Joint Source-Channel Coding of Images with (not very) Deep Learning

David Burth Kurka and Deniz Gündüz
Department of Electrical and Electronic Engineering, Imperial College London, London, UK
{d.kurka, d.gunduz}@imperial.ac.uk

***Abstract*—Almost all wireless communication systems today are designed based on essentially the same digital approach, that separately optimizes the compression and channel coding stages. Using machine learning techniques, we investigate whether end-to-end transmission can be learned from scratch, thus using joint source-channel coding (JSCC) rather than the separation approach. This paper reviews and advances recent developments on our proposed technique, *deep-JSCC*, an autoencoder-based solution for generating robust and compact codes directly from images pixels, being comparable or even superior in performance to state-of-the-art (SoA) separation-based schemes (BPG+LDPC). Additionally, we show that deep-JSCC can be expanded to exploit a series of important features, such as graceful degradation, versatility to different channels and domains, variable transmission rate through successive refinement, and its capability to exploit channel output feedback.**

I. Introduction

Wireless communication systems have traditionally followed a modular model-based design approach, in which highly specialized blocks are designed separately based on expert knowledge accumulated over decades of research. This approach is partly motivated by Shannon's *separation theorem* [1], which gives theoretical guarantees that the separate optimization of source compression and channel coding can, in the asymptotic limit, approach the optimal performance. In this way, we have available today highly specialized source codes, e.g., JPEG2000/BPG for images, MPEG-4/WMA for audio, or H.264 for video, to be used in conjunction with near-capacity-achieving channel codes, e.g., Turbo, LDPC, polar codes.

However, despite its huge impact, optimality of separation holds only under unlimited delay and complexity assumptions; and, even under these assumptions, it breaks down in multi-user scenarios [2], [3], or non-ergodic source or channel distributions [4], [5]. Moreover, unconventional communication paradigms have been emerging, demanding extreme end-to-end low latency and low power (e.g., IoT, autonomous driving, tactile Internet), and operating under more challenging environments that might not follow the traditional models (e.g., channels under bursty interference).

In light of above, our goal is to rethink the problem of wireless communication of lossy sources by using machine learning techniques, focusing particularly on image transmission. For this, we replace the modular *separation-based* design with a single neural network component for encoder and decoder (see Fig.1 for an illustrative diagram), thus performing JSCC, whose parameters are trained from data, rather than being designed. Our solution, the *deep-JSCC*, is applied to the problem of image transmission and can learn strictly from data in an unsupervised manner, as we model our system as an autoencoder [6], [7] with the communication channel incorporated as a non-trainable layer. This approach is motivated by the recent developments in machine learning through the use of deep learning (DL) techniques, and their applications to communication systems in recent years [8]. Autoencoders, in particular, due to the similarity between its architecture and digital communication systems [9], [10] have been used in related problems and pushing the boundaries of communications [11]–[16]. The use of DL for the separate problems of channel coding and image compression have been showing promising results, achieving performance in some cases superior to handcrafted algorithms [17], [18]. We show, however, that by performing JSCC, we can further improve the end-to-end performance.

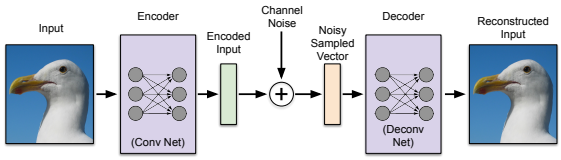


Fig. 1. Machine learning based communication system.

This paper reviews different features that were shown to be achieved with deep-JSCC, namely (a) performance comparable or superior to SoA separation-based schemes; (b) graceful degradation upon deterioration of channel conditions [19]; (c) versatility to adapt to different channels and domains [19]; (d) capacity of successive refinement [20] and (e) ability to exploit channel output feedback in order to improve the communication [21]. Thus, deep-JSCC presents itself as a powerful solution for the transmission of images, enabling communications with excellent performance while achieving low-delay and low-energy, being robust to channel changes, and allowing small and flexible bandwidth transmissions, thus advancing the field of communications by improving existing JSCC and separation-based methods.

This work was supported by the European Research Council (ERC) through project BEACON (No. 677854).

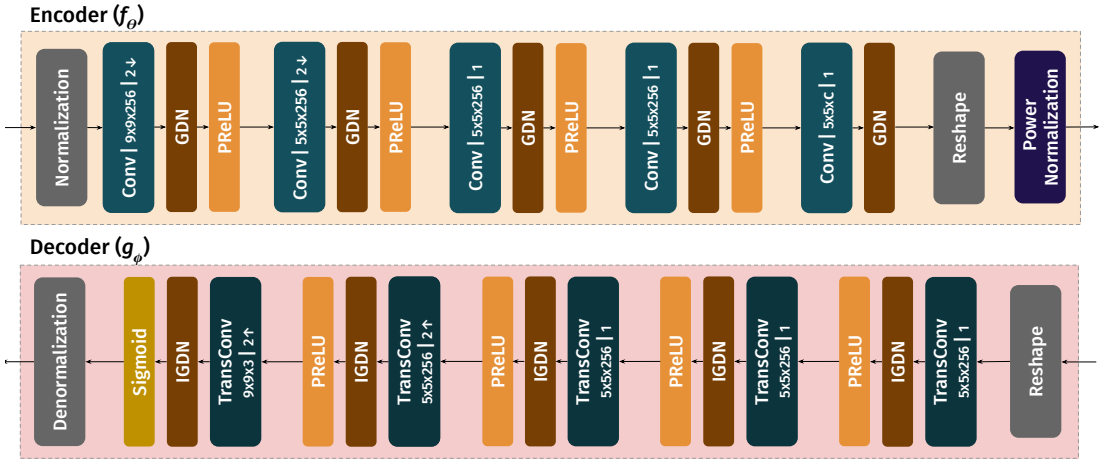


Fig. 2. Encoder and decoder architectures used in experiments.

II. Problem Formulation and Model Description

Consider an input image with height H, width W and C color channels, represented as a vector of pixel intensities $\boldsymbol{x} \in \mathcal{R}^n$; $n = H \times W \times C$ to be transmitted over k uses of a noisy channel, where k/n is the *bandwidth ratio*. An encoder $f_{\theta_i} : \mathcal{R}^n \to \mathcal{C}^{k_i}$ maps $\boldsymbol{x}$ into channel input symbols $\boldsymbol{z}_i \in \mathcal{C}_i^k$ in L blocks, where $\sum_{i=1}^{L} k_i = k$. These symbols are transmitted over a noisy channel, characterized by a random transformation $\eta : \mathcal{C}^{k_i} \to \mathcal{C}^{k_i}$, which may model physical impairments such as noise, fading or interference, resulting in the corrupted channel output $\hat{\boldsymbol{z}}_i = \eta(\boldsymbol{z}_i)$. We consider L distinct decoders, where the channel outputs for the first i blocks are decoded using $g_{\phi_i} : \mathcal{C}^{k_I} \to \mathcal{R}^n$ (where $I = \sum_{j=0}^{i} k_j$), creating reconstructions $\hat{\boldsymbol{x}}_i = g_{\phi_i}(\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_i) \in \mathcal{R}^n$, for $i \in 1, \ldots, L$.

The encoder and decoder(s) are modelled as fully convolutional networks, using generalized normalization transformations (GDN/IGDN) [22], followed by a parametric ReLU (PReLU) [23] activation function (or a sigmoid, in the last decoder block). The channel is incorporated into the model as a non-trainable layer. Fig. 2 presents the architecture and the hyperparameters used in the experiments. Before transmission, the latent vector $\boldsymbol{z}_i'$ generated at the encoder's last convolutional layer is normalized to enforce an average power constraint so that $\frac{1}{k_i}\mathbb{E}[\boldsymbol{z}_i^* \boldsymbol{z}_i] \leq P$, by setting $\boldsymbol{z}_i = \sqrt{k_i P}\frac{\boldsymbol{z}_i'}{\sqrt{\boldsymbol{z}_i'^* \boldsymbol{z}_i'}}$. The model can be optimized to minimize the average distortion between input $\boldsymbol{x}$ and its reconstructions $\hat{\boldsymbol{x}}_i$ at each layer i:

$$(\theta_i^*, \phi_i^*) = \arg\min_{\theta_i, \phi_i} \mathbb{E}_{p(\boldsymbol{x}, \hat{\boldsymbol{x}})}[d(\boldsymbol{x}, \hat{\boldsymbol{x}}_i)], \tag{1}$$

where $d(\boldsymbol{x}, \hat{\boldsymbol{x}}_i)$ is a specified distortion measure, usually the mean squared error (MSE), although other metrics are also considered. When $L > 1$, we have a multi-objective problem. However, we simplify it so that the optimization of multiple layers is done either jointly, by considering a weighted combination of losses, or greedily, by optimizing (θ_i, ϕ_i) successively. Please see [20], [21] for more details.

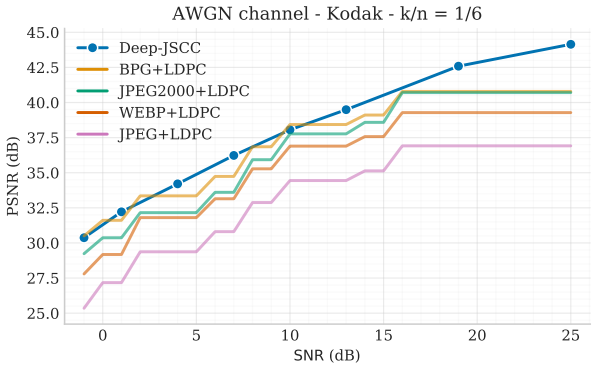


Fig. 3. Deep-JSCC performance compared to digital schemes.

III. Deep-JSCC

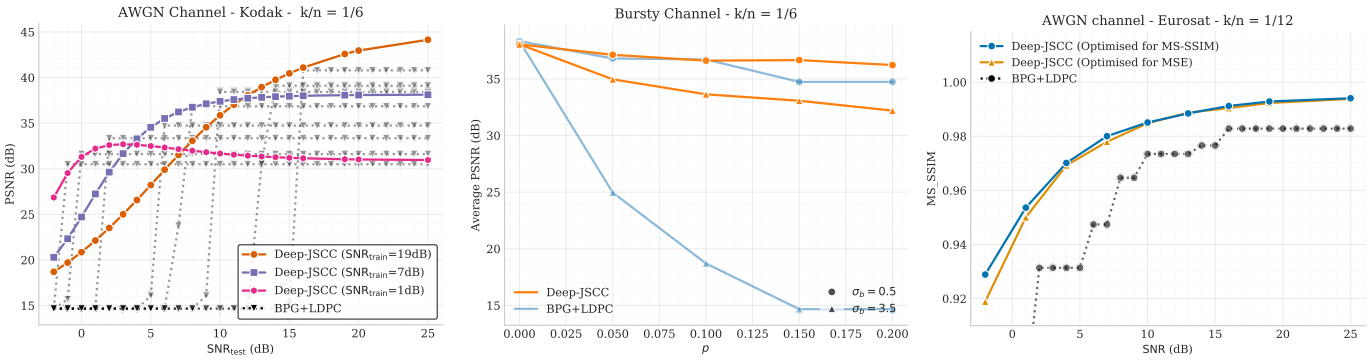
Our first set of results demonstrate the base case when, an image $\boldsymbol{x}$ is encoded by a single encoder and a single decoder, thus $L = 1$. We consider a complex AWGN channel with transfer function given by:

$$\eta_n(\boldsymbol{z}) = \boldsymbol{z} + \boldsymbol{n}, \tag{2}$$

where $\boldsymbol{n} \in \mathbb{C}^k$ is independent and identically distributed (i.i.d.) with $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma^2 I)$, where σ^2 is the average noise power. We measure the quality of the channel by the average signal-to-noise ratio (SNR) given by $\text{SNR} = 10\log_{10}\frac{1}{\sigma^2}(dB)$ when $P = 1$ and the systems' performance by the peak SNR (PSNR), given by $\text{PSNR} = 10\log_{10}\frac{255^2}{||\boldsymbol{x}-\hat{\boldsymbol{x}}_i||^2}(dB)$.

Fig. 3 compares deep-JSCC with other well established codecs (BPG, JPEG2000, WebP, JPEG) followed by LDPC channel coding (see [19], [24] for more information on the experimental setup, dataset and alternative schemes considered). We see that the performance of deep-JSCC is either above or comparable to the performance of the SoA schemes, for a wide range of channel SNRs.

These results are obtained by training a different encoder/decoder model for each SNR value evaluated in the case of deep-JSCC, and considering the best performance achieved by the separation-based scheme at each SNR. In



(a) (b) (c)

Fig. 4. (a) effects of graceful degradation for deep-JSCC compared to cliff effect in separation-based scheme; (b) performance of deep-JSCC on a bursty interference channel (c) performance of deep-JSCC trained with MS-SSIM as objective function.

Fig. 4a, we experiment training models at a specific channel SNR, but evaluating it on several SNR_{test} values, also for the separation-based schemes. It can be clearly seen that deep-JSCC presents *graceful degradation*, that is, the performance gradually decreases as channel deteriorates, while the digital scheme presents a *cliff-effect* when the quality of the channel goes below the capacity for which the code was designed, losing all transmission output. Thus, we can see that deep-JSCC not only produces high performing transmissions, but also *analog behavior*, being more robust to non-ergodic channels.

A. Versatility

A big advantage of deep-JSCC being data-driven is the possibility of training for different channel models, objective functions, or specific domains. Previous work [19] show deep-JSCC is able to learn how to operate on a Rayleigh fading channel, which models variations in channel quality over time, due to physical changes in the environment. Remarkably, the model could learn to operate in a fading channel without the need of channel estimation or feedback, which are both common practice in separation-based systems.

We can also consider a channel with 'bursty' noise, which can model the presence of a high variance noise with probability p in addition to the AWGN noise $\boldsymbol{n}$, modeling in practice, an occasional random interference from a nearby transmitter. Formally, this is a Bernoulli-Gaussian noise channel with transfer function:

$$\eta_w(\boldsymbol{z}) = \boldsymbol{z} + \boldsymbol{n} + B(k,p)\boldsymbol{w}, \tag{3}$$

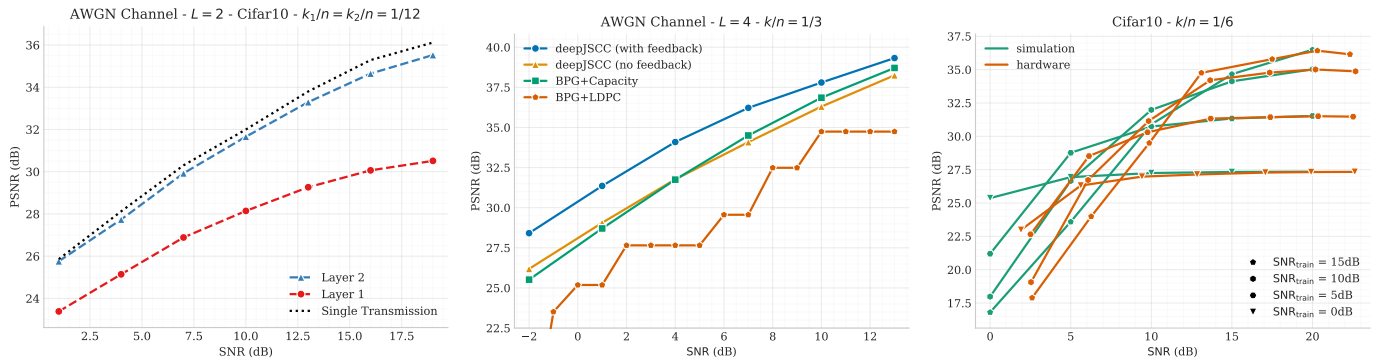
where $B(k,p)$ is the binomial distribution, and $\boldsymbol{w} \sim \mathcal{CN}(0, \sigma_b^2 I)$ the high variance noise component ($\sigma_b^2 >> 0$). Fig. 4b shows the effect of the probability p on the performance when the AWGN component's SNR is 10dB. We consider both a low-power ($\sigma_b = 0.5$) and a high-power ($\sigma_b = 3.5$) burst, and compare the performance with a digital scheme with BPG+LDPC. As expected, the performance degrades as p increases, but deep-JSCC is much more robust against the increasing power of the burst noise. A high-power burst degrades the performance of the digital scheme very quickly, even if the burst probability is very low, completely destroying the signal when $p > 0.15$. Deep-JSCC exhibits graceful degradation even in the presence of bursty noise, another important advantages in practical scenarios, particularly for communications over unlicensed bands, where occasional burst noise is common.

We also experimented training our model to a domain specific task, namely the transmission of satellite image data [25], a plausible application of our model. Here we use the distortion measure of multi-scale structural similarity (MS-SSIM) [26] – a widely accepted image quality measure that better represents human visual perception than pixel-wise differences. Our results, shown in Fig. 4c show that, when considering more specific domains, our model can better adapt to it, significantly increasing the performance gap between deep-JSCC and separation-based schemes.

B. Successive Refinement

Yet another advantage of deep-JSCC is the flexibility to adapt the transmission to different paths or stages. Consider a model with $L > 1$, in which a same image is transmitted progressively in blocks of size k_i, $i = 1, \dots L$ and $\sum_{i=1}^{L} k_i = k$. We aim to be able to reconstruct the complete image after each transmission, with increasing quality, thus performing *successive refinement* [27]–[29]. Progressive transmission can be applied to scenarios in which communication is either expensive or urgent. For example, in surveillance applications, it may be beneficial to quickly send a low-resolution image to detect a potential threat as soon as possible, while a higher resolution description can be later received for further evaluation or archival purposes. Or, in a multi-user communication setting, one could send different number of component for different users, depending on the available bandwidth.

We therefore expand our system, by creating L encoder and decoder pairs, each responsible for a partial transmission $\boldsymbol{z}_i$ and trained jointly (see [20] for implementation details and alternative architectures). Fig. 5a presents results for the case $L = 2$, for $k_1/n = k_2/n = 1/12$ and



(a) (b) (c)

Fig. 5. (a) Successive refinement with $L = 2$; (b) Layered transmission with channel output feedback, for $L = 4$; (c) Comparison between simulated and hardware performance.

shows the performance of each layer for different channel SNRs, for the AWGN channel. Results show that the loss of dividing the transmission into multiple stages is not significant; when compared to a single transmission with $k/n = 1/6$ (dotted black curve in Fig. 5a), the model performs with approximately the same quality for most channel conditions. Moreover, we observe that every layer of the layered transmission scheme preserves all features of the single transmission, such as graceful degradation and adaptability to different channel models.

C. Channel Output Feedback

Another interesting direction to be explored by deep-JSCC is the use of channel output feedback, when it is available. Suppose that alongside the *forward* communication channel considered so far, there is also a *feedback* channel, able to send back to the transmitter an estimation of the channel output $\tilde{\boldsymbol{z}}_{\boldsymbol{i}}$ after its realization. In a multi-layered transmission, this information can be used to inform subsequent layers and enhance the reconstruction at the receiver. Thus, a transmission of a source $\boldsymbol{x}$ is done sequentially in L steps, in which each step i a channel input $\boldsymbol{z}_i$ is generated from input x and feedback $\tilde{\boldsymbol{z}}_{\boldsymbol{i-1}}$ (for $i > 1$), transmitted and decoded to generate successively refined representations $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$ (see [21] for specific architecture and implementation details). There has also been recent advances in the use of channel output feedback to improve the performance of channel coding [30]; however, the design is for a specific blocklength and code rate, whereas the proposed deep-JSCC scheme can transmit large content, such as images.

Fig. 5b shows the results for a scenario considering noiseless feedback (i.e. $\tilde{\boldsymbol{z}}_{\boldsymbol{i}} = \hat{\boldsymbol{z}}_{\boldsymbol{i}}$) and three uses of the feedback channel ($L = 4$), for channel inputs with size $k_i/n = 1/12$, $i = 1, \ldots, 4$. We see that by exploiting the feedback information, deep-JSCC can further increase its performance, establishing its superiority to other schemes. Note that we compare deep-JSCC with feedback with a theoretical capacity achieving channel code, and can still outperform the separation-based scheme.

This architecture enables other communication strategies, such as variable length coding, in which a minimum number of layers $\boldsymbol{z}_{\boldsymbol{i}}$ are transmitted and the quality of the reconstruction is estimated through feedback, until a target quality is achieved and the further transmission is interrupted. This scheme can provide gains of over 50% in bandwidth, when compared to separation-based approaches [21]. Further experiments also demonstrate that our model successfully operates under noisy feedback channels, and even present graceful degradation when the feedback channel changes between training and evaluation.

D. Hardware Implementation

Finally, to validate the real world performance of the proposed architecture, we implemented our basic deep-JSCC on software defined radio platform. We used models trained on the AWGN model, with different SNRs. Results can be seen in Fig. 5c and show that the simulated performance closely matches the hardware performance, especially in higher SNRs.

We also analyzed the execution time of our model. We observed that the average encoding and decoding time per image with deep-JSCC is 6.40ms on GPU, or 15.4ms on CPU, while a scheme with JPEG2000+LDPC and BPG+LDPC takes on average 4.53 and 69.9ms respectively. This shows that, although our model can be further optimized for speed, it already presents competitive times, given its outstanding performance.

IV. Conclusion

This paper reviewed and explored different features of a DL-based architecture for JSCC of images over wireless channels, the deep-JSCC. We have shown that our architecture is extremely versatile to channel models, objective functions and even transmission configurations, being able to perform multi-layered transmission and exploit channel feedback. When compared to traditional digital schemes of transmission, deep-JSCC has shown outstanding performance in different metrics and scenarios, therefore presenting itself as a viable and superior alternative, particularly for low-latency and low-power applications.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[2] ——, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, vol. 1, Berkeley, CA, 1961, pp. 611–644.

[3] D. Gündüz, E. Erkip, A. Goldsmith, and H. V. Poor, "Source and channel coding for correlated sources over multiuser channels," *IEEE Trans. on Information Theory*, vol. 55, no. 9, pp. 3927–3944, Sep. 2009.

[4] S. Vembu, S. Verdu, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 44–54, Jan 1995.

[5] D. Gunduz and E. Erkip, "Joint source-channel codes for MIMO block-fading channels," *IEEE Trans. on Information Theory*, vol. 54, no. 1, pp. 116–134, Jan 2008.

[6] Y. Bengio, "Learning deep architectures for AI," *Found. and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, Jan. 2009.

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

[8] D. Gunduz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar, "Machine learning in the air," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2184–2199, Oct 2019.

[9] T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *Proc. of IEEE Int. Symp. on Signal Processing and Information Technology (ISSPIT)*, Dec. 2016, pp. 223–228.

[10] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec 2017.

[11] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Be'ery, "Deep learning methods for improved decoding of linear codes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 119–131, Feb 2018.

[12] H. Kim, Y. Jiang, R. Rana, S. Kannan, S. Oh, and P. Viswanath, "Communication algorithms via deep learning," in *Proc. of Int. Conf. on Learning Representations (ICLR)*, 2018.

[13] N. Samuel, T. Diskin, and A. Wiesel, "Deep mimo detection," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2017, pp. 1–5.

[14] M. B. Mashhadi, Q. Yang, and D. Gunduz, "Cnn-based analog csi feedback in fdd mimo-ofdm systems," 2019.

[15] A. Felix, S. Cammerer, S. Dorner, J. Hoydis, and S. ten Brink, "OFDM autoencoder for end-to-end learning of communications systems," in *Proc. IEEE Int. Workshop Signal Proc. Adv. Wireless Commun. (SPAWC)*, Jun. 2018.

[16] A. Caciularu and D. Burshtein, "Blind channel equalization using variational autoencoders," in *Proc. IEEE Int. Conf. on Comms. Workshops, Kansas City, MO*, May 2018, pp. 1–6.

[17] Y. Jiang, H. Kim, H. Asnani, S. Kannan, S. Oh, and P. Viswanath, "Turbo autoencoder: Deep learning based channel codes for point-to-point communication channels," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 2754–2764.

[18] 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*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Curran Associates, Inc., 2018, pp. 10 771–10 780.

[19] E. Bourtsoulatze, D. Burth Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

[20] D. B. Kurka and D. Gündüz, "Successive refinement of images with deep joint source-channel coding," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2019, pp. 1–5.

[21] D. B. Kurka and D. Gündüz, "Deepjscc-f: Deep joint-source channel coding of images with feedback," 2019.

[22] J. Ballé, V. Laparra, and E. P. Simoncelli, "Density modeling of images using a generalized normalization transformation," *arXiv preprint arXiv:1511.06281*, 2015.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[24] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 4774–4778.

[25] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019.

[26] 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. Ieee, 2003, pp. 1398–1402.

[27] Y. Steinberg and N. Merhav, "On hierarchical joint source-channel coding," in *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, Jun. 2004, pp. 365–365.

[28] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Transactions on Information Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.

[29] K. R. Sloan and S. L. Tanimoto, "Progressive refinement of raster images," *IEEE Transactions on Computers*, vol. 28, no. 11, pp. 871–874, 1979.

[30] H. Kim, Y. Jiang, S. Kannan, S. Oh, and P. Viswanath, "Deepcode: Feedback codes via deep learning," in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Curran Associates, Inc., 2018, pp. 9436–9446.