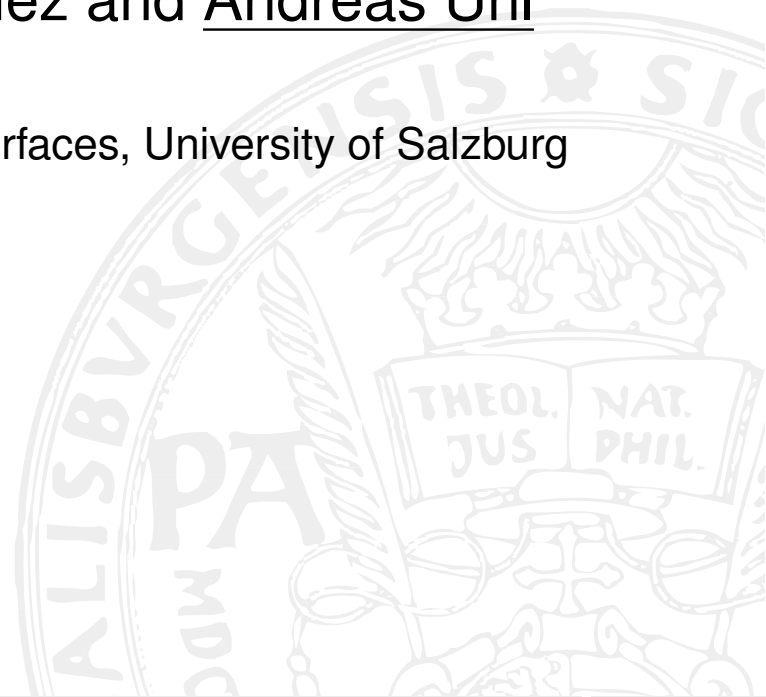
Impact of Deep Learning on Facial Data Security

Heinz Hofbauer, Simon Kirchgasser, Lukas Lamminger, Yoanna Martinez-Diaz, Heydi Mendez-Vazquez and Andreas Uhl

Department of Artificial Intelligence and Human Interfaces, University of Salzburg

June 27th, 2023



Biometric Backgound

- In face recognition, several techniques have been developed over the years to protect facial templates as well as facial sample data.
- With the rise of deep learning techniques, face recognition as seen a revolution in terms of achievable recognition accuracy.
- These developments also had impact on the security techniques applied to facial data:
  - Early template protection schemes do not provide security any longer
  - The security of existing selective encryption schemes developed for facial data has to be re-considered
- We show that learning-based inpainting schemes can be used to successfully attack selectively encrypted facial data.

This talk is a blend of several (4) presentations held over the last 3 years:

Papers Covered

- Is Warping-based Cancellable Biometrics (still) Sensible for Face Recognition ? Simon Kirchgasser, Yoanna Martínez Díaz, Heydi Mendez-Vazquez, Andreas Uhl. Proceedings of the IAPR/IEEE International Joint Conference on Biometrics (IJCB '20).
- Highly Efficient Protection of Biometric Face Samples with Selective JPEG2000 Encryption. Heinz Hofbauer, Yoanna Martínez-Díaz, Simon Kirchgasser, Heydi Méndez-Vázquez, Andreas Uhl. Proc. of the IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP '21).
- Utilizing CNNs for Cryptanalysis of Selective Biometric Face Sample Encryption. Heinz Hofbauer, Yoanna Martínez-Díaz, Luis Santiago Luevano, Heydi Méndez-Vázquez, Andreas Uhl. Proceedings of the 26th International Conference on Pattern Recognition (ICPR'22).
- First Learning Steps to Recognize Faces in the Noise. Lukas Lamminger, Heinz Hofbauer, and Andreas Uhl. Proceedings of the 11th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'23).

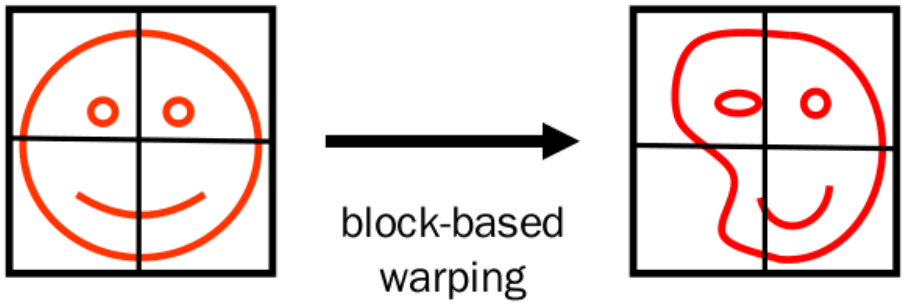
Template Protection

Cancelable Biometrics

- Template is generated using a system-specific or user-specific key (so we result in a two factor authentication)
- Template comparison is done on protected templates
- In case of compromise, the template is revoked and re-issued with a different key
- Non-invertability is a must (knowing the key), thus, homomorphic encryption is not template protection
- For most techniques, we see a trade-off between security, computational cost, and recognition accuracy

Here we consider block-based warping, which was introduced by IBM (together with block re-mapping) about 20 years back as a fist class of cancellable biometrics.

Block-based Warping



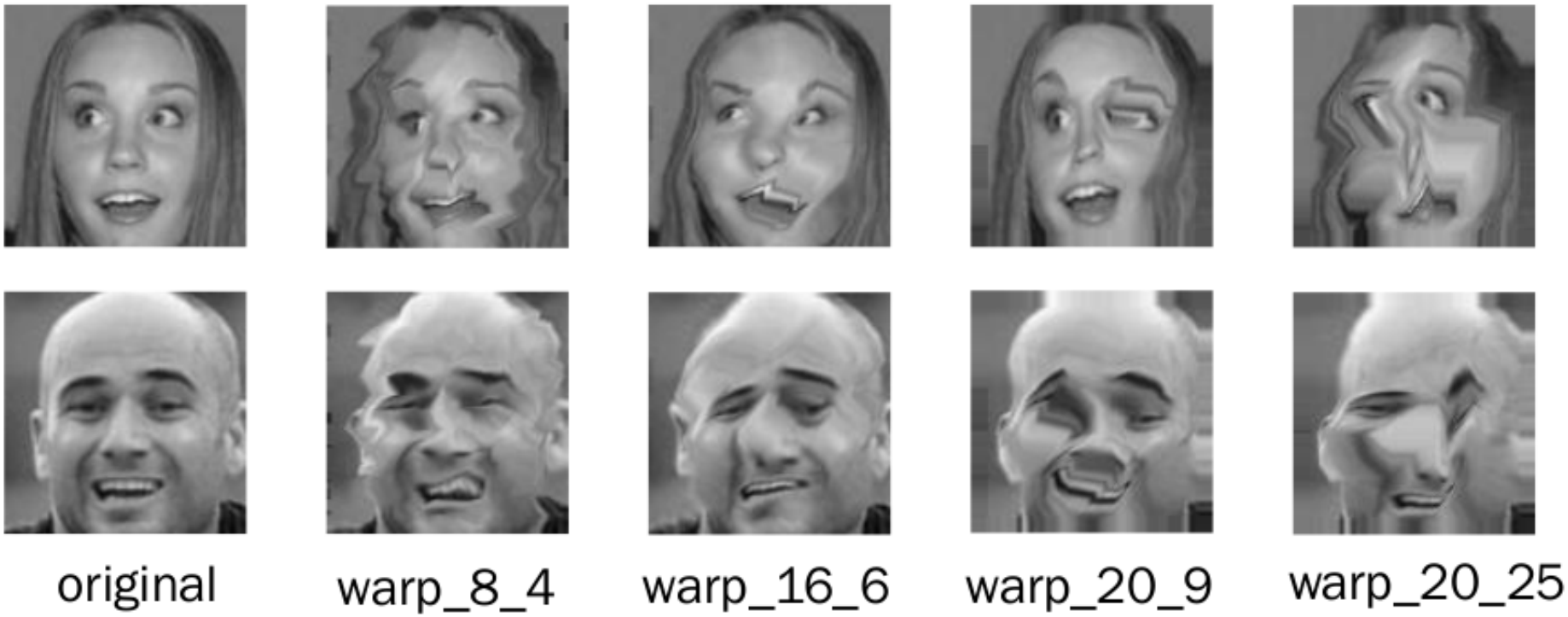


Figure: Block-based warping shown schematically and applied to “Labeled Faces in the Wild”.

Template Protection: Properties

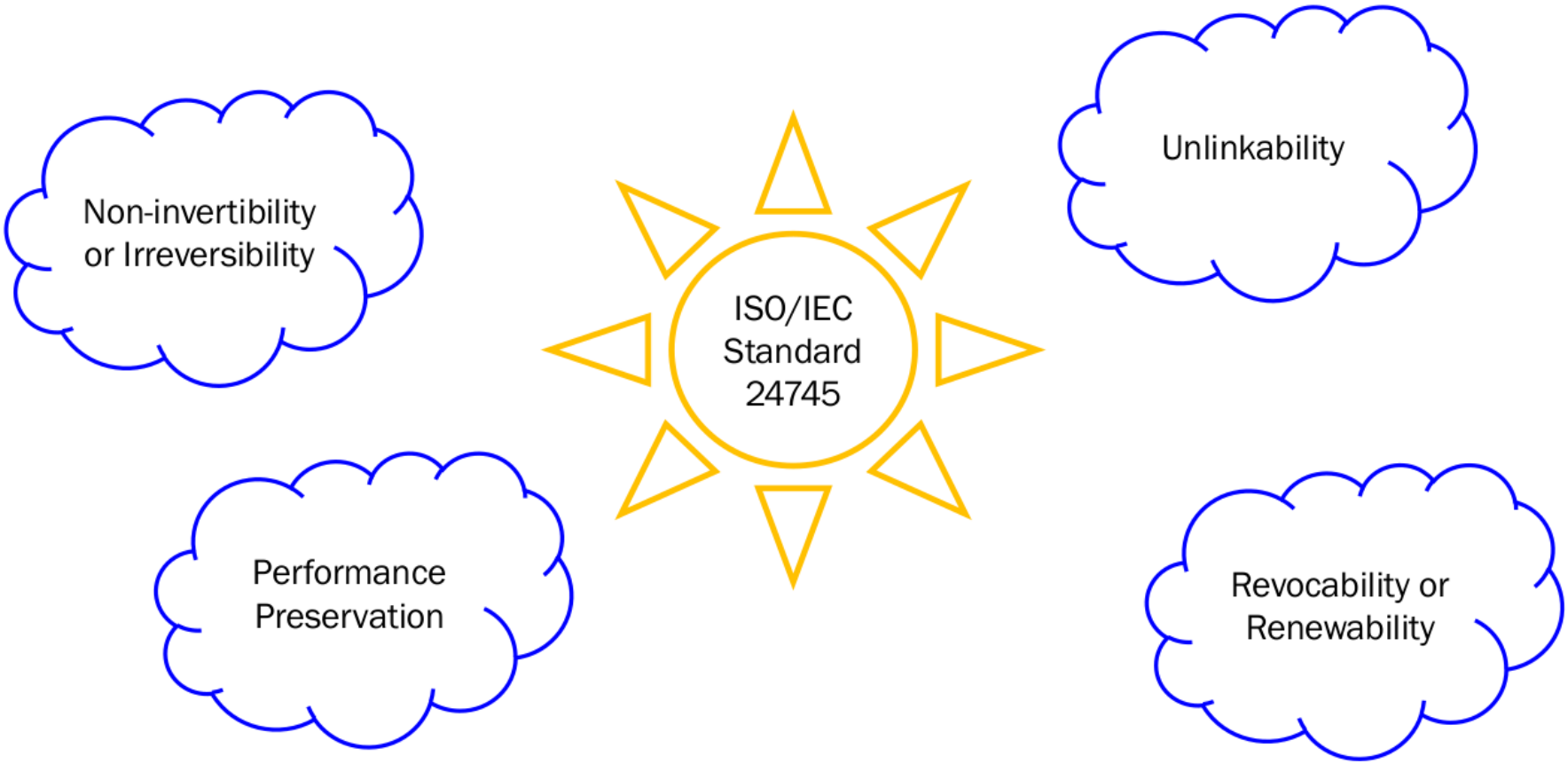


Figure: ISO/IEC 24745.

Template Protection: Evaluation

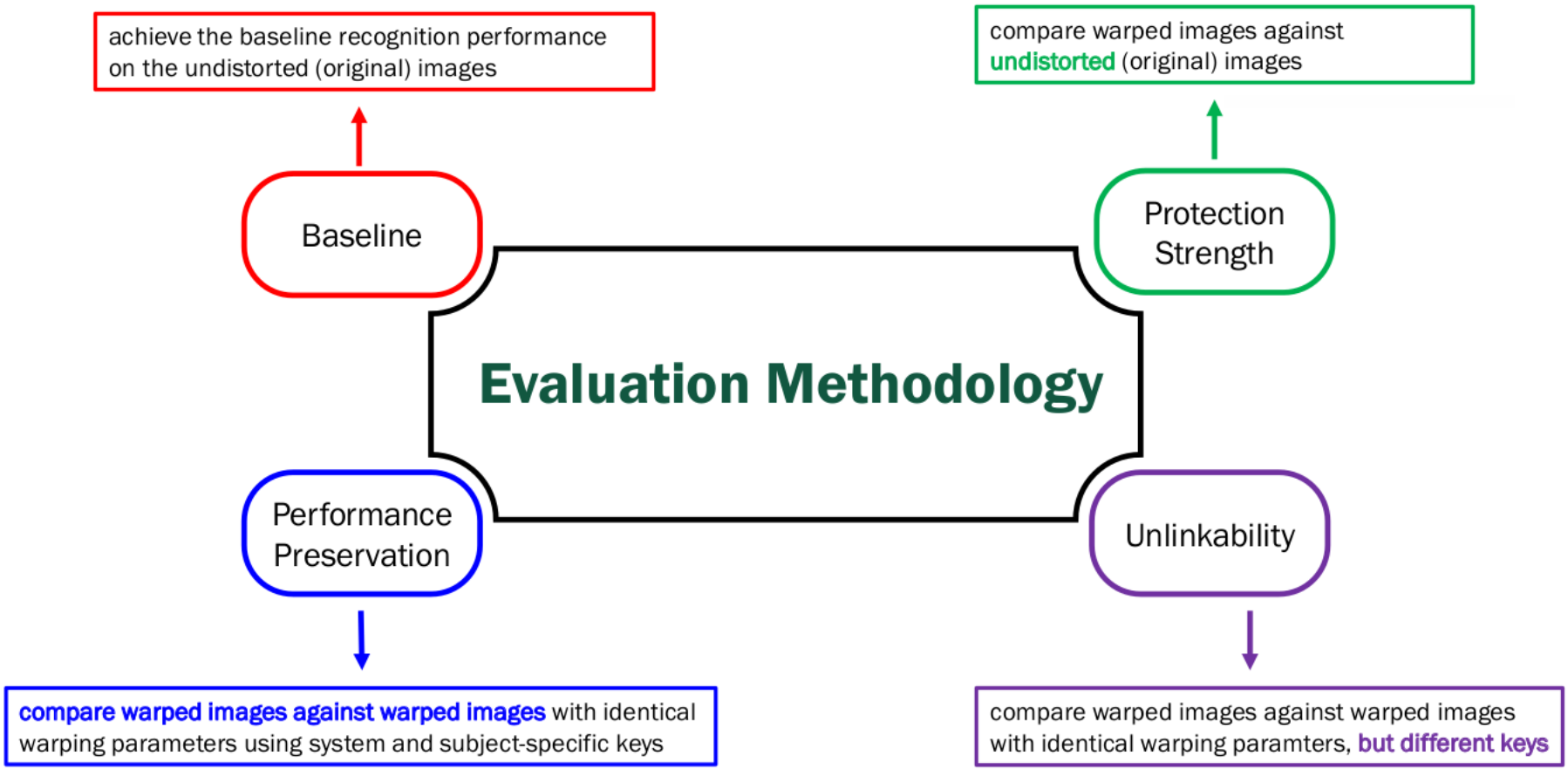


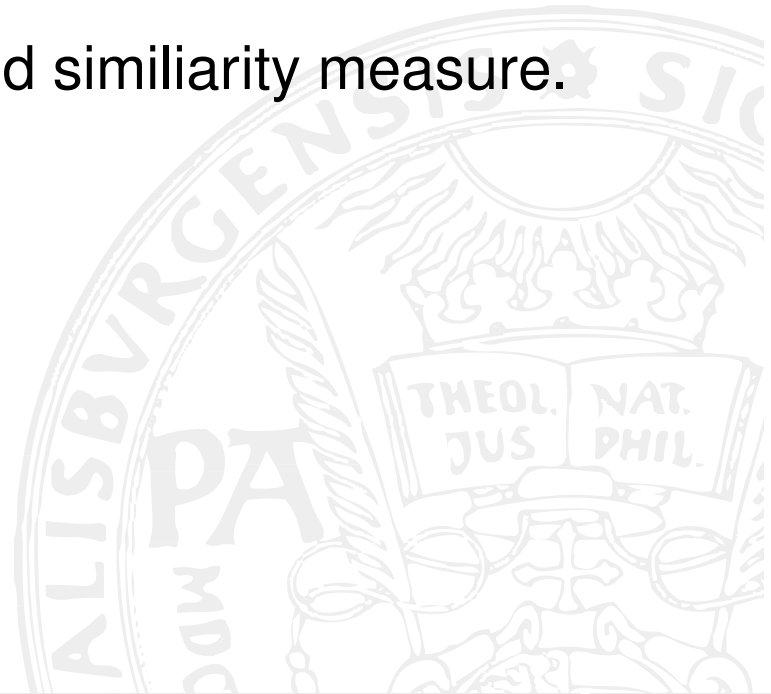
Figure: Evaluation dimensions.

Face Recognition Methods

- Traditional
  - Local Binary Patterns (LBP)
  - Multi-Block LBP (MBLBP)
- Advanced
  - FisherVector SIFT
  - FisherVector BRIEF

Histogram per 14x14 cell region, chi-squared similiarity measure.

- CNN based methods
  - ResNet-ArcFace (ArcFace)
  - MobileFaceNet (MobileFace)
  - ShuffleFaceNet (ShuffleFace)



DL-based vs. Descriptor-based Face Recognition

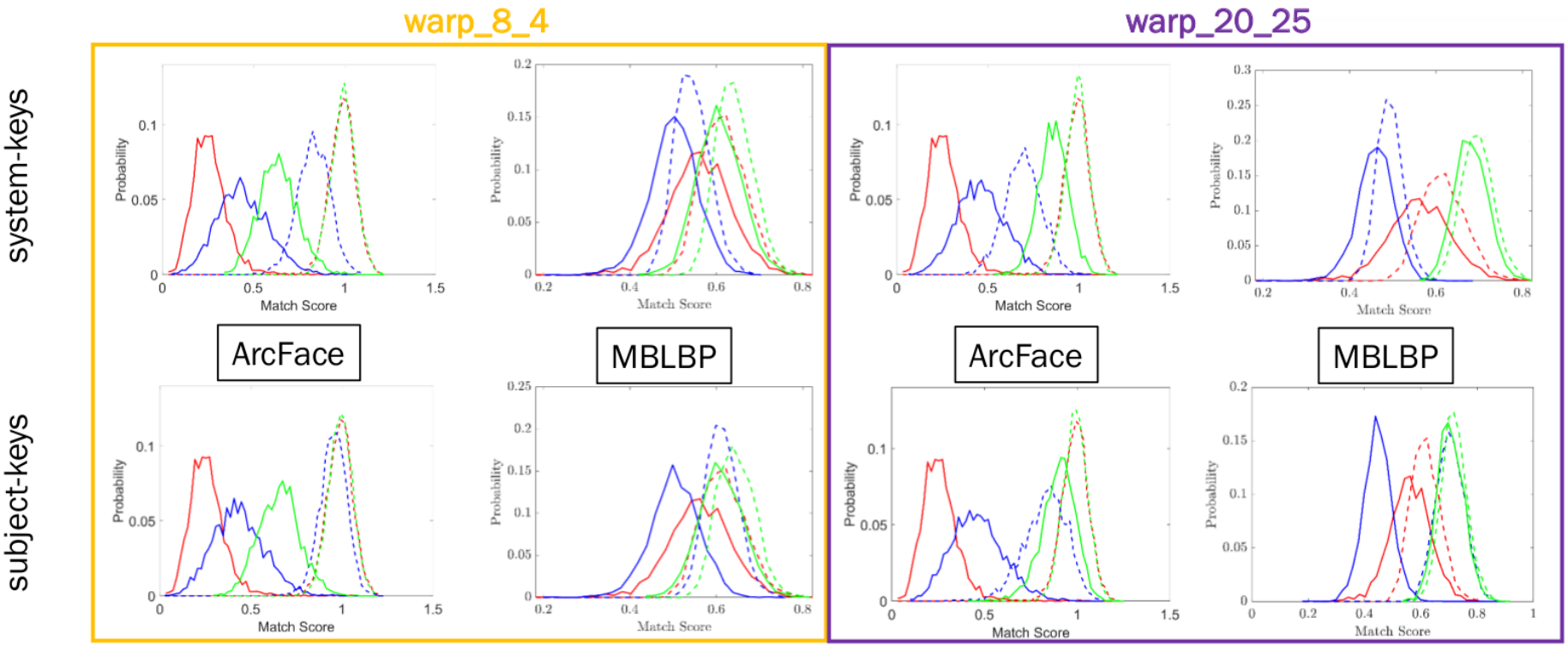


Figure: Red - baseline, blue - performance preservation, green - protection strength; genuine scores - solid lines, impostor scores - dashed lines.

Performance preservation (blue): Distributions should stay apart

Protection strength (green): Distributions should overlap

Unlinkability

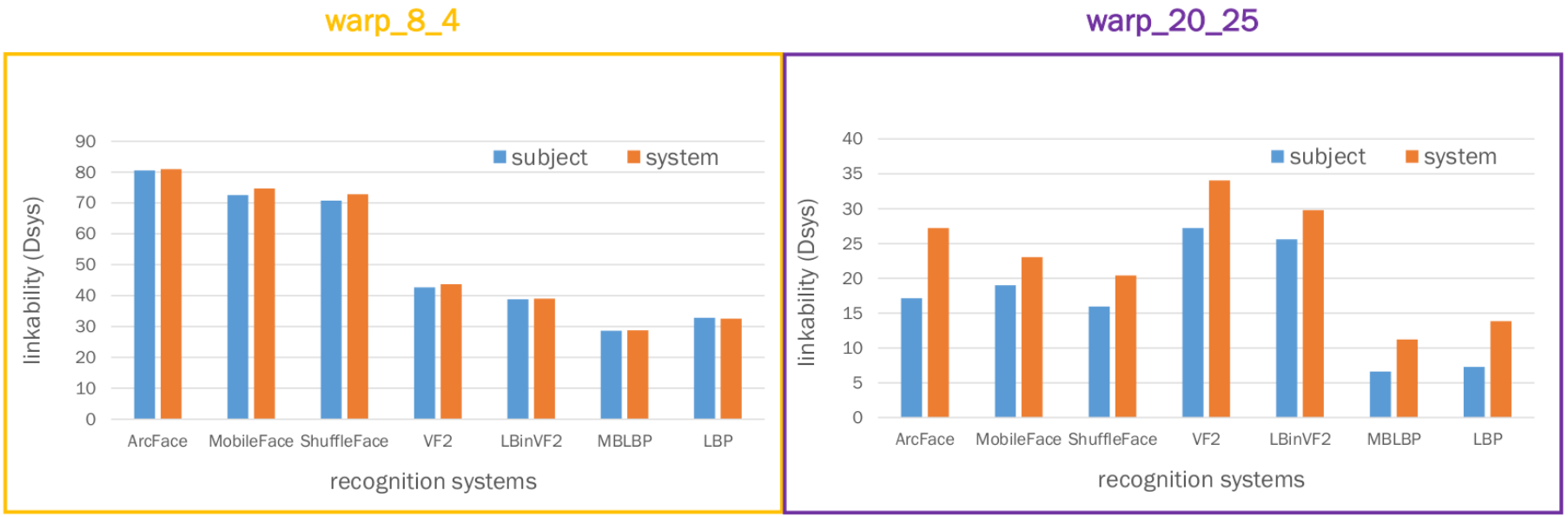


Figure: High value $\Longrightarrow$ high linkability (bad property).

Overall, all template protection performance indicators are significantly impared when migrating from descriptor-based to DL-based face recognition !!

Selective Encryption of Facial Samples

Biometric Backgound

- ISO specifies biometric data to be recorded and stored in (raw) image form (ISO/IEC FDIS 19794), i.e., sample images, and not only in extracted templates.
- The (off-line) database of user enrollment samples typically contains such data, the ISO/IEC Standard 19794-5:2011 suggests JPEG or JPEG2000 encoding for facial sample images.
- Such deployments benefit from future improvements which can be incorporated without re-enrollment of registered users, thereby increasing interoperability and vendor neutrality.
- The (JPEG2000) compressed biometric sample data will be optimally protected via encryption by a state of the art cipher.

Here we look at an efficient selective encryption approach for facial samples by applying strong encryption techniques to specific parts of the data.

Relevant Components of a Face Recognition System

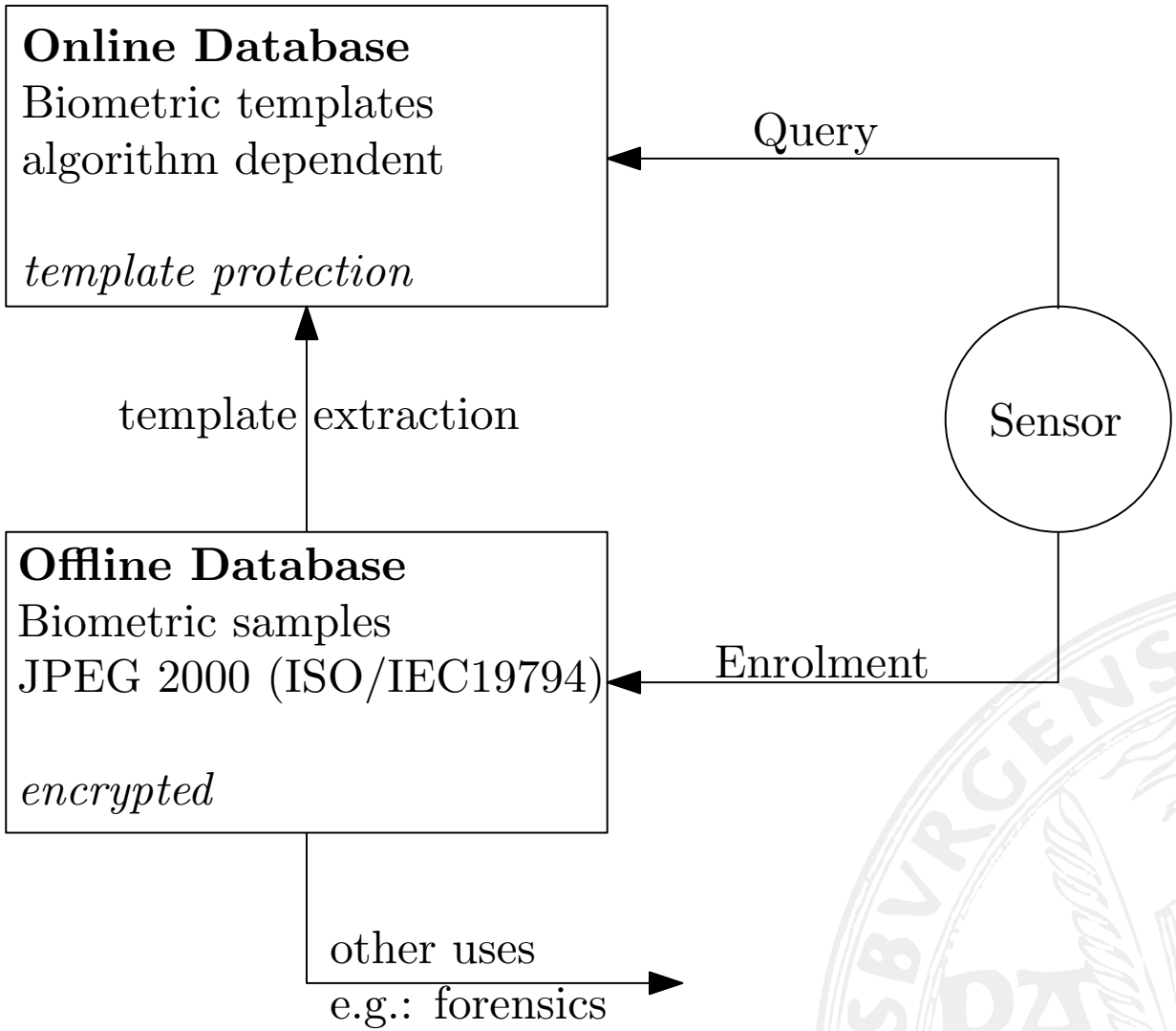


Figure: Illustrating the difference between Off-line and On-line database.

Why using selective encryption for facial sample data ?

Whenever a sample from the offline database needs to be accessed it needs to be *decrypted* first, which is time consuming, in particular when the whole database needs to be accessed, if e.g.

- the biometric comparison or template extraction technique is changed (the alternative would be a re-enrollment of all users);
- the key used in the employed TP scheme needs to be changed due to a periodic update (as a preventive measure to guard against undetected loss of a key) or because it has been lost in an attack or data breach (to reestablish security).

Other scenarios are given if one or only a few samples in the offline database need to be decrypted, e.g.:

- Template regeneration for single users, e.g., in case an ATM card with biometric template is lost and a replacement card has to be issued, or a user specific TP key needs to be changed;
- Explicit sample comparison in forensic identification or in context of de-duplication.

$\Longrightarrow$ Therefore, computationally efficient encryption is important.

Selective Encryption of JPEG2000 Data

For encryption we used a format compliant JPEG2000 (J2K) encryption scheme:

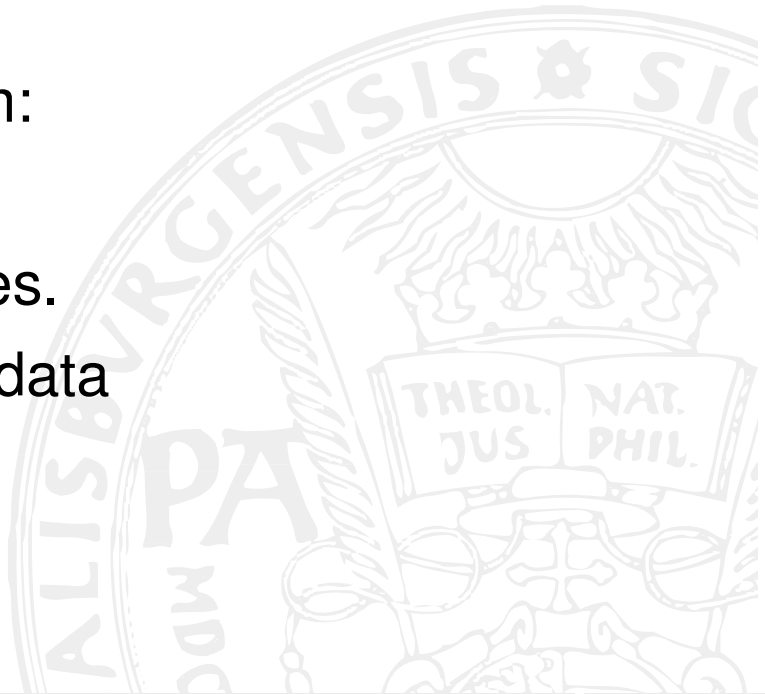
- Format compliance is important as it allows to decode the image by a standards compliant J2K decoder.
- Error-concealment attack applied (best effort attack): There is an inbuilt error correction in a J2K decoder which was used in all experiments and during training.
- We use a sliding window encryption where a window of variable size (in % of the total bitstream) is encrypted and the offset where the encryption starts (again in % of the bitstream) varies: Offset where encryption starts (o) and the window size (w).
- Images were then encoded with lossless J2K, differing only in the progression type: Layer (l) and resolution (r) progression.
- As a certain amount of data is left in plain text, security analysis is required !

Which Bitstream Parts should be Encrypted ?

- Where is the most relevant information for the face recognition algorithms?
- What is the minimum amount of encryption required to protect the biometric face sample?

Focus is on the beginning of the codestream:

- Beginning: Structural data
- Towards end: refinement for fine textures.
- face information depends on structural data



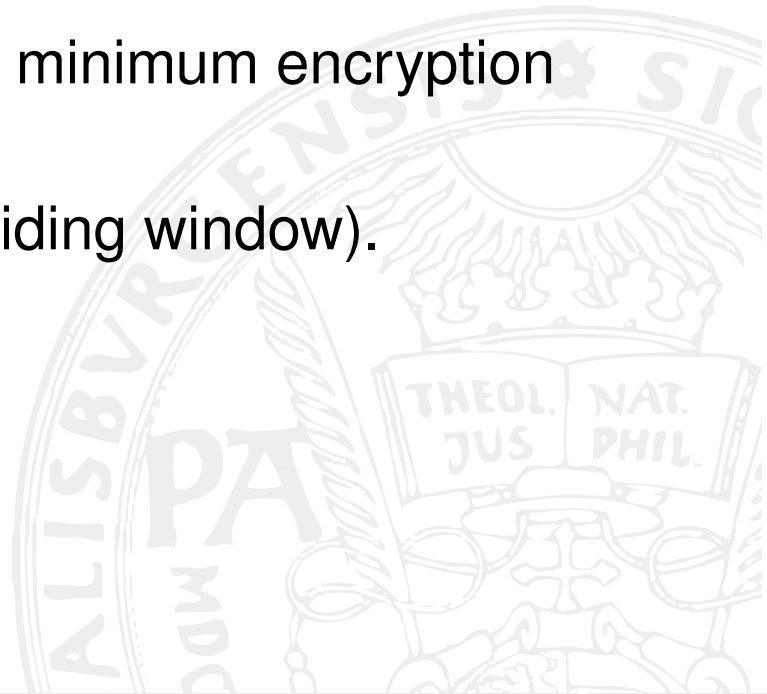
Evaluation Methodology I

Sliding Window Encryption to detect the location of relevant data.

- A small part of the codestream is encrypted (window)
- Offset is varied (sliding window)

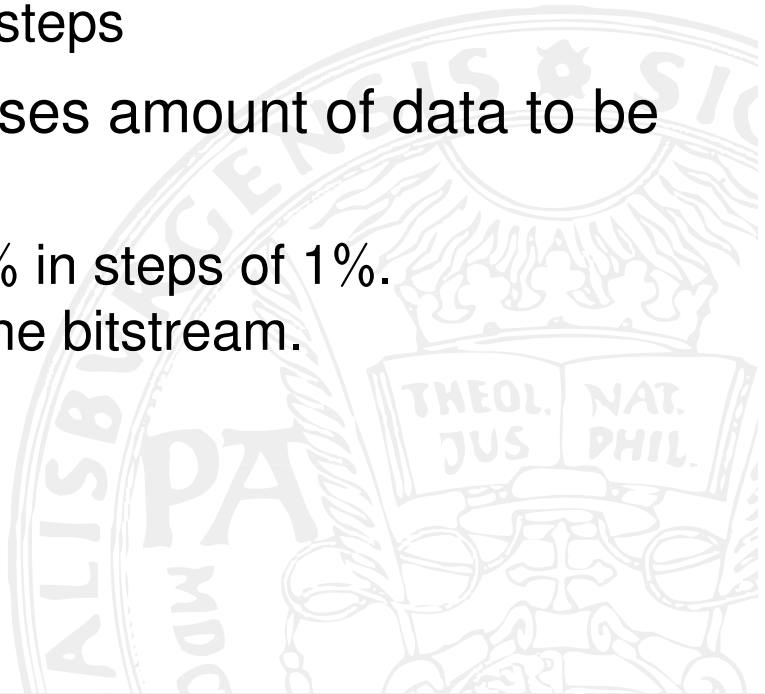
Increasing Window Encryption to find the minimum encryption emount.

- Offset is fixed from the beginning (no sliding window).
- Encryption amount is increased.



Evaluation Methodology II

- **small encryption window** is a sliding window encryption
  - window size is 0.5% of the bitstream
  - offset varies from 0% to 15% in steps of 1%
- **large encryption window** is a sliding window encryption
  - window size is 4%
  - offset is varied from 0% to 20% in 2% steps
- **increasing encryption window** increases amount of data to be encrypted encryption
  - window size increases from 1% to 15% in steps of 1%.
  - Encryption starts at the beginning of the bitstream.



Visual Examples

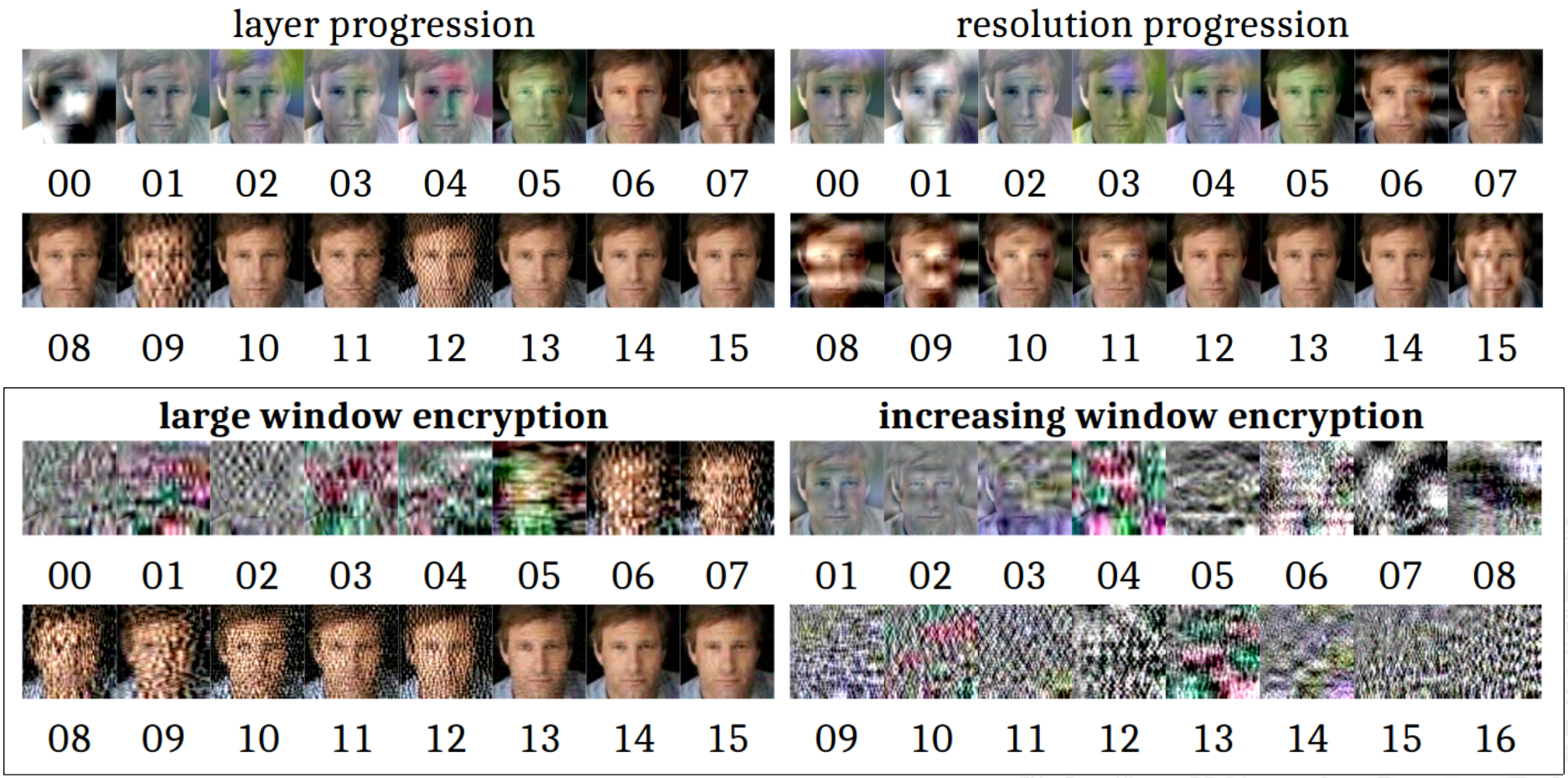


Figure: Illustrating the difference between JPEG2000 layer progressive and resolution progressive (1st & 2nd rows for small window encryption, increasing offset) and the difference between large window encryption (increasing the offet) and increasing window encryption.

Results: Small Window Encryption

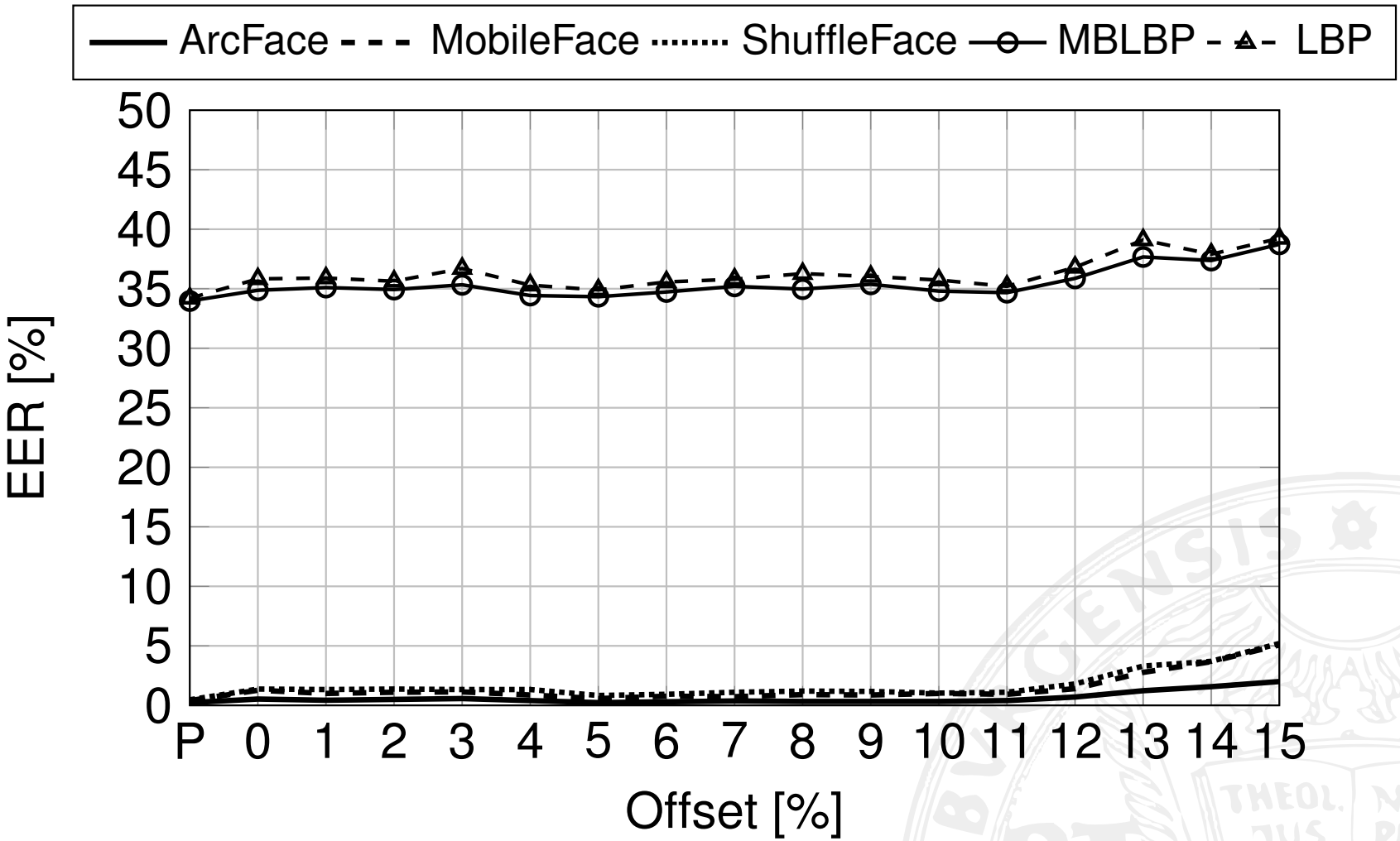


Figure: Resolution progression with error correction.

Results: Small Window Encryption

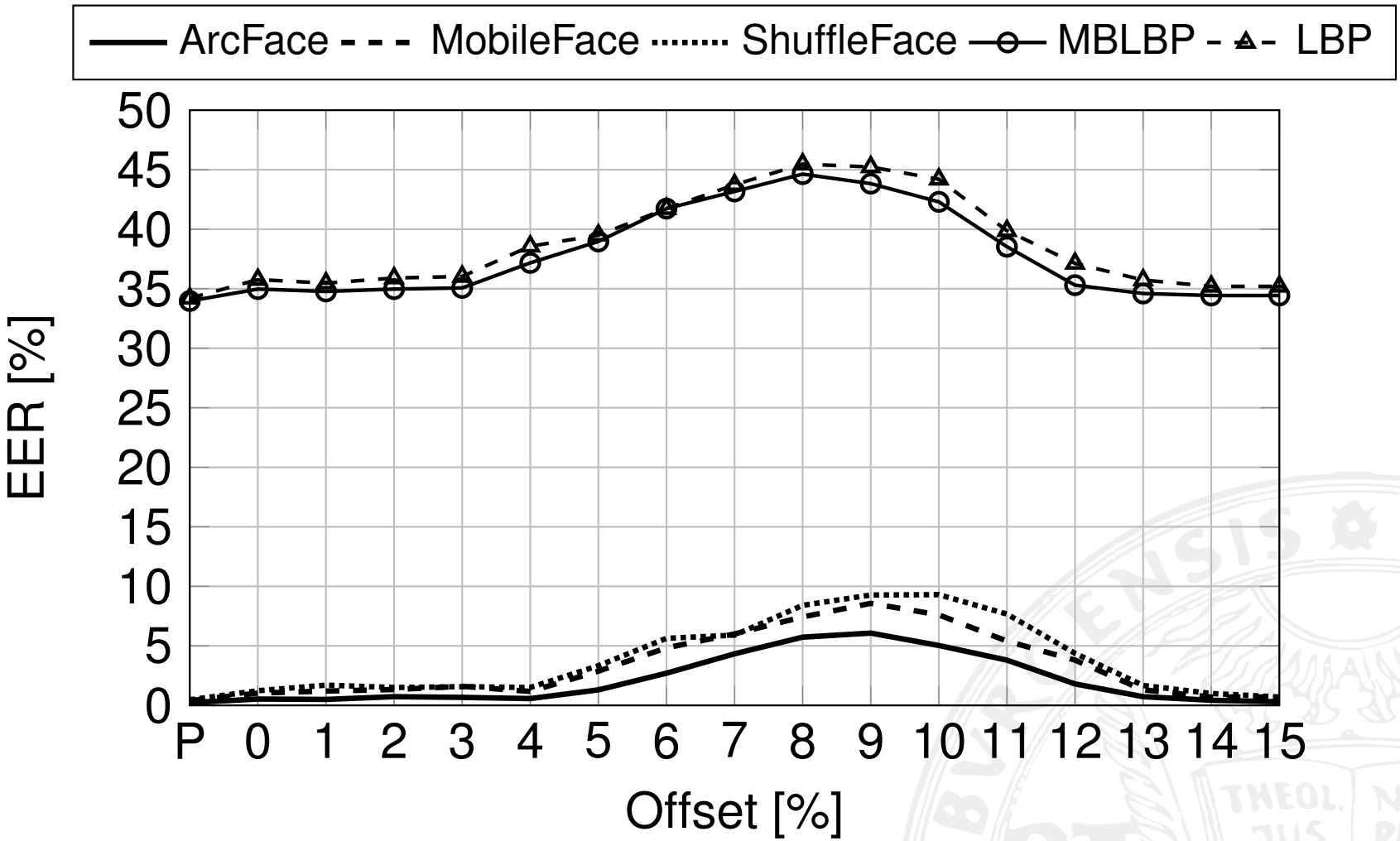


Figure: Layer progression with error correction.

Results: Large Window Encryption

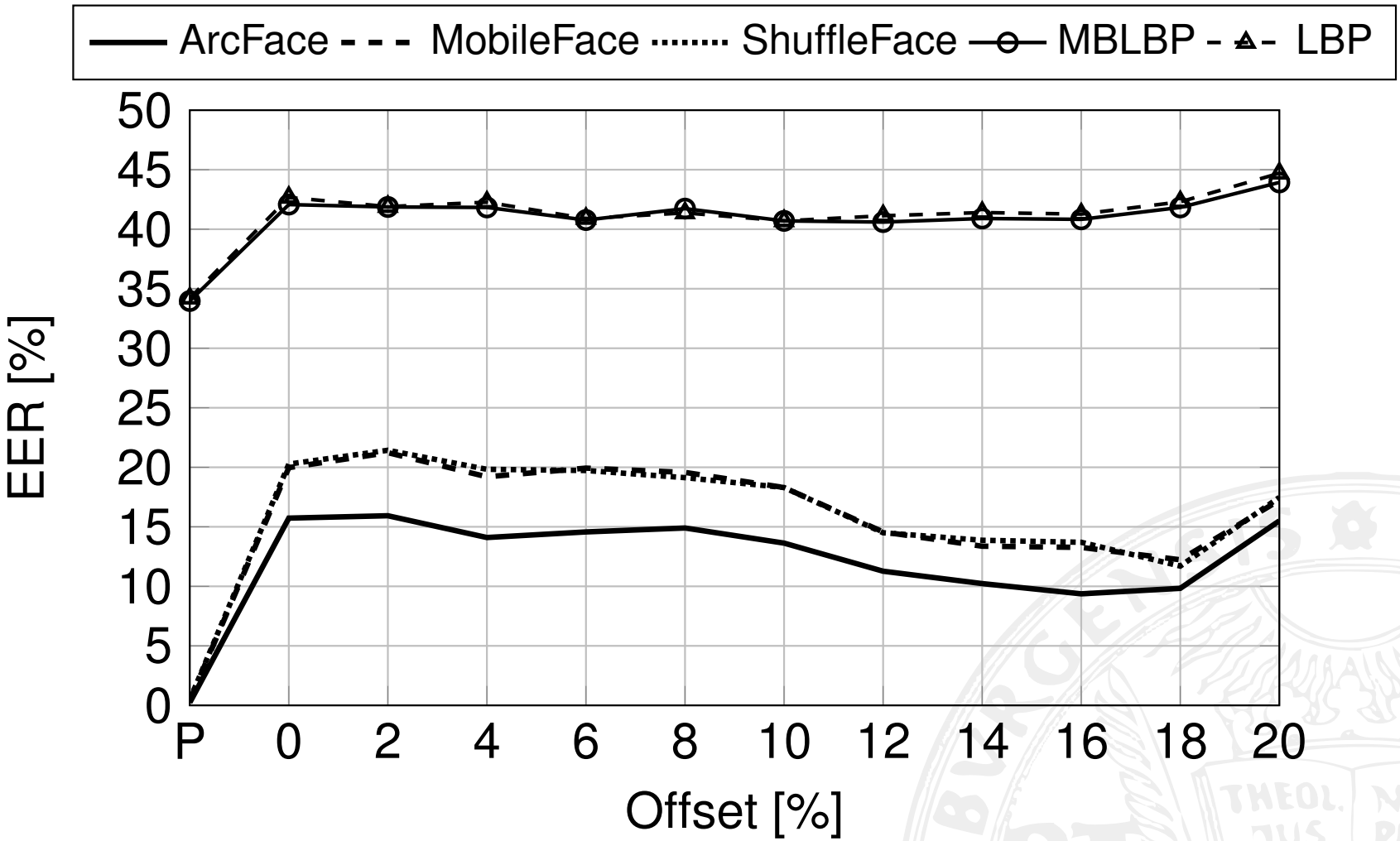


Figure: Resolution progression with error correction.

Results: Large Window Encryption

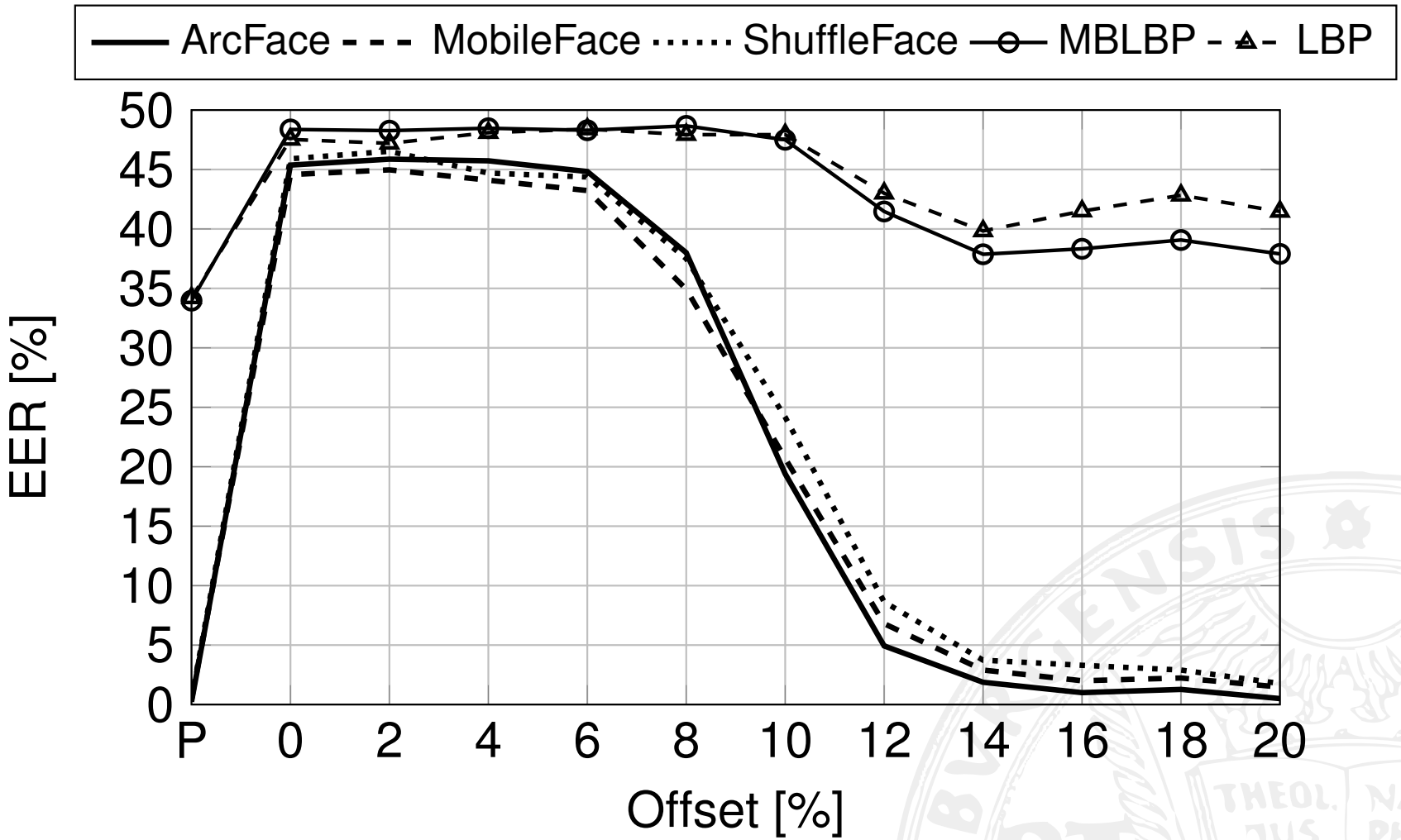


Figure: Layer progression with error correction.

Results: Increasing Window Encryption

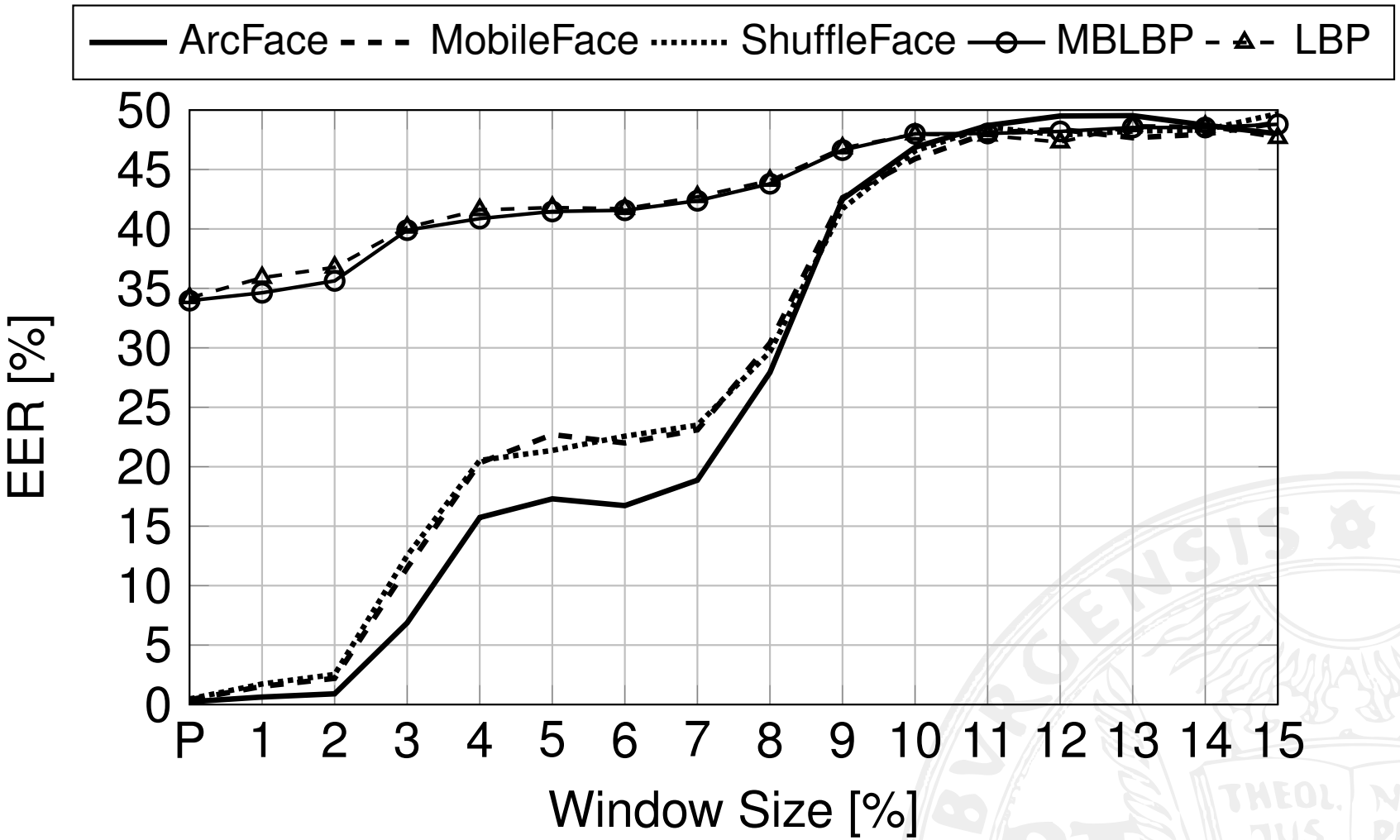


Figure: Resolution progression with error correction.

Results: Increasing Window Encryption

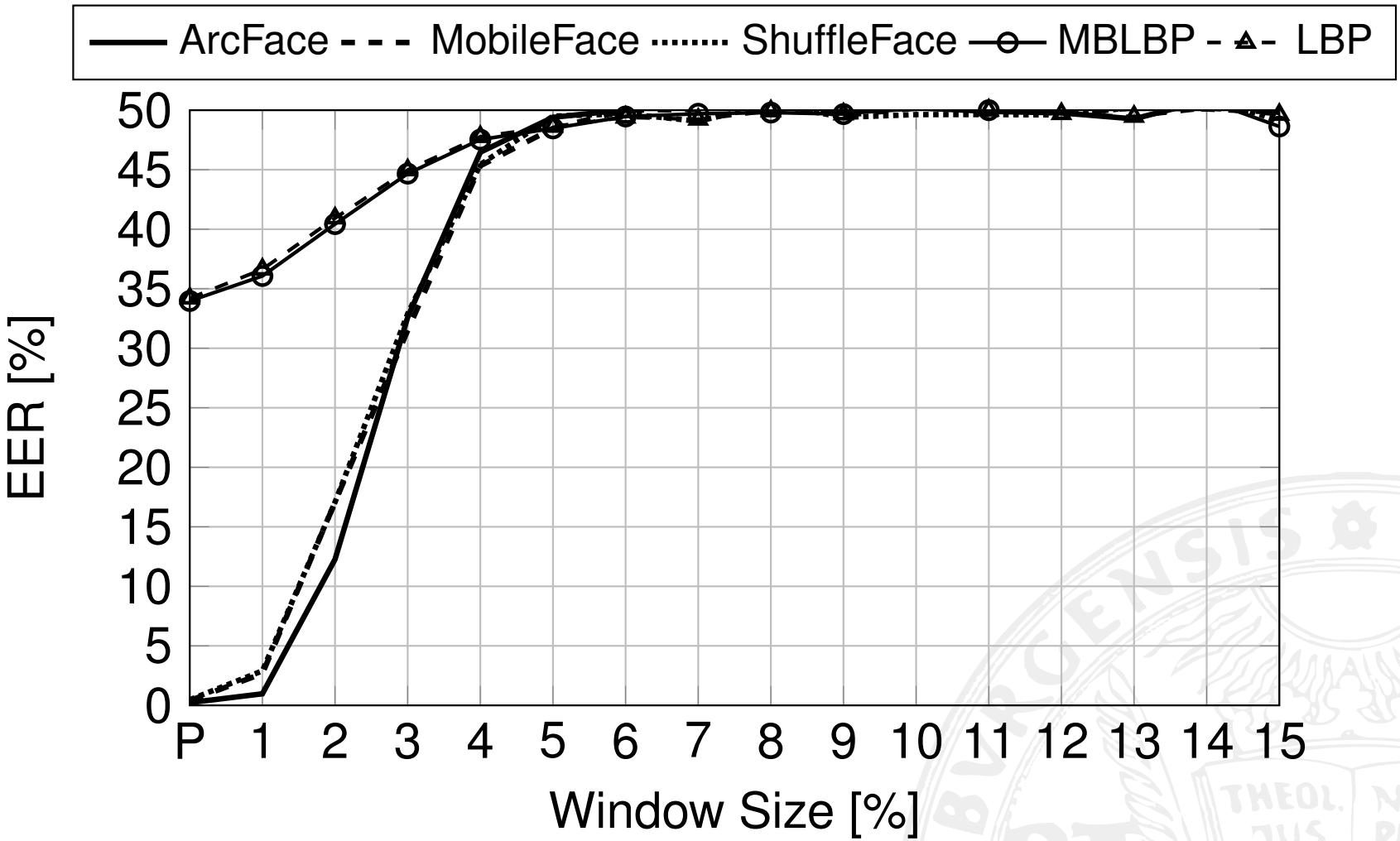


Figure: Layer progression with error correction.

Findings

- Traditional and deep learning based methods exhibit a very different behavior regarding recognition capabilities on “protected” data.
- Strong protection can only achieved when encrypting from the start of the bitstream for DL-based schemes, far more options are available when considering traditional face recognition.
- The relevant part for biometric face recognition is at around 4–12% of the total layer progressive codestream.
- The most secure method for encryption is to start at the beginning and at least include the first 12% (for traditional techniques much less is required to be encrypted).

Cryptoanalysis 2.0 using CNNs

Re-training Face Recognition

- Cryptoanalysis so far judges its success with face recognition trained on pristine facial data.
- We apply a potentially stronger attack for cryptanalysis by enhancing the training of the face recognition CNN.
- We will take a closer look at the amount of data in J2K encoded and selectively encrypted images which can be used for the facial biometric recognition under **optimal** attack conditions.
- We assume that it is known at training time of the system that such protected face data might be presented to the system.
- Refinement training is done using partially encrypted samples, i.e. CNN is trained for recognition based on protected samples.

Attack Scenario

Does this approach, i.e. refinement training of the face recognition CNN, represent a realistic attack scenario ?

Rather no - how should the attacker get access for re-training the face recognition system ?

However, if information which can lead to a correct biometric recognition is still contained in the encrypted images and can be exploited in some way, then it is conceivable that an attacker could extract it as well by other means without retraining the system (as we shall see in the last part of this presentation, where such an attack is presented and successfully applied).

Selection of Experimental Settings

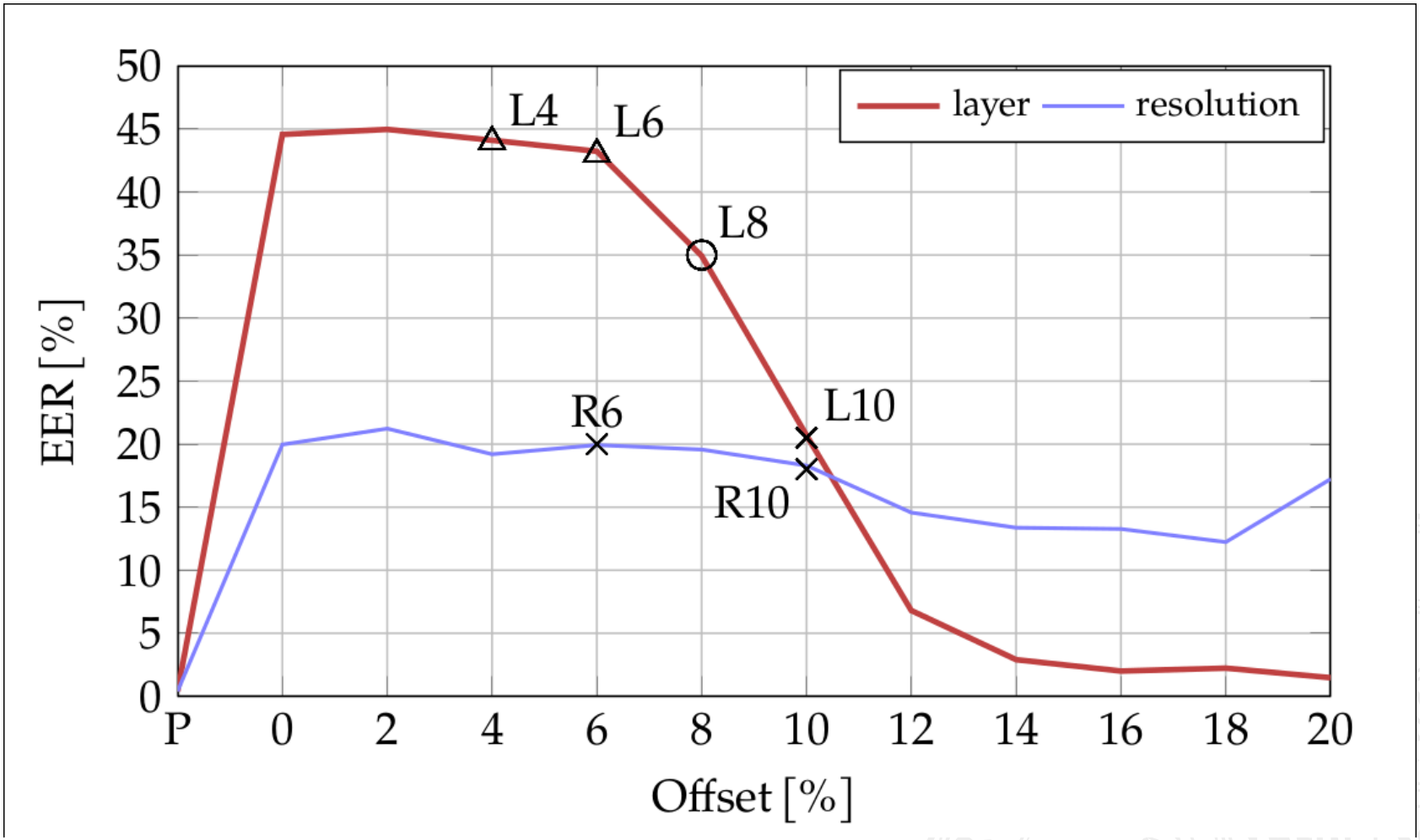


Figure: Results of the sliding window encryption (window size 4%) and the given offset (P is for original) on “Labeled Faces in the Wild” (using MobileFaceNet). Settings with comparable EER are chosen, these are used for training.

Experimental Settings

We aim at a setting with reasonable generalisability.

- For training, we select CASIA-WebFace in the encryption settings as specified (window encryption with 4% data encrypted in resolution or layer progressive JPEG2000 mode, starting at varying positions in the bitstream (offset).
- Pre-trained MobileFaceNet is fine-tuned with differently encrypted data as desribed.
- Protected data to be recognised: Increasing window encryption (i.e. starting from the bitstream beginning) is applied to LFW data, both encrypted in resolution or layer progressive JPEG2000 mode.

$\Longrightarrow$ mismatch between facial datasets (subjects, imaging) and encryption type (increasing window vs. sliding window), partially also between JPEG2000 progression.

Results: Encryption of Layer Progressive JPEG2000

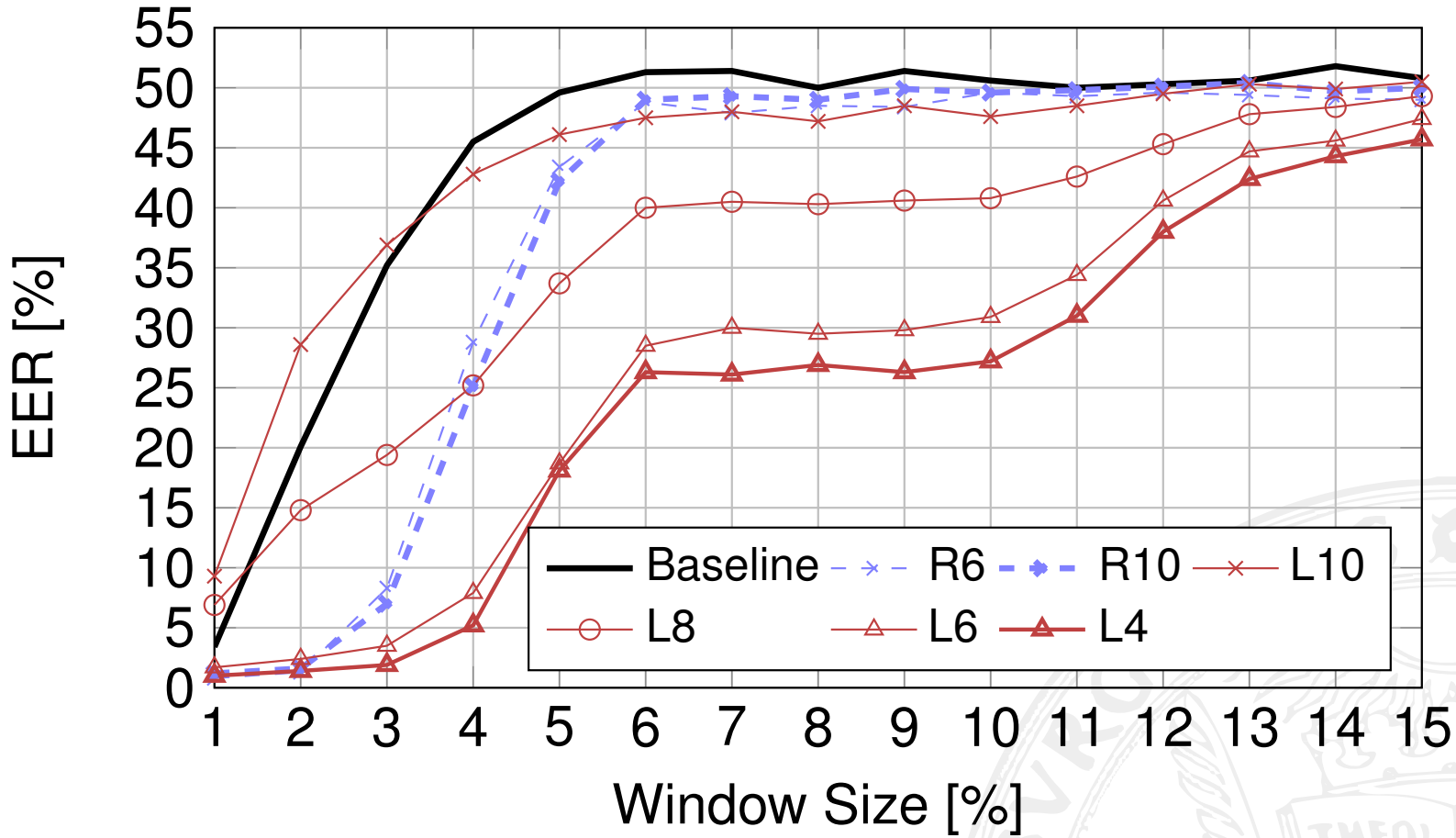


Figure: Finetuning MobileFaceNet Face Recognition.

Results: Encryption of Resolution Progressive JPEG2000

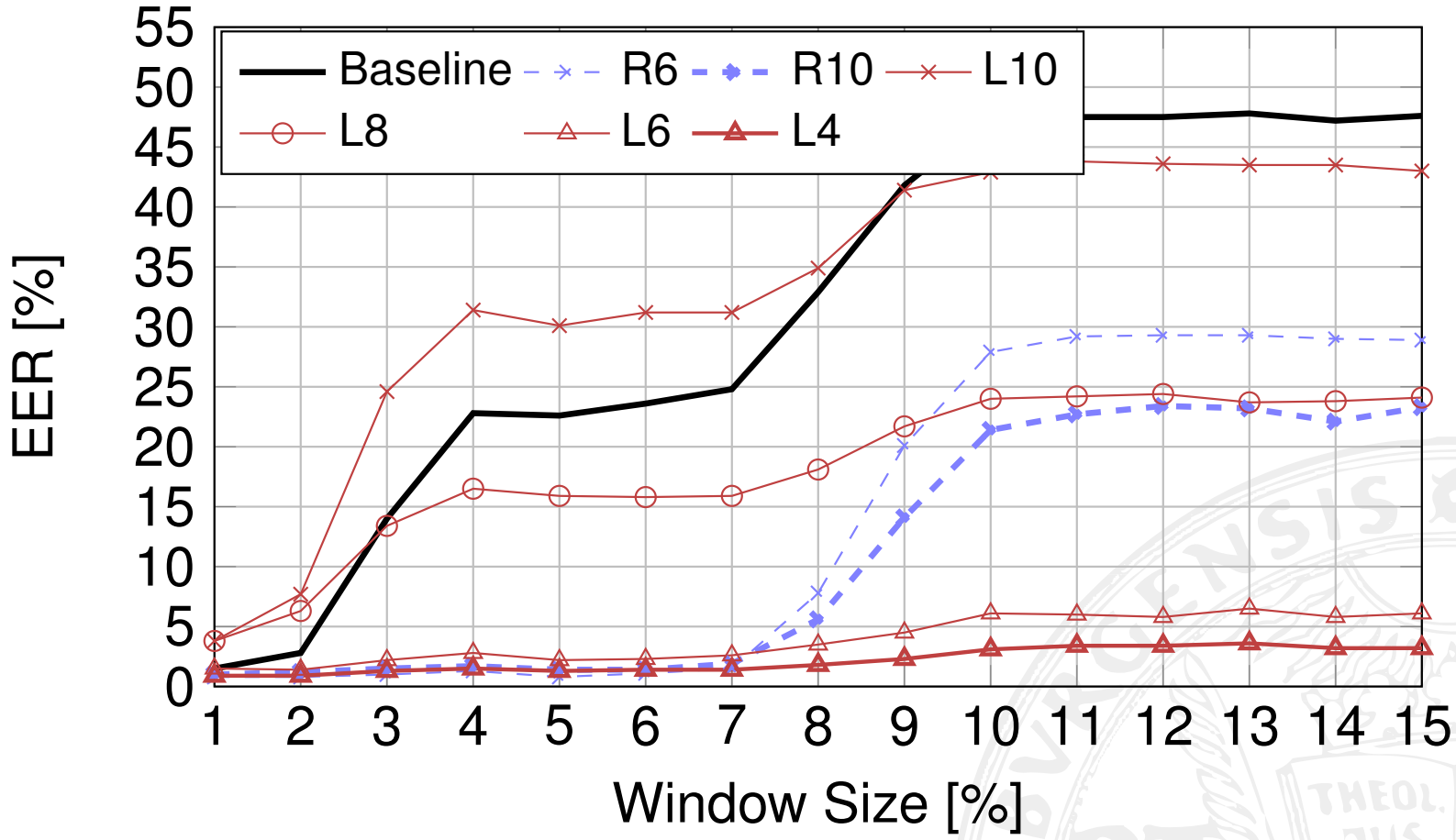


Figure: Finetuning MobileFaceNet Face Recognition.

Reducing Generalisability: Identical Encryption

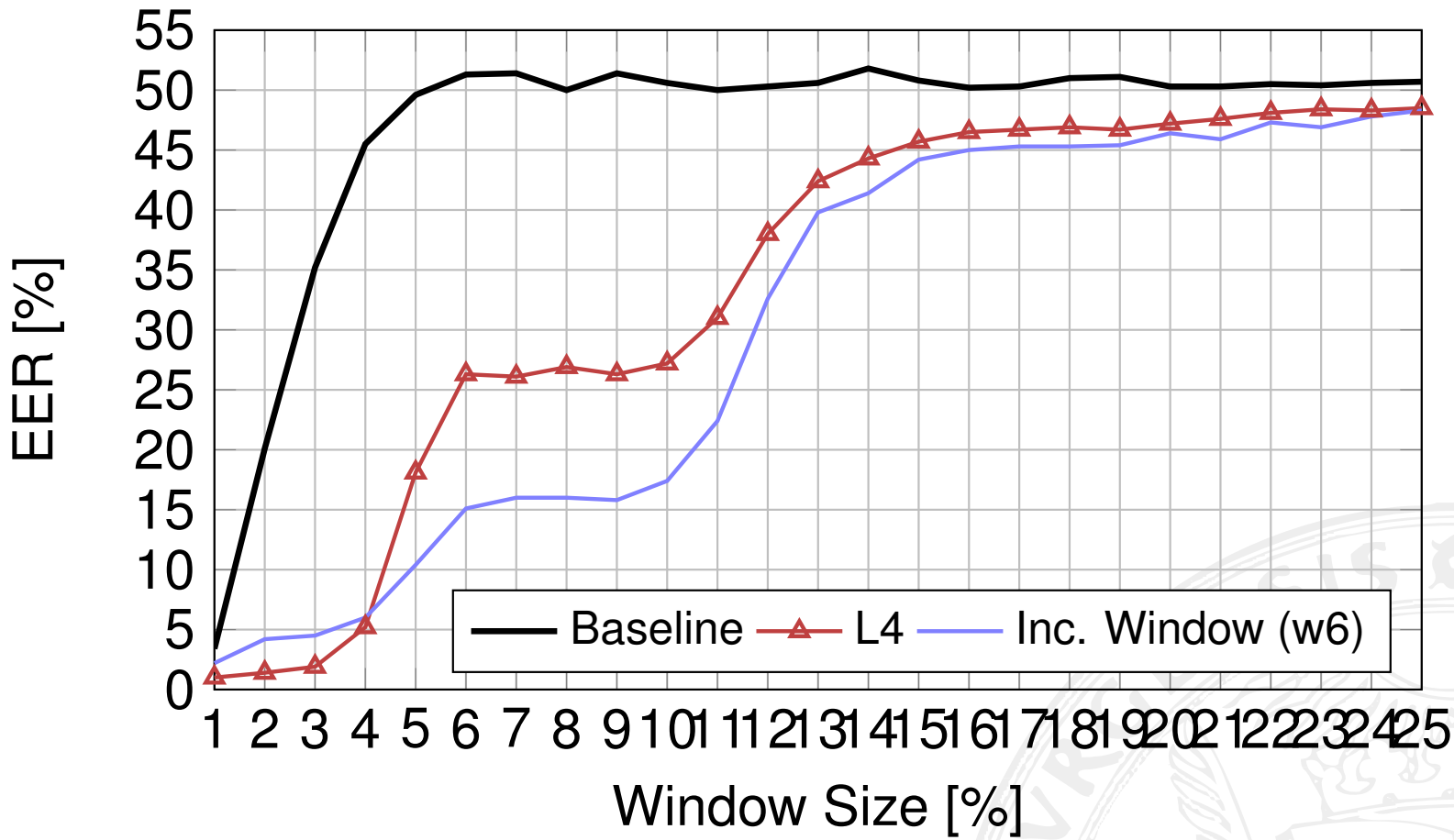


Figure: Refinement training with increasing window encrytion (layer progressive JPEG2000).

Maximum Generalisability: Different Coding Format (JPEG Encryption)

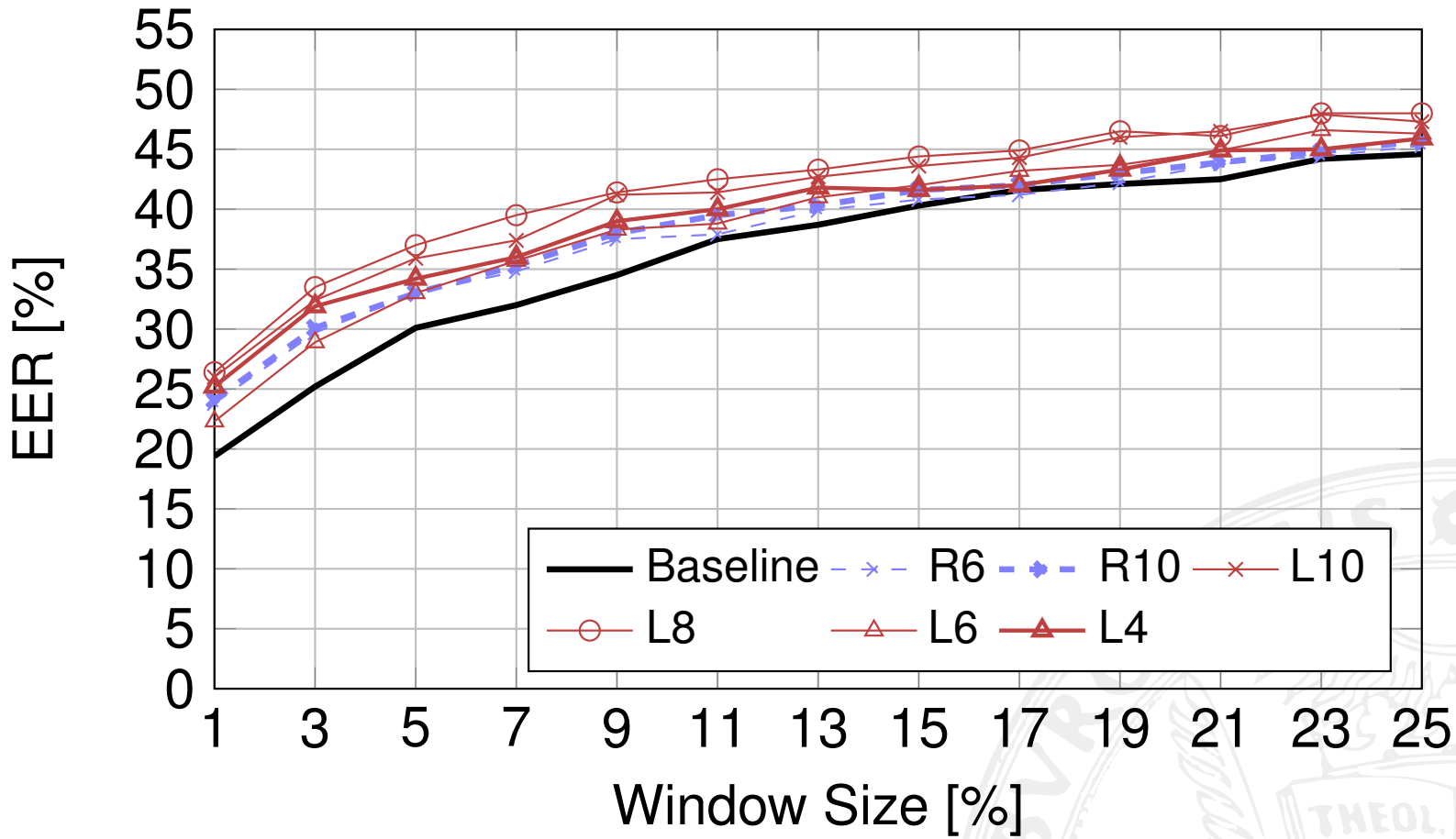


Figure: Evaluation on LFW with increasing window encryption applied to JPEG data.

Findings

- Fine-tuning of face recognition with encrypted data does in fact significantly weaken the protection strength of selective encryption.
- This works with a certain amount of generalisability, e.g. across different datasets and across different JPEG2000 progression orders.
- By specifically traing a specific encryption setting, results can be improved.
- A severe mismatch between encryption types cannot be tolerated, in this case (JPEG vs. JPEG2000) re-training does not work.

However, those findings can hardly be exploited in an actual attack !

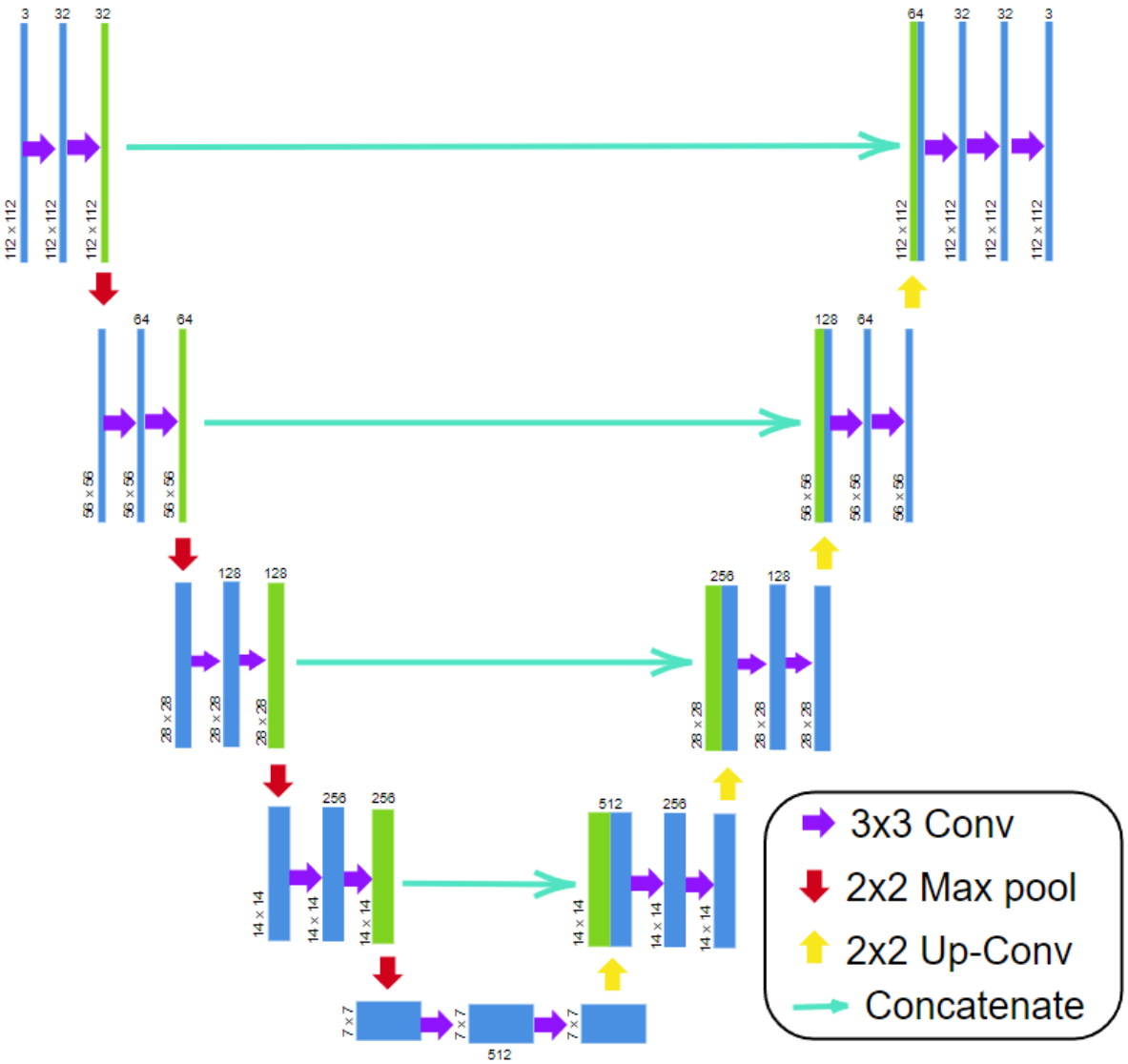
Realistic Attack against Selective JPEG2000 Encryption

Backgound

- We have seen that selectively encrypted image data contains informations, which can be exploited by a learning-based scheme.
- Artifacts consist of noise, but due to bitstream oriented encryption this noise can be coarse grained.
- When de-noising is applied (only), larger gaps remain in the images that have to be filled with reasonable content.
- This is the classical field of “image inpainting”. Many learning-based schemes do exist.

We employ a simple architecture to see if the appraoch works in principle, training with pairs of original and selectively encrypted portraits.

Inpainting Architecture



- Encoder-decoder network with skip connections between mirrored layers in the encoder and decoder stacks.
- UNet-“like” architecture: structure and the skip connections, but layers do not mimic exactly the real UNet (intended for segmentation).
- MSE as loss

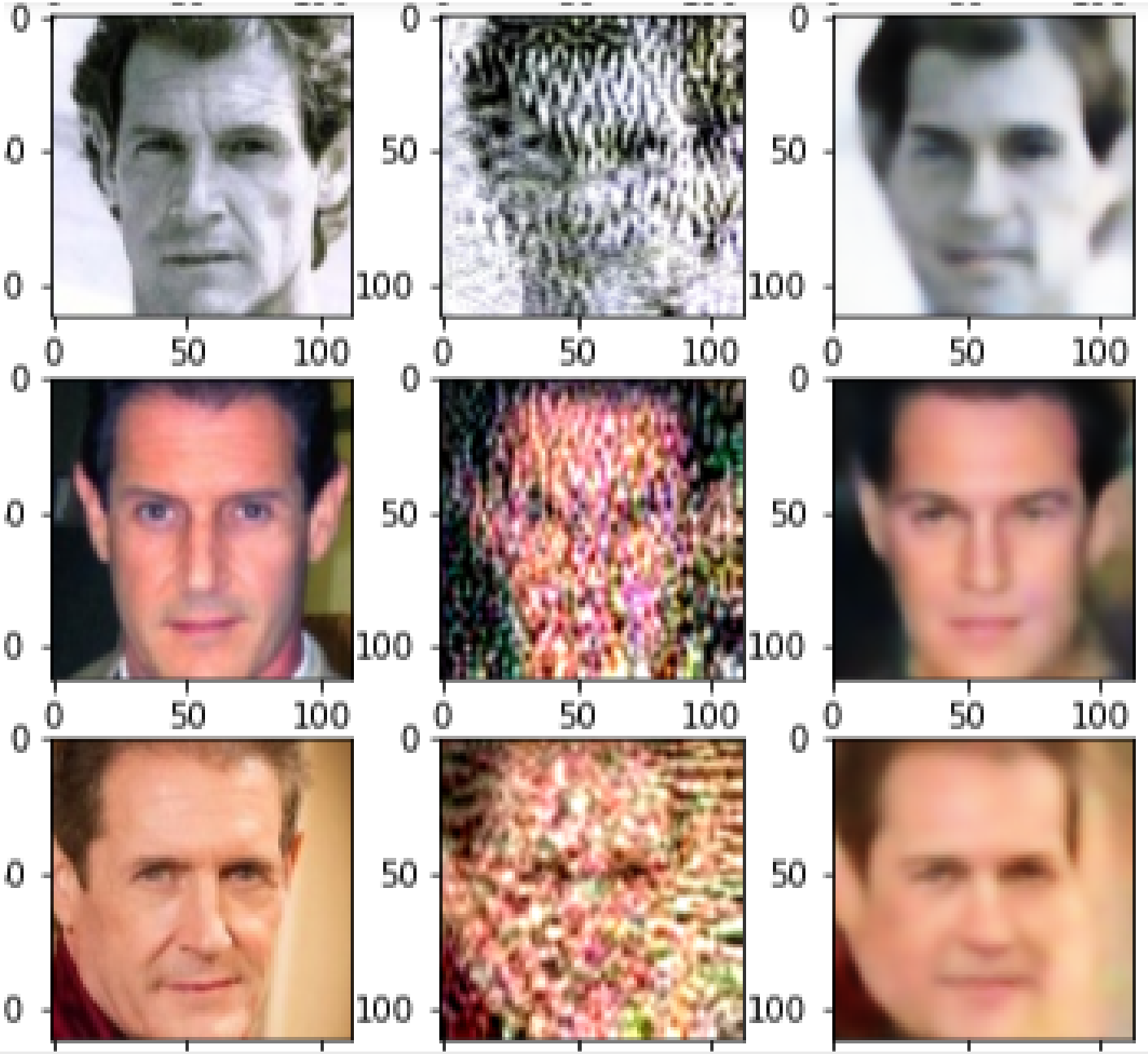
Experimental Settings

Data CASIA-WebFace database containing 494414 images of 10575 identities, with variations in pose, age, ethnicity and illumination. RetinaFace face detection is used to detect face landmark points, which are used to align, normalize, and crop each face to a size of 112×112. As our used reference dataset, we use 106 IDs from the database, in total 3963 images in plaintext. Three sliding window encryption variants in layer progression mode, i.e. lw4o6, lw6o6, and lw8o4, one in resolution progressive mode, i.e. rw4o6. A train/test (75/25%) split was created by partitioning subjects into 4 folds.

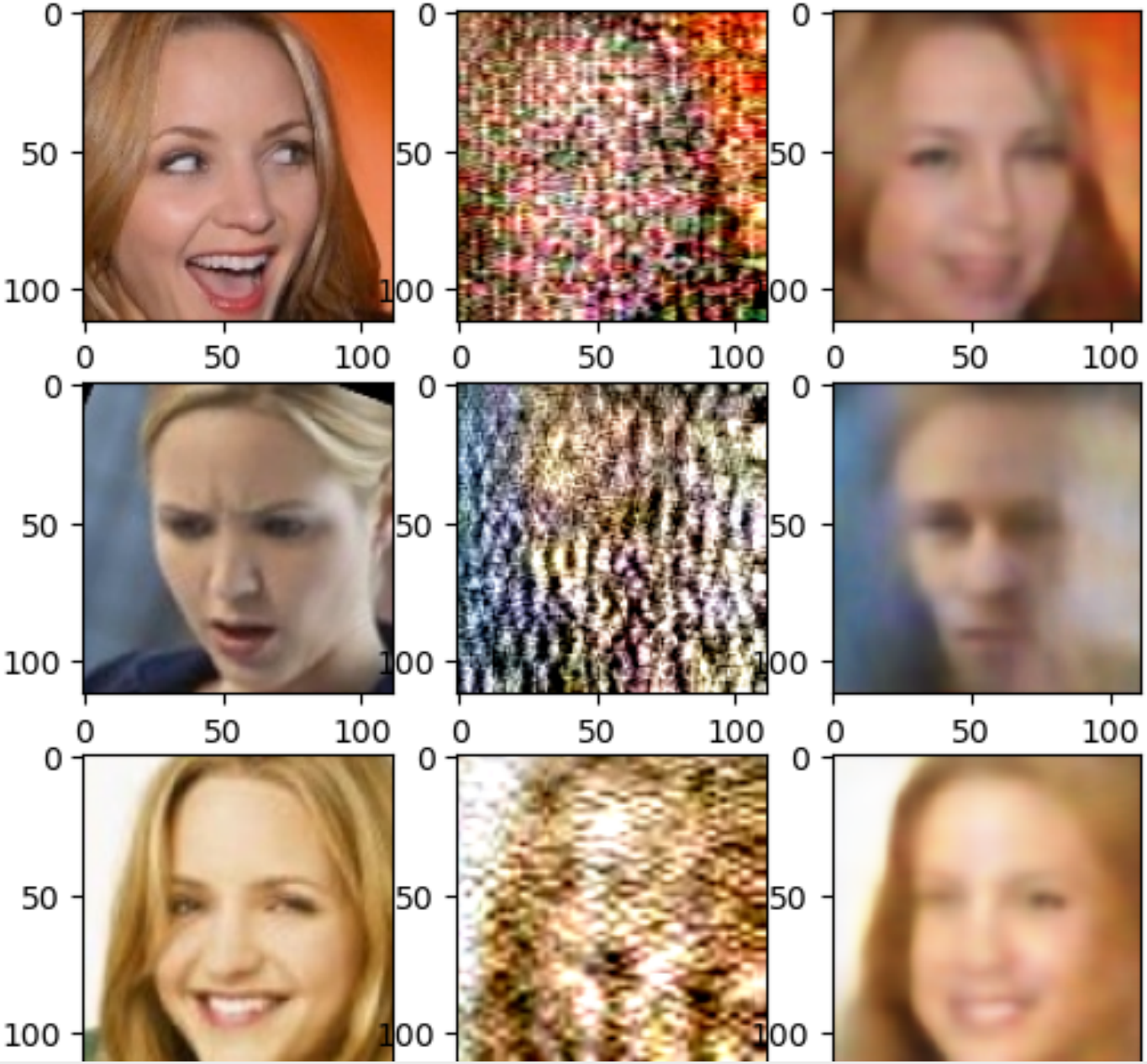
Evaluation

1. Qualitative: Human visual inspection.
2. Quantitative A: FaceQnet as No-Reference, end-to-end Quality Assessment system for face recognition.
3. Quantitative B: ArcFace biometric face recognition (ERR) - 7.16% ERR on reference dataset with plaintext probe and gallery, respectively.

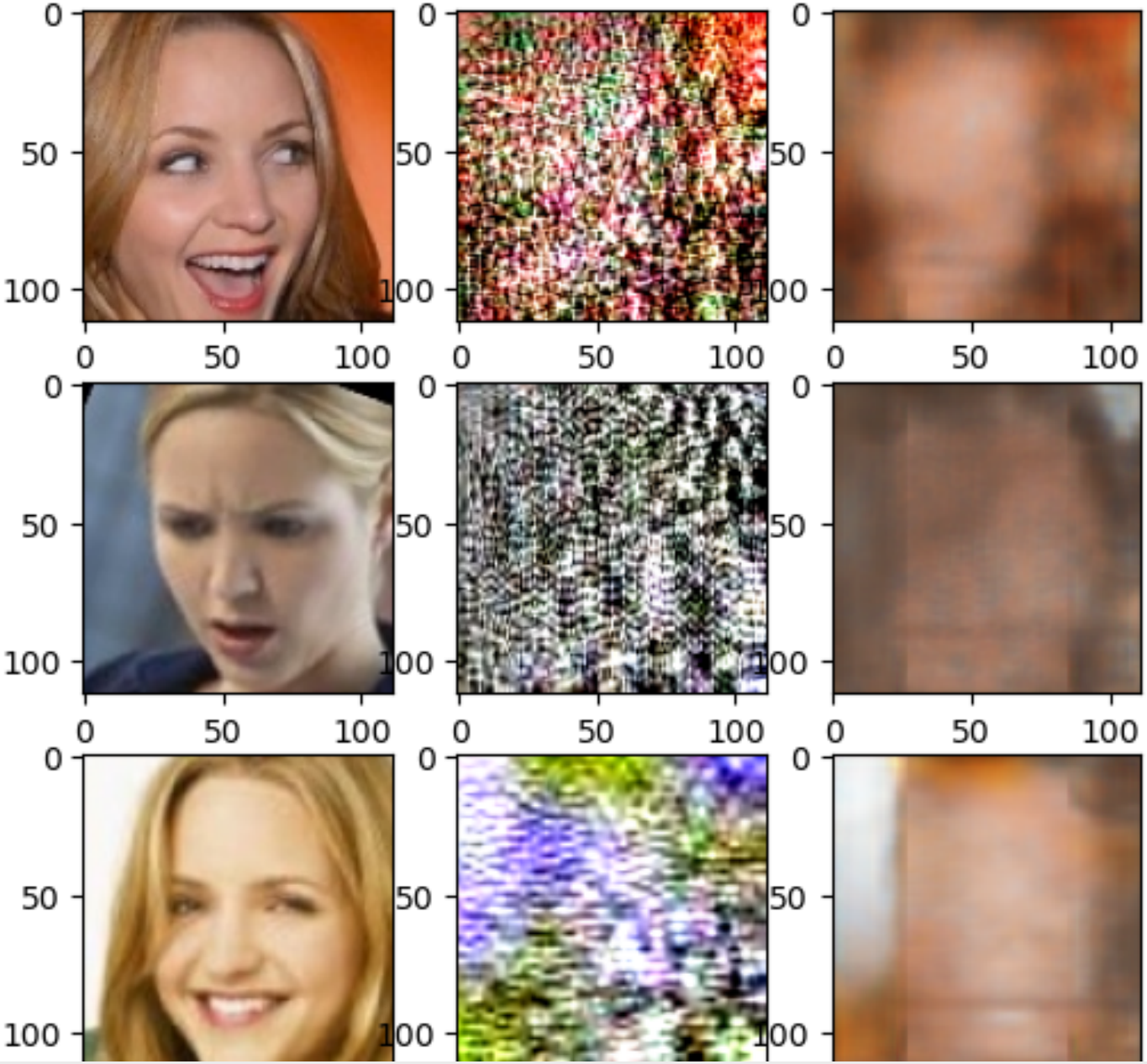
Examples: lw4o6



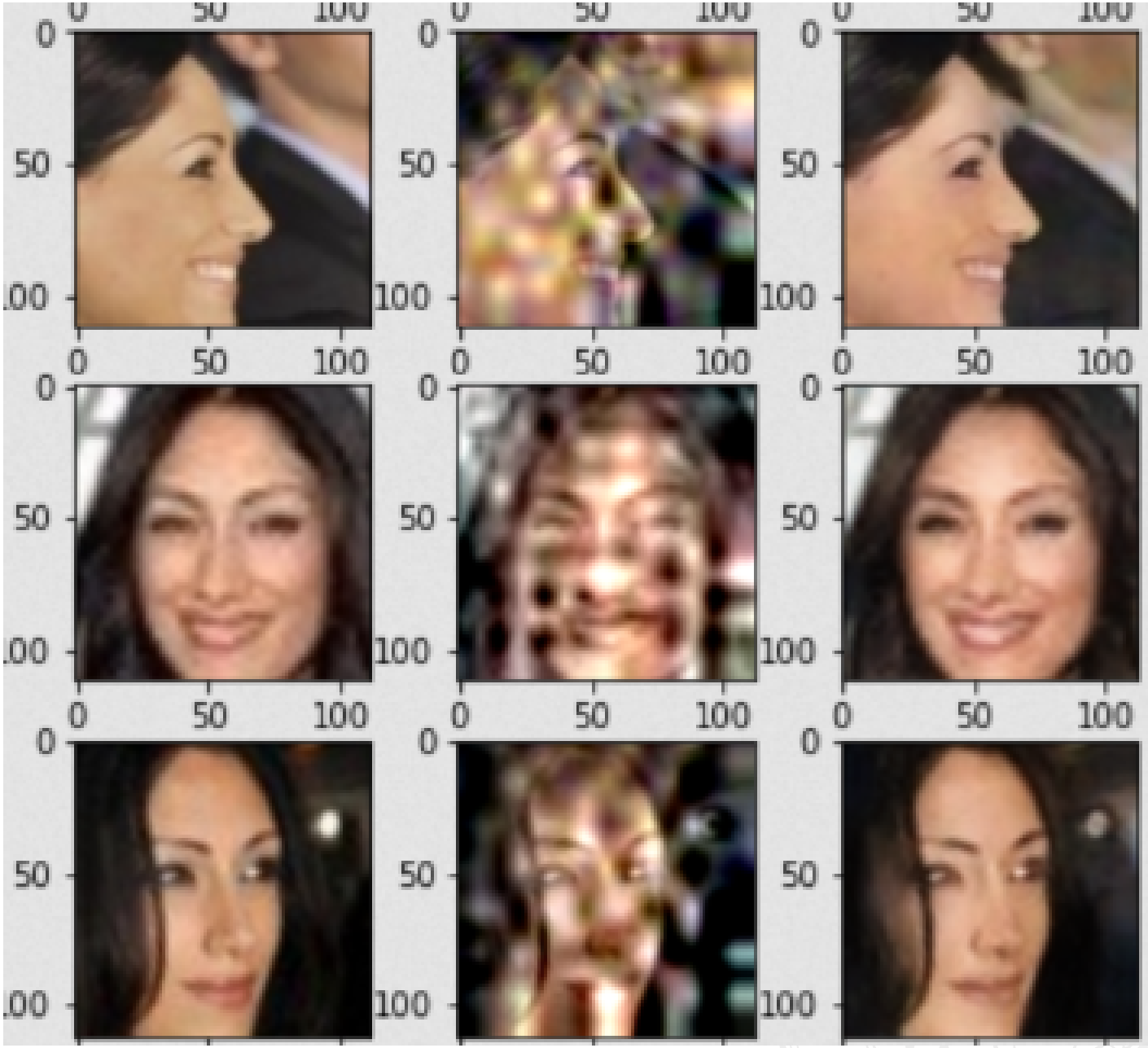
Examples: lw6o6



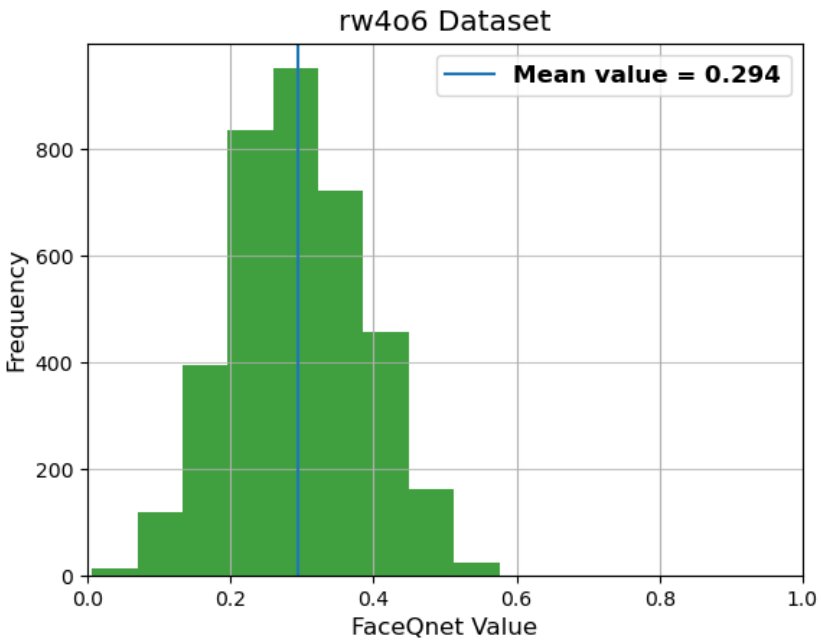
Examples: lw8o4



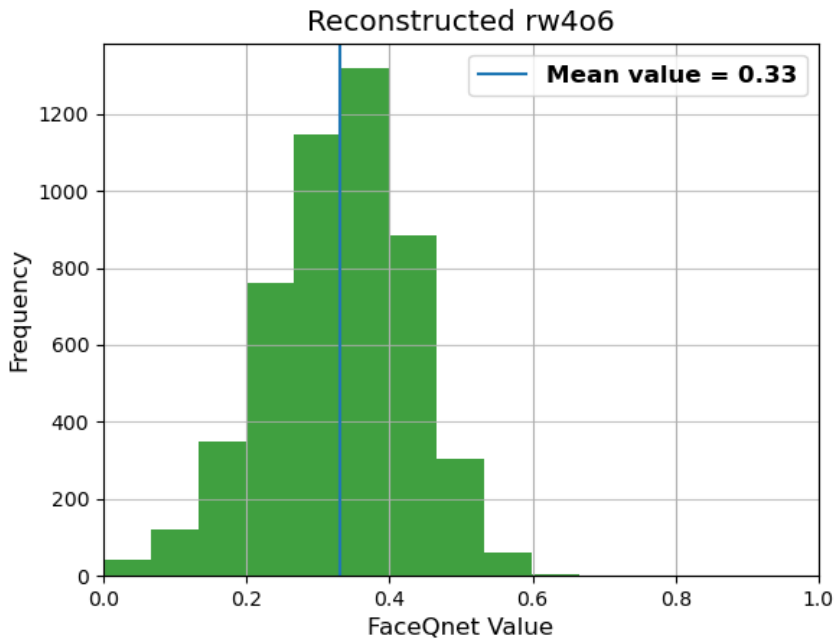
Results: rw4o6



Facial Image Quality



(a) protected data



(b) after inpainting

Figure: Distribution of FaceQNet values before and after inpainting (rw4o6 encryption).

ArcFace Recognition

	original	l4o6	l6o6	l8o4	r4o6
Protected	7.16%	47.55%	46.83%	48.49%	33.52%
Inpainted	7.16%	33.57%	31.93%	45.73%	11.09%

Table: Arcface recognition accuracy (EER in %) applied to protected (line 1) and inpainted (line 2) samples, respectively (where recognition on unencrypted plaintext originals results in an EER of 7.16%).

$\Longrightarrow$ The weaker the protection is, the better inpainting is able to improve recognition results.

Conclusions

Observations

- Due to the higher robustness of learning-based face recognition schemes, security mechanisms developed and proposed in the context of traditional face recognition need to be revisited.
- Block-based warping template protection as well as selective sample encryption have shown to be vulnerable under learning-based face recognition.
- Not at all limited to the biometric field - recommendations how to apply selective encryption schemes securely need to be revised.
- Learning based attacks (i.e. re-training using encrypted samples and inpainting) have shown to severely impact protection strenth of selective encryption.

Future work $\Longrightarrow$ better generalisability of the inpainting learning process (i.e. training with different encrypted schemes and data);
$\Longrightarrow$ usage of more advanced inpainting architectures.

Thank you for your attention!

Questions?