Deep Learning-based Compression Artifacts Reduction for JPEG Image Classification

Andrey O. O. dos Reis
Dept. of Electrical Engineering (ENE)
University of Brasilia (UnB)
Brasilia, Brazil
andrey.reis@aluno.unb.br

Edson M. Hung
Dept. of Electrical Engineering (ENE)
University of Brasilia (UnB)
Brasilia, Brazil
mintsu@unb.br

Daniel G. Silva
Dept. of Electrical Engineering (ENE)
University of Brasilia (UnB)
Brasilia, Brazil
danielgs@unb.br

***Abstract*—Computer vision has been one of the main application areas for Deep Learning (DL) techniques. Image classification, in particular, has been widely culminating in systems increasingly capable of replacing human visual analysis. However, having high-quality images is one of the main requirements for high performance of DL-based classification, but often factors associated with the models application condition make this ideal scenario impossible. Lossy compression fits very well in these conditions because it is a type of coding technique widely used in image storage and transmission. In this paper, we propose compression artifact reduction (CAR) as a way to circumvent the degradation problem for lossy compression. To this end, we perform experiments using JPEG with very low levels of quality factor (QF) compressing the test dataset for classification. Then, using a DL-based CAR model, we restore this same dataset in order to investigate a possible improvement in the classifiers performance. The two evaluated datasets presented positive results: in Flowers-102 we reached an average 0.52% increase in accuracy for 10 QFs values, whereas for Cub-200 this value was more expressive, around 34.57%. Those findings reinforce that DL-based CAR may increase performance in classification models degraded by drastic levels of signal corruption.**

***Index Terms*—deep learning, image classification, image restoration, compression artifacts, JPEG**

I. Introduction

Deep Learning (DL) is a category of Machine Learning models within the context of Artificial Intelligence that presents a higher learning capacity. Recently, DL has been widely used for several signal processing applications, either in predictive models or in the treatment of these signals. Image classification is inserted in this set of applications as the most prominent, because it already has high efficiency models, ready to be used, in the market. However, lossy compression, a widely used type of coding which reduces signal quality, raises some concern for being detrimental to the classifiers performance.

Among image encoders, JPEG [1] is certainly one of the most widely used. The compressor high efficiency is associated with its structure, which involves frequency domain transformation, scalar quantization and entropic coding of zigzaged coefficients, as Fig. 1 shows. Initially, the discrete cosine transform (DCT) is applied on 8x8 blocks of the image. Resulting coefficients are quantized according to a quantization table (Q-Table) that is established according to the quality factor (QF) assigned to the encoder. The higher the QF, the more quantization levels are offered to each of the coefficients, on the other hand, the lower the QF, the fewer levels each coefficient will have access to. The artifacts generated during the compression process are directly linked to the choice of QF. Blocking is caused by discontinuity observed between the image processed blocks. Attenuation of high frequencies in the quantization, in turn, provides ringing (analogous to the Gibbs effect) and the loss of these components, i.e. blurring.

In the real world, photographs captured by smartphones are compressed and stored with a high quality factor (QF 95) and then may undergo more aggressive compression when transmitted to another device, in order to save bandwidth. This scenario highlights the existence of doubly compressed signals. Moreover, datasets generally used as pre-training for Convolutional Neural Networks (CNNs) such as Imagenet [2] and CIFAR [3] are already available in JPEG, i.e. they have a certain level of compression. This may indicate presence of signal degradation when these datasets are employed as the training set of classification models.

In this context, artifact reduction methods can be applied in order to enhance images and, thus, positively interfere in the performance of image classifiers. Initially, transform-based techniques [4] were able to increase the quality of images, but in spite of reducing artifact blocking, limitations of filters led to an increase of blurring and ringing similar effects. Thus, CNN-based techniques have been recently explored and showed results superior to classical signal processing techniques.

In this paper, we propose to improve DL-based image classifiers performance in face of lossy compressed images, by firstly using CNN-based Compression Artifact Reduction (CAR). We use one such CNN-based image restoration model as a pre-processing step in order to increase the generalization performance of classifiers. The central idea is to improve the input signal quality of the classification models, which has been intensively compressed, in order to preserve original classifier performance as much as possible. Our experiments further address a more challenging scenario, i.e. that of doubly compressed images. The conclusions drawn from our results show a certain effectiveness of CAR even under extreme signal

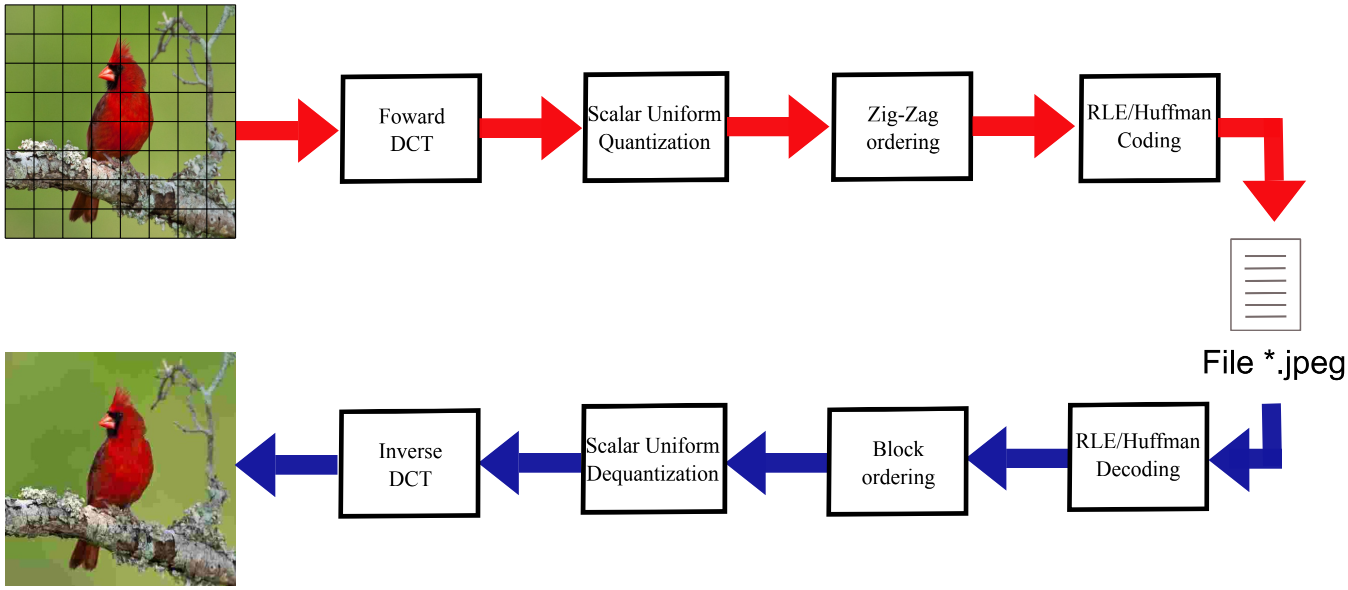


Fig. 1: Structural model of the JPEG encoder. The encoding path is in red and the decoding path is in blue. Quantization is the step responsible for inserting losses into the image.

corruption conditions.

The rest of this paper is structured as follows. Section II presents the related work. In Section III, we state the proposed CAR model as well as implementation details of the respective classifiers used. Section IV describes performed experiments and an analysis of the obtained results. Finally, in Section V we present the final considerations.

II. Related Work

A. Compression Impacts on Image Classification

In the context of training Deep Learning models, the problem of image transmission and storage is very common due to the need of large databases for the success of the learning process. The evaluation of compression impacts on mammographic images is done by Jo et al. [5], showing the decrease of DL model performance for cancer diagnosis as the compression rate increases. A similar behavior occurs in steel surface classification [6], which is mitigated with data augmentation based on compressed images. The problem also extends to object detection [7], which imposes an additional degree of difficulty for the models, because it requires localization within the image. Overall, these works indicate the possibility of preserving some performance of CNNs even under considerable compression rates.

In contrast, strategies have begun to appear to curb model degradation in the face of compression. The Highest Rank Selector proposed by Yang et al. [8] presents a technique for choosing the JPEG quality factor for a given predictor. Optimization of the JPEG quantization process for image classification, on the other hand, is presented in [9]. In this technique, the encoder Q-Table is treated as another classifier hyperparameter, which is selected via random search and Bayesian optimization during the model training. Results show an increase in performance for similar compression ratios.

B. Deep Learning for Compression Artifacts Reduction

Since the emergence of lossy image coders, the incidence of artifacts has always been a concern. At first, methods exclusively based on signal processing started to be developed to improve signal quality, however, with the advance of CNNs, learning-based techniques also started to be explored. The “Artifacts Reduction Convolutional Neural Network (ARCNN)” proposed by Dong et al. [10] was one of the first CNN-based models designed to remove JPEG artifacts. Based on the super resolution technique, the four-layer CNN is trained with both compressed and uncompressed images.

Usually, QF is a parameter that must be established in order to choose the adequate restorer, but recently models such as [11], [12] are able to perform CAR without the need of QFs prior knowledge. Called blind CARs, the performance achieved by these networks is superior than those of conventional networks. This type of approach provides the use of a single model trained for all distinct levels of compression.

C. Image Restoration with Semi-Supervised Learning

As indicated earlier, the main idea of restoring images by Deep Learning consists in training a CNN with pairs of clean and corrupted images, e.g. raw and JPEG images. However Noise2Noise (N2N) developed by Lehtinen et al. [13] is an approach where only corrupted images are used in learning. Research reveals the possibility of obtaining higher quality images just by looking at low quality images. To do so, the model training must contain pairs of images with two distinct levels of corruption, i.e. one signal with little corruption and the other with too much corruption.

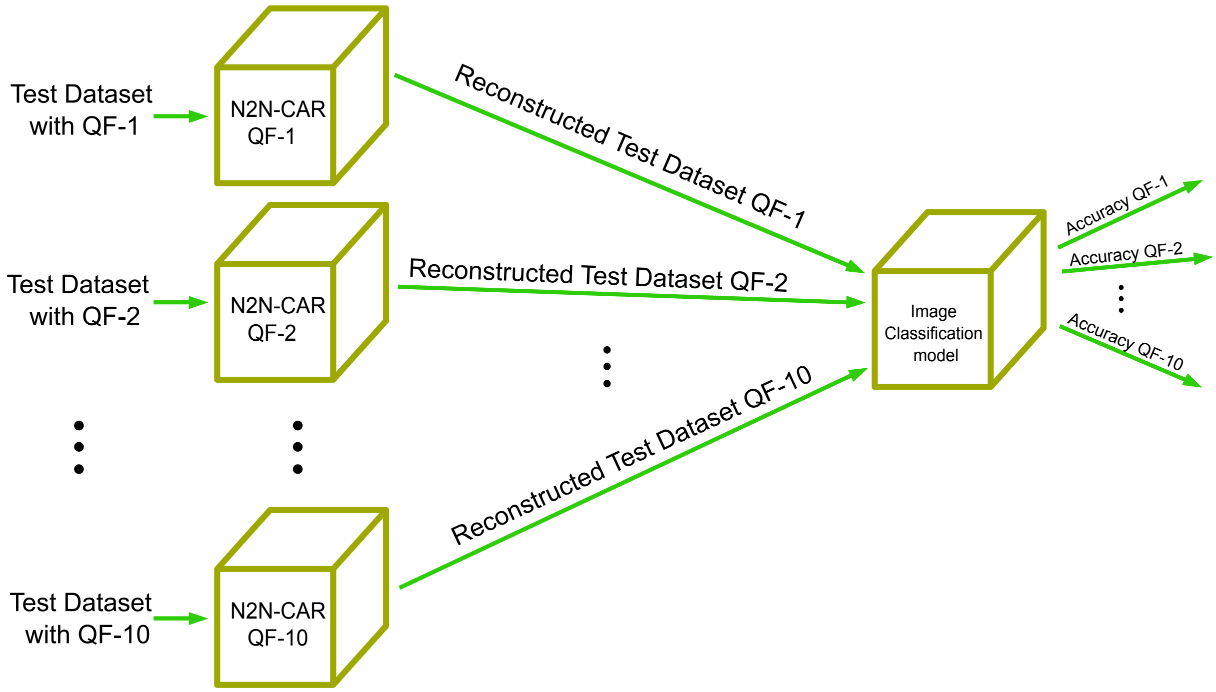


Fig. 2: Schematic of the experiment performed. Initially, the test dataset is compressed to different QFs (1 to 10). Then it is restored by N2N-CAR trained on the compressed training dataset with its respective QF. Finally, validation is performed with the classification model trained on the original training dataset.

Several techniques based on N2N started to appear, exactly because it is a relatively efficient method in situations where high quality data is scarce. Medical images (MRI, CT and PET scans) are one of the main focuses of application, as they need improvement in the face of the low sampling problem. In addition, images that have noise sources similar to those indicated in the research (Gaussian, Poisson and Bernoulli) can greatly benefit from this strategy, e.g. despeckle of SAR satellite images [14]. Even in other types of signals, such as audio, this approach shows promising results [15].

In the case of double compressed images, a scenario where the images undergo a first compression in the capture device and a second one for signal transmission. Deep Learning models that operate only on these corrupted signal are just beginning to appear [16]. But the treatment of doubly compressed images has already started to show results in the CAR state of the art. The main problem in this case concerns images where $Q_2 > Q_1$, i.e. the first compression ratio (Q_1) is greater than the second (Q_2), which is the exact problem we propose to solve.

III. Methodology

Our tests were performed on two very popular image classification datasets: Oxford Flowers-102 [17] and Caltech Cub-200 [18]. Flowers-102 is divided into training, validation and test datasets, with 6149, 1020 and 1020 images for each dataset, respectively. Cub-200 dataset is divided into 5994 images for training and 5794 for testing. First we trained the reconstruction models and compressed the test datasets with different compression levels (QFs). Then we used the CAR on the compressed test datasets and performed the test on previously trained classifiers as shown in Fig. 2. In both steps both the CAR and classification models were trained on the same images.

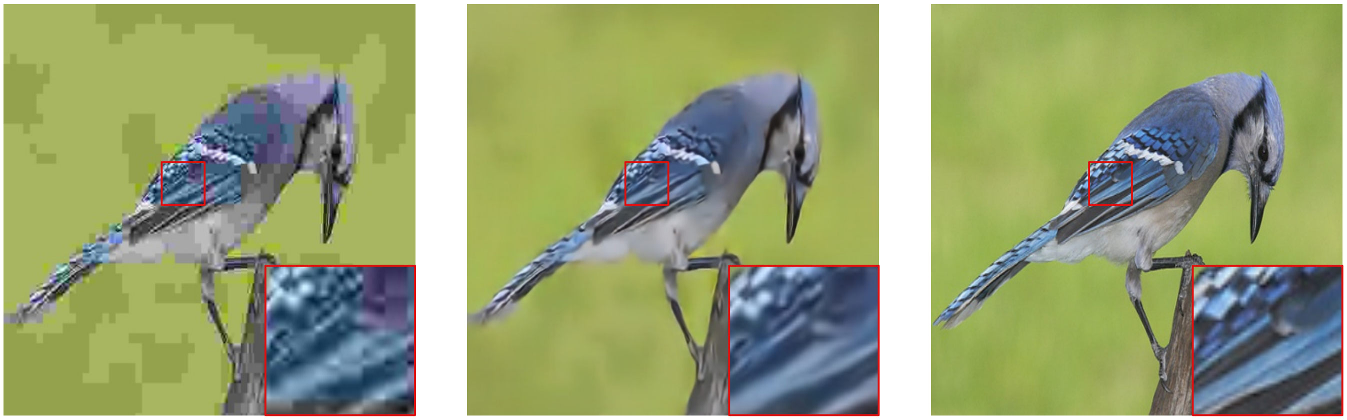
A. Compression Artifacts Reduction with N2N U-Net

The N2N image restoration model is structured on the U-Net architecture [19]. This CNN has 23 convolution layers divided into a contraction path and an expansion path. In the first path, the image goes through successive pairs of 3×3 convolutions followed by a rectified linear unit (ReLU) and shrunk by 2×2 max pooling with stride 2. In the second path, the feature map is augmented by 2×2 convolution, concatenated with the cropped feature map of the associated contraction path, followed by the pair of 3×3 convolution with ReLU. The output has the same dimension of the input image, i.e. it represents the restored image. In the case of RGB images, three input channels result in three output channels.

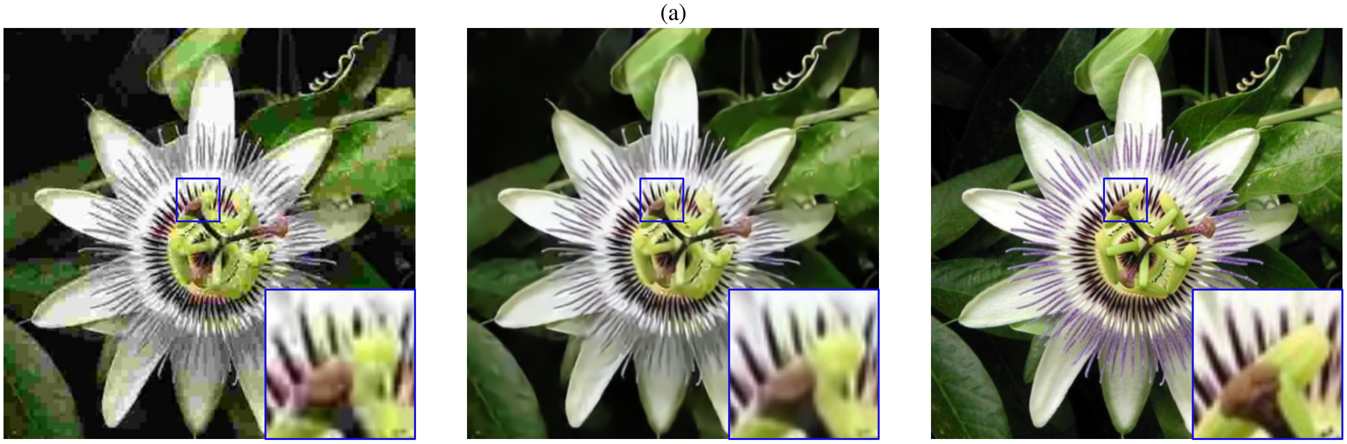
We train the network for both datasets separately, i.e. using their respective training datasets. In both cases we set the learning rate to 10^{-3} and the ADAM optimizer with the parameters $\beta_1 = 0.9$, $\beta_2 = 0.99$, $\epsilon = 10^{-8}$. The network weights θ were updated by minimizing the l_2 loss function:

$$\theta^{l_2} \in \arg\min_{\theta} \sum_{i=1}^{N} (y_i - f_\theta(x_i))^2 \tag{1}$$

where, x_i is the compressed signal, f_θ is the neural network, y_i is the original uncompressed signal and N is the batch-



(a)



(b)

Fig. 3: Left: JPEG-compressed image with QF equal to 5. Center: image restored by N2N-CAR. Right: original image. (a) Cub-200 samples with respective PSNR values: 30.53 dB and 31.43 dB. (b) Flowers-102 samples with respective PSNR values: 29.99 dB and 31.06 dB. In both (a) and (b) it is possible to observe a reduction in compression artifacts and an increase in objective quality.

TABLE I: Training parameters for the classification models.

Classification Models	**Implementation Details**					
	Input size	***Optimizer***	***Optimizer parameters***	***Learning rate***	***Batch-size***	***Epochs***
CCT-14/7x2	384x384	ADAM	$\beta_1 = 0.9$, $\beta_2 = 0.999$	0.001	20	300
MMAL-Net	448x448	SGD	$\mu = 0.9$, $\lambda = 0.0001$	0.001	6	144

size. For Flowers-102 we assigned a batch size of 13 and for Cub-200 a size of 9.

In addition, we trained our own model for the compressed images at each compression level, with 10 distinct QFs in the range [1,10]. Each model served to restore its corresponding compressed test dataset to the same QF as the training dataset. Figure 3 shows the example of CAR on a Cub-200 and a Flowers-102 image compressed with QF of value 10, the three types of artifacts reduction is remarkable. We used approximately 15 hours of an NVIDIA GeForce RTX 2080 Ti GPU for the complete experiment, i. e. one and a half hours for each model.

B. Image Classification Models

In this section, we present the classification models used for each of the mentioned datasets. In both cases, we loaded the models with the weights that were previously trained by their designers, in order to reproduce their results. Training specifications are shown in Table I. It is important to note that the datasets fall into a special type of image classification, the Fine-Grained Visual Categorization (FGVC). In this group of Computer Vision tasks the number of classes is larger and the visual differences between them is smaller.

1) Flowers-102: For this dataset we use a classifier structured on the Compact Convolution Transformer (CCT) [20]. This network is a compact version of the Vision Transformer (ViT) [21], a transformer-based self attention model. Its architecture is divided into Convolutional Layer, Pooling, Reshape, Transformer Encoder, Sequential Pooling and Linear Layer. Despite the convolutional layer present in the network, there is no similarity with a CNN, because in this model this layer is inserted in a process called Tokenization. The advantage of this type of architecture is the ability to be trained with a smaller dataset while maintaining the high performance of ViTs. In the Flowers-102 classification the CCT-14/7x2 is used, which according to the specifications has 14 encoder transformers and 7 convolutional layers with 2×2 kernel. With the weights provided by the CCT designers, a 99.9% accuracy on the test set was obtained.

2) Cub-200: Unlike CCT, the Multi-branch and Multi-scale Attention Learning Network (MMAL-Net) [22], used in Cub-200, has CNNs as its backbone. The architecture consists of three branches, each one has its own CNN and Fully Connected Layer that shares the same weights. Raw branch, object branch and part branch are connected by two modules: the Attention Object Location Module and the Attention Part Proposal Module. These modules are responsible for locating the main object in the image and for highlighting important parts of these objects, respectively. The CNN present in MMAL-Net is ResNet50 [23]. This classifier achieves a 89.61% accuracy.

IV. Practical Experiments

We compressed both the training and test datasets using JPEG with QFs in the range [1, 10]. These very low QF values were chosen to force some reduction in classification performance, otherwise a higher QF does not impact in the classifiers accuracy. After training the CAR models for both datasets, we were left with a total of 20 CAR models that were subsequently applied to the compressed test datasets, in order to increase the quality of the images degraded by compression. This scenario addresses drastic signal degradation conditions, on the other hand, it is easily observed where there are computational limitations, either in relation to storage or in relation to bandwidth and energy consumption for transmission.

Test step for CCT and MMAL-Net consisted in using both compressed dataset and CAR-restored dataset, to obtain the models performance in both situations. Fig. 4 shows the accuracy obtained by the classifiers in these two conditions for the cited QFs. The first plot involves the CCT model. There are two curves with very similar behavior, however with one slightly overlapping the other. Due to the application of CAR in one of the datasets it is possible to observe the performance elevation. Similarly, Fig. 4b deals with MMAL-Net which is much more benefited by the CAR at the same time that it suffers more damage due to compression.

Compared to the compressed images only, the CAR on average provided a 0.52% increase in accuracy for CCT and a 34.57% increase for MMAL-Net. This is a performance increase in both cases, but it was much higher for the CNN-based model. Some factors that may be associated with this result are: network architecture, larger number of classes and dataset quality. Regarding architecture, transformer models may have a higher robustness to lossy compression, and this already makes the harmful impacts of artifacts less severe. Regarding the number of categories, Cub-200 has almost twice as many classes as Flowers-102, which makes the classification task more arduous. Finally, the fact that datasets are pre-compressed indicates the possibility of further degradation of Cub-200.

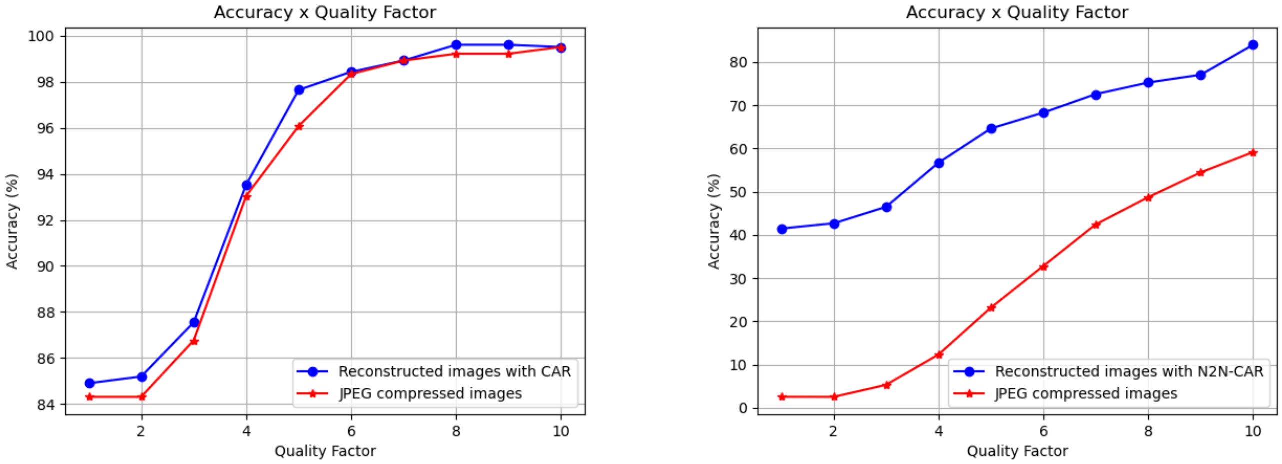
We also calculate average peak signal to noise ratio (PSNR) of each dataset relative to the original dataset, which is depicted in Fig. 5. By analyzing the curves, it is possible to extract important information about the results. Firstly, N2N-CAR curve (green) shifted to the right indicates a better objective quality of the restored datasets, higher PSNR. Secondly, in Fig. 5a accuracy values on the compressed dataset only (cyan curve) in some cases are higher for the same PSNRs, i.e., even with increasing PSNR the accuracy is lower with the Flowers-102 images generated by N2N-CAR (green curve below the cyan curve). Finally, in Cub-200 (Fig. 5b) the opposite happens at the same PSNR values the accuracy using the restored datasets is higher (cyan curve below the green curve). These findings show that apparently lossy compression PSNR reduction can be more impactful in some models than in others.

It is worth noting that besides increasing the objective quality (PSNR), the main goal of CAR is associated with increasing the signal subjective quality. Removing artifacts does not necessarily imply an increase in PSNR. Thus, each model may behave differently when facing the DL-based CAR, including losing performance with a higher PSNR and a lower amount of artifacts. Apart from the observed divergences, overall the positive impact prevailed for both classification models and datasets.

Despite the performance increase, Figure 6 highlights a possible concern, the negative impact of image restoration on classification. Such cases certainly occur less frequently than otherwise, e.g. in this dataset (Cub-200 with quality factor 9) only one image was classified right for compression only and wrong for compression and restoration. But it warns against the following situation: despite the better signal quality, if CAR is not considerably efficient, the effect can be the opposite of what is expected. And as observed for CCT, in many cases there is no increase in accuracy, evidencing that increased signal quality will not always result in improved classification performance.

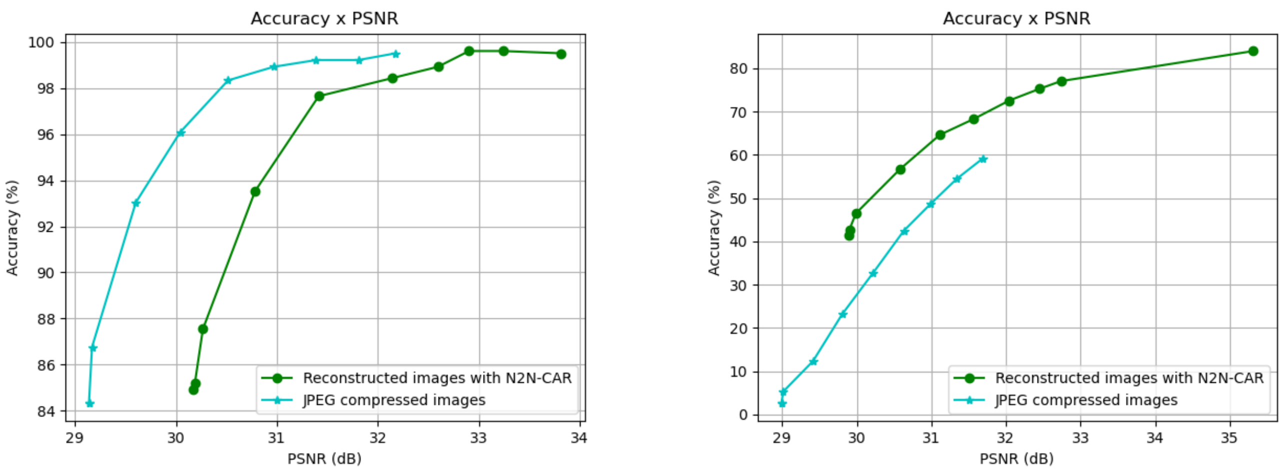
V. Conclusion

This paper presented a possible method to circumvent the problem of compressed image classification. The use of CAR for this purpose relies on the fact that the dataset quality is strictly related to the performance of DL-based classifiers. We performed N2N-CAR training for different compression levels on each dataset evaluated and then we analyzed the behavior for each situation (compressed and restored images). In both datasets increased objective (PSNR) and visual (artifact reduction) quality were the targeted aspects to reduce model degradation due to compression.



(a) (b)

Fig. 4: Validation accuracy on the test dataset compressed and restored by N2N-CAR for different JPEG QFs. (a) Flowers-102. (b) Cub-200.



(a) (b)

Fig. 5: Performance of classification models with respect to the objective quality (PSNR) of compressed and restored test datasets by N2N-CAR. (a) Flowers-102. (b) Cub-200.

From the results presented and the aspects about which we have discussed, it is possible to conclude that there are noticeable advantages of N2N-CAR in the JPEG image classification. Certainly, the results we have presented cannot be extended to all models and datasets. However, in cases where the DL-based CAR can be applied with some efficiency, there are great chances of reducing classifier degradation. It can even be used to treat the training dataset itself in a semi-supervised approach.

One of the limitations of the experiment we performed is that a CAR model is required for each QF, which can be very costly and makes it unfeasible to apply this method in certain situations. Ideally, it would not be necessary to know the QF assigned to each dataset in advance, so a single model trained for a diversity of QFs would be sufficient to restore any dataset. One of the more recent CAR models mentioned in the related works (FBCNN or DPCNN) could solve this issue in the future. Overall, the possibility of obtaining increased accuracy of DL-based image classification models under extreme conditions of signal corruption is evidenced by our results. Furthermore, the implementation of more versatile techniques to address the problem can be developed in future work.

References

[1] G. K. Wallace, "The jpeg still picture compression standard," *IEEE transactions on consumer electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

Fig. 6: Left: image compressed with JPEG with QF equal to 9 (29.52 dB). Center: image restored by N2N-CAR (29.77). Right: original image. Situation where MMAL-Net correctly classifies the compressed image but incorrectly classifies the restored image, despite the quality improvement.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[3] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.

[4] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images," *IEEE transactions on image processing*, vol. 16, no. 5, pp. 1395–1411, 2007.

[5] Y.-Y. Jo, Y. S. Choi, H. W. Park, J. H. Lee, H. Jung, H.-E. Kim, K. Ko, C. W. Lee, H. S. Cha, and Y. Hwangbo, "Impact of image compression on deep learning-based mammogram classification," *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021.

[6] T. Benbarrad, L. Eloutouate, M. Arioua, F. Elouaai, and M. D. Laanaoui, "Impact of image compression on the performance of steel surface defect classification with a cnn," *Journal of Sensor and Actuator Networks*, vol. 10, no. 4, p. 73, 2021.

[7] T. Gandor and J. Nalepa, "First gradually, then suddenly: understanding the impact of image compression on object detection using deep learning," *Sensors*, vol. 22, no. 3, p. 1104, 2022.

[8] E.-H. Yang, H. Amer, and Y. Jiang, "Compression helps deep learning in image classification," *Entropy*, vol. 23, no. 7, p. 881, 2021.

[9] Z. Li, C. De Sa, and A. Sampson, "Optimizing jpeg quantization for classification networks," *arXiv preprint arXiv:2003.02874*, 2020.

[10] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 576–584.

[11] G. Amaranageswarao, S. Deivalakshmi, and S.-B. Ko, "Blind compression artifact reduction using dense parallel convolutional neural network," *Signal Processing: Image Communication*, vol. 89, p. 116009, 2020.

[12] J. Jiang, K. Zhang, and R. Timofte, "Towards flexible blind jpeg artifacts removal," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4997–5006.

[13] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[14] E. Dalsasso, L. Denis, and F. Tupin, "Sar2sar: A semi-supervised despeckling algorithm for sar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4321–4329, 2021.

[15] M. M. Kashyap, A. Tambwekar, K. Manohara, and S. Natarajan, "Speech denoising without clean training data: A noise2noise approach," *arXiv preprint arXiv:2104.03838*, 2021.

[16] J. Yoon and N. I. Cho, "Jpeg artifact reduction based on deformable offset gating network controlled by a variational autoencoder," *IEEE Access*, vol. 11, pp. 30 282–30 291, 2023.

[17] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*. IEEE, 2008, pp. 722–729.

[18] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The caltech-ucsd birds-200-2011 dataset," California Institute of Technology, Tech. Rep. CNS-TR-2011-001, 2011.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[20] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li, and H. Shi, "Escaping the big data paradigm with compact transformers. arxiv 2021," *arXiv preprint arXiv:2104.05704*.

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, and T. Unterthiner, "Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[22] F. Zhang, M. Li, G. Zhai, and Y. Liu, "Multi-branch and multi-scale attention learning for fine-grained visual categorization," in *MultiMedia Modeling: 27th International Conference, MMM 2021, Prague, Czech Republic, June 22–24, 2021, Proceedings, Part I 27*. Springer, 2021, pp. 136–147.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.