

An Image Semantic Communication Method Based on SwinJSCC with Explicit MIMO and Hybrid Channel Modeling

Jiale Yu¹, Yuling Zhang^{1*}, Wenwei He², Yujuan Sun¹, Tao Yao¹

¹ School of Computer and Artificial Intelligence, Ludong University, Yantai 264025, China

² School of Integrated Circuits, Ludong University, Yantai 264025, China
zhangyuling@ldu.edu.cn

Abstract. To address the limited channel coverage of the original SwinJSCC in complex wireless scenarios, the lack of a MIMO transmission mechanism, and the heavy training burden caused by separately training independent models for different channel modes, this paper proposes a SwinJSCC-based image semantic communication method for image transmission over MIMO channels. Based on SwinJSCC, the proposed method preserves the original SNR adaptation and rate adaptation capabilities while unifying AWGN, Rayleigh fading, MIMO transmit diversity, MIMO spatial multiplexing, and large-scale path-loss extended channels into a single end-to-end training and testing framework. To avoid training a large number of separate models for different channel types and MIMO structures, a mode-oriented conditional modulation mechanism is introduced, enabling the encoder and decoder to adjust features according to the current transmission mode while sharing the same backbone parameters. In channel modeling, the proposed method does not approximate multi-antenna gain by a simple equivalent SNR. Instead, 2×1 , 4×1 , and 8×1 are modeled as transmit diversity, while 2×2 and 4×2 are modeled as spatial multiplexing, further combined with large-scale path-loss factors. This work provides a scalable unified single-model solution for image semantic transmission in complex wireless environments.

Keywords: SwinJSCC, semantic communication, MIMO, mode-adaptive model, channel adaptation.

1 Introduction

1.1 Motivation

In the field of image semantic communication, the research focus has gradually shifted from “accurately transmitting all pixels” to “preserving task-relevant semantic information under limited resources” [1]. Transformer-based image modeling methods are good at capturing global dependencies and are particularly suitable for high-resolution image scenarios. SwinJSCC (Swin Transformer [2] based Deep Joint Source-Channel Coding [3]) introduces the Swin Transformer into the end-to-end JSCC framework [2], and enables adaptation to different SNRs and target coding rates within a single model through hierarchical window attention and plug-and-play adaptive modules [4]. Therefore, extending SwinJSCC toward complex wireless channels and multi-mode multi-

antenna transmission is of both theoretical significance and practical engineering value [5], [6].

From an engineering perspective, future wireless image transmission systems will not operate solely under the ideal AWGN scenario, but must also cope with Rayleigh fading, multi-antenna transmission and reception, large-scale path loss, low-resolution remote monitoring, and high-resolution fine reconstruction. If a separate network is trained for each channel type and each antenna configuration, the number of models will increase rapidly, making training time, memory consumption, and deployment complexity unacceptable. Therefore, constructing a unified framework that preserves the original SA/RA capabilities of SwinJSCC while enabling a single model to cover multiple channel modes is the central motivation of this work.

1.2 Contributions

This paper extends SwinJSCC from a single-channel model with SNR/rate adaptation to a unified image semantic communication framework for MIMO channels under diverse channel conditions. First, a unified channel family is constructed. The transmit diversity modes, including 2×1 , 4×1 , and 8×1 , are modeled based on Alamouti-type schemes [7], while the spatial multiplexing modes, including 2×2 and 4×2 , are modeled with MMSE detection instead of approximating MIMO gains as equivalent SNR improvement [8]. Large-scale path-loss effects are also incorporated to reflect propagation distance and environmental attenuation.

Second, a mode-oriented conditional modulation mechanism is introduced. Based on the original SNR adaptation and rate adaptation of SwinJSCC, discrete channel-mode conditions are further incorporated, enabling a single model to adapt to AWGN, Rayleigh fading, different MIMO modes, and their large-scale variants while sharing the same backbone parameters. The proposed framework provides a scalable single-model solution for image semantic transmission over realistic wireless links.

2 Overall Framework

The overall framework of the proposed method is illustrated in Fig. 1.

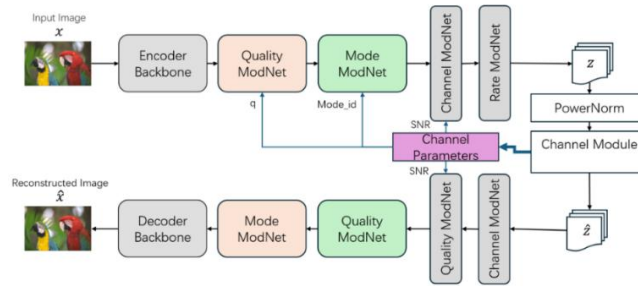


Fig. 1. Overall framework of the proposed SwinJSCC-based image semantic communication method for MIMO channels under multiple channel conditions.

The input image x is first fed into the Encoder Backbone, where hierarchical semantic features are extracted and compressed into a latent semantic representation. Based on the original SwinJSCC framework, two additional conditional modulation mechanisms are introduced to improve the adaptability of the model to complex channel conditions and multiple transmission modes. Specifically, the Channel Module provides a continuous quality vector q to describe the current channel quality state and a discrete mode identifier, mode ID, to indicate the active channel mode. Meanwhile, the original Channel ModNet and Rate ModNet are retained to receive SNR and rate information, respectively, so that the basic SNR adaptation and rate adaptation capabilities of SwinJSCC are preserved. Therefore, the encoder output is not directly transmitted through the channel, but is successively modulated by multiple conditional factors to generate the transmitted semantic feature z . After transmission over the corresponding MIMO channel, the received feature $\hat{z}$ is obtained and then symmetrically demodulated and restored at the decoder side to reconstruct the input image.

To enable the network to perceive both large-scale path-loss and fading conditions, the logarithmic path-loss term $P_{L,dB}$ and the fading indicator ξ are further incorporated into the quality vector, which is defined as $q = [\gamma, R, P_{L,dB}, \xi]$. Here, γ denotes the current SNR, R denotes the rate-related bottleneck setting, $P_{L,dB}$ denotes the logarithmic path-loss term, and ξ denotes the fading indicator. When the current mode belongs to the large-scale or large-scale MIMO family, the linear path-loss factor is calculated as $P_L = d^{-\alpha}$, and its logarithmic form is given by $P_{L,dB} = 10\log_{10}(P_L + \varepsilon)$, where d is the propagation distance [9], α is the path-loss exponent, and ε is a small constant for numerical stability. When the current mode does not belong to the large-scale family, $P_{L,dB}$ is set to zero. In addition, $\xi = 0$ is used for AWGN or no fading, while $\xi = 1$ is used for Rayleigh fading. In this way, the quality vector keeps a unified dimension across all channel modes, although its last two entries only contain nonzero effective information in the corresponding scenarios.

Based on this unified quality description, the quality condition is mapped into a gating vector through a conditional branch and fused with the feature branch by element-wise modulation. Thus, large-scale path-loss and fading factors do not directly enter the physical channel layer, but indirectly affect the encoder and decoder responses through feature reweighting. Furthermore, to adapt to different discrete channel modes, the active channel mode $chan_mode$ is mapped into a discrete identifier $mode_id = M(chan_mode)$. AWGN, Rayleigh fading, 2×1 , 2×2 , 4×1 , 4×2 , and 8×1 MIMO modes, together with their large-scale variants, are uniformly encoded as integer indices. The mode condition is then transformed into another gating vector and applied to the encoder and decoder features in the same conditional modulation manner. As a result, different channel types correspond to different feature reweighting patterns, enabling a single model to adapt to multiple discrete channel modes while sharing the same backbone parameters.

3 Channel Model

To support multiple wireless channel modes within a unified framework, the internal structure of the proposed channel module is illustrated in Fig. 2.

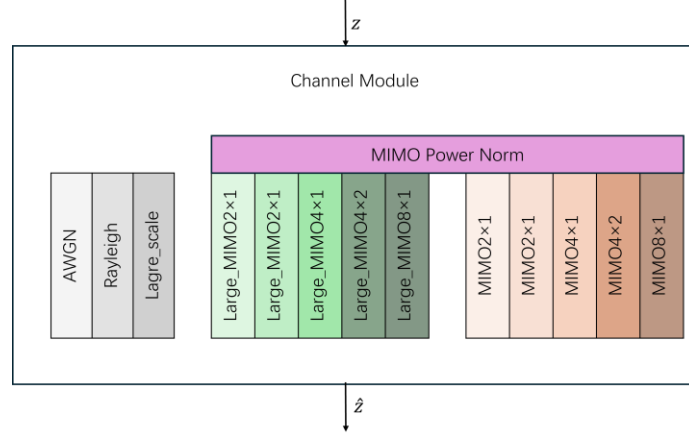


Fig. 2. Internal structure of the channel module.

The channel input and output are represented using a unified complex-valued form. Before transmission, the real-valued network output is split into two parts, which are respectively treated as the real and imaginary components of a complex signal. After transmission, the received complex signal is converted back into real-valued tensors for subsequent decoding. In this way, the backbone network remains real-valued, while the channel layer can model AWGN channels, fading channels, and MIMO transmission processes. This unified interface enables the model to switch among different channel modes without modifying the backbone architecture.

3.1 SISO Channel Modeling

In the single-input single-output (SISO) case, three basic channels are considered, namely the AWGN channel, the Rayleigh fading channel, and the large-scale path-loss extended channel [10]. For the AWGN channel, let s denote the transmitted complex symbol. The received signal is given by

$$y = s + n \quad (1)$$

where $n \sim \mathcal{CN}(0, \sigma^2)$ is complex Gaussian white noise.

For the Rayleigh fading channel, let the complex fading coefficient satisfy

$$h \sim \mathcal{CN}(0, 1) \quad (2)$$

Then, the received signal can be written as

$$y = hs + n \quad (3)$$

and an idealized equalization step is further adopted in the implementation to obtain the equalized feature:

$$\tilde{y} = \frac{y}{h + \varepsilon} \quad (4)$$

where ε is a small constant introduced to avoid numerical instability.

For the large-scale path-loss extended channel, the transmitted symbol first undergoes path-loss attenuation and is then combined with AWGN or Rayleigh small-scale fading. The received signal can be expressed as

$$y = \sqrt{P_L} hs + n \quad (5)$$

where

$$P_L = d^{-\alpha} \quad (6)$$

denotes the linear path-loss factor, d is the propagation distance, and α is the path-loss exponent. Its logarithmic form is defined as

$$P_{L,dB} = 10 \log_{10}(P_L + \varepsilon) \quad (7)$$

where ε is a small constant used to avoid numerical underflow. In the current implementation, the large-scale path-loss term is computed from `pl_alpha` and `pl_d`, and is written into the third entry of the quality vector $\mathbf{q}$. If the current mode does not belong to the large-scale family, this term is set to zero.

To unify different channel types within the same framework, any channel mode can be abstracted as

$$y = \mathcal{C}_m(s; \gamma, \eta) \quad (8)$$

where γ denotes quality-related parameters such as SNR, and η denotes mode-specific parameters, such as the antenna configuration, the path-loss coefficient, and the fading type. With this unified channel interface, the training process can switch among different channel modes without changing the input-output form of the backbone network.

3.2 MIMO Channel Modeling

The supported training channel modes include Rayleigh-based MIMO extensions and large-scale MIMO variants. In the proposed method, MIMO gain is not simplified as an equivalent SNR increase. Instead, transmit diversity and spatial multiplexing are modeled as two different transmission mechanisms. Specifically, MIMO2×1, MIMO4×1, and MIMO8×1 are regarded as transmit diversity modes, while MIMO2×2 and MIMO4×2 are regarded as spatial multiplexing modes. This categorization is consistent with the physical meanings of different MIMO structures [11].

For transmit diversity modes, an Alamouti-inspired design is adopted. In the 2×1 case, two adjacent complex symbols s_0 and s_1 are arranged into the following transmit matrix over two time slots:

$$\mathbf{S} = \begin{bmatrix} s_0 & -s_1^* \\ s_1 & s_0^* \end{bmatrix} \quad (9)$$

For the 4×1 and 8×1 cases, multiple Alamouti pairs are further constructed in parallel, and their received branches are jointly combined at the receiver. The corresponding received signals over two time slots can be written as

$$y_1 = \beta \sum_p (h_{1,p} s_0 + h_{2,p} s_1) + n_1 \quad (10)$$

$$y_2 = \beta \sum_p (-h_{1,p} s_1^* + h_{2,p} s_0^*) + n_2 \quad (11)$$

where β denotes the power normalization factor. The combined estimates are given by

$$\hat{s}_0 = \frac{\sum_p (h_{1,p}^* y_1 + h_{2,p} y_2^*)}{\sum_p (|h_{1,p}|^2 + |h_{2,p}|^2) + \varepsilon} \quad (12)$$

$$\hat{s}_1 = \frac{\sum_p (h_{2,p}^* y_1 - h_{1,p} y_2^*)}{\sum_p (|h_{1,p}|^2 + |h_{2,p}|^2) + \varepsilon} \quad (13)$$

For the spatial multiplexing modes, the proposed method considers the 2×2 and 4×2 structures. Let the two transmitted complex symbol streams be denoted by

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (14)$$

In the 2×2 case, an identity precoding matrix is directly used, whereas in the 4×2 case, a fixed 4×2 precoding matrix is adopted to map the two streams onto four transmit antennas. The received signal model can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{n} \quad (15)$$

where $\mathbf{H}$ is the complex channel matrix, $\mathbf{F}$ is the precoding matrix, and $\mathbf{n}$ is the noise vector. To recover the transmitted streams, MMSE detection is adopted at the receiver.

Compared with simply approximating multiple transmission branches as a higher equivalent SNR, explicit MIMO modeling can better reflect the differences in coupling relationships, detection difficulty, and performance gains among different MIMO structures. Transmit diversity mainly improves reliability and is more suitable for low-SNR or strong-fading scenarios, while spatial multiplexing mainly enhances parallel transmission capability and is more suitable for better channel conditions and higher throughput requirements. Therefore, introducing both mechanisms into the semantic communication framework helps the model learn transmission-aware feature scheduling patterns that are more consistent with practical physical channels.

4 Training Procedure

To describe the training process of the proposed framework, Algorithm 1 summarizes the overall end-to-end optimization procedure. In each iteration, the model samples a channel condition, generates the corresponding quality vector and mode identifier, and

performs conditional modulation at the encoder and decoder. Through this process, a single model can learn to adapt to multiple channel modes.

Algorithm 1. Training procedure of the proposed SwinJSCC-based mode-adaptive framework for MIMO channels

```

1: Input: training dataset  $D$ , SNR set  $\Gamma$ , rate set  $\mathcal{R}$ , channel mode set  $\mathcal{M}$ , initial parameters  $\theta^{(0)}$ ;
2: Output: optimized parameters  $\theta^{(T)}$ ;
3: Initialize  $\theta^{(0)}$ ;
4: for  $t = 1$  to  $T$  do
5:   \ Sample a mini-batch  $x \sim D$ , and sample  $\gamma$ ,  $R$ , and  $\text{chan\_mode}$ ;
6:   \ Build  $\mathbf{q} = [\gamma, R, P_{L,AB}, \xi]$  and obtain  $\text{mode\_id}$  from  $\text{chan\_mode}$ ;
7:   \ Encode  $x$  with the Encoder Backbone to obtain latent features;
8:   \ Perform encoder-side conditional modulation using Quality ModNet, Mode ModNet, Channel ModNet, and Rate ModNet to generate  $z$ ;
9:   \ Transmit  $z$  through PowerNorm and the Channel Module to obtain  $\hat{z}$ ;
10:  \ Perform decoder-side conditional modulation and reconstruct  $\hat{x}$  with the Decoder Backbone;
11:  \ Compute  $\mathcal{L}(\hat{x}, x)$  and update  $\theta^{(t)}$ ;
12: end for

```

Algorithm 1 summarizes the unified training process of the proposed framework under different channel conditions. By introducing both quality-aware and mode-aware conditioning into the encoder-decoder pipeline, the model can learn channel-dependent feature modulation within a shared backbone. This enables single-model adaptation to multiple MIMO channel scenarios.

5 Experimental Results

5.1 Experimental Setup

The proposed SwinJSCC-based image semantic communication method was implemented in PyTorch [12]. CIFAR10 was used for both training and testing [13], which contains 60,000 color images with a resolution of 32×32 , including 50,000 training images and 10,000 test images. The experiments were conducted on an NVIDIA GeForce RTX 4090D GPU. The learning rate was set to 1×10^{-4} , the total number of epochs was 41, the evaluated model was selected at epoch 40, and the batch size was set to 128.

In the CIFAR10 experiments, the bottleneck channel dimension was set to $C = 192$, corresponding to $\text{CBR} = 2$. Unless otherwise specified, $\text{CBR} = 2$ is used in the following CIFAR10 experiments. The original SA/RA modules of SwinJSCC were retained, and the framework was extended to multiple channel modes, including AWGN, Rayleigh fading, MIMO transmit diversity, spatial multiplexing, and large-scale path-loss extended channels. With mode adaptation, the model can be jointly trained over multiple channel modes within a unified framework.

5.2 Cross-Channel Adaptation Performance

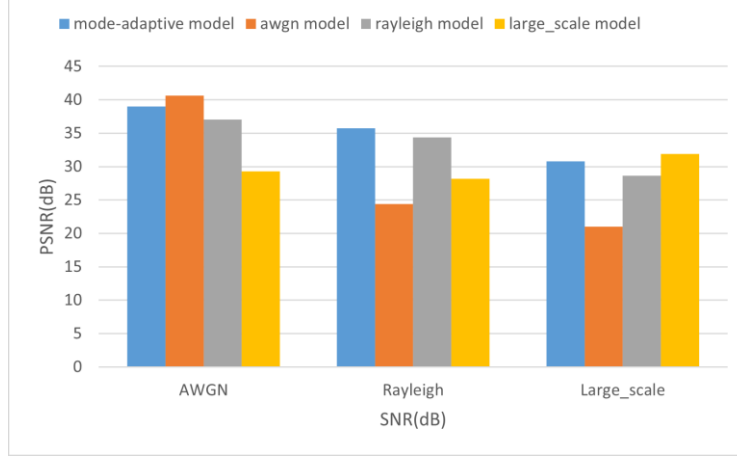


Fig. 3. Comparison of the adaptation capability of the CIFAR10-trained mixed-channel mode-adaptive model, AWGN model, Rayleigh model, and large-scale model under different test channel modes at fixed CBR = 2 and SNR = 4 dB.

At the fixed test condition of SNR = 4 dB and CBR = 2, Fig. 3 compares the mode-adaptive model with three single-channel trained models under different test channels. The proposed mode-adaptive model achieves PSNR values of 39.0 dB, 35.7 dB, and 30.8 dB under AWGN, Rayleigh, and large-scale channels, respectively, showing good reconstruction quality and cross-channel robustness.

Although single-channel models perform well when the test channel matches the training channel, their performance drops significantly under mismatched channels. For example, the AWGN model achieves 40.6 dB under the AWGN test channel, but decreases to 24.4 dB and 21.0 dB under Rayleigh and large-scale channels, respectively. In contrast, the mode-adaptive model maintains more stable performance and even outperforms the Rayleigh-only model under the Rayleigh test channel. These results indicate that mixed-channel training with mode adaptation enables a single model to achieve more balanced performance across multiple channel environments.

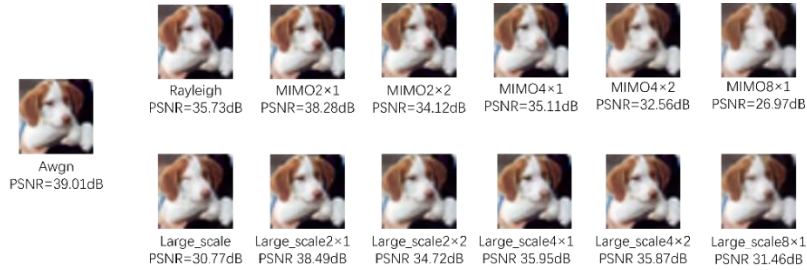


Fig. 4. Visualization of the reconstructed CIFAR10 image under different MIMO test channels for the mode-adaptive model at fixed CBR = 2 and SNR = 4 dB.

Visualization of the reconstructed CIFAR10 dog image for the mode-adaptive model under different MIMO test channels at fixed CBR = 2 and SNR = 4 dB. The visual results show that the proposed model is able to preserve the main object structure and overall texture details under different MIMO test channels, while slight differences can still be observed in local reconstruction quality.

5.3 Performance under MIMO Families

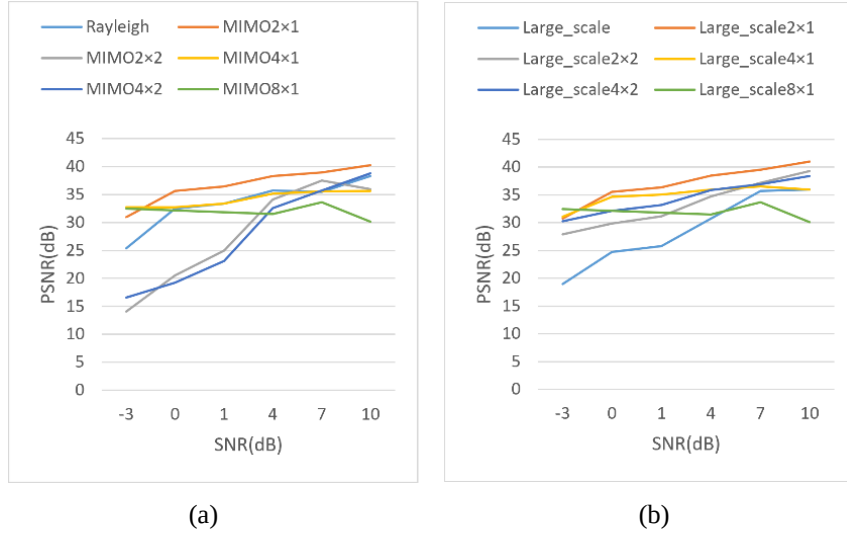


Fig. 5. Performance comparison of the CIFAR10-trained mixed-channel mode-adaptive model under different MIMO antenna configurations at fixed CBR = 2 and varying SNRs. (a) Rayleigh channel and its MIMO family. (b) Large-scale channel and its MIMO family.

At a fixed CBR of 2, Fig. 5 compares the PSNR performance of the mode-adaptive model under different SNRs for the Rayleigh MIMO and large-scale MIMO channel families. Overall, PSNR increases with SNR, while different MIMO modes show different trends due to the difference between diversity and multiplexing mechanisms.

For the Rayleigh MIMO family, the 2×1 diversity mode achieves the best overall performance. Its PSNR reaches 30.9 dB, 38.3 dB, and 40.2 dB at −3 dB, 4 dB, and 10 dB, respectively, all higher than the corresponding Rayleigh fading results. The large-scale 2×1 mode shows a similar trend, indicating that the 2×1 diversity structure can effectively improve transmission reliability under the fixed-CBR setting.

However, more transmit antennas do not always lead to better performance. Although 4×1 and 8×1 also belong to diversity-based modes, their performance is affected by total transmit-power normalization. Excessive diversity branches may split the available power and reduce effective semantic transmission efficiency.

Multiplexing-based modes, such as 2×2 and 4×2, are more sensitive to SNR. They perform poorly at −3 dB but improve significantly at higher SNRs. This indicates that spatial multiplexing can exploit parallel transmission when channel quality is good, but is more vulnerable to noise and inter-stream interference under low-SNR conditions.

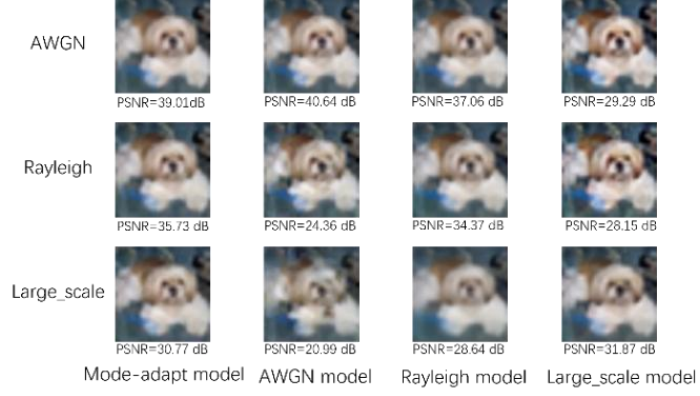


Fig. 6. Visualization of the reconstructed CIFAR10 image for the CIFAR10-trained mixed-channel mode-adaptive model, AWGN-trained model, Rayleigh-trained model, and large-scale-trained model under AWGN, Rayleigh, and large-scale test channels at fixed CBR = 2 and SNR = 4 dB.

The visual results show that the mode-adaptive model maintains consistently good reconstruction quality under different test channels, indicating strong channel adaptation capability. In contrast, the single-channel-trained AWGN, Rayleigh, and large-scale models perform well mainly under their matched test channels, but their reconstruction quality degrades noticeably under mismatched channels. This degradation is clearly visible, with more blur, loss of texture details, and stronger local distortions.

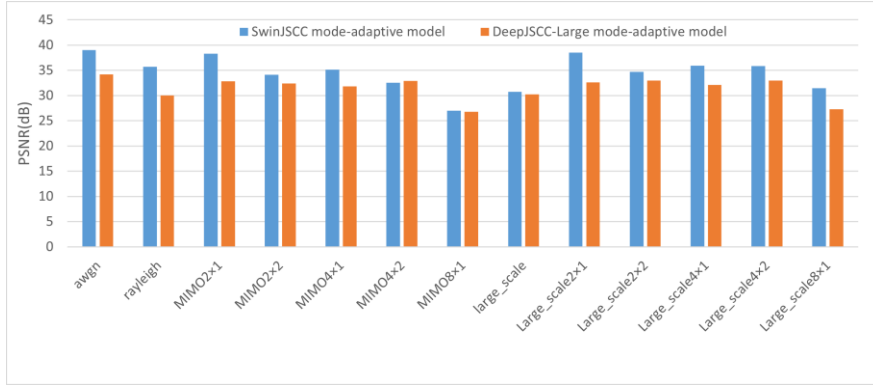


Fig. 7. Comparison of performance between the proposed mode-adaptive SwinJSCC model and the mode-adaptive DeepJSCC-Large model.

Overall, the SwinJSCC mode-adaptive model achieves higher PSNR than the DeepJSCC-Large mode-adaptive model under most channel modes, showing stronger feature representation and robustness. For example, under AWGN, SwinJSCC reaches 39.02 dB, while DeepJSCC-Large reaches 34.2 dB.

Compared with the corresponding non-MIMO channels, MIMO2×1 improves the performance of both models. The PSNR of SwinJSCC increases from 35.73 dB under Rayleigh fading to 38.28 dB under MIMO2×1, while DeepJSCC-Large increases from 30.0 dB to 32.8 dB. Although both models degrade under challenging channels such as MIMO8×1 and large-scale MIMO8×1, the proposed adaptive SwinJSCC model still maintains better overall performance and stronger cross-channel robustness.

5.4 Computational Complexity and Inference Latency Analysis

Table 1. Complexity and Inference Latency Comparison of DeepJSCC-Large and SwinJSCC Models

Transmission scheme	Adaptive module	FLOPs (G)	Parameters (M)	Inference time
Original DeepJSCC-Large	None	4.05	25.84	2.80 ms
Original SwinJSCC	SA/RA	1.69	16.96	9.87 ms
Proposed DeepJSCC-Large	Mode-adapt	4.05	25.85	3.29 ms
Proposed SwinJSCC	Mode ModNet + Quality ModNet	2.12	29.61	12.47 ms

As shown in Table 1, all models are evaluated under the same profiling setting on CIFAR10 with SNR = 4 dB and C = 192. The mimo2×1 mode is fixed as the profiling channel mode. For DeepJSCC-Large, ratio = 4 corresponds to the inner channel dimension C = 192, while for SwinJSCC, C = 192 corresponds to CBR = 2 under the CIFAR10 setting.

DeepJSCC-Large is introduced to provide a fairer comparison with SwinJSCC by enlarging the CNN-based DeepJSCC backbone. The original DeepJSCC-Large contains 25.83M parameters and requires 4.05G FLOPs. After introducing the mode-adaptive module, the parameter count slightly increases to 25.85M, and the FLOPs increase to 4.05G. This indicates that the mode-adaptive module introduces only marginal additional complexity.

For SwinJSCC, the original model with SA/RA modules contains 16.96M parameters and requires 1.69G FLOPs. After incorporating Mode ModNet and Quality ModNet, the proposed SwinJSCC model contains 29.61M parameters and requires 2.12G FLOPs. The inference time increases from 9.87 ms to 12.47 ms per image. Although the adaptive modules increase the computational cost, the inference latency remains within a low millisecond-level range, indicating that the proposed adaptive SwinJSCC model maintains feasible computational efficiency.

Conclusion

This paper proposes a SwinJSCC-based adaptive image semantic communication method for MIMO channels to improve image transmission adaptability in complex wireless environments. By introducing channel mode information into the end-to-end transmission process, the proposed method enables conditional modulation and feature

adaptation under AWGN, Rayleigh fading, MIMO, and large-scale path-loss extended channels.

Experimental results show that the proposed adaptive model achieves favorable reconstruction performance and cross-channel robustness. In particular, the MIMO 2×1 diversity mode provides a clear performance gain, indicating that transmit diversity can improve transmission reliability and reconstruction quality. Compared with models trained for a single channel or a specific MIMO mode, the proposed method provides more balanced overall performance and is more suitable for unified deployment in complex wireless environments.

Future work will further explore multiplexing-based MIMO, semantic feature allocation, and mode selection strategies to improve the efficiency and robustness of semantic communication systems.

References

1. Xie, H., Qin, Z., Li, G.Y., Juang, B.-H.: Deep Learning Enabled Semantic Communication Systems. *IEEE Transactions on Signal Processing* 69, 2663–2675 (2021).
2. Liu, Z., Lin, Y., Cao, Y., et al.: Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9992–10002 (2021).
3. Bourtsoulatzé, E., Kurka, D.B., Gündüz, D.: Deep Joint Source-Channel Coding for Wireless Image Transmission. *IEEE Transactions on Cognitive Communications and Networking* 5(3), 567–579 (2019).
4. Yang, K., Wang, S., Dai, J., Qin, X., Niu, K., Zhang, P.: SwinJSCC: Taming Swin Transformer for Deep Joint Source-Channel Coding. *IEEE Transactions on Cognitive Communications and Networking* 11(1), 90–104 (2025)
5. Zhang, H., Shao, S., Tao, M., Bi, X., Letaief, K.B.: Deep Learning-Enabled Semantic Communication Systems With Task-Unaware Transmitter and Dynamic Data. *IEEE Journal on Selected Areas in Communications* 41(1), 170–185 (2023).
6. Zhang, H., Tao, M., Sun, Y., Letaief, K.B.: Improving Learning-Based Semantic Coding Efficiency for Image Transmission via Shared Semantic-Aware Codebook. *IEEE Transactions on Communications* 73(2), 1217–1232 (2025).
7. Alamouti, S.M.: A Simple Transmit Diversity Technique for Wireless Communications. *IEEE Journal on Selected Areas in Communications* 16(8), 1451–1458 (1998).
8. Tse, D., Viswanath, P.: *Fundamentals of Wireless Communication*. Cambridge University Press, Cambridge (2005).
9. Rappaport, T.S.: *Wireless Communications: Principles and Practice*. 2nd edn. Prentice Hall, Upper Saddle River (2002).
10. Goldsmith, A.: *Wireless Communications*. Cambridge University Press, Cambridge (2005).
11. Foschini, G.J.: Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multi-Element Antennas. *Bell Labs Technical Journal* 1(2), 41–59 (1996).
12. Paszke, A., Gross, S., Massa, F., et al.: PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, pp. 8024–8035 (2019).
13. Krizhevsky, A., Hinton, G.: Learning Multiple Layers of Features from Tiny Images. Technical report, University of Toronto, Toronto (2009).