Utilizing CNNs for Cryptanalysis of Selective Biometric Face Sample Encryption

Heinz Hofbauer[1] • Yoanna Martínez-Díaz[2] • Luis Santiago Luevano[3] • Heydi Méndez-Vázquez[2] • Andreas Uhl[1]

[1]Paris Lodron University of Salzburg, Austria ({hofbauer, uhl}@cs.sbg.ac.at)
[2]Advanced Technologies Application Center (CENATAV), Cuba ({ymartinez,hmendez}@cenatav.co.cu)
[3]Tecnologico de Monterrey, School of Engineering and Science, Mexico (luis.s.luevano@tec.mx)

September 13, 2022

Abstract

When storing face biometric samples in accordance with ISO/IEC 19794 as JPEG2000 encoded images, it is necessary to encrypt them for the sake of users' privacy. Literature suggests selective encryption of JPEG2000 images as fast and efficient method for encryption, the trade-off is that some information is left in plaintext. This could be used by an attacker, in case the encrypted biometric samples are leaked. In this work, we will attempt to utilize a convolutional neural network to perform cryptanalysis of the encryption scheme. That is, we want to assess if there is any information left in plaintext in the selectively encrypted face images which can be used to identify the person. The chosen approach is to train CNNs for biometric face recognition not only with plaintext face samples but additionally conduct a refinement training with partially encrypted data. If this system can successfully utilize encrypted face samples for biometric matching, we can show that the information left in encrypted biometric face samples is information actually usable for biometric recognition.

The method works and we can show that a supposedly secure biometric sample still contains identifying information on average over the whole database.

Contents

1 Introduction

Selective encryption refers to selecting only a part some data to be encrypted with state of the art ciphers (e.g. AES). The motivation is to speed up encryption and decryption by reducing the encryption amount. If it can be ascertained that the selective encryption prevents abuse of the data, an overall speedup was gained.

A recent paper [1] uses selective encryption to protect face biometric samples. While this not new by itself, as selective encryption was proposed with different encryption methods and for different biometric modalities before [2, 3, 4, 5], it is interesting insofar as that face recognition uses CNNs for biometric recognition. One takeaway from that paper is that the recognition performance of CNNs is so strong, that techniques considered safe (in terms of encryption) for traditional biometric recognition methods do not hold up when used in systems utilizing CNNs.

In [1] the authors analyse the performance of selective encryption of face biometric samples by attempting to conduct the recognition process on ciphertext data with traditional and CNN based methods. The findings are such that information contained in the plaintext which was hitherto, i.e. for recognition with traditional methods, considered safe is now insecure due to the higher recognition performance of a CNN based method. This is likely the case also for other selective encryption methods and other application fields when recognition is moved to CNNs. After these findings, the authors evaluated the threshold for a save amount of encryption based on the improved, i.e., CNN based, recognition scheme, concluding that information relevant for faces is condensed in the first 12% of a layer progressive JPEG2000 encoded bitstream.

However, this topic was not thoroughly investigated, specifically, the authors used a classical CNN trained on an unencrypted face database and then tested for authentication by comparing encrypted and unencrypted face images (this is possible due to format compliance of the encryption scheme, see Section 4). Thus, one might argue that using ciphertext directly in a classical face recognition scheme is not the strongest cryptanalysis thinkable. In this work, we apply a potentially stronger attack for cryptanalysis by enhancing the training of the face recognition CNN - refinement training is done using partially encrypted samples, i.e. CNN is trained for recognition based on protected samples.

The rest of the paper is structured as follows: Section 2 reviews the state-of-the-art in cryptoanalysis of encryption schemes for visual data, while Section 3 explains the role of selective encryption in the context of a biometric system. Section 4 describes the methods, experimental setup, and databases used and how they are combined for the experiments. Section 5 presents and discusses the experimental evaluation. Finally, Section 6 succinctly recaps the findings and concludes the paper.

2 Security of Encryption for Visual Data

The assessment of encryption schemes for visual data, using selective encryption or weaker encryption approaches like chaos-based encryption, is known to be difficult. On the one hand, traditional image quality metrics are not useful to determine visual security [6], and the establishment of proper metrics is difficult [7, 8]. On the other hand, the methodology of cryptanalysis being applied turns out to be faulty in many cases [9]. For selective encryption, specific attack techniques do exist, which exploit information from the plaintext part to attack the ciphertext part (i.e. replacement [10] and reconstruction attack [11], respectively). A specific reconstruction attack applies error-concealment functionality of compression formats, developed for storage or transmission errors, to encryption artefacts, thus termed error-concealment attack (see [12] for an example applied to J2K selective encryption). Another reason to not use image quality metrics or similar methods for assessment of residual information is that we have an objective way of doing so. Specifically, we want to assess the security in regards to biometric recognition, as such we can (and should!) use the biometric recognition error as an inverse measure of residual information, i.e., the lower the recognition error the more pertinent information remains.

Caused by the recent advancement in machine learning, a new application field for image encryption has emerged, and new techniques for cryptanalysis have been developed. In distributed machine learning, privacy preserving CNNs have been introduced for various applications including video/image classification [13], in the context of which also new learning-based encryption techniques have been proposed [14]. Many of those techniques have been found to be insecure, partially using new attack types [15, 14, 16, 17], like GAN or Inverse Transformation-based attacks, respectively [14]. Other attacks look into the intermediate CNN representations for cryptanalysis [13], which is a related problem to reconstructing face images from deep templates [18]. However, deep learning-based attacks have also been successfully applied to classical chaos-based encryption [19, 20].

In this work, we consider a distinct pathway for cryptanalysis. Contrasting to the papers referenced, we do not aim to break the images' encryption, but we aim to train the face recognition network to also cope with selectively encrypted face data, thereby revealing the presence of relevant plaintext parts, and thus conducting successful cryptanalysis. The experiments in [1], as discussed in the Introduction, already show a baffling ability of the CNN to extract faces from noisy (i.e. encrypted) information. However, the experiments do not take into account that a CNN can also be trained on noisy data, very likely improving its performance. To train the CNN on noisy data suggests itself whenever a potential noisy image, e.g., outdoor access system with low light, might otherwise impact recognition performance. The conclusion about the required encryption amount (stated as 12% in [1]) was based on the idea that this is where the recognition performance was impacted by the encryption to such a degree that the biometric recognition results were random (i.e. the contained information could not get better results than guessing the identity). However, this analysis only confirms that the protected biometric face samples can not be successfully directly presented to the system. The claim that all the relevant data is contained in the first 12% is not checked beyond that assumption.

Here, 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. This means that we assume that it is known at training time of the system that such protected face data might be presented to the system. Of course, this is not typical for an actual system in use. However, if information which can lead to a correct biometric recognition is still contained in the encrypted images, then it is conceivable that an attacker could extract it as well, e.g. using a GAN-based attack. While it is difficult to anticipate all possible attacks, we can utilize the learning capabilities of CNNs to see, given encrypted data available during the training process, whether a correct biometric recognition is possible.

Thus our approach is to train CNNs to recognise encrypted facial images. If there is no information left in the encrypted image the matching performance should be similar to random guessing, i.e., 50% equal-error rate (EER). However, if the EER deviates significantly from this rate then information related to the biometric trait is retained in the encrypted face samples. In such a case we have a proof of falsification to the claim in [1].

We would like make clear that this is of course not a practical attack, it merely shows that information is left in the plaintext which an attacker could potentially exploit by any of the attacks discussed.

In summary our contribution is to show the use of CNNs as application-oriented cryptanalysis tool, which to our knowledge was not done before. We use this tool to assess the amount of information retained in the plaintext portion of the method introduced in [1]. This is of course equally applicable to other selective encryption methods, but to the best of our knowledge, the only work in the literature proposing selective encryption of face biometric data which takes into account CNNs is [1].

3 Selective Encryption and it's Place in the Biometric System

In this section, we briefly give an overview of the relevant literature, and standards, regarding the storage of biometric (face in our case) samples.

The International Organisation for Standardisation (ISO) specifies biometric data to be recorded and stored in image form (ISO/IEC 19794-5:2011 [21]), i.e., biometric samples, not only extracted templates. The standard allows two file formats for facial images: JPEG and JPEG2000. Several studies recommend JPEG2000 (J2K) due to better performance, [22, 23]. Storing the facial images during enrolment (in addition to the extracted templates in the biometric database) has the benefit of being future proof, i.e., extraction methods can be easily changed without the need for the costly and cumbersome re-enrollment. While this increases interoperability and vendor neutrality [24] on the positive side, it also generates a problem, if this database is compromised as the biometric traits of all enrolled users are leaked.

One method to circumvent this is to split storage of the extracted biometric templates (online in the system) from the facial images (offline database). In addition, both databases can be protected: biometric templates with template protection schemes, and the offline database either with traditional or selective encryption schemes (both using state of the art ciphers).

During the process of using the offline (biometric sample) database to regenerate the online (biometric templates) database the system can not be used, thus, this operation should be fast. Typically, the online database (with protected templates) is regenerated for three reasons: (i) changes in the biometric template extraction or matching, (ii) a known breach of the online database, (iii) a periodical change of the online database to cover the case of an unknown breach. The last two usually rely on a change in the extraction key of a given template protection scheme. In either case, the downtime should be minimized, which requires a fast decryption method for the offline database. This is usually the reason to utilize selective encryption.

A recent biometric sample protection method relying on J2K selective encryption, and the only one which considers Deep learning-based recognition for security assessment to the best of our knowledge, is suggested in [1]. There are a number of findings in that paper, but the pertinent one is that the layer progressive encoding of J2K compresses the information required for face biometric recognition (and resolution progression should not be used for security reasons). It is found that the relevant information for face biometric recognition is contained in the first 12% of a J2K compressed images with layer progression. Furthermore, results also show the ability of convolutional neural networks (CNNs) to conduct facial recognition despite strong (noise-like) interference caused by partial encryption.

The aim of this method is to prevent the use of the biometric face samples in case the offline database is leaked. In such a case the still usable information in the database (resulting from unencrypted J2K codestream parts) has to be assessed to ascertain that it does not allow for an attack on the biometric recognition system, i.e., to use the protected face samples to conduct successful authentication.

4 Methodology and Experimental Setup

Our goal is to determine the information left in a selectively encrypted biometric face sample. Engineering every potential attack is not possible so we utilize CNNs to learn to recognize encrypted face samples. In order to do this, we will keep the training and testing dataset and encryption as separate as possible, the methods used and how they are set up follows.

We will use methods from [1], so we can use the results from that paper to guide our training approach. We only use a single CNN based method from those evaluated since we need to retrain the models in our experiments and this can be time consuming. For the same reason, we chose the MobileFaceNet [25] over the best performing ResNet-ArcFAce [26]. For the layer progression encoding of J2K MobileFace has a comparable performance [1] and a high efficiency in terms of parameters, FLOPs, model size and computation time [27].

For training, we select CASIA-WebFace [28] (*CASIA*) which contains 494414 images of 10575 identities, with variations in pose, age, ethnicity and illumination. For preprocessing, RetinaFace detector [29] is used to detect face landmark points, which are used to align, normalize, and crop each face into a template of size 112×112.

These images were then encoded with lossless J2K, differing only in the progression type, of which there are two: layer and resolution progression. The actual progression used is given in the experiments.

For encryption we used the format compliant J2K encryption scheme introduced in [30]. Format compliance is important as it allows to decode the image by a standards compliant J2K decoder. The encryption introduces strong distortions into the image but the output can directly be used for biometric recognition. It should be noted that there is an inbuilt error correction in a J2K decoder which was used in all experiments and during training, which was shown to slightly improves results in [1].

Two encryption setups were used. The first is the increasing window encryption, where encryption starts at the beginning of the J2K bitstream and a window of a given size is encrypted. The other is a sliding window encryption where a fixed window (4% of the total bitstream) is encrypted but the offset varies. Both methods can be described with two parameters, the offset where encryption starts (o) and the window size (w), i.e., how much of the bitstream is encrypted. These parameters will be given in the experiments. Examples of different offsets for the window encryption on the CASIA database, which is used for training, can be seen in Figure 1. Except for one test, we used different encryption setups for training (sliding window) and for evaluation (increasing window) to increase generalisability.

We fine-tuned the pre-trained MobileFaceNet on differently encrypted CASIA datasets, with the settings that will be given in the next section in conjunction with the discussion of the experiments. As optimization solver we adopt stochastic gradient descent with the batch sizes of 128, 256 and 512. The learning rate was initialized to 0.1 and decreased by a factor of 10 periodically at 100 k, 140 k, and 160 K iterations. The final iteration step was 200 k. The momentum parameter was set to 0.9 with a weight decay of 5×10^{-4}. The parameter initialization for convolution is Xavier with random sampling from a Gaussian normal distribution and the loss function used is ArcFace [26] with an angular margin $m = 0.5$.

For testing we used a separate dataset, the Labeled Faces in the Wild (*LFW*) database [31], which is well known as a public benchmark for unconstrained face verification. It contains 13233 face images are from 5749 different identities, with large variations in pose, expression and illumination. The same preprocessing as for the training data, i.e., face detection with RetinaFace, cropping and normalization, was used.

For evaluation we use the increasing window encryption since we already know that the sliding window encryption is insecure [1]. Figure 2 shows samples of the encryption used for testing, a single image encrypted with various window sizes. It can be seen that the resolution encoding allows the remnants of low-frequency data, i.e., coarse object outlines, to be present much longer than for layer progression.

5 Experiments and Discussion

For encryptions we used the results from [1] to steer our examination. The assumption is that a similar strength encryption would have a similar impact on the results, therefore, we used the EER as reported in [1] to find encryptions of similar strength. We want to see if there is a difference of using the same strength of encryption on resolution vs. layer progression for learning. Further, we want to see whether there is a maximum encryption strength to use during training for a positive impact. That is, if is there a point where the encryption is so strong that nothing can be learned from it or that it even becomes detrimental to learning.

The basis for learning is the sliding window encryption with a window size of 4%. The progression mode, layer (L) or resolution (R), is specified in conjunction with the offset. For example, *L6* describes an image encoded with layer progression with a sliding window encryption with window size 4% and starting offset of 6%.

Figure 3 shows the performance of MobileFaceNet on the sliding window encryption of the LFW database as discussed in [1] without any specific refinement of the CNN. The encryption settings which will be used for training are marked and labeled. Figure 1 shows examples of the CASIA database, used during training, encoded and encrypted with the marked settings. The marks will be consistently used for the rest of the plots as well to make reference to Figure 3 easier. The parameters R6, R10 and L10 are of equal EER performance but different progression modes (L10 and R10) or of the same progression but different encryption settings (R6 and R10). For the encrypted layer progression, which spans a larger range of EER rates, we also chose two different encryption settings with similar EER performance (L4 and L6). Finally, we also included an intermediary between the two EER rates represented so far (L8).

5.1 Refinement with Fixed Window Size Encrypted Data

The results of the evaluation on the LFW database with increasing window size after refinement of the CNN, and the two progression modes, are presented in Figure 4. The color reflects the progression type during training, the same mark (e.g. △) represent similar EER performance (without refinement) of that encryption type in Figure 3. The bold black line is the baseline recognition performance of the system without refinement. Note that 50% EER is guessing, this is the goal for a security evaluation. The baseline of the resolution progression encoded J2K does not reach 50% and should not be considered secure, it is only included to see if there are differences between layer and resolution progression when it comes to refinement learning. While the absolute values are different of course, the overall behaviour, as discussed in the following, for layer progression also holds for resolution progression. Apart from not being secure the two progression modes otherwise behave similarly.

The first thing to note is that refinement generally improves the biometric recognition, meaning there is a more information still contained in the encrypted face samples than it originally appeared in [1]. This also means that the use of CNNs for cryptanalysis has merit.

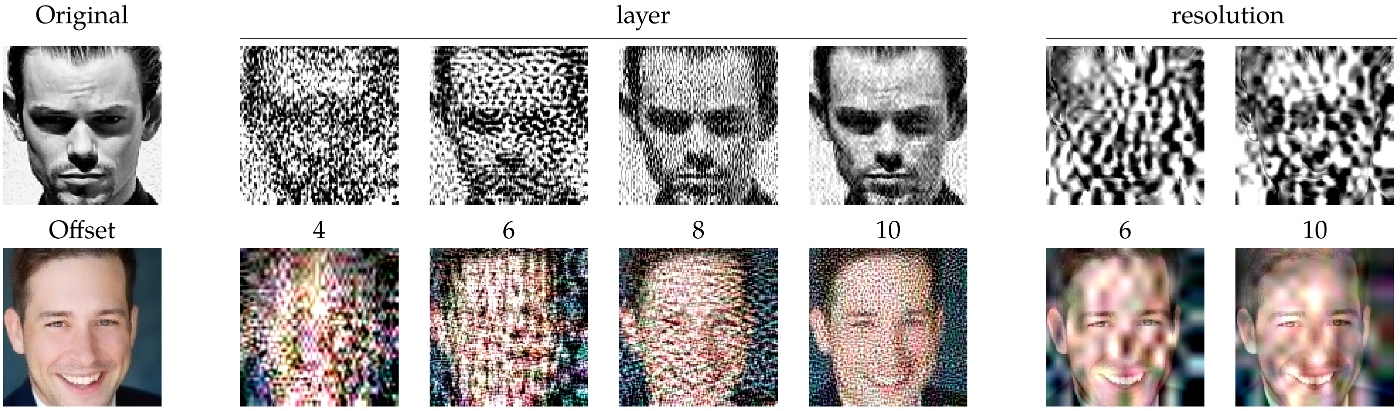
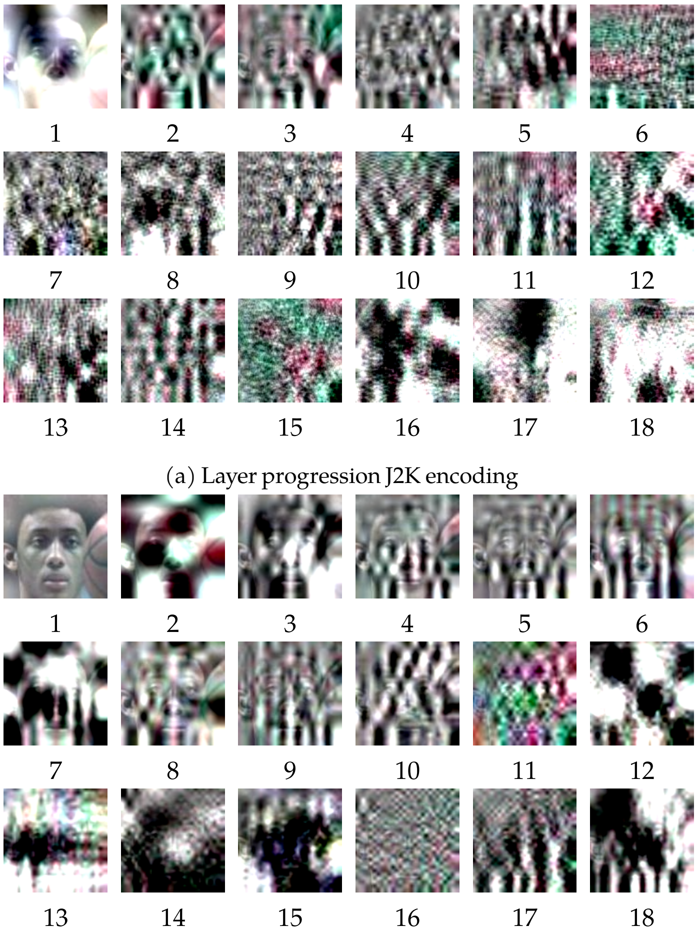


Figure 1: Samples of the encryption on the CASIA database which was used for training. The J2K progression mode is given as well as the encryption offset for the sliding window encryption (window size is 4%).



(a) Layer progression J2K encoding

(b) Resolution progression J2K encoding

Figure 2: Encryption examples from the faces in the wild database with the given window size (increasing window) and the given progression for J2K encoding.

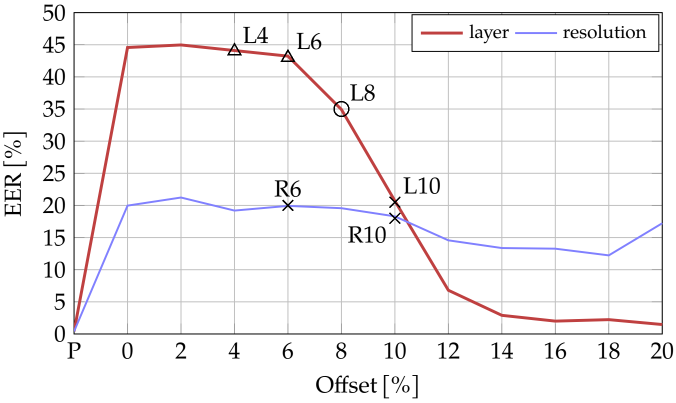
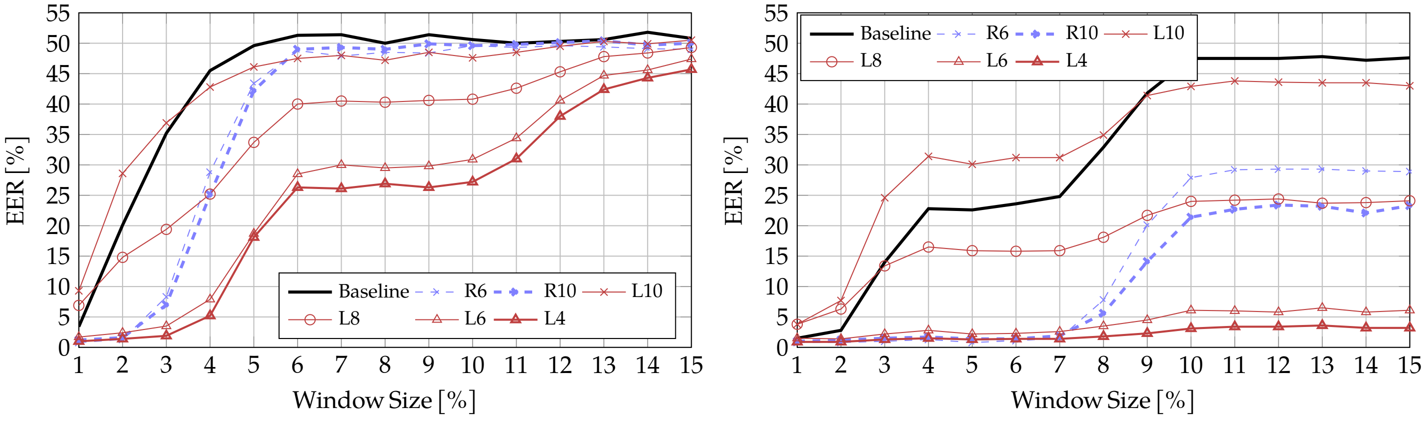


Figure 3: Results of the sliding window encryption (window size 4%) and the given offset (P is for original) without refinement.

Similar EER performance of an encryption type, for the same progression mode, show a similar performance on the increased window encryption test set, these are the R6/R10 and L4/L6 pairs. However, there is a distinct difference when it comes to different progression types: R10 and L10 had a similar EER for the sliding window encryption, see Figure 3. However, when used for refinement and evaluated on the increasing window encryption the behaviour is very different, Figure 4. Not only is the resulting performance different, refinement with L10 (and at the start also with L8) actually decreases performance. This is due to the impact of the encryption on the encoding. The farther away from the start of the bitstream the higher the encoded frequencies become. This is only a rule of thumb and not totally true for layer progression where amplitude as well as frequency of the signal determine the ordering, but it is mostly true, see Figure 1. The increasing window encryption starts at the beginning, so encrypts primarily low frequency content, see Figure 2. So, the CNNs are trained to deal with high frequency noise and are presented low frequency noise, which leads to the results in Figure 4. Only when the increasing window also produces higher frequency noise does the refinement of the CNNs kick in and improve the results. This is reflected in Figure 4 by the better performance of those networks when the encryption becomes stronger.

The same argument applies to the differences of the R6/R10 and L4/L6 pairs, where the lower offset always performs better. This is again due to the CNNs being trained on lower frequencies. The better performance of CNNs refined on layer progressive encoding has a similar reason. A finding in [1] was that the layer progressive encoding compresses the information which is relevant to the biometric comparison. This means that a 4% sliding



(a) Layer progressive encoding

(b) Resolution progressive encoding

Figure 4: Results for the evaluation on LFW with increasing window encryption, window size as given.

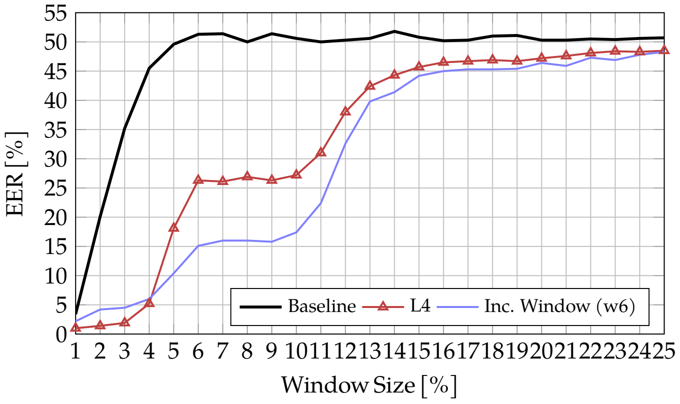


Figure 5: Evaluation on LFW with increasing window encryption for layer progressive encoding.

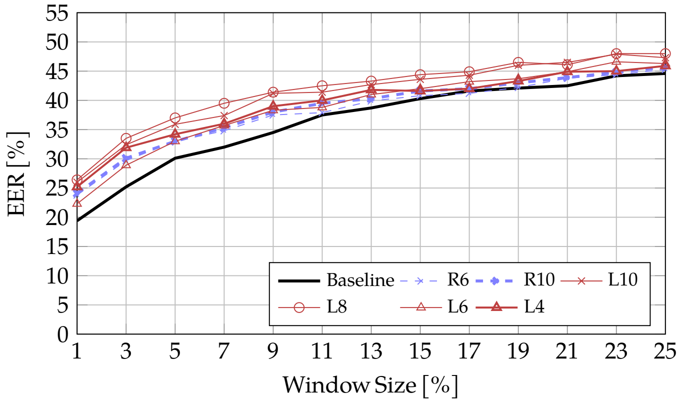


Figure 6: Evaluation on LFW with increasing window encryption for JPEG encryption.

window encrypts more information for layer than for resolution progressive J2K coding. In other words, CNNs trained on layer progression and encrypted images are trained to perform matching with less information than those trained on the resolution progression.

5.2 Refinement with Increasing Window Size Encrypted Data

As a final experiment about J2K encryption, we have used the same encryption for training that is used for testing, an increasing window encryption with a window size of 6, which is the first EER plateau reached by the evaluation of the prior tests, see Figure 4. The results of the evaluation, together with the baseline and the best performing result from Figure4, is shown in Figure 5. We have also extended the maximum size of the window during testing to 25% of the bitstream, which would be a speedup of 4 over a full encryption. We can see that overall training on the increased window encryption improves the results but performance is slightly worse for smaller window sizes. This is likely because the CNN does not learn to handle lower encryption amounts. The next step to improve the results would be to combine different encryption strengths during training. However, this is somewhat moot as we can see that even at a window size of 25% we have not reached the 50% EER. But the figure makes it clear that, should we further increase the window size, both retrained CNNs would converge on the 50% EER.

5.3 JPEG Encryption and Cross Encryption Refinement

While JPEG is not recommended for storage of face images for biometric comparison [22, 23], we can use it to see if the refined (on J2K) CNNs also improve on different encryption artefacts. The results are found in Figure 6, where the EER is given for the unrefined CNN (as baseline), as well as those of the refined CNNs. We won't go into details about the JPEG encryption [32] but apart from the change in encryption we followed the same workflow as with J2K increasing window encryption.

Two facts are readily apparent: (1) JPEG encryption, even with a 25% window, is still far from secure (the 50% EER range). So the choice to use J2K for storage of face biometric samples is also good in light of securing those samples. And (2), retraining does not improve the recognition in case of JPEG encryption artefacts, the contrary is true. So cross encryption refinement, i.e., transfer learning, does not work.

Neither of these is terribly surprising, the artefacts are totally different from J2K artefacts, compare Figures 2 and 7 for

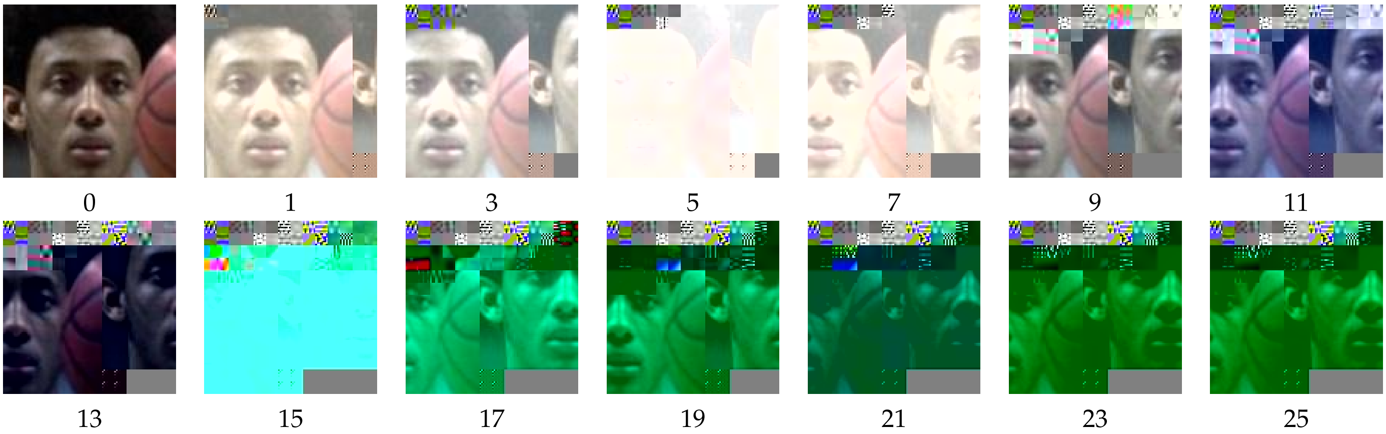


Figure 7: JPEG encryption examples from the faces in the wild database with the given window size (increasing window).

encrypted J2K and JPEG images respectively. So learning one does not help with the other. Further the coding of JPEG is in spatial ordering, so encryption of any percentage is leaves the rest very recognizable. In contrast JPEG2000 orders its data by impact on the whole image, compressing data relevant to the whole image into the earlier parts of the bitstream.

What is more, decoding problems lead to skipped blocks for the JPEG encrypted images resulting in a circular image shift. This splits the face, even though the results are recognizable to a degree, it should be clear that an attacker which reverses this shift could easily produce an equal error rate which deviates farther from 50%.

6 Conclusion

We have shown, by using encrypted data during training, that a CNN can learn to correctly match encrypted face samples. This means that information pertaining to the identity of a protected biometric sample is leaked despite the selective encryption. As stated this is a proof by falsification of the claim from [1] that the information required for biometric face recognition is contained in the first 12% of a layer progressively encoded JPEG2000 bitstream.

We have also shown that JPEG, the other standards approved storage variant for face biometric samples [21], is less secure than JPEG2000.

The observed leak of biometric information can be alleviated by increasing the encryption amount of the bitstream, but the trade-off between performance and risk becomes troublesome to a point where it is likely better to use a traditional encryption method. However, this is a theoretical result, in that usually the network used for the biometric recognition will not be trained on the encrypted data itself. It is unclear how this data can be extracted in practice, and how an attack can be mounted with it. We have shown however that the relevant data is present and therefore a potential security leak.

We have also shown that when using CNNs as we did here we can utilize them to asses the usefulness of the remnants of information left in the plaintext part of a selectively encrypted image. In this way a CNN can be used to cryptanalyse a selective encryption scheme. While it is not an actual attack, it certainly showcases whether or not information left to an attacker to use is remaining in plaintext.

Acknowledgment

This project received funding from EU Horizon 2020 research program, grant agreement No. 690907, and from the Austrian Science Fund, project No. P27776.

References

[1] H. Hofbauer, Y. Martínez-Díaz, S. Kirchgasser, H. Méndez-Vázquez, and A. Uhl, “Highly efficient protection of biometric face samples with selective jpeg2000 encryption,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP ’21*, 2021 (cit. on pp. 2–5, 7).

[2] M. Rieger, J. Hämmerle-Uhl, and A. Uhl, “Efficient iris sample data protection using selective jpeg2000 encryption of normalised texture,” in *Proceedings of the 6th International Workshop on Biometrics and Forensics (IWBF’18)*, 2018 (cit. on p. 2).

[3] M. Rieger, J. Hämmerle-Uhl, and A. Uhl, “Selective JPEG2000 Encryption of Iris Data: Protecting Sample Data vs. Normalised Texture,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’19)*, 2019 (cit. on p. 2).

[4] M. Draschl, J. Hämmerle-Uhl, and A. Uhl, “Sensor dependency in efficient fingerprint image protection using selective jpeg2000 encryption,” in *Proceedings of the 5th International Workshop on Biometrics and Forensics (IWBF’17)*, 2017 (cit. on p. 2).

[5] S. Shekhawat, H. Hofbauer, B. Prommegger, and A. Uhl, “Efficient fingervein sample image encryption,” in *Proceedings of the 8th International Workshop on Biometrics and Forensics (IWBF’20)*, Shortlisted for best paper award, 2020 (cit. on p. 2).

[6] H. Hofbauer and A. Uhl, “Identifying deficits of visual security metrics for images,” *Signal Processing: Image Communication*, vol. 46, 2016, ISSN: 0923-5965 (cit. on p. 2).

[7] H. Hofbauer, F. Autrusseau, and A. Uhl, “To recognize or not to recognize — a database of encrypted images with subjective recognition ground truth,” *Information Sciences*, no. 551, 2020 (cit. on p. 2).

[8] S. Guo, T. Xiang, X. Li, and Y. Yang, "Peid: A perceptually encrypted image database for visual security evaluation," *IEEE Transactions on Information Forensics and Security*, vol. 15, 2020 (cit. on p. 2).

[9] M. Preishuber, T. Hütter, S. Katzenbeisser, and A. Uhl, "Depreciating motivation and empirical security analysis of chaos-based image and video encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, 2018 (cit. on p. 2).

[10] Mark M. Fisch, H. Stögner, and A. Uhl, "Layered encryption techniques for DCT-coded visual data," in *Proceedings (CD-ROM) of the European Signal Processing Conference, EUSIPCO '04*, paper cr1361, 2004 (cit. on p. 2).

[11] M. Podesser, H.-P. Schmidt, and A. Uhl, "Selective bitplane encryption for secure transmission of image data in mobile environments," in *CD-ROM Proceedings of the 5th IEEE Nordic Signal Processing Symposium (NORSIG 2002)*, file cr1037.pdf, 2002 (cit. on p. 2).

[12] T. Stütz and A. Uhl, "On JPEG2000 error concealment attacks," in *Advantages in Image and Video Technology: Proceedings of the 3rd Pacific-Rim Symposium on Image and Video Technology, PSIVT '09*, ser. Lecture Notes in Computer Science, 2009 (cit. on p. 2).

[13] H. Benkraouda and K. Nahrstedt, "Image reconstruction attacks on distributed machine learning models," in *Proceedings of the 2nd ACM International Workshop on Distributed Machine Learning*, 2021, ISBN: 9781450391344 (cit. on p. 2).

[14] W. Sirichotedumrong and H. Kiya, "Visual security evaluation of learnable image encryption methods against ciphertext-only attacks," in *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2020 (cit. on p. 2).

[15] H. Oh and Y. Lee, "Exploring image reconstruction attack in deep learning computation offloading," in *The 3rd International Workshop on Deep Learning for Mobile Systems and Applications*, 2019 (cit. on p. 2).

[16] W. Sirichotedumrong, Y. Kinoshita, and H. Kiya, "On the security of pixel-based image encryption for privacy-preserving deep neural networks," in *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, 2019 (cit. on p. 2).

[17] A. H. Chang and B. M. Case, "Attacks on image encryption schemes for privacy-preserving deep neural networks," *CoRR*, vol. abs/2004.13263, 2020 (cit. on p. 2).

[18] G. Mai, K. Cao, P. C. Yuen, and A. K. Jain, "On the reconstruction of face images from deep face templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 5, 2019 (cit. on p. 2).

[19] X. Bai, J.-X. Li, Z. Yu, Z.-Z. Yang, Y.-J. Wang, X.-Y. Chen, and X. Zhou, "Reconstruction of chaotic grayscale image encryption based on deep learning," in *2021 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2021 (cit. on p. 2).

[20] C. He, K. Ming, Y. Wang, and Z. J. Wang, "A deep learning based attack for the chaos-based image encryption," *CoRR*, vol. abs/1907.12245, 2019 (cit. on p. 2).

[21] *ISO/IEC 19794-5:2011 information technology — biometric data interchange formats — part 5: Face image data*, 2011 (cit. on pp. 3, 7).

[22] K. Delac, M. Grgic, and S. Grgic, "Effects of JPEG and JPEG2000 compression on face recognition," in *Proceedings of ICAPR 2005*, ser. LNCS, vol. 3687, 2005 (cit. on pp. 3, 6).

[23] A. K. John, O. O. Williams, and M. A. Adewale, "Evaluating the effect of jpeg and jpeg2000 on selected face recognition algorithms.," *International Journal of Modern Education & Computer Science*, vol. 6, no. 1, 2014 (cit. on pp. 3, 6).

[24] C. Rathgeb, A. Uhl, and P. Wild, *Iris Recognition: From Segmentation to Template Security*, ser. Advances in Information Security. 2013, vol. 59 (cit. on p. 3).

[25] S. Chen, Y. Liu, X. Gao, and Z. Han, "Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices," in *Biometric Recognition*, 2018 (cit. on p. 3).

[26] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019 (cit. on pp. 3, 4).

[27] Y. Martínez-Díaz, M. Nicolás-Díaz, H. Méndez-Vázquez, L. S. Luevano, L. Chang, M. Gonzalez-Mendoza, and L. E. Sucar, "Benchmarking lightweight face architectures on specific face recognition scenarios," *Artificial Intelligence Review*, 2021 (cit. on p. 3).

[28] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *CoRR*, vol. abs/1411.7923, 2014. arXiv: `1411.7923` (cit. on p. 4).

[29] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-stage dense face localisation in the wild," *CoRR*, vol. abs/1905.00641, 2019. arXiv: `1905.00641` (cit. on p. 4).

[30] F. Dufaux, S. Wee, J. Apostolopoulos, and T. Ebrahimi, "JPSEC for secure imaging in JPEG2000," in *Applications of Digital Image Processing XXVII*, vol. 5558, 2004 (cit. on p. 4).

[31] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, 2007 (cit. on p. 4).

[32] T. Stütz and A. Uhl, "Transparent image encryption using progressive JPEG," in *Information Security. Proceedings of the 9th Infomation Security Conference (ISC'06)*, ser. Lecture Notes on Computer Science, vol. 4176, 2006 (cit. on p. 6).