On the Robustness of Deep Learning Based Face Recognition

Werner Bailer, Martin Winter

JOANNEUM RESEARCH DIGITAL

AI4TV '19, Nice, FR

Motivation

Context

- Identifying persons is a crucial task in video analysis
  - broadcast content/archive documentation
  - media monitoring, open source intelligence
  - processing of user generated content
- Annotating training data is costly
  - large databases cover internationally known celebrities etc.
  - for many applications less known people are of interest, for which it is not easy to find samples on the web
  - resources for collecting and annotating samples are limited, identification must work with few examples

Motivation

Specific problems

- Persons of interest not known in advance
  - add quickly new persons to be identified
  - add additional samples of a person not reliably detected in some cases
- Analyse only once
  - re-running (total) analysis is costly
  - it shall be possible to find people added later in already annotated content and link them to names
  - requires **robust detection**
- Handling unknown persons
  - in many media applications, a small set of persons is relevant
  - most persons detected are unknown
  - **requires robust method to decide about known/unknown**
  - not well covered by existing benchmarks

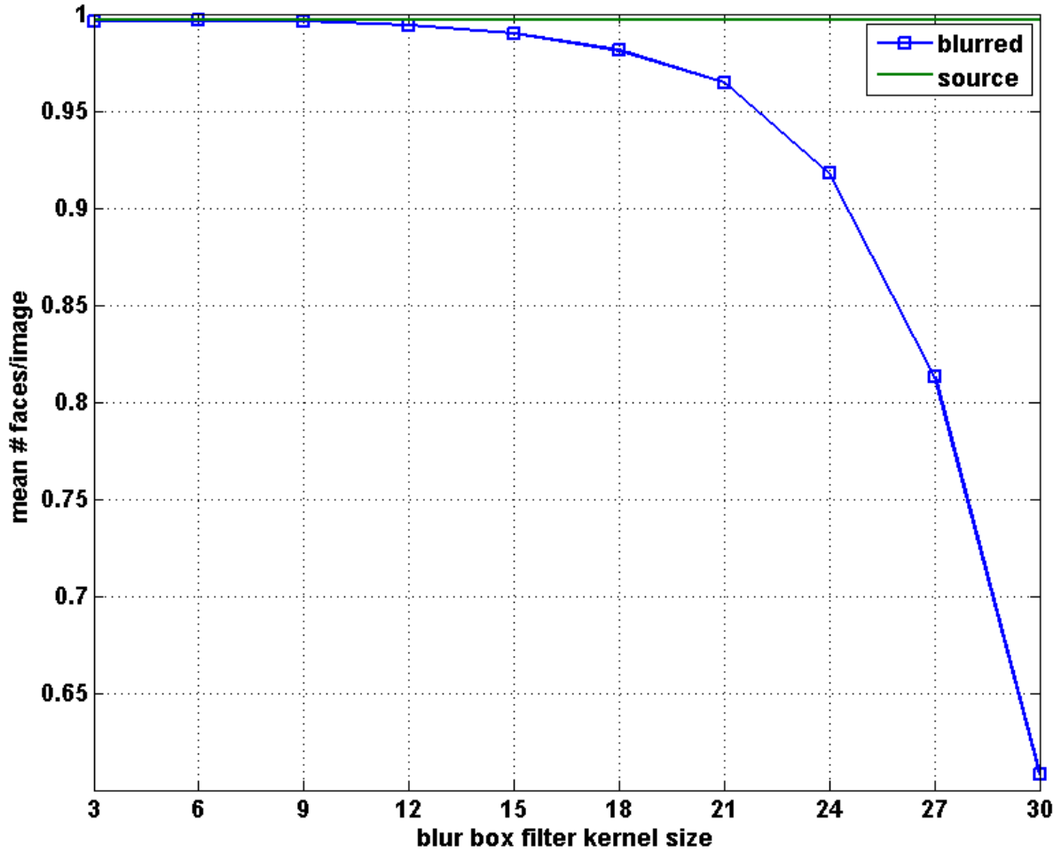
Robustness of face detection

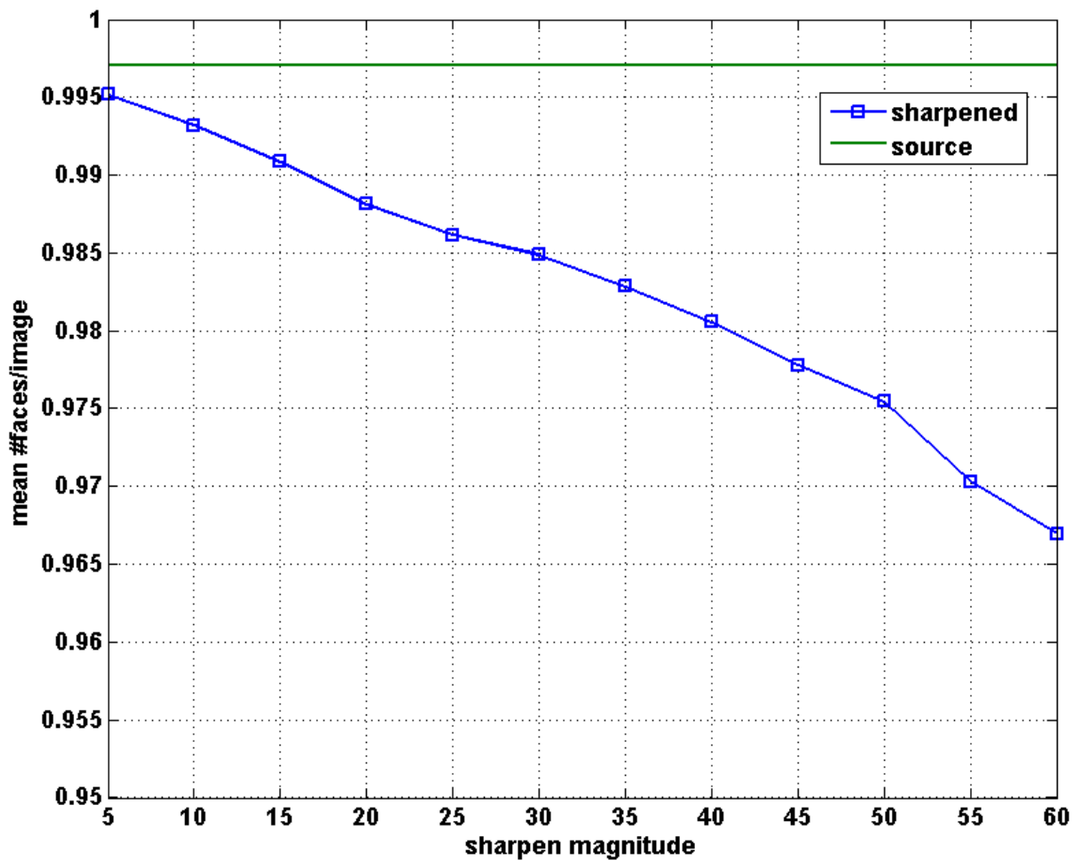
Test detection under presence of impairments

- SOA face detection method using Multi-task CNNs (Zhang et al., 2016)
- Impairments
  - Blurring: box filter of size *k* x *k*
  - Sharpening: unsharp masking, add difference of blurring with 3×3 binomial filter with magnitude m_s
  - JPEG compression: recompress with JPEG quality factor *q*
  - JPEG compression concealment: two passes with 4×4 box filter
- Datasets
  - Labelled Faces in the Wild (LFW): still images only (JPEG compressed)
  - Youtube Faces (YTF): labelled face tracks in video of various quality

Face detection on LFW

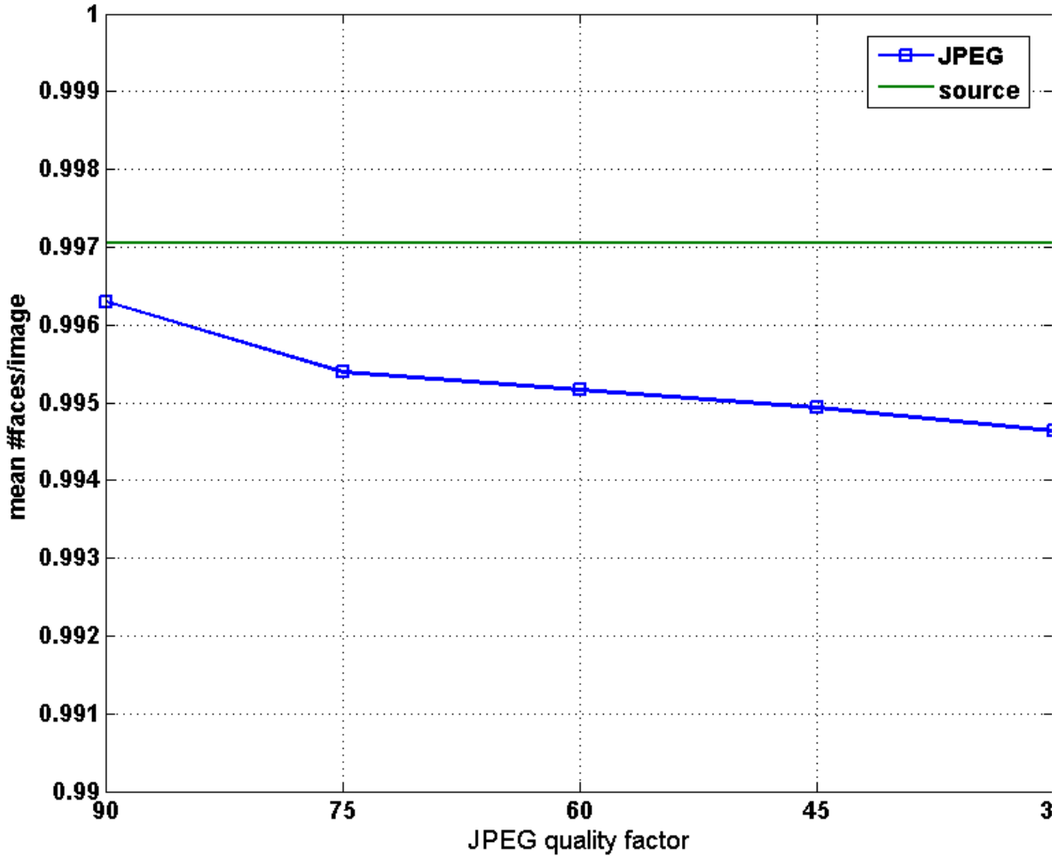
Blur and sharpen

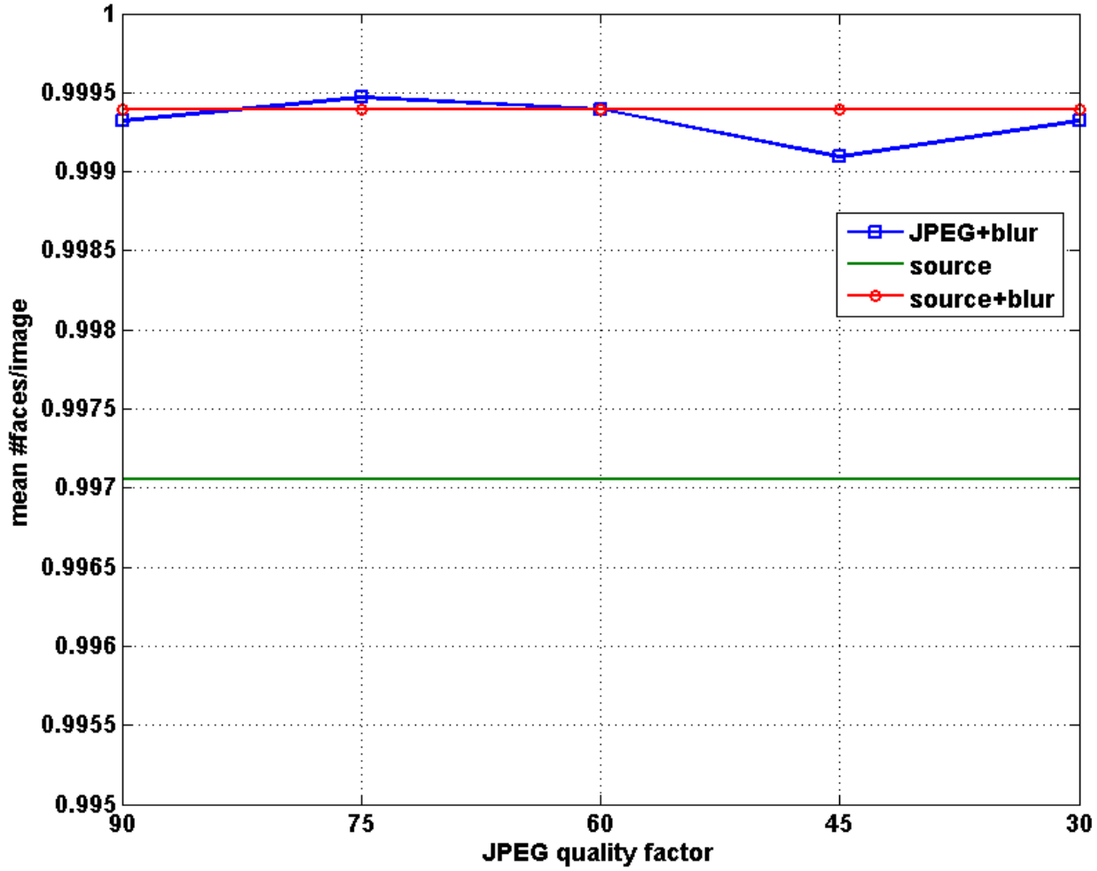




Face detection on LFW

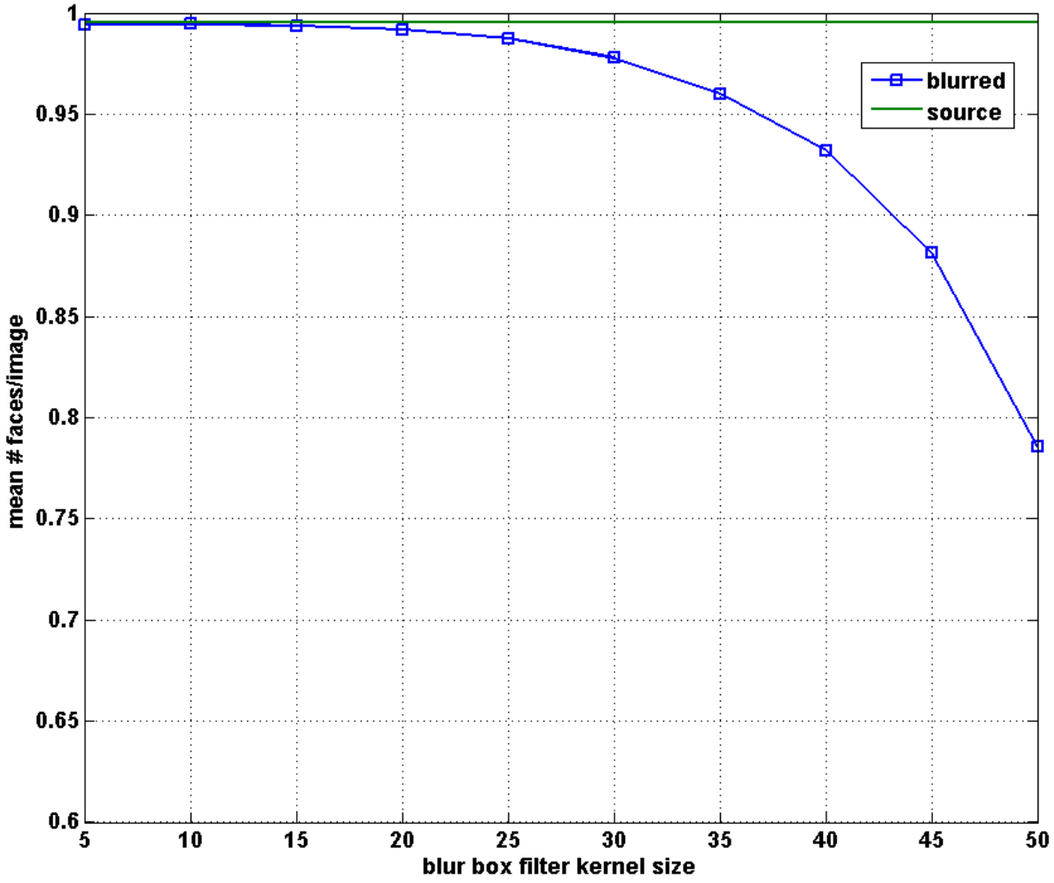
Compression

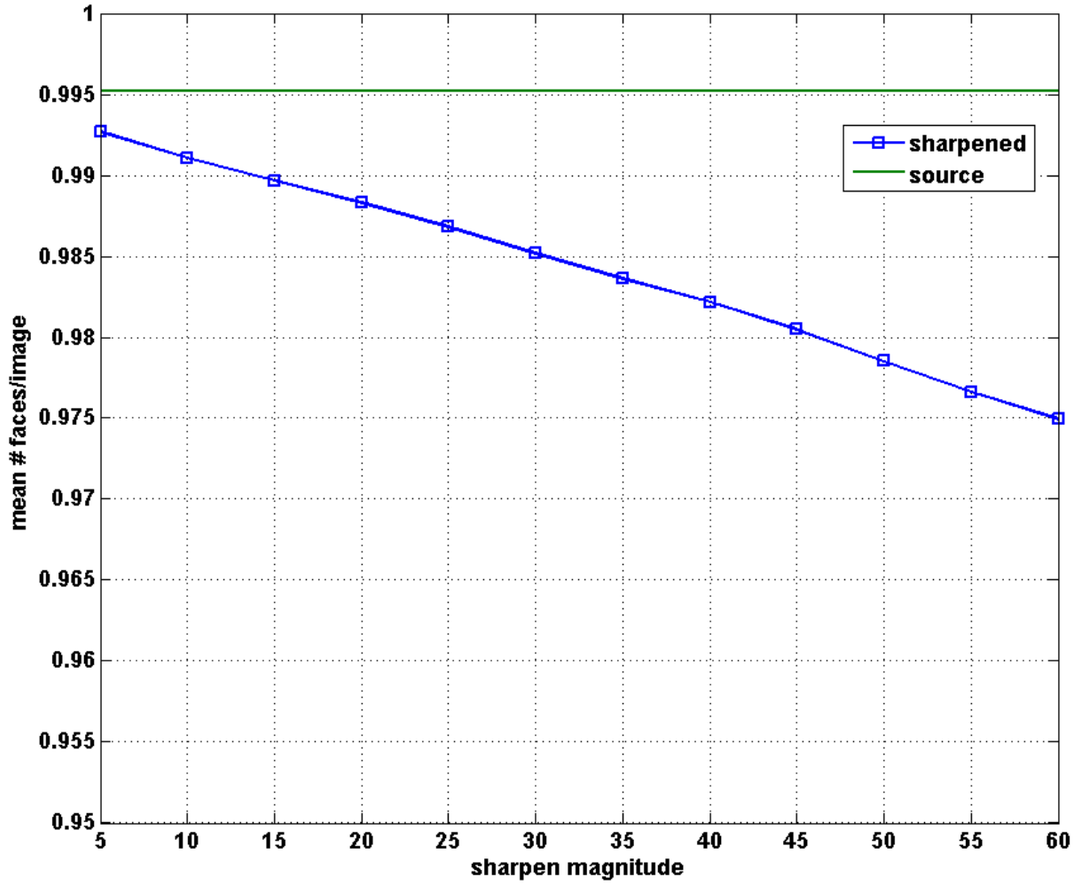




Face detection on YTF

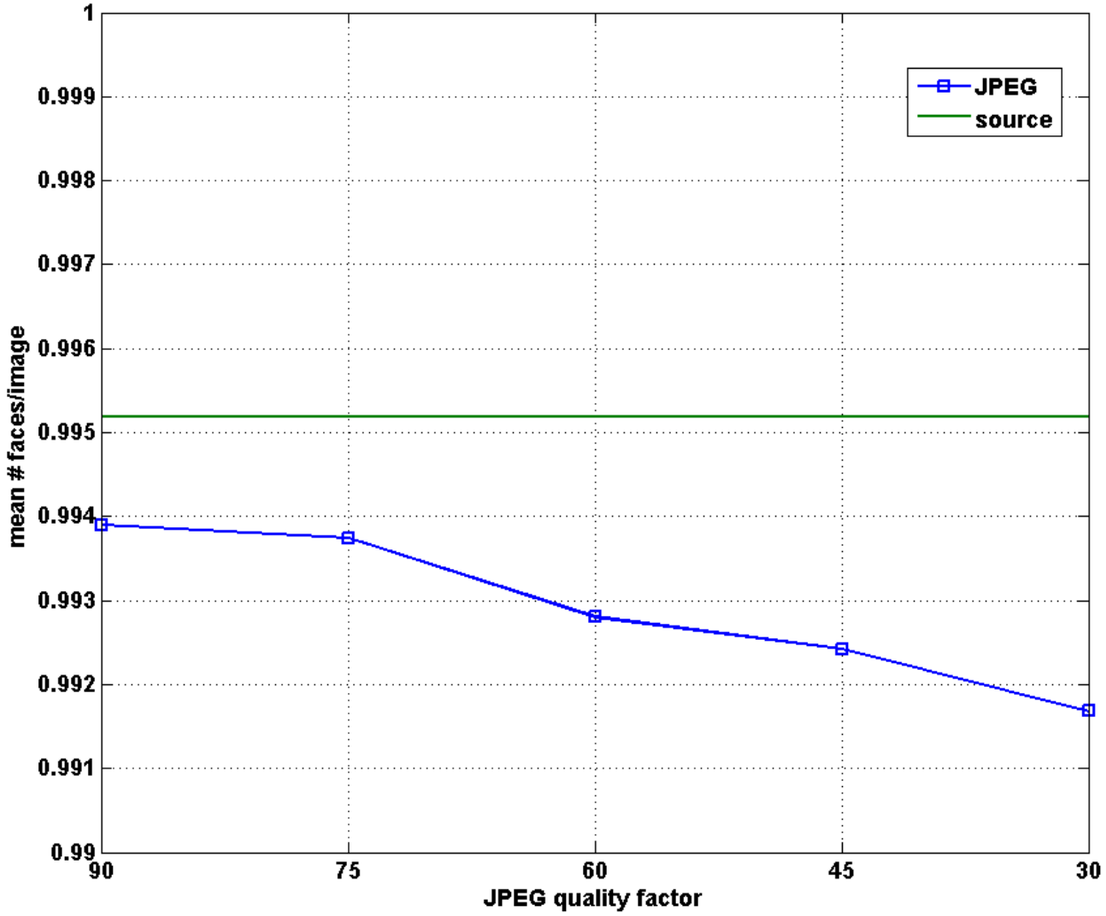
Blur and sharpen

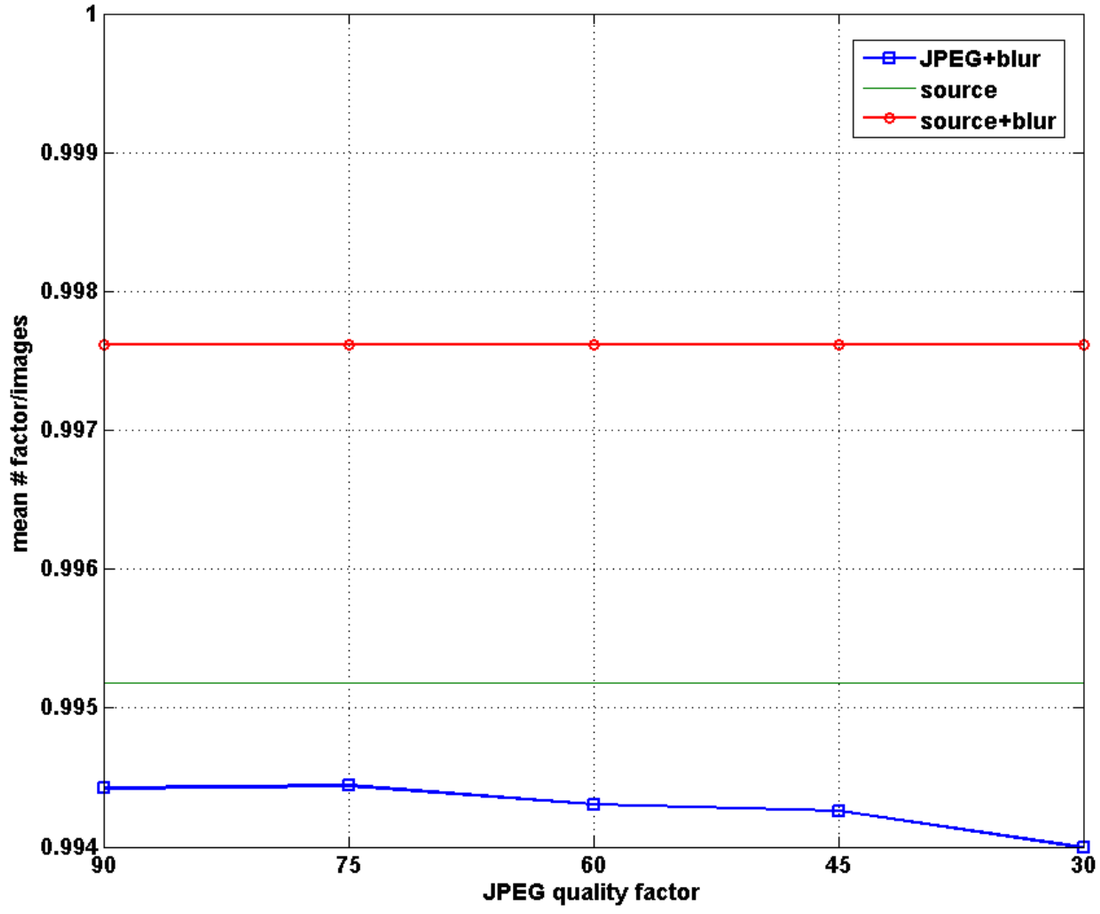




Face detection on YTF

Compression





Handling unknown faces

Open vs. closed set face recognition

- Most methods (and benchmarks) focus on closed set recognition, i.e. find best matching class
- Open set recognition also needs to decide whether face is known at all
- Results significantly lower than for closed set (75-85% vs. > 95%)
- Using approach with FaceNet features and Online Random Forest as classifier (Winter, 2019)
- Two improvements to decide whether candidate face is known in the database
  - Change calculation of classification confidence
  - Use correlation of feature of class prototypes and new sample
- Datasets
  - Labelled Faces in the Wild (LFW)

Handling unknown faces

Calculation of classification confidence

- Common approach in RF:

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

- Use ratio of first (=$C_{x,tot}$) and second class:

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

- Use ratio of first and mean of all classes:

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

- Combine the two:
  (β=0.75 in our experiments)

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

Handling unknown faces

Correlation

- Keep feature prototypes for each class (bag of features)
- If confidence >θ, determine correlation of features of sample and prototype:

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

- Consider a match, if correlation with at least one feature in the bag >τ
- In our experiments: θ=0.55, τ=0.75

Open set face recognition

Results on LFW

Measure	Baseline		ImprovedConf		+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%

Conclusion

Robust face detection

- Blurring and sharpening cause as expected performance loss proportional to impairments
- Compression has non-negligible impact on performance
- At high to moderate JPEG quality factors, performance loss is not due to a loss of information, but due to quantization noise
- Slight blurring does not cause reduction of detection performance, but can reduce JPEG quantization noise
- On high quality content concealment reaches same performance when starting from source or compressed version
- On content with higher compression (video) concealment improves results, but cannot go beyond the original quality

Conclusion

Open set face recognition

- Applying CNN-based features with incremental machine learning reaches SOA performance in open set recognition
- Improved confidence measure increases classification performance for known faces, but also number of falsely classified unknown faces
- Combination with a correlation-based check reduces the falsely classified unknown faces (at small cost in terms of true known face classifications)

Questions ?

JOANNEUM RESEARCH
DIGITAL – Institute for Information and Communication Technologies

werner.bailer@joanneum.at

www.joanneum.at/en/digital

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 761802, MARCONI and no. 761934, Hyper360.