Towards an Efficient Facial Image Compression with Neural Networks

Maria Ausilia Napoli Spatafora[1] [0000−0001−5259−7353], Alessandro Ortis[1] [0000−0003−3461−4679], and Sebastiano Battiato[1] [0000−0001−6127−2470]

Department of Mathematics and Computer Science
University of Catania, Italy
maria.napolispatafora@phd.unict.it, {ortis,battiato}@dmi.unict.it

Abstract. Digital images are more and more part of everyday life. Efficient compression methods are needed to reduce the disk-space usage for their storage and the bandwidth for their transmission while keeping the resolution and the visual quality of the reconstructed images as close to the original images as possible. Not all images have the same importance. The facial images are being extensively used in many applications (e.g., law enforcement, social networks) and require high efficient facial image compression schemes in order to not compromise face recognition and identification (e.g., for surveillance and security scenarios). For this reason, we propose a promising approach that consists of a custom loss that combines the two tasks of image compression and face recognition. The results show that our method compresses efficiently face images guaranteeing high perceptive quality and face verification accuracy.

Keywords: Convolutional autoencoder · Face images compression · Custom loss function

1 Introduction

The use of images is increasing in many different applications. Digital images require huge amounts of space for storage and large bandwidth for transmissions. Image compression is a solution and so it is a fundamental problem in computer vision and image processing, which finds several applications. Image compression techniques consist in two tasks [23]:

- coding: it reduces the number of bits required to represent an image by taking advantage of these redundancies.
- decoding or decompression: it is the inverse process; it is applied to the compressed data to get the reconstructed image.

It is very crucial to reduce the size of disk-space used as well as reduce the amount of internet bandwidth used for the transmission of images while keeping the resolution and the visual quality of the reconstructed image as close to the original image as possible. The signal-processing-oriented compression has a long history studded with algorithms that are still widely employed today (e.g.,

JPG). Nevertheless, the attainable performances have already achieved the maximum. For this reason, it is taking new approaches to image processing such as the Neural Networks due to the proliferation of specific hardware (e.g., GPUs) and the meaningful results in other computer vision fields. The increasing interest in image compression with Neural Networks is proven by international workshops and challenges on this subject and by human and financial efforts of lead companies in this area. Such is the case of NVIDIA Corporation that has recently developed a compression algorithm for video conferencing based on Neural Networks that is able to outperform competing methods [29]. In addition, the potential in this area has been demonstrating by profitable literature. During the year, autoencoders [1, 24], Recurrent Neural Networks [25, 26] and Generative Adversarial Networks [21] achieve good and promising performance. Particularly, facial images are being extensively used and stored in large databases by social networks, web services, or various organizations such as states, law enforcement, schools, universities, and private companies. Thus, efficient storage of such images is very useful. For this reason, face-specific image compression schemes are developed. This motivated us to explore the possibility to build new architectures for images specifically addressed for compression of face images. Facial image compression can be regarded as a special application of general image compression but this task has the need of high efficient facial image compression algorithms (e.g., in surveillance system) in order to not compromise face recognition and identification that have been developing rapidly in recent years. This drove us to propose a new approach for the compression of facial images through a custom loss function that joins efficiently the image compression need with the face recognition task.
The rest of the paper is organised as follow: in Section 2, we present related works on image compression with neural networks; Section 3 describes the proposed method; Section 4 reports experimental settings and results; Section 5 reports conclusions and future works.

2 Related Works

2.1 Image Compression

Image compression is an important research topic in the field of signal processing to achieve efficient image storage and transmission. During the past decades, many *ad hoc* algorithms and heuristics [28, 19, 14, 13, 32] have been proposed and some of them became common standards such as JPG.
Recently, computer vision is applied in many research fields showing high capabilities including image compression in order to obtain perceptual and disk-space performance better than those of previous approaches. Different neural approaches have been adopted such as Autoencoder [1, 24], Recurrent Neural Network [25, 26] and Generative Adversarial Network [21].
Autoencoder architectures have been adopted also for purposes that bring forward the compression image. Autoencoder has been employed to reduce the dimensionality of images [10], to convert images to compressed binary codes

for retrieval [15] and to extract compact visual representations for several purposes [27]. More recently, the authors in [1] and [24] both exploited methods based on convolutional autoencoder architectures; Ballè et al. [1] used Generalized Divise Normalization (GDN) for joint nonlinearity and approximated the quantized values with additive uniform noise for entropy rate loss; while Theis et al. [24] employed a smooth approximation of the derivative of the rounding function and upper-bound the discrete entropy rate loss.
Works in [25] and [26] explored various transformations for binary features extraction based on different types of RNNs that adopt stochastic quantization, and finally, the binary representations are entropy-coded. Toderici et al. [25] proposed a RNN to compress 32×32 images; while, the work in [26] further introduced a set of full-resolution compression methods for progressive encoding and decoding of images.
Finally, GAN architectures [9] have been applied in the field of image compression, too. Santurkar et al. [21] described the concept of generative compression that is the compression of data using generative models; this approach produces accurate and visually pleasing reconstructions at much deep compression levels for image data.

2.2 Facial Image Compression

Although previous approaches have been widely used and designed for general purposes, facial images are specific objects and play a very important role in technology and science, military, medical and other industries. Facial images differ from other natural images and have specific prior knowledge as presented by [16] that analyzes statistically gradient features of facial images. The gradient features of facial images are compared with those of other natural images showing that they are different in symmetry and geometry structure. For this reason, there are also some face-specific image compression schemes such as works [7, 3, 20] that design dictionary-based coding schemes on this specific type of image. Liu et al. [17] proposed a PCA method to realize facial image compression extracting common features from this type of image in order to reduce redundancy to achieve compression.
In recent years, some approaches based on deep learning have been proposed to represent and compress facial images. Hu et al. [11] proposed a content-aware facial image compression method to compress facial images at low bit rates using Convolutional Autoencoder. Chen et al. [4] proposed a GAN architecture that is able to automatically optimize codec configuration according to the minimization of semantic distortion. Finally, Bian et al. [2] developed a Convolutional Autoencoder combined with a standard lossless image code PNG. This framework is followed by the measurement of performance for facial recognition on compressed images.
Differently from previous works, we propose a new approach for facial image compression that is able to simultaneously save enough space and ensure good performance in face recognition task. The method is summarised in the next section.

3 Proposed Method

In this section, we describe the developed convolutional autoencoder architecture for face image compression and the workflow of the algorithm.

The proposed architecture is made by three parts: the encoder (*Enc*), the decoder (*Dec*), and the binarizer (*Bin*). The overall pipeline of the process is:

1. *Enc* generates a compact representation of the input image for the encoding: it takes RGB PNG image of any size as input and produces a compressed representation in `xz`[1] format exploiting LZMA algorithm [8];
2. *Bin* is used for quantifying the output of *Enc*;
3. *Dec* is the mirror symmetry of *Enc* and reconstructs an image in PNG format from its binary representation.

The input of the binarizer - marked as x - is the output feature vector from the encoder part and consists of continuous floating-point numbers. The output of the binarizer - marked as $Bin(x)$ - is the final compressed data after the quantization step provided by the binarizer, and it is composed of discrete binary values. The binarizer follows the rule defined in Equation (1), where i is the i-th element of encoding output vector and of P that is a vector x-like filled by values drawn from a uniform probability.

$$Bin(x[i]) = \begin{cases} -1, \texttt{ if } & \frac{1-x[i]}{2} <= P[i] \\ +1, & \texttt{otherwise} \end{cases} \tag{1}$$

The Fig. 1 shows the structure of the neural network employed. For the encoding process, input images pass through two convolutional structures containing a convolutional layer, followed by batch normalization and Rectified Linear Unit (ReLU) activation function each. The first convolutional structure contains 32 filters of size 8×8 (with stride of 4 and padding of 2) while the second contains 64 filters of size 2×2 (with stride of 2). After these two convolutional structures, the processed input image goes through the binarizer that consists of a single convolutional layer with 128 filters with a size of 1×1 followed by the hyperbolic tangent (tanh function). This processed data vector, finally, passes through the function detailed in Equation (1). At this point, the network produces the compressed archive in `xz` format.

For the decoding process, input data pass through two transposed convolutional structures containing transposed convolutional layer, batch normalization, and ReLU activation function. The first transposed convolutional structure contains 32 filters of size 8×8 (with stride of 4 and padding of 2) while the second contains 3 filters of size 2×2. The output is the image in PNG format.
We choose to use batch normalization after convolution and transposed convolution. While we adopt the ReLU activation function.

The architecture becomes specific for facial image compression with a custom loss function whose goal is to bind the task of image compression and the task

[1] This acronym is a file format for LZMA archives.

Fig. 1. Structure of the employed neural network.

of face recognition. The custom loss is a linear combination of two losses: the one works on pixels space and the other works on features loss

$$\texttt{custom loss} = a \cdot \texttt{pixel loss} + b \cdot \texttt{feature loss} \tag{2}$$

Here, a and b are the weights assigned to the two parts of the loss. The pixel loss aims to obtain a reconstructed image more and more similar to the original. While, the feature loss has the goal of achieving a face-embedding of reconstructed image more and more near to that of the original. We select FaceNet [22] as the face-embedding. It is a neural-network-based tool[2] that maps face to a compact Euclidean space. Such space amplifies distances of faces from distinct people, while reduce distances of faces from the same person. This model is pre-trained with triplet loss and center loss.

4 Experimental settings and results

In this section, we will introduce the datasets and the conducted experiments and comparisons for facial compression. The results are encouraging and promising. The training of the proposed convolutional autoencoder happens in two phases:

1. *Image Compression*: the above network is first trained on dataset provided by the "3rd Workshop and Challenge on Learned Image Compression" [5] that has taken place during the "CVPR 2020" [6] (the Fig. 2 shows some images from this dataset). The organizers of this workshop and challenge, provide a set of images which represent a very realistic view of the types of images that are widely available nowadays to ensure realism. This dataset contains 568 images. For the purpose of the training, we adopt the L1 loss, the Multi Scale Structural Similarity (MS-SSIM) as learning metric (it is a convenient way to incorporate image details at different resolutions [30] and it is suggested by the challenge [5]) and the following hyper-parameters: dynamic learning rate 0.0001; momentum 0.99. Finally, we use Adam optimizer.
2. *Face Image Compression*: it is used the model pre-trained in the previous phase. The employed dataset is CelebA aligned [18] converted to PNG. CelebA contains 10177 number of identities and 202599 number of facial images. We eliminated the faces that cannot be detected by MTCNN [31]. Here, the combined loss (Equation 2) is employed together with MTCNN-FaceNet pipeline for the features-space loss in the following way:
   - MTCNN provides the bounding-boxes of the face in the groundtruth image;
   - these bounding-boxes are used to extract faces in the groundtruth image and in the output image of the network;
   - the patches corresponding to the faces are fed to FaceNet that computes two face-embeddings;
   - the face-embeddings are compared with the features-space loss.

   For this loss, we adopt L2 Loss for both parts and the weights $a = b = 1$. The other hyperparameters are static learning rate 0.003, momentum 0.99, weight decay 0.001 and we use Adam as optimizer.

After the training of our autoencoder, we test it on another dataset containing only faces. For this purpose, we choose the Labeled Faces in the Wild (LFW) [12]. This benchmark dataset contains 13233 face images of 5749 people with a picture size of 112×96.
In order to properly evaluate the validity of the method, we use on the one hand

[2] It is available here https://github.com/timesler/facenet-pytorch

Fig. 2. Some images from dataset provided by the "3rd Workshop and Challenge on Learned Image Compression"[5]

compression ratio, bit-per-pixels (bpp), Peak signal-to-noise ratio (PSNR), and MS-SSIM as metrics for the image compression task, and on the other average cosine similarity (1 indicates two vectors with the same orientation; -1 indicates two vectors diametrically opposed) and face verification through 1-NN with the cosine distance as metric for the face recognition task. We compare our results in the face recognition task with the JPG to have a point of reference. Moreover, we compare our method injected with the custom loss (Equation 2) with the same method without the custom loss. All these results are reported in the Tables 1, 2 respectively for the image compression task and for the face recognition task with the performances of JPG at different compression quality levels. The results prove the validity of the proposed loss that improves our performance enough to reach JPG's performance. Actually, the custom loss function decreases the number of faces not detected in reconstructed images and increases both the average cosine similarity and the face verification accuracy. It should be noted that JPG is a little bit better than our method with custom loss only for the face verification accuracy. The Fig. 3 shows some faces compressed with our method and JPG (quality factor 80) proving that our method with the custom loss preserves good perceptive quality of the images. These results releave interesting insights in the research topic of the facial image compression with neural networks.

Table 1. Results of the image compression task with our proposed approach about LFW dataset.

	Average	**Maximum**	**Minimum**
bpp	0.761	0.864	0.507
Compression ratio	5.892	8.065	3.336
PSNR	30.905	35.064	28.671
MS-SSIM	0.989	0.998	0.857

Table 2. Results of the face recognition task after compression and comparison with the approach without the custom loss and JPG about LFW dataset.

	Compressed dataset (MB)	**Average cosine similarity**	**Face Verification Accuracy**	**Faces not detected**
Our method without custom loss	34	0.422	0.306	68
Our method with custom loss	40	0.801	0.944	1
JPG (quality factor 70)	35	0.768	0.986	2
JPG (quality factor 80)	41	0.780	0.986	2
JPG (quality factor 90)	56	0.793	0.990	2

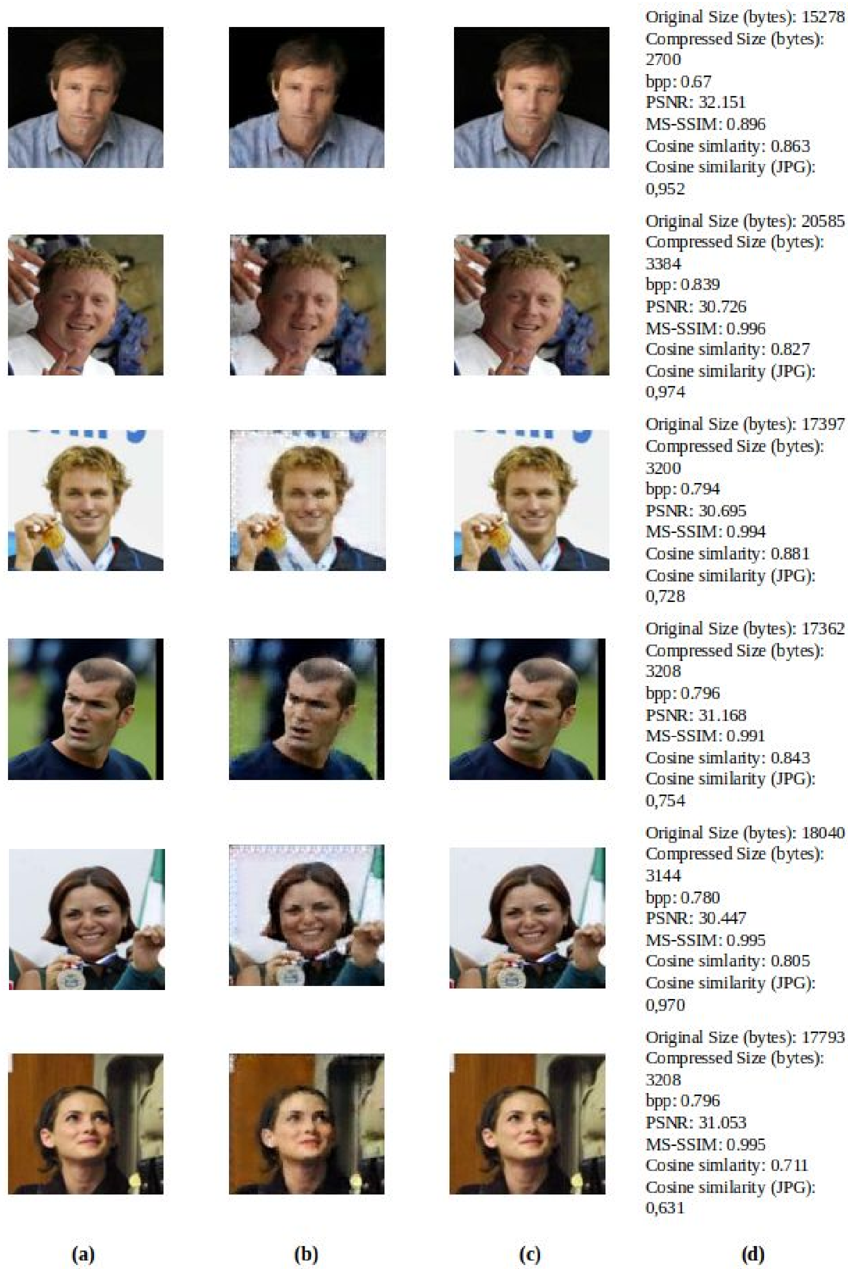

Fig. 3. **(a)** Original image; **(b)** Compressed and reconstructed image by our convolutional autoencoder; **(c)** Image in JPG with quality factor 80; **(c)** Metrics.

5 Conclusions

In this paper, we explore the problem of image compression for a specific type of image such as those containing only faces. Facial image compression is a significant topic because surveillance and security scenarios need efficient compression algorithms to not compromise facial recognition and identification. We develop simplistic convolutional autoencoder architecture that is trained first on images of various genres and next on facial images after the injection of a custom loss that is able to guarantee the face recognition and face verification. Experimental results validate our proposed method achieved a state-of-the-art performance compared to the coding standard of JPG from the point of view of objective quality metrics and face verification task.

The experiments reported in this paper represent a first investigation attempt on this research topic, which outcomes revealed insights suggesting that there is room for improvements in this research direction. Moreover, there are not many studies in literature for facial image compression and those existing do not have a unique touchstone that makes a fair comparison difficult with the state-of-the-art for the face image compression field. For this reason, we are planning to build a benchmark dataset of facial images with the indication of different quality factors of some algorithms (e.g., JPG), and the face recognition performance with some face embedding algorithms (e.g., FaceNet). Future works will be also devoted to the improvement of the compression ratio preserving the face recognition performance in order to outperform JPG. This could be done with a more complex neural network architecture characterized by residual blocks and entropy coding that have already been applied in the image compression topic.

References

1. Ballé, J., Laparra, V., Simoncelli, E.: End-to-end Optimized Image Compression. In: Proceedings of the Internation Conference on Learning Representation (ICLR) (11 2017)
2. Bian, N., Liang, F., Fu, H., Lei, B.: A Deep Image Compression Framework for Face Recognition. In: 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI). pp. 99–104 (2019). https://doi.org/10.1109/CCHI.2019.8901914
3. Bryt, O., Elad, M.: Compression of Facial Images Using the K-SVD Algorithm. J. Vis. Comun. Image Represent. **19**(4), 270–282 (2008). https://doi.org/10.1016/j.jvcir.2008.03.001
4. Chen, Z., He, T.: Learning based Facial Image Compression with semantic fidelity metric. Neurocomputing **338**, 16–25 (2019). https://doi.org/https://doi.org/10.1016/j.neucom.2019.01.086
5. CLIC: CLIC, Workshop and Challenge on Learned Image Compression. http://www.compression.cc/, online; last accessed 26th April 2021
6. CVPR: CVPR 2020. http://cvpr2020.thecvf.com/, online; last accessed 26th April 2021
7. Elad, M., Goldenberg, R., Kimmel, R.: Low Bit-Rate Compression of Facial Images. IEEE Transactions on Image Processing **16**, 2379–2383 (2007)
8. Epiphany, J.L., Danasingh, A.A.: Hardware Implementation of LZMA Data Compression Algorithm. International Journal of Applied Information Systems **5**, 52–56 (2013)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, S., Bengio, Y.: Generative adversarial networks. In: Proceedings of the 27th International Conference on Neural Information Processing Systems (ICONIP). vol. 63, pp. 2672–2680 (2014)
10. Hinton, G.E., Salakhutdinov, R.: Reducing the Dimensionality of Data with Neural Networks. Science **313**, 504 – 507 (2006)
11. Hu, S., Duan, Y., Tao, X., Liu, Y., Zhang, X., Lu, J.: Content-aware Facial Image Compression with Deep Learning Method. In: 2020 International Conference on Wireless Communications and Signal Processing (WCSP). pp. 516–521 (2020). https://doi.org/10.1109/WCSP49889.2020.9299680
12. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. rep., University of Massachusetts, Amherst (2007)
13. Hurtik, P., Perfilieva, I.: A hybrid image compression algorithm based on jpeg and fuzzy transform. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). pp. 1–6 (2017). https://doi.org/10.1109/FUZZ-IEEE.2017.8015614
14. Jassim, F., Qassim, H.: Five Modulus Method For Image Compression. Signal and Image Processing : An International Journal **3** (2012). https://doi.org/10.5121/sipij.2012.3502
15. Krizhevsky, A., Hinton, G.: Using Very Deep Autoencoders for Content-Based Image Retrieval. In: Proceedings of the 19th European Symposium on Artificial Neural Networks (ESANN) (2011)
16. Liu, X., Gan, Z., Liu, F.: Hierarchical Subspace Regression for Compressed Face Image Restoration. In: 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP). pp. 1–6 (2018). https://doi.org/10.1109/WCSP.2018.8555682

17. Liu, Y., Kau, L.: Scalable face image compression based on Principal Component Analysis and arithmetic Coding. In: 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW). pp. 265–266 (2017). https://doi.org/10.1109/ICCE-China.2017.7991097
18. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 3730–3738 (2015). https://doi.org/10.1109/ICCV.2015.425
19. Rabbani, M., Joshi, R.: An overview of the JPEG 2000 still image compression standard. Signal Processing: Image Communication **17**, 3–48 (2002)
20. Ram, I., Cohen, I., Elad, M.: Facial Image Compression using Patch-Ordering-Based Adaptive Wavelet Transform. IEEE Signal Processing Letters **21**(10), 1270–1274 (2014). https://doi.org/10.1109/LSP.2014.2332276
21. Santurkar, S.and Budden, D., Shavit, N.: Generative Compression. In: Proceedings of the Picture Coding Symposium (PCS) (2017)
22. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 815–823 (2015). https://doi.org/10.1109/CVPR.2015.7298682
23. Subramanya, A.: Image compression technique. IEEE Potentials **20**(1), 19–23 (2001). https://doi.org/10.1109/45.913206
24. Theis, L., Shi, W., Cunningham, A., Huszár, F.: Lossy Image Compression with Compressive Autoencoders. In: Proceedings of the Internation Conference on Learning Representation (ICLR) (2017)
25. Toderici, G., O'Malley, S., Hwang, S., Vincent, D., Minnen, D., Baluja, S., Covell, M., Sukthankar, R.: Variable Rate Image Compression with Recurrent Neural Networks. In: Proceedings of the Internation Conference on Learning Representation (ICLR) (2016)
26. Toderici, G., Vincent, D., Johnston, N., Hwang, S., Minnen, D., Shor, J., Covell, M.: Full Resolution Image Compression with Recurrent Neural Networks. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR). pp. 5435–5443 (2017). https://doi.org/10.1109/CVPR.2017.577
27. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning. pp. 1096–1103 (2008). https://doi.org/10.1145/1390156.1390294
28. Wallace, G.K.: The JPEG still picture compression standard. IEEE Transactions on Consumer Electronics **38**(1), xviii–xxxiv (1992). https://doi.org/10.1109/30.125072
29. Wang, T., Mallya, A., Liu, M.: One-shot free-view neural talking-head synthesis for video conferencing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2021)
30. Wang, Z., Simoncelli, E., Bovik, A.: Multiscale structural similarity for image quality assessment. In: Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems and Computers. vol. 2, pp. 1398 – 1402 (2003). https://doi.org/10.1109/ACSSC.2003.1292216
31. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters **23**(10), 1499–1503 (2016). https://doi.org/10.1109/LSP.2016.2603342
32. Zhang, Y., Cai, Z., Xiong, G.: A New Image Compression Algorithm Based on Non-uniform Partition and U-System. IEEE Transactions on Multimedia **23**, 1069–1082 (2021). https://doi.org/10.1109/TMM.2020.2992940