On the Robustness of Deep Learning Based Face Recognition

Werner Bailer, Martin Winter
JOANNEUM RESEARCH, DIGITAL – Institute for Information and Communication Technologies
Graz, Austria
{firstname.lastname}@joanneum.at

ABSTRACT

Identifying persons using face recognition is an important task in applications such as media production, archiving and monitoring. Like other tasks, also face recognition pipelines have recently shifted to Deep Convolutional Neural Network (DNNs) based approaches. While they show impressive performance on standard benchmark datasets, the same performance is not always reached on real data from media applications. In this paper we address robustness issues in a face detection and recognition pipeline. First, we analyze the impact of image impairments (in particular compression) on face detection, and how to conceal them in order to improve face detection performance. This is studied both on face samples originating from still image and video data. Second, we propose approaches to improve open-set face recognition, i.e., handling of "unknown" persons, in particular to reduce false positive recognitions. We provide experimental results on image and video data and provide conclusions that help to improve the performance in practical applications.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Applied computing** → Digital libraries and archives.

KEYWORDS

robustness; face detection; face recognition; open-set recognition; compression

ACM Reference Format:
Werner Bailer, Martin Winter. 2019. On the Robustness of Deep Learning Based Face Recognition. In *1st International Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV '19), October 21, 2019, Nice, France.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3347449.3357483

1 INTRODUCTION

The identifiable persons appearing in audiovisual content are one of the most important cues for content understanding and description. In many tasks in media production, media archiving and media monitoring tagging the appearance of known persons via recognition of their faces is needed for describing content, determining its relevance or indexing it for search, to name just a few use cases. Depending on the application, the set of persons of interest may vary, but it is usually a small set compared to the number of faces appearing in video content. This poses practical face recognition as an open-set recognition problem, i.e., it is a classification into one from a set of known persons – or "unknown", with the latter having a high prior probability. For example, recognizing persons of interest in media monitoring or surveillance, or annotating persons in audiovisual media production and archiving are all applications that fall into this category.

Many tasks for video understanding have shifted from traditional approaches to ones that rely on Deep Convolutional Neural Networks (DNNs) in recent years. This is also true for a face recognition pipeline, which usually consists of a face detection and a face classification step. For example, it has been shown that multitask CNNs [18] achieve very good performance for face detection, including detection of partly occluded faces. In order to enable fast and efficient training of new faces, there are several recognition pipelines that use DNNs as a feature extractor and then use a classifier that can be trained with a moderate number of examples per person, such as support vector machines (e.g., [12]) or online random forests (e.g., [15]). These approaches became sufficiently mature for practical use in applications in the media industry and beyond. However, the performance achieved on standard benchmark datasets is not always reached on real data from media production and archiving. There are number of issues related to the robustness of deep learning approaches that need to be considered.

A general issue of neural networks is that the training data has certain characteristics in terms of image quality (noise level, compression artifacts, etc.), and the network may learn some of these characteristics even if they are irrelevant to the task. For practical applications the robustness against variations in these parameters is crucial. The problem is also related to that of using so-called *adversarial samples*, i.e., samples that add noise not visible to humans to the image in order to cause misclassifications by the neural network, exploiting the patterns it has learned. In practical applications, some level of compression artifacts is always present, and also other types of impairments might occur.

Another practical problem rarely discussed in the literature is the robust distinction of faces never presented to the classifier from those already learned [5]. This is a non-trivial task often neglected in state of the art face recognition benchmarks which usually focus on optimization of classification accuracy (true-positives and false-positives values) of "known" faces in the entire database, but do not consider the robust separation of "unknown" faces. In practical applications, such as media production and archiving, the majority of faces in the content is likely not in the dataset. This produces a large number of false detections, even at low to moderate false positive rates of the algorithm.

AI4TV '19, October 21, 2019, Nice, France

ACM ISBN 978-1-4503-6917-6/19/10…$15.00
https://doi.org/10.1145/3347449.3357483

In this paper we address robustness issues in a face detection and recognition pipeline. First, we analyze the impact of image impairments (in particular compression) on face detection, and how to conceal them in order to improve the face detection performance. This is studied both on face samples originating from still image and video data. Second, we propose approaches to improve the classification of faces into "known" and "unknown", in particular to reduce false positives recognitions.

The rest of this paper is organized as follows. Section 2 discusses related work and Section 3 presents the experiments we have performed in order to analyze the problem and possible solutions. We report and discuss the results in Section 4, and Section 5 concludes the paper.

2 RELATED WORK

The impact of quality impairments in input images has only been studied in few works. [4] analyzed the impact of different distortions (blur, noise, contrast, JPEG and JPEG2000 compression) on the performance of image classification using DNNs. They used rather strong compression, i.e., JPEG compression parameters ≤ 20. For all the modifications there is a positive correlation between stronger distortion and decreased DNN performance, with different non-linear relations. [3] studied the relation between adversarial attacks and JPEG compression. They argue that JPEG compression tends to reduce high-frequency components in images, and thus has the potential to eliminate noise that could be used in adversarial attacks on the DNN. A somewhat related work is [13], which aims to develop image quality metrics targeting alignment with machine performance rather than human perception. They use face detection as an example task to derive a metric that aligns with detection performance. However, their face detector is not DNN-based. In summary we can thus make the following observations from the existing literature. None of the works addresses particularly face detection, and all address only on data originating from still images. Also, no strategies for robustness against compression artifacts have been studied in existing works.

Robust distinction between faces never presented to the classifier and those that were trained, has been rarely discussed in literature so far [5]. One earlier work dealing with the issue of the so-called open-set recognition of faces and objects is Scheirer et al. [11]. In particular, the authors propose a novel "1-vs-set machine" using modified marginal distances from a linear, binary SVM. Although the work primarily focuses on object recognition tasks, there is also an evaluation on the Labeled Faces in the Wild (LFW) [6] dataset. More recent work on open-set face recognition is the work of Günther et al. [5], where the authors compared several algorithms for assessing similarity of deep feature approaches and concluded that only extreme value machines (EVM) [9] can sufficiently discriminate between known and unknown persons on the LFW database. However, the reported performance of EVMs is not sufficient for real-time surveillance applications (only 60% correctly classified faces at a false alarm rate of 0.01) and additional research is recommended for practical applications. Another interesting work reporting also results on the popular YouTube Faces (YTF) [16] database is the one from Sun et al. [14], where a stacked set of deep convolutional networks (25 DeepID2+ networks) has been proposed to achieve state of the art results. Similar to the observations in other work cited above the performance on closed-set evaluations is 99.5% and 93.2% on LFW and YTF respectively, but the performance for face identification degrades to 80.7% in the open-set evaluation benchmark. The work of Liu et al. [7] is also noteworthy, where the authors proposed a novel loss function (A-Softmax loss) for a CNN architecture (termed 'SphereFace') in order to replace Euclidean metrics based margins by a proper Face-manifold metric. This metric can be used to recognize faces with a nearest neighbor classifier while coevally using distance-thresholding in the hypersphere manifold. Reported accuracies on the LFW and YTF datasets are 99.4% and 95.0% in a closed-set benchmark protocol respectively. Performance for open-set evaluation is only reported for the Mega-Face data [8], but naturally shows lower performance between 72% and 75% depending on the SphereFace variant implemented. Regarding open-set face recognition we can conclude that the performance reported in literature is below the needs for practical applications, and research for novel algorithms is needed.

3 EXPERIMENTS

We use two datasets for the experiments. Labeled Faces in the Wild (LFW) [6] is a commonly used dataset for face recognition, containing 13K still images (JPEG compressed) of faces collected from the web and showing 1,680 labeled persons. YouTube Faces (YTF) [16] is a dataset created from web video and thus containing various video compression qualities. The data set contains 3,425 videos of 1,595 different people, which amounts to about 600K frames with face detections.

3.1 Face detection under distortions

For both datasets, we run face detection using multi-task CNNs [18] on each of the images. The images usually contain a single face, however, in particular some of the LFW images contain small faces in the background. Thus we limit the number of detected faces to 1, using the largest face only. We apply the detector to the original image, which will indicate where detection fails due to the compression of the source or the insufficient performance of the face detector, and to distorted versions of the images. We apply the following set of distortions:

Blurring. A box filter with size $k \times k$ is applied, with $k \in \{3, 6, 9, 12, 15, 18, 21, 27, 30\}$.

Sharpening. A sharpening approach based on unsharp masking is applied. The source image is blurred with a 3×3 binomial filter, and the difference between the source and blurred source image is multiplied the magnitude m_s and added again to the image. The magnitudes used are $m_s = \in \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60\}$.

JPEG compression. The image is recompressed with a JPEG quality factor of q, with $q \in \{90, 75, 60, 45, 30\}$. In contrast to [4] we use high to moderate compression settings which better reflect the content in professional media production.

JPEG compression concealment. The analysis of the differences between images compressed with different JPEG compression factors shows that most differences are quantization noise, with rather small absolute pixel differences. The results of the blurring experiments (for details see Section 4) indicate that moderate blurring never harms the face detection performance. We empirically determined on a small set of sample images, that the compression artifacts can be well suppressed by blurring with a 4 × 4 box filter, that is applied in two passes. Blurring only once or with a smaller kernel did not sufficiently reduce the artifacts. Thus we apply this blurring to the source image (to address cases where the compression of the source already prevents successful detection) and to each of the recompressed JPEG images.

3.2 Unknown face classification

As shown e.g. by [7] one of the best performing approaches to closed-set face recognition is FaceNet [12]. Recent modifications of this approach for media production and archiving use a combination of FaceNet features with incremental machine learning approaches to automatically train classifiers for unknown persons [15]. As a first experiment we start with their approach and use the unknown person detection method proposed to evaluate the performance in a closed-set LFW database scenario.

In particular, we first divide the LFW database into a set of known and unknown faces based on the minimum number of available samples for each of the 5,750 persons. As there is a minimum number of 5 training images required to achieve good performance [15], we treat all instances with at least 6 face-samples as known faces and the remaining part as unknown ones. Thus we end up with an evaluation set containing 3,665 known and 8,042 unknown faces respectively.

Based on the results and analysis of the baseline experiment (see Sections 4.3 and 4.4 for more details) we implemented two improvements to overcome the lack of performance for the usage of the algorithm in real practical applications. The first improvement focuses on the limited discriminative power of the classification confidence for a class obtained from the incremental random forest classifier.

A random forest classifier [1] is a well known and studied classification and regression method based on an ensemble of (binary) decision trees. It has been successfully applied to a large number of applications in machine learning and computer vision. Even unified frameworks for easy adaptation to regression, density estimation and manifold learning (e.g. Criminisi et al. [2]) as well as variants to online adaptation exist [10].

Usually, the classification decision of a forest is obtained by a simple majority voting of all the individual leaf nodes' class predictions. Given the total number of leaf-node votes (N_{tot}), the confidence values (C_x) for a class (x) is usually calculated by the relative leaf node vote frequency (eq. 1) when a certain test sample is traversing through the ensemble of classification trees:

$$C_{x,tot} = \frac{N_x}{N_{tot}} \tag{1}$$

As we found that this is not enough to reliably discriminate between known and unknown faces, especially when the number of different classes is higher, we improve the confidence value by combining the ratio between most ($C_{first} = C_{x,tot}$) and second most class confidence (C_{sec})

$$C_{x,rel} = \frac{C_{first}}{C_{sec}} \tag{2}$$

with the ratio of maximum class confidence and mean of other class confidences ($C_{x,mean}$)

$$C_{x,mean} = \frac{C_{x,tot}}{mean\left([C_1, ..., C_n] \setminus C_{n=x}\right)} \tag{3}$$

by a weighted sum (with $\beta = 0.75$ in our experiments) to get an improved prediction score (C_x).

$$C_x = \beta * C_{x,rel} + (1 - \beta) * C_{x,mean} \tag{4}$$

The second improvement is a sample excluding pre-processing step based on the calculation of a simple correlation measure between the actual test sample and the prototypical face feature samples stored during the training phase. In particular, if a detected face is considered as a known-face, i.e., its classification confidence C_x is above a 'database membership' threshold $\theta = 0.55$, we compare the feature vector $F(x)$ with all the n prototypical features vectors $F_p(x)$ for class x and dimension d according to their correlation measure

$$Corr_{x,p} = \frac{\sum_{i=1}^{d}(x_i - \overline{x})(p_i - \overline{p})}{\sqrt{\sum_{i=1}^{d}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{d}(p_i - \overline{p})^2}} \tag{5}$$

and accept it only if any of the correlation measures are above a minimal correlation threshold ($\tau = 0.75$ in our experiments).

4 RESULTS

4.1 Face detection results on LFW

The results for LFW are shown in the plots in Figure 1. The first observation is that the face detection rate for the source images is at just above 0.997, i.e., the face in nearly 0.3% of source images is not detected. For blurring, there is hardly any impact up to a kernel size of $k = 9$, and then the performance start to decline quickly. For sharpening, there is already a small performance loss with the smallest magnitude, and then the performance loss is approximately linear with the magnitude. The results for JPEG compression show a similar behavior of roughly linear performance reduction, but the performance loss is much smaller.

When compression artifact concealment by blurring is applied, the detection rate on the source images increases to 0.9994, i.e., the miss rate drops to one fifth of the one without concealment. When applying concealment to the re-encoded JPEG images with lower quality factors, the results oscillate around the detection rate for the source images with concealment. This result indicates that this level of JPEG compression does not cause loss of relevant information on LFW. Any missed detections due to JPEG artifacts are due to the quantization noise, thus reconstruction to the original performance level can be reached again.

4.2 Face detection results on YTF

The results of YTF are shown in the plots in Figure 2. Overall they are similar to those on LFW, with a decline of performance for blurring with kernel sizes above $k = 9$, and roughly linearly declines for sharpen and JPEG compression (with a smaller magnitude for the JPEG compression). However, the performance level for the

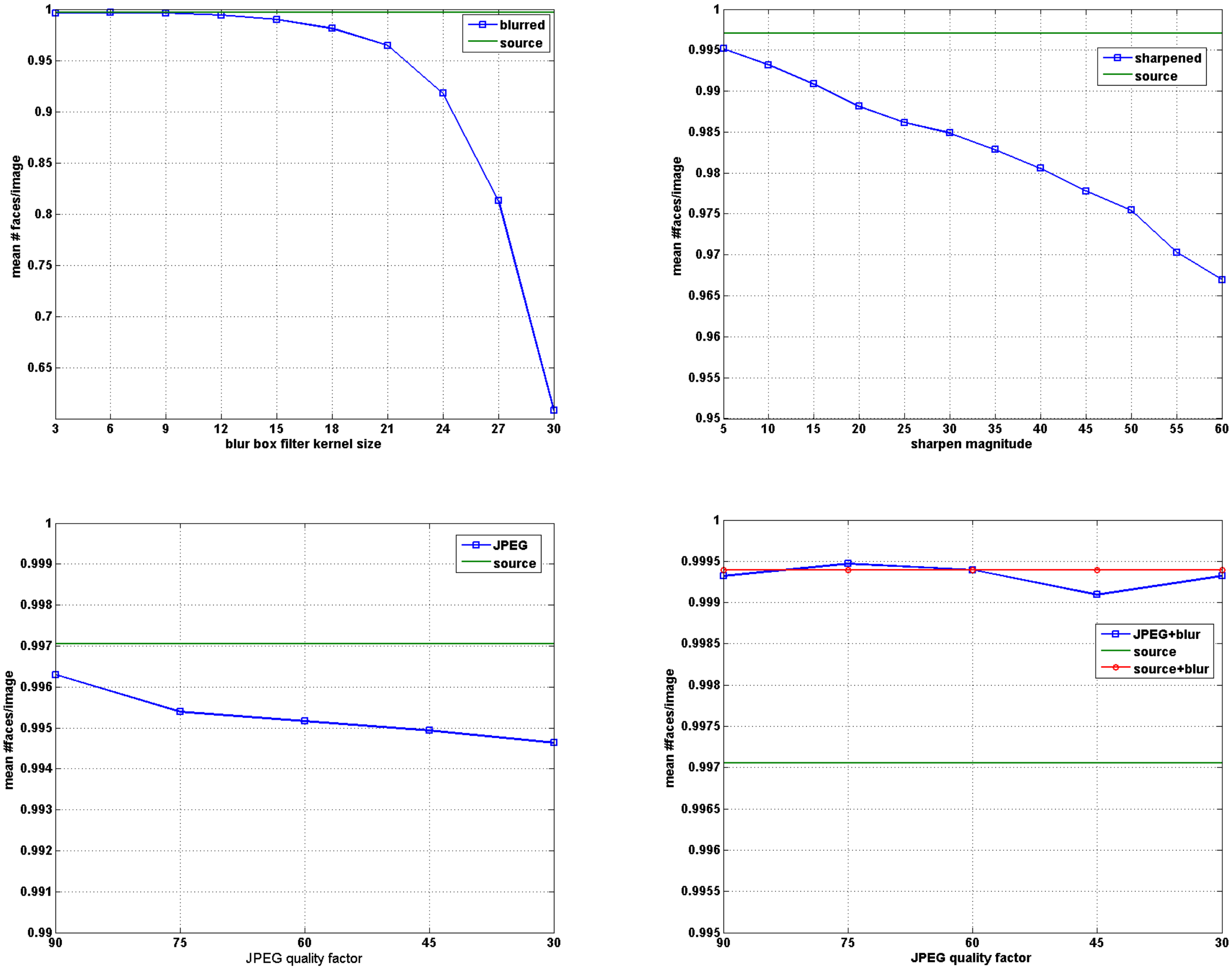

Figure 1: Face detection results on LFW.

source images is just above 0.995, i.e., the face in nearly 0.5% of source images is not detected.

When compression artifact concealment by blurring is applied, the detection rate clearly increases. It does not reach the same level as on LFW, but still the miss rate is halved. When applying concealment to the re-encoded JPEG images with lower quality factors, the detection performance stays nearly constant, with a very small decline for quality factors below 45. However, the performance stays about 0.1% below that of the source images without concealment. The main difference is that the source content on YTF has already undergone video compression, while LFW source images are moderately JPEG compressed. Thus not only the overall performance on the source content is lower, but also the effects of further compression cannot be entirely eliminated.

4.3 Classifying unknown faces on LFW database

The results for the experiments using the LFW dataset in an open-set evaluation scenario are summarized in Table 1. We split the evaluation into two parts in order to see the individual contributions of either the improved probability measure calculation (ImprovedPROB) and correlation check (+CorrCHECK) proposed.

The evaluation measures are split into the true/false (T/F) and positive/negative (P/N) parts in order to get better insight and allow for individual discussions for known ($_k$) and unknown ($_u$) faces. Relative results (%) are always calculated with respect to the absolute number of known/unknown faces respectively. [1]

[1] Please note that for the baseline evaluation total number of faces checked is insignificantly lower (<2 ‰), as the improved correlation probability measure also affects the training procedure for a few samples.

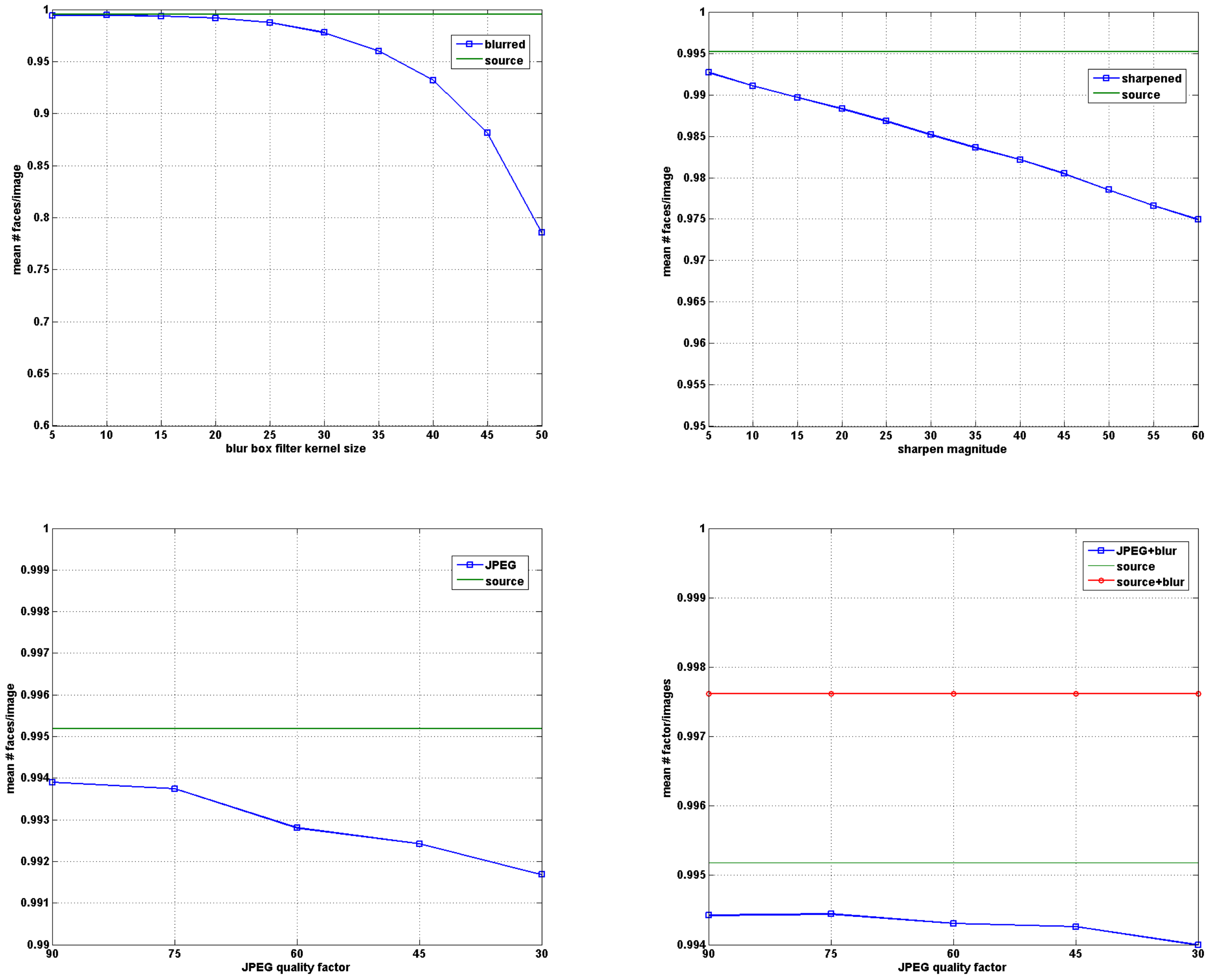

Figure 2: Face detection results on YTF.

Measure	Baseline		ImprovedPROB		+CorrCHECK	
TP_k	1857	48.00%	3441	93.89%	3206	87.48%
FP_k	4	0.10%	5	0.14%	1	0.03%
FN_k	2008	51.90%	219	5.98%	458	12.50%
TN_u	7791	99.65%	7877	97.95%	8022	99.75%
FP_u	27	0.35%	165	2.05%	20	0.25%
CCR		82.55%		96.68%		95.91%

Table 1: Results for the experiments using the LFW dataset in an open-set evaluation benchmark scenario.

It is obvious, that the basic approach is not applicable to practical applications as the true classifications for the known faces is <50% in the open-set scenario of LFW. This is not feasible for practical applications although the total number of false classifications for the known faces (FP_k) is low and most of the faces are erroneously classified as unknowns.

Applying the improved probability measure, the situation changes significantly. There is an improvement (from 48% to 94%) introduced for the correct classifications while keeping the number of false classifications (FP_k) low. As the number of correctly classified unknown faces does not change, too, the improved probability measure causes a shift of almost all erroneously classified unknown faces from the base experiment to the known face group. Moreover, all of this examples are also correctly classified leading to the high number of TP_k.

The main drawback is that some of the remaining unknown face samples are erroneously classified as known ones, thus leading to a substantial number of classified faces although not in the database

(FP_u). Although their absolute number (165) is low in relation to the more than 1,500 correct classifications, this is not acceptable with respect to usability. Consider for example a video-surveillance or media-investigation application where raising too much false alarms/detections immediately disqualifies the solution. Fortunately the pre-correlation check contributes to lowering those cases to 1/10 of the number without the check, while keeping the correct classifications high and moreover improving the overall ratio of identified unknown faces (TN_u) to almost 100%.

4.4 Discussion

Face detection under distortions. We can gain the following insights from the experiments for face detection on the two datasets.

- Blurring and sharpening cause as expected performance loss that is proportional to the strength of the defect. However, the relation between the parameter of the defect and the performance impact is quite different, as is the effect at small strengths (some tolerance up to certain strength vs. immediate impact).
- Compression does have a non-negligible impact on the performance of face detection, and also the unmodified source images of common datasets are affected.
- At high to moderate JPEG quality factors, this performance loss is not due to a loss of information, but due to quantization noise that is independent of the quality factor.
- Slight blurring does not cause reduction of detection performance, but can reduce JPEG quantization noise in at least half of the cases where detection on the original source images fails.
- The findings of [3] are only partly confirmed by our analysis. While the compression will remove high-frequency noise from the source content, which could be used for adversarial attacks, the compression process will also produce quantization noise. This may have an impact on the detection results, though it may be more difficult to use it for an adversarial attack.
- On content with high quality, the additional compression will not cause information loss, and concealment reaches the same performance level whether starting from the source or a compressed version. On content with already higher source compression, concealment always provides improvement, though not beyond that of the source content.

Unknown face classification. For unknown face classification we can gain the following insights from the experiments.

- Robust distinction of faces never presented to the classifier from those already learned is a problem for practical applications. Applying a state of the art combination of a CNN-based combined with an incremental machine learning approach for classification and detection of unknown persons in a closed-set scenario provided similar results than those reported for extreme value machines (EVM) [9].
- Applying the improved probability measure proposed, it is possible to dramatically increase the classification performance of known faces. However, the most critical measure of falsely classified unknown faces requires additional measures.
- The combination of the former with a correlation-based pre-check method reduces the number of unknown faces erroneously treated as known ones by coevally keeping the other measures in feasible range, albeit at a small, tolerable cost of true known face classifications.

5 CONCLUSION

In this paper, we have proposed strategies for improving robustness of face detection and classification of unknown faces in a face recognition pipeline.

For handling compressed content, the use of slight blurring as a concealment strategy seems useful. In the experiments, the missed detections were at least halved, and no negative impact of blurring could be observed. This is a very efficient and easy to apply preprocessing step. Alternative approaches would require retraining of the respective face detector. Data augmentation by providing more compressed samples could be one option. However, given the observations on the nature of the distortion, augmentation would benefit from a large number of differently compressed samples to eliminate any statistical patterns in the quantization noise rather than covering a broad range of quality factors. The relatedness of the quantization noise to adversarial samples has been mentioned. Thus a recently proposed approach for adversarial training called ME-Net [17] might be applicable, which uses matrix estimation to replace the original training data with an approximated version in order to eliminate noise while preserving larger structures.

To achieve robust distinction of faces never presented to the classifier from those already learned, a novel two-step approach – based on a state of the art combination of deep-learning based face-features and an incremental machine learning algorithm – has been proposed. Our experiments show that the improved probability measure and correlation-based pre-check increase the classification rate for faces in an open-set benchmark. Moreover we get an almost perfect separation capability between known and unknown faces.

ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreements n° 761802 MARCONI ("Multimedia and Augmented Radio Creation: Online, iNteractive, Individual") and n° 761934, Hyper360 ("Enriching 360 media with 3D storytelling and personalisation elements")

REFERENCES

[1] Leo Breimann. 2001. Random Forests. *Machine Learning* 45 (2001), 5–32.

[2] Antonio Criminisi, Jamie Shotton, and Ender Konukoglu. 2012. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *Found. Trends. Comput. Graph. Vis.* 7 (Feb. 2012), 81–227. https://doi.org/10.1561/0600000035

[3] Nilaksh Das, Madhuri Shanbhogue, Shang-Tse Chen, Fred Hohman, Siwei Li, Li Chen, Michael E Kounavis, and Duen Horng Chau. 2018. Shield: Fast, practical defense and vaccination for deep learning using jpeg compression. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 196–204.

[4] Samuel Dodge and Lina Karam. 2016. Understanding how image quality affects deep neural networks. In *2016 eighth international conference on quality of multimedia experience (QoMEX)*. IEEE, 1–6.

[5] Manuel Günther, Steve Cruz, Ethan M Rudd, and Terrance E Boult. 2017. Toward open-set face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 71–80.

[6] Erik Learned-Miller, Gary B Huang, Aruni RoyChowdhury, Haoxiang Li, and Gang Hua. 2016. Labeled faces in the wild: A survey. In *Advances in face detection and facial image analysis*. Springer, 189–248.

[7] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. 2017. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 212–220.

[8] Daniel Miller, Evan Brossard, S Seitz, and Ira Kemelmacher-Shlizerman. 2015. Megaface: A million faces for recognition at scale. *arXiv preprint arXiv:1505.02108* (2015).

[9] Ethan M Rudd, Lalit P Jain, Walter J Scheirer, and Terrance E Boult. 2017. The extreme value machine. *IEEE transactions on pattern analysis and machine intelligence* 40, 3 (2017), 762–768.

[10] Amir Saffari, Christian Leistner, Jakob Santner, Martin Godec, and Horst Bischof. 2009. On-line random forests. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on.* IEEE, 1393–1400.

[11] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult. 2013. Toward Open Set Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 7 (July 2013), 1757–1772.

[12] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 815–823.

[13] Rajiv Soundararajan and Soma Biswas. 2019. Machine vision quality assessment for robust face detection. *Signal Processing: Image Communication* 72 (2019), 92–104.

[14] Yi Sun, Xiaogang Wang, and Xiaoou Tang. 2015. Deeply learned face representations are sparse, selective, and robust. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2892–2900.

[15] Martin Winter and Werner Bailer. 2019. Incremental Training for Face Recognition. In *MultiMedia Modeling - 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8-11, 2019, Proceedings, Part I.* 289–299. https://doi.org/10.1007/978-3-030-05710-7_24

[16] L Wolf, T Hassner, and I Maoz. 2011. Face recognition in unconstrained videos with matched background similarity. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE Computer Society, 529–534.

[17] Yuzhe Yang, Guo Zhang, Zhi Xu, and Dina Katabi. 2019. ME-Net: Towards Effective Adversarial Robustness with Matrix Estimation. In *International Conference on Machine Learning.* 7025–7034.

[18] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters* 23, 10 (2016), 1499–1503.