Realism and Fidelity: Two Sides of a Coin in Deep Joint Source-Channel Coding

Haotian Wu, Weichen Wang, Di You✉, Pier Luigi Dragotti, Deniz Gündüz
Department of Electrical and Electronic Engineering
Imperial College London, London SW7 2AZ, U.K.
{haotian.wu17, ww18, dy22, d.gunduz}@imperial.ac.uk
https://eedavidwu.github.io/W2-DeepJSCC/

Abstract

Deep joint source-channel coding (DeepJSCC) offers a promising approach to improving transmission efficiency by jointly leveraging source semantics and channel conditions. While prior work has focused on fidelity under varying channel conditions, recent diffusion-based approaches improve perceptual quality at the cost of high complexity and limited adaptability. In this work, we reveal that fidelity and perceptual realism can be unified in an adaptive DeepJSCC scheme through SNR-aware optimization, eliminating the need for separate models. Specifically, we propose W^2-DeepJSCC, a unified, channel-adaptive framework that dynamically balances fidelity and perceptual realism based on channel conditions. It introduces two key innovations: a saliency-guided perception–fidelity adapter (SG-PFA) and wavelet Wasserstein distortion (WA-WD). SG-PFA enables a single model to adapt across varying channel conditions, preserving semantic realism under poor channel conditions while enhancing fidelity under good ones. WA-WD, inspired by foveal and peripheral vision, provides fine-grained control in the wavelet domain. As a plug-and-play module, W^2-DeepJSCC integrates seamlessly with existing DeepJSCC architectures. Experiments show that W^2-DeepJSCC significantly outperforms baselines in perceptual metrics while maintaining strong fidelity at high SNRs. Prototype verification further highlights its advantages, demonstrating that the proposed method delivers competitive fidelity and perception with low complexity, making it a promising alternative for future deployments. Additionally, a user study further confirms that WA-WD aligns more closely with human perception than existing metrics.

1 Introduction

Advances in wireless communication and machine learning have driven emerging applications such as streaming and edge intelligence, requiring efficient, low-latency transmission on resource-limited devices. While Shannon's separation theorem [25] proves the optimality of independent source and channel coding under ideal conditions, suboptimality of separation is well known under practical constraints such as finite block lengths and dynamic channels. This motivates joint source-channel coding (JSCC) as a promising alternative [12], though classical JSCC design remains challenging. Recently, deep neural networks have been leveraged to learn a direct mapping from source data to channel symbols, known as DeepJSCC [7, 43, 35]. DeepJSCC has demonstrated versatility across diverse data types and channel models [38, 30, 39, 26], positioning it as a key foundation for next-generation semantic communication systems [11].

Correspondence to: Di You <dy22@ic.ac.uk>

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI and ML for Next-Generation Wireless Communications and Networking (AI4NextG).

Figure 1: Visual comparison between our W^2-DeepJSCC, ADJSCC, and the BPG-Capacity scheme when SNR=0dB and R=1/24. W^2-DeepJSCC achieves best perceptual realism, closely matching the original image, while others lose significant texture detail. Many more results, including a prototype validation, across SNRs and datasets are available in Appendix and the supplementary materials.

However, most existing DeepJSCC schemes are optimized for pixel-wise fidelity metrics, such as peak signal-to-noise ratio (PSNR), which often fail to capture true perceptual quality, especially under challenging channel scenarios where the textures and structures of the source degrade. In such cases, these metrics can even lose relevance, making perceptual quality increasingly critical. Yet, evaluating perceptual quality remains an open challenge, as widely studied in image compression [20, 3] and restoration [19, 48]. Perceptual quality, also referred to as 'realism' in the literature [13], refers to the reconstructions that comes from the same distribution as natural images. In general, there is a more general trade-off between the rate, distortion, and perception [6], which highlights that optimizing the rate-distortion trade-off alone does not guarantee the preservation of natural image characteristics.

This has motivated perception-enhanced DeepJSCC designs, with recent work leveraging generative models to better model natural image distributions. Early methods adopted generative-adversarial networks (GANs) [20] for perceptual enhancement [32, 10, 45], while more recent approaches utilize diffusion models (DMs) [16] to significantly boosting realism [49, 34, 45, 46, 47, 40, 22, 33, 8]. However, these methods are computationally expensive, typically tailored to low-SNR scenarios, and often require separate models for fidelity and perception objectives, posing storage and scalability challenges on edge devices. Meanwhile, diffusion-based DeepJSCC is also prone to hallucinations, even with semantic guidance [49], and is difficult to integrate with existing model-driven frameworks. These challenges highlight the need for a unified, low-complexity DeepJSCC approach that adapts to channel conditions and user priorities of fidelity-perception while maintaining compatibility.

To address these challenges, we propose W^2-DeepJSCC, a novel unified and channel-adaptive DeepJSCC framework. Two key challenges that we strive to address are: (1) how to explicitly control the perception–fidelity trade-off across channel conditions? and (2) how to design a diagnostic or human-aligned metric for adaptive DeepJSCC? At its core, W^2-DeepJSCC employs a saliency-guided perception–fidelity adapter (SG-PFA), which uses wavelet Wasserstein distortion (WA-WD) metric to dynamically shift optimization between perceptual quality at low SNR and high-fidelity reconstruction at high SNR. This plug-and-play module eliminates the need for multiple task-specific models and integrates seamlessly into existing DeepJSCC architectures, enabling an all-in-one model that delivers strong perceptual quality at significantly lower complexity than generative approaches.

Our main contributions are: (1) We propose W^2-DeepJSCC, an all-in-one DeepJSCC framework that unifies fidelity and perceptual realism via SNR-aware optimization. It introduces a saliency-guided perception–fidelity adapter based on wavelet Wasserstein distortion, enabling dynamic control across SNRs. As a plug-and-play module, it integrates with most existing DeepJSCC architectures, achieving

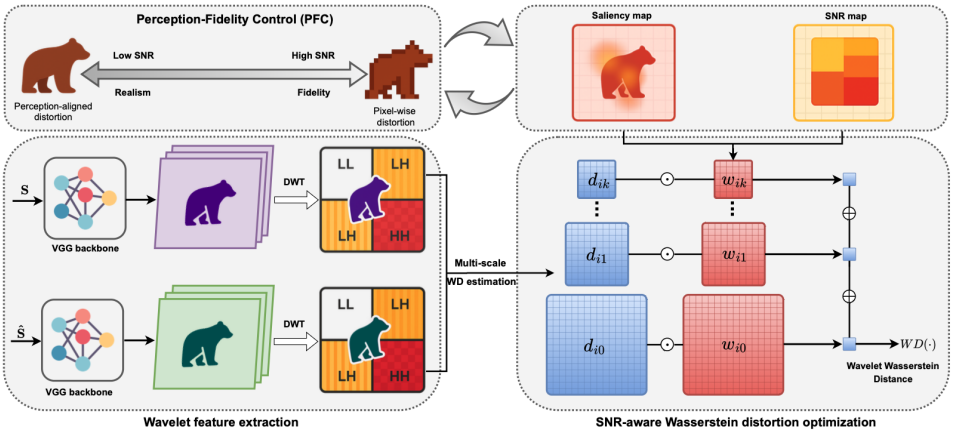

Figure 2: Illustration of the proposed W^2-DeepJSCC scheme, highlighting its key feature of perception–fidelity control. W^2-DeepJSCC enables fine-grained modulation of perceptual and fidelity objectives in the wavelet domain across varying channel conditions, guided by both saliency and SNR maps.

human-preferred perceptual quality at low SNR and high-fidelity at high SNR. (2) Numerical experiments demonstrate that W^2-DeepJSCC achieves strong perception–fidelity trade-offs across all SNRs, significantly outperforming baselines in perception metrics, while maintaining reasonable fidelity at high SNRs. These results align with the R-D-P theory. (3) Notably, the proposed wavelet Wasserstein distortion serves as an independent perceptual metric and diagnostic indicator for DeepJSCC, capturing both fidelity and realism under varying channels. User studies show it is more aligned with human rating results.

2 Related work

2.1 Deep joint source-channel coding

The first DeepJSCC scheme for image transmission was proposed in [7], outperforming standard separation-based approaches. It was later extended to accommodate various channel conditions[38, 42], as well as a range of communication scenarios, including orthogonal frequency division multiplexing (OFDM) [38, 44], multiple-input multiple-output (MIMO) [36], relay [5], and feedback channels [39]. Recently, diffusion models have gained increasing attention for enhancing perceptual quality in DeepJSCC. Early efforts explored hybrid designs, such as conditioning DMs on low-resolution images [21] or refining DeepJSCC outputs through post-processing [47]. More recent approaches incorporate semantic guidance [45, 49], channel denoising models [40], posterior sampling [22, 41, 33], and invertible networks [9, 8]. These works reflect a growing emphasis on perceptual quality in transmission, even at the cost of significantly increased computation.

2.2 Wasserstein Distortion

Wasserstein distance [31], grounded in optimal transport theory, has proven effective for guiding perceptual optimization [27, 14]. Its practical use began with image retrieval in [24], and later evolved through models such as Wasserstein GANs [1] and Wasserstein Autoencoders [29], which leveraged it to enhance perceptual quality and balance reconstruction fidelity. More recently, inspired by the structure of foveal and peripheral vision, Wasserstein distortion (WD) [23] was introduced as a unified metric that controls the fidelity–perception trade-off via spatial pooling. Building on this, [3] showed that optimizing lightweight codecs with WD can achieve perceptual quality comparable to generative models, but with significantly lower decoding complexity. These findings highlight the potential to unify perception and fidelity within a single model, achieving human-preferred visual quality with reduced storage and computational complexity.

3 Methodology

3.1 W^2-DeepJSCC architecture

The architecture of W^2-DeepJSCC (see Fig. 9 for an illustration) includes four residual and four dual-attention blocks in a comb structure, enabling feature modulation at multiple scales based on channel conditions and saliency maps. Each residual module consists of 2D convolution/deconvolution, GDN [2], and PReLU layers. The dual-attention block [38] modulates the generated features using two components: a channel-attention (CA) block, which adapts to channel conditions, and a spatial-attention (SA) block, which jointly considers SNR and saliency-based spatial importance. In the following section, we introduce the proposed Wavelet Wasserstein Distortion, which serves as the training objective for W^2-DeepJSCC under the SG-PFA mechanism.

3.2 Wavelet Wasserstein distortion.

Wavelet Wasserstein Distortion (WA-WD) compares features using a spatially varying σ-map in the wavelet domain. Images and reconstructions are passed through a VGG backbone and Haar wavelet transform to produce multiscale features $\boldsymbol{f_i^{gt}}$ and $\hat{\boldsymbol{f}}_{\boldsymbol{i}}^{re}$, where the superscript with color denote the ground truth and reconstructions, respectively. For each band, WA-WD computes the local 2-Wasserstein distance over patches centered at (x, y) with kernel size $\sigma(x, y)$, modulated by saliency and channel conditions with SG-PFA (which will be detailed later in Section 3.3). To reduce WD computational cost, we follow [3] to discretize σ into powers of two and estimate the WD. For each feature $\boldsymbol{f_i}$, the reference WD, $d_{i\alpha}$, is computed between reconstruction and ground truth via local means $\mu_{i\alpha}$ and standard deviations $\nu_{i\alpha}$ (element-wise) over each pooling scale α, as follows:

$$d_{i\alpha} = \sqrt{\left(\boldsymbol{\mu}_{\boldsymbol{i\alpha}}^{\mathbf{re}} - \boldsymbol{\mu}_{\boldsymbol{i\alpha}}^{\mathbf{gt}}\right)^2 + \left(\boldsymbol{\nu}_{\boldsymbol{i\alpha}}^{\mathbf{re}} - \boldsymbol{\nu}_{\boldsymbol{i\alpha}}^{\mathbf{gt}}\right)^2}. \tag{1}$$

The spatial σ-map is downsampled into $\sigma_{i\alpha}$ to match α for a linearly interpolated weight computation: $w_{i\alpha} = \max\left(0,\, 1 - |\log_2 \sigma_{i\alpha} - \alpha|\right)$. The final WA-WD is estimated via aggregation across different feature scales and bands [23, 3] as:

$$\text{WA-WD} = \sum_{b\in\{\text{LL,LH,HL,HH}\}} \sum_{i,j} \left(w_{ij}^b \odot d_{ij}^b\right), \tag{2}$$

where $b \in \{LL, LH, HL, HH\}$ denotes the wavelet sub-band (Please refer to Fig. 2 and Appendix C.2 for details). With loss in Eqn. 2 and SG-PFA guidance, W^2-DeepJSCC enables fine-grained wavelet-domain control of perceptual and fidelity objectives under varying channels.

3.3 Saliency-Guided Perception-Fidelity Adapter

Inspired by [23], we propose a saliency-guided perception–fidelity adapter for DeepJSCC. SG-PFA dynamically adjusts the wavelet-WD based on channel conditions: prioritizing perceptual quality at low SNRs while preserving fine details at high SNRs for fidelity. As illustrated in Fig. 3, this trade-off is controlled by assigning different σ maps to the wavelet-WD loss based on varying SNRs and the corresponding saliency maps.

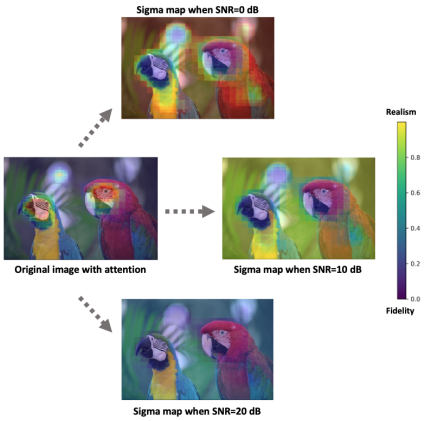

Figure 3: Illustration of SG-PFA.

Specifically, we obtain saliency predictions from EML-Net [17], with scores $s \in [0, 1]$. These scores are then converted into a spatial likelihood p using: $p = p_{\min} + (1 - p_{\min}) \cdot \frac{s}{\bar{s}}$, where $p_{\min} = 0.5$ serves as a lower bound, and $\bar{s}$ denotes the spatial mean of the saliency scores. This formulation ensures p is always positive and spatially averaged to one. Finally, the saliency-derived spatial likelihhod p is mapped to the sigma field $\sigma(x, y)$:

$$\sigma(x, y) = \kappa(\mu) \cdot \frac{p_{\min}}{p(x, y)}, \tag{3}$$

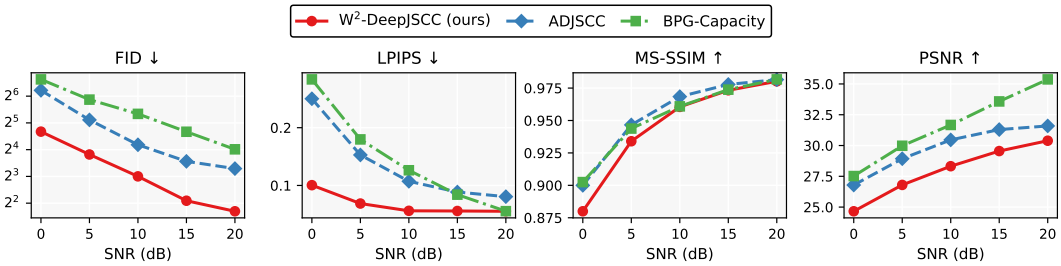

Figure 4: Performance of W^2-DeepJSCC across various SNRs on the Kodak dataset (R = 1/24). Arrows in the titles indicate whether lower (↓) or higher (↑) values are preferred.

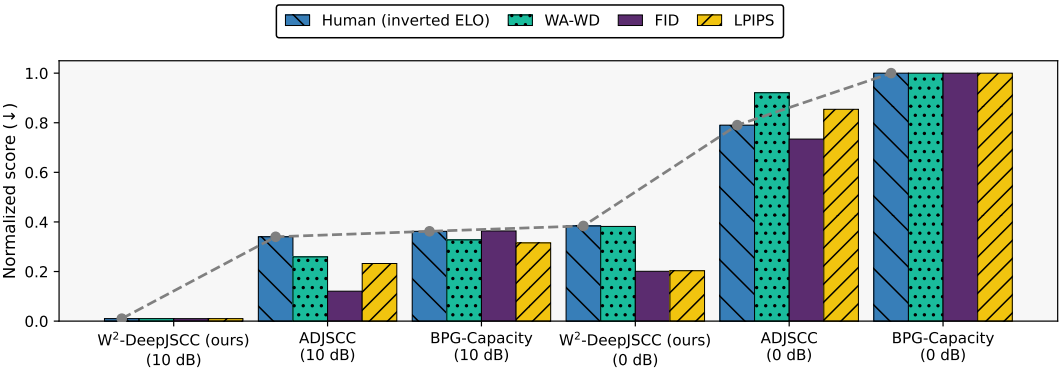

Figure 5: Normalized perceptual scores across SNRs from the user study. Human scores are inverted so that lower is better across all metrics. W^2-DeepJSCC significantly outperforms all baselines in perceptual quality. The proposed WA-WD metric shows strong alignment with human ratings, suggesting its effectiveness as a perceptual indicator across various SNRs.

where μ denotes the channel SNR, and $\kappa(\mu)$ is an SNR-adaptive scaling function defined as:

$$\kappa(\mu) = \sigma_{\max} - \frac{(\mu - \mathrm{SNR}_{\min})}{(\mathrm{SNR}_{\max} - \mathrm{SNR}_{\min})} \cdot (\sigma_{\max} - \sigma_{\min}). \tag{4}$$

Here, $\mathrm{SNR}_{\min}$ and $\mathrm{SNR}_{\max}$ define the SNR range, while $\sigma_{\min}$ and $\sigma_{\max}$ specify the bounds for sigma field. These parameters should be selected based on image resolution and perception requirements.

4 Experimental results

We evaluate W^2-DeepJSCC on the Kodak [18], comparing it with its counter-pair ADJSCC [42, 38] and a traditional separation-based scheme (BPG-Capacity) [37][2]. To quantify the results, we evaluate them using various metrics. For perceptual evaluation, we use FID [15], learned perceptual image patch similarity (LPIPS)[50], multi-scale structural similarity (MS-SSIM), and a human-rated ELO score [28]. For fidelity, we report PSNR. Additional experiments and visualizations are provided in the Appendix and supplementary materials.

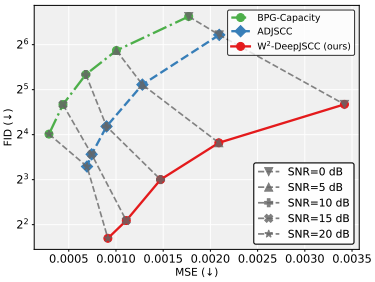

Figure 6: P-D trade-off.

General performance. As shown in Fig. 4, W^2-DeepJSCC dominates all other baselines in terms of all perceptual quality indices across all SNRs, despite lower PSNR and MS-SSIM at low SNRs. Notably, it achieves drastically improved FID and LPIPS results as SNR decreases, remarkably outperforming other methods at 10 dB, even with its results at 5 dB. As SNR increases,

[2]Diffusion- and ViT-based baselines are omitted due to their high complexity; similar gains are expected, as our unified framework can be integrated into them directly.

(a) Ground truth (b) BPG-Capacity (c) ADJSCC (d) W^2-DeepJSCC

Figure 7: Visual comparisons at SNR = 0 dB and R =1/24, where our method demonstrates clear advantages across the entire image. Additional examples are available in Appendix D.

the PSNR gap narrows, showing that W^2-DeepJSCC shifts towards improving the fidelity. Fig. 6 shows the detailed perception-distortion trade-off analysis, where we can observe that its pixel-wise fidelity improves steeply and catches up with the baselines, while others saturate. FID remains strong but improves more slowly. In contrast, ADJSCC quickly converges in fidelity, but perceptual gains saturate more slowly.

User study. Since metrics alone cannot fully capture perceptual quality, we conducted a user study using the CLIC rating protocol [3, 20] (details in Appendix B). As shown in Fig. 5, wavelet Wasserstein distortion (WA-WD) aligns better with human ratings than other metrics. Notably, WD indicator predicts correctly that ADJSCC and BPG are perceived similarly around 10 dB, consistent with user preferences, whereas FID and LPIPS often fail or even contradict subjective judgments.

Visualizations. As shown in Fig. 7, W^2-DeepJSCC shows clear perceptual advantages over all baselines, preserving finer texture details (e.g., brick, grass, hair). Interestingly, minor differences in repetitive non-salient regions can be observed, with minimal impact on overall visual quality. More examples can be seen in Appendix D.

Prototype verification. While we already show significant improvement by W^2-DeepJSCC on several metrics, to further validate our method in reality, we also conduct a prototype verification, where more pronounced advantages can be observed. More implementation details are provided in our Appendix E.

5 Conclusion

This paper proposed W^2-DeepJSCC, a unified, channel-adaptive DeepJSCC framework that explicitly balances fidelity and perceptual realism through a saliency-guided adapter based on wavelet

Wasserstein distortion. By leveraging SNR-aware optimization, our method shifts between preserving perceptual semantics at low SNRs and restoring fine-grained details at high SNRs, all within a single, low-complexity model. The proposed wavelet Wasserstein distortion serves both as a tunable optimization objective and a perceptually aligned indicator of adaptation and perception. Extensive experiments demonstrate that W^2-DeepJSCC significantly outperforms baseline methods in human preference and perceptual metrics. Our findings highlight the importance of unified realism–fidelity control in future semantic communication systems, and we hope this work marks a step toward perceptually aware, adaptive, and efficient end-to-end communication design.

Limitations and future work. While this paper introduces a novel wavelet-based Wasserstein distortion metric to unify realism and fidelity in channel-adaptive DeepJSCC, the overall design remains somewhat ad hoc and could greatly benefit from future advances in perceptual quality assessment. One promising direction is to replace the current VGG-based backbone with more powerful feature extractors to further enhance performance. Another is to integrate the proposed distortion or adaptation scheme into the diffusion-based sampling process, which holds potential for achieving stronger perceptual quality under stringent bandwidth constraints.

6 Acknowledgments and Disclosure of Funding

We acknowledge funding from the UKRI for the projects AI-R (ERC Consolidator Grant, EP/X030806/1) and SNS JU project 6G-GOALS under the EU's Horizon program (Grant Agreement No. 101139232).

References

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[2] Johannes Ballé, Valero Laparra, and Eero P Simoncelli. Density modeling of images using a generalized normalization transformation. *[Online]. Available: https://arxiv.org/abs/1511.06281*, 2015.

[3] Jona Ballé, Luca Versari, Emilien Dupont, Hyunjik Kim, and Matthias Bauer. Good, cheap, and fast: Overfitted image compression with wasserstein distortion. *arXiv preprint arXiv:2412.00505*, 2024.

[4] Fabrice Bellard. Bpg image format. *URL https://bellard. org/bpg*, 1(2):1, 2015.

[5] Chenghong Bian, Yulin Shao, Haotian Wu, Emre Ozfatura, and Deniz Gündüz. Process-and-forward: Deep joint source-channel coding over cooperative relay networks. *IEEE Journal on Selected Areas in Communications*, 2025.

[6] Yochai Blau and Tomer Michaeli. Rethinking lossy compression: The rate-distortion-perception tradeoff. In *International Conference on Machine Learning*, pages 675–685. PMLR, 2019.

[7] Eirina Bourtsoulatze, David Burth Kurka, and Deniz Gündüz. Deep joint source-channel coding for wireless image transmission. *IEEE Transactions on Cognitive Communications and Networking*, 5(3):567–579, 2019.

[8] Jiakang Chen, Selim F. Yilmaz, Di You, Pier Luigi Dragotti, and Deniz Gündüz. Sing: Semantic image communications using null-space and inn-guided diffusion models. *arXiv:eess.IV:2503.12484*, 2025.

[9] Jiakang Chen, Di You, Deniz Gündüz, and Pier Luigi Dragotti. Commin: Semantic image communications as an inverse problem with inn-guided diffusion models. In *IEEE Int'l Conf. on Acous., Speech and Sig. Proc. (ICASSP)*, pages 6675–6679, Seoul, Korea, 2024.

[10] Ecenaz Erdemir, Tze-Yang Tung, Pier Luigi Dragotti, and Deniz Gündüz. Generative joint source-channel coding for semantic image transmission. *IEEE J. Sel. Areas Commun.*, 41(8):2645–2657, 2023.

[11] Deniz Gündüz, Zhijin Qin, Inaki Estella Aguerri, Harpreet S Dhillon, Zhaohui Yang, Aylin Yener, Kai Kit Wong, and Chan-Byoung Chae. Beyond transmitting bits: Context, semantics, and task-oriented communications. *IEEE Journal on Selected Areas in Communications*, 41(1):5–41, 2022.

[12] Deniz Gündüz, Michèle A. Wigger, Tze-Yang Tung, et al. Joint source–channel coding: Fundamentals and recent progress in practical designs. *Proceedings of the IEEE*, pages 1–32, 2024.

[13] Yassine Hamdi, Aaron B. Wagner, and Deniz Gündüz. Rate-distortion-perception trade-off with strong realism constraints: Role of side information and common randomness. *arXiv:cs.IT:2507.14825*, 2025.

[14] Eric Heitz, Kenneth Vanhoey, Thomas Chambon, and Laurent Belcour. A sliced wasserstein loss for neural texture synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9412–9420, 2021.

[15] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Proc. Adv. in Neural Inf. Proc. Sys. (NeurIPS)*, pages 6840–6851, 2020.

[17] Sen Jia and Neil DB Bruce. Eml-net: An expandable multi-layer network for saliency prediction. *Image and vision computing*, 95:103887, 2020.

[18] Kodak. Kodak dataset. 1991.

[19] Haichuan Ma, Dong Liu, and Feng Wu. Rectified wasserstein generative adversarial networks for perceptual image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3648–3663, 2022.

[20] Fabian Mentzer, George D Toderici, Michael Tschannen, and Eirikur Agustsson. High-fidelity generative image compression. *Advances in neural information processing systems*, 33:11913–11924, 2020.

[21] Xueyan Niu, Xu Wang, Deniz Gündüz, Bo Bai, Weichao Chen, and Guohua Zhou. A hybrid wireless image transmission scheme with diffusion. In *IEEE Int'l Wrks. on Sig. Proc. Adv. in Wireless Comms. (SPAWC)*, pages 86–90, 2023.

[22] Li Qiao, Mahdi Mashhadi, Zhen Gao, Chuan Heng Foh, Pei Xiao, and Mehdi Bennis. Latency-aware generative semantic communications with pre-trained diffusion models. *[Online]: https://arxiv.org/abs/2403.17256*, 2024.

[23] Yang Qiu, Aaron B Wagner, Johannes Ballé, and Lucas Theis. Wasserstein distortion: Unifying fidelity and realism. In *2024 58th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2024.

[24] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

[25] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[26] Yulin Shao, Chenghong Bian, Li Yang, Qianqian Yang, Zhaoyang Zhang, and Deniz Gunduz. Point cloud in the air. *IEEE Communications Magazine*, pages 1–7, 2025.

[27] Guillaume Tartavel, Gabriel Peyré, and Yann Gousseau. Wasserstein loss for image synthesis and restoration. *SIAM Journal on Imaging Sciences*, 9(4):1726–1755, 2016.

[28] George Toderici, Wenzhe Shi, Radu Timofte, Lucas Theis, Johannes Ballé, Eirikur Agustsson, Nick Johnston, and Fabian Mentzer. Workshop and challenge on learned image compression (clic2020). In *CVPR*, 2020.

[29] Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. Wasserstein auto-encoders. *arXiv preprint arXiv:1711.01558*, 2017.

[30] Tze-Yang Tung and Deniz Gündüz. Deepwive: Deep-learning-aided wireless video transmission. *IEEE Journal on Selected Areas in Communications*, 40(9):2570–2583, 2022.

[31] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2008.

[32] Jun Wang, Sixian Wang, Jincheng Dai, Zhongwei Si, Dekun Zhou, and Kai Niu. Perceptual learned source-channel coding for high-fidelity image semantic transmission. *[Online]. Available: https://arxiv.org/abs/2205.13120*, 2022.

[33] Sixian Wang, Jincheng Dai, Kailin Tan, Xiaoqi Qin, Kai Niu, and Ping Zhang. Diffcom: Channel received signal is a natural condition to guide diffusion posterior sampling. *[Online]. Available: https://arxiv.org/abs/2406.07390*, 2024.

[34] Xinfeng Wei, Haonan Tong, Nuocheng Yang, and Changchuan Yin. Language-oriented semantic communication for image transmission with fine-tuned diffusion model. In *2024 16th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1456–1461. IEEE, 2024.

[35] Haotian Wu et al. Deep Joint Source and Channel Coding. *Foundations of Semantic Communication Networks*, pages 61–110, 2025.

[36] Haotian Wu, Yulin Shao, Chenghong Bian, Krystian Mikolajczyk, and Deniz Gündüz. Deep joint source-channel coding for adaptive image transmission over MIMO channels. *IEEE Transactions on Wireless Communications*, 2024.

[37] Haotian Wu, Yulin Shao, Chenghong Bian, Krystian Mikolajczyk, and Deniz Gündüz. Vision transformer for adaptive image transmission over MIMO channels. In *ICC 2023 - IEEE International Conference on Communications*, pages 3702–3707, 2023.

[38] Haotian Wu, Yulin Shao, Krystian Mikolajczyk, and Deniz Gündüz. Channel-adaptive wireless image transmission with OFDM. *IEEE Wireless Communications Letters*, 11(11):2400–2404, 2022.

[39] Haotian Wu, Yulin Shao, Emre Ozfatura, Krystian Mikolajczyk, and Deniz Gündüz. Transformer-aided wireless image transmission with channel feedback. *IEEE Transactions on Wireless Communications*, pages 1–1, 2024.

[40] Tong Wu, Zhiyong Chen, Dazhi He, Liang Qian, Yin Xu, Meixia Tao, and Wenjun Zhang. CDDM: Channel denoising diffusion models for wireless semantic communications. *IEEE Trans. Wireless Commun.*, 2024.

[41] Bingxuan Xu, Shujun Han, Xiaodong Xu, Weizhi Li, Rui Meng, Chen Dong, and Ping Zhang. Semantic prior aided channel-adaptive equalizing and de-noising semantic communication system with latent diffusion model. *IEEE Transactions on Wireless Communications*, pages 1–1, 2025.

[42] Jialong Xu, Bo Ai, Wei Chen, Ang Yang, Peng Sun, and Miguel Rodrigues. Wireless image transmission using deep source channel coding with attention modules. *IEEE Trans. Circuits Syst. Video Technol.*, 32(4):2315–2328, 2021.

[43] Jialong Xu, Tze-Yang Tung, Bo Ai, Wei Chen, Yuxuan Sun, and Deniz Gündüz. Deep joint source-channel coding for semantic communications. *IEEE Communications Magazine*, 61(11):42–48, 2023.

[44] Mingyu Yang, Chenghong Bian, and Hun-Seok Kim. OFDM-guided deep joint source channel coding for wireless multipath fading channels. *IEEE Transactions on Cognitive Communications and Networking*, 8(2):584–599, 2022.

[45] Mingyu Yang, Bowen Liu, Boyang Wang, and Hun-Seok Kim. Diffusion-aided joint source channel coding for high realism wireless image transmission. *arXiv preprint arXiv:2404.17736*, 2024.

[46] Ruihan Yang and Stephan Mandt. Lossy image compression with conditional diffusion models. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, 2024.

[47] Selim F Yilmaz, Xueyan Niu, Bo Bai, Wei Han, Lei Deng, and Deniz Gündüz. High perceptual quality wireless image delivery with denoising diffusion models. In *IEEE INFOCOM 2024-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 1–5. IEEE, 2024.

[48] Di You and Pier Luigi Dragotti. Indigo+: A unified inn-guided probabilistic diffusion algorithm for blind and non-blind image restoration. *IEEE Journal of Selected Topics in Signal Processing*, 18(6):1108–1122, 2024.

[49] Maojun Zhang, Haotian Wu, Guangxu Zhu, Richeng Jin, Xiaoming Chen, and Deniz Gündüz. Semantics-guided diffusion for deep joint source-channel coding in wireless image transmission. *IEEE Transactions on Wireless Communications*, pages 1–1, 2025.

[50] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

Appendix Contents

A Broader impacts

This work presents a technical contribution in joint source and channel coding with no immediate societal impact. Nonetheless, potential downstream effects may arise depending on the applications:

A.1 Positive impacts

- Efficiency: Its high realism at low complexity supports energy-efficient, scalable deployment in bandwidth-constrained settings like IoT and mobile devices, significantly reducing the energy footprint of edge devices in the future.
- Semantic-aware content delivery: Saliency-based coding schemes can benefit adaptive streaming and semantic-level rendering with high realism in AR/VR.

A.2 Potential negative impacts

However, W^2-DeepJSCC may also be misused to selectively manipulate image content, or potentially leak information from training data. To mitigate such risks, we encourage future work to incorporate safeguards such as explainable DeepJSCC transmission mechanisms. All datasets used in our study are publicly available, and we release our models and code solely to support transparency and reproducibility.

B Baseline implementation

We train all DeepJSCC schemes on the ImageNet dataset using randomly cropped 256×256 patches. Channel SNRs are uniformly sampled from 0 dB to 20 dB, with a batch size of 16 and a learning rate of $1e^{-4}$ (following the original papers of baselines).

For baselines, W^2-DeepJSCC is compared against separation-based baselines, including BPG-Capacity [38], as well as its counterpart, ADJSCC [42]. We note that the BPG-Capacity scheme represents a loose upper bound for separation-based methods, as it assumes a capacity-achieving channel code, and BPG [4] itself is a competitive standard for image compression. Note that we do not include diffusion- or ViT-based methods as baselines, since they incur significantly higher complexity. Our approach is designed as a unified framework that can be integrated into such schemes to provide further enhancements.

For human rating, we set up a two-alternative forced-choice (2AFC) [3]user study following [3] on the full Kodak dataset [18]. As shown in Fig. 8, for each individual rating (among five), two images are available to the rater: On one side of the screen, a random 512 × 512 pixel crop of the original. On the other, the corresponding crop of two reconstructed images, between which the rater can flip by pressing a key. The rater is asked to select the reconstruction that looks more similar to the original.

C W^2-DeepJSCC implementation

C.1 Detailed W^2-DeepJSCC architecture

The detailed architecture of W^2-DeepJSCC is illustrated in Fig. 9. The architecture of W^2-DeepJSCC includes four residual and four dual-attention blocks in a comb structure, enabling feature modulation

[3]We compute Elo scores using the open source implementation of the CLIC rating model `https://github.com/google-research/googleresearch/tree/master/elo_rater_model`.

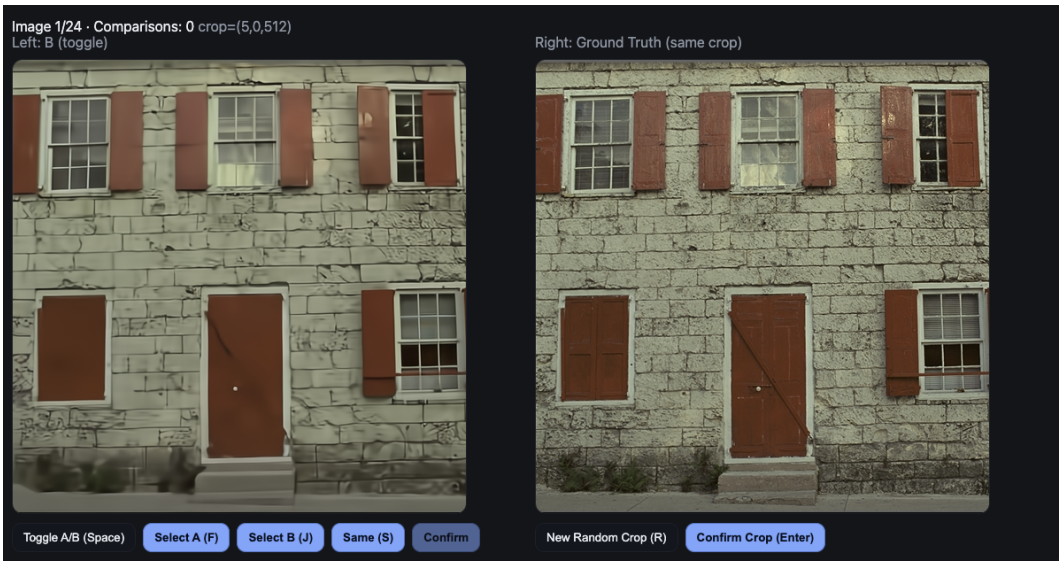

Figure 8: Example screenshots of the rating interface. The user is asked to select their preferred transmission result on the left, based on the comparison with the result on the right.

at multiple scales based on channel conditions and saliency maps. Each residual module consists of 2D convolution/deconvolution, GDN [2], and PReLU layers. The dual-attention block [38] modulates the generated features using two components: a channel-attention (CA) block, which adapts to channel conditions, and a spatial-attention (SA) block, which jointly considers SNR and saliency-based spatial importance. In the following section, we introduce the proposed Wavelet Wasserstein Distortion, which serves as the training objective for W^2-DeepJSCC under the SG-PFA mechanism.

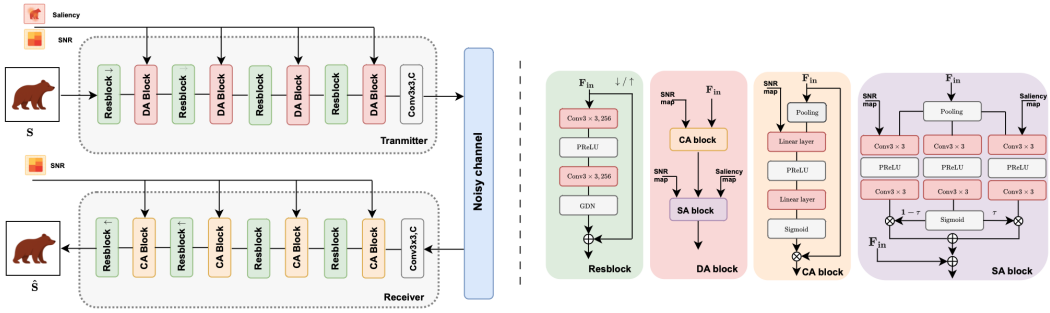

Figure 9: Illustration of the W^2-DeepJSCC architecture, where residual blocks and dual-attention modules are integrated. Each dual-attention block combines channel-aware and saliency-guided spatial modulation to adaptively modulate features based on both SNR and semantic importance.

C.2 Wavelet-WD computation

For each band, WA-WD computes a local 2-Wasserstein distance over patches centered at (x, y) with size $\sigma(x, y)$, derived from channel and saliency maps to adapt spatial sensitivity. As $\sigma \to 0$, it reduces to pointwise distance (foveal vision), while a larger σ captures peripheral, texture-level deviations. In WA-WD, a larger σ makes the metric more permissive to texture resampling—allowing replacement with statistically similar textures.

For DeepJSCC, we assign a smaller σ in salient regions likely to be directly observed, and a larger σ elsewhere. Under low SNR, the overall σ range is enlarged, guiding the codec to allocate more resources to preserve important salient content while still maintaining perceptual quality for others. For high SNR, the overall σ is reduced, guiding a more fidelity-preferred optimization. In the wavelet domain, such a design becomes more fine-grained by explicitly separating feature bands, allowing

more precise control over low- and high-frequency components under different channel conditions. To reduce the computational cost of spatially varying pooling, we follow [3] and discretize σ into powers of two for efficient approximation. Specifically, for each feature $\boldsymbol{f_i}$, the WD is computed as:

$$d_{i\alpha} = \sqrt{\left(\mu_{i\alpha}^{\mathrm{re}} - \mu_{i\alpha}^{\mathrm{gt}}\right)^2 + \left(\nu_{i\alpha}^{\mathrm{re}} - \nu_{i\alpha}^{\mathrm{gt}}\right)^2}, \tag{5}$$

where $\mu_{i\alpha}^{\mathrm{gt}}$ and $\nu_{i\alpha}^{\mathrm{re}}$ denote the first moment and the standard deviation (element-wise) of each feature map, for the i-th feature and equivalent pooling size α.

For robust performance, d is computed through k downsampling processes, yielding multi-resolution distances $d_{i,k}$. Meanwhile, σ is downsampled accordingly, giving $\sigma_{i,\alpha}$. The interpolation estimation weight is then defined as:

$$w_{i,\alpha} = \max\big(0,\, 1 - |\log_2 \sigma_{i\alpha} - \alpha|\big), \tag{6}$$

which equals 1 when σ matches the reference resolution, and linearly fades to 0 for nearby resolutions.

The final WA-WD is estimated via an aggregation across different features and scales ([23, 3], Fig. 2, and Appendix C.2 for more details) as:

$$\text{WA-WD} = \sum_{b\in\{\text{LL,LH,HL,HH}\}} \sum_{i,j} \big(w_{ij}^b \odot d_{ij}^b\big). \tag{7}$$

More details can be seen in [3, 23].

D Additional experiments and visualizations

This section provides more visulizations across Kodak and CLIC2020 dataset, as shown in Figs. 10, 11, and 12.

E Prototype verification

To further verify the effectiveness of our method, we conducted a prototype validation using a software-defined radio (SDR) setup. The transmitter executes the encoder on an NVIDIA Jetson Xavier NX, paired with a USRP-2922 for radio-frequency transmission. The receiver is a personal computer equipped with a LimeSDR Mini 2.0, running GNU Radio 3.10. All model inference is performed directly on the Jetson device. Experiments were conducted in an indoor corridor with a distance of 5 meters between transmitter and receiver, significantly longer than the $1\breve{}2$ meters typically used in prior DeepJSCC prototype experiments, introducing realistic multipath and shadowing effects. Both devices used omnidirectional antennas positioned at equal height. To adjust the signal-to-noise ratio (SNR), only the transmitter gain was varied (from 17 dB to 25 dB), while receiver-side gain remained constant. This approach ensures changes in SNR result solely from the transmitted signal strength. Visualized results are provided in Figs. 13, 14

Figure 10: Visual performance at SNR = 0dB for Kodak dataset. W^2DJSCC shows great performance on water texture and text preservation. Note that for reversed text the perceptual quality lowers compared to normal ones.

Figure 11: Visual performance (image horizontally cropped) at SNR = 0dB for CLIC 2020 dataset. The textures of walls and structures of buildings are largely preserved with W^2DJSCC. Details in the low contrast region are particularly perceptually visible.

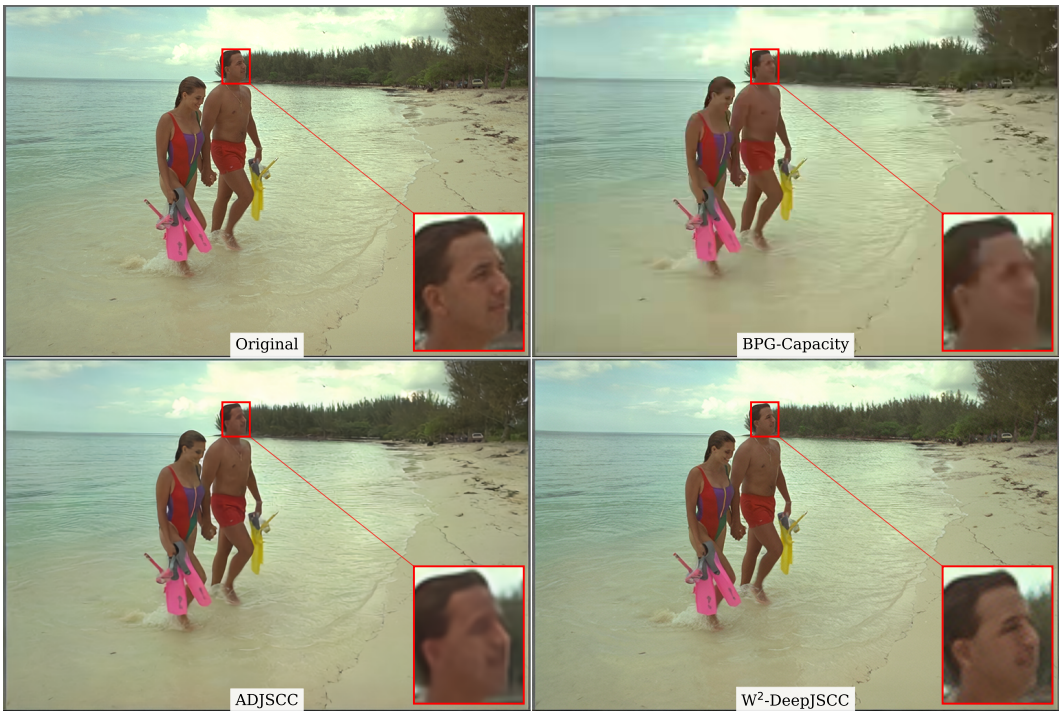

Figure 12: Visual performance at SNR = 10dB for Kodak dataset. We can see that for relatively high SNR there is still visible difference for perceptually sensitive objects such as human face.

(a) Original kodim01

(b) ADJSCC ($P_s = 17dB, \hat{\mu} = 3dB$)

(c) W^2-DeepJSCC ($P_s = 17dB, \hat{\mu} = 3dB$)

(d) ADJSCC ($P_s = 21dB, \hat{\mu} = 7dB$)

(e) W^2-DeepJSCC ($P_s = 21dB, \hat{\mu} = 7dB$)

(f) ADJSCC ($P_s = 25dB, \hat{\mu} = 11dB$)

(g) W^2-DeepJSCC ($P_s = 25dB, \hat{\mu} = 11dB$)

Figure 13: Prototype verification under different transmitter power for Kodak dataset, where the bandwidth ratio is $1/12$. W^2-DeepJSCC enables more fine-grained and robust transmission, preserving detailed textures such as curtains, wall patterns, and even grass on the ground.

(a) Original kodim01

(b) ADJSCC ($P_s = 17dB, \hat{\mu} = 3dB$)

(c) W^2-DeepJSCC ($P_s = 17dB, \hat{\mu} = 3dB$)

(d) ADJSCC ($P_s = 21dB, \hat{\mu} = 7dB$)

(e) W^2-DeepJSCC ($P_s = 21dB, \hat{\mu} = 7dB$)

(f) ADJSCC ($P_s = 25dB, \hat{\mu} = 11dB$)

(g) W^2-DeepJSCC ($P_s = 25dB, \hat{\mu} = 11dB$)

Figure 14: Prototype verification under different transmitter power for Kodak dataset, where the bandwidth ratio is $1/12$. W^2-DeepJSCC offers a clear advantage, such as textures on trees and even interior details beneath the river surface that are lost in baseline reconstructions.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims in the abstract and introduction are consistent with the core contributions and scope of our paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We provide a section for limitations and future work.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: We study a learning-based neural codec without relying on formal assumptions or theoretical proofs. The conclusions of the paper are based on extensive numerical experiments, which are consistently supported throughout the paper.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes] ,

Justification: We provide parameters and architecture necessary for this paper. A more complete version of the codebase, including detailed documentation and usage instructions, will be released publicly after the longer version of this paper is accessible.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We provide parameters and architecture necessary for this paper. A more complete version of the codebase, including detailed documentation and usage instructions, will be released publicly after the longer version of this paper is accessible.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide all details in the appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Not applicable to our setting, where a deterministic coding scheme is used.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix provides all the platform details.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper conform with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We provide this in our appendix.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [No]

    Justification: Our work does not involve models or datasets considered high risk for misuse. The models and datasets used are standard compression dataset and do not pose notable security, privacy, or misuse concerns.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All assets used in this work, including datasets, pretrained models, and code libraries, are properly cited in the paper. Their licenses and terms of use are respected.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: Not applicable to our work.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

    Justification: Our work includes a small-scale human preference study assessing the perceptual quality of compression and transmission results. All participants were adult volunteers from our research group, who provided informed consent and were not monetarily compensated. The full instructions and example screenshots of the rating interface are provided in the Appendix.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This work does not involve risk for participants, and therefore does not require IRB approval or equivalent ethical review.

    Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: No large language models (LLMs) are used as part of the core methodology or experimental design in this research. Any LLM usage, if any, was limited to minor writing support and did not influence the scientific content or originality of the work.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.