), or to *key frames*.
- **face detection** : based on libfacedetection (Peng & Yu, 2021) or OpenCV CNN, with customizable detection confidency and minimal size of faces to be returned. Table 3 show best predictions performances are obtained for face resolutions above 75 pixels.
- **face tracking (optional)** : built on the top of DLib's correlation tracker (Danelljan et al., 2014). Allows to limit the amount of calls to time-consuming face detection methods, and to smooth tracked faces prediction results.
- **face preprocessing** : Facial landmarks are extracted using Dlib, and faces are rotated so that the eyes lie in a horizontal line. A parametric scaling factor is used to extend face bounding boxes (we obtained best results on gender classification with a scaling factor of 1.1 (Rezgui, 2019), while (Rothe et al., 2018) used a factor of 1.4). Finally, faces are resized to the dimensions required by classifiers and stored in a FIFO structure, allowing to take advantage of GPU architectures by providing large batches of faces.
- **face classification** : Our default and best model is based on ResNet50 architecture (He et al., 2016) and implemented in Keras (Chollet et al., 2015) using input faces of size (224*224). It was trained on FairFace (Karkkainen & Joo, 2021) using data augmentation strategies and a multi-objective loss allowing to predict simultaneously gender, age and race. This strategy allows to learn different concepts jointly, and optimize inference computation time.

Results are returned as pandas DataFrames that can be exported to tabular formats. Export functions allows to generate videos with embed results, or ASS subtitles that can be displayed in VLC. Lastly, command line scripts allow to process list of files with custom parameters.

Execution time and accuracy may vary according to input video resolution and analysis parameters. Using a GeForce RTX 2080 Ti GPU, it can process a 10 minutes 960*540 video in 26 seconds, using a 1 FPS analysis rate without tracking.

Evaluation of face gender and age classification models

Four publicly available databases with gender, age and racial annotations are used : FairFace (Karkkainen & Joo, 2021), ColorFerret (Phillips et al., 2000), UTK Faces (Zhang & Qi, 2017) and MAFA (Ge et al., 2017). MAFA is particularly challenging since it is composed of masked faces, which can be representative of audiovisual content during the Covid-19 pandemic. MAFA is mostly composed of asian faces and do not contain age annotations.

Tables 1 and 2 compare the results obtained with inaFaceAnalyzer, to those obtained with the open-source frameworks DeepFace (Serengil & Ozpinar, 2020) and DEX (Rothe et al., 2018). DeepFace and DEX were evaluated using inaFaceAnalyzer's face detection and preprocessing pipeline¹. Gender classification is evaluated using accuracy and age prediction using Root Mean Square Error (RMSE). In order to neutralize corpora biases (some categories of people may be more represented - impacting accuracy and RMSE), we present the results together with gender (male vs female), racial (white vs non white) and age biases (adults from 20 to 50 vs adults over 50 years old). These estimates were obtained by calculating evaluation metric for each of the 8 (2*2*2) categories separately (ie: accuracy for non white females over 50 years old) and reporting their averaged difference (ex: positive accuracy age bias means the system is more accurate for people under 50).

¹DEX won ChaLearn LAP 2015 challenge on apparent age estimation and do not provide face detection. DeepFace's face detection systems do not detect all faces we're using in the evaluation. We report the best results obtained by these systems using 10 different bounding box scaling factors varying between 1 and 2.

inaFaceAnalyzer is associated to the best performances in gender and age prediction, with very low gender, age and racial biases compared to the other systems. Gender classification results are also higher than those reported in the literature for FairFace (+1 accuracy) and MAFA (+1.2 accuracy) (Islam et al., 2021; Karkkainen & Joo, 2021). The large negative gender bias obtained on MAFA indicates a tendency to associate masked faces to female faces, which should be improved in future releases.

Table 1: Gender prediction evaluation based on accuracy / gender bias / racial bias / age bias

System/Dataset	FairFace	ColorFerret	MAFA	UTK Faces
inaFaceAnalyzer	95.3/0.8/-1.5/-0.2	99.2/0.8/0.3/-0.2	84.9/-8.9/6.4/?	94.8/-1.5/0.4/0.2
DeepFace	81.7/26.0/5.7/7.7	87.4/33.9/7.3/4.0	77.7/6.3/10.1/?	90.3/9.8/3.9/5.2
DEX	83.6/17.1/6.1/5.2	93.2/14.6/5.1/1.3	80.8/6.2/7.9/?	91.6/4.1/2.1/2.7

Table 2: Age prediction evaluation based on RMSE / gender bias / racial bias / age bias

System/Dataset	FairFace	ColorFerret	UTK Faces
inaFaceAnalyzer	6.8/-0.8/-0.6/-2.0	7.2/-0.8/-0.0/0.1	7.7/-2.7/0.1/-3.5
DeepFace	11.8/-0.4/-1.5/-5.0	9.5/-1.0/0.0/-0.4	10.7/-2.7/-0.9/-5.1
DEX	8.7/-1.4/-1.0/-2.7	7.2/-0.6/-0.2/0.3	8.4/-2.9/-1.2/-4.1

Table 3: Fairface Gender and age prediction evaluation for face resolutions varying between (10*10) and (224*224) pixels. Lower resolution faces were obtained artificially by resizing detected faces to target dimensions using OpenCV resize function

reso- lu- tion (pix- els)	10	15	20	25	30	40	50	75	100	150	224
Gen- der accu- racy	63.7	76.9	83.5	88.5	90.9	93.3	94.3	95.0	95.4	95.4	95.3
Age RMSE	14.9	13.2	10.9	9.3	8.4	7.6	7.2	6.9	6.8	6.8	6.8

Distribution

inaFaceAnalyzer source code and trained models are available on [github](#), [PyPI](#) and [Dockerhub](#); and distributed under MIT license. After consultation of French CNIL (French data protection authority) and DDD (French Rights Defender), racial classification layers of our models were removed from their public distribution in order to prevent their use for non ethical purposes. These models can however be provided for free after examination of each demand.

Acknowledgements

This work has been partially funded by the French National Research Agency (project GEM: ANR-19-CE38-0012) and European Union (project MeMAD : H2020 grant agreement No 780069). We acknowledge contributions from Zohra Rezgui wrote the first piece of code that lead to inaFaceAnalyzer during her internship at INA (Rezgui, 2019).

References

- Baudry, N., Claverie, C., Nibert, D., Poupardin, E., Segault, A., & Quesada, E. V. (2020). Le genre dans les imaginaires de l'informatique : Voix et représentations dans les publicités (1980-2010). *Billets de Recherche Inatheque*.
- Calvez, C. (2020). Place des femmes dans les médias en temps de crise. *Public Report*.
- Chollet, F.others. (2015). Keras. GitHub. <https://github.com/fchollet/keras>
- CSA. (2021). La représentation des femmes à la télévision et à la radio: Rapport sur l'exercice 2020. *Public Report*.
- Danelljan, M., Häger, G., Khan, F., & Felsberg, M. (2014). Accurate scale estimation for robust visual tracking. *British Machine Vision Conference, Nottingham*. <https://doi.org/10.5244/c.28.65>
- Doukhan, D., Méadel, C., & Coulomb-Gully, M. (2020). En période de coronavirus, la parole d'autorité dans l'info télé reste largement masculine. *La Revue Des Médias*.
- Doukhan, D., Poels, G., Rezgui, Z., & Carrive, J. (2018). Describing gender equality in french audiovisual streams with a deep learning approach. *VIEW Journal of European Television History and Culture*, 7(14), 103–122. <https://doi.org/10.18146/2213-0969.2018.jethc156>
- Doukhan, D., Rezgui, Z., Poels, G., & Carrive, J. (2019). Estimer automatiquement les différences de représentation existant entre les femmes et les hommes dans les médias. *Journée DAHLIA: "Informatique Et Humanités Numériques: Quelles Problématiques Pour Quels Domaines?"*.
- Ge, S., Li, J., Ye, Q., & Luo, Z. (2017). Detecting masked faces in the wild with LLE-CNNs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2682–2690. <https://doi.org/10.1109/cvpr.2017.53>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. <https://doi.org/10.1109/cvpr.2016.90>
- Islam, M. M., Tasnim, N., & Baek, J.-H. (2021). Gender prediction from masked face images using deep learning towards smart store customer management. *International Conference on Future Information & Communication Engineering*, 12, 68–71. <https://doi.org/10.1109/cvpr.2017.53>
- Karkkainen, K., & Joo, J. (2021). FairFace: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1548–1558. <https://doi.org/10.1109/wacv48630.2021.00159>
- Peng, H., & Yu, S. (2021). A systematic IoU-related method: Beyond simplified regression for better localization. *IEEE Transactions on Image Processing*, 30, 5032–5044. <https://doi.org/10.1109/tip.2021.3077144>
- Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10), 1090–1104. <https://doi.org/10.1109/34.879790>
- Rezgui, Z. (2019). *Détection et classification de visages pour la description de l'égalité femme-homme dans les archives télévisuelles* [Msc. Thesis]. Higher School of Statistics; Information Analysis, University of Carthage.
- Rothe, R., Timofte, R., & Van Gool, L. (2018). Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2), 144–157. <https://doi.org/10.1007/s11263-016-0940-3>

- 149 Serengil, S. I., & Ozpinar, A. (2020). LightFace: A hybrid deep face recognition framework.
150 *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 23–27. <https://doi.org/10.1109/asyu50717.2020.9259802>
151
152 Zhang, S., Zhifei, & Qi, H. (2017). Age progression/regression by conditional adversarial
153 autoencoder. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
154 <https://doi.org/10.1109/cvpr.2017.463>

DRAFT