



HAL
open science

Learning on JPEG-LDPC Compressed Images: Classifying with Syndromes

Ahcen Aliouat, Elsa Dupraz

► **To cite this version:**

Ahcen Aliouat, Elsa Dupraz. Learning on JPEG-LDPC Compressed Images: Classifying with Syndromes. 32nd European Signal Processing Conference (EUSIPCO), Aug 2024, Lyon, France. hal-04682219

HAL Id: hal-04682219

<https://imt-atlantique.hal.science/hal-04682219v1>

Submitted on 30 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning on JPEG-LDPC Compressed Images: Classifying with Syndromes

Ahcen Aliouat and Elsa Dupraz

IMT Atlantique, CNRS UMR 6285, Lab-STICC, Brest, France

Abstract—In goal-oriented communications, the objective of the receiver is often to apply a Deep-Learning model, rather than reconstructing the original data. In this context, direct learning over compressed data, without any prior decoding, holds promise for enhancing the time-efficient execution of inference models at the receiver. However, conventional entropic-coding methods like Huffman and Arithmetic break data structure, rendering them unsuitable for learning without decoding. In this paper, we propose an alternative approach in which entropic coding is realized with Low-Density Parity Check (LDPC) codes. We hypothesize that Deep Learning models can more effectively exploit the internal code structure of LDPC codes. At the receiver, we leverage a specific class of Recurrent Neural Networks (RNNs), specifically Gated Recurrent Unit (GRU), trained for image classification. Our numerical results indicate that classification based on LDPC-coded bit-planes surpasses Huffman and Arithmetic coding, while necessitating a significantly smaller learning model. This demonstrates the efficiency of classification directly from LDPC-coded data, eliminating the need for any form of decompression, even partial, prior to applying the learning model.

Index Terms—Goal-oriented communications, Image coding for Machines, Entropic coding, LDPC codes, RNN.

I. INTRODUCTION

Over the past few years, learning on coded data has emerged as an important area of research due to the increasing amount of images and videos that need to be processed in current and next-generation networks. This emerging paradigm is often referred to as Image and Video Coding for Machines (VCM) [1]. A key desired feature of this paradigm is that the receiver may apply the considered learning task without any prior decoding of the data. Such a feature would present a compelling advantage for the time-efficient execution of fundamental learning tasks such as classification, segmentation, or recommendation, directly on compressed data.

In this context, this work considers the important case of classification over compressed images. Recently, full end-to-end Deep Learning approaches have been proposed for this purpose, where the receiver applies learning on a well-designed latent space [2]–[4]. However, these approaches often lack the capability to reconstruct images when necessary, and they deviate from compliance with established standards for image compression, such as JPEG. Alternatively, in [5]–[8], learning methods are applied over JPEG-compressed images. Yet, these methods necessitate partial decoding steps,

such as reconstructing the Discrete Cosine Transform (DCT) coefficients, before the application of the learning task. In addition, some other studies, such as [9], aim to apply machine vision or object segmentation over compressed videos. These approaches encounter similar limitations as the methods proposed for images, either lacking compatibility with prevalent video coding standards like MPEG or HEVC or mandating partial decoding before task execution.

Interestingly, [10], [11] make a step toward overcoming these limitations by investigating classification directly over entropy-coded images. These works encompass usual entropy coding techniques, such as Huffman and Arithmetic coding. They demonstrate that established Deep Learning models for image classification, such as ResNet or VGG, can be directly applied to entropy-coded images without necessitating any partial decoding. However, this comes at a cost of reduced classification accuracy, attributed to the disruptive nature of conventional entropy coders on image structure, notably compromising spatial closeness between pixels.

In this paper, our objective is to explore alternative entropy-coding techniques that offer greater relevance from a learning standpoint while retaining the capability for data reconstruction. Specifically, we concentrate on Low-Density Parity Check (LDPC) codes, a family of channel codes that have also demonstrated efficiency for entropic coding at large [12] as well as in distributed source coding [13], and the context of 360-degree images [14]. Our rationale for examining this unconventional entropic-coding approach stems from the key idea that its inherent structure may be leveraged more efficiently by Deep Learning models for image classification.

We consider standard JPEG compression, with its typical steps such as DCT and quantization, while substituting the final entropic coder with LDPC coding applied over bit planes. The later step generates distinct codewords, termed syndromes, for each bit-plane. At the receiver, we employ Recurrent Neural Networks (RNN), specifically the Gated Recurrent Unit (GRU) model, to classify images based on their LDPC-coded bit-plane syndromes. This choice of RNN draws inspiration from the work of [15], which initially demonstrated the ability of RNN structures in approaching Maximum-Likelihood channel decoding of linear block codes.

Our experimental results demonstrate the feasibility of applying image classification directly over LDPC-coded syndromes without the need for any partial decoding. Furthermore, this approach yields superior accuracy compared to classification over Huffman or Arithmetic entropic coding,

This work has received a French government support granted to the Cominlabs excellence laboratory and managed by the National Research Agency in the ‘Investing for the Future’ program under reference ANR-10-LABX-07-01.

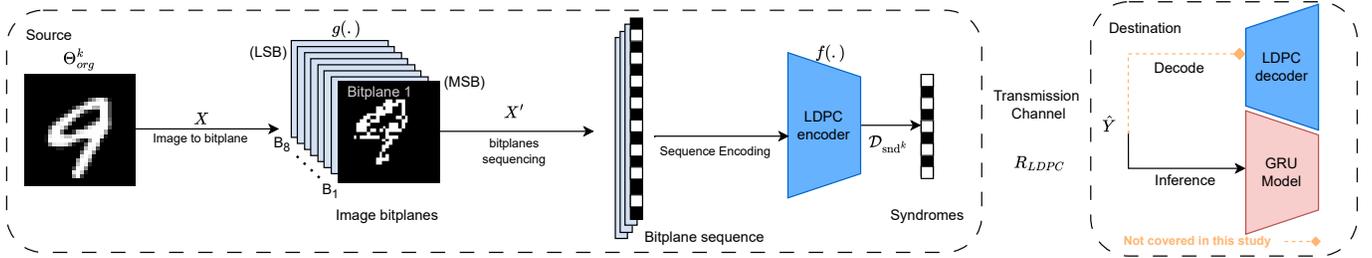


Fig. 1: **First setup:** Syndromes obtained from the LDPC-coded bit-planes are fed as input of a GRU model for classification

while employing models of significantly reduced complexity. This approach has the potential to streamline the learning pipeline by eliminating the need for decoding. Additionally, it offers a new perspective on the synergy between compression techniques and Deep Learning models.

The remainder of the paper is organized as follows. Section II describes JPEG-compressed images with entropic-coder based on LDPC codes. Section III introduces the proposed classification method. Section IV details the experiments conducted. Section V presents and discusses the results.

II. JPEG CODING WITH LDPC CODES

This section briefly describes JPEG compression and shows how LDPC codes can be used for bit-plane entropy-coding in this context.

A. JPEG Coding

The conventional JPEG compression workflow starts with a DCT transform, followed by a quantization step. Subsequently, Zig-Zag scanning as well as Run-Length encoding are applied to rearrange the DCT coefficients, enhancing data suitability for entropic coding. Finally, Huffman or Arithmetic entropy-coding methods are employed to minimize the size of the compressed data based on symbol probability. For a comprehensive understanding of JPEG compression, readers are referred to [16].

The works presented in [10], [11] show that image classification over Huffman or Arithmetic coded data is possible. However, this comes at the price of a significant loss in classification accuracy, due to the fact that such entropic coding methods break the semantic and closeness structure of the pixels. Nevertheless, these properties are essential for usual Deep Learning models dedicated to image classification. This motivates the need to explore other entropic-coding methods that are more appropriate for learning on coded data.

B. JPEG Coding with LDPC Codes

In this work, we propose to use LDPC codes for entropic coding. Indeed, we hypothesize that the LDPC coder may be more relevant for image classification, since it may better preserve the structure of the images features, through the LDPC code structure. Further, it was shown in [12] that LDPC codes are appropriate for entropic-coding, at the price of a slightly higher coding rate. To effectively implement this approach, we now describe the specific requirements

for applying binary LDPC codes to n non-binary pixels or quantized DCT coefficients. First, we transform the successive non-binary symbols into I bit-planes, where the bit-plane with the most significant bits (MSBs) contains the majority of an image's information. Previous studies have demonstrated that LDPC bit-plane encoding is practical and does not significantly impact the coding rate compared to non-binary LDPC encoding [14].

Next, let H be the binary parity check matrix of size $m \times n$, where ($m < n$) of a binary LDPC code [13], [17]. If H is full rank, the source coding rate is defined as $R = m/n$. The parity check matrix H can be equivalently described as a Tanner graph, which is bipartite in nature, connecting n source nodes with m syndrome nodes. In our scheme, each bit plane \mathbf{x}_i , $i \in \llbracket 1, I \rrbracket$ has the same length n and it is compressed by an LDPC code using the following formula [13]:

$$\mathbf{s}_i = H\mathbf{x}_i \quad (1)$$

where \mathbf{s}_i is a binary sequence of length m called the syndrome. In our scheme, the I syndromes will be sent to the decoder as the compressed information. At the receiver, these syndromes will be used either for data reconstruction as in [14], or as inputs to a learning model.

III. CLASSIFICATION OVER LDPC-CODED IMAGES

In this section, we describe the GRU model we consider for classification over LDPC-coded images.

A. GRU for compressed image sequences classification

The GRU model simplifies the Long Short Term Memory (LSTM) framework by combining the forget and input gates into a single update gate and introducing a reset gate, thereby reducing complexity without significantly sacrificing performance. Unlike LSTMs, GRUs manage the flow of information without using a separate memory cell, simplifying the processing of temporal sequences. Consider a GRU with J units. The activation variable h_t^j for the unit j of the GRU at time t is computed as a weighted average between the previous activation variable h_{t-1}^j and the candidate activation variable \tilde{h}_t^j , as follows:

$$h_t^j = (1 - z_t^j) h_{t-1}^j + z_t^j \tilde{h}_t^j. \quad (2)$$

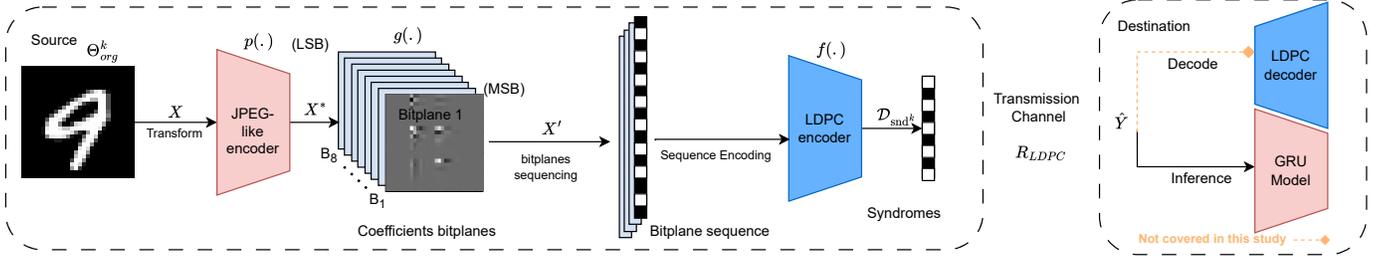


Fig. 2: **Second setup**: Syndromes of the DCT-LDPC coefficients bit-planes are fed as inputs of a GRU model for classification

This equation simplifies the information update process within the unit. The update gate z_t^j determines the extent to which the unit updates its activation and is calculated using:

$$z_t^j = \sigma(W_z \mathbf{u}_t + U_z \mathbf{h}_{t-1}), \quad (3)$$

where \mathbf{u}_t is the input signal, and $\sigma(\cdot)$ is the sigmoid activation function. Furthermore, the candidate activation \tilde{h}_t^j is computed like the update gate, but incorporates the reset gate r_t^j to potentially discard information that is no longer relevant:

$$\tilde{h}_t^j = \tanh(W \mathbf{u}_t + U(r_t^j \odot \mathbf{h}_{t-1})), \quad (4)$$

where \odot denotes an element-wise multiplication. Here, r_t^j represents a set of reset gates, and W_z , W , U_z , and U are the weights associated with the update gate and the candidate activation respectively.

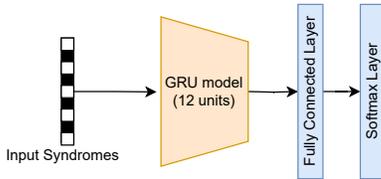


Fig. 3: Considered learning model for classification

The GRU model introduced by Cho et al. [18], includes a reset gate r_t^j that allows for the selective forgetting of prior information. This capability simulates the scenario of encountering the initial element of an input sequence whenever the reset gate is inactive ($r_t^j = 0$). Such a mechanism enables the model to dynamically adapt its memory focus. The reset gate's operation is governed by the equation:

$$r_t^j = \sigma(W_r \mathbf{u}_t + U_r \mathbf{h}_{t-1}). \quad (5)$$

The update gate z_t^j plays a pivotal role in modulating the influence of previous states on the current state, allowing the model to balance between short-term and long-term dependencies. This balance is achieved through the combined activity of the reset and update gates. The model involving the considered GRU architecture is summarized in Fig. 3.

B. Learning setups

We consider an initial dataset of images Θ_{org} , and investigate two setups. In the first setup, we do not consider JPEG

compression: for each image $\theta_k \in \Theta_{org}$, the pixels are directly transformed into bit-planes and encoded using LDPC codes. The I resulting syndromes $\mathbf{s}_{k,i}$, $i \in \llbracket 1, I \rrbracket$, form a new data $\mathbf{d}_k^{(1)}$. We refer to the corresponding new dataset as $\mathcal{D}_{snd}^{(1)}$, where the generation of one element $\mathbf{d}_k^{(1)}$ of the dataset $\mathcal{D}_{snd}^{(1)}$ can be described as follows:

$$\mathbf{d}_k^{(1)} = f(g(\theta_k)). \quad (6)$$

Here, the function $g(\cdot)$ represents the bit-plane constructor, and the function $f(\cdot)$ is the LDPC source encoding according to (1).

For the second setup, we mimic JPEG compression and first apply the DCT transform followed by standard JPEG quantization. We do not consider Zig-Zag scanning nor RLE (since they were specific to Huffman coding) and directly transform the DCT coefficients into bit-planes. In this case, each element $\mathbf{d}_k^{(2)}$ of the new dataset $\mathcal{D}_{snd}^{(2)}$ is obtained as

$$\mathbf{d}_k^{(2)} = f(g(p(\theta_k))). \quad (7)$$

The function $p(\cdot)$ represents the JPEG-like operation, which carries out 8×8 DCT transform and quantization. For the quantization step, we employ the standard JPEG quantization matrix. The quality factor is fixed as $QF = 95$.

TABLE I: Hyper-parameters for the learning model

parameter	Value	parameter	Value
Learning rate	0.001	Dropout rate	0.2
No. epochs	30	L2 Regularization	0.002
No. GRU Units	12 (32)	Activation function	Softmax
Batch size	64	Optimizer	Adam

The two datasets $\mathcal{D}_{snd}^{(1)}$ and $\mathcal{D}_{snd}^{(2)}$ will serve as inputs to two GRU learning models for training. In what follows, we aim to quantitatively evaluate the efficiency of applying such classification models directly onto LDPC-coded syndromes, without any prior decoding.

IV. DATA AND EXPERIMENTS

We now describe the considered datasets, as well as the experimental setups for our performance evaluation. The experiments are divided into two parts, aligned with the considered two setups. The first set of experiments evaluates the ability of the GRU model to learn from the LDPC syndromes of one or multiple bit-planes of the image. The second set of

TABLE II: Classification accuracy of the GRU on LDPC coded data compared to the state-of-the-art for multiple datasets.

Dataset	Model	No coding		Coding on Orig. (Setup1)			Coding on JPEG (Setup2)				
		None	None MSB	Huff [10]	Arith [10]	LDPC	JPEG [10]	DCT -tr. [8]	J-L 8bp	J-L MSB	J-L MSB+1bp
MNIST	GRU12(proposed)	0.9439	0.8842	-	-	0.8192	-	-	0.9060	0.6548	0.8791
	GRU32(proposed)	0.9799	0.9154	-	-	0.8556	-	-	0.9237	0.6843	0.8849
	VGG11 [10]	0.9891	-	0.8323	0.6313	-	-	-	-	-	-
	URESNET18 [10]	0.9875	-	0.7450	0.5949	-	-	-	-	-	-
	FullyConn [8]	0.9200	-	-	-	-	-	0.9000	-	-	-
Fashion-MNIST	GRU12	0.8616	0.8052	-	-	0.8166	-	-	0.8332	0.5222	0.8325
	GRU32	0.8750	0.8314	-	-	0.8306	-	-	0.8434	0.5395	0.8414
	VGG11 [10]	0.9018	-	0.7634	0.6898	-	-	-	-	-	-
	URESNET18 [10]	0.8497	-	0.6862	0.6116	-	-	-	-	-	-
	FullyConn [8]	0.9200	-	-	-	-	-	0.9000	-	-	-
YCIFAR-10	GRU12	0.3127	0.3249	-	-	0.4023	-	-	0.4234	0.1350	0.3537
	GRU32	0.3596	0.3560	-	-	0.4023	-	-	0.4316	0.1403	0.3544
	VGG11 [10]	0.5657	-	0.3606	0.2976	-	0.3245	-	-	-	-
	URESNET18 [10]	0.3836	-	0.2591	0.2432	-	-	-	-	-	-
	FullyConn [8]	0.3800	-	-	-	-	-	0.3000	-	-	-

*J-L: JPEG-LDPC, MSB+1bp: Sign bit-plane of the DCT coefficients + the next bit-plane after the sign bit-plane

experiments investigates the model classification performance over the considered JPEG-like chain where Huffman coding is replaced by LDPC coding.

A. Datasets

The tests are conducted on three standard datasets: MNIST [19], Fashion-MNIST [20], and CIFAR-10 [21]. For MNIST and Fashion-MNIST, images are grayscale-converted and processed as outlined in Section III-B. While CIFAR images are converted to YCrCb, only the Y channel is considered for coding, benefiting from its ability to retain most information from the original RGB images.

B. Code parameters

In our experiments, we consider a regular (3,6)-LDPC parity-check matrix H of size 1024×512 with rate $1/2$. In order to perform a consistent evaluation, the same parity-check matrix is used in all the experiments. Accordingly, the sizes of the images of the three datasets are resized from 28×28 to 32×32 by padding zeros at the end of the columns and rows. This adjustment also facilitates the application of the 8×8 DCT transform.

C. Model parameters

In the experiments, the GRU model’s input size is set to $512 \times I$ when applied over the I LDPC-coded syndromes, and to $1024 \times I$ when applied over the original I bit-planes. The latter corresponds to learning from the uncoded data, and it is considered for comparison purposes. For GRU model training, datasets are divided into training/validation sets at ratios of 84%-16% for MNIST and Fashion-MNIST, and 86%-14% for YCIFAR-10. The used hyper-parameters are shown in Table I.

D. Rate gain

We propose to measure the achieved gain while considering LDPC bit-planes coding using the following equation:

$$\Gamma = \frac{R_{LDPC}}{N_{bp}}. \quad (8)$$

Where R_{LDPC} is the LDPC coding rate, and N_{bp} is calculating as the number I of considered bit-planes divided by the maximum number of bit-planes η_{bp} , ie: $N_{bp} = I/\eta_{bp}$.

V. RESULTS AND DISCUSSION

In this section, the GRU model’s performance is reported in terms of classification accuracy and model complexity.

A. Accuracy comparison

Table II shows the classification accuracy over LDPC-coded syndrome under different conditions, including: (i) without coding for reference, (ii) by applying LDPC coding directly on the original image bit-planes (Setup 1), (iii) with JPEG-like compression (Setup 2). Results are compared to those of Huffman and Arithmetic coding methods from [10], and to the "truncated DCT" technique of [8]. For Setup 1, learning is applied over the $I = 8$ bit-planes. For Setup 2, we also report on the impact of the number of bit-planes on the model performances, when learning over 1 bit-plane (J-L MSB), when considering 2 bit-planes (J-L MSB + 1 bp), and when considering the 8 bit planes (DL). It is worth noting that the results of [11] for Huffman and Arithmetic coding also considered learning over the 8 bit-planes. It is remarkable that Setup 1 significantly outperforms other methods in accurately classifying entropic-coded images. We observe a performance improvement of about 15% for the CIFAR-10, 10% for Fashion-MNIST, and MNIST, compared to the best results of the previously mentioned studies. The results also reveal that Setup 2, which combines DCT and

quantization with LDPC coding, surpasses the performance of *Setup 1*, which solely relies on LDPC coding, as shown in Table II. This superiority most probably stems from the DCT’s efficacy in feature representation through signal frequencies. Consequently, integrating DCT with LDPC coding enhances syndrome classification accuracy, highlighting the proposed GRU model’s performance improvements due to the synergistic effect of these techniques. Furthermore, examining the impact of utilizing limited bit-planes (Columns J-L-MSB and J-L-MSB+bp) reveals their unexpected efficiency. While learning from only the MSB bit-plane ($I = 1$) degrades the performance, learning from the MSB bit-plane and one additional bit-plane ($I = 2$) yields results nearly equivalent to using all original bit-planes. This finding shows the feasibility of compressing data through selective bit-plane usage without compromising learning efficacy.

B. Accuracy-Complexity balance

According to the complexity analysis shown in Table III, our proposed method requires only 19k learnable parameters to outperform 2D-CNN, 1D-CNN, and “1024 Fully Connected” models. When using GRU on compressed images as feature sequences instead of image pixels (or coefficients), complexity reduction is about 1600%, which is significant. This confirms the inherent ability of LDPC codes to preserve the structure of the data, which significantly simplifies the learning process for the model, compared to Huffman and Arithmetic coding.

Setup	Best accuracy	No. of learnables
Huffman on VGG [10]	0.8323	138M
Huffman on Resnet [10]	0.6313	60M
DCT on FullyConnected [8]	0.9000	1M
JPEG-like LDPC on 12 units GRU	0.9060	19k
JPEG-like LDPC on 32 units RGU	0.9237	52.6k

TABLE III: Reported Accuracy vs weights for MNIST dataset

C. Compression gain

The same LDPC parity-check matrix, with a rate $R_{LDPC} = 1/2$ was used in all the experiments. By encoding only one bit-plane (MSB) ($N_{bp} = 1/8$), the compression ratio can reach up to 0.5 bits per pixel. On the other hand, encoding a maximum of 8 bit-planes yields a compression ratio of 4 bits per pixel. It is noteworthy that incorporating a pruning step before LDPC coding can significantly improve the compression gain. However, its impact on accuracy also warrants investigation.

VI. CONCLUSION

In this paper, we have proposed a lightweight learning model based on GRU for learning from LDPC source entropy-coded data. We have demonstrated the efficiency of this approach in terms of classification accuracy and model complexity. As a result, integrating LDPC codes within a JPEG-like chain better preserves the data structure, and opens the way to learning over coded data, without any prior decoding. Future work will include evaluating the impact on the learning

performance of JPEG quality factors, the number of bit planes used for learning, and pruning strategies, which all change the compression ratio. We will also investigate the performance of various regular and irregular LDPC codes in such learning contexts and explore the idea of designing LDPC codes specifically for learning from syndromes.

REFERENCES

- [1] L. Duan, J. Liu, W. Yang, T. Huang, and W. Gao, “Video coding for machines: A paradigm of collaborative compression and intelligent analytics,” *IEEE Transactions on Image Processing*, vol. 29, pp. 8680–8695, 2020.
- [2] 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.
- [3] Z. Duan, Z. Ma, and F. Zhu, “Unified architecture adaptation for compressed domain semantic inference,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.
- [4] Y. Zhang, C. Jia, J. Chang, and S. Ma, “Machine perception-driven image compression: A layered generative approach,” *arXiv preprint arXiv:2304.06896*, 2023.
- [5] P. R. Hill and D. R. Bull, “Transform and bitstream domain image classification,” *arXiv preprint arXiv:2110.06740*, 2021.
- [6] M. Pistono, G. Coatrieux, J.-C. Nunes, and M. Cozic, “Training machine learning on jpeg compressed images,” in *DCC*, p. 388, 2020.
- [7] J. Liu, H. Sun, and J. Katto, “Improving multiple machine vision tasks in the compressed domain,” in *2022 26th International Conference on Pattern Recognition (ICPR)*, pp. 331–337, 2022.
- [8] D. Fu and G. Guimaraes, “Using compression to speed up image classification in artificial neural networks,” *Technical report*, 2016.
- [9] L. D. Chamain, F. Racapé, J. Bégin, A. Pushparaja, and S. Feltman, “End-to-end optimized image compression for multiple machine tasks,” *arXiv preprint arXiv:2103.04178*, 2021.
- [10] R. Piau, T. Maugey, and A. Roumy, “Learning on entropy coded images with cnn,” in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023.
- [11] R. Piau, T. Maugey, and A. Roumy, “Predicting cnn learning accuracy using chaos measurement,” in *HMM-QoE 2023 (2023 IEEE International Conference on Acoustics, Speech and Signal Processing Satellite Workshop)*, 2023.
- [12] G. Caire, S. Shamai, and S. Verdú, “Noiseless data compression with low-density parity-check codes,” *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 66, pp. 263–284, 2004.
- [13] A. D. Liveris, Z. Xiong, and C. N. Georghiades, “Compression of binary sources with side information at the decoder using ldpc codes,” *IEEE communications letters*, vol. 6, no. 10, pp. 440–442, 2002.
- [14] F. Ye, N. M. Bidgoli, E. Dupraz, A. Roumy, K. Amis, and T. Maugey, “Bit-plane coding in extractable source coding: Optimality, modeling, and application to 360° data,” *IEEE Communications Letters*, vol. 25, no. 5, pp. 1412–1416, 2021.
- [15] A. Benmaman, Y. Choukroun, and P. Kisilev, “Deep learning for decoding of linear codes—a syndrome-based approach,” in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 1595–1599, IEEE, 2018.
- [16] G. Wallace, “The jpeg still picture compression standard,” *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.
- [17] F. Ye, E. Dupraz, Z. Mheich, and K. Amis, “Optimized rate-adaptive protograph-based ldpc codes for source coding with side information,” *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 3879–3889, 2019.
- [18] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” *arXiv preprint arXiv:1409.1259*, 2014.
- [19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [20] H. Xiao, K. Rasul, and R. Vollgraf, “Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms,” *arXiv preprint arXiv:1708.07747*, 2017.
- [21] A. Krizhevsky, G. Hinton, *et al.*, “Learning multiple layers of features from tiny images,” 2009.