

Distributed Compression in the Era of Machine Learning: A Review of Recent Advances

Ezgi Özyıldan and Elza Erkip

Dept. of Electrical and Computer Engineering

New York University

NY 11201, USA

{ezgi.ozyildan, elza}@nyu.edu

Abstract—Many applications from camera arrays to sensor networks require efficient compression and processing of correlated data, which in general is collected in a distributed fashion. While information-theoretic foundations of distributed compression are well investigated, the impact of theory in practice-oriented applications to this day has been somewhat limited. As the field of data compression is undergoing a transformation with the emergence of learning-based techniques, machine learning is becoming an important tool to reap the long-promised benefits of distributed compression. In this paper, we review the recent contributions in the broad area of learned distributed compression techniques for abstract sources and images. In particular, we discuss approaches that provide interpretable results operating close to information-theoretic bounds. We also highlight unresolved research challenges, aiming to inspire fresh interest and advancements in the field of learned distributed compression.

Index Terms—Distributed source coding, Wyner-Ziv coding, lossy compression, binning, neural networks, rate-distortion theory, learning.

I. INTRODUCTION

Data compression, also known as source coding, is a fundamental and extensively studied problem in both information theory and engineering disciplines. Shannon established that the *entropy* of a source serves as a fundamental limit of lossless data compression [1]. Achieving lossless compression near the Shannon limit requires that multiple samples of the information source to be compressed jointly, allowing for a small but vanishing probability of error as the blocklength increases. Counterpart to his lossless source coding theorem, Shannon's *lossy* source coding theorem [1], [2] provides theoretical limits on how efficiently one can compress an information source while allowing for a given level of *distortion* in the reconstruction. The so-called rate-distortion theory is particularly relevant in practical scenarios where the source is often continuous-valued, such as vectors representing image pixel intensities. For such sources, it is common in practice to initially quantize the data to a finite set of discrete values, thereby introducing a certain level of error. This gives rise to the challenge of lossy compression problems, where a trade-off must be struck between two competing goals: the entropy of the discretized representation (rate) and the error stemming from quantization (distortion).

The optimal trade-off in lossy compression is determined by information-theoretic rate-distortion function [3], which

is fully characterized only for specific source distributions. Given the inherent complexity of lossy compression, practical codes are often tailored to specific domains, such as image, audio, or video. In the context of lossy image compression, well-known codecs typically adopt a two-step process: first, a linear transformation followed by quantization, and then lossless compression using entropy coding, in the form of arithmetic coding, to efficiently transmit the induced quantized representation. This two-step approach is exemplified in widely used standards like JPEG [4] and JPEG 2000.

In recent years, there has been a significant surge in research on data-driven image compression algorithms enhanced by deep neural networks (DNN) [5]–[8]. These methods have demonstrated impressive performance, surpassing classical standard compression techniques like JPEG 2000 and BPG [9]. As argued in [5], [7], the effectiveness of DNN-aided compressors lies in their use of learned *nonlinear* transforms, aligned with the universal function approximation capability of artificial neural networks (ANNs) [10]. Unlike traditional methods that rely on handcrafted representations (e.g., discrete cosine transform as in the JPEG algorithm) to capture natural image statistics, neural compressors learn the underlying representations in an end-to-end fashion directly from the data itself.

Distributed source coding (DSC) addresses the challenge of efficiently compressing information from physically separated sources that are then decoded jointly. High correlation between distributed data streams naturally arises in many current practical setups, such as camera arrays and sensor networks. While the theory of DSC predicts substantial improvements in compression efficiency compared to the point-to-point setup [3], practical distributed compressors have not been fully developed to this day mainly due to the difficulty of handling complex correlations between information sources. We posit that learning-based methods could be helpful in this regard, given their quick adaptability to arbitrary data sources and to new modalities of multimedia.

The goal of this review article is to summarize recent results in the literature on DNN-aided DSC, with a specific focus on a special and simpler distributed lossy compression case where the side information is available only at the decoder side, also known as the Wyner-Ziv problem [11]. To understand the efficacy of learning-based techniques and the conditions under

which they perform competitively, we also discuss setups involving abstract sources in addition to practice-oriented ones that involve images.

The subsequent sections of this paper are structured as follows. Section II offers a brief overview of information-theoretic foundations of distributed data compression. Transitioning to neural/learned compression schemes, we recap the current state-of-the-art framework in Section III. Section IV delves into practical examples of various learned DSC scenarios, mostly involving images as information sources, and identifies gaps in the deep learning literature. Section V provides a compilation of recent neural distributed compression techniques considering decoder-only side information, with a focus on abstract sources. Lastly, Section VI outlines avenues for future research.

II. INFORMATION-THEORETICAL BACKGROUND

In this section, we provide the well-known result on lossless compression of separately encoded sources, often referred to as the *Slepian–Wolf* theorem. Subsequently, we shift our focus to distributed lossy compression, specifically considering the scenario with decoder-only side information that is known as the *Wyner–Ziv* theorem.

A. Lossless Distributed Source Coding (*Slepian–Wolf*)

The lossless distributed compression problem was first studied by Slepian and Wolf for two sources [12]. The main result is presented below. For achievability and converse parts of the theorem, we refer the readers to [13] as well as the original paper in [12].

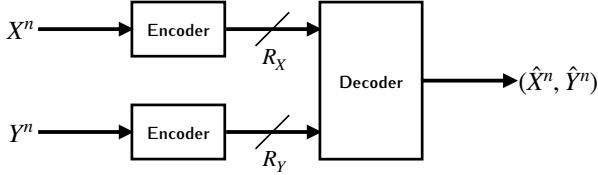


Fig. 1: Distributed lossless compression with two sources, also known as *Slepian–Wolf* coding.

Theorem 1. (*Slepian–Wolf Theorem [1973]*) *The optimal rate region for distributed lossless source coding of a pair of discrete memoryless sources (X, Y) is the set of pairs (R_X, R_Y) such that:*

$$R_X \geq H(X | Y), \quad (1)$$

$$R_Y \geq H(Y | X), \quad (2)$$

$$R_X + R_Y \geq H(X, Y). \quad (3)$$

The achievability part of the theorem involves a coding procedure using *random bins*, due to Cover [14]. As in many information theory problems, the achievability assumes asymptotically large blocklengths. The converse part invokes Fano's inequality and chain rule arguments. Note that the Slepian–Wolf theorem can be straightforwardly extended to distributed lossless compression with an arbitrary number of sources [3].

Let's now consider a corner point of the achievable Slepian–Wolf rate region, with rate $R_Y = H(Y)$, $R_X = H(X | Y)$. This operationally corresponds to the scenario that the source Y is losslessly available at the decoder side. Surprisingly, one can observe that an encoder that does not have access to the correlated side information can asymptotically achieve the same compression rate with the case when side information is available at both the encoder and the decoder. In other words, two distributed (or isolated) encoders can losslessly compress their sources as efficiently as if they were communicating with one another.

B. Lossy Source Coding with Side Information (*Wyner–Ziv*)

A general distributed lossy compression setup involving two encoders and two distortion measures was studied by Berger and Tung [15]. The rate region of the quadratic-Gaussian setup involving two encoders was explicitly characterized in [16]. Here, we only focus on the Wyner–Ziv setup as it captures the simplest distributed lossy compression setting, assuming decoder-only side information. For the achievability and converse parts of the theorem, we refer the readers to [13], as well as the original paper in [11].

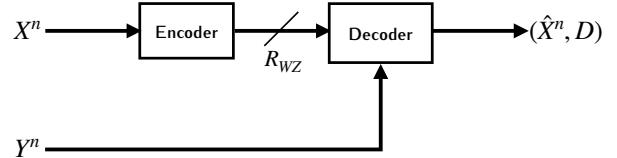


Fig. 2: Rate–distortion with side information, also known as *Wyner–Ziv* coding.

Theorem 2. (*Wyner–Ziv Theorem [1976]*) *Let (X, Y) be correlated sources, drawn i.i.d. $\sim p(x, y)$, and let $d(x, \hat{x})$ be a single-letter distortion measure. The rate–distortion function for X with side information Y available (non-causally) at the decoder side is as follows:*

$$R_{WZ}(D) = \min(I(X; U) - I(Y; U)), \quad (4)$$

where the minimization operation is over all conditional probability distribution functions $p(u|x)$, and all functions $g(u, y)$ such that $\mathbb{E}_{p(x,y)p(u|x)} d(x, g(u, y)) \leq D$.

The achievability part of the Wyner–Ziv theorem, which again relies on asymptotically large blocklengths, invokes the covering lemma, resulting in the rate of $I(X; U)$, followed by a random binning argument based on joint typicality, which yields the rate discount of $I(Y; U)$ in Eq. (4) [3]. This achievability, which is shown to be tight, assumes a Markov chain constraint $U - X - Y$.

In general, the optimal $p(u|x)$ and $g(u, y)$ are not known in closed form, and, similar to the point-to-point rate–distortion function, the Wyner–Ziv rate–distortion function in Eq. (4) has a closed-form expression only in a few special cases.

The literature contains several constructive codes for the Wyner–Ziv problem, which assume synthetic setups and specific correlation settings. Zamir et al. [17], [18] provide

optimal constructive mechanisms for binary and Gaussian sources, respectively, using nested linear and lattice codes. The authors of DIstributed Source Coding Using Syndromes (DISCUS) [19] formulate the Wyner–Ziv setup as a dual quantizer–channel coding problem in the finite blocklength regime. In this setting, the source is first quantized based on its marginal density, and the quantization codebook space is divided into *cosets*, determined by the virtual channel between the side information and the quantized source. Instead of transmitting the quantization index, the encoder sends the coset index, leading to further rate reduction. The decoder then disambiguates the coset index with the side information, recovers the quantization index, and estimates the source based on the deduced index and side information according to the distortion criterion. This systematic partitioning of the quantized source space with cosets resembles the random binning procedure seen in the proofs of the Slepian–Wolf and Wyner–Ziv theorems that consider the asymptotic blocklength regime [11], [12]. The proposed method achieves performance close to the asymptotic Wyner–Ziv rate–distortion bound, for Gaussian sources and linear correlation structures.

III. LEARNED DATA COMPRESSION

The general problem of optimal quantization in high-dimensional spaces, which is the case for many practical compression problems, is known to be intractable without further constraints [20]. Consequently, many current practical compressors rely on a technique called *transform coding* [21], which operates by linearly transforming the data vector into a continuous-valued representation, quantizing its elements independently, and subsequently encoding the induced discrete representation using a lossless entropy code [22].

The conventional wisdom once considered designing codes with nonlinear transforms as computationally infeasible. However, this premise has changed with the resurgence of ANNs, which can approximate arbitrary functions when equipped with the right set of parameters [10]. Leveraging these tools for function approximation in the context of source coding, the popular class of neural compression methods can now be framed through the perspective of *nonlinear transform coding* (NTC) [7].

Classic learned compression literature typically focuses on improving point-to-point (with no side information) rate–distortion performances of popular practical information sources, such as Kodak image dataset [23]. To this end, NTC seeks to find an *analysis* transform, g_a , that maps the space of source values to the *latent* space, and a *synthesis* transform, g_s , that maps from the latent space back into the source/reconstruction space. Typically, the latent variables, $\hat{\mathbf{u}}$, are quantized with uniform scalar quantization (i.e., rounding to the nearest integers) that yields $\hat{\mathbf{u}} = \lfloor \mathbf{u} \rfloor$. The quantized values are then entropy coded using likelihoods generated from an entropy model, $p_{\hat{\mathbf{u}}}(\hat{\mathbf{u}})$. The goal is to minimize the operational rate–distortion trade-off, in combination with

stochastic optimization methods (e.g., stochastic gradient descent), which can be compactly expressed by the following objective function:

$$\mathbb{E}_{\mathbf{x}}[-\log_2 p_{\hat{\mathbf{u}}}(\lfloor g_a(\mathbf{x}) \rfloor)] + \lambda \cdot \mathbb{E}_{\mathbf{x}}[d(\mathbf{x}, g_s(\lfloor g_a(\mathbf{x}) \rfloor))], \quad (5)$$

where $\lambda > 0$ controls the trade-off, and $d(\cdot, \cdot)$ denotes some distortion measure that quantifies the discrepancy of the reconstructed vector with respect to the input source. The transforms, g_a and g_s , as well as the entropy model $p_{\hat{\mathbf{u}}}(\hat{\mathbf{u}})$ are all parametrized using neural networks. In the NTC case, unlike traditional transform coding that involves handcrafted transformations, the analysis and synthesis transforms are equipped with nonlinear transforms (e.g., variations of the rectified linear unit activation function). These transforms, along with the entropy model, can be optimized for any differentiable distortion measure $d(\cdot, \cdot)$ in an end-to-end fashion by building onto the rate–distortion Lagrangian in Eq. (5).

Note that the rounding operation in Eq. (5) is non-differentiable, which poses a challenge for optimizing the encoder. This means that optimizing the parameters of the analysis function, g_a , directly via *backpropagation*, a crucial component of the learning process that neural methods rely on, is not possible. To address this, a common strategy is to substitute quantization with a differentiable approximation during training, typically through additive uniform noise, while retaining the quantization as is during inference time [5]. Recently, training lossy compressors without relying on any quantization approximations has gained popularity in the compression community [24], which draws inspiration from the literature on *reverse channel coding*, also known as *channel simulation* [25]. Instead of using additive uniform noise as a differentiable proxy during training, this line of work advocates for a differentiable channel that is also directly employed during inference time, in place of switching to hard quantization [24]. Although the mismatch between training and inference phases is eliminated as such by foregoing quantization, it remains unclear how well-suited such methods are to practical compression scenarios as the computational complexity of the proposed methods grows exponentially with the number of bits to be transmitted [24], [26]–[28].

IV. DISTRIBUTED COMPRESSION WITH REAL-LIFE SOURCES

As explained in Section III, conventional learned compression frameworks aim at enhancing point-to-point rate–distortion performance across various practical image datasets. While stereo image compression (jointly compressing the two images of a stereo pair by a single encoder) has already garnered attention from the compression community [29], [30], DSC approaches, concerning independent encoding and joint decoding of correlated images, have remained vastly unexplored until recently.

One of the early works that introduced a neural image compression framework fusing decoder-only side information is the approach proposed by Ayzik and Avidan [31]. In this work, they first reconstruct the original image using a learned

point-to-point compression framework, followed by additional refinement of the reconstructed image using side information image available at the decoder side. Hence, their method can be categorized as a “post-processing” technique of the original image with the help of side information, as it does not rely on the side information image to compress the original one.

Exploiting the side information in a more efficient way has since been investigated in [32]–[34]. Mital et al. employ cross-attention modules at the decoder side, in order to further align the features in [32], and explicitly model the common information between the input and side information in [33]. The most recent work in [34] uses a conditional Vector-Quantized Variational Autoencoder (VQ-VAE) [35], in lieu of having a learned entropy model as in NTC [7], and establishes a conceptual link between the decoder-only side information setting and an optimization objective based on a variant of the evidence lower bound. The empirical results provided in [31]–[34] demonstrate that incorporating the decoder-only side information, in the form of a correlated image, into the compressor significantly improves the operational rate-distortion performance across various practical stereo image datasets, affirming that the neural compressors can successfully exploit the correlation between the images. The benefit of decoder-only side information has also been studied in the context of data-driven joint source channel coding (JSCC) for correlated image transmission [36]. In this work, the authors demonstrated that the learned JSCC framework which exploits decoder-only side information outperforms both the learned point-to-point JSCC scheme [37] and also the separation-based transmission scheme that makes use of side information by adopting a two-step source and channel coding approach. Recently, a learning-based multi-view image coding that involves independent encoding of correlated images followed by joint decoding was proposed in [38].

Another line of research involving a learned DSC framework was explored in [39], where the authors introduced a *task-aware* distributed compression method focusing solely on downstream tasks, such as object detection, instead of reconstructing the original input source as in works that were discussed previously [31]–[34], [36], [38]. The proposed objective function in this case is not based on a rate–distortion Lagrangian (as in Eq. (5)), but rather on minimizing the final composite task loss subjected to a fixed latent dimension that the authors term as the *bandwidth*. Relying on *principal component analysis* as a means of linear dimensionality reduction, which is coupled with a neural autoencoder aiming to extract task-relevant representations to be compressed, the approach learns to allocate respective bandwidths to each source based on the importance of the task at hand.

While learning-based methods offer a pathway to develop practical distributed compressors, which has been challenging with earlier approaches (Section II), relying on neural networks as a “black box” in real-world scenarios may have its drawbacks. As the explicit source distributions emerging in these aforementioned practical setups are unknown, the optimal rate–distortion bounds in such distributed compres-

sion scenarios are not formally characterized. Consequently, a comprehensive assessment of the actual performance of these proposed learned distributed compressors and a thorough understanding of their deviation from the theoretical optimum remain challenging. By examining exemplary abstract sources, for which the theoretical optimal performance is known (e.g., quadratic-Gaussian model in the Wyner–Ziv case), it becomes more feasible to gauge the effectiveness of neural distributed compressors and to understand whether they actually recover the optimum, or near optimal, solutions. Considering such abstract settings allows us to devise well-informed architectural design choices, which can also potentially improve compression efficiency for high-dimensional sources such as images.

V. LEARNED WYNER–ZIV CODING WITH ABSTRACT SOURCES

Several works have investigated how learned compressors perform on abstract sources to gain a better understanding of the conditions that lead to optimal performance. The study in [7] validated that the NTC-based models can achieve the same compression performance as the optimal (point-to-point) entropy-constrained scalar quantizer for the Laplace distribution, whose performance was established by Sullivan [40]. Similarly, Wagner and Ballé [41] showed that the NTC-based schemes can also optimally compress sources that exhibit a low-dimensional manifold structure in a high-dimensional ambient space, such as a specific random process named the *sawbridge*. Notably, the recent work in [42] demonstrated that the same class of learned compressors are, surprisingly, unable to optimally compress certain complex sources exhibiting a circular topology, such as a particular random process named the *ramp*. The authors hypothesized that this is due to the NTC-based schemes having difficulty learning sufficiently steep functions, a phenomenon known as *spectral bias* [43] in learning literature. These analyses offer valuable insights into the fundamental learning capabilities of ANN-based compressors, and have the potential to exert a significant impact on image compression. However, all of these works only considered point-to-point information sources as inputs, without accounting for neither any presence of side information nor any other distributed compression setup in their formulation.

As one of the initial studies on a learned approach for the Wyner–Ziv problem for abstract sources, the authors of [44] considered a data-driven iterative optimization scheme for Gauss–Markov source and side information sequences with arbitrary memory. Building their model structure onto the Wyner–Ziv coding architectures proposed in [45], [46], which involve one-dimensional scalar quantization and scalar binning of quantization indices, this work showcased that a lower-complexity data-driven method can match the performance of a quantizer optimized using an exhaustive search.

Inspired by lossy compression techniques rooted in reverse channel coding, the recent work in [47] established a novel one-shot coding scheme for the Wyner–Ziv problem, where

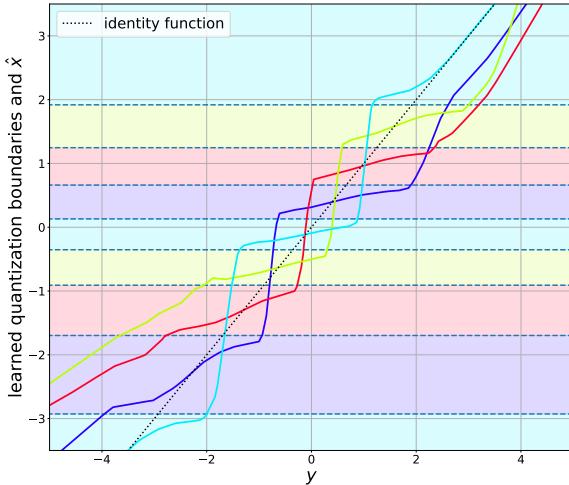


Fig. 3: Figure taken from [49]. Visualization (best viewed in color) of the learned encoder and decoder of the operational scheme proposed in [49], for the quadratic-Gaussian Wyner-Ziv setup. The dashed horizontal lines are quantization boundaries, and the colors between boundaries represent unique values of bin indices. The decoding function is depicted as separate plots for each value of bin index, using the same color assignment. Color coding reveals that the model learns discontiguous quantization bins.

the authors introduced an *importance sampling* based compression method that is a variant of random coding [28]. As such, this approach suggests entirely omitting the discretization/quantization step, in contrast to the NTC-based compressors (see Section III), and instead assumes the existence of a shared source of randomness available to both the encoder and the decoder, similar to the recently introduced Poisson functional representation lemma [48].

While the NTC-based compressors seem to be a good candidate for the Wyner-Ziv problem, our analysis in [50] revealed that surprisingly, they fail to exploit the side information most efficiently. We argued that this popular class of neural compressors are unable to recover any flexible and efficient binning schemes even considering a simple test case, where side information is the sign function of input realization. As an alternative to the NTC-based compressors, our work in [49]–[51] proposed a more generic learning-based algorithm, which represents the first unstructured entropy-constrained vector quantizer that makes use of side information. We demonstrated that such a learned compressor can rediscover some principles of the optimum theoretical solution for the Wyner-Ziv problem, such as optimal combination of the quantization index and side information at the decoder, for various exemplary sources. In particular, we illustrated that these learned distributed compressors are capable of exhibiting different types of interpretable binning mechanisms, such as periodic-like mappings (see Fig. 3) for the quadratic-Gaussian case and symmetric mappings for the sign function as side information, akin to the achievability part of the Wyner-Ziv theorem (see Section II-B). This provides empirical evidence that the ANN-based methods can learn constructive solutions very similar to some of the handcrafted compressors proposed

for the Wyner-Ziv problem, such as DISCUS [19], without requiring a priori knowledge of source statistics. We also justify how these proposed models are interpretable as operational schemes by picking a suitable entropy coding technique for each. These learned distributed compressors, coupled with a high-order entropy coding scheme (either classic or Slepian-Wolf) yield a better performance compared to the point-to-point rate-distortion functions, and in some cases achieve the optimal information-theoretic entropy-distortion bounds with side information [50]. We note that our proposed entropy-constrained compression frameworks are more amenable to practical use, in comparison to the scheme proposed in [47] that is based on communicating samples from continuous distributions. However, in their current form, the operational schemes in [49]–[51] inherit the limitations of vector quantization, that is naïvely increasing the codebook size renders these methods computationally infeasible.

VI. DISCUSSION

In this article, we have reviewed the literature on learned distributed compressors, with a specific focus on the schemes that assessed their compression performance on abstract sources. Recent findings demonstrate that ANN-based compressors, even without the explicit knowledge of source distributions, have emerged as competitive candidates for addressing open problems in distributed compression and offer promising constructive solutions to long-standing challenges in the field, such as adaptability to complex correlations.

Possible avenues for future work include analyzing the robustness of these neural schemes across various correlation structures and also extending them to fully distributed compression settings. Exploring this research direction has the potential to greatly enhance compression efficiency for transmitting practical signals in communication networks, such as multi-view image and video. The development of learned distributed compressors can also play a role in the practical design of fundamental building blocks of cooperative communications, such as the relay channel [52].

ACKNOWLEDGMENTS

We would like to thank Johannes Ballé for helpful discussions and valuable comments on this manuscript.

REFERENCES

- [1] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [2] ——, *Coding Theorems for a Discrete Source With a Fidelity Criterion*. Institute of Radio Engineers, International Convention Record, vol. 7, 1959., 1993, pp. 325–350.
- [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.
- [4] G. Wallace, “The JPEG still picture compression standard,” *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.
- [5] J. Ballé, V. Laparra, and E. P. Simoncelli, “End-to-end optimized image compression,” in *International Conference on Learning Representations*, 2017.

- [6] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, “Variational image compression with a scale hyperprior,” in *International Conference on Learning Representations*, 2018.
- [7] J. Ballé, P. A. Chou, D. Minnen, S. Singh, N. Johnston, E. Agustsson, S. J. Hwang, and G. Toderici, “Nonlinear transform coding,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 2, pp. 339–353, 2021.
- [8] D. C. Minnen, J. Ballé, and G. Toderici, “Joint autoregressive and hierarchical priors for learned image compression,” in *Neural Information Processing Systems*, 2018.
- [9] F. Bellard, “BPG image format,” <https://bellard.org/bpg/>, 2014.
- [10] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, “Multilayer feedforward networks with a nonpolynomial activation function can approximate any function,” *Neural Networks*, vol. 6, no. 6, pp. 861–867, Jan. 1993. [Online]. Available: [https://doi.org/10.1016/s0893-6080\(05\)80131-5](https://doi.org/10.1016/s0893-6080(05)80131-5)
- [11] A. Wyner and J. Ziv, “The rate–distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1 – 10, 1976.
- [12] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471 – 480, 1973.
- [13] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. USA: Cambridge University Press, 2012.
- [14] T. Cover, “A proof of the data compression theorem of Slepian and Wolf for ergodic sources (corresp.),” *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 226–228, 1975.
- [15] T. Berger, *Multiterminal Source Coding*. In G. Longo (Ed.), *The Information Theory Approach to Communications*. USA: New York, NY, USA: Springer-Verlag, 1977.
- [16] A. B. Wagner, S. Tavildar, and P. Viswanath, “Rate region of the quadratic gaussian two-encoder source-coding problem,” in *2006 IEEE International Symposium on Information Theory*, 2006, pp. 1404–1408.
- [17] R. Zamir and S. Shamai, “Nested linear/lattice codes for Wyner-Ziv encoding,” in *1998 Information Theory Workshop (Cat. No.98EX131)*, 1998, pp. 92–93.
- [18] R. Zamir, S. Shamai, and U. Erez, “Nested linear/lattice codes for structured multiterminal binning,” *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.
- [19] S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): Design and construction,” *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 626–643, 2003.
- [20] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. USA: Kluwer Academic Publishers, 1991.
- [21] V. Goyal, “Theoretical foundations of transform coding,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 9–21, 2001.
- [22] A. Netravali and J. Limb, “Picture coding: A review,” *Proceedings of the IEEE*, vol. 68, no. 3, pp. 366–406, 1980.
- [23] E. Kodak, “Kodak lossless true color image suite (PhotoCD PCD0992),” <http://r0k.us/graphics/kodak>.
- [24] E. Agustsson and L. Theis, “Universally quantized neural compression,” in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.
- [25] C. Bennett, P. Shor, J. Smolin, and A. Thapliyal, “Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem,” *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2637–2655, 2002.
- [26] M. Havasi, R. Peñar, and J. M. Hernández-Lobato, “Minimal random code learning: Getting bits back from compressed model parameters,” in *International Conference on Learning Representations*, 2019.
- [27] G. Flamich, M. Havasi, and J. M. Hernández-Lobato, “Compressing images by encoding their latent representations with relative entropy coding,” in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.
- [28] L. Theis and N. Y. Ahmed, “Algorithms for the communication of samples,” in *Proceedings of the 39th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 162. PMLR, 17–23 Jul 2022, pp. 21308–21328.
- [29] P. Merkle, K. Muller, A. Smolic, and T. Wiegand, “Efficient compression of multi-view video exploiting inter-view dependencies based on H.264/MPEG4-AVC,” in *2006 IEEE International Conference on Multimedia and Expo*, 2006, pp. 1717–1720.
- [30] J. Liu, S. Wang, and R. Urtasun, “DSIC: Deep stereo image compression,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, October 2019.
- [31] S. Ayzik and S. Avidan, “Deep image compression using decoder side information,” in *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII*. Berlin, Heidelberg: Springer-Verlag, 2020, p. 699–714.
- [32] N. Mital, E. Özyilkan, A. Garjani, and D. Gunduz, “Neural distributed image compression with cross-attention feature alignment,” in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision*, Jan 2023, pp. 2497–2506.
- [33] N. Mital, E. Özyilkan, A. Garjani, and D. Gündüz, “Neural distributed image compression using common information,” in *2022 Data Compression Conference*, 2022, pp. 182–191.
- [34] J. Whang, A. Nagle, A. Acharya, H. Kim, and A. G. Dimakis, “Neural distributed source coding,” 2023. [Online]. Available: <https://arxiv.org/abs/2106.02797>
- [35] A. van den Oord, O. Vinyals, and K. Kavukcuoglu, “Neural discrete representation learning,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, p. 6309–6318.
- [36] S. F. Yilmaz, E. Özyilkan, D. Gunduz, and E. Erkip, “Distributed deep joint source-channel coding with decoder-only side information,” to appear in *2024 IEEE International Conference on Machine Learning for Communication and Networking*, 2023. [Online]. Available: <https://arxiv.org/abs/2310.04311>
- [37] E. Bourtsoulatz, D. Burth Kurka, and D. Gunduz, “Deep joint source-channel coding for wireless image transmission,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, p. 567–579, Sep. 2019.
- [38] X. Zhang, J. Shao, and J. Zhang, “LDMIC: Learning-based distributed multi-view image coding,” in *The Eleventh International Conference on Learning Representations*, 2023.
- [39] P. Li, S. K. Ankireddy, R. Zhao, H. N. Mahjoub, E. M. Pari, U. Topcu, S. P. Chinchali, and H. Kim, “Task-aware distributed source coding under dynamic bandwidth,” in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- [40] G. Sullivan, “Efficient scalar quantization of exponential and Laplacian random variables,” *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1365–1374, 1996.
- [41] A. B. Wagner and J. Ballé, “Neural networks optimally compress the sawbridge,” in *2021 Data Compression Conference*, 2021, pp. 143–152.
- [42] S. Bhadane, A. B. Wagner, and J. Ballé, “Do neural networks compress manifolds optimally?” in *2022 IEEE Information Theory Workshop*, 2022, pp. 582–587.
- [43] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, “On the spectral bias of neural networks,” in *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, 2019, pp. 5301–5310.
- [44] E. Domanovitz, D. Severo, A. Khisti, and W. Yu, “Data-driven optimization for zero-delay lossy source coding with side information,” in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 5203–5207.
- [45] E. Tuncel, “Predictive coding of correlated sources,” in *Information Theory Workshop*, 2004, pp. 111–116.
- [46] X. Chen and E. Tuncel, “Low-delay prediction- and transform-based Wyner-Ziv coding,” *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 653–666, 2011.
- [47] B. Phan, A. Khisti, and C. Louizos, “Importance matching lemma for lossy compression with side information,” 2024. [Online]. Available: <https://arxiv.org/abs/2401.02609>
- [48] C. T. Li and V. Anantharam, “A unified framework for one-shot achievability via the Poisson matching lemma,” *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2624–2651, May 2021.
- [49] E. Özyilkan, J. Ballé, and E. Erkip, “Learned Wyner-Ziv compressors recover binning,” in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 701–706.
- [50] E. Özyilkan, J. Ballé, and E. Erkip, “Neural distributed compressor discovers binning,” 2023. [Online]. Available: <https://arxiv.org/abs/2310.16961>
- [51] E. Özyilkan, J. Ballé, and E. Erkip, “Neural distributed compressor does binning,” in *ICML 2023 Workshop Neural Compression: From Information Theory to Applications*, 2023.
- [52] T. Cover and A. Gamal, “Capacity theorems for the relay channel,” *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.