

Article

Network Coding Enhanced Semantic Communications in Internet of Vehicles

Yanzhou Wang ^{1,2}, Jiahang Zhong ²  and Congduan Li ^{2,*}¹ International School of BUPT, Beijing University of Posts and Telecommunications, Beijing 100876, China; 2686705997@bupt.edu.cn² School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 510006, China; zhongjh58@mail2.sysu.edu.cn

* Correspondence: licongd@mail.sysu.edu.cn

Abstract

Low-latency visual information sharing is a key enabler for cooperative perception in vehicular networks. Network coding (NC) can exploit wireless superposition and side information to improve spectral efficiency in bidirectional relaying. This paper presents an end-to-end learned framework for Roadside Units (RSU)-assisted bidirectional view sharing that integrates joint source-channel coding (JSCC) with a feature-domain, self-information-assisted NC scheme over learned semantic representations, referred to as semantic network coding (semantic NC). In the proposed framework, two vehicles encode their camera images into compact semantic features and simultaneously transmit them to the RSU. The RSU exploits signal additivity to form a feature-domain mixture and broadcasts the mixed representation back to both vehicles. Each vehicle then uses its own transmitted semantic feature as self-information to cancel its contribution from the received mixture and reconstruct the other vehicle's view through a neural decoder. Experiments under AWGN and Rayleigh fading channels show that the proposed semantic NC scheme achieves stable reconstruction performance across different SNRs. Compared with semantic transmission without NC, the proposed semantic NC incurs about 0.3–1.5 dB PSNR loss in the KITTI high-resolution setting and about 0.9–2.2 dB PSNR loss in the CIFAR-10 low-resolution setting, while reducing the required bidirectional relay transmission phases from four time slots to two. These results demonstrate that the proposed scheme achieves a favorable reconstruction–latency trade-off and has potential for low-latency, reconstruction-oriented view sharing in vehicular networks.

Keywords: vehicular view sharing; semantic communication; network coding; joint source-channel coding; roadside unit; cooperative perception



Academic Editor: Alessandro Lo Schiavo

Received: 14 April 2026

Revised: 18 June 2026

Accepted: 30 June 2026

Published: 7 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Background and Related Work

In modern intelligent networks supporting cooperative perception and edge intelligence, efficient bidirectional visual information exchange is crucial for vehicular view sharing. In Internet of Vehicles (IoV) [1,2] scenarios, vehicles are no longer isolated sensing units; instead, they are expected to exchange visual observations with neighboring vehicles and roadside units (RSUs) [3] to extend the perception range, alleviate occlusion-induced blind spots, and support more reliable environment understanding for intelligent driving [4]. Such cooperative visual information sharing is particularly important for

safety-critical applications, where timely access to complementary views can improve situational awareness under complex traffic conditions [5]. Traditional communication systems grounded in Shannon's separation theorem [6] typically employ separate source and channel coding [7,8], which can be suboptimal for bandwidth-limited and latency-constrained scenarios, exhibiting pronounced cliff effects under degraded channel conditions [9]. This limitation becomes more evident in vehicular networks, where wireless channels may vary rapidly due to vehicle mobility, fading, interference, and unstable link quality, while visual data usually require a large number of bits for reliable transmission. To address these challenges, semantic communication has emerged as a novel paradigm [10–13], leveraging deep learning for JSCC [14]. Different from conventional bit-oriented transmission, semantic communication aims to preserve task-relevant or reconstruction-relevant information, thereby allowing the transmitter and receiver to jointly learn compact and channel-adaptive representations. Early works introduced CNN-based JSCC methods [15,16] that map images directly onto channel symbols, showing that end-to-end learned transmission can provide graceful performance degradation under noisy channels. Given the advantages of Swin Transformer [17] in computational efficiency and hierarchical structure modeling, ref. [18] employed it to replace CNN backbones, achieving superior performance across multiple datasets. These studies suggest that learned semantic representations are promising for image transmission, especially when communication resources are limited and the receiver is more concerned with perceptual reconstruction quality than exact bit-level recovery.

Although semantic communication has shown promising performance for image transmission, most existing works focus on point-to-point links and do not exploit the potential of relay-assisted vehicular networks. Network coding (NC), which allows intermediate nodes to combine multiple information flows before forwarding, has been widely adopted in conventional relay networks to improve spectral efficiency and reduce latency. Integrating semantic communication with NC in RSU-assisted vehicular view sharing, however, remains largely unexplored.

1.2. Motivation and Contribution

Motivated by the complementary strengths of semantic communication and NC, this paper investigates low-latency vehicular view sharing via an RSU by unifying learned JSCC with feature-domain NC. Instead of enforcing bit-perfect decoding of network-coded packets, we operate on learned semantic features and exploit side information at each receiving vehicle for interference cancellation, aiming to achieve robust reconstruction quality under diverse channel conditions.

The main contributions are summarized as follows:

- We propose an end-to-end framework for RSU-assisted vehicular view sharing, where semantic feature extraction, wireless transmission, feature-domain network coding at the RSU, and receiver-side reconstruction are jointly learned.
- We develop a feature-domain NC mechanism that performs network coding directly on learned semantic features, rather than on detected bits or explicitly decoded XOR packets. The RSU exploits wireless superposition to form a broadcast mixture, while each receiving vehicle uses its own transmitted semantic feature as self-information to perform semantic-level interference cancellation and disentangle the other view for reconstruction.
- We evaluate the proposed method under AWGN and Rayleigh fading channels using PSNR as the reconstruction-quality metric. In addition, we introduce a latency-oriented comparison based on the required number of time slots, showing that the proposed semantic NC scheme reduces the bidirectional exchange from four time slots to two time slots compared with the non-NC relaying scheme.

2. Related Work

2.1. Basic Knowledge of Semantic Communication

Semantic communication has emerged as a promising paradigm for overcoming the limitations of conventional communication systems in bandwidth-constrained and dynamically varying wireless environments. Unlike traditional communication schemes that pursue accurate bit-level recovery, semantic communication focuses on preserving task-relevant information through end-to-end optimization. Early studies mainly adopted deep joint source-channel coding (JSCC) frameworks, where source compression and channel transmission are jointly learned by neural networks. Representative works such as Deep-JSCC [15] and its subsequent variants [16] demonstrated that end-to-end learned transmission can achieve graceful performance degradation under noisy channel conditions and outperform conventional separate source and channel coding schemes in low-SNR scenarios. To further improve semantic representation capability and transmission efficiency, recent research has introduced Transformer-based architectures into semantic communication systems. Benefiting from their hierarchical feature extraction and long-range dependency modeling capabilities, Transformer-based frameworks can better capture high-level semantic information from images and adapt more effectively to varying channel conditions. For example, WITT [18] employs the Swin Transformer architecture [17] to enhance semantic feature learning and has demonstrated superior reconstruction quality compared with conventional CNN-based approaches across multiple image transmission benchmarks.

More recently, the rapid development of generative artificial intelligence has stimulated increasing interest in generative semantic communication. By incorporating powerful generative priors, such as diffusion models [19] and large-scale generative networks, these approaches aim to reconstruct perceptually faithful images from highly compressed semantic representations. Representative studies have explored stable-diffusion-assisted semantic communication and multimodal semantic generation frameworks, showing that high-quality image reconstruction can be achieved even under extremely limited communication resources. These advances suggest that semantic communication is gradually evolving from reconstruction-oriented transmission toward semantic understanding and content generation.

In parallel, semantic communication has attracted growing attention in vehicular networks, where efficient exchange of sensory information is essential for cooperative perception and intelligent driving applications. Existing studies have investigated semantic-aware image transmission, edge intelligence, and cooperative perception mechanisms to reduce communication overhead while maintaining perception performance. By transmitting semantic features rather than raw visual data, these methods can significantly improve spectrum utilization and reduce latency in Internet of Vehicles (IoV) scenarios. However, most existing semantic communication frameworks are developed for point-to-point transmission systems and do not explicitly consider relay-assisted bidirectional view sharing. Moreover, integrating semantic communication with network coding mechanisms to improve relay efficiency and exploit side information available at receiving vehicles remains largely unexplored.

2.2. Network Coding for Relay-Assisted Communications

Network coding (NC) [20] provides an important theoretical basis for improving the efficiency of relay-assisted information exchange. Unlike conventional store-and-forward routing, where intermediate nodes only forward received packets independently, NC allows relay nodes to combine multiple incoming information flows before transmission. Such combinations can be implemented through bit-wise XOR operations or linear combi-

nations [21] over a finite field. For a network represented by a directed graph $G = (V, E)$, a general linear NC operation at an intermediate node v can be written as:

$$Z_e(v, i) = \sum_{e' \in \text{In}(v)} \alpha_{v,e,e'} Y_{e'}(v, i) + \beta_{v,e} X(v, i), \quad (1)$$

where $Z_e(v, i)$ is the symbol transmitted on outgoing edge e , $Y_{e'}(v, i)$ is the symbol received from incoming edge e' , $X(v, i)$ is the local source symbol if available, and $\alpha_{v,e,e'}$ and $\beta_{v,e}$ denote coding coefficients. Classical NC theory [22] shows that, by allowing intermediate nodes to mix information flows, multicast throughput can approach the max-flow min-cut bound, and linear network coding is sufficient for achieving multicast capacity in many network settings.

In wireless two-way relaying, NC is particularly attractive because the superposition property of wireless signals can be exploited to reduce the number of required transmission phases. Instead of forwarding two source streams separately, the relay can broadcast a coded representation that contains information from both sources, while each receiver uses its own side information to recover the desired message. Nevertheless, conventional bit-level or symbol-level NC usually depends on accurate decoding, synchronization, and channel estimation, which may be difficult in vehicular environments with mobility, fading, and interference. Therefore, in this work, NC is extended from the bit domain to the semantic feature domain. The RSU forms a feature-domain mixture through wireless superposition, and each receiving vehicle uses locally available side information to perform semantic-level cancellation and view reconstruction. This design preserves the spectral efficiency advantage of NC while relaxing the requirement of strict bit-wise network-code recovery.

On the other hand, in RSU-assisted vehicular view sharing, communication efficiency is strongly constrained by the relay transmission protocol. In conventional store-and-forward relaying, the RSU first receives the information from one vehicle and then forwards it to the other vehicle in separate transmission phases. For bidirectional view sharing, this process usually requires multiple time slots, which increases transmission latency and reduces spectrum efficiency. Wireless-superposition-enabled NC [22,23] provides a promising solution by allowing simultaneously transmitted signals to be naturally superimposed over the wireless channel and combined at the RSU [24]. In typical two-way relay networks [24], this mechanism can reduce the required number of time slots from four to two, thereby improving relay efficiency and resource utilization [25,26]. Despite these advantages, conventional bit-level NC usually depends on accurate symbol- or bit-level recovery at the relay or the end nodes. As a result, its performance can be sensitive to practical wireless impairments, such as relative phase offsets, imperfect synchronization, and channel estimation errors among simultaneously transmitted signals. These impairments may distort the superimposed signals and cause significant degradation in decoding accuracy [27]. To alleviate these limitations, recent studies have begun to explore semantic communication-empowered NC, in which the coding and recovery processes are performed over learned semantic representations rather than raw bit streams [27,28]. By focusing on task-relevant semantic information, these methods can better preserve reconstruction- or task-relevant content during relay transmission. In addition, semantic communication over multi-hop relay networks has been investigated to mitigate semantic attenuation during long-distance transmissions [29]. These studies indicate that semantic-aware relay design can better preserve task-relevant information across intermediate nodes. However, most existing frameworks are developed for general relay or point-to-point scenarios and may not fully exploit the unique characteristics of vehicular view sharing. In vehicular environments, receiving vehicles naturally know their own transmitted semantic features, which can be used as self-information for canceling their own contribution from the mixed signal. Such

side information can provide useful semantic context for separating desired features from interfering components. Nevertheless, how to effectively incorporate this side information into network-coded semantic transmission remains underexplored. In contrast to existing methods, this paper proposes a feature-domain NC framework tailored for RSU-assisted vehicular view sharing. Instead of recovering and recombining bit streams, the proposed framework directly combines semantic features at the RSU. Each receiving vehicle then exploits locally available side information to perform semantic-level interference cancellation and recover the desired view-related information. In this way, the proposed design not only inherits the time-slot efficiency of NC-assisted relaying but also improves the robustness and adaptability of semantic feature recovery in vehicular communication scenarios.

3. System Model

Time-slot scheduling and the role of network coding. We consider an RSU-assisted bidirectional vehicular view sharing scenario for cooperative perception, where two participating vehicles, denoted by V_1 and V_2 , exchange visual information through an RSU. It should be noted that V_1 and V_2 act as both transmitters and receivers in the proposed exchange. During the uplink phase, the two vehicles transmit their own semantic features to the RSU; during the downlink phase, they receive the RSU broadcast and use their locally available self-information to reconstruct the other vehicle's view. Specifically, vehicle V_k , $k \in \{1, 2\}$, captures an image denoted by $x_k \in [0, 1]^{3 \times H \times W}$ where H and W represent the image height and width, respectively. The two images x_1 and x_2 contain complementary visual information, such as objects, road structures, or occluded regions observed from different viewpoints.

Different from conventional relay-assisted transmission that forwards the received information separately, the RSU in our considered system receives a superimposed feature-domain signal from V_1 and V_2 , and then broadcasts the mixed representation back to both vehicles. Meanwhile, each vehicle naturally possesses its own transmitted semantic feature as locally available self-information. Since this self-information is generated by the vehicle's own encoder, no additional wireless side-information link is required. After receiving the RSU broadcast, vehicle V_k uses its self-information to suppress its own contribution from the mixed representation and reconstruct the other vehicle's view.

In practical deployments, the RSU functionality can be realized by two infrastructure nodes, where one node is responsible for receiving uplink transmissions and the other is responsible for broadcasting downlink signals. These two nodes can be connected through a wired or high-capacity wireless backhaul link. For clarity of system modeling and without loss of the main transmission logic, we model them as a single half-duplex RSU entity. Following the learned joint source-channel coding (JSCC) paradigm, each source vehicle maps its input image directly into channel symbols through an end-to-end trainable encoder, instead of performing explicit source coding, channel coding, and modulation as separate modules. This design allows the transmitted representation to preserve task-relevant visual semantics while adapting to wireless channel conditions.

3.1. Feature-Domain Transmission and Channel Model

Let $f_\theta(\cdot)$ denote the semantic encoder adopted at the source vehicles. For source vehicle $k \in \{1, 2\}$, the input image x_k is first mapped into a compact real-valued semantic representation:

$$u_k = f_{\theta_k}(x_k) \in \mathbb{R}^{2m}, \quad (2)$$

where the dimension $2m$ corresponds to the real-valued representation of m complex channel symbols. This feature-domain representation is learned to preserve reconstruction-relevant visual information while adapting to the wireless transmission process. Following

the learned JSCC paradigm, u_k is divided into two equal parts and interpreted as the real and imaginary components of complex symbols:

$$s_k = u_k^{(r)} + ju_k^{(i)}, u_k = [u_k^{(r)}, u_k^{(i)}], s_k \in \mathbb{C}^m \quad (3)$$

Before transmission, s_k is normalized to satisfy a unit average power constraint, so that the channel quality can be consistently characterized by the signal-to-noise ratio (SNR). For consistency with Figure 1, the semantic feature generated by the encoder is denoted by u_k , while the complex-valued transmitted semantic symbol after feature-to-symbol mapping and power normalization is denoted by s_k . The locally available self-information at vehicle V_k is denoted by z_k . In the simplified architecture shown in Figure 1, z_k represents the known self-contribution constructed from the vehicle's own transmitted semantic symbol. In the detailed channel model, this selfcontribution corresponds to $h_{Rk}x_R$, where h_{Rk} is the uplink channel coefficient from vehicle V_k to the RSU. Therefore, z_k is not obtained from an additional wireless link but is locally constructed from the vehicle's own transmitted feature and the corresponding channel coefficient. The RSU broadcast mixture shown in Figure 1 is denoted by r . In the simplified architectural description, r represents the mixed feature-domain representation received by the vehicles after RSU forwarding and channel equalization. In the detailed channel model, this quantity corresponds to the effective received mixture $\tilde{y}_k$ at vehicle V_k .

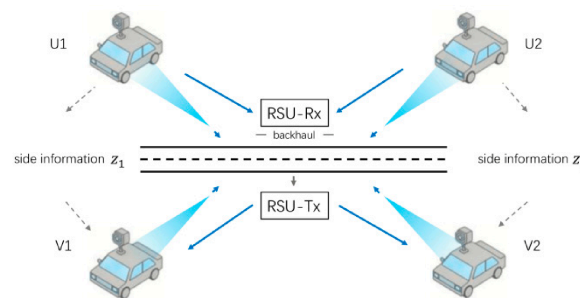


Figure 1. Overall architecture of the proposed semantic NC framework for RSU-assisted bidirectional vehicular view sharing. The two participating vehicles, V_1 and V_2 , act as both transmitters and receivers. Vehicle V_1 encodes its image x_1 into semantic feature u_1 , while vehicle V_2 encodes its image x_2 into semantic feature u_2 , using Swin Transformer encoders. The two vehicles simultaneously transmit the corresponding semantic representations to the RSU. The RSU exploits wireless superposition to form a feature-domain mixture r and broadcasts it back to both vehicles. Each vehicle uses its locally stored self-information, i.e., z_1 at V_1 or z_2 at V_2 , to suppress its own contribution from the mixed representation and reconstruct the other vehicle's view through Swin Transformer decoders.

We denote the wireless channel operator by $\mathcal{H}(\cdot; \gamma)$, where γ is the SNR. In the AWGN case, the transmitted complex symbols are corrupted by additive complex Gaussian noise whose variance is determined by γ . In the Rayleigh fading case, the channel additionally introduces i.i.d. multiplicative fading coefficients before adding Gaussian noise. These two channel models are used to evaluate the robustness of the learned feature-domain transmission under both noise-only and fading-plus-noise conditions.

3.2. RSU-Assisted Exchange with Side Information

The RSU leverages wireless superposition to form a feature-domain network-coded representation. Different from conventional non-NC relaying, where the two directions are forwarded separately, the proposed scheme allows the two vehicles to transmit simultaneously to the RSU. The RSU does not decode the individual semantic streams. Instead, it normalizes and broadcasts the superimposed feature-domain signal to both vehicles.

During the multiple-access phase, the two vehicles simultaneously transmit their normalized semantic symbols s_1 and s_2 to the RSU. The received signal at the RSU is modeled as

$$\mathbf{y}_R = h_{1R}\mathbf{s}_1 + h_{2R}\mathbf{s}_2 + \mathbf{n}_R \quad (4)$$

where h_{kR} denotes the uplink channel coefficient from vehicle V_k to the RSU, and $\mathbf{n}_R$ is complex Gaussian noise.

The RSU then normalizes and forwards the superimposed feature-domain signal:

$$\mathbf{x}_R = \beta \mathbf{y}_R, \quad (5)$$

where β is a power-normalization factor satisfying the RSU transmit-power constraint. During the broadcast phase, vehicle V_k receives

$$\mathbf{y}_k = h_{Rk}\mathbf{x}_R + \mathbf{n}_k, \quad (6)$$

where h_{Rk} denotes the downlink channel coefficient from the RSU to vehicle V_k , and $\mathbf{n}_k$ is the receiver noise.

After channel equalization, the effective received mixture at vehicle V_k can be written as

$$\tilde{\mathbf{y}}_k = h_{1R}\mathbf{s}_1 + h_{2R}\mathbf{s}_2 + \tilde{\mathbf{n}}_k. \quad (7)$$

Since vehicle V_k knows its own transmitted semantic feature $\mathbf{s}_k$, it uses this self-information as local side information and removes its own contribution from the mixed representation:

$$\mathbf{e}_k = \tilde{\mathbf{y}}_k - h_{kR}\mathbf{s}_k = h_{\bar{k}R}\mathbf{s}_{\bar{k}} + \tilde{\mathbf{n}}_k, \quad (8)$$

where $\bar{k}$ denotes the other vehicle index, i.e., $1 = 2$ and $2 = 1$. The residual feature $\mathbf{e}_k$ is then fed into the semantic decoder to reconstruct the other vehicle's view:

$$\hat{\mathbf{x}}_{k \leftarrow \bar{k}} = g_{\phi_k}(\mathbf{e}_k; \gamma). \quad (9)$$

Therefore, the proposed exchange performs side-information-assisted cancellation in the learned semantic feature domain rather than relying on exact bit-level XOR recovery. This formulation also clarifies that the side information does not require an additional wireless link, because each vehicle naturally knows its own transmitted semantic feature.

3.3. Bandwidth Ratio and Optimization Objective

We adopt the channel bandwidth ratio (CBR) to characterize the transmission resource consumed by the semantic representation. Since the transmitted semantic symbol vector $\mathbf{s}_k$ contains m complex channel symbols for the source image $\mathbf{x}_k$, the CBR is defined as

$$\text{CBR} = \frac{m}{\text{numel}(\mathbf{x}_k)}, \quad (10)$$

where $\text{numel}(\mathbf{x}_k)$ denotes the number of source image elements. A smaller CBR indicates that fewer channel symbols are used to transmit each image, corresponding to a more bandwidth-efficient semantic representation. The training objective minimizes the average distortion over the four reconstructions produced by the view sharing procedure:

$$\mathcal{L} = \frac{1}{4}(d(\mathbf{x}_1, \hat{\mathbf{x}}_{1 \leftarrow 1}) + d(\mathbf{x}_2, \hat{\mathbf{x}}_{1 \leftarrow 2}) + d(\mathbf{x}_2, \hat{\mathbf{x}}_{2 \leftarrow 2}) + d(\mathbf{x}_1, \hat{\mathbf{x}}_{2 \leftarrow 1})) \quad (11)$$

where $d(\cdot, \cdot)$ is set to the mean squared error (MSE) in our experiments. The terms $\hat{\mathbf{x}}_{1 \leftarrow 1}$ and $\hat{\mathbf{x}}_{2 \leftarrow 2}$ are introduced as auxiliary self-reconstruction constraints during training while

the main bidirectional view sharing task focuses on the cross-view reconstructions $\hat{\mathbf{x}}_{1 \leftarrow 2}$ and $\hat{\mathbf{x}}_{2 \leftarrow 1}$.

4. Proposed Semantic Network Coding Framework

4.1. Overall Bidirectional Semantic Exchange

The proposed framework integrates learned JSCC with feature-domain NC to enable low-latency vehicular view sharing assisted by an RSU. The key idea is to perform network coding directly on continuous-valued semantic features, rather than on detected bits, and to exploit each vehicle's own transmitted semantic feature as self-information for interference cancellation.

Specifically, each vehicle maps its image to a compact semantic feature sequence and transmits the corresponding representation over the wireless channel under a unit power constraint. Due to the additive nature of wireless superposition, the RSU receives a superimposed signal that contains contributions from both sources. Instead of attempting to decode individual semantic streams or generate a bit-wise XOR packet, the RSU forwards a feature-domain mixture formed from the superposition. After the RSU broadcast, vehicle V_k has two pieces of information: (i) its own transmitted semantic representation z_k , which is naturally available as self-information, and (ii) the RSU broadcast mixture r . The receiver then constructs a residual representation by suppressing its own contribution from the RSU broadcast. In the simplified notation of Figure 1, this operation can be written as $\mathbf{e}_k = r - z_k$. In the detailed channel model, it corresponds to $\mathbf{e}_k = \tilde{\mathbf{y}}_k - h_{kR}\mathbf{s}_k$, where the channel coefficient is explicitly considered. This residual representation is then fed into a learned semantic decoder to reconstruct the other vehicle's view.

From an information-processing viewpoint, the RSU broadcast and the local side information constitute a form of side-information assisted separation: the side information anchors one source's semantics, while the residual emphasizes the other source's semantics. The overall system is trained end-to-end to optimize reconstruction quality under noise and fading, so that the encoder learns representations that are amenable to superposition and cancellation and the decoder learns to robustly disentangle the mixture without requiring explicit symbol-level or bit-level network-coding mapping rules.

4.2. Encoder, RSU Operation, and Decoder Architecture

Semantic encoder/decoder. We adopt a hierarchical encoder–decoder architecture based on Swin Transformer blocks. At the transmitter side, the encoder first converts the input image into patch tokens through patch embedding. The tokens are then processed by multiple Swin Transformer stages, where window-based self-attention captures local contextual dependencies and shifted-window attention enables cross-window information interaction. Patch merging is used between stages to gradually reduce spatial resolution and increase feature dimensions, thereby producing hierarchical multi-scale semantic representations.

A final linear projection layer maps the extracted features into a bottleneck feature sequence for wireless transmission. The output is organized as a real-valued tensor, whose first and second halves are interpreted as the real and imaginary parts of complex channel symbols, as defined in System Model. Therefore, the encoder directly learns channel-input representations without explicit bit-level source coding, channel coding, or modulation. The decoder mirrors the encoder by using Swin Transformer blocks and patch reverse merging to progressively recover spatial resolution and reconstruct the image. In the proposed view sharing procedure, each receiving vehicle uses its decoder twice: once for decoding the side-information observation and once for decoding the residual after cancellation.

Channel-aware modulation (implementation detail). To incorporate channel state information into the learned representations, we implement a lightweight modulation network that takes the scalar SNR as input and outputs token-wise scaling coefficients. These coefficients are multiplicatively applied to intermediate features in both the encoder and decoder. This conditional modulation enables the model to adjust feature representations according to channel quality. Under high-SNR conditions, more detailed visual information can be preserved, whereas under low-SNR conditions, the model can emphasize more robust semantic components to improve reconstruction stability.

RSU-side feature-domain network coding and side-information cancellation. Consistent with the forward pass in our implementation, the RSU-side operation is performed on noisy feature observations. Let u_1 and u_2 denote the transmitted feature tensors from the two source vehicles. The RSU forms a superposition of the two uplink-corrupted features and forwards the resulting mixture through the downlink channel, yielding the broadcast mixture r . Unlike conventional bit-level NC, the RSU does not need to recover individual bit streams or explicitly generate XOR-coded packets.

Each receiving vehicle also has access to its own transmitted semantic symbol z_k from a direct link, which serves as self-information and does not require an additional wireless link. The receiver uses this side information as a reference and forms the residual $\mathbf{e}_k = \tilde{\mathbf{y}}_k - h_{kR}\mathbf{s}_k$ to suppress the known feature contribution. The residual is then decoded to reconstruct the other source image. Since cancellation and reconstruction are performed in the learned feature domain, the proposed method can tolerate imperfect channel observations and does not rely on exact symbol-level recovery. The entire exchange is optimized end-to-end with the four-way reconstruction objective defined in System Model, allowing the encoder and decoder to jointly learn representations suitable for superposition, cancellation, and robust image reconstruction.

The effectiveness of the feature-domain cancellation operation comes from both the additive structure of the wireless superposition and the end-to-end learned semantic representation. Since the RSU broadcast contains a superimposed mixture of the semantic features transmitted by the two vehicles, each vehicle can use its own transmitted feature as self-information to suppress the known component from the received mixture. Although the learned semantic features are continuous-valued representations rather than exact digital codewords, the encoder and decoder are jointly optimized under the reconstruction loss, which encourages the transmitted features to be compatible with superposition, side-information-assisted cancellation, and subsequent decoding. In this sense, the residual feature after cancellation is not required to be a perfectly separated signal in the conventional bit-level sense. Instead, it only needs to preserve sufficient reconstruction-relevant information about the other vehicle's view. The neural decoder can then learn to compensate for residual interference and channel noise, enabling robust view reconstruction from the imperfectly canceled feature-domain representation.

5. Experimental Results

5.1. Experimental Setup

Implementation and hardware. All experiments were implemented in PyTorch 3.10 and trained on a single NVIDIA GPU. Unless otherwise specified, we use the Adam optimizer with a learning rate of 1×10^{-4} [30]. The random seed is fixed for reproducibility.

Datasets and preprocessing. We evaluate the proposed JSCC-based semantic NC framework on two representative image transmission settings: a low-resolution benchmark and a high-resolution benchmark. For low-resolution experiments, we use CIFAR-10 [16] for both training and testing with images resized/cropped to 32×32 . For high-resolution experiments, we use DIV2K [31] for training with images cropped to 256×256 , and report

results on KITTI for testing. During training and testing, two images are randomly sampled to form an exchange Pair $r(x_1, x_2)$.

Channel models, SNR and scope. We consider additive white Gaussian noise (AWGN) and Rayleigh fading channels [32,33], which are standard benchmark models for wireless image transmission and relay communication systems [25]. AWGN is used to isolate the effect of additive noise, while Rayleigh fading is used to evaluate the robustness of the proposed framework under fading-plus-noise conditions. The SNR is specified in dB. During training, the SNR is sampled from a predefined set, e.g., {1, 4, 7, 10, 13} dB, to expose the model to a range of channel conditions; during evaluation, results are reported at each tested SNR.

It should be emphasized that the purpose of this work is not to explicitly compensate for all practical propagation impairments in highly dynamic IoV environments, such as Doppler spread, time-varying multipath, or complex V2V/V2I link variations. Instead, this paper focuses on reconstruction-oriented semantic view sharing and investigates whether the proposed JSCC-based semantic NC framework can achieve effective image reconstruction and low-latency bidirectional exchange under representative wireless distortions. More realistic vehicular channel models, including Doppler-sensitive time-varying fading, imperfect CSI, synchronization errors, and complex V2V/V2I multipath propagation, will be considered in future work.

Model configuration and rate control. The bottleneck dimension C controls the transmission rate and determines the effective CBR computed from the encoder output, consistent with the definition in System Model. For the CIFAR-10 low-resolution setting and the DIV2K-trained, KITTI-tested high-resolution setting, we use the corresponding encoder/decoder stage configurations and window sizes as specified in our implementation, following the Swin Transformer architecture [17].

Baselines. We compare the proposed semantic NC with a two-hop point-to-point JSCC relaying baseline without network coding. The baseline uses the same encoder/decoder architecture as the proposed method. For source vehicle k , the baseline forward pass is $z_k^{\text{baseline}} = \mathcal{H}(\mathcal{H}(u_k; \gamma); \gamma)$, where the RSU separately forwards each source vehicle's features over an uplink and a downlink channel, so the end-to-end signal undergoes two channel realizations (equivalently, two noise injections) before decoding at the receiving vehicle. The baseline is trained with the same MSE objective over the two reconstructions (one per source vehicle).

Training objective and metrics. The model is trained by minimizing the MSE averaged over the four reconstructions produced by the view sharing exchange. For reporting transmission quality, we use peak signal-to-noise ratio (PSNR) as the primary reconstruction-quality metric. PSNR is widely adopted in image transmission and JSCC-based semantic communication because it directly measures pixel-level distortion between the reconstructed image and the original image.

In addition to PSNR, we report the number of required time slots for completing bidirectional view sharing. This latency-oriented metric directly reflects the relay efficiency of the proposed semantic NC scheme. Therefore, the proposed method is evaluated from two complementary perspectives: reconstruction fidelity, measured by PSNR, and bidirectional exchange efficiency, measured by the required number of time slots.

5.2. Results Analysis

Figure 2 compares the reconstruction quality of the proposed semantic NC scheme under AWGN and Rayleigh fading channels. The high-resolution results are reported on the KITTI test set with $\text{CBR} = 1/16$, while the low-resolution results are reported on CIFAR-10 with $\text{CBR} = 1/2$. Overall, the PSNR increases with SNR in most cases, indicating

that improved channel quality benefits both semantic transmission without NC and the proposed semantic NC scheme. However, the relative performance differs across datasets, channel models, and transmission schemes.

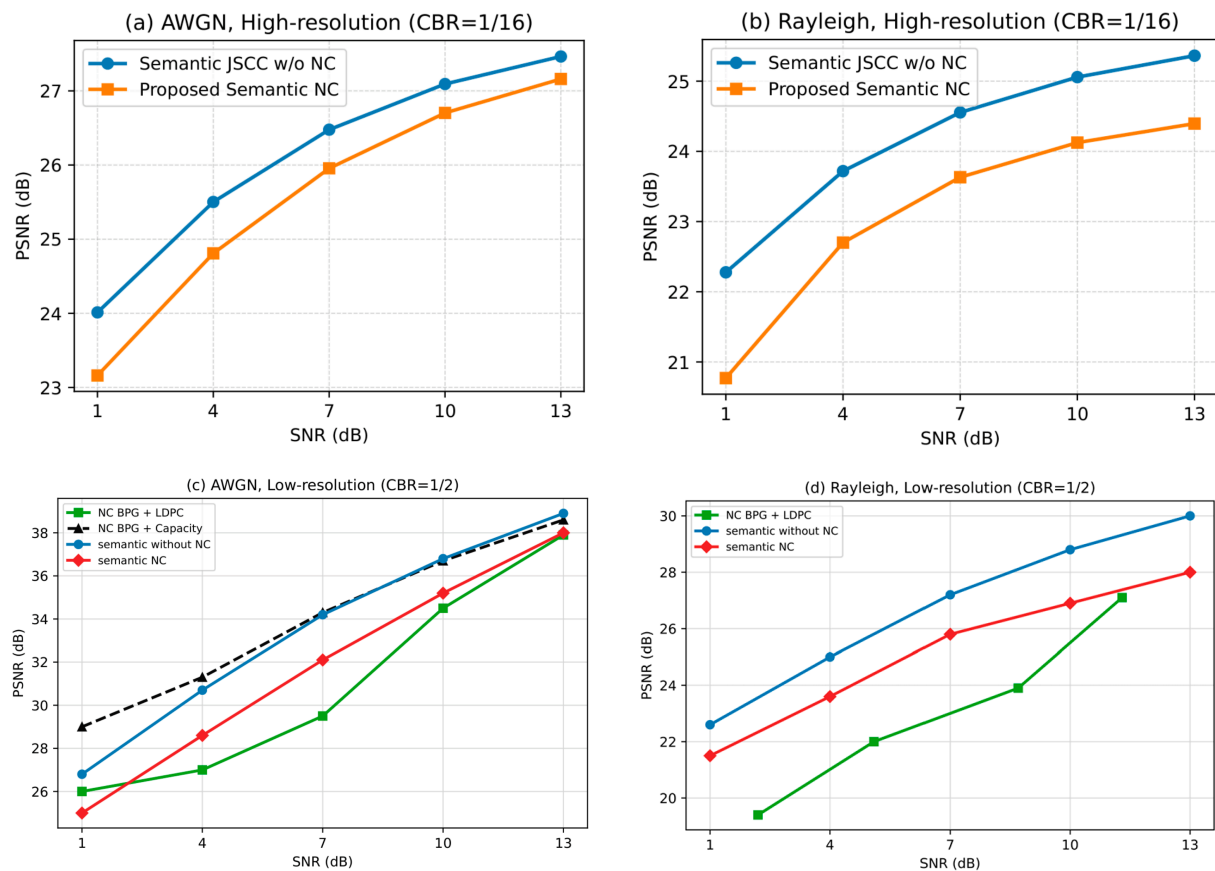


Figure 2. PSNR comparison under AWGN and Rayleigh fading channels at different SNRs. The high-resolution results are reported on the KITTI test set with CBR = 1/16, while the low-resolution results are reported on CIFAR-10 with CBR = 1/2. The high-resolution subfigures compare semantic transmission without NC and the proposed semantic NC scheme, and the low-resolution subfigures further include NC-assisted BPG-based digital baselines.

In the KITTI high-resolution AWGN setting, semantic transmission without NC achieves the highest PSNR because it avoids the residual distortion introduced by feature-domain superposition and side-information-assisted cancellation. The proposed semantic NC scheme shows a moderate PSNR loss compared with semantic transmission without NC, with an approximate gap of 0.3–0.9 dB over the tested SNR range. Although this gap is caused by the additional feature mixing and cancellation process, the proposed semantic NC still maintains stable reconstruction quality as SNR increases. This result indicates that the proposed scheme can preserve reconstruction-relevant information while reducing the bidirectional relay transmission procedure from four time slots to two.

In the KITTI high-resolution Rayleigh fading setting, the overall PSNR is lower than that under AWGN because multiplicative fading introduces stronger channel distortion. Semantic transmission without NC still provides the best reconstruction quality, while the proposed semantic NC remains close to it with an approximate PSNR gap of 0.9–1.5 dB. Compared with the AWGN case, the performance gap becomes slightly larger under Rayleigh fading, which is reasonable because the mixed feature domain representation is affected by both fading and noise before side-information-assisted cancellation. Neverthe-

less, the proposed semantic NC still shows a smooth PSNR improvement with increasing SNR, demonstrating its robustness under fading-plus-noise distortions.

For the CIFAR-10 low-resolution AWGN setting, all schemes achieve higher PSNR values as SNR increases. Semantic transmission without NC serves as a strong reference because it does not involve network-coded feature mixing. The proposed semantic NC incurs a moderate PSNR loss compared with semantic transmission without NC, but it remains competitive over the tested SNR range. Compared with the NC-assisted BPG + LDPC baseline, the proposed semantic NC achieves better reconstruction quality at most SNR points, especially under low and medium SNR conditions. This suggests that the learned semantic representation is more tolerant of channel noise than separated digital transmission when the channel quality is limited. The BPG + Capacity baseline provides an idealized digital reference and remains competitive when the coding rate is matched to the channel capacity.

Under the CIFAR-10 low-resolution Rayleigh fading setting, the performance degradation becomes more pronounced for all schemes. Semantic transmission without NC achieves the highest PSNR, while the proposed semantic NC maintains a consistent gap due to the additional distortion introduced by wireless superposition and residual interference after cancellation. Nevertheless, semantic NC still outperforms or remains competitive with the NC-assisted BPG + LDPC baseline over most tested SNRs. This result suggests that the proposed semantic NC scheme can preserve useful reconstruction-relevant information even when the received feature-domain mixture is affected by fading and noise.

Figure 3 further evaluates the impact of transmission rate by varying the CBR under AWGN at 10 dB on the KITTI high-resolution test set. For semantic transmission without NC, increasing the CBR generally improves PSNR because a larger bottleneck allows more visual and semantic information to be preserved. For the proposed semantic NC scheme, the PSNR remains competitive across different CBR values, but the trend is not strictly monotonic. This indicates that a larger bottleneck does not always directly translate into better reconstruction performance in the semantic NC setting, because the receiver must recover the desired view from a superimposed feature-domain mixture using side-information-assisted cancellation. Therefore, the performance is jointly affected by the representation capacity, feature superposition, residual interference, and decoder robustness. Overall, the CBR results show that semantic NC can maintain competitive reconstruction quality while enabling a more efficient bidirectional exchange protocol.

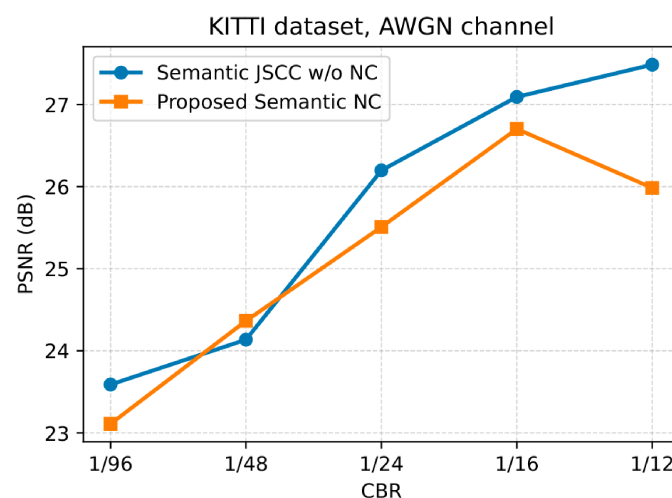


Figure 3. PSNR comparison between semantic transmission without NC and the proposed semantic NC scheme under different CBR values on the KITTI high-resolution test set over the AWGN channel at SNR = 10 dB.

Figure 4 presents representative visual reconstruction results on the KITTI high-resolution test set under the AWGN channel at 7 dB. Compared with semantic transmission without NC, the proposed semantic NC reconstruction is slightly affected by feature-domain mixing and imperfect cancellation, but it still preserves the main road layout, vehicle contours, lane structures, and scene-level semantic information. The visual results are consistent with the PSNR results: semantic NC may introduce a moderate reconstruction-quality loss, but the reconstructed images remain visually coherent and suitable for reconstruction-oriented view sharing.



Figure 4. Representative visual reconstruction results on the KITTI high-resolution test set under the AWGN channel at 7 dB. The first column shows the original KITTI test images, while the remaining columns compare the reconstructed images obtained by semantic transmission without NC and the proposed semantic NC scheme, respectively.

Overall, the proposed semantic NC scheme should be interpreted as a reconstruction–latency trade-off method rather than a pure PSNR-maximization scheme. Compared with semantic transmission without NC, semantic NC introduces a moderate PSNR degradation because the RSU forwards a superimposed feature-domain mixture and each vehicle relies on side-information-assisted cancellation to recover the desired view. However, this degradation is exchanged for a significant reduction in bidirectional relay transmission phases, from four time slots to two. Therefore, the proposed scheme provides a favorable balance between reconstruction quality and time-slot efficiency, which is particularly suitable for latency-sensitive vehicular view sharing.

5.3. Latency-Oriented Comparison

In addition to PSNR-based reconstruction evaluation, we further compare the latency efficiency of the proposed scheme and the conventional non-NC relaying scheme. For bidirectional vehicular view sharing, both schemes aim to exchange two views between vehicles V_1 and V_2 through the RSU. Here, a view refers to the visual image captured from one vehicle’s perspective, e.g., x_1 from V_1 and x_2 from V_2 . The key difference lies in whether the two directions are forwarded independently or combined through feature-domain network coding.

Without NC, the two directions are transmitted separately. The exchange of x_1 from V_1 to V_2 requires two time slots, namely $V_1 \rightarrow R$ and $R \rightarrow V_2$. Similarly, the exchange of x_2 from V_2 to V_1 requires another two time slots, namely $V_2 \rightarrow R$ and $R \rightarrow V_1$. Therefore, four half-duplex time slots are required in total.

In contrast, the proposed semantic NC scheme requires only two time slots. In the first time slot, V_1 and V_2 simultaneously transmit their semantic features to the RSU. In the second time slot, the RSU broadcasts the mixed feature-domain representation to both vehicles. Each vehicle then uses its own semantic feature as side information to recover the other vehicle’s view.

As shown in Table 1, both schemes complete the exchange of two views, where each view corresponds to one vehicle-captured image. However, the proposed semantic NC scheme reduces the required number of time slots from four to two, corresponding to a 50% reduction in relay exchange latency and a twofold improvement in bidirectional exchange

efficiency. Therefore, the proposed method should be interpreted as a reconstruction–latency trade-off scheme rather than a pure PSNR-maximization method.

Table 1. Latency comparison of non-NC relaying and semantic NC.

Scheme T	Transmission Procedure	Time Slots	Views/Slot
Without NC	NC $V_1 \rightarrow R, R \rightarrow V_2, V_2 \rightarrow R, R \rightarrow V_1$	4	2/4
Proposed semantic NC	$V_1, V_2 \rightarrow R, R \rightarrow V_1, V_2$	2	2/2

6. Conclusions

This paper investigated a semantic communication framework for low-latency vehicular view sharing assisted by an RSU. By integrating learned JSCC with feature-domain NC, the RSU broadcasts a superimposed feature mixture, and each vehicle exploits its own transmitted semantic feature as self-information to cancel interference and reconstruct the other vehicle’s view. The proposed system is trained end-to-end under noisy channels with an MSE objective, and its rate is characterized by the CBR defined from the encoder output, consistent with the implementation. Comprehensive experiments on CIFAR-10 for low-resolution image transmission and on the DIV2K-trained, KITTI-tested high-resolution setting validate the robust image reconstruction performance of the proposed semantic NC scheme under AWGN and Rayleigh fading channels across different CBR and SNR conditions. Compared with semantic transmission without NC, the proposed semantic NC scheme introduces only a modest PSNR degradation while reducing the required bidirectional relay transmission phases from four time slots to two. In the low-resolution experiments, compared with separated BPG-based digital transmission baselines, including BPG with LDPC coding and the idealized BPG with capacity-based transmission, the proposed semantic NC scheme shows competitive reconstruction performance and better robustness in several low-SNR and fading scenarios. These results demonstrate that the proposed scheme achieves a favorable reconstruction–latency trade-off and has potential for low-latency, reconstruction-oriented view sharing in vehicular networks.

Author Contributions: Conceptualization, Y.W.; Methodology, Y.W.; Software, Y.W.; Validation, Y.W.; Formal analysis, Y.W.; Investigation, Y.W.; Resources, Y.W. and C.L.; Data curation, Y.W. and J.Z.; Writing—original draft, Y.W.; Writing—review & editing, J.Z. and C.L.; Visualization, Y.W.; Project administration, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Science Foundation of China (NSFC) with grant No. 62271514, in part by the Foundation of State Key Laboratory of Public Big Data with grant No. PBD2023-01.

Institutional Review Board Statement: This study does not involve human or animal subjects, nor does it collect any personally identifiable information. All experiments are conducted on publicly available datasets. Therefore, no ethical approval is required.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusion of this article will be made available by the authors on request.

Acknowledgments: Y. Wang and J. Zhong contributed equally to this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, C.; Shan, G.; Roh, B.-H. Fmd-iov: Security and robust enhancement for federated multi-domain learning–based iov. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 14225–14236.

2. Cheng, J.; Yuan, G.; Zhou, M.; Gao, S.; Liu, C.; Duan, H.; Zeng, Q. Accessibility analysis and modeling for iov in an urban scene. *IEEE Trans. Veh. Technol.* **2020**, *69*, 4246–4256. [\[CrossRef\]](#)
3. Ma, S.; Wen, F.; Wang, Z. An efficient GPS-free vehicle localization algorithm using single roadside unit and receiver. In Proceedings of the 2018 International Conference on Networking and Network Applications (NaNA), Xi'an, China, 12–15 October 2018; pp. 310–313.
4. Arnold, E.; Dianati, M.; de Temple, R.; Fallah, S. Cooperative perception for 3d object detection in driving scenarios using infrastructure sensors. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 1852–1864.
5. Caillot, A.; Ouerghi, S.; Vasseur, P.; Boutteau, R.; Dupuis, Y. Survey on cooperative perception in an automotive context. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 14204–14223. [\[CrossRef\]](#)
6. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [\[CrossRef\]](#)
7. Richardson, T.; Kudekar, S. Design of low-density parity check codes for 5g new radio. *IEEE Commun. Mag.* **2018**, *56*, 28–34. [\[CrossRef\]](#)
8. Arikan, E. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inf. Theory* **2009**, *55*, 3051–3073.
9. Shi, Y.; Zhou, Y.; Wen, D.; Wu, Y.; Jiang, C.; Letaief, K.B. Task-oriented communications for 6g: Vision, principles, and technologies. *IEEE Wirel. Commun.* **2023**, *30*, 78–85. [\[CrossRef\]](#)
10. Weaver, W. Recent contributions to the mathematical theory of communication. *ETC Rev. General. Semant.* **1953**, *10*, 261–281.
11. Gündüz, D.; Qin, Z.; Aguerri, I.E.; Dhillon, H.S.; Yang, Z.; Yener, A.; Wong, K.K.; Chae, C.-B. Beyond transmitting bits: Context, semantics, and task-oriented communications. *IEEE J. Sel. Areas Commun.* **2022**, *41*, 5–41. [\[CrossRef\]](#)
12. Yang, W.; Du, H.; Liew, Z.Q.; Lim, W.Y.B.; Xiong, Z.; Niyato, D.; Chi, X.; Shen, X.; Miao, C. Semantic communications for future internet: Fundamentals, applications, and challenges. *IEEE Commun. Surv. Tutor.* **2022**, *25*, 213–250. [\[CrossRef\]](#)
13. Xie, H.; Qin, Z.; Li, G.Y.; Juang, B.-H. Deep learning enabled semantic communication systems. *IEEE Trans. Signal Process.* **2021**, *69*, 2663–2675. [\[CrossRef\]](#)
14. Chen, Q.; Wang, L.; Hong, S.; Xiong, Z. Performance improvement of jscc scheme through redesigning channel code. *IEEE Commun. Lett.* **2016**, *20*, 1088–1091. [\[CrossRef\]](#)
15. Bourtsoulatz, E.; Kurka, D.B.; Gündüz, D. Deep joint source-channel coding for wireless image transmission. *IEEE Trans. Cogn. Commun. Netw.* **2019**, *5*, 567–579. [\[CrossRef\]](#)
16. Kurka, D.B.; Gündüz, D. Deepjscc-f: Deep joint source channel coding of images with feedback. *IEEE J. Sel. Areas Inf. Theory* **2020**, *1*, 178–193.
17. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
18. Yang, K.; Wang, S.; Dai, J.; Tan, K.; Niu, K.; Zhang, P. Witt: A wireless image transmission transformer for semantic communications. In Proceedings of the ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; IEEE: New York, NY, USA, 2023; pp. 1–5.
19. Zhang, M.; Wu, H.; Zhu, G.; Jin, R.; Chen, X.; Gündüz, D. Semantics-guided diffusion for deep joint source-channel coding in wireless image transmission. *IEEE Trans. Wirel. Commun.* **2026**, *25*, 1547–1564.
20. Ahlswede, R.; Cai, N.; Li, S.-Y.; Yeung, R. Network information flow. *IEEE Trans. Inf. Theory* **2000**, *46*, 1204–1216. [\[CrossRef\]](#)
21. Ho, T.; Medard, M.; Koetter, R.; Karger, D.; Effros, M.; Shi, J.; Leong, B. A random linear network coding approach to multicast. *IEEE Trans. Inf. Theory* **2006**, *52*, 4413–4430. [\[CrossRef\]](#)
22. Li, S.-Y.R.; Sun, Q.T.; Shao, Z. Linear network coding: Theory and algorithms. *Proc. IEEE* **2011**, *99*, 372–387. [\[CrossRef\]](#)
23. Bassoli, R.; Marques, H.; Rodriguez, J.; Shum, K.W.; Tafazolli, R. Network coding theory: A survey. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1950–1978. [\[CrossRef\]](#)
24. Koike-Akino, T.; Popovski, P.; Tarokh, V. Optimized constellations for two-way wireless relaying with physical network coding. *IEEE J. Sel. Areas Commun.* **2009**, *27*, 773–787. [\[CrossRef\]](#)
25. Zhang, S.; Liew, S.C.; Lam, P.P. Hot topic: Physical-layer network coding. In Proceedings of the 12th Annual International Conference on Mobile Computing and Networking, Los Angeles, CA, USA, 23–29 September 2006; pp. 358–365.
26. Liew, S.C.; Lu, L.; Zhang, S. *A primer on Physical-Layer Network Coding*; Springer Nature: Berlin/Heidelberg, Germany, 2022.
27. Pan, H.; Yang, S.; Chan, T.-T.; Wang, Z.; Leung, V.C.; Li, J. SC-PNC: Semantic communication-empowered physical layer network coding. *IEEE Trans. Cogn. Commun. Netw.* **2024**, *10*, 1160–1174.
28. Yang, S.; Pan, H.; Chan, T.-T.; Wang, Z. Semantic communication-empowered physical-layer network coding. In Proceedings of the 2023 IEEE Wireless Communications and Networking Conference (WCNC), Glasgow, UK, 26–29 March 2023; IEEE: New York, NY, USA, 2023; pp. 1–6.

29. Yin, Y.; Huang, L.; Li, Q.; Tang, B.; Pandharipande, A.; Ge, X. Semantic communication on multi-hop concatenated relay networks. In *Proceedings of the 2023 IEEE/CIC International Conference on Communications in China (ICCC), Dalian, China, 10–12 August 2023*; IEEE: New York, NY, USA, 2023; pp. 1–6.
30. Cawley, G.C.; Talbot, N.L. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.
31. Ignatov, A.; Timofte, R.; Van Vu, T.; Luu, T.M.; Pham, T.X.; Van Nguyen, C.; Kim, Y.; Choi, J.-S.; Kim, M.; Huang, J.; et al. Pirm challenge on perceptual image enhancement on smartphones: Report. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018*.
32. Rappaport, T.S. Wireless communications—principles and practice, (the book end). *Microw. J.* **2002**, *45*, 128–129.
33. Sklar, B. Rayleigh fading channels in mobile digital communication systems. i. characterization. *IEEE Commun. Mag.* **2002**, *35*, 90–100.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.