High Perceptual Quality Image Denoising via Neural Compression

Nam Nguyen
Oregon State University, Oregon, United States
{nguynam4}@oregonstate.edu

Abstract

Image denoising aims to recover a clean image from a single noisy observation, yet modern deep denoisers typically rely on paired clean-noisy supervision that is often unavailable in practical imaging pipelines. Motivated by the information-theoretic connection between lossy compression and denoising, we revisit compression-based denoising using learned neural codecs trained directly on noisy inputs, where the rate bottleneck implicitly regularizes the reconstruction. However, optimizing only rate-distortion commonly yields perceptually over-smoothed outputs, especially at low bitrates. To address this, we propose a perception-enhanced neural compression denoiser *that augments the rate-distortion objective with an adversarial distribution-matching term implemented via a Wasserstein GAN (WGAN). Using an unpaired clean-image dataset, the discriminator encourages reconstructions to lie on the clean-image manifold while preserving the compression-induced denoising effect. Experiments on natural images (KODAK-24) with synthetic Gaussian noise demonstrate improved perceptual quality compared to distortion-driven baselines, and additional evaluations on fluorescence microscopy (Mouse Nuclei) and real smartphone photographs (SIDD) show that the method generalizes beyond controlled synthetic settings. Overall, our framework provides a simple and effective way to obtain perceptually faithful denoising from a single noisy image without requiring paired supervision.*

1. Introduction

Image denoising is a fundamental problem in low-level vision, aiming to recover a clean image from a single noisy observation. While classical approaches relied on handcrafted priors and sparse signal representations (e.g., filtering, wavelets, and sparse coding) [20–22, 32, 36, 38, 49], modern deep denoisers learn rich image statistics directly from data and achieve strong benchmark performance [28, 30, 54, 56, 58]. However, most state-of-the-art methods require large-scale paired clean-noisy supervision, which is often unavailable in practical imaging pipelines such as microscopy, remote sensing, and astronomical imaging [7, 9, 39, 51].

To relax the need for paired data, self-supervised denoising methods train using only noisy observations [8, 24, 26, 29, 41]. While these approaches broaden applicability, they can underperform fully supervised denoisers and may require stronger assumptions (e.g., multiple noisy captures or noise independence) and more delicate training strategies [8, 24, 26, 29]. These limitations motivate alternative paradigms that can denoise from a single noisy image without paired supervision.

A promising and principled direction stems from lossy compression. When compressing a noisy signal at an appropriate operating point, the rate bottleneck can suppress noise while preserving the underlying structure, effectively acting as an implicit prior. This connection between compression and denoising has been studied in information theory and universal denoising, where rate constraints and description-length principles yield denoising behavior under broad conditions [19, 48]. Historically, however, practical compression-based denoising has been most effective for simplified source models, limiting its impact on high-dimensional natural images.

Recent advances in neural image compression make this connection practically actionable. Learned codecs optimize nonlinear analysis/synthesis transforms under explicit rate constraints and have substantially improved rate-distortion performance over traditional hand-engineered designs [4–6, 33, 52]. When trained directly on noisy inputs, the compression bottleneck prevents trivial identity mappings and can induce denoising even without clean supervision, enabling denoising from a single noisy observation as demonstrated by recent compression-based denoisers [53].

However, distortion-driven compression objectives alone are often insufficient for perceptually satisfying denoising. Pixel-wise criteria such as MSE can yield over-smoothed reconstructions, especially at low bitrates where aggressive bottlenecking removes fine details. To address this, we propose a *perception-enhanced neural compression denoiser* that augments the standard rate-distortion objective with an adversarial distribution-matching term implemented via a Wasserstein GAN (WGAN) [3]. Using an unpaired clean-

image dataset, the discriminator encourages reconstructions to align with the clean-image manifold, with the Lipschitz constraint enforced through a gradient penalty [23]. This formulation preserves the compression-induced denoising effect while improving perceptual realism.

We validate the proposed framework on both synthetic and real-world denoising benchmarks. Experiments on KODAK-24 with additive Gaussian noise show improved perceptual quality over distortion-only compression baselines and competitive performance among unsupervised methods [16, 46, 53]. Additional evaluations on fluorescence microscopy (Mouse Nuclei) and real smartphone photographs (SIDD) demonstrate robustness beyond controlled synthetic settings [1, 13].

2. Related Work

Classical and Learning-Based Image Denoising. Early image denoising methods relied on handcrafted priors and signal representations, including linear filtering, wavelet thresholding, sparse coding, and Markov random field models [20, 21, 32, 36, 38, 49]. These approaches established fundamental principles for noise suppression but struggled to generalize across diverse image statistics. Deep learning has since revolutionized denoising by learning rich image priors from data, with CNN- and transformer-based models achieving state-of-the-art distortion performance under supervised training [28, 30, 54, 56, 58]. Despite their effectiveness, such methods typically require large paired clean-noisy datasets, limiting their applicability in many real-world imaging scenarios.

Self-Supervised and Unsupervised Denoising. To reduce reliance on paired supervision, self-supervised denoising methods train using only noisy observations. Noise2Noise [29] exploits multiple independent noisy realizations, while Noise2Void [26] and Noise2Self [8] leverage blind-spot architectures to avoid learning the identity mapping. More recently, score-based approaches such as Noise2Score [24] estimate the score of the noisy data distribution for denoising. Although these methods broaden applicability, they often underperform supervised counterparts and may require restrictive assumptions or careful architectural design. In contrast, our approach denoises from a single noisy observation without blind-spot constraints or repeated measurements.

Compression-Based Denoising and Information-Theoretic Foundations. The relationship between compression and denoising has deep roots in information theory. Early work showed that description-length and rate constraints can induce denoising behavior, particularly for finite-alphabet and stationary sources [19, 48]. Practical compression-based denoising methods include MDL-based wavelet selection [40], the Occam filter [34], and adaptive wavelet thresholding schemes [14, 15]. While these methods are theoretically grounded, their effectiveness on high-dimensional natural images is limited.

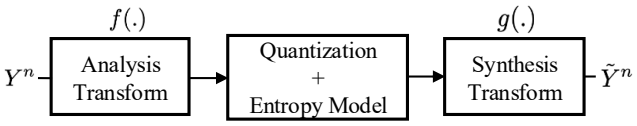


Figure 1. Neural compression architecture used for denoising.

Recent work revisits compression-based denoising using learned neural codecs. Zafari *et al.* [53] demonstrate that neural image compression models trained on noisy inputs can act as effective denoisers when the rate-distortion trade-off is properly tuned. Unlike traditional self-supervised methods, this approach requires only a single noisy observation and no clean supervision. Our work builds on this paradigm by addressing a key limitation: distortion-only optimization leads to perceptually over-smoothed reconstructions.

Neural Image Compression. Neural image compression has emerged as a powerful alternative to traditional codecs, achieving superior rate-distortion performance through end-to-end learning [5, 6, 33]. Subsequent work introduced more expressive entropy models [4], attention and transformer-based architectures [31, 59], and theoretical analyses of neural codecs [10, 35, 45]. While most compression methods focus on efficient representation of clean images, several works jointly address compression and denoising [12, 37, 43, 50]. In contrast, we leverage compression as an implicit regularizer for denoising rather than explicitly optimizing for compression of noisy content.

Perceptual Restoration and Adversarial Learning. Optimizing pixel-wise distortion metrics alone often yields reconstructions that are perceptually suboptimal. Prior work in image restoration has incorporated perceptual losses and adversarial learning to improve visual fidelity, including GAN-based super-resolution and denoising [11, 27, 47]. The Wasserstein GAN framework provides a principled means of distribution matching with improved training stability [3, 23]. Our approach integrates adversarial distribution matching into a neural compression framework, enabling perception-aware denoising from unpaired clean images while preserving the compression-induced denoising effect.

3. Perceptual Image Denoising via Neural Compression

3.1. Neural Compression as a Denoising Mechanism

Despite strong information-theoretic motivation, compression-based denoising has historically seen limited

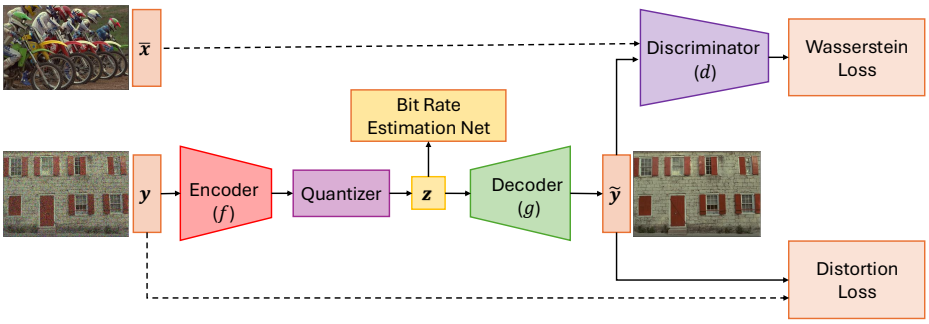


Figure 2. Proposed framework combining a neural compression autoencoder with a WGAN-based discriminator.

success on high-dimensional signals such as natural images, largely due to the restricted expressiveness of classical compression models. Recent advances in neural lossy compression provide a powerful data-driven alternative, enabling highly nonlinear representations that adapt to complex image statistics. Building on this insight, Zafari *et al.* [53] demonstrate that neural compression networks can be repurposed as effective denoisers when trained directly on noisy observations.

As illustrated in Figure 1, a neural compression model maps a noisy input image Y^n to a reconstruction $\tilde{Y}^n$ through learned analysis and synthesis transforms f and g. Importantly, training requires access only to noisy images: the reconstruction loss is computed between the input and its compressed reconstruction. In contrast to conventional denoising networks such as DnCNN [54], the objective is not to explicitly estimate the clean image. Instead, the compression bottleneck implicitly regularizes the mapping, preventing the network from learning a trivial identity function despite the absence of clean supervision.

Following standard neural compression frameworks [6], an entropy model is learned to approximate the distribution of the latent representation. Let $C^m = f(Y^n)$ denote the latent code and $\lfloor C^m \rceil$ its quantized version. The expected rate is defined as

$$R = \mathbb{E}[-\log_2 \mathbb{P}(\lfloor C^m \rceil)]\,,$$

and the reconstructed output is given by $\tilde{Y}^n = g(\lfloor C^m \rceil)$. Training minimizes the standard rate-distortion objective

$$\mathbb{E}\left[\frac{1}{n}\|\tilde{Y}^n - Y^n\|_2^2\right] + \lambda_r R, \tag{1}$$

where λ_r controls the rate-distortion trade-off. When λ_r is appropriately matched to the noise level, the compression bottleneck suppresses noise while preserving salient image structure, yielding effective denoising from a single noisy observation.

4. Perception-Enhanced Neural Compression Denoiser

While neural compression provides a principled and unsupervised mechanism for denoising, objectives based solely on distortion metrics often yield perceptually suboptimal reconstructions. In particular, pixel-wise losses such as mean squared error favor overly smooth outputs, especially in low-bitrate regimes where aggressive bottlenecking removes fine details and textures.

To address this limitation, we introduce a perception-enhanced neural compression framework that augments the rate-distortion objective with a perceptual regularization term. As shown in Figure 2, we integrate a Wasserstein GAN (WGAN) discriminator [3] that encourages reconstructed images to align with the distribution of clean natural images. Crucially, the discriminator is trained using an *unpaired* clean-image dataset, preserving the unsupervised nature of the denoising task.

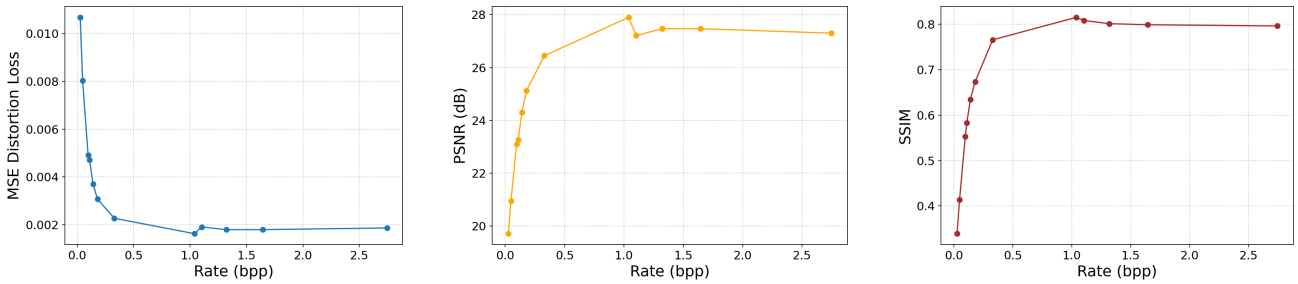
The resulting objective is

$$\mathbb{E}\left[\frac{1}{n}\|\tilde{Y}^n - Y^n\|_2^2\right] + \lambda_r R + \lambda_p\, W_1(p_{\bar{X}}, p_{\tilde{Y}})\,, \tag{2}$$

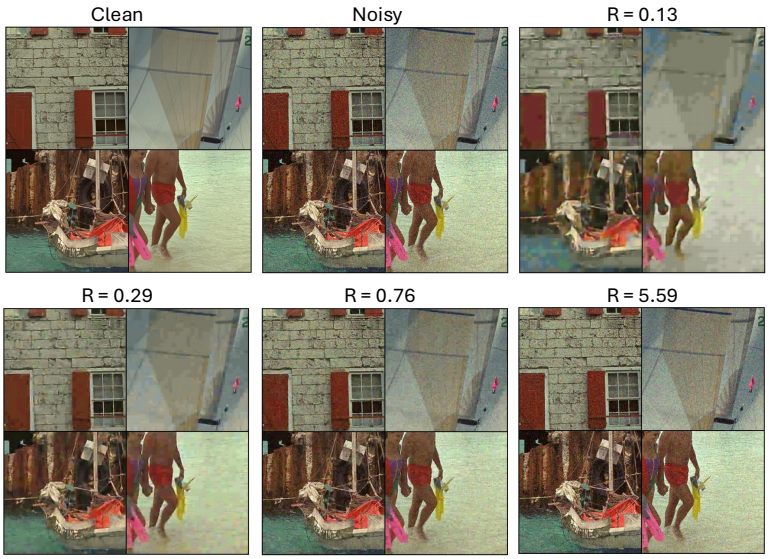
where the Wasserstein-1 term enforces distributional alignment between reconstructed outputs $\tilde{Y}$ and clean images $\bar{X}$. The hyperparameters λ_r and λ_p control the trade-off between rate efficiency, distortion, and perceptual fidelity. This formulation preserves the denoising effect induced by the compression bottleneck while producing reconstructions that are perceptually sharper and more realistic.

5. Experiments

We evaluate our perception-enhanced neural compression denoiser using a stochastic autoencoder architecture composed of an encoder f, quantizer Q, decoder g, and a WGAN discriminator d, as illustrated in Figure 2. Reconstruction distortion is measured using the mean squared error (MSE). Given a noisy observation Y, the reconstructed output is defined as $\tilde{Y} = g(Q(f(Y)))$. The discriminator d is trained



(a) MSE-rate curve. **(b)** PSNR-rate curve. **(c)** SSIM-rate curve.



(d) Reconstruction and denoising examples.

Figure 3. Denoising results on the Kodim01 image from the KODAK-24 dataset corrupted by Gaussian noise $\mathcal{N}(0, \sigma^2)$ with $\sigma = 25$. Increasing the representation rate leads to improved reconstruction fidelity and reduced residual noise.

to align the distribution of reconstructed outputs $p_{\tilde{Y}}$ with that of clean images $p_{\bar{X}}$ through a Wasserstein-1 objective, while the autoencoder and discriminator are optimized alternately following the standard WGAN training protocol [3].

Loss Function. The model is trained by minimizing a composite objective that jointly balances reconstruction distortion, compression rate, and perceptual fidelity:

$$\mathcal{L} = \mathbb{E}\Big[\|Y - \tilde{Y}\|^2\Big] + \lambda_r \log \mathbb{P}(Q(f(Y))) + \lambda_p W_1(p_{\bar{X}}, p_{\tilde{Y}}), \tag{3}$$

where the first term penalizes reconstruction error, the second term promotes compact latent representations by regularizing the compression rate, and the third term enforces perceptual alignment between reconstructed outputs and clean images via the Wasserstein-1 distance. Notably, the clean images $\bar{X}$ are drawn from an unpaired dataset, enabling perception-aware denoising without requiring paired clean–noisy supervision. The hyperparameters λ_r and λ_p control the trade-off between rate efficiency and perceptual quality.

5.1. Results

5.1.1 Natural Images with Synthetic Noise

For training, we use the BSDS500 dataset [2], which consists of 481×321 natural images with rich texture and structural diversity. All methods are evaluated on the KODAK-24 dataset [16], containing 768×512 high-quality color images that are widely adopted for benchmarking image denoising and compression performance.

Figure 3 illustrates the rate-distortion behavior of our method on a representative KODAK image. As the representation rate increases, the model progressively suppresses noise while recovering fine textures and edges, resulting in smooth yet visually plausible reconstructions. Compared to low-rate regimes, higher-rate operating points better preserve structural details while avoiding over-smoothing artifacts, highlighting the effectiveness of the learned rate-distortion-perception trade-off.

Table 1 reports quantitative denoising results on KODAK-24 [16] under additive Gaussian noise with $\sigma = 25$. As

Category	Method	PSNR (dB)↑	SSIM↑	PI↓
Non-learning	JPEG-2K [42]	26.4408	0.7357	7.4794
	BM3D [17]	31.8757	0.8687	2.6503
Supervised	N2C [55]	<u>32.2114</u>	<u>0.8865</u>	2.5446
	N2N [29]	**32.2723**	**0.8877**	2.5439
Unsupervised	DeCompress [53]	27.8315	0.7519	2.7979
	OTDenoising [46]	31.2893	0.8677	**2.0095**
	Ours	28.0435	0.8035	<u>2.1670</u>

Table 1. Quantitative comparison on the KODAK-24 dataset corrupted by Gaussian noise $\mathcal{N}(0, \sigma^2)$ with $\sigma = 25$. Best results are shown in **bold**, and second-best results are <u>underlined</u>.

Dataset	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	Rate (bpp)↓
Mouse Nuclei ($\sigma = 10$)	33.0337	0.8052	0.0442	0.1400	0.1908
Mouse Nuclei ($\sigma = 20$)	30.5880	0.8028	0.0734	0.1675	0.1162
SIDD (real noise)	33.6047	0.9038	0.3233	0.2365	0.2194

Table 2. Real-world denoising results on fluorescence microscopy (Mouse Nuclei) and smartphone images (SIDD). For microscopy, synthetic Gaussian noise with standard deviation σ is added following prior work.

expected, fully supervised methods (N2C and N2N) achieve the highest distortion-based metrics, benefiting from access to paired clean-noisy training data [29, 55]. Classical non-learning baselines such as JPEG-2000 and BM3D [17, 42] provide strong distortion performance, yet may yield less visually pleasing reconstructions when evaluated with perception-oriented criteria.

Among unsupervised approaches, distortion-only compression-based denoising (DeCompress) [53] exhibits weaker perceptual scores, highlighting the limitation of optimizing only the rate-distortion objective. In contrast, OTDenoising [46] attains the best Perceptual Index (PI) [11] among unsupervised methods, consistent with its explicit emphasis on perceptual alignment. Our method substantially improves PI over DeCompress while retaining the simplicity of compression-based denoising, achieving the second-best PI among unsupervised methods. Overall, these results indicate that incorporating an adversarial distribution-matching term via WGAN [3, 23] improves perceptual fidelity without requiring paired supervision.

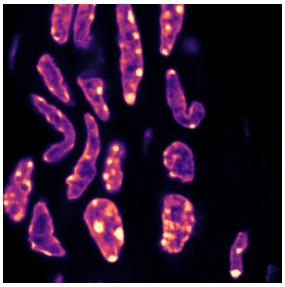
5.1.2 Fluorescence Microscopy and Real Camera Images

Beyond standard natural-image benchmarks, we evaluate the proposed framework on two real-world denoising scenarios with noise characteristics that deviate from idealized synthetic assumptions: fluorescence microscopy images from the Mouse Nuclei dataset [13] and real smartphone photographs from the SIDD dataset [1]. These datasets represent challenging acquisition conditions with domain-specific structures and complex noise statistics, particularly in microscopy and camera ISP pipelines [1, 9].

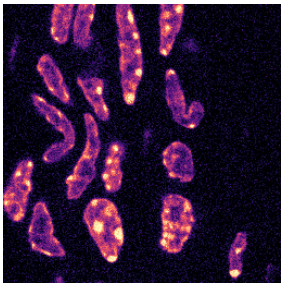
For fluorescence microscopy, we follow prior work and corrupt clean Mouse Nuclei images with additive Gaussian noise $\mathcal{N}(0, \sigma^2)$ at multiple noise levels [13]. This dataset exhibits sparse, high-contrast structures and statistics that differ substantially from natural photographic images [9, 13]. As reported in Table 2, our method achieves consistently strong denoising performance across noise levels, yielding high PSNR and SSIM while maintaining low perceptual distortion as measured by LPIPS and DISTS [18, 57], averaged over 67 test images.

To further assess robustness under realistic photographic noise, we evaluate on the SIDD dataset, which contains signal-dependent noise arising from real camera pipelines and in-camera processing [1]. Despite the absence of explicit modeling for such noise, our method attains competitive distortion and perceptual metrics on SIDD, averaged over 10 test images, indicating strong generalization beyond controlled synthetic settings [1].

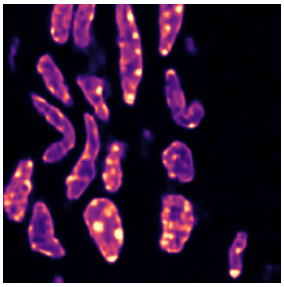
Overall, these results demonstrate that the proposed perception-enhanced compression framework extends effectively to real-world restoration scenarios, including microscopy imaging and consumer photography, while preserving a favorable balance between distortion and perceptual quality [11].



(a) Clean image.



(b) Noisy image.



(c) Denosing image.

Figure 4. Reconstruction/denoising fluorescence microscopy images.

6. Conclusion

We presented a perception-enhanced neural compression framework for image denoising from a single noisy observation. Our approach leverages the compression bottleneck as an implicit prior for suppressing noise, and further improves visual realism by integrating a WGAN-based perceptual regularizer that aligns reconstructed outputs with an unpaired clean-image distribution. Experiments on KODAK-24 under synthetic Gaussian noise validate the learned rate-distortion-perception trade-off, showing that our method achieves competitive restoration quality while substantially improving perceptual fidelity relative to distortion-only compression baselines. We further demonstrated robustness on fluorescence microscopy (Mouse Nuclei) and real camera noise (SIDD), indicating that the proposed framework extends beyond natural-image synthetic settings.

References

[1] Abderrahim Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1692–1700, 2018. 2, 5

[2] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(5):898–916, 2011. 4

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223, 2017. 1, 2, 3, 4, 5

[4] Johannes Ballé, Philip A Chou, David Minnen, Saurabh Singh, Nick Johnston, Eirikur Agustsson, Sung Jin Hwang, and George Toderici. Nonlinear transform coding. *IEEE Journal of Selected Topics in Signal Processing*, 15(2):339–353, 2020. 1, 2

[5] Johannes Ballé, Valero Laparra, and Eero P. Simoncelli. End-to-end optimized image compression. In *International Conference on Learning Representations*, 2017. 2

[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. 1, 2, 3

[7] Dalya Baron. Machine learning in astronomy: A practical overview. *arXiv preprint arXiv:1904.07248*, 2019. 1

[8] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97 of *Proceedings of Machine Learning Research*, pages 524–533. PMLR, 2019. 1, 2

[9] Chinmay Belthangady and Loic A Royer. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature methods*, 16(12):1215–1225, 2019. 1, 5

[10] Sourbh Bhadane, Aaron B Wagner, and Johannes Ballé. Do neural networks compress manifolds optimally? In *2022 IEEE Information Theory Workshop (ITW)*, pages 582–587. IEEE, 2022. 2

[11] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6228–6237, 2018. 2, 5

[12] Benoit Brummer and Christophe De Vleeschouwer. On the importance of denoising when learning to compress images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2440–2448, 2023. 2

[13] Tim-Oliver Buchholz, Mangal Prakash, Deborah Schmidt, Alexander Krull, and Florian Jug. Denoiseg: Joint denoising and segmentation. In *European Conference on Computer Vision (ECCV)*, pages 324–337. Springer, 2020. 2, 5

[14] S Grace Chang, Bin Yu, and Martin Vetterli. Image denoising via lossy compression and wavelet thresholding. In *Proceedings of International Conference on Image Processing*, volume 1, pages 604–607. IEEE, 1997. 2

[15] S Grace Chang, Bin Yu, and Martin Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE transactions on image processing*, 9(9):1532–1546, 2000. 2

[16] Eastman Kodak Company. Kodak photocd true color image suite (24 high resolution images). Dataset, 1991. 2, 4

[17] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 5

[18] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12113–12122, 2020. 5

[19] David Leigh Donoho. *The kolmogorov sampler*. Department

of Statistics, Stanford University, 2002. 1, 2

[20] David L Donoho and Iain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455, 1994. 1, 2

[21] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006. 2

[22] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014. 1

[23] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *Advances in neural information processing systems*, pages 5767–5777, 2017. 2, 5, 8

[24] Kwanyoung Kim and Jong Chul Ye. Noise2score: tweedie's approach to self-supervised image denoising without clean images. *Advances in Neural Information Processing Systems*, 34:864–874, 2021. 1, 2

[25] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 8

[26] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void – learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2129–2137. IEEE, 2019. 1, 2

[27] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Alykhan Aitken, Aadil Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4681–4690, 2017. 2

[28] Stamatios Lefkimmiatis. Universal denoising networks: a novel cnn architecture for image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3204–3213, 2018. 1, 2

[29] Jaakko Lehtinen, Jonas Munkberg, Janne Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pages 4620–4631, 2018. 1, 2, 5

[30] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 1, 2

[31] Jinming Liu, Heming Sun, and Jiro Katto. Learned image compression with mixed transformer-cnn architectures. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14388–14397, 2023. 2

[32] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999. 1, 2

[33] David Minnen, Johannes Ballé, and George D Toderici. Joint autoregressive and hierarchical priors for learned image compression. In *Advances in Neural Information Processing Systems*, pages 10771–10780, 2018. 1, 2

[34] Balas K Natarajan. Filtering random noise from deterministic signals via data compression. *IEEE transactions on signal processing*, 43(11):2595–2605, 1995. 2

[35] Ezgi Ozyilkan, Jona Ballé, Sourbh Bhadane, Aaron B Wagner, and Elza Erkip. Breaking smoothness: The struggles of neural compressors with discontinuous mappings. In *Workshop on Machine Learning and Compression, NeurIPS 2024*, 2024. 2

[36] Javier Portilla, Vasily Strela, Martin J Wainwright, and Eero P Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image processing*, 12(11):1338–1351, 2003. 1, 2

[37] Saeed Ranjbar Alvar, Mateen Ulhaq, Hyomin Choi, and Ivan V Bajić. Joint image compression and denoising via latent-space scalability. *Frontiers in Signal Processing*, 2:932873, 2022. 2

[38] Stefan Roth and Michael J Black. Fields of experts. *International Journal of Computer Vision*, 82:205–229, 2009. 1, 2

[39] Anastasiia Safonova, Gohar Ghazaryan, Stefan Stiller, Magdalena Main-Knorn, Claas Nendel, and Masahiro Ryo. Ten deep learning techniques to address small data problems with remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 125:103569, 2023. 1

[40] Naoki Saito. Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion. In *Wavelet Analysis and Its Applications*, volume 4, pages 299–324. Elsevier, 1994. 2

[41] Shakarim Soltanayev and Se Young Chun. Training deep learning based denoisers without ground truth data. *Advances in neural information processing systems*, 31, 2018. 1

[42] David S. Taubman, Michael W. Marcellin, and Majid Rabbani. Jpeg2000: Image compression fundamentals, standards and practice. *Journal of Electronic Imaging*, 11(2):286–287, 2002. 5

[43] Michela Testolina, Evgeniy Upenik, and Touradj Ebrahimi. Towards image denoising in the latent space of learning-based compression. In *Applications of Digital Image Processing XLIV*, volume 11842, pages 412–422. SPIE, 2021. 2

[44] Cédric Villani. *Optimal Transport: Old and New*, volume 338. Springer, 2009. 8

[45] Aaron B Wagner and Johannes Ballé. Neural networks optimally compress the sawbridge. In *2021 Data Compression Conference (DCC)*, pages 143–152. IEEE, 2021. 2

[46] Wei Wang, Feng Wen, Zhen Yan, and Peng Liu. Optimal transport for unsupervised denoising learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(2):2104–2118, 2023. 2, 5

[47] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yi Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *European Conference on Computer Vision (ECCV) Workshops*, 2018. 2

[48] Tsachy Weissman and Erik Ordentlich. The empirical distribution of rate-constrained source codes. *IEEE Transactions on Information Theory*, 2005. 1, 2

[49] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*. The MIT press, 1964. 1, 2

[50] Yuxin Xie, Li Yu, Farhad Pakdaman, and Moncef Gabbouj.

Joint end-to-end image compression and denoising: Leveraging contrastive learning and multi-scale self-onns. *arXiv preprint arXiv:2402.05582*, 2024. 2

[51] Fuyong Xing, Yuanpu Xie, Hai Su, Fujun Liu, and Lin Yang. Deep learning in microscopy image analysis: A survey. *IEEE transactions on neural networks and learning systems*, 29(10):4550–4568, 2017. 1

[52] Yibo Yang, Stephan Mandt, Lucas Theis, et al. An introduction to neural data compression. *Foundations and Trends® in Computer Graphics and Vision*, 15(2):113–200, 2023. 1

[53] Ali Zafari, Xi Chen, and Shirin Jalali. Decompress: Denoising via neural compression. In *2025 IEEE International Symposium on Information Theory (ISIT)*, pages 1–6, 2025. 1, 2, 3, 5

[54] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 1, 2, 3

[55] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. In *IEEE Transactions on Image Processing (TIP)*, volume 26, pages 3142–3155, 2017. 5

[56] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 1, 2

[57] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018. 5

[58] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2480–2495, 2020. 1, 2

[59] Yinhao Zhu, Yang Yang, and Taco Cohen. Transformer-based transform coding. In *International conference on learning representations*, 2022. 2

A. Architecture

A.1. Training Details

We adopt a Wasserstein GAN (WGAN) framework for distributional alignment, jointly training the encoder f, decoder g, and discriminator d. By the Kantorovich–Rubinstein duality [44], the Wasserstein-1 distance is expressed as

$$W_1(p_Y, p_{\tilde{Y}}) = \sup_{\|\nabla d\| \le 1} \mathbb{E}[d(Y)] - \mathbb{E}[d(\tilde{Y})],$$

where $\tilde{Y} = g(Q(f(X)))$. The Lipschitz constraint on the discriminator is enforced via a gradient penalty following WGAN-GP [23]. Unless otherwise specified, we use $\lambda_{\text{p}} = 0.02$, gradient penalty coefficient $\lambda_{\text{GP}} = 10$, $n_{\text{discriminator}} = 5$ discriminator updates per generator update, and classification weight $\lambda_{\text{c}} = 0.01$.

All components are optimized using Adam [25] with learning rates $5{\times}10^{-3}$ for the autoencoder, 10^{-4} for the entropy bottleneck, and 10^{-4} for the discriminator, and momentum parameters $(\beta_1, \beta_2) = (0.5, 0.999)$. We apply gradient clipping with a maximum norm of 2.0, employ mixed-precision training, and compute the exact bit-per-pixel (bpp) rate during evaluation.

A.2. Neural Network Architectures

We use the same network architectures across all datasets, summarized in Table 3. The neural codec consists of an encoder f composed of three convolutional layers followed by a learnable entropy bottleneck, and a decoder g with three deconvolutional layers and GDN activations. The adversarial branch is implemented as a WGAN-GP discriminator d comprising three strided convolutional blocks with LeakyReLU activations, followed by global average pooling and a linear output layer.

Table 3. Network architectures for the autoencoder and discriminator.

Encoder f
Conv2D, stride 2, GDN
Conv2D, stride 2, GDN
Conv2D, stride 2
Entropy Bottleneck (EB)

Decoder g
Deconv2D, stride 2, GDN
Deconv2D, stride 2, GDN
Deconv2D, stride 2

Discriminator d **(WGAN-GP)**
Conv2D, stride 2, LeakyReLU
Conv2D, stride 2, LeakyReLU
Conv2D, stride 2, LeakyReLU
Global Average Pooling