JPEG Artifacts Removal via Contrastive Representation Learning:Appendices

Xi Wang, Xueyang Fu*, Yurui Zhu, and Zheng-Jun Zha

University of Science and Technology of China, Hefei, China
{wangxxi, zyr}@mail.ustc.edu.cn, {xyfu, zhazj}@ustc.edu.cn

1 More visual comparison images

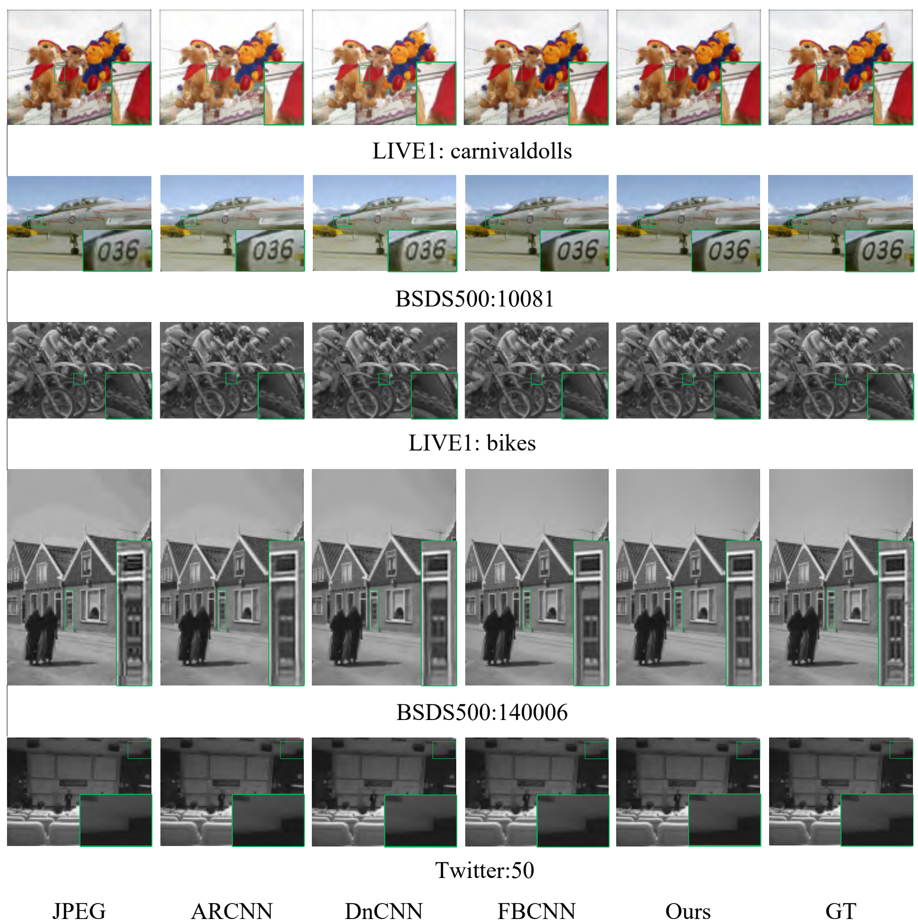


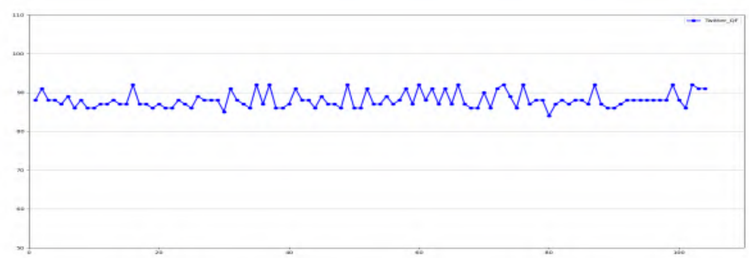
Fig. 1: The images are from synthetic datasets(the first four) and the real-world dataset(last one).

* Corresponding author

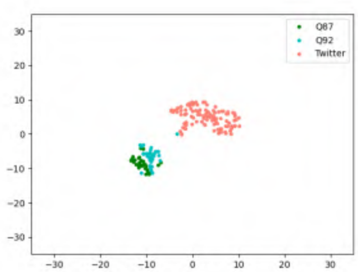
2 Generalization to the real-world dataset

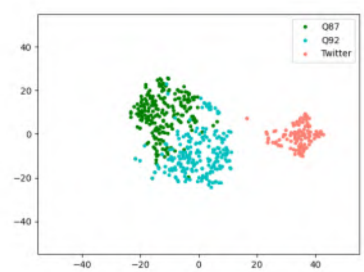
Table 1: Quantitative comparisons of different methods on **Twitter** dataset. PSNR / SSIM / PSNR-B format. The best results are **boldfaced**.

Dataset	ARCNN	DnCNN	DCSC	RNAN	RDN	FBCNN	Ours
Twitter	27.54/0.730/27.49	27.63/0.729/27.54	27.63/0.731/27.43	27.43/0.718/27.42	27.44/0.719/27.39	28.67/0.765/28.67	28.72 /0.769 /28.70



(a)



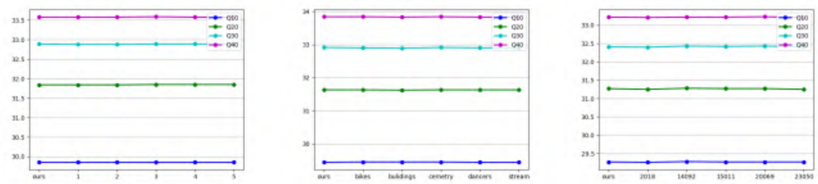


(b) (c)

Fig. 2: Compression quality estimation of the real-world dataset. (a) FBCNN uses a direct predictive quality factor (QF) approach and predicts QF values between 87-93 for the majority of the real-world dataset (Twitter). (b)(c) Our unsupervised approach clearly distinguishes the compression qualities of the Twitter dataset and synthetic datasets with QF of 87 and 92. (b) Twitter and LIVE1 with QF of 87 and 92. (c) Twitter and BSDS500 with QF of 87 and 92.

3 Network architecture of Compression Quality Encoder

The compression quality encoder aims to extract the discriminative compression quality representations. We use a multi-scale feature extraction network as an encoder network. The first scale uses only one convolutional layer. Subsequently,



(a) Classic5 (b) LIVE1 (c) BSDS500

Fig. 3: **PSNR-B** results were achieved by using compression quality representations learned from different image contents.

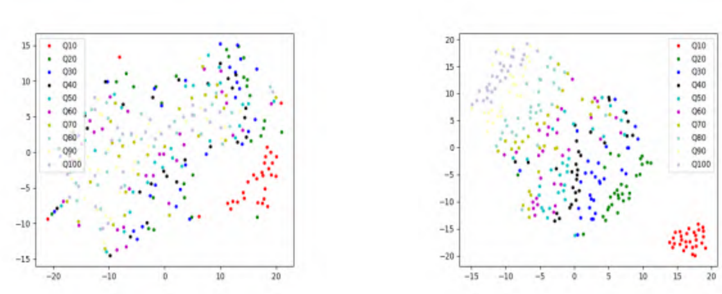
each scale adopts two residual blocks and each block contains two 3×3 convolution layers, two batch normalization layers and one ReLU activation function, both were also downsampled by a factor of 1/2. The number of channels for each scale is set to 64, 64, 128 and 256 respectively. So the output dimensions for each scale are 64×H×W, 64×H/2×W/2, 128×H/4×W/4 and 256×H/8×W/8, respectively. Then a global pooling layer is used to obtain a compression quality representation. The output of each scale of the compression quality encoder is denoted as E_0, E_1, E_2 and E_3 respectively. The front two scale extracts feature, E_0 and E_1, are general features, and the latter two scales extract E_2 and E_3 are discriminative compression quality features.

4 Study of Compression Quality Representations

To demonstrate that the compression quality encoder learns useful information about compression quality, we designed a series of experiments. Specifically, we randomly select an image I with compression quality Q, and then replace both E_2 and E_3 of the entire test dataset with E_2 and E_3 extracted from I. Note that the test dataset has the same compression quality and different image content as I. For the sake of comprehensiveness and convincingness of the experiment, we conducted this experiment on Classis5, LIVE1 and test sets of BSDS500, with five randomly selected images on each dataset of Q10, Q20, Q30 and Q40 for extracting compression quality representations. Because the PSNR-B is specifically designed for image deblocking, we compare the impact of this implementation on the PSNR-B. The results are shown in Fig. 3. The stable performance of our model with this setup demonstrates that our compression quality encoder learns the compression quality information of discriminable images quite well, without interference from the image content.

5 Effect of whether the Compression Quality Encoder is pre-trained or not

We show the visualization of decoder feature maps with or without the compression quality encoder pre-trained weights for LIVE1 with the quality set to



(a) Without. (b) With.

Fig. 4: Visualization of decoder feature maps.

[Q10, Q90] at step 10. It can be seen that with the pre-trained compression quality encoder, the JPEG artifacts removal network produces discriminative compression quality feature representations.

6 Comparison of model parameters

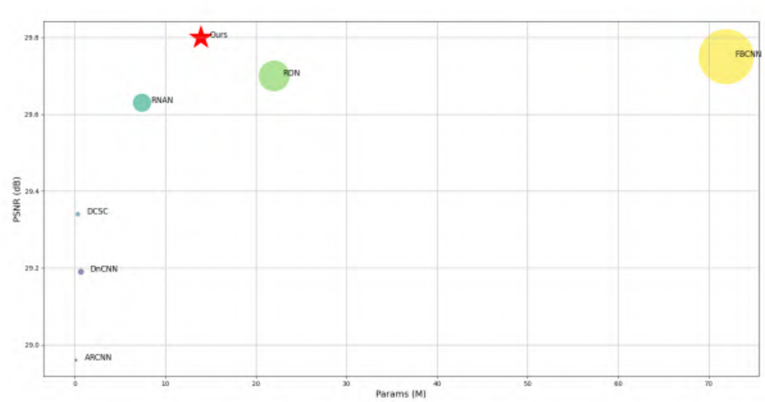


Fig. 5: Testing the PSNR values on Q10 of the LIVE1 dataset, our network achieved the best results with a tolerable number of parameter numbers.