An Introduction to Neural Data Compression

Other titles in Foundations and Trends® in Computer Graphics and Vision

Vision-Language Pre-Training: Basics, Recent Advances, and Future Trends
Zhe Gan, Linjie Li, Chunyuan Li, Lijuan Wang, Zicheng Liu and Jianfeng Gao
ISBN: 978-1-63828-132-0

Semantic Image Segmentation: Two Decades of Research
Gabriela Csurka, Riccardo Volpi and Boris Chidlovskii
ISBN: 978-1-63828-076-7

Video Summarization Overview
Mayu Otani, Yale Song and Yang Wang
ISBN: 978-1-63828-078-1

A Comprehensive Review of Modern Object Segmentation Approaches
Yuanbo Wang, Unaiza Ahsan, Hanyan Li and Matthew Hagen
ISBN: 978-1-63828-070-5

Deep Learning for Image/Video Restoration and Super-resolution
A. Murat Tekalp
ISBN: 978-1-68083-972-2

Deep Learning for Multimedia Forensics
Irene Amerini, Aris Anagnostopoulos, Luca Maiano and Lorenzo Ricciardi Celsi
ISBN: 978-1-68083-854-1

An Introduction to Neural Data Compression

Yibo Yang
University of California, Irvine
yibo.yang@uci.edu

Stephan Mandt
University of California, Irvine
mandt@uci.edu

Lucas Theis
Google Research
theis@google.com

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

Y. Yang, S. Mandt and L. Theis. *An Introduction to Neural Data Compression.* Foundations and Trends® in Computer Graphics and Vision, vol. 15, no. 2, pp. 113–200, 2023.

ISBN: 978-1-63828-175-7

Foundations and Trends® in Computer Graphics and Vision
Volume 15, Issue 2, 2023

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Computer Graphics and Vision publishes survey and tutorial articles in the following topics:

- Rendering
- Shape
- Mesh simplification
- Animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and video retrieval
- Video analysis and event recognition
- Medical image analysis
- Robot localization and navigation

Information for Librarians

Foundations and Trends® in Computer Graphics and Vision, 2023, Volume 15, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Contents

An Introduction to Neural Data Compression

Yibo Yang[1], Stephan Mandt[2] and Lucas Theis[3]

[1] *University of California, Irvine, USA; yibo.yang@uci.edu*
[2] *University of California, Irvine, USA; mandt@uci.edu*
[3] *Google Research, USA; theis@google.com*

ABSTRACT

Neural compression is the application of neural networks and other machine learning methods to data compression. Recent advances in statistical machine learning have opened up new possibilities for data compression, allowing compression algorithms to be learned end-to-end from data using powerful generative models such as normalizing flows, variational autoencoders, diffusion probabilistic models, and generative adversarial networks. This monograph aims to introduce this field of research to a broader machine learning audience by reviewing the necessary background in information theory (e.g., entropy coding, rate-distortion theory) and computer vision (e.g., image quality assessment, perceptual metrics), and providing a curated guide through the essential ideas and methods in the literature thus far.

Yibo Yang, Stephan Mandt and Lucas Theis (2023), "An Introduction to Neural Data Compression", Foundations and Trends® in Computer Graphics and Vision: Vol. 15, No. 2, pp 113–200. DOI: 10.1561/0600000107.

1

Introduction

The goal of data compression is to reduce the number of bits needed to represent useful information. *Neural*, or *learned* compression, is the application of neural networks and related machine learning techniques to this task. This monograph aims to serve as an entry point for machine learning researchers interested in compression by reviewing the prerequisite background and representative methods in neural compression.

The basic idea of learning-based data compression has long existed in various forms before the current era of deep learning [224][154][37][60]. Many of the tools and techniques for neural compression, especially for images, also draw on a rich history of learning-based approaches in computer vision. Indeed, many problems in image processing and restoration can be viewed as lossy image compression; e.g., image super-resolution can be solved by learning a decoder for a fixed encoder (the image downsampling process) [49][105]. In fact, neural networks have already been applied to image compression in the late 1980s and 1990s [170][61], and even an early review article [96] exists. Compared to early work, modern methods differ markedly in their scale, neural architectures, and encoding schemes.

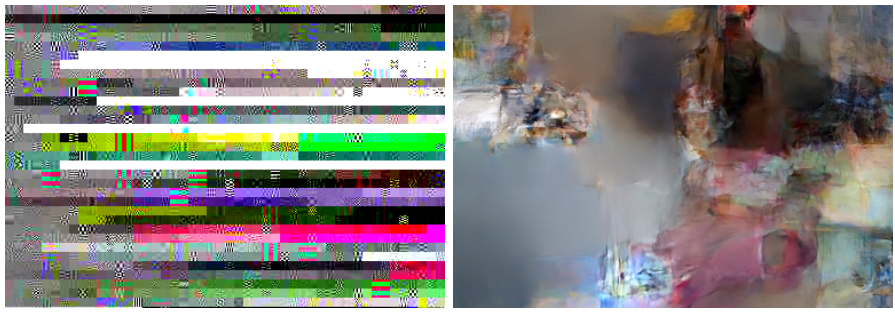

Figure 1.1: Compression as generative modeling. *Left:* A sample drawn from the probabilistic model underlying JPEG, which betrays an assumption of independence among neighboring 8 by 8 pixel blocks (except for the DC components within each row). *Right:* A sample generated by a recent neural compression model by Minnen *et al.* [132].

Current research in neural compression is heavily inspired by advances in deep generative modeling, such as GANs [65], VAEs [99][151], normalizing flows [104], and autoregressive models [180][140]. While these models allow us to capture complex data distributions from samples (a key to neural compression), the research tends to focus on generating realistic data samples [139][142] or achieving high data log-density [151][101], objectives not necessarily aligned with data compression.

Arguably the first work exploring deep generative models for data compression appeared in 2016 [70], and the topic of neural compression has grown considerably since then. Multiple researchers have identified connections between variational inference and lossless [59][118] as well as lossy [12][184][6][209] compression. This monograph hopes to further facilitate such exchange between these fields, raising awareness of compression as a fruitful application of generative modeling along with the associated challenges.

Instead of surveying the vast literature, we aim to cover the essential concepts and methods in neural compression, with a reader in mind who is versed in machine learning but not necessarily data compression. We hope to complement existing surveys that have a more specialized or applied focus [10][117][111] by highlighting the connections to generative modeling and machine learning in general. In most of this monograph, we make essentially no assumption on the data other than that it

is independently and identically distributed (i.i.d.), a typical setting for machine learning and statistics. We center our discussions around image compression, where most neural compression methods were first developed, but the basic ideas we present here are data agnostic. Towards the end, in Section 3.7, we lift the i.i.d. assumption and consider video compression, which can be seen as an extension of the existing ideas along the temporal dimension.

Neural compression can ease the development and optimization of data compression algorithms in a data-driven fashion. This can be especially useful for new or domain-specific data types, such as VR/AR content or scientific data, where developing custom codecs may otherwise be expensive. Indeed, learning-based approaches are being applied to emerging data types, such as point clouds [147][72][89], implicit 3D surfaces [178], and neural radiance fields [22]. Effectively compressing such data may require new neural architectures [178] and/or domain knowledge to convert the data into neural-network-friendly representations [89]. However, the essential ideas and techniques introduced here for reducing the entropy, or bit-rate cost, of learned representations remain the same.

JPEG [92] serves as a good motivating example of the lossy compression pipeline (depicted in Figure 1.2). First introduced in 1992, it is still one of the most widely used image compression standards [90]. At the heart of JPEG are linear mappings which losslessly transform pixels into coefficients and back. The coefficients are first quantized to integers, incurring some information loss. Then they are further compressed losslessly by a combination of run-length encoding and entropy coding (the latter is discussed in Section 2.1.1).

The linear portion of the encoding process consists of several steps. First, each pixel is transformed from RGB to YCC coefficients consisting of a luma component (Y) and two color components (C). After this color transform, each channel is treated independently, and optional downsampling is applied to the color channels. Next, each channel is divided into 8×8 pixel blocks, and each block independently undergoes a *discrete cosine transform* (DCT). The transform coefficients are then

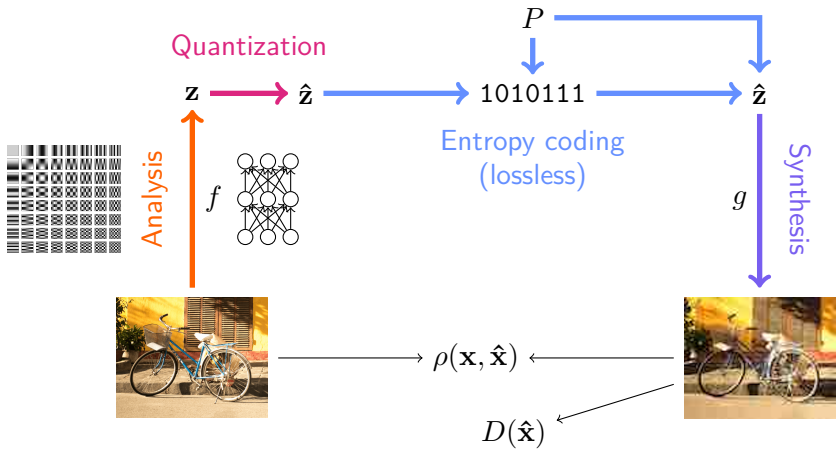

Figure 1.2: A typical pipeline in both neural and classical lossy image compression. An encoder transformation f (for example, the DCT or a neural network) maps images to coefficients $\mathbf{z}$, which are first quantized to $\hat{\mathbf{z}}$, and then entropy encoded into bits using an entropy model P. A reconstruction $\hat{\mathbf{x}}$ is obtained using a decoder g that aims for a small distortion ρ between the data $\mathbf{x}$ and its lossy reconstruction $\hat{\mathbf{x}}$. In addition, neural compression can also involve an adversarial critic D, encouraging realism and high perceptual quality.

linearly scaled and finally rounded to integers. Given an image $\mathbf{x}$, the encoder thus performs

$$\hat{\mathbf{z}} = \lfloor \mathbf{DACx} \rceil, \tag{1.1}$$

where $\mathbf{C}$ is the pixelwise color transform, $\mathbf{A}$ is the block- and channelwise DCT, and $\mathbf{D}$ is a diagonal matrix scaling the coefficients. The decoder applies the transforms in reverse,

$$\hat{\mathbf{x}} = \mathbf{C}^{-1}\mathbf{A}^{\top}\mathbf{D}^{-1}\hat{\mathbf{z}}. \tag{1.2}$$

Readers familiar with machine learning will be reminded of autoencoders [29][158] and it is natural to consider learned neural networks in place of the linear transforms. As we will see later, there are indeed close connections between lossy compression and variational autoencoders (VAEs) [12][184][6][211], though other generative models have a role to play as well. What we call "coefficients" in the context of compression are often called "latent variables" in the context of generative models.

Like generative models, JPEG defines a probability distribution over coefficients which represents assumptions about the latent representation. Just as in VAEs, we can use this distribution to draw samples from the model underlying JPEG, with an example shown in Figure 1.1.

Overview. This introduction is organized into two main parts, lossless (Section 2) and lossy (Section 3) compression, with the latter relying on the former for compressing lossy representations of the data (see Figure 1.2). We begin by reviewing basic *coding theory* (Section 2.1), and learn how we can turn the problem of lossless compression into learning a discrete data distribution, with the help of *entropy-coding*. For this to work in practice, we decompose the potentially high-dimensional data distribution using tools from generative modeling, including *autoregressive models* (Section 2.2), *latent-variable models*, (Section 2.3), and other models (Section 2.4). Each model class differs in its compatibility with different entropy-coding algorithms, and offers a different trade-off between the compression bit-rate and computational efficiency. *Lossy* compression introduces additional desiderata, the most common being the *distortion* of reconstructions, based on which the classical *rate-distortion theory* and algorithms such as *vector quantization* and *transform coding* are reviewed (Section 3.1). We then introduce *neural lossy compression* as a natural extension of transform coding (Section 3.2) and discuss the techniques necessary for end-to-end learning of quantized representations (Section 3.3), as well as lossy compression schemes that attempt to bypass quantization (Section 3.4). We then explore additional desiderata, such as the *perceptual quality* of reconstructions (Section 3.5), and the usefulness of learned representations for downstream tasks (Section 3.6), before briefly reviewing video compression (Section 3.7). Finally, we conclude in Section 4 with the challenges and open problems in neural compression that may drive its future advances.

References

[1] E. Agustsson, F. Mentzer, M. Tschannen, L. Cavigelli, R. Timofte, L. Benini, and L. V. Gool, "Soft-to-Hard Vector Quantization for End-to-End Learning Compressible Representations," in *Advances in Neural Information Processing Systems 30*, pp. 1141–1151, 2017.

[2] E. Agustsson and L. Theis, "Universally Quantized Neural Compression," in *Advances in Neural Information Processing Systems 33*, 2020.

[3] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool, "Generative adversarial networks for extreme learned image compression," in *The IEEE International Conference on Computer Vision (ICCV)*, pp. 221–231, 2019.

[4] E. Agustsson, D. Minnen, N. Johnston, J. Ballé, S. J. Hwang, and G. Toderici, "Scale-space flow for end-to-end optimized video compression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8503–8512, 2020.

[5] N. Ahmed, T. R. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. C-23, 1974, pp. 90–93.

[6] A. Alemi, B. Poole, I. Fischer, J. Dillon, R. A. Saurous, and K. Murphy, "Fixing a broken elbo," in *International Conference on Machine Learning*, PMLR, pp. 159–168, 2018.

[7] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, "Deep variational information bottleneck," *arXiv preprint arXiv:1612.00410*, 2016.

[8] D. Amir and Y. Weiss, "Understanding and simplifying perceptual distances," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12 226–12 235, Jun. 2021.

[9] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, 1972, pp. 14–20.

[10] J. Ballé, P. A. Chou, D. Minnen, S. Singh, N. Johnston, E. Agustsson, S. J. Hwang, and G. Toderici, "Nonlinear transform coding," *IEEE Trans. on Special Topics in Signal Processing*, vol. 15, 2021.

[11] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimization of nonlinear transform codes for perceptual quality," in *Picture Coding Symposium*, 2016.

[12] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end Optimized Image Compression," in *International Conference on Learning Representations*, 2017.

[13] J. Ballé, N. Johnston, and D. Minnen, "Integer networks for data compression with latent-variable models," in *International Conference on Learning Representations*, 2018.

[14] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational Image Compression with a Scale Hyperprior," *International Conference on Learning Representations*, 2018.

[15] R. Bamler, "Understanding Entropy Coding With Asymmetric Numeral Systems (ANS): a Statistician's Perspective," *arXiv preprint arXiv:2201.01741*, 2022.

[16] F. Bellard, *Lossless Data Compression with Neural Networks*, 2019.

[17] Y. Bengio, N. Leonard, and A. Courville, *Estimating or propagating gradients through stochastic neurons for conditional computation*, 2013.

[18] C. H. Bennett and P. W. Shor, "Entanglement-Assisted Capacity of a Quantum Channel and the Reverse Shannon Theorem," *IEEE Transactions on Information Theory*, vol. 48, no. 10, 2002.

[19] T. Berger and J. D. Gibson, "Lossy source coding," *IEEE Transactions on Information Theory*, vol. 44, no. 6, 1998, pp. 2693–2723.

[20] T. G. Bever and D. Poeppel, "Analysis by synthesis: A (re-) emerging program of research for language and vision," *Biolinguistics*, vol. 4, no. 2-3, 2010, pp. 174–200.

[21] S. Bhardwaj, I. Fischer, J. Ballé, and T. Chinen, "An Unsupervised Information-Theoretic Perceptual Quality Metric," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 13–24, Curran Associates, Inc., 2020.

[22] T. Bird, J. Ballé, S. Singh, and P. A. Chou, "3d scene compression through entropy penalized neural representation functions," in *2021 Picture Coding Symposium (PCS)*, IEEE, pp. 1–5, 2021.

[23] C. M. Bishop, *Pattern Recognition and Machine Learning*, M. Jordan, J. Kleinberg, and B. Schölkopf, Eds., ser. Information science and statistics. Springer, 2006, ch. Graphical. DOI: 10.1117/1.2819119.

[24] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, 1972, pp. 460–473. DOI: 10.1109/TIT.1972.1054855.

[25] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6228–6237, 2018.

[26] G. E. Blelloch, *Introduction to Data Compression*. Carnegie Mellon University, 2013.

[27] L. Boltzmann, "Vorlesungen über Gastheorie,(2 volumes)," *Leipzig (1895, 1898)*, 1895.

[28] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3773–3777, 2016. DOI: 10.1109/ICIP.2016.7533065.

[29] H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," *Biological cybernetics*, vol. 59, no. 4, 1988, pp. 291–294.

[30] J. Bruna, P. Sprechmann, and Y. LeCun, "Super-resolution with deep convolutional sufficient statistics," in *International Conference on Learning Representations*, Jan. 2016.

[31] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, 2013, pp. 1872–1886. DOI: 10.1109/TPAMI.2012.230.

[32] Y. Burda, R. Grosse, and R. Salakhutdinov, "Importance weighted autoencoders," *arXiv preprint arXiv:1509.00519*, 2015.

[33] T. Chen, H. Liu, Q. Shen, T. Yue, X. Cao, and Z. Ma, "Deepcoder: A deep neural network based video compression," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, IEEE, pp. 1–4, 2017.

[34] X. Chen, N. Mishra, M. Rohaninejad, and P. Abbeel, "PixelSNAIL: An improved autoregressive generative model," in *Proceedings of the 35th International Conference on Machine Learning*, J. Dy and A. Krause, Eds., pp. 864–872, PMLR, Jul. 2018.

[35] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Learned image compression with discretized gaussian mixture likelihoods and attention modules," *arXiv preprint arXiv:2001.01568*, 2020. arXiv: 2001.01568 [eess.IV].

[36] Y. Choi, M. El-Khamy, and J. Lee, "Variable rate deep image compression with a conditional autoencoder," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019.

[37] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, 1989, pp. 31–42.

[38] R. Clausius, *Mechanical Theory of Heat*. Taylor and Francis, 1850.

[39] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, vol. 2. John Wiley & Sons, 2006.

[40] C. Cremer, X. Li, and D. Duvenaud, "Inference suboptimality in variational autoencoders," in *International Conference on Machine Learning*, pp. 1078–1086, 2018.

[41] P. Cuff, "Communication requirements for generating correlated random variables," in *2008 IEEE International Symposium on Information Theory*, pp. 1393–1397, 2008.

[42] P. W. Cuff and E. C. Song, "The likelihood encoder for source coding," in *2013 IEEE Information Theory Workshop*, 2013.

[43] J. Degrave and I. Korshunova, *How we can make machine learning algorithms tunable*. URL: https://www.engraved.blog/how-we-can-make-machine-learning-algorithms-tunable/.

[44] E. L. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a Laplacian pyramid of adversarial networks," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 1486–1494, 2015.

[45] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems," *International Journal of Computer Vision*, no. 129, 2021, pp. 1258–1281.

[46] L. Dinh, D. Krueger, and Y. Bengio, "Nice: Non-linear independent components estimation," *arXiv preprint arXiv:1410.8516*, 2014.

[47] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real nvp," *arXiv preprint arXiv:1605.08803*, 2016.

[48] A. Djelouah, J. Campos, S. Schaub-Meyer, and C. Schroers, "Neural inter-frame compression for video coding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6421–6429, 2019.

[49] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, 2015, pp. 295–307.

[50] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," in *Advances in Neural Information Processing Systems*, vol. 29, Curran Associates, Inc., 2016. URL: https://proceedings.neurips.cc/paper/2016/file/371bce7dc83817b7893bcdeed13799b5-Paper.pdf.

[51] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2758–2766, 2015.

[52] Z. Duan, M. Lu, Z. Ma, and F. Zhu, "Lossy image compression with quantized hierarchical vaes," *arXiv preprint arXiv:2208.F180313056*, 2022.

[53] Y. Dubois, B. Bloem-Reddy, K. Ullrich, and C. J. Maddison, "Lossy compression for lossless prediction," in *Neural Information Processing Systems*, 2021.

[54] J. Duda, "Asymmetric numeral systems," *arXiv preprint arXiv:0902.0271*, 2009.

[55] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Communications*, vol. 43, 1995, pp. 2959–2965.

[56] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12 873–12 883, 2021.

[57] G. Flamich, M. Havasi, and J. M. Hernández-Lobato, *Compressing Images by Encoding Their Latent Representations with Relative Entropy Coding*, 2020.

[58] G. Flamich, S. Markou, and J. M. Hernandez-Lobato, "Fast relative entropy coding with a* coding," in *Proceedings of the 39th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 162, pp. 6548–6577, PMLR, Jul. 2022.

[59] B. J. Frey, *Bayesian networks for pattern classification, data compression, and channel coding.* Citeseer, 1998.

[60] B. J. Frey and G. E. Hinton, "Efficient stochastic source coding and an application to a bayesian network source model," *The Computer Journal*, vol. 40, no. 2_and_3, 1997, pp. 157–165.

[61] S. M. G.L. Sicuranza G. Ramponi, "Artificial neural network for image compression," *Electronics Letters*, vol. 26, 7 Mar. 1990, 477–479(2).

[62] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2016. URL: http://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Gatys_Image_Style_Transfer_CVPR_2016_paper.html.

[63] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, vol. 159. Springer Science & Business Media, 2012.

[64] A. Golinski, R. Pourreza, Y. Yang, G. Sautiere, and T. S. Cohen, "Feedback recurrent autoencoder for video compression," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[65] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., vol. 27, 2014.

[66] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[67] V. K. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Processing Magazine*, vol. 18, no. 5, 2001, pp. 9–21.

[68] V. K. Goyal, J. Zhuang, and M. Veiterli, "Transform coding with backward adaptive updates," *IEEE Transactions on Information Theory*, vol. 46, no. 4, 2000, pp. 1623–1633.

[69] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE transactions on information theory*, vol. 44, no. 6, 1998, pp. 2325–2383.

[70] K. Gregor, F. Besse, D. Jimenez Rezende, I. Danihelka, and D. Wierstra, "Towards conceptual compression," *Advances In Neural Information Processing Systems*, vol. 29, 2016, pp. 3549–3557.

[71] P. D. Grünwald, *The minimum description length principle*. MIT press, 2007.

[72] A. F. Guarda, N. M. Rodrigues, and F. Pereira, "Point cloud coding: Adopting a deep learning-based approach," in *2019 Picture Coding Symposium (PCS)*, IEEE, pp. 1–5, 2019.

[73] A. Habibian, T. v. Rozendaal, J. M. Tomczak, and T. S. Cohen, "Video compression with rate-distortion autoencoders," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7033–7042, 2019.

[74] J. Han, S. Lombardo, C. Schroers, and S. Mandt, "Deep generative video compression," in *Neural Information Processing Systems*, 2019.

[75] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The Communication Complexity of Correlation," in *Twenty-Second Annual IEEE Conference on Computational Complexity*, pp. 10–23, 2007.

[76] M. Havasi, R. Peharz, and J. M. Hernández-Lobato, "Minimal Random Code Learning: Getting Bits Back from Compressed Model Parameters," in *International Conference on Learning Representations*, 2019.

[77] L. Helminger, A. Djelouah, M. Gross, and C. Schroers, "Lossy image compression with normalizing flows," *arXiv preprint arXiv: 2008.10486*, 2020.

[78] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[79] A. Hines and N. Harte, "Speech Intelligibility Prediction using a Neurogram Similarity Index Measure," *Speech Communication*, vol. 54, no. 2, 2012, pp. 306–320.

[80] A. Hines, J. Skoglund, A. C. Kokaram, and N. Harte, "ViSQOL: an objective speech quality model," *EURASIP Journal on Audio, Speech, and Music Processing*, 2015. DOI: 10.1186/s13636-015-0054-9.

[81] G. E. Hinton and D. Van Camp, "Keeping the neural networks simple by minimizing the description length of the weights," in *Proceedings of the sixth annual conference on Computational learning theory*, pp. 5–13, 1993.

[82] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *NeurIPS*, 2020.

[83] J. Ho, E. Lohn, and P. Abbeel, "Compression with flows via local bits-back coding," in *Advances in Neural Information Processing Systems*, pp. 3874–3883, 2019.

[84] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, 1997, pp. 1745–1780.

[85] A. Honkela and H. Valpola, "Variational learning and bits-back coding: An information-theoretic view to bayesian learning," *IEEE transactions on Neural Networks*, vol. 15, no. 4, 2004, pp. 800–810.

[86] E. Hoogeboom, A. A. Gritsenko, J. Bastings, B. Poole, R. v. d. Berg, and T. Salimans, "Autoregressive diffusion models," *arXiv preprint arXiv:2110.02037*, 2021.

[87] E. Hoogeboom, J. Peters, R. van den Berg, and M. Welling, "Integer discrete flows and lossless compression," in *Advances in Neural Information Processing Systems*, pp. 12 134–12 144, 2019.

[88] R. Hosseini, F. Sinz, and M. Bethge, "Lower bounds on the redundancy of natural images," *Vision Research*, vol. 50, no. 22, 2010, pp. 2213–2222.

[89] L. Huang, S. Wang, K. Wong, J. Liu, and R. Urtasun, "Octsqueeze: Octree-structured entropy model for lidar compression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1313–1323, 2020.

[90] G. Hudson, A. Léger, B. Niss, I. Sebestyén, and J. Vaaben, "JPEG-1 standard 25 years: Past, present, and future reasons for a success," *Journal of Electronic Imaging*, vol. 27, no. 4, 2018, p. 040 901.

[91] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, 1952, pp. 1098–1101.

[92] ITU-T, *Recommendation ITU-T T.81: Information technology – Digital compression and coding of continuous-tone still images – Requirements and guidelines*, 1992.

[93] ITU-T, *Recommendation ITU-T T.800.2: Methods for objective and subjective assessment of speech and video quality*, 2016.

[94] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, 1991.

[95] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," *arXiv preprint arXiv:1611.01144*, 2016.

[96] J. Jiang, "Image compression with neural networks–a survey," *Signal processing: image Communication*, vol. 14, no. 9, 1999, pp. 737–760.

[97] J. Kim, A.-D. Nguyen, and S. Lee, "Deep cnn-based blind image quality predictor," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, 2019, pp. 11–24. DOI: 10.1109/TNNLS.2018.2829819.

[98] Y. Kim, S. Wiseman, A. Miller, D. Sontag, and A. Rush, "Semi-amortized variational autoencoders," in *International Conference on Machine Learning*, PMLR, pp. 2678–2687, 2018.

[99] D. Kingma and M. Welling, "Auto-encoding variational Bayes," in *International Conference on Learning Representations*, 2014.

[100] D. P. Kingma and P. Dhariwal, "Glow: Generative Flow with Invertible 1x1 Convolutions," in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., pp. 10 215–10 224, 2018.

[101] D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, "Improved variational inference with inverse autoregressive flow," *Advances in neural information processing systems*, vol. 29, 2016.

[102] B. Knoll, *Cmix*. URL: https://www.byronknoll.com/cmix.html.

[103] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25, Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[104] H. Larochelle and I. Murray, "The Neural Autoregressive Distribution Estimator," *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, 2011.

[105] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, 2017.

[106] J. Lee, S. Cho, and S.-K. Beack, "Context-adaptive entropy model for end-to-end optimized image compression," in *International Conference on Learning Representations*, May 2019.

[107] C. T. Li and A. E. Gamal, "Strong Functional Representation Lemma and Applications to Coding Theorems," in *IEEE International Symposium on Information Theory*, 2017.

[108] M. Li, W. Zuo, S. Gu, D. Zhao, and D. Zhang, *Learning convolutional networks for content-weighted image compression*, 2017. arXiv: 1703.10553 [cs.CV].

[109] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, *Toward a practical perceptual video quality metric*, 2016. URL: https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652.

[110] A. Liu, S. Mandt, and G. V. d. Broeck, "Lossless compression with probabilistic circuits," in *International Conference on Learning Representations*, 2022.

[111] D. Liu, Y. Li, J. Lin, H. Li, and F. Wu, "Deep learning-based video coding: A review and a case study," *ACM Computing Surveys (CSUR)*, vol. 53, no. 1, 2020, pp. 1–35.

[112] H. Liu, T. Chen, P. Guo, Q. Shen, X. Cao, Y. Wang, and Z. Ma, "Non-local attention optimized deep image compression," *arXiv preprint arXiv:1904.09757*, 2019.

[113] T.-J. Liu, Y.-C. Lin, W. Lin, and C.-C. J. Kuo, "Visual quality assessment: Recent developments, coding applications and future trends," *APSIPA Transactions on Signal and Information Processing*, vol. 2, 2013. DOI: 10.1017/ATSIP.2013.5.

[114] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, 1982, pp. 129–137.

[115] G. Lu, W. Ouyang, D. Xu, X. Zhang, C. Cai, and Z. Gao, "Dvc: An end-to-end deep video compression framework," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11 006–11 015, 2019.

[116] G. Lu, X. Zhang, W. Ouyang, L. Chen, Z. Gao, and D. Xu, "An end-to-end learning framework for video compression," *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[117] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, and S. Wang, "Image and video compression with neural networks: A review," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, 2019, pp. 1683–1698.

[118] D. J. MacKay and D. J. Mac Kay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[119] C. J. Maddison, A. Mnih, and Y. W. Teh, "The concrete distribution: A continuous relaxation of discrete random variables," *arXiv preprint arXiv:1611.00712*, 2016.

[120] M. Mahoney, *Large Text Compression Benchmark*. URL: http://mattmahoney.net/dc/text.html.

[121] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 2813–2821, 2017. DOI: 10.1109/ICCV.2017.304.

[122] G. Martin, "Range encoding: An algorithm for removing redundancy from a digitised message," in *Video and Data Recording Conference, Southampton, 1979*, pp. 24–27, 1979.

[123] Y. Matsubara, R. Yang, M. Levorato, and S. Mandt, "Supervised compression for resource-constrained edge computing systems," *arXiv preprint arXiv:2108.11898*, 2021.

[124] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool, "Conditional probability models for deep image compression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4394–4402, 2018.

[125] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool, "Practical full resolution learned lossless image compression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[126] F. Mentzer, G. Toderici, D. Minnen, S.-J. Hwang, S. Caelles, M. Lucic, and E. Agustsson, "Vct: A video compression transformer," *arXiv preprint arXiv:2206.07307*, 2022.

[127] F. Mentzer, G. D. Toderici, M. Tschannen, and E. Agustsson, "High-fidelity generative image compression," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[128] B. Meyer and P. Tischer, "Glicbawls – Grey Level Image Compression By Adaptive Weighted Least Squares," in *Data Compression Conference*, 2001.

[129] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, 2021, pp. 99–106.

[130] T. Minka, "Divergence measures and message passing," Microsoft Research, Tech. Rep. TR-2005-173, 2005.

[131] D. Minnen, J. Ballé, and G. D. Toderici, "Joint Autoregressive and Hierarchical Priors for Learned Image Compression," in *Advances in Neural Information Processing Systems 31*, 2018.

[132] D. Minnen and S. Singh, "Channel-wise autoregressive entropy models for learned image compression," in *IEEE International Conference on Image Processing (ICIP)*, 2020.

[133] D. Minnen, *Current Frontiers In Neural Image Compression: The Rate-Distortion-Computation Trade-Off And Optimizing For Subjective Visual Quality*, Plenary talk at ICIP 2021, 2021. URL: https://www.youtube.com/watch?v=84XkOlBhOas (accessed on 10/26/2022).

[134] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, Mar. 2013, pp. 209–212. DOI: 10.1109/LSP.2012.2227726.

[135] D. Mukherjee, "Challenges in incorporating ML in a mainstream nextgen video codec," *CLIC Workshop and Challenge on Learned Image Compression*, 2022. URL: https://storage.googleapis.com/clic2022_public/slides/Challenges%20in%20incorporating%20ML%20in%20a%20practical%20Nextgen%20Video%20Codec.pdf.

[136] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *arXiv preprint arXiv:2201.05989*, 2022.

[137] X. L. Nguyen, M. J. Wainwright, and M. I. Jordan, "On surrogate loss functions and f-divergences," *The Annals of Statistics*, vol. 37, no. 2, 2009, pp. 876–904. DOI: 10.1214/08-AOS595.

[138] S. Nowozin, B. Cseke, and R. Tomioka, "f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization," in *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[139] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio," *CoRR*, vol. abs/1609.03499, 2016. URL: http://arxiv.org/abs/1609.03499.

[140] A. van den Oord and N. Kalchbrenner, "Pixel Recurrent Neural Networks," in *Proceedings of the 33rd International Conference on Machine Learning*, 2016.

[141] A. van den Oord, N. Kalchbrenner, O. Vinyals, A. G. L. Espeholt, and K. Kavukcuoglu, "Conditional Image Generation with PixelCNN Decoders," in *Advances in Neural Information Processing Systems 29*, pp. 4790–4798, 2016.

[142] A. v. d. Oord, O. Vinyals, and K. Kavukcuoglu, "Neural discrete representation learning," 2017.

[143] N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, and D. Tran, "Image transformer," in *Proceedings of the 35th International Conference on Machine Learning*, J. Dy and A. Krause, Eds., ser. Proceedings of Machine Learning Research, vol. 80, pp. 4055–4064, Stockholmsmässan, Stockholm Sweden: PMLR, Jul. 2018.

[144] R. Peharz, S. Lang, A. Vergari, K. Stelzner, A. Molina, M. Trapp, G. Van den Broeck, K. Kersting, and Z. Ghahramani, "Einsum networks: Fast and scalable learning of tractable probabilistic circuits," in *International Conference on Machine Learning*, 2020.

[145] J. C. Platt and A. H. Barr, "Constrained differential optimization for neural networks," California Institute of Technology, Tech. Rep., 1988.

[146] M. Quach, J. Pang, D. Tian, G. Valenzise, and F. Dufaux, "Survey on deep learning-based point cloud compression," *Frontiers in Signal Processing*, 2022.

[147] M. Quach, G. Valenzise, and F. Dufaux, "Learning convolutional transforms for lossy point cloud geometry compression," in *2019 IEEE international conference on image processing (ICIP)*, IEEE, pp. 4320–4324, 2019.

[148] M. Quach, G. Valenzise, and F. Dufaux, "Improved deep point cloud geometry compression," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, IEEE, pp. 1–6, 2020.

[149] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *International Conference on Machine Learning*, PMLR, pp. 8821–8831, 2021.

[150] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *International Conference on Machine Learning*, 2014.

[151] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *International conference on machine learning*, PMLR, pp. 1530–1538, 2015.

[152] O. Rippel and L. Bourdev, "Real-time adaptive image compression," in *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[153] O. Rippel, S. Nair, C. Lew, S. Branson, A. G. Anderson, and L. Bourdev, "Learned video compression," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3454–3463, 2019.

[154] J. Rissanen and G. G. Langdon, "Arithmetic coding," *IBM Journal of research and development*, vol. 23, no. 2, 1979, pp. 149–162.

[155] L. G. Roberts, "Picture Coding Using Pseudo-Random Noise," *IRE Transactions on Information Theory*, 1962.

[156] T. van Rozendaal, G. Sautière, and T. S. Cohen, *Lossy compression with distortion constrained optimization*, 2020. arXiv: 2005.04064 [cs.LG].

[157] Y. Ruan, K. Ullrich, D. Severo, J. Townsend, A. Khisti, A. Doucet, A. Makhzani, and C. J. Maddison, "Improving lossless compression rates via monte carlo bits-back coding," in *International Conference on Machine Learning*, 2021.

[158] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.

[159] C. Saharia, W. Chan, H. Chang, C. A. Lee, J. Ho, T. Salimans, D. J. Fleet, and M. Norouzi, "Palette: Image-to-Image Diffusion Models," *CoRR*, vol. abs/2111.05826, 2021. arXiv: 2111.05826.

[160] T. Salimans, A. Karpathy, X. Chen, and D. P. Kingma, "PixelCNN++: A pixelcnn implementation with discretized logistic mixture likelihood and other modifications," in *International Conference on Learning Representations*, 2017.

[161] S. Santurkar, D. Budden, and N. Shavit, "Generative compression," in *2018 Picture Coding Symposium (PCS)*, pp. 258–262, 2018.

[162] K. Sayood, "Vector quantization," in *Introduction to Data Compression (Fourth Edition)*, ser. The Morgan Kaufmann Series in Multimedia Information and Systems, K. Sayood, Ed., 4th ed., Boston: Morgan Kaufmann, 2012, pp. 295–344. DOI: https://doi.org/10.1016/B978-0-12-415796-5.00010-7.

[163] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, 1948, pp. 379–423.

[164] C. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec., March 1959*, vol. 4, 1959, pp. 142–163.

[165] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, 2006, pp. 3440–3451. DOI: 10.1109/TIP.2006.881959.

[166] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[167] S. Singh, S. Abu-El-Haija, N. Johnston, J. Ballé, A. Shrivastava, and G. Toderici, "End-to-end learning of compressible features," in *ICIP*, IEEE, pp. 3349–3353, 2020.

[168] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International Conference on Machine Learning*, PMLR, pp. 2256–2265, 2015.

[169] C. K. Sønderby, T. Raiko, L. Maaløe, S. K. Sønderby, and O. Winther, "Ladder variational autoencoders," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29, Curran Associates, Inc., 2016. URL: https://proceedings.neurips.cc/paper/2016/file/6ae07dcb33ec3b7c814df797cbda0f87-Paper.pdf.

[170] N. Sonehara, M. Kawato, S. Miyake, and K. Nakane, "Image data compression using a neural network model," *International 1989 Joint Conference on Neural Networks*, 1989, 35–41 vol.2.

[171] E. C. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Transactions on Information Theory*, vol. 62, 4 2016.

[172] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, and G. R. G. Lanckriet, "On integral probability metrics, ϕ-divergences and binary classification," *arXiv*, arXiv:0901.2698, Jan. 2009, arXiv:0901.2698.

[173] M. Stern, N. Shazeer, and J. Uszkoreit, "Blockwise parallel decoding for deep autoregressive models," in *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[174] K. Storrs, S. V. Leuven, S. Kojder, L. Theis, and F. Huszár, "Adaptive paired-comparison method for subjective video quality assessment on mobile devices," in *Picture Coding Symposium*, 2018. URL: https://arxiv.org/abs/1807.02175.

[175] M. Strathern, "Improving ratings: audit in the British University system," *European Review*, vol. 5, 3 1997, pp. 305–321.

[176] G. J. Sullivan, "Efficient scalar quantization of exponential and laplacian random variables," *IEEE Transactions on Information Theory*, vol. 42, no. 5, 1996, pp. 1365–1374. DOI: 10.1109/18.532878.

[177] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *arXiv*, vol. 1512.00567, 2015.

[178] D. Tang, S. Singh, P. A. Chou, C. Hane, M. Dou, S. Fanello, J. Taylor, P. Davidson, O. G. Guleryuz, Y. Zhang, *et al.*, "Deep implicit volume compression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1293–1303, 2020.

[179] L. Theis and E. Agustsson, "On the advantages of stochastic encoders," 2021. URL: https://arxiv.org/abs/2102.09270.

[180] L. Theis and M. Bethge, "Generative image modeling using spatial LSTMs," in *Advances in Neural Information Processing Systems 28*, 2015.

[181] L. Theis and J. Ho, "Importance weighted compression," in *Neural Compression Workshop at ICLR 2021*, 2021. URL: https://openreview.net/forum?id=n6skss_9-v3.

[182] L. Theis, R. Hosseini, and M. Bethge, "Mixtures of conditional Gaussian scale mixtures applied to multiscale image representations," *PLoS ONE*, vol. 7, no. 7, 2012.

[183] L. Theis, A. van den Oord, and M. Bethge, "A note on the evaluation of generative models," in *International Conference on Learning Representations*, Apr. 2016. URL: http://arxiv.org/abs/1511.01844.

[184] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy Image Compression with Compressive Autoencoders," in *International Conference on Learning Representations*, 2017.

[185] L. Theis, T. Salimans, M. D. Hoffman, and F. Mentzer, *Lossy compression with gaussian diffusion*, 2022.

[186] L. Theis and N. Yosri, "Algorithms for the communication of samples," in *Proceedings of the 39th International Conference on Machine Learning*, 2021.

[187] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," *arXiv preprint physics/0004057*, 2000.

[188] G. Toderici, S. M. O'Malley, S. J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, and R. Sukthankar, "Variable rate image compression with recurrent neural networks," in *International Conference on Learning Representations*, 2016.

[189] G. Toderici, D. Vincent, N. Johnston, S. Jin Hwang, D. Minnen, J. Shor, and M. Covell, "Full resolution image compression with recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5306–5314, 2017.

[190] R. Torfason, F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool, "Towards image understanding from deep compression without decoding," *arXiv preprint arXiv:1803.06131*, 2018.

[191] J. Townsend, "A tutorial on the range variant of asymmetric numeral systems," *arXiv preprint arXiv:2001.09186*, 2020.

[192] J. Townsend, T. Bird, J. Kunze, and D. Barber, "Hilloc: Lossless image compression with hierarchical latent variable models," *arXiv preprint arXiv:1912.09953*, 2019.

[193] J. Townsend, T. Bird, and D. Barber, "Practical lossless compression with latent variables using bits back coding," *arXiv preprint arXiv:1901.04866*, 2019.

[194] J. Townsend and I. Murray, "Lossless compression with state space models using bits back coding," *arXiv preprint arXiv:2103.10150*, 2021.

[195] D. Tran, K. Vafa, K. Agrawal, L. Dinh, and B. Poole, "Discrete flows: Invertible generative models of discrete data," in *Advances in Neural Information Processing Systems*, pp. 14 719–14 728, 2019.

[196] K. Tsubota and K. Aizawa, "Comprehensive comparisons of uniform quantizers for deep image compression," in *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 2089–2093, 2021. DOI: 10.1109/ICIP42928.2021.9506497.

[197] B. Uria, I. Murray, and H. Larochelle, "RNADE: the real-valued neural autoregressive density-estimator," in *Advances in Neural Information Processing Systems 26*, vol. 26, 2013.

[198] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[199] A. B. Wagner and J. Ballé, "Neural networks optimally compress the sawbridge," in *2021 Data Compression Conference (DCC)*, IEEE, pp. 143–152, 2021.

[200] C. S. Wallace, "Classification by minimum-message-length inference," in *International Conference on Computing and Information*, Springer, pp. 72–81, 1990.

[201] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, 1398–1402 Vol.2, 2003. DOI: 10.1109/ACSSC.2003.1292216.

[202] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine learning*, vol. 8, no. 3, 1992, pp. 229–256.

[203] C.-Y. Wu, N. Singhal, and P. Krahenbuhl, "Video compression through image interpolation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 416–431, 2018.

[204] X. Wu, K. U. Barthel, and W. Zhang, "Piecewise 2D Autoregression for Predictive Image Coding," in *ICIP*, pp. 901–904, IEEE Computer Society, 1998.

[205] Y. Xie, K. L. Cheng, and Q. Chen, "Enhanced invertible encoding for learned image compression," in *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 162–170, 2021.

[206] R. Yang, F. Mentzer, L. Van Gool, and R. Timofte, "Learning for video compression with recurrent auto-encoder and recurrent probability model," *IEEE Journal of Selected Topics in Signal Processing*, 2020.

[207] R. Yang and S. Mandt, *Lossy image compression with conditional diffusion models*, 2022.

[208] R. Yang, Y. Yang, J. Marino, and S. Mandt, "Hierarchical autoregressive modeling for neural video compression," in *International Conference on Learning Representations*, 2020.

[209] Y. Yang, R. Bamler, and S. Mandt, "Improving inference for neural image compression," in *Neural Information Processing Systems (NeurIPS), 2020*, 2020.

[210] Y. Yang, R. Bamler, and S. Mandt, "Variational Bayesian Quantization," in *International Conference on Machine Learning*, 2020.

[211] Y. Yang and S. Mandt, "Towards empirical sandwich bounds on the rate-distortion function," in *International Conference on Learnning Representations*, 2022.

[212] A. Yu, R. Li, M. Tancik, H. Li, R. Ng, and A. Kanazawa, "Plenoctrees for real-time rendering of neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5752–5761, 2021.

[213] J. Yu, X. Li, J. Y. Koh, H. Zhang, R. Pang, J. Qin, A. Ku, Y. Xu, J. Baldridge, and Y. Wu, "Vector-quantized image modeling with improved vqgan," *arXiv preprint arXiv:2110.04627*, 2021.

[214] J. Yu, Y. Xu, J. Y. Koh, T. Luong, G. Baid, Z. Wang, V. Vasudevan, A. Ku, Y. Yang, B. K. Ayan, *et al.*, "Scaling autoregressive models for content-rich text-to-image generation," *arXiv preprint arXiv:2206.10789*, 2022.

[215] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge University Press, 2014.

[216] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizers," *IEEE Transactions on Information Theory*, vol. 38, no. 2, 1992, pp. 428–436.

[217] C. Zhang, J. Bütepage, H. Kjellström, and S. Mandt, "Advances in variational inference," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, 2018, pp. 2008–2026.

[218] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Transactions on Image Processing*, vol. 24, no. 8, 2015, pp. 2579–2591. DOI: 10.1109/TIP.2015.2426416.

[219] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018. DOI: 10.1109/CVPR.2018.00068.

[220] L. Zhou, C. Cai, Y. Gao, S. Su, and J. Wu, "Variational Autoencoder for Low Bit-rate Image Compression," in *Challenge on Learned Image Compression*, 2018.

[221] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004, pp. 600–612. DOI: 10.1109/TIP.2003.819861.

[222] Y. Zhu, Y. Yang, and T. Cohen, "Transformer-based transform coding," in *International Conference on Learning Representations*, 2021.

[223] J. Ziv, "On universal quantization," *IEEE Transactions on Information Theory*, vol. 31, no. 3, 1985, pp. 344–347.

[224] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Transactions on information theory*, vol. 23, no. 3, 1977, pp. 337–343.