UDC 621.391.1

SEMANTIC COMMUNICATION SYSTEM FOR VEHICLE DETECTION

QIKAI WANG

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 08, 2026

Abstract. This paper presents a semantic communication system for vehicle detection over noisy wireless channels. A Convolutional Neural Network (CNN) semantic encoder compresses image features with multiple compression ratios to transmit task-related information only, which is then processed by a pre-trained You Only Look Once version 12 (YOLOv12) object detector. A Supernet with shared encoder weights and Sandwich Rule training is proposed to reduce redundancy. Experiments on Additive White Gaussian Noise (AWGN) channels show that the semantic system significantly outperforms traditional transmission at low signal-to-noise ratio (SNR), and the Supernet achieves comparable performance with independent models.

Keywords: semantic communication, vehicle detection, YOLOv12, CNN encoder, supernet, AWGN channel, compression ratio.

Introduction

The rapid proliferation of intelligent transportation systems (ITS) has driven growing demand for reliable vehicle detection over wireless communication channels. In bandwidth-constrained or high-noise environments, conventional image transmission methods degrade significantly because they transmit the full raw pixel stream regardless of its task relevance. When channel quality is poor, compressed image artifacts and AWGN noise corrupt pixel values irreversibly before reaching the inference engine, causing catastrophic detection failure.

Semantic communication offers a fundamentally different paradigm: instead of transmitting the image, only the features necessary to accomplish the downstream task are encoded and sent [1, 2]. This approach can provide strong robustness to channel noise because the encoder learns compact, task-aware representations that are far less sensitive to moderate channel distortion than raw pixel values [3, 4]. In this work, we design a CNN-based semantic encoder for vehicle detection and systematically evaluate its robustness across five SNR levels under AWGN conditions. We further introduce a Supernet architecture that unifies three compression ratios (CH4, CH8, CH16) into a single model, reducing deployment overhead while preserving detection performance.

System Model

The proposed semantic communication pipeline consists of four components: a CNN semantic encoder, an AWGN channel simulation, a CNN semantic decoder, and a pre-trained and frozen YOLOv12 object detector. The overall architecture is illustrated in Figure 1. Given an input image, the encoder maps it to a compact feature vector. This compressed feature is then corrupted by additive white Gaussian noise, where the noise variance is determined by the target SNR level. Internally, the CNN semantic encoder is constructed using a series of convolutional layers paired with batch normalization and Leaky ReLU activation functions [5]. This hierarchical structure allows the network to progressively filter out irrelevant background details and focus strictly on high-level semantic features, such as vehicle edges, shapes, and textures. By compressing the raw image into a lower-dimensional semantic space, the feature map becomes inherently more robust to random variations caused by the channel noise. The decoder reconstructs a feature representation which is then passed to the pre-trained and frozen YOLOv12 detector to produce bounding box predictions. For the traditional baseline, the raw image is

JPEG-compressed, pixel values are directly corrupted by AWGN, and the noisy image is forwarded to YOLOv12 without any semantic encoding. The YOLOv12 model was specifically chosen as the downstream inference engine due to its state-of-the-art balance between detection accuracy and computational speed [6]. In intelligent transportation applications, real-time processing is just as critical as accuracy. By keeping the YOLOv12 weights frozen during the training of the semantic communication system, we ensure that the encoder learns to generate features that are perfectly aligned with the standard input requirements of advanced object detectors, making our system highly versatile and easy to deploy.

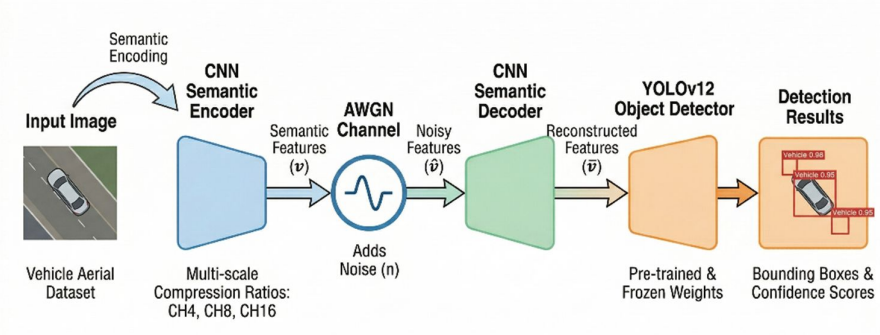

Figure 1. Architecture of the proposed semantic communication system

Supernet Architecture

Training three independent semantic encoders for different compression ratios is computationally redundant. We therefore design a Supernet architecture that shares encoder and decoder backbone weights across all three sub-networks, differing only in the final channel projection layer. Training is performed using the Sandwich Rule strategy: in each iteration, the largest sub-network, the smallest sub-network, and a randomly sampled intermediate sub-network are each evaluated and their gradients accumulated before a single optimizer step [8]. This ensures that all sub-networks receive balanced gradient updates and prevents the shared backbone from being dominated by any single compression configuration. The Supernet was trained for 50 epochs on a cloud server equipped with an NVIDIA RTX 4090, completing training in approximately 53 minutes-the same wall-clock time as a single independent encoder.

Experiments

In this section, we evaluate the performance of the proposed semantic communication system for vehicle detection under various channel conditions. Experiments are conducted on a multi-class vehicle aerial image dataset containing 5,171 training images. Evaluation is performed on a held-out validation set under five SNR conditions: 30, 20, 10, 5, and 0 dB. To provide a comprehensive evaluation, multiple performance metrics are tracked during testing. The primary metric is the Mean Average Precision at an intersection-over-union threshold of 0.5 (mAP@0.5), which effectively balances both bounding box localization accuracy and classification confidence. Additionally, Precision and Recall metrics are analyzed to understand the model's behavior in extreme noise, specifically tracking whether the system suffers more from false positives or missed detections as the signal quality degrades.

Performance Comparison: Semantic vs. Traditional Transmission

The traditional method achieves the highest average precision at 30 dB due to lossless high-SNR conditions, but collapses sharply below 10 dB, reaching a near-zero detection rate at 0 dB. In contrast, the semantic transmission scheme degrades gracefully. As illustrated in Figure 2, the CH16 semantic encoder retains a mean average precision of 0.414 at 0 dB-a relative decrease of only 8.2 percent from its 30 dB performance. At 0 dB, the semantic system provides a massive performance improvement over traditional transmission.

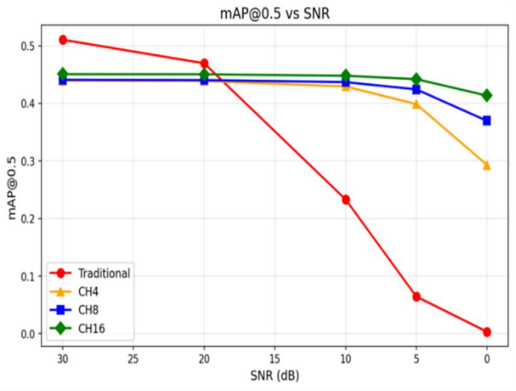

Figure 2. Semantic vs traditional transmission performance vs SNR

From a qualitative perspective, the visual difference between the two methods at 0 dB is striking. Under traditional transmission at this extreme noise level, the reconstructed images are completely overwhelmed by colorful static noise, rendering vehicle outlines entirely indistinguishable to both human eyes and computer vision algorithms. Consequently, the detector fails to propose any meaningful bounding boxes. Conversely, the feature maps generated by the semantic system preserve the essential structural outlines of the vehicles. Even though the channel is severely degraded, the YOLOv12 detector consistently maintains confidence scores above 0.5 for most vehicles, demonstrating the distinct advantage of task-oriented feature transmission.

Performance Comparison: Supernet vs. Independent Models

We further compare the performance of the proposed Supernet with independently trained encoder models to verify its efficiency. The Supernet sub-networks track the independently trained models closely across all SNR conditions. At the most challenging 0 dB condition, the highest-capacity sub-network achieves an average precision that is extremely close to the independent model, confirming the effectiveness of the Sandwich Rule training. The Supernet effectively replaces three independent models with one, resulting in a parameter cost reduction of approximately 67 percent with minimal performance degradation.

Conclusion

This paper presents a semantic communication system for vehicle detection that demonstrates strong robustness to AWGN channel noise. The proposed CNN semantic encoder significantly reduces detection degradation in extreme noise conditions compared to traditional image transmission. Furthermore, the Supernet architecture successfully consolidates multiple compression-ratio variants into a single model with minimal performance loss. Future work will investigate adaptive compression ratio selection based on estimated channel quality and the end-to-end joint optimization of the encoder-decoder-detector pipeline. The proposed system is highly suitable for low-SNR scenarios in intelligent transportation systems.

References

1. Qin Z., et al. // arXiv:2201.01389. 2022.
2. Bourtsoulatze E., et al. // IEEE Trans. Cogn. Commun. Netw. 2019. Vol. 5. P. 567–579.
3. Tung T.-Y., Gündüz D. // IEEE J. Sel. Areas Inf. Theory. 2022. Vol. 3. P. 720–731.
4. Wang X., et al. // IEEE J. Sel. Areas Commun. 2023. Vol. 41. P. 214–229.
5. Shi W., et al. // arXiv:2111.10044. 2021.
6. Redmon J., Farhadi A. // arXiv:1804.02767. 2018.
7. Cai H., et al. // ICLR. 2020.
8. Yu J., Yang L., Xu N., Yang J., Huang T. // ICLR. 2019.