An SNR-Adaptive Deep Joint Source-Channel Coding Scheme for UAV Semantic Image Transmission

Quang Tuan Do, Tung Son Do, Thanh Phung Truong, Dongwook Won, and Sungrae Cho
School of Computer Science and Engineering
Chung-Ang University
Seoul, South Korea
Email: dqtuan@uclab.re.kr; tsdo@uclab.re.kr; tptruong@uclab.re.kr; dwwon@uclab.re.kr; srcho@cau.ac.kr

***Abstract*—Deep learning-based joint source-channel coding (Deep JSCC) has demonstrated substantial promise for robust image transmission in Unmanned Aerial Vehicle (UAV) systems. However, conventional Deep JSCC models operate exclusively at fixed Signal-to-Noise Ratios (SNRs) and exhibit severe performance degradation when channel conditions deviate from their design parameters. Furthermore, deploying multiple specialist models to accommodate varying channel qualities becomes impractical for resource-constrained UAVs due to storage and computational limitations. To address this challenge, we adopt the Attention DL-based JSCC (ADJSCC) mechanism, which conditions channel-wise soft attention on SNR to enable a single model to maintain high performance across diverse channel conditions, and tailor it for UAV semantic image transmission. In particular, we integrate ADJSCC's Attention Feature (AF) modules throughout both encoder and decoder to modulate features based on instantaneous SNR, and provide a UAV-focused training and evaluation protocol. Our comprehensive evaluation on the CIFAR-10 dataset demonstrates that this ADJSCC-based model achieves robust performance across all tested SNRs, consistently matching or surpassing specialist models within their optimal operating regions while significantly outperforming them under mismatched conditions, validating the practical viability of ADJSCC for reliable UAV image transmission in dynamic wireless environments.**

***Index Terms*—Semantic Communication, Deep Joint Source-Channel Coding, Unmanned Aerial Vehicle, Image Transmission, SNR-Adaptive, Attention Mechanism.**

I. Introduction

In mission-critical applications such as post-disaster surveillance, environmental monitoring, and precision agriculture, Unmanned Aerial Vehicles (UAVs) function as indispensable mobile sensing platforms that must transmit captured visual data to ground control stations for real-time analysis [1]. Related multi-UAV surveillance and cooperative control studies further underscore the need for reliable airborne sensing and communication [2], [3].

These operations occur under challenging and dynamically varying wireless conditions characterized by path loss variations, interference, and temporal channel fluctuations. Applications such as autonomous drone-delivery also require cooperative mobility under tight energy budgets, reinforcing the need for robust links and adaptive communication [3]. Deep learning-based Joint Source-Channel Coding (Deep JSCC), also referred to as semantic communication, has emerged as a revolutionary paradigm that addresses these challenges by fundamentally rethinking traditional communication system design.

Unlike conventional approaches that strictly adhere to Shannon's separation theorem [4], Deep JSCC employs a unified deep neural network (DNN) autoencoder architecture to learn direct mappings from source data to channel-robust symbol representations. This end-to-end optimization strategy has demonstrated exceptional performance characteristics, particularly in low Signal-to-Noise Ratio (SNR) regimes where traditional methods often fail [5]. Moreover, Deep JSCC systems exhibit graceful degradation properties and can prioritize semantically important information during transmission, making them particularly suitable for bandwidth-limited UAV applications [6]. Complementary quality-aware video delivery research likewise emphasizes end-to-end adaptation to network dynamics in mobile environments [7], [8].

Despite these advantages, a fundamental limitation constrains the practical deployment of existing Deep JSCC models: their inherently static nature. Contemporary approaches are typically trained and optimized for specific, predetermined channel SNR values [9]. When actual channel conditions deviate from these design points—a frequent occurrence for mobile UAVs operating in dynamic environments—the model's performance deteriorates dramatically. While deploying multiple specialist models, each optimized for different channel conditions, might theoretically address this limitation, such approaches become prohibitively expensive in terms of storage requirements, memory consumption, and computational overhead for resource-constrained UAV platforms [10]. Parallel advances in energy-efficient adaptive communication and learning further reinforce the imperative of lightweight, unified designs in resource-constrained settings [11], [12].

To overcome this adaptability limitation, Xu et al. introduced Attention DL-based JSCC (ADJSCC), which employs SNR-conditioned, channel-wise soft attention to dynamically recalibrate intermediate features and enable a single model to operate robustly over a wide SNR range; in this work, this attention mechanism is adopted and specialized for UAV semantic image transmission [13], [14]. This Attention Feature (AF) block accepts estimated channel SNR as a conditional input and generates channel-aware attention scores that dynamically recalibrate feature map activations. Consequently, the network learns to implement context-aware transmission strategies—for instance, emphasizing protection of essential features under adverse channel conditions while transmitting detailed information when channels permit higher fidelity reconstruction.

Our principal contributions are threefold:

1) We adopt the ADJSCC mechanism of Xu et al. and integrate its SNR-conditioned Attention Feature modules into a UAV-oriented pipeline, documenting design choices and implementation details for aerial image links [13].
2) We develop a comprehensive evaluation methodology that systematically compares the unified ADJSCC-based model against multiple specialist baselines trained for fixed SNR operating points, across extensive channel condition ranges.
3) We demonstrate through PSNR, SSIM, and LPIPS that the ADJSCC-based approach provides consistently robust image reconstruction, underscoring its practical utility for UAVs in fluctuating channel environments.

II. System Model and Problem Formulation

We consider a UAV-to-ground image transmission system wherein the channel SNR information is available at both transmitter and receiver to guide adaptive encoding and decoding processes, as assumed by ADJSCC's SNR-feedback setting [13]. This assumption reflects practical scenarios where channel state information can be estimated through pilot signals or feedback mechanisms [15]. Analogous adaptivity demands appear in mobility-centric streaming systems, motivating SNR-aware policies for stable perceptual quality under motion [16].

A. Transmitter Model and Channel Characterization

At the UAV transmitter, a source image $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$ and the instantaneous channel SNR γ are jointly processed by the semantic encoder $f_{enc}(\cdot, \gamma)$. The encoder, implemented as a deep neural network with learnable parameters θ_{enc}, maps the input to a compact latent representation:

$$\mathbf{z} = f_{enc}(\mathbf{x}, \gamma; \theta_{enc}) \tag{1}$$

Subsequently, the latent tensor $\mathbf{z}$ undergoes flattening and power normalization to satisfy average transmission power constraints [17]. The resulting normalized vector $\mathbf{s}$ is transmitted over the wireless channel. For this investigation, we model the communication link as an Additive White Gaussian Noise (AWGN) channel, representing scenarios with stable Line-of-Sight propagation paths. This choice enables direct analysis of the model's SNR adaptability without complications from fading effects. The received signal at the ground station is:

$$\mathbf{y} = \mathbf{s} + \mathbf{n} \tag{2}$$

where $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ represents additive Gaussian noise with variance $\sigma^2 = \frac{P_s}{10^{\gamma/10}}$, where P_s denotes the signal power.

B. Receiver Model and Optimization Objective

At the ground station receiver, the reconstruction objective is to recover the original image from the noisy received vector $\mathbf{y}$. Both the received signal and the same SNR value γ are input to the semantic decoder $g_{dec}(\cdot, \gamma)$:

$$\hat{\mathbf{x}} = g_{dec}(\mathbf{y}, \gamma; \theta_{dec}) \tag{3}$$

The fundamental optimization problem involves designing a single encoder-decoder pair with parameters $\theta = \{\theta_{enc}, \theta_{dec}\}$ that minimizes expected reconstruction error across distributions of both images and SNR values:

$$\min_{\theta} \mathbb{E}_{\mathbf{x},\gamma} \left[d(\mathbf{x}, g_{dec}(f_{enc}(\mathbf{x}, \gamma) + \mathbf{n}, \gamma)) \right] \tag{4}$$

where $d(\cdot, \cdot)$ represents a distortion measure such as Mean Squared Error (MSE). This formulation explicitly incorporates SNR variability into the learning objective, enabling the network to develop adaptive strategies across channel conditions, contrasting with traditional separate source-channel coding approaches [4].

III. SNR-Adaptive Framework

Our design follows the ADJSCC architecture by Xu et al., which augments a convolutional autoencoder backbone with AF modules to achieve SNR-conditioned feature recalibration across the encoder and decoder [13].

A. Backbone Architecture Design

The encoder architecture comprises a sequence of convolutional residual blocks that progressively downsample input images while extracting hierarchical feature representations [18]. Each residual block incorporates two convolutional layers with GDN activations, which have proven effective for decorrelating feature responses in image compression applications [19]. The symmetric decoder employs transposed convolutional residual blocks to upsample encoded features and reconstruct output images.

The encoder configuration consists of five residual blocks with the following channel progressions: $3 \to 128 \to 128 \to 128 \to 128 \to 48$, where the final output maintains spatial dimensions of 8×8 pixels. The decoder mirrors this architecture in reverse, ultimately producing RGB images with sigmoid activation to ensure pixel values within $[0, 1]$. This design choice leverages the proven effectiveness of residual connections in enabling training of deep networks while avoiding vanishing gradient problems [18].

B. Attention Feature (AF) Block for Dynamic SNR Adaptation

The AF block represents the core component enabling channel-aware adaptation and is strategically inserted after each residual block in both encoder and decoder paths. Drawing inspiration from successful attention mechanisms in computer vision and natural language processing [14], the AF block operation proceeds as follows:

1) **Feature Aggregation**: Given input feature map $\mathbf{X}_{in} \in \mathbb{R}^{B \times C \times H \times W}$ and scalar SNR value γ, the block computes channel-wise feature statistics:

$$\boldsymbol{\mu} = \frac{1}{H \cdot W} \sum_{h=1}^{H} \sum_{w=1}^{W} \mathbf{X}_{in}[:, :, h, w] \tag{5}$$

2) **Context Integration**: The mean vector $\boldsymbol{\mu} \in \mathbb{R}^{B \times C}$ is concatenated with the SNR value to form a joint feature-channel representation:

$$\mathbf{c} = [\boldsymbol{\mu}, \gamma] \in \mathbb{R}^{B \times (C+1)} \tag{6}$$

3) **Attention Score Generation**: A compact two-layer Multi-Layer Perceptron (MLP) with ReLU and sigmoid activations learns the mapping from joint features to attention scores:

$$\mathbf{a} = \text{Sigmoid}(\mathbf{W_2} \cdot \text{ReLU}(\mathbf{W_1} \cdot \mathbf{c} + \mathbf{b_1}) + \mathbf{b_2}) \tag{7}$$

where $\mathbf{W_1} \in \mathbb{R}^{C/2 \times (C+1)}$, $\mathbf{W_2} \in \mathbb{R}^{C \times C/2}$ are learnable weight matrices.

4) **Feature Modulation**: The attention scores are applied channel-wise to modulate the input features:

$$\mathbf{X}_{out} = \mathbf{X}_{in} \odot \text{reshape}(\mathbf{a}, [B, C, 1, 1]) \tag{8}$$

This design enables the network to learn content- and channel-aware transmission strategies within a unified model, dynamically emphasizing or suppressing feature channels based on both semantic content and channel quality. The lightweight nature of the AF blocks ensures minimal computational overhead while providing significant adaptability benefits. This approach also aligns with quality-aware delivery principles that balance fidelity and robustness under varying network conditions [8].

IV. Experimental Evaluation

A. Experimental Setup and Configuration

We conduct comprehensive evaluations using the CIFAR-10 dataset [20], which provides standardized benchmarks for image transmission tasks with 60,000 32×32 color images across 10 classes. The experimental configuration encompasses the following components:

- **Proposed ADJSCC Model**: Our SNR-adaptive architecture employs $N_C = 128$ channels throughout the network with convolutional kernels of size 5×5. The model undergoes end-to-end training on the CIFAR-10 training set using a batch size of 64 for 50 epochs with Adam optimizer and learning rate of 10^{-4}. During training, SNR values are uniformly sampled from the continuous range $[0, 20]$ dB for each batch to encourage adaptation across diverse channel conditions.
- **Baseline Specialist Models**: We implement three baseline vanilla Deep JSCC models using standard CNN autoencoder architectures without AF blocks. Each specialist model is trained exclusively for a single, fixed SNR value from the set $\{0, 10, 20\}$ dB, representing low, medium, and high-quality channel conditions respectively.
- **Evaluation Protocol**: All models undergo evaluation across a comprehensive discrete SNR range of $[-10, -8, ..., 18, 20]$ dB using the CIFAR-10 test set to assess generalization capabilities under diverse channel conditions.
- **Performance Metrics**: We employ three complementary image quality assessment metrics [21]:
  1) **Peak Signal-to-Noise Ratio (PSNR)**: Quantifies reconstruction fidelity in decibels (higher values indicate better quality).
  2) **Structural Similarity Index Measure (SSIM)**: Evaluates perceptual similarity considering luminance, contrast, and structural information (values range from 0 to 1, with 1 representing perfect similarity) [21].
  3) **Learned Perceptual Image Patch Similarity (LPIPS)**: Employs pre-trained deep networks to measure perceptual distances between images (lower values indicate greater perceptual similarity) [22].

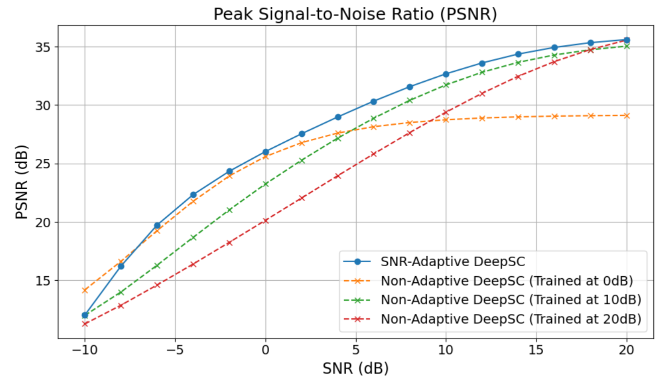

Fig. 1. Peak Signal-to-Noise Ratio (PSNR) versus SNR for the proposed SNR-Adaptive DeepSC and vanilla CNN baselines trained at fixed SNRs.

B. Results and Analysis

The experimental results, illustrated in Figures 1, 2, and 3, demonstrate the substantial advantages of our SNR-adaptive approach over fixed-SNR specialist models across all evaluated metrics.

Peak Signal-to-Noise Ratio Analysis: Fig. 1 reveals that each specialist vanilla model achieves optimal performance only within a narrow range surrounding its training SNR, consistent with findings in traditional Deep JSCC literature [5]. Specifically, the model trained at 0 dB demonstrates superior performance in low-SNR conditions (approximately -10 to 5 dB) but experiences rapid degradation as channel quality improves. Conversely, the model trained at 20 dB excels in high-SNR environments but exhibits catastrophic failure under noisy conditions, with PSNR values dropping below 15 dB at negative SNRs. The model trained at 10 dB provides

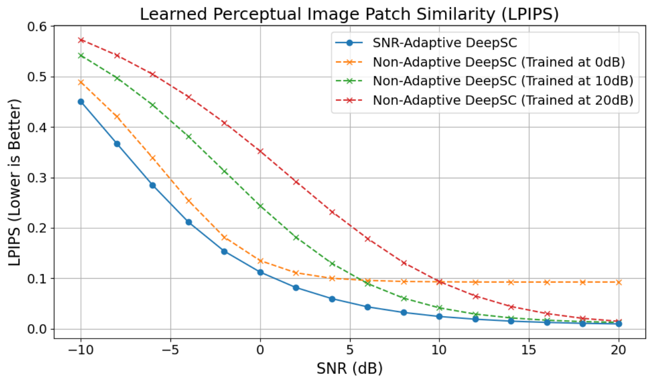

Fig. 2. Learned Perceptual Image Patch Similarity (LPIPS) versus SNR comparing the adaptive model with fixed-SNR baselines.

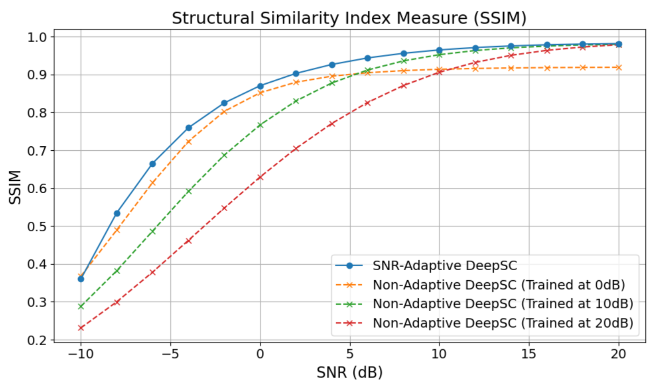

Fig. 3. Structural Similarity Index Measure (SSIM) versus SNR across the same set of models and evaluation points.

moderate performance across medium-range SNRs but fails to achieve the peak performance of either extreme specialist. In contrast, our SNR-adaptive model maintains consistently high PSNR values across the entire evaluated range, achieving near-optimal performance comparable to each specialist within their respective optimal regions while providing significantly superior robustness under mismatched conditions, similar to the adaptive approaches demonstrated in recent Deep JSCC research [23].

Perceptual Quality Assessment: Fig. 2 shows that the adaptive model consistently achieves lower LPIPS values (indicating better perceptual quality) across most SNRs, whereas fixed-SNR models incur noticeably larger perceptual distances when operated away from their design points. The LPIPS metric, which employs pre-trained deep networks to measure perceptual distances between images [22], provides critical insights into perceptual reconstruction quality beyond pixel-level metrics. The results demonstrate that specialist models suffer from significant perceptual artifacts when operating outside their design parameters, with LPIPS values exceeding 0.5 in extreme cases. Our adaptive model consistently maintains low LPIPS values below 0.1 across the majority of tested conditions, indicating superior perceptual quality preservation. This performance characteristic proves particularly important for UAV applications where visual interpretation by human operators or computer vision algorithms requires high perceptual fidelity [10], [16].

Structural Similarity Analysis: Fig. 3 confirms that structural similarity gains follow the same trend as the PSNR results. The SSIM metric, which evaluates perceptual similarity considering luminance, contrast, and structural information [21], demonstrates that our adaptive model reaches high SSIM values at moderate-to-high SNR without collapsing at low SNR, whereas each specialist saturates near its training point and drops markedly otherwise. This stability is consistent with the encoder-decoder residual design enhanced with GDN layers [19] and the channel-aware AF blocks applied after each stage, which together enable selective emphasis of structure-preserving channels in adverse conditions while restoring fine details as the channel improves. The adaptive model maintains SSIM values above 0.8 across most evaluated conditions and approaches 0.95 at high SNRs, indicating excellent preservation of structural information regardless of channel quality.

Robustness and Practical Implications: The comprehensive evaluation across all three metrics establishes that our single SNR-adaptive model provides a practical and efficient solution for real-world UAV deployments, addressing the channel variability challenges identified in recent semantic communication research [9], [24]. Unlike specialist models that require a priori knowledge of channel conditions and potentially multiple model deployments, our approach offers consistent performance with a single unified model. The graceful degradation characteristics and superior adaptation capabilities make it particularly suitable for dynamic wireless environments typical of UAV operations [1]. The attention mechanism integration enables dynamic feature modulation based on channel conditions, providing the flexibility necessary for robust performance across diverse operating environments while maintaining computational efficiency suitable for UAV platforms with limited processing capabilities. Similar adaptation needs also arise in integrated terrestrial/non-terrestrial networks with highly dynamic access, underscoring the value of channel-aware scheduling and control [25].

V. Conclusion

This paper addresses the fundamental challenge of channel variability in UAV image transmission systems by proposing a novel SNR-adaptive Deep JSCC framework. Our approach employs an innovative attention mechanism that conditions network behavior on instantaneous channel quality, enabling a single model to adapt dynamically across diverse operating conditions. Through comprehensive experimental validation, we demonstrate substantial improvements in robustness compared to traditional static models trained for fixed SNRs.

The key innovation of our Attention Feature blocks enables the network to learn context-aware feature modulation strategies that optimize transmission for specific channel conditions. Our experimental results conclusively show that the adaptive model maintains high reconstruction quality across extensive

SNR ranges, providing a far more practical solution for real-world UAV operations where channel quality fluctuates unpredictably. The single-model approach significantly reduces storage and computational requirements compared to deploying multiple specialist models, making it particularly attractive for resource-constrained UAV platforms.

Future research directions include investigating the integration of realistic UAV mobility models and channel fading effects beyond AWGN, extending the framework to more complex time-varying fading channels, and exploring joint optimization of UAV trajectory planning with adaptive communication strategies. Additionally, the incorporation of semantic segmentation and object detection tasks could further enhance the practical utility of the proposed framework for autonomous UAV missions. The extension to other image datasets and evaluation under practical channel estimation errors represents another promising avenue for future work.

ACKNOWLEDGMENTS

This work was supported in part by the IITP (Institute of Information Communications Technology Planning Evaluation) - ITRC (Information Technology Research Center) (IITP-2026-RS-2022-00156353, 50%) grant funded by the Korea government (Ministry of Science and ICT) and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00453301).

REFERENCES

[1] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on uavs for wireless networks: Applications, challenges, and open problems," *IEEE communications surveys & tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.

[2] W. J. Yun, S. Park, J. Kim, M. Shin, S. Jung, D. A. Mohaisen, and J.-H. Kim, "Cooperative multiagent deep reinforcement learning for reliable surveillance via autonomous multi-uav control," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7086–7096, 2022.

[3] S. Park, C. Park, and J. Kim, "Learning-based cooperative mobility control for autonomous drone-delivery," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 4, pp. 4870–4885, 2023.

[4] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[5] E. Bourtsoulatze, D. B. 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, 2019.

[6] X. Gu and G. Zhang, "A survey on uav-assisted wireless communications: Recent advances and future trends," *Computer Communications*, vol. 208, pp. 44–78, 2023.

[7] J. Kim, G. Caire, and A. F. Molisch, "Quality-aware streaming and scheduling for device-to-device video delivery," *IEEE/ACM Transactions on Networking*, vol. 24, no. 4, pp. 2319–2331, 2015.

[8] M. Choi, J. Kim, and J. Moon, "Wireless video caching and dynamic streaming under differentiated quality requirements," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1245–1257, 2018.

[9] K. Chi, Y. He, Q. Yang, Z. Jiang, Y. Shu, Z. Wang, J. Luo, and J. Chen, "Deepstream: Prototyping deep joint source-channel coding for real-time multimedia transmissions," *arXiv preprint arXiv:2509.05971*, 2025.

[10] F. Wan, M. B. Yaseen, M. B. Riaz, A. Shafiq, A. Thakur, and M. O. Rahman, "Advancements and challenges in uav-based communication networks: A comprehensive scholarly analysis," *Results in Engineering*, vol. 24, p. 103271, 2024.

[11] J. Oh, D. Lee, D. S. Lakew, and S. Cho, "Dacode: Distributed adaptive communication framework for energy efficient industrial iot-based heterogeneous wsn," *ICT Express*, vol. 9, no. 6, pp. 1085–1094, 2023.

[12] M. C. Ho, A. T. Tran, D. Lee, J. Paek, W. Noh, and S. Cho, "A ddpg-based energy efficient federated learning algorithm with swipt and mc-noma," *ICT Express*, vol. 10, no. 3, pp. 600–607, 2024.

[13] J. Xu, B. Ai, W. Chen, A. Yang, P. Sun, and M. Rodrigues, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 2315–2328, 2021.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[15] J. Meng, Z. Wei, Y. Zhang, B. Li, and C. Zhao, "Machine learning based low-complexity channel state information estimation," *EURASIP Journal on Advances in Signal Processing*, vol. 2023, no. 1, p. 98, 2023.

[16] W. J. Yun, D. Kwon, M. Choi, J. Kim, G. Caire, and A. F. Molisch, "Quality-aware deep reinforcement learning for streaming in infrastructure-assisted connected vehicles," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2002–2017, 2021.

[17] J. G. Proakis and M. Salehi, *Digital communications*. McGraw-hill New York, 2001, vol. 4.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

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

[20] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[22] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[23] H. Wu, Y. Shao, C. Bian, K. Mikolajczyk, and D. Gündüz, "Deep joint source-channel coding for adaptive image transmission over mimo channels," *IEEE Transactions on Wireless Communications*, 2024.

[24] M. Tang, C. Feng, and T. Q. Quek, "Decentralized semantic communication and cooperative tracking control for a uav swarm over wireless mimo fading channels," *IEEE Transactions on Vehicular Technology*, 2025.

[25] S. Park, G. S. Kim, Z. Han, and J. Kim, "Quantum multi-agent reinforcement learning is all you need: Coordinated global access in integrated tn/ntn cube-satellite networks," *IEEE Communications Magazine*, vol. 62, no. 10, pp. 86–92, 2024.