Samarathunga, P., Fernando, T., Ganearachchi, Y., Pollwaththage, N., & Fernando, A. (2026). Semantic communication and residual deep learning-driven bitplane coding for scalable image compression and communications. In 2025 IEEE International Conference on Image Processing Workshops (ICIPW) (pp. 568-573). IEEE. https://doi.org/10.1109/icipw68931.2025.11385999.

SEMANTIC COMMUNICATION AND RESIDUAL DEEP LEARNING-DRIVEN BITPLANE CODING FOR SCALABLE IMAGE COMPRESSION AND COMMUNICATIONS

Prabhath Samarathunga, Thanuj Fernando, Yasith Ganearachchi, Nimesh Pollwaththage and Anil Fernando

Department of Computer and Information Sciences
University of Strathclyde, Glasgow, U.K.

ABSTRACT

Digital imaging applications which require both reconstruction and semantic accuracy pose challenges to both conventional codecs, which fail to preserve semantics in very noisy conditions, and learned codecs, which fail to achieve reconstruction accuracy even though they preserve semantics. To address this challenge, we propose a semantic autoencoder that preserves high-level image structures in a compact latent representation combined with a progressive bitplane refinement mechanism employing Enhancement Networks and Bitplane Correction Functions to minimize reconstruction errors, which also effectively prevents error propagation across bitplanes while enabling flexible rate adaptation. Experimental results demonstrate significant improvements over conventional image compression codecs, achieving up to 10 dB higher PSNR and 0.12 higher SSIM at low bitrates. In particular, the proposed system reaches quality saturation at just 3.2 bpp, while conventional codecs require 8 bpp to achieve comparable performance. These advances establish a new paradigm for scalable compression, where latent semantics guide iterative refinement.

Index Terms— Bitplane coding, Bitplane refinements, Scalable image compression, Semantic communication

1. INTRODUCTION

The exponential growth of digital imaging across applications ranging from mobile photography to medical diagnostics has created an unprecedented demand for efficient compression technologies that preserve both perceptual quality and semantic content. While conventional image compression formats such as the Joint Photography Experts Group (JPEG) codecs including JPEG2000, and the High Efficiency Image Format (HEIF) have served as industry standards for decades, their reliance on handcrafted transforms and fixed compression paradigms limits their effectiveness in modern use cases where they often fail to maintain critical image semantics at aggressive compression ratios. In contrast, deep learning-based alternatives can maintain image semantics, but frequently neglect the precise bit-level control needed for scalable compression.

This paper introduces a groundbreaking image compression framework that bridges the gap between semantic-aware representation and precision bitplane coding. The proposed codec makes three fundamental contributions. Firstly, we present a semantic autoencoder architecture that learns to preserve high-level image structures in an extremely compact latent representation. Secondly, we develop a novel bitplane refinement system combining Enhancement Networks (EN) and Bitplane Correction Functions (BCF) that progressively reconstructs image details to achieve very high bitplane accuracy. Third, we demonstrate how context-adaptive binary arithmetic coding (CABAC) can optimally compress both semantic and residual information in a unified framework.

The proposed system has the potential to revolutionize scalable image compression by simultaneously achieving two critical advances that elude conventional methods, which are structural preservation at ultralow bitrates and seamless progression to lossless quality reconstruction at higher bitrates. Through comprehensive evaluation, our framework demonstrates consistent improvements across all scalable quality layers, delivering up to 10 dB improvements in peak signal-to-noise ratio (PSNR) compared to established standards at equivalent bitrates. A particularly noteworthy achievement is reaching visual quality saturation at merely 3.2 bpp, whereas traditional codecs require more than 8 bpp to attain comparable fidelity.

These performance improvements stem from our novel approach of integrating semantic-aware base coding with precision bitplane enhancement, enabling highly efficient bandwidth utilization. This creates new opportunities for its use in demanding applications, including high-precision telemedicine diagnostics, adaptive mobile visual communication, and bandwidth-constrained professional imaging workflows. The progressive refinement capability of the proposed framework ensures optimal quality at every decode point, making it particularly valuable for scenarios where transmission interruptions or bandwidth fluctuations might otherwise compromise image utility.

2. RELATED WORK

Classical image compression has been led for decades by standards such as JPEG [1], which uses the Discrete Cosine Transform (DCT), quantization, and Huffman coding. Although effective, JPEG suffers from blocking artifacts at low bitrates and lacks scalability. JPEG2000 [2] improved upon this with wavelet-based compression and progressive transmission, offering better rate-distortion performance. However, both rely on fixed, hand-crafted transforms that may not generalize well to modern image content. More recent standards such as HEIF based on High Efficieny Video Coding (HEVC)[3], and Versatile Video Coding (VVC) [4] integrate advanced prediction and filtering techniques from video coding, achieving better compression efficiency, but still adhering to the conventional transform-based paradigm.

The advent of deep learning has transformed image compression. Ballé et al. [5] introduced a variational autoencoder (VAE) approach with learned entropy models, outperforming classical methods. Follow-up works have incorporated hyperprior models [6] and attention mechanisms to improve adaptability. Recently, transformer-based architectures [7] have been proposed, extending VAE frameworks with enhanced encoder-decoder structures.

Scalable image compression enables progressive quality refinement by supporting multiple layers of resolution, quality, or bit-depth. The JPEG2000 embedded block coding with optimized transcription of embedded bit streams (EBCOT) [8] achieves fine-grained rate scalability by bit plane coding of wavelet coefficients, while scalable video coding (SVC) [9] allows spatial, temporal, and quality scalability through interlayer prediction. Traditional scalable codecs use hierarchical transform decomposition, but face challenges like error propagation and inefficient rate allocation. Progressive coding methods [10] partially address these issues through adaptive quantization and prioritized bit allocation.

Recent progress in semantic image communication has markedly improved the efficiency and dependability of image transmission systems through context-aware compression methods and advanced deep learning techniques. Foundational research on region-of-interest scalable image compression [11] and context-sensitive semantic coding [12] has paved the way for the incorporation of joint source-channel coding frameworks [13] and edge-supported processing architectures [14] to achieve reliable, low-latency communication. Furthermore, the use of reinforcement learning for optimized resource allocation [15] and innovative coding approaches designed for machine vision applications [16] has driven significant performance gains. Privacy-preserving techniques that leverage federated learning [17] and perceptually focused quality assessment methods [18] have addressed essential security and user-oriented concerns. A comprehensive end-to-end wireless image transmission system that embodies these integrated semantic communication principles highlights the transformative potential of these advances for next-generation networks [19].

Although the existing literature has made significant contributions to individual aspects of image compression, such as deep learning-based coding, semantic communication, scalable compression, and bitplane processing, there remains a critical gap in unifying these approaches into a coherent framework. Traditional scalable codecs suffer from error propagation and suboptimal semantic preservation, while semantic communication approaches have not been effectively integrated with practical scalable compression requirements. Our work addresses these limitations by introducing the first integrated framework that combines semantic-aware representation learning with neural bitplane coding, enabling both semantic interpretability and fine-grained rate adaptation while maintaining compression efficiency.

3. PROPOSED SYSTEM

The proposed system introduces a scalable bitplane compression framework that combines semantic feature extraction with hierarchical bitplane refinement. The architecture consists of an encoder-decoder pipeline designed to process RGB images through two complementary stages: a semantic-aware base compression layer and a multistage bitplane enhancement mechanism.

The initial stage employs a semantic encoder implemented using the encoding layers of an Autoencoder (AE) to compress the input image into a latent representation, capturing high-level semantic features through a series of convolutional and downsampling layers [20]. This latent vector is quantized and encoded using context-adaptive binary arithmetic coding (CABAC) [21] for efficient transmission. The corresponding semantic decoder implemented using the decoding layers of the same autoencoder (AD) reconstructs a base approximation of the image from this latent representation, serving as the foundation for subsequent bitplane refinements. By prioritizing semantic content in this initial stage, the system ensures robust preservation of structural features (e.g., edges, textures) even under aggressive compression. After generating the semantic-aware base layer reconstruction, we calculate the residual image by subtracting this reconstruction from the original input image. This residual image then undergoes a sophisticated bitplane processing procedure, which makes use of BCF and EN blocks to achieve improved performance compared to conventional scalable image coding methods.

3.1. Bitplane Correction Function (BCF)

For a given bitplane i (where $i \in \{8, 7, \ldots, 1\}$), we first calculate the accumulated value of all higher bitplanes (already corrected) as shown in (1).

$$V_{corr}^{(i)} = \sum_{i=k}^{8} b_k \cdot 2^{k-1} \tag{1}$$

where $b_k \in \{0,1\}$ represents the corrected binary value of the k-th bitplane.

Let V_{pred} be the pixel value predicted by AD for the current stage. We define the bitplane correction rule as follows: let the corrected pixel value be defined by $V_{final}^{(i)}$ and it can be determined by (2).

$$V_{final}^{(i)} = \begin{cases} V_{corr}^{(i)} & \text{if } V_{pred} < V_{corr}^{(i)} \\ V_{pred} & \text{if } V_{pred} \geq V_{corr}^{(i)} \end{cases} \tag{2}$$

When $V_{\text{pred}} < V_{\text{corr}}^{(i)}$, this indicates that the predicted value is inconsistent with the higher bitplanes already corrected, so we enforce consistency by using $V_{corr}^{(i)}$. When $V_{\text{pred}} \geq V_{\text{corr}}^{(i)}$, we trust the decoder's prediction for the remaining uncorrected bits. The floor/modification operation in (3) properly extracts the corrected bit value while preserving higher bitplane corrections. This formulation mathematically captures your described logic while maintaining proper bitplane arithmetic and correction propagation. Based on the above, we can update the bitplane for the i^{th} bitplane as given in (3).

$$b_i^{\text{corrected}} = \left\lfloor \frac{V_{\text{final}}^{(i)}}{2^{i-1}} \right\rfloor \mod 2 \tag{3}$$

3.2. Enhancement Network (EN)

The proposed Enhancement Network (EN) is a deep learning module designed for progressive refinement of compressed images in a scalable bitplane coding framework. Based on a modified U-Net architecture, the EN predicts and corrects residual errors between bitplanes using both the base layer reconstruction from a semantic autoencoder and previously decoded higher bitplanes. The U-Net encoder-decoder structure with skip connections enables it to capture multiscale features, preserving global and local details during reconstruction. Each EN is trained on the CIFAR dataset. For example, EN8 is trained on the auto-decoder output images, and EN7 is trained on the corrected images of bitplane 8. Each stage consists of pairs of convolutional layers, followed by batch normalization and ReLU activation, with a sigmoid activation applied at the final layer.

The EN and BCF integrate with CABAC, which compresses only the corrected residuals. This efficient pipeline achieves high compression with minimal quality loss. The progressive structure allows stopping at any bitplane, enabling flexible and scalable compression.

The end-to-end system decomposes the residual image into 8 bitplanes, processing them sequentially from the Most Significant Bitplane (MSB) to the Least Significant Bitplane (LSB). The processing begins with the MSB (8th bitplane), where our EN predicts the MSB values from the base layer reconstruction. This is then compared with the original MSB to generate an error signal which is losslessly encoded using CABAC. This encoded error signal ensures perfect reconstruction of the MSB on the decoder side. The system utilizes error-corrected MSB to predict subsequent bitplanes through a specialized U-Net architecture [22].

For each bitplane from the 7th to the 1st, the process follows a rigorous refinement pipeline. First, the current bitplane is predicted using the proposed U-Net enhancement network. This prediction then undergoes correction through our BCF before the resulting residual is encoded using CABAC. This dual-stage prediction-correction approach ensures optimal accuracy before final encoding.

The enhancement networks employ an U-Net architecture specifically optimized for bitplane prediction. Each network takes as input both the current reconstruction state and the preceding bitplane data, enabling it to predict the residual errors between the intermediate reconstruction and the target bitplane with high precision. The BCF then applies an additive correction to these predictions, incorporating error feedback from previously processed bitplanes to progressively refine the reconstruction.

This cascaded correction mechanism addresses the susceptibility to error propagation across bitplanes, which is a critical limitation of traditional scalable image compression methods. Our system's iterative refinement at each bitplane level provides three key advantages: immediate error correction prevents the accumulation of artifacts, each correction informs and improves subsequent predictions, and the process maintains consistent performance across varying scalable compression ratios.

The hierarchical nature of this approach ensures a differentiated treatment of visual information. Higher bitplanes, containing more critical structural data, benefit from direct correction and precise encoding. Lower bitplanes, representing finer details, gain accuracy through the cumulative effect of preceding corrections. This strategic allocation of processing resources contributes significantly to our codec's superior rate-distortion performance while maintaining computational efficiency.

The complete pipeline is illustrated in Fig. 1, forming a closed-loop system where each stage's output progressively refines the next stage's input. This iterative refinement mechanism enables reconstruction quality that consistently exceeds conventional compression methods. The semantic encoder-decoder pair in our architecture is built on [20], optimized here for scalable bitplane processing. The following two sections detail the bitplane refinement process - the core innovation enabling our codec's superior scalability and precision.

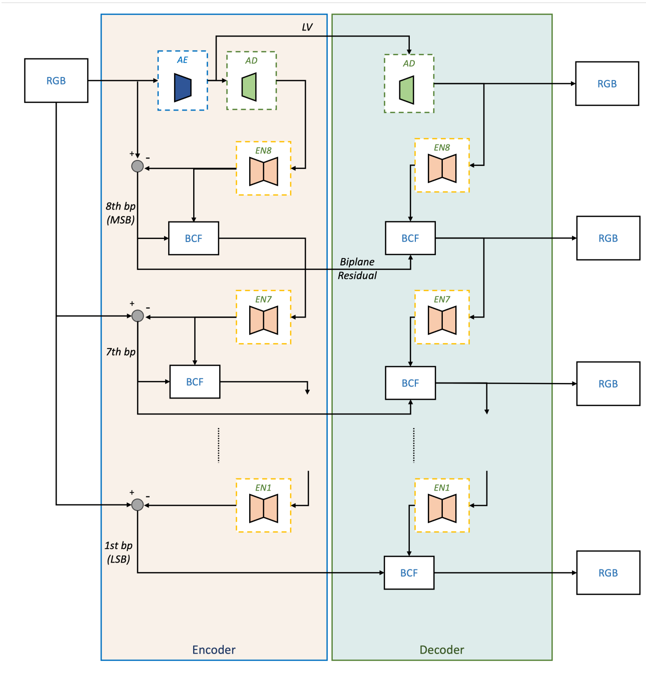
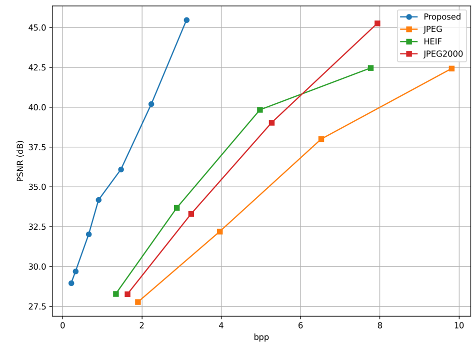
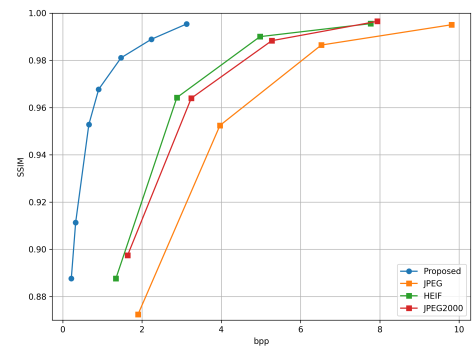


Fig. 1: Architecture of Proposed System



(a) bpp vs. PSNR



(b) bpp vs. SSIM

Fig. 2: Rate Distortion Performance of the Proposed Scalable Codec

4. RESULTS AND DISCUSSION

To evaluate the performance of the proposed scalable bitplane compression system, we performed experiments comparing it against JPEG, HEIF, and JPEG2000. The evaluation metrics used are Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), which assess the quality of reconstructed images across varying bits per pixel (bpp) rates. The experiments were performed on the CIFAR-10 dataset [23], which consists of 60,000 RGB images, ensuring a diverse range of image types for robust evaluation.

4.1. Rate-Distortion Performance Evaluation

The proposed system demonstrates superior compression efficiency across all tested bitrates, as evidenced by its dominant performance in both PSNR and SSIM metrics as shown in Fig. 2. At low bitrates (0-2 bpp), where conventional codecs typically struggle to maintain fidelity, the proposed method achieves a remarkable 10 dB PSNR improvement over JPEG and 8 dB gains over both HEIF and JPEG2000. In terms of perceptual quality, the system provides SSIM improvements of 0.12, 0.09, and 0.10 against JPEG, HEIF, and JPEG2000, respectively, which is a clear testament to its effective prioritization of perceptually critical data through advanced bitplane correction.

The proposed system also maintains its performance advantage as bitrates increase from 2 to 6 bpp, demonstrating near-linear quality improvement that reaches SSIM > 0.98 by 4 bpp, significantly better than conventional codecs. Quality saturation occurs at just 6 bpp (PSNR $\approx$45 dB, SSIM $\approx$1.0), showing the proposed method achieves optimal reconstruction quality while requiring approximately half the bitrate of HEIF and JPEG2000, and one-third the bitrate of JPEG for equivalent fidelity. This early saturation point highlights the remarkable efficiency of our hybrid architecture, where the semantic base layer effectively preserves global image structures while the bitplane enhancement mechanism precisely allocates the remaining bandwidth to critical local details.

Visual evaluations provide compelling evidence to support these quantitative metrics. Across all tested bitrates, the proposed reconstructions exhibit noticeably sharper edges and cleaner textures compared to conventional codecs. At 2 bpp, a particularly challenging operating point for bandwidth-constrained applications, the advantages become especially apparent. The proposed method completely avoids the characteristic blocking artifacts of JPEG and the excessive smoothing exhibited by HEIF, while maintaining superior preservation of fine textural details and object boundaries that are often lost in traditional approaches.

The consistent performance gaps observed in both PSNR and SSIM metrics demonstrate that the proposed architecture successfully overcomes the fundamental trade-off between pixel-level accuracy and perceptual quality that has long challenged image compression systems. This breakthrough is directly attributable to the proposed novel combination of semantic-aware encoding and systematic bitplane refinement, which work in concert to optimize objective and subjective quality measures throughout the entire bitrate range.

4.2. Reduction in Bitplane Errors Using BCF and EN

Table 1 presents the percentage reduction in bitplane errors achieved by BCF and EN on bitplanes 1 to 8, compared to the initial bitplane errors. These results demonstrate the effectiveness of the proposed system in mitigating errors during the compression process, which is critical to maintaining image quality in a scalable bit-plane compression framework. This suggests that the proposed system is particularly effective in correcting errors in bitplanes that contribute significantly to image quality.

Bitplane	BCF Reduction (%)	EN Reduction (%)
8	-	10.53%
7	26.61%	4.58%
6	64.72%	9.24%
5	43.42%	13.16%
4	30.09%	14.42%
3	20.45%	15.45%
2	11.11%	13.60%
1	2.33%	5.12%

Table 1: Bitplane-wise BCF and EN Reduction Percentages

4.3. Comparative Insights

The superior performance of the proposed system can be attributed to its use of an auto-encoder-based semantics encoder and bit-plane correction mechanism, which effectively capture and preserve semantic features during compression. The EN and BCF further refine the bitplane data, improving the reconstruction quality, which is needed for a scalable image codec. In contrast, traditional methods like JPEG and JPEG2000 struggle at lower bpp rates due to their reliance on frequency-domain transformations, which can introduce noticeable distortions. HEIF, although more advanced, does not match the scalability and fidelity of the proposed system, particularly at lower bit rates. These results validate the effectiveness of the proposed architecture in achieving high compression efficiency without sacrificing image quality, making it a promising solution for scalable image compression applications.

4.4. Complexity Analysis

JPEG, JPEG2000, and HEIF exhibit progressively higher computational complexity, with JPEG remaining the simplest due to its block-based DCT and entropy coding architecture. JPEG2000 increases complexity through wavelet transforms and EBCOT, achieving better rate distortion performance than JPEG while supporting scalable decoding. HEIF introduces the highest complexity among traditional codecs through advanced prediction modes and transform coding, delivering superior compression efficiency.

Although JPEG offers the fastest processing, it suffers from blocking artifacts. JPEG2000 provides a balance of speed and quality with inherent scalability, and HEIF achieves the best compression at the cost of greater computational demands. The proposed semantic-aware codec introduces marginally higher complexity than traditional codecs during inference due to its neural network components, but this is substantially outweighed by its breakthrough performance across all scalable layers, demonstrating unmatched compression efficiency while maintaining perceptual quality even at ultra-low bitrates. The architectural efficiency of its pre-trained models ensures this complexity increase remains practical for real-world deployment.

5. CONCLUSION

This paper presented a novel scalable image compression framework that synergistically combines semantic-aware encoding with hierarchical bitplane refinement. The proposed system establishes new state-of-the-art performance, demonstrating consistent superiority over conventional codecs (JPEG, HEIF, JPEG2000) across all evaluated bitrates. Key innovations of the propsed system include an autoencoder-based semantic compression layer that preserves critical structural information and a progressive bitplane correction mechanism that systematically reduces reconstruction errors through EN and BCF. The success of the proposed approach comes from its unique ability to combine the strengths of semantic compression with precise bitplane processing. The semantic base layer efficiently encodes global features, while the bitplane enhancement mechanism sequentially allocates residual bandwidth to local details through an iterative bitplane correction. This dual strategy effectively resolves the traditional trade-off between compression ratio and reconstruction quality. Future work will focus on extending the framework in exploring scalable video compression and other media coding tasks where scalable, quality-adaptive compression is crucial. The code and models will be made publicly available to support further advancements in learned image compression.

6. REFERENCES

[1] Gregory K. Wallace, "The jpeg still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, Apr. 1991.

[2] David S. Taubman and Michael W. Marcellin,

JPEG2000 Image Compression Fundamentals, Standards and Practice, The Springer International Series in Engineering and Computer Science. Springer New York, NY, 1 edition, 2002, Copyright Kluwer Academic Publishers 2002.

[3] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[4] Benjamin Bross, Jianle Chen, Jens-Rainer Ohm, Gary J. Sullivan, and Ye-Kui Wang, "Developments in international video coding standardization after avc, with an overview of versatile video coding (vvc)," *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1463–1493, 2021.

[5] Johannes Ballé, Valero Laparra, and Eero P. Simoncelli, "End-to-end optimized image compression," 2017.

[6] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston, "Variational image compression with a scale hyperprior," 2018.

[7] Ming Lu, Peiyao Guo, Huiqing Shi, Chuntong Cao, and Zhan Ma, "Transformer-based image compression," 2021.

[8] D. Taubman, "High performance scalable image compression with ebcot," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, 2000.

[9] Heiko Schwarz, Detlev Marpe, and Thomas Wiegand, "Overview of the scalable video coding extension of the h.264/avc standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[10] A. Said and W.A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.

[11] P. Samarathunga, V. Gowrisetty, T. Fernando, Y. Ganearachchi, and A. Fernando, "Region of interest scalable image compression using semantic communications," in *2024 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2024, pp. 1–5.

[12] A. Zhang, B. Li, and C. Wang, "Deep learning for semantic image compression: A context-aware approach," *IEEE Transactions on Image Processing*, vol. 34, no. 2, pp. 876–889, Feb. 2025.

[13] D. Chen, E. Wu, and F. Liu, "Semantic image transmission in 6g: A joint source-channel coding framework," *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 5, pp. 1123–1136, May 2025.

[14] G. Zhao, H. Yang, and I. Park, "Edge-assisted semantic image processing for real-time surveillance systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 6, pp. 1450–1463, June 2025.

[15] J. Lee, K. Sun, and L. Zhang, "Reinforcement learning-based resource allocation for semantic image transmission in iot networks," *IEEE Internet of Things Journal*, vol. 12, no. 8, pp. 7654–7667, Aug. 2025.

[16] M. Xu, N. Ding, and O. Tang, "From pixels to semantics: Advances in image coding for machine vision," *IEEE Transactions on Multimedia*, vol. 27, no. 9, pp. 2345–2358, Sept. 2025.

[17] P. Zhou, Q. Yu, and R. Guo, "Privacy-preserving semantic image sharing using federated learning," *IEEE Transactions on Information Forensics and Security*, vol. 20, no. 11, pp. 3210–3223, Nov. 2025.

[18] S. Kim, T. Wang, and U. Nguyen, "Perceptual quality assessment for semantic image communications: A deep learning approach," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 21, no. 4, pp. 1–24, Dec. 2025.

[19] M. U. Lokumarambage, V. S. S. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva, and A. Fernando, "Wireless end-to-end image transmission system using semantic communications," *IEEE Access*, vol. 11, pp. 37149–37163, 2023.

[20] Prabhath Samarathunga, Hossein Rezaei, Maheshi Lokumarambage, Thushan Sivalingam, Nandana Rajatheva, and Anil Fernando, "An autoencoder-based task-oriented semantic communication system for m2m communication," *Algorithms*, vol. 17, no. 11, 2024.

[21] Christian Rohlfing, Thibaut Meyer, Jens Schneider, and Jan Voges, "Python wrapper for context-based adaptive binary arithmetic coding," in *Visual Communications and Image Processing VCIP '23*, Jeju, South Korea, Dec. 2023, IEEE, Piscataway.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[23] Alex Krizhevsky, "Learning multiple layers of features from tiny images," https://www.cs.toronto.edu/ kriz/cifar.html, 2009, Accessed: 2025-05-28.