On the Importance of Denoising when Learning to Compress Images

Benoit Brummer
intoPIX, University of Louvain
Mont-Saint-Guibert, Belgium
benoit.brummer@uclouvain.be

Christophe De Vleeschouwer
University of Louvain
Louvain-la-Neuve, Belgium
christophe.devleeschouwer@uclouvain.be

Abstract

Image noise is ubiquitous in photography. However, image noise is not compressible nor desirable, thus attempting to convey the noise in compressed image bitstreams yields sub-par results in both rate and distortion. We propose to explicitly learn the image denoising task when training a codec. Therefore, we leverage the Natural Image Noise Dataset, which offers a wide variety of scenes captured with various ISO numbers, leading to different noise levels, including insignificant ones. Given this training set, we supervise the codec with noisy-clean image pairs, and show that a single model trained based on a mixture of images with variable noise levels appears to yield best-in-class results with both noisy and clean images, achieving better rate-distortion than a compression-only model or even than a pair of denoising-then-compression models with almost one order of magnitude fewer GMac operations.

1. Introduction

Image sensors capture noise along with useful image information. This noise increases with the camera's ISO sensitivity setting, but noise is virtually always present to some extent and it is both incompressible and undesirable. Lossy image compressors inherently perform some image denoising because removing random noise is often the most effective way to reduce entropy in a signal, but without proper training (or algorithm design) the resulting image size is still inflated and the results look sub-par, as shown in Figure 1 (and also attested by Figure S1 in Supplementary Material, and by our experiments in Figure 3a). This increase in bitrate is readily observable in both conventional and learned codecs. A learned lossy image compression scheme is trained by forwarding the image through an autoencoder (AE) and backpropagating from the loss, whose components are the bitrate and the distortion [5]. The bitrate is computed from an optimized, i.e. trained, cumulative distribution function, and the distortion quantifies the difference between the output of the autoencoder and the input image,

typically by computing the mean square error (MSE).

Any image compression scheme can attain better rate-distortion by having the noise removed first. An image denoiser can typically be trained to reconstruct clean images from noisy inputs, using a dataset of paired images where static scenes are captured using a progressively faster shutter speed [8].

In this work, we consider joint compression and denoising. Adding a denoising functionality essentially comes down to feeding the network with a potentially noisy image and comparing its output with a clean image that may have a better quality than what was initially input to the network. The goal is to generate an image that is of higher quality than the one used as input, while decreasing the necessary bitrate close to that of a clean image. Meanwhile, there is no added complexity because the inference process and the network architecture remain unchanged. The network is trained with both noisy images and some clean images as input such that it removes noise while retaining the ability to compress clean input images efficiently.

Our experiments analyzes the impact of image noise on the rate-distortion of different standard and learned compression methods, the benefit of performing denoising prior to compression, and denoising while compressing. Our original supervision strategy, introduced to promote the reconstruction of clean images when the learned codec is fed with noisy ones, appears to be effective. The resulting joint denoising and compression models perform properly on clean images as well as noisy ones, effectively replacing a standard compression models for general-purpose image compression and substantially improving the rate-distortion on noisy images. As illustrated in the second line of Figure 1 (comparison between second and third columns), it is shown to significantly improve rate-distortion performance (using non-noisy images as ground-truth) compared to relying on the implicit noise removal induced by the conventional adoption of a perception-based loss function during training. It also reaches slightly better rate-distortion than a computationally heavy two-step procedure, involving one AE for denoising, followed by another AE-based network for compression.

Figure 1: Visualization of a clean (top) and noisy (bottom) test image from NIND [8]. From left to right: (i) ground-truth/noisy input, (ii) compression autoencoder trained with standard supervision [9], relying on the adoption of a perception-based loss function to mitigate the impact of noise, (iii) our proposed joint denoising and compression model trained with Natural Image Noise Removal supervision using both clean and low noise images ("JDC-Cn.8").

This paper is organized as follows: Section 2 summarizes the work on which this paper builds. The main concepts behind our joint denoising and compression supervision are introduced in Section 3. The implementation details are given in Section 4 followed by the results, and Section 5 summarizes the impact of the present work.

2. Background

Learned Lossy Image Compression is typically based on the seminal work of Johannes Ballé et al. [5]; a convolutional autoencoder [17] with generalized divisive normalization (GDN) [4], and an entropy model which is jointly optimized to capture the latent distribution. This model has been extended with a parametrized hyperprior [6] or with a competition between multiple priors [9], which allows for manipulating an image-dependent latent distribution. Our experiments build onto the initial architecture from [5] completed with multiple sets of latent distributions learned in [9].

Image Noise occurs as the sensitivity of an image sensor is increased to make up for non-ideal lighting conditions. When insufficient light is provided or the dynamic range is too wide, the ISO and/or shutter speed settings are increased accordingly. Pixels take on random, less accurate values, and fewer detail is visible as a result. Different image restoration techniques have been developed to tackle image denoising, including Wavelet [22, 13] and non-local means based methods [11], BM3D [15], and recent deep-learning based methods [8, 14, 18, 12]. Image noise does not reflect the physical components of the observed scene. It is a consequence of the imperfect acquisition process, and thereby should be ignored when possible. Hence, **targeting the reconstruction of a denoised image is the proper way to proceed to get a faithful representation of reality** (even if it implies to not perfectly render the captured signal).

The **Natural Image Noise Dataset** (NIND) [8] and Smartphone Image Denoising Dataset (SIDD) [3] provide sets of clean–noisy image pairs which are appropriate to train a denoising neural network. NIND is made of multiple pictures of many static scenes captured on a tripod to ensure spatial consistency; the clean ground-truth images are taken

in ideal conditions with a camera's base ISO sensitivity to capture as much light as is necessary and to obtain the best possible quality, and matching versions of the same scene are captured with a variety of increasing shutter speed and ISO settings, which result in increased noise and lower image quality. These noisy images are typically fed as the input of the denoising neural network while training it to reconstruct the scene as if it were taken in ideal conditions.

Denoising and Compression have been studied as a combined problem in the wavelet domain [13, 10], and more recently the idea of learning denoising and compression jointly with neural networks was approached by Testolina et al. [23] in the context of the JPEG AI codec. The decoder in [23] is extended such that it consists of twice as many layers, and Poissonian-Gaussian noise is applied to the input training data. This approach delegates the denoising task to the decoder, in line with the JPEG AI requirements of using a universal encoder and specialized decoders. However, as shown in our experimental section, this architecture results in no appreciable bitrate reduction because it is only the encoder that can ensure that incompressible noise does not reach the bitstream. Moreover, training a denoiser with synthetic noise tends to produce a poor model on real data [8, 20]. Testolina et al. introduce a promising joint denoising and compression (JDC) scheme but the resulting rate-distortion of their model falls short of that obtained using our proposed supervision based on pairwise naturally noisy / clean images.

3. Jointly Learned Denoising and Compression

An effective denoising network can be trained to reconstruct clean images given noisy images as input, using a dataset of paired noisy–clean images such as NIND [8] and SIDD [3]. We propose to adopt a similar principle to train an autoencoder originally designed for image compression.

A joint denoising and compression autoencoder [9] is trained to generate a clean image from either a matching noisy image in a paired dataset or from the same clean image. The aim is to obtain a decoded image whose quality is potentially higher than that of the input image, while saving the space that would otherwise be wasted in encoding noise. Different methods are proposed to train such a joint denoising and compression model using Natural Image Noise Removal (NINR) supervision. They are described in Section 3.1 and Figure 2 illustrates the general training process.

3.1. Our Proposed NIN Supervision Strategies

Four different strategies are envisioned and compared to implement our novel Natural Image Noise Removal (NINR) supervision paradigm. They are listed in Table 1 and described below.

Noisy Pairs (JDC-N) The simplest joint denoising and compression implementation consists of training with all noisy–clean image pairs available in the dataset(s).

Clean and Noisy Pairs (JDC-CN) This method considers some clean–clean image pairs, in addition to the noisy–clean image pairs, to ensure that the network's performance does not degrade when the input images contain no unwanted noise. The dataset of clean images can be selected as a set of images which has been assessed and promoted by human reviewers, such as the Wikimedia Commons Featured Pictures [1], then further refined by eliminating images whose metadata indicates a high ISO value in order to ensure the absence of noise.

Clean and Low-noise Pairs (JDC-Cn) To specialize the model to the most frequent input image noise levels, we have also considered placing a threshold on the training data noise level. Our experiments reveal that it is beneficial to filter out the most noisy input training images because the overall rate-distortion degrades when the network is trained to perform more extreme denoising. Such extreme denoising would require added complexity on the encoder and, although possible, extreme denoising is outside the scope of a combined denoiser whose aim is to improve rate-distortion by removing noise that is inherently present in most photographs rather than learning to see in the dark, as proposed in [14]. The paired image noise dataset is analyzed prior to training such that the multi-scale structural similarity (MS-SSIM) [24] score between each noisy crop and its clean ground-truth is stored in a file, and the training dataset can be initialized such that all training crops exceed a set quality threshold. The effect of different noise thresholds is analyzed in the ablation study.

Building Pairs from a Universal Denoiser (JDC-UD) A fourth training method consists of running a pre-trained blind denoising model [21, 8] on all the training data to generate the ground-truth images, and computing the training loss between the input images and the denoised images. This method effectively performs knowledge distillation [16] from a powerful universal denoising network to the joint denoising and compression network. All input images are considered noisy and the training dataset is virtually limitless because the ground-truth images are generated (in advance), thus entire image datasets are used without filtering.

4. Experiments

4.1. Practical Implementation Details

These experiments are based on the PyTorch implementation of the autoencoder base codec introduced in [9]. Source code is provided as Supplementary Material and available on `https://github.com/trougnouf/compression`. The training loss of the compression autoencoder is computed as Loss $=$ bitrate$(\hat{x}) + \lambda \times \mathrm{MSE}(\hat{x}, x)$, where $\hat{x}$ is the decoded image and x is the clean ground-truth which, as explained in Section 3, may differ from the input image, and λ balances
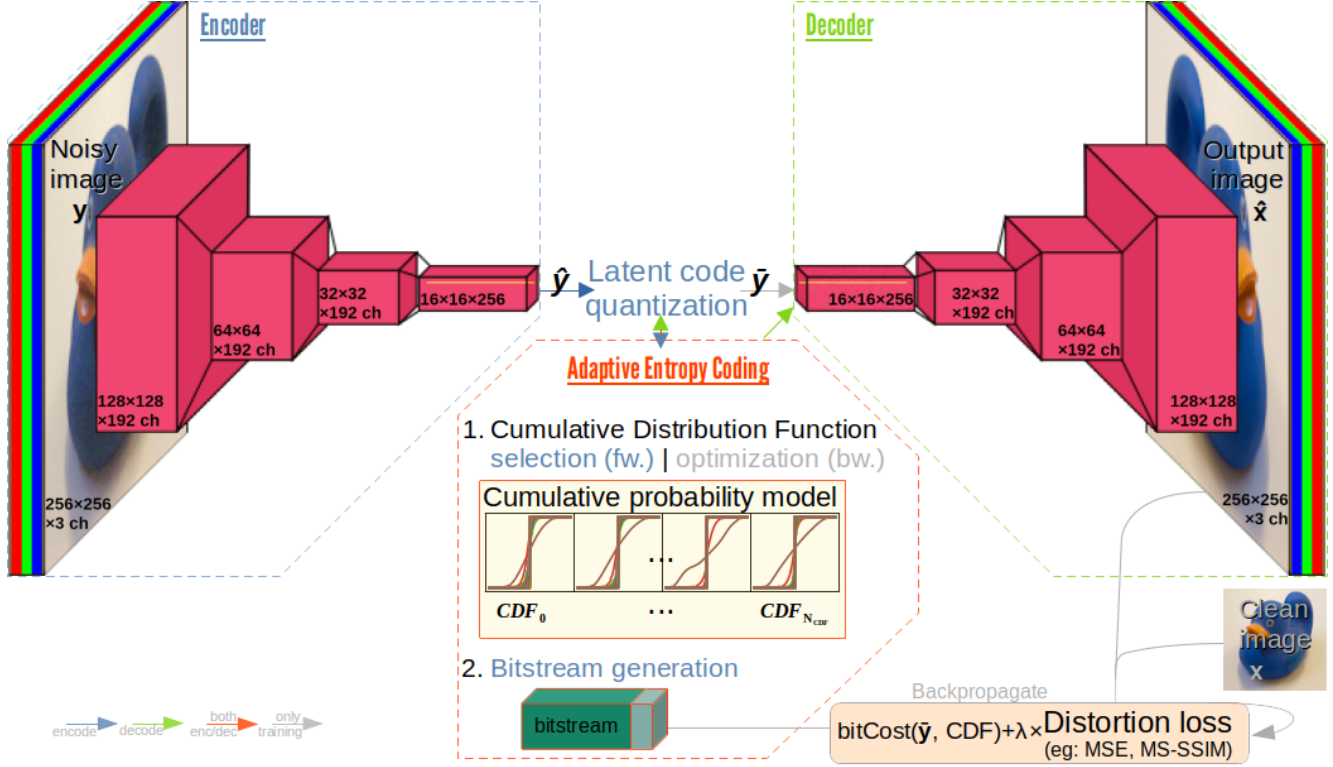
Figure 2: Denoising and compression joint training: the distortion loss is computed between the reconstructed image $\hat{x}$ and a clean image x. The input image y may be noisy. The network [9] is made of four (transposed) convolutions with stride of 2 and kernel size of 5, each followed by a GDN activation [4] except for the last (transposed) convolution. (Best viewed in color.)

Table 1: Data pairs considered in this paper to train a joint denoising and compression model. JDC-Cn is also referred to with its training noise threshold (e.g. JDC-Cn.8 is trained with MS-SSIM ≥ 0.8, see the text for details).

Method	Training input	Expected output
JDC-N	Noisy image from clean–noisy paired dataset [8]	Clean ground-truth
JDC-CN	Noisy image from clean–noisy paired dataset, clean image from high quality dataset [1]	Clean ground-truth, clean input
JDC-Cn	Weakly noisy image from clean–noisy paired dataset, clean image from a high quality dataset	Clean ground-truth, clean input
JDC-UD	Arbitrary input image [1, 8]	Provided by a universal denoiser
Testolina [23]	Clean image from high quality dataset + artificial noise	Clean input

the rate/distortion trade-off of the model. The combined denoising and compression autoencoder [9] is trained with batches of four noisy images from NIND [8] and one clean image from the Wikimedia Commons Featured Pictures [1] dataset whose ISO value does not exceed 200, with a crop size of 256 as is typically done in the learned compression literature [9, 5, 6]. The pre-trained "**universal denoiser**" used to train the JDC-UD model is the U-Net-like blind denoising model published with NIND, which was updated such that its training includes clean–clean image pairs. The

CLIC professional test set [2] is used to assess the models on other clean images.

The JDC model defined in Testolina et al. (2021) [23] is trained entirely as described with Poissonian-Gaussian artificial noise [19] (with noise parameters $a = 0.2^2, b = 0.04^2$), the encoder from Balle et al. (2018) [6], and their proposed decoder which has twice as many layers as the one recommended by Ballé et al. An additional JDC-Cn model is trained with the larger decoder proposed by Testolina et al. in order to assess their proposed network architecture

separately from the training method.

Most models are trained for six million steps with $\lambda = 4096$ to yield the highest bitrate, then the λ value is halved and training continues for three million steps for each λ value all the way down to $\lambda = 256$, like in [9]. Both the JDC-Cn.8-Tdec method that is matched with the decoder defined by Testolina et al. and the JDC-N model trained with only noisy–clean image pairs have had an additional model trained with $\lambda = 8192$ in order to reach the other methods' highest bitrate. Likewise, the whole method defined by Testolina et al. [23] was trained with up to $\lambda = 16384$.

The standard codecs comparisons are made by encoding images using GraphicsMagick 1.3.37 (JPEG), the JPEG XL encoder v0.6.1, and the BPG Image Encoder version 0.9.8. The "**standard autoencoder**" is that defined [9].

4.2. Results
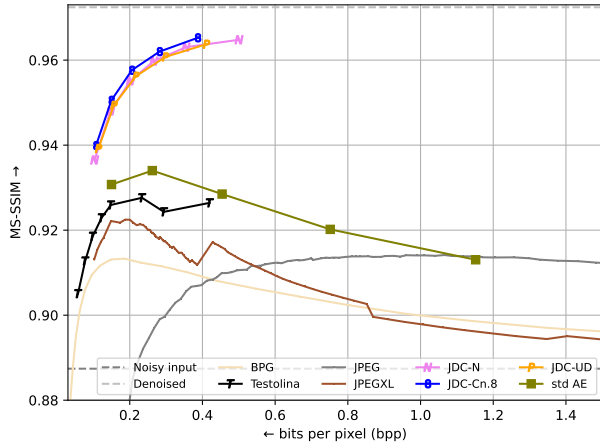
4.2.1 On the Importance of Denoising.

The first experiment measures the impact of denoising prior to compression with different compression methods. Figure 3a plots the rate-distortion curves obtained without specific denoising of the input images, for a variety of codecs. We observe that for all codecs compression is an effective denoising method at the lowest bitrate. However, at reasonable bitrates, all conventional codecs (learned or not) tend to reproduce the noise. This is in contrast with our proposed joint denoising and compression paradigm, which continuously increases quality when the bitrate increases. Denoising before compression might be considered to solve the quality issue when using conventional codecs. As shown in Figure 3b, this bridges (most of) the quality gap compared to our proposed JDC method, but at the cost of a significantly increased complexity (see Section 4.2.3).
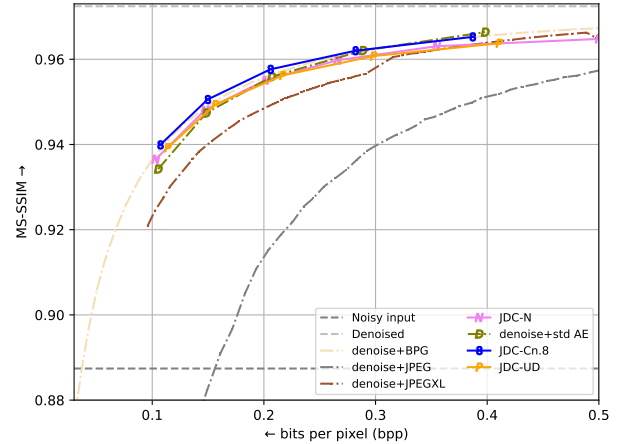
4.2.2 Our Joint Denoising and Compression Scheme

Figure 3 also introduces the proposed joint denoising and compression models, JDC-CN (no noise threshold) and JDC-Cn.8 (training noise limited to MS-SSIM ≥ 0.8). These models are trained like the dedicated denoising model in that the input training batch is made of four noisy images and one clean image, and the model is tasked with reconstructing the clean ground-truths. This single JDC model generally achieves better rate-distortion than a duo of denoising then compression neural networks (except at high bitrate), while using significantly less computational overhead.

The best results shown are obtained with the "JDC-Cn.8" model which is trained with paired images whose input noise is limited to MS-SSIM ≥ 0.8 as well as with unpaired clean images to promote generalization. T is the method trained with artificial noise described by Testolina et al. [23], which performs worse than a duo of models as is also shown in their results.

(a) Compression when taking the original (and thus noisy) image as input.

(b) Applying CNN-based denoising before compression (except for our JDC which inputs the originally captured image).

Figure 3: Lossy compression of noisy (MS-SSIM $\in [0.7, 1.0)$) test images from NIND [8] with respect to their matching clean ground-truth.
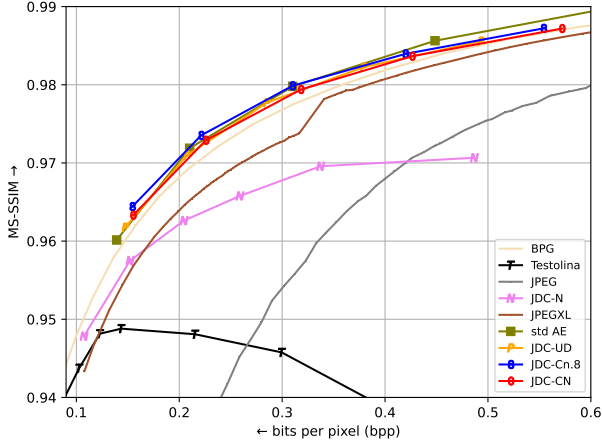**(a)** Original (noisy) images are provided as input. Standard methods (JPEG, JPEG XL, BPG, and a standard compression autoencoder [9]) perform some implicit denoising at low bitrates, but image quality degrades as the bitrate increases since the noisy signal is reconstructed.
(b) A universal denoiser is applied before presenting the image to the standard method. This greatly improves rate-distortion but adds of an order of magnitude of complexity. Our joint denoising and compression (JDC) autoencoder with Natural Image Noise Removal supervision allows noise-free reconstruction without prior denoising.
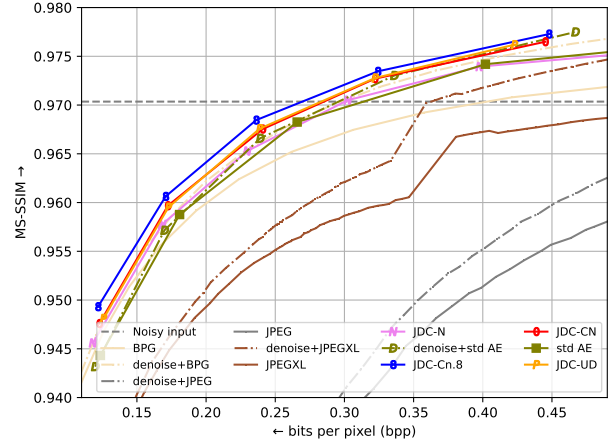
4.2.3 Computational Complexity

Computational cost is measured in terms of billion multiply-accumulate operations (GMac) and runtime on an AMD Threadripper 3960X CPU. The dedicated denoising U-Net performs 812 billion multiply-accumulate operations (GMac) per megapixel (MP) in 65.8 sec./MP, whereas the JDC model's compression encoder performs 92.8 GMAC/MP [9] in 2.9 sec./MP. The dual model approach (denoising then compression) thus performs a total of 904.8 GMac/MP, whereas a single joint denoising and compression model operates with 10.3% of that computational complexity.
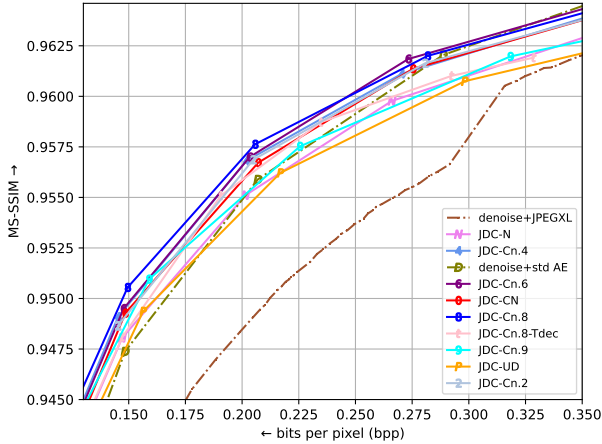

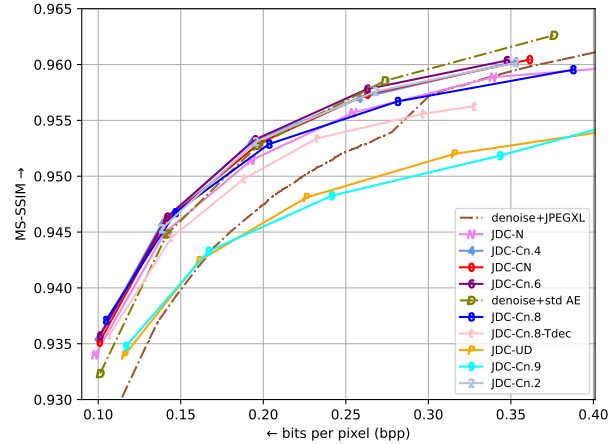(a) High quality (unpaired) test images

(b) Test images with (MS-SSIM $\in [0.95, 1.0)$, MS-SSIM$_\mu = 0.97$)

Figure 4: MS-SSIM rate-distortion curve of different compression methods on (a) high quality images from the CLIC pro test set [2] and (b) nearly noiseless test images from NIND [8].

(a) Test images with MS-SSIM $\in [0.7, 1.0)$, MS-SSIM$_\mu = 0.89$

(b) Test images with MS-SSIM $\in [0.5, 1.0)$, MS-SSIM$_\mu = 0.84$

Figure 5: Lossy compression of noisy test images from NIND [8] with different joint denoising and compression methods including JDC-Cn trained with different MS-SSIM thresholds, JDC-CN trained with no such threshold, JDC-N trained with no clean images, JDC-UD trained with knowledge distillation from a universal denoiser, and JDC-Cn.8 trained with the larger decoder defined by Testolina et al. [23]. **(a)** The JDC-Cn methods tend to perform well, especially when the quality threshold is set between 0.6 and 0.8, but the rate-distortion worsens when the training quality threshold increases to 0.9 (i.e. little noise is seen by the model during training). The JDC-UD model yields similarly lower performance, and so does the larger decoder despite using the same training scheme as JDC-Cn.8. **(b)** The noise level is more extreme than what is likely to occur in photographs. This further shows that compressing with more noise than is ever seen during training (e.g. JDC-Cn.9) yields worse rate-distortion. The model trained without clean image pairs (JDC-N) does not perform as well even when the test images are noisier.

Figure 6: Visualization of a clean (top) and noisy (bottom) test image from NIND [8] encoded with a target bitrate of 0.13 bpp using different trained compression autoencoders. From left to right: ground-truth/noisy input, standard autoencoder [9], autoencoder from Testolina et al. (2021) [23] (trained on artificial noise; increasing bitrate did not yield quality improvements on the ground-truth), joint model trained with knowledge distillation from a universal denoiser (JDC-UD), and joint model trained with both clean and low noise images (JDC-Cn.8). Images are best visualized after zooming in on a screen.

4.2.4 Handling Clean Images

JDC-C models are trained with both noisy–clean and clean-clean paired images in order to better generalize and maintain good rate-distortion on clean images. Figure 4a and Figure 7 show the behavior of different JDC training strategies when compressing clean images. JDC models trained with some clean input images or with the JDC-UD knowledge distillation technique yield a rate-distortion similar to the model trained for compression only, even when no minimum training noise threshold is set, thus incorporating clean images in the training data (JDC-CN) reinstates a good rate-distortion. Limiting the input noise to MS-SSIM ≥ 0.8 (JDC-Cn.8) further improves rate-distortion such that it is slightly better with a JDC model than a standard model at low bitrates, and slightly worse at high bitrates due to the perception-distortion tradeoff [7] where only reconstruction fidelity matters. The JDC-N model trained with only noisy–clean image pairs performs significantly worse on clean images, and the model trained with artificial noise ("T") performs worst.

Figure 4b shows a common use-case where the amount of noise is low (MS-SSIM $\in [0.95, 1)$). Prior denoising still improves rate-distortion, and joint denoising and compression methods yield the most significant rate-distortion benefits. All compression methods benefit from prior or joint denoising even when the level of noise is minor. Traditional compression schemes benefit the most from prior denoising, and joint denoising outperforms prior denoising in learned methods.

4.2.5 Ablation Study

The effect of different training noise thresholds is analyzed when compressing noisy images. In Figure Figure 5a the test noise is limited to MS-SSIM $\in [0.7, 1)$ which is qualitatively fairly noisy, as shown in Figure 1. None of the methods perform significantly better or worse than the denoise and compress duo of models. It is worth noting that the three worst JDC training schemes are the knowledge distillation JDC-UD model, the model trained with a quality threshold of MS-SSIM ≥ 0.9, and the decoder defined by Testolina et al. which contains twice as many layers. The JDC-Cn

models trained with an MS-SSIM threshold of 0.6 and 0.8 yield the best rate-distortion. A visualization of the different denoising and compression methods at low bitrates is shown as Figure 6.

In Figure 5b, the testing noise is increased to MS-SSIM $\in [0.5, 1)$, showing how the models behave under extreme input noise. The results are largely the same; the model trained with MS-SSIM ≥ 0.9 struggles even more due to the increased noise and its performance is close to that of the JDC-UD method, the model trained with MS-SSIM ≥ 0.8 does not perform as well whereas the model trained with MS-SSIM ≥ 0.6 is still competitive, and it remains beneficial to train with clean image pairs as well.

5. Conclusion

Denoising images improves rate-distortion whenever there is noise present, regardless of the compression method. Denoising can be performed prior to compression using a dedicated denoiser (as is typically done in professional image development workflow) with no adaptation to the compression scheme. A joint model that is trained to perform denoising and compression simultaneously yields further improvements in rate-distortion. As a result, a joint denoising and compression model performs 8.9 times fewer GMAC operations than a U-Net denoiser followed by a compression encoder. Since the JDC model only differs from standard learned compression models by the adopted supervision strategy, it can be implemented using any of the compression architectures available in the literature (such as [5, 6, 9]). Our proposed Natural Image Noise Removal supervision strategy thus provides a fundamental and generic contribution that is expected to become popular in future works related to learned compression.

In practice, joint denoising and compression models may be trained using a dataset of noisy–clean image pairs with natural noise, such as NIND [8] and SIDD [3]. Performance is improved by setting a quality threshold on the training images, such as MS-SSIM ≥ 0.8 or MS-SSIM ≥ 0.6 depending on the maximum expected noise. The rate-distortion curve is preserved on clean images and improved in any case by incorporating clean-clean image pairs in the training data.

An alternative method consists of performing knowledge distillation [16] by using the output of a dedicated denoiser as ground-truth images during training. This has the benefit of allowing a virtually limitless training dataset because a paired dataset is no longer required, but it requires preprocessing of the training images and results in a slightly worse rate-distortion.

6. Acknowledgements

Figure 7: Visualization of a clean test image compressed using different methods with a target bitrate of 0.23 bpp. Standard methods (JPEG, JPEG XL, BPG) tend to produce block artifacts. The method defined by Testolina et al. [23] produces oversmoothed results on clean images. Training with only noisy images can produce the same level of quality at the cost of increased bitrate. Other methods—training with only clean images, training with both clean and noisy images, and knowledge distillation from a powerful denoiser—perform well on clean images. The input image has been processed in the darktable software with (among other methods) non-local means [11] profiled denoising.

References

[1] Featured pictures on wikimedia commons. `https://commons.wikimedia.org/wiki/Category:Featured_pictures_on_Wikimedia_Commons`. Accessed: 2020-04-03.

[2] Challenge on learned image compression. `http://clic.compression.cc/2021/tasks`, 2020.

[3] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, June 2018.

[4] Johannes Ballé. Efficient nonlinear transforms for lossy image compression. In *2018 Picture Coding Symposium (PCS)*, 2018.

[5] Johannes Ballé, Valero Laparra, and Eero Simoncelli. End-to-end optimized image compression. In *5th International Conference on Learning Representations, ICLR 2017*, Toulon, France, April 2017.

[6] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. In *International Conference on Learning Representations*, 2018.

[7] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[8] Benoit Brummer and Christophe De Vleeschouwer. Natural image noise dataset. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1777–1784, June 2019.

[9] Benoit Brummer and Christophe De Vleeschouwer. End-to-end optimized image compression with competition of prior distributions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1890–1894, June 2021.

[10] V. Bruni and D. Vitulano. Combined image compression and denoising using wavelets. *Image Commun.*, 22(1):86–101, jan 2007.

[11] A. Buades, B. Coll, and J. . Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65 vol. 2, June 2005.

[12] Haoming Cai, Jingwen He, Qiao Xiao Yu, and Chao Dong. Toward interactive modulation for photo-realistic image restoration. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 294–303, 2021.

[13] S.G. Chang, B. Yu, and M. Vetterli. Image denoising via lossy compression and wavelet thresholding. In *Proceedings of International Conference on Image Processing*, volume 1, pages 604–607 vol.1, 1997.

[14] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3291–3300, 06 2018.

[15] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising with block-matching and 3d filtering. *Proceedings of SPIE - The International Society for Optical Engineering*, 6064:354–365, 02 2006.

[16] Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*, 2015.

[17] Mark A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2):233–243, 1991.

[18] Yao Li, Xueyang Fu, and Zheng-Jun Zha. Cross-patch graph convolutional network for image denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4651–4660, October 2021.

[19] Joseph Meola, Michael T. Eismann, Randolph L. Moses, and Joshua N. Ash. Modeling and estimation of signal-dependent noise in hyperspectral imagery. *Appl. Opt.*, 50(21):3829–3846, Jul 2011.

[20] Tobias Plötz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR)*, pages 2750–2759, 2017.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[22] Dr. Sachin Ruikar and Dharmpal Doye. Image denoising using wavelet transform. In *2010 International Conference on Mechanical and Electrical Technology*, pages 509–515, 09 2010.

[23] Michela Testolina, Evgeniy Upenik, and Touradj Ebrahimi. Towards image denoising in the latent space of learning-based compression. In Andrew G. Tescher and Touradj Ebrahimi, editors, *Applications of Digital Image Processing XLIV*, volume 11842, pages 412 – 422. International Society for Optics and Photonics, SPIE, 2021.

[24] Zhou Wang, Eero P. Simoncelli, and Alan C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, Nov 2003.