F. Boutros, N. Damer, M. Fang, M. Gomez-Barrero, K. Raja,
C. Rathgeb, A. Sequeira, and M. Todisco (Eds.): BIOSIG 2024,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2024

Evaluating the Impact of Traditional Anonymization Filters on State-of-the-Art Facial Recognition Performance

Gabriele Turetta[1]

Abstract: In the digital age, anonymization filters are crucial for protecting individuals' privacy across platforms like media, social networks, mapping services, and surveillance systems. Traditional techniques such as blurring, pixelization, and black bands are widely used, but more advanced methods like morphing, warping, and scrambling offer enhanced privacy. This study explores the effects of these traditional anonymization methods on the performance of MagFace, a leading facial recognition model. By applying filters like Gaussian blur, Median Blur, Pixelization and Black Bands to a subset of the CelebA dataset, we assess how different types and intensities of anonymization impact recognition accuracy.

Keywords: Facial anonymization, Image obfuscation, Facial recognition systems, Biometric performance, Privacy-preserving techniques, DET curves

1 Introduction

In today's digital landscape, anonymization filters play a vital role in safeguarding individuals' privacy across various platforms, including media, social networks, mapping services like Google Maps, and surveillance systems. Traditional facial anonymization techniques, such as blurring, pixelization, and black bands, have been commonly employed to obscure identifiable features. However, beyond these traditional methods, more sophisticated approaches like morphing [KE14], warping [KE14], and scrambling [MD14][KE14] also exist.

As demonstrated by Korshunov et al. [Ko13], increasing the strength of anonymization filters generally leads to a decrease in recognition accuracy, thus enhancing privacy but also reducing detection accuracy, thereby decreasing intelligibility.

Maintaining a certain level of visual coherence in anonymized images is crucial, especially in systems that need to recognize and categorize human actions. In such environments where both privacy and intelligibility are critical, an effective anonymization filter must anonymize a face by removing features that make it identifiable, while ensuring that the performance of event recognition systems is not compromised by the anonymization process.

Recent research efforts, such as those by Ren et al. [RLR18], have focused on enhancing the intelligibility for camera systems that need to recognize and categorize human actions

[1] Human Centered Artificial Intelligence, Danmarks Tekniske Universitet, Anker Engelunds Vej 1, Bygning 101A, 2800 Kongens Lyngby, s233124@dtu.dk

while preserving privacy. Ren's method employs a convolutional neural network (CNN) to implement pixel-level modifications of faces. Instead of simply applying a mask or blur to faces, this technique modifies individual pixels of faces to render them unrecognizable to humans, while preserving useful information such as the general shape of the face, the position of eyes, and mouth, for action recognition.

Similarly Chen et al. [CKI18] propose a Privacy-Preserving Representation-Learning Variational Generative Adversarial Network (PPRL-VGAN) to learn an image representation that is explicitly disentangled from the identity information.

It is evident that achieving a trade-off between privacy preservation and activity intelligibility under video surveillance is extremely important in such systems. Thus, understanding the necessary level of filtering to anonymize a face against a state-of-the-art facial recognition model is essential for effective privacy protection.

Previous studies have highlighted the effects of anonymization filters on performances of face recognition systems. The first was Newton et al. [NSM05] who explored the low efficacy of conventional facial anonymization techniques in protecting privacy against advanced facial recognition models, demonstrating significant vulnerabilities. Dufaux et al. [DE10] proposed a framework to evaluate the performance of facial recognition algorithms on images subjected to traditional anonymization filters, revealing their ineffectiveness in preserving privacy. Korshunov et al. [KE14] investigated various privacy protection filters with differing strengths using multiple face recognition algorithms implemented in OpenCV: based on Principal Component Analysis (PCA) [TP91], based on Linear Discriminant Analysis (LDA) [BHK97], and based on local features (LBP) [AHP06]. They illustrated the challenges in achieving optimal privacy protection.

This study specifically examines the impact of traditional anonymization filters on the performance of a state-of-the-art facial recognition model, MagFace [Me21]. By focusing solely on privacy rather than intelligibility, the study seeks to understand how classic filters affect the accuracy of facial recognition systems. MagFace represents a cutting-edge advancement in face recognition technology by introducing novel loss functions that facilitate the learning of a universal feature embedding. This embedding serves a dual purpose: it quantifies the quality of acquired facial data and provides a measure of the likelihood of successful recognition. Unlike traditional methods, MagFace ensures that the magnitude of the feature embedding increases monotonically with the likelihood of recognition, enhancing reliability in varied acquisition conditions. Furthermore, MagFace incorporates an adaptive mechanism that refines within-class feature distributions by emphasizing representative samples and mitigating the influence of noisy or low-quality data, thereby improving recognition accuracy in real-world scenarios. Extensive experimental validation consistently demonstrates MagFace's superiority over existing state-of-the-art methods, underscoring its role as a leading model in contemporary face recognition research.

For this study classic filters like Gaussian Blur, Median Blur, Pixelization and Black Bands were applied to a subset of the CelebA dataset [Li15].

The experimental setup and dataset creation are detailed in the subsequent chapter, followed by a presentation of the results. The conclusion section will summarize the findings and propose directions for further research to develop more effective privacy protection methods in an increasingly surveilled world.

2 Experimental Setup

The study began by structuring two subdatasets derived from the extensive CelebA public dataset, which comprises 202,599 photos of 10,177 different celebrities.

Due to computational and time constraints, the number of photos in the subdatasets was significantly reduced from the original dataset's size. This reduction was carried out to establish the baseline performance of the facial recognition model on original photos without any filter applied:

- For non-mated comparisons, a subdataset was created containing 200 images, one per celebrity, resulting in a total of 19,900 non-mated comparisons. This step was crucial in determining the model's ability to distinguish between faces of different individuals.
- For mated comparisons, a subdataset was formed with 1,000 images from 50 different celebrities, providing 20 images per celebrity. The total number of mated comparisons performed by the model is therefore:

$$\binom{20}{2} \cdot 50 = 9500$$

  This setup was used to establish the model's baseline performance in recognizing paired faces from the same individual.

The number of comparisons is summarized in the following table:

Comparison Type	**Number of Comparisons**
Non-Mated	19,900
Mated	9,500
Total	**29,400**

Tab. 1: Summary of comparisons for baseline performance

Following the baseline assessment, each anonymization filter was applied to all images in the mated dataset to observe its impact on the model's recognition accuracy and performance.

2.1 Baseline performance

The cosine similarity metric was used to calculate the correlation between pairs of feature vectors returned by MagFace. Cosine similarity was chosen because it is robust to differences in the magnitudes of the feature vectors and is calculated as:

$$\text{Cosine similarity} = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\|\|\vec{B}\|} \tag{1}$$

This metric ranges from -1 to 1, where:

- 1 indicates that the vectors are identical.
- 0 indicates that the vectors are orthogonal, meaning they are completely dissimilar.
- -1 indicates that the vectors are diametrically opposed.

The correlations' statistics on the original photo datasets are summarized in the following table.

Statistics	**Mated**	**Non Mated**
Observations	9500	19900
Minimum	-0.12749	-0.21133
Maximum	0.98446	0.50220
Mean	**0.50323**	**0.09731**
St. Dev.	0.15585	0.08911
Skewness	-0.42125	0.31458
Ex. Kurtosis	0.03081	0.37775

Tab. 2: Model recognition statistics on original photo datasets

After that a Kernel Density Estimation (KDE) was applied to the normalized data to estimate the probability density function of the random variables and evaluate the separation between the mated and non mated comparisons, as visible in Figure 1.

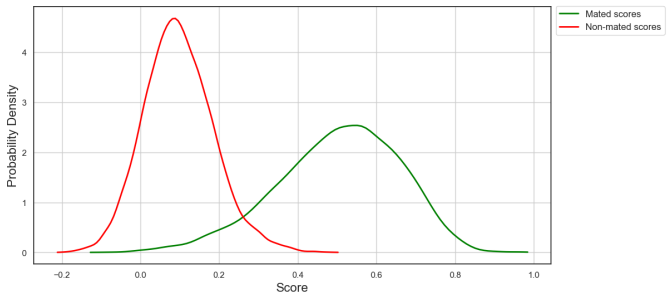


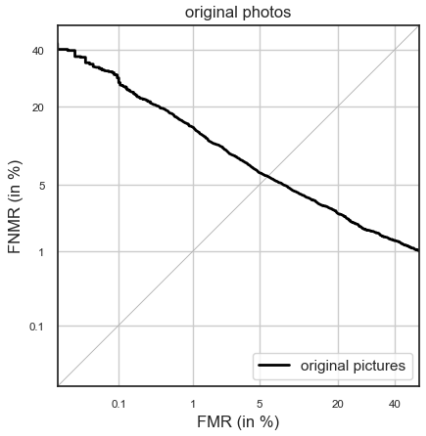
Fig. 1: Kernel Density Estimation plot for the normalized datasets

There's a decent separation between the two distribution, confirmed by the sensitivity index, a dimensionless metric which measures the separation between two distributions:

$$d' = \frac{\|\mu_{\text{mated}} - \mu_{\text{non-mated}}\|}{\sqrt{\frac{1}{2}\left(\sigma^2_{\text{mated}} + \sigma^2_{\text{non-mated}}\right)}}$$

The sensitivity index returns a value of 3.1977, that can be considered generally good. This means that if we set a correlation threshold, we can classify a pair of faces as "mated" (belonging to the same identity) if the correlation value is above this threshold. Conversely, we can classify two faces as "non-mated" (belonging to different identities) if the correlation value is below this threshold.

Using the DET curve [Ma97], we can visualize how the False Matching Rate (FMR), the percentage of face comparisons that are incorrectly identified as matches, and the False Non-Matching Rate (FNMR), which represents the percentage of face comparisons that are incorrectly identified as non-matches, vary with the decision threshold. The DET curve for the baseline performance is visualized in Figure 2a.



(a) DET curve illustrating baseline performance of the system with original photos.

(b) TP,TP,FP,FN visualized for the threshold corresponding to the EER point

Fig. 2

In Figure 2a the function y=x represents the Equal Error Rate (EER) line, for which FMR = FNMR. For our system, the EER point falls at 5.97%. This operating point is significant because it indicates the threshold that minimizes overall error, effectively balancing the sensitivity (ability to detect mated pairs) and specificity (ability to detect non-mated pairs) of the recognition system. This operating point serves as an excellent starting point for further evaluations.

At the EER point of 5.9%, the found threshold is 0.2398 (Figure 2b) and our system's performance metrics are as follows:

Metric	Value
True Accepts (TP)	8933
True Rejects (TN)	18714
False Accepts (FP)	1186
False Rejects (FN)	567
False Match Rate (FMR)	5.97%
False Non-Match Rate (FNMR)	5.97%

Tab. 3: Performance metrics at EER point (5.9%)

2.2 Anonymization Filters

The following anonymisation filters were applied to varying degrees to all photos in the mated dataset. For each level of anonymisation and for each anonymised photo, the correlation with each original photo of the same celebrity was then calculated, resulting in 20,000 correlations for each gradation of the filter.

2.2.1 Gaussian Blur Filter

Fig. 3: different levels of gaussian blur filter applied

The Gaussian Blur filter offered by the OpenCV library was used with 9 different intensities, varying the Gaussian kernel size parameter linearly. The filter calculates the value of each pixel as a weighted average of the surrounding pixels within a kernel defined by the kernel size, with the weights calculated using a Gaussian function. The results of applying the filter can be seen in the figure 3. As mentioned in the introduction to the subsection, the mated correlations were repeated for each filter intensity. On the other hand, the original non-mated correlations have been retained.

2.2.2 Median Blur Filter

Fig. 4: different levels of median blur filter applied

The median blur filter offered by the OpenCV library was used with 9 different intensities, varying the kernel size parameter linearly. Unlike the Gaussian blur filter, the median blur filter calculates the value of each pixel by replacing it with the median of the pixel

values within a kernel of specified size. The results of applying the filter can be seen in the figure 4.

2.2.3 Pixelization Filter

Fig. 5: different levels of pixelization filter applied

The pixelization filter provided by the OpenCV library was applied using 9 different intensities, linearly varying the block size parameter. The pixelization filter processes each pixel by replacing it with the average color of a block of pixels defined by the block size, employing the `cv2.INTER_NEAREST` interpolation method. This method resizes the image by selecting the nearest pixel value for each new pixel. The results of applying the filter can be seen in Figure 5.

2.2.4 Face Masking Filter

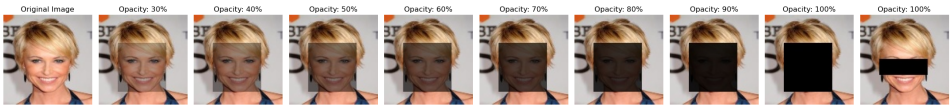


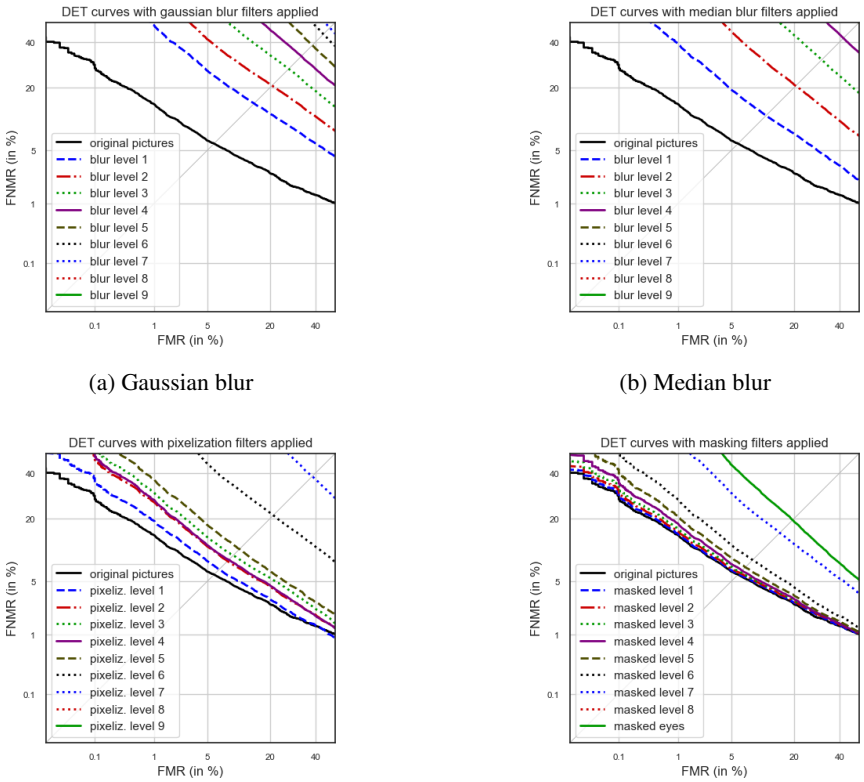
Fig. 6: different levels of face masking filter applied

The face masking filter implemented in this study utilizes OpenCV's capabilities to selectively obscure facial features in images. The filter applies a black band over detected faces at 8 opacity levels, ranging from 30% to 100%. A 9th image with a black band at 100% opacity was added, with the black band obscuring only the middle third of the face, to analyze the difference in recognition performance between obscuring the whole face and obscuring only the eyes region.
The application employs the `haarcascade_frontalface_default.xml` classifier, a pretrained classifier that utilizes the Haar Cascade algorithm [VJ01]. Figure 6 illustrates the impact of the filter, showcasing how different opacity settings alter the visibility of facial features."

3 Results

After applying filters at various intensities to the original photos, the DET curves were plotted again to analyze the degradation of system performance.

As visible in Figure 7, increasing the filter intensity linearly results in linear degradation of model performance for blur filters, both Gaussian and median, with the latter appearing



(a) Gaussian blur

(b) Median blur

(c) Pixelization

(d) Masking

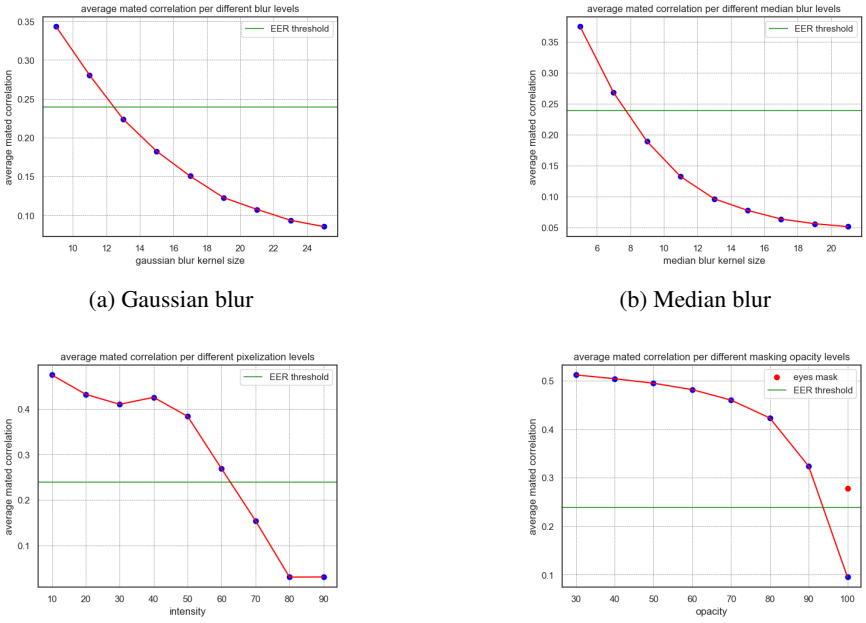
Fig. 7: DET Curves for different filters

more impactful on system performance and thus more effective from an anonymization standpoint.

It is also interesting to observe in pixelization how well the recognition model responds to anonymization, with its performance minimally affected by the presence of a filter. Only at high pixelization intensities (level 6 and beyond), FMR and FNMR increase significantly.

The system also reacts remarkably well to a black face mask, delivering compelling performance up to 90% opacity of the mask. It's noteworthy that a small mask at 100% opacity just on the eye area significantly degrades the model. To understand if such masks, as well as other filters, are adequate for rendering the face unrecognizable, we consider the previously found EER point with a decision threshold of 0.2398. As mentioned earlier, at this threshold, the rates of FMR and FNMR are identical, indicating a balance point in error trade-offs.

Figure 8 shows the trend of average mated comparisons for each filter as the filter intensity varies. Our plots exhibit characteristics reminiscent of an e^{-x} shape for Gaussian blur and



(a) Gaussian blur

(b) Median blur

(c) Pixelization

(d) Masking

Fig. 8: Average mated correlation per different intensities of the filters

median blur, while pixelization and black mask display an inverse $-e^x$ profile, suggesting a greater robustness of the model to the application of pixelization and black mask.

It is interesting to note how at intensity 40, the average correlation for pixelization anomalously increases compared to intensity 30, before resuming the negative trend from intensity 50 onwards. This could be due to the moderate application of a pixelization filter reducing noise in the images, temporarily enhancing the average correlation. Alternatively, the pixelization filter may interact specifically with image features, accentuating or attenuating certain details or patterns that influence the average correlation.

It is also important to highlight how applying a black mask filter only over the eyes at 100%, as indicated by the red dot in Figure 8.d, is insufficient at the chosen decision threshold to anonymize the face, demonstrating the robustness of MagFace.

The following table shows the values of the intensities for the different filters where the recognition system is more likely to fail to recognize the original face at the EER point, with an accuracy of 94.03%. An example of corresponding images with the filters applied is shown in Figure 9.

Filter	Parameter	Value
Gaussian blur	Kernel size	13
Median blur	Kernel size	9
Pixelization	Intensity	70
Face masking	Opacity	100%

Tab. 4: Values of filter intensities where the recognition system fails to recognize the original face, at the EER point

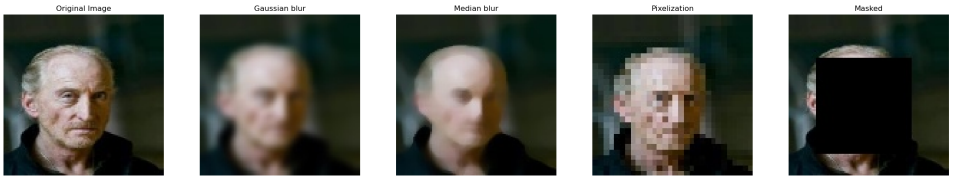


Fig. 9: Lowest filters intensity for which the model is less likely to recognize the original faces, at the EER point

4 Conclusion

This research explored the influence of traditional anonymization filters on the MagFace facial recognition model's performance. Common filters were applied—Gaussian blur, median blur, pixelization, and black bands—to a subset of the CelebA dataset to assess how these methods affect recognition accuracy.

Our analysis revealed that traditional anonymization techniques can indeed diminish the effectiveness of facial recognition systems, though the degree of impact varies with the type and strength of the filter used. Notably, the median blur filter was more effective than the Gaussian blur in impairing the system's performance. Pixelization maintained some resilience at lower intensities but significantly hampered recognition accuracy at higher intensities. Additionally, the face masking filter demonstrated that obscuring key facial features substantially reduced the model's identification capabilities only with a 100%-opacity masks. Furthermore, the mask covering only the eyes at 100% opacity proved to be ineffective at anonymizing the face, at the considered EER point.

These findings underscore the necessity of carefully choosing anonymization methods to ensure robust privacy protection. While sophisticated facial recognition systems like MagFace show resilience against some conventional methods, intensifying these filters can strike a balance between privacy and usability. Future studies should delve into more advanced anonymization techniques and evaluate their effects on a wider array of facial recognition models.

Another interesting area to explore is to study how the model's performance changes in different lighting conditions and environments, as well as how the model's performance changes depending on the subject being photographed, to understand if MagFace performs better with certain ethnicities, genders, or ages.

References

[AHP06] Ahonen, Timo; Hadid, Abdenour; Pietikainen, Matti: Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence, 28(12):2037–2041, 2006.

[BHK97] Belhumeur, Peter N.; Hespanha, Joao P; Kriegman, David J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on pattern analysis and machine intelligence, 19(7):711–720, 1997.

[CKI18] Chen, Jiawei; Konrad, Janusz; Ishwar, Prakash: VGAN-Based Image Representation Learning for Privacy-Preserving Facial Expression Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. June 2018.

[DE10] Dufaux, Frédéric; Ebrahimi, Touradj: A framework for the validation of privacy protection solutions in video surveillance. In: 2010 IEEE International Conference on Multimedia and Expo. pp. 66–71, 2010.

[KE14] Korshunov, Pavel; Ebrahimi, Touradj: Towards optimal distortion-based visual privacy filters. In: 2014 IEEE International Conference on Image Processing (ICIP). pp. 6051–6055, 2014.

[Ko13] Korshunov, Pavel; Melle, Andrea; Dugelay, Jean-Luc; Ebrahimi, Touradj: Framework for objective evaluation of privacy filters. In (Tescher, Andrew G., ed.): Applications of Digital Image Processing XXXVI. volume 8856. International Society for Optics and Photonics, SPIE, p. 88560T, 2013.

[Li15] Liu, Ziwei; Luo, Ping; Wang, Xiaogang; Tang, Xiaoou: Deep Learning Face Attributes in the Wild. In: Proceedings of International Conference on Computer Vision (ICCV). December 2015.

[Ma97] Martin, Alvin F; Doddington, George R; Kamm, Terri; Ordowski, Mark; Przybocki, Mark A: The DET curve in assessment of detection task performance. In: Eurospeech. volume 4, pp. 1895–1898, 1997.

[MD14] Melle, Andrea; Dugelay, Jean-Luc: Scrambling faces for privacy protection using background self-similarities. In: 2014 IEEE International Conference on Image Processing (ICIP). pp. 6046–6050, 2014.

[Me21] Meng, Qiang; Zhao, Shichao; Huang, Zhida; Zhou, Feng: MagFace: A Universal Representation for Face Recognition and Quality Assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14225–14234, June 2021.

[NSM05] Newton, E.M.; Sweeney, L.; Malin, B.: Preserving privacy by de-identifying face images. IEEE Transactions on Knowledge and Data Engineering, 17(2):232–243, 2005.

[RLR18] Ren, Zhongzheng; Lee, Yong Jae; Ryoo, Michael S.: Learning to Anonymize Faces for Privacy Preserving Action Detection. In: Proceedings of the European Conference on Computer Vision (ECCV). September 2018.

[TP91] Turk, Matthew A; Pentland, Alex P: Face recognition using eigenfaces. In: Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition. IEEE Computer Society, pp. 586–587, 1991.

[VJ01] Viola, Paul; Jones, Michael: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. volume 1. Ieee, pp. I–I, 2001.