Department of Electrical and Computer Engineering
Cullen College of Engineering

Generative AI Enabled Semantic Communication

Zhu Han
IEEE/AAAS/ACM Fellow
University of Houston, TX USA

Thanks to Zhijin Qin, Weilong Chen, Weimin Yuan, Faheem Quazi, Loc X. Nguyen, Choong Seon Hong, Yiru Wang, Zehui Xiong, and US National Science Foundation

Outline

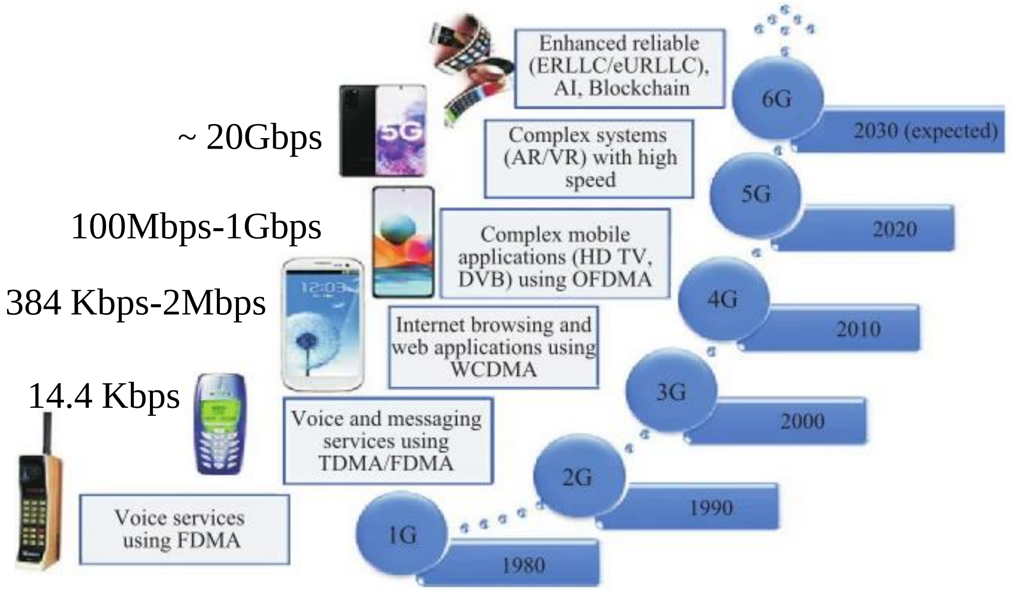
- **Overview: Semantic Communications and GAI**
- **Research Applications**
  - Swin-Transformer-Based Dynamic Semantic Communication for Multi-User With Different Computing Capacity
  - An Efficient Federated Learning Framework for Training Semantic Communication Systems.
  - AI-Generated Content for SCM (AIGC-SCM)
- **Demo of Generative AI Enabled Semantic Communication**
- **Conclusion and Future Direction**

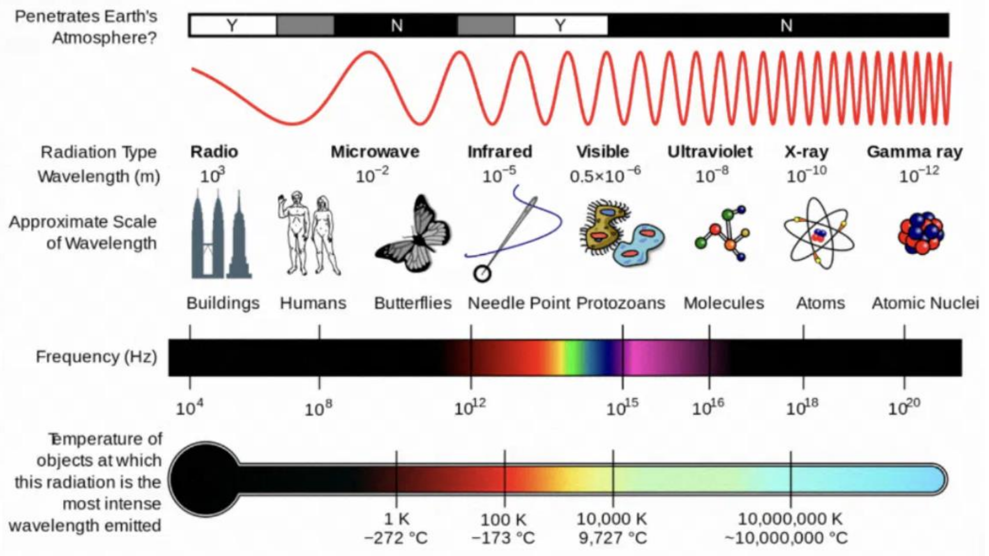
Overview: Conventional Communications

- **Conventional communications**
  - ➢ Bit/symbol accurate delivery
  - ➢ Regardless of content

- **Performance improvement**
  - ➢ Increase the dimension of information
    - ✓e.g., massive MIMO
  - ➢ Optimize resource allocation

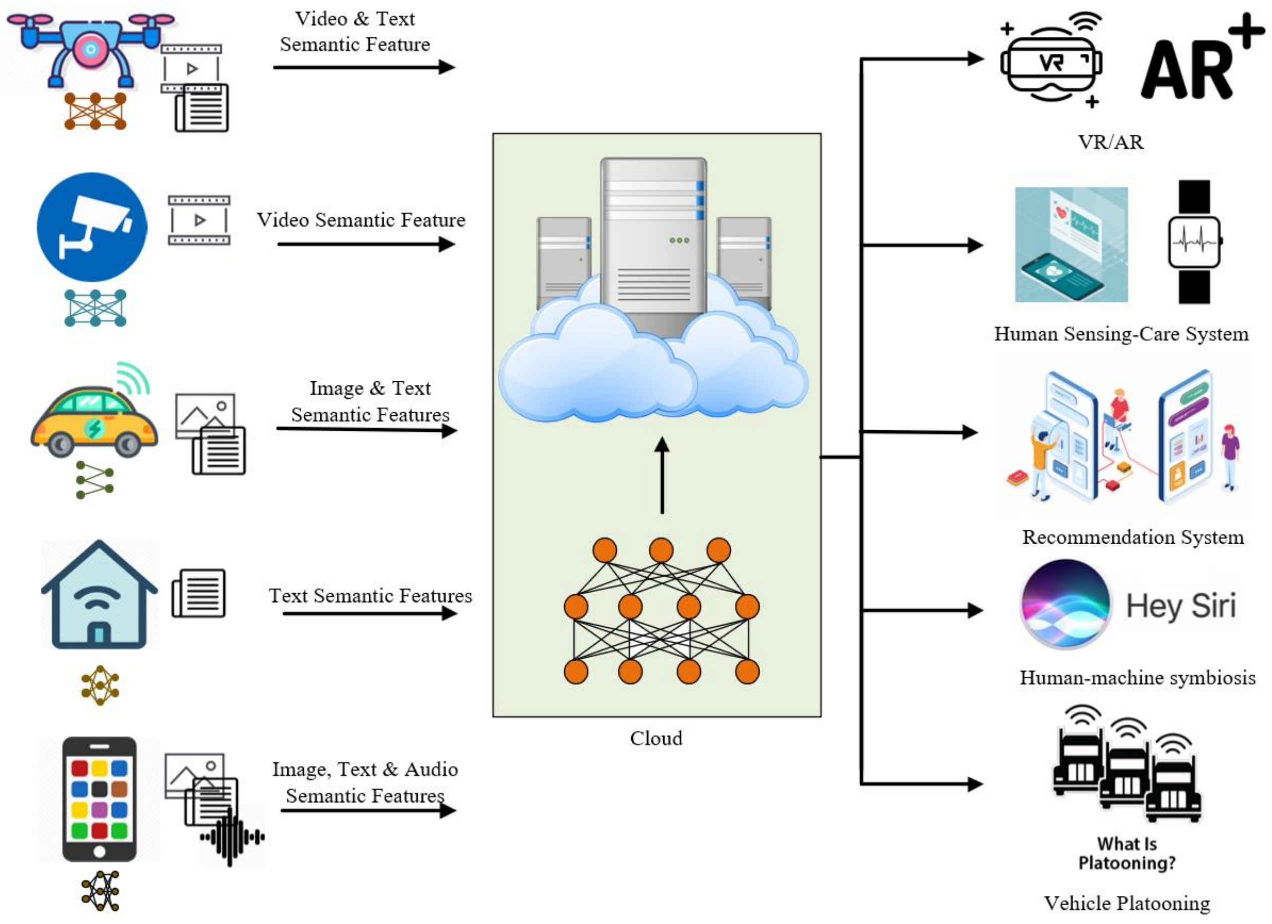
- **Bottlenecks**
  - ➢ Approaching the Shannon limit
  - ➢ The large power consumption
  - ➢ The spectrum shortage





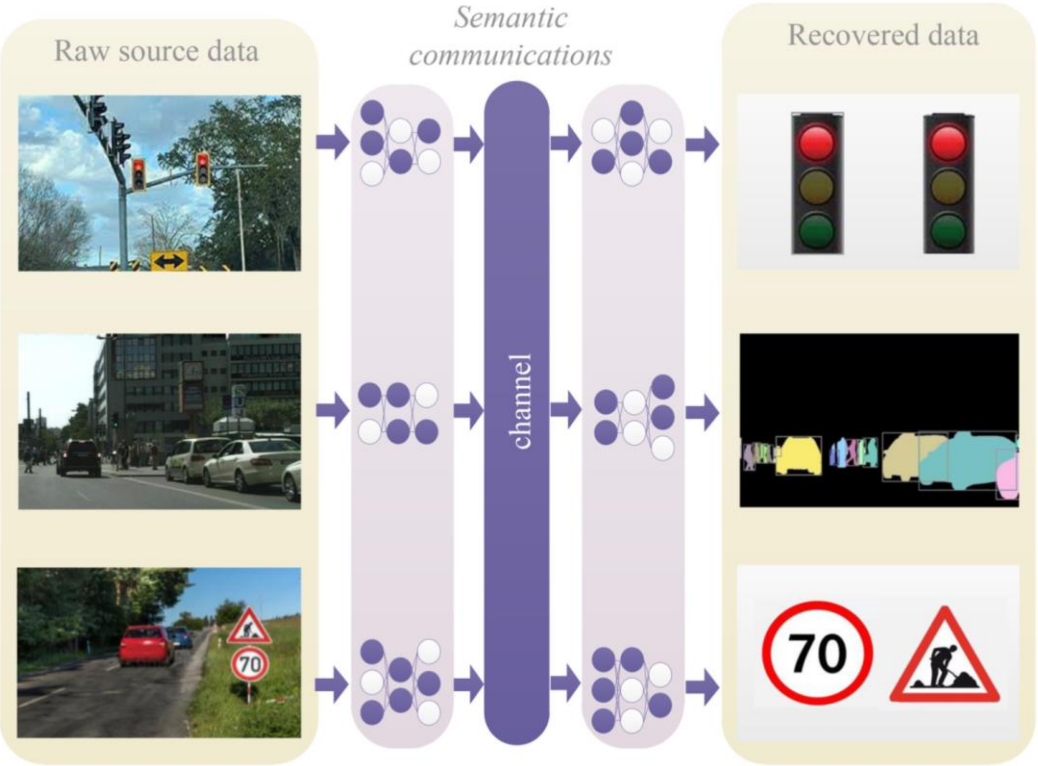
Semantic Communications in 6G and Beyond

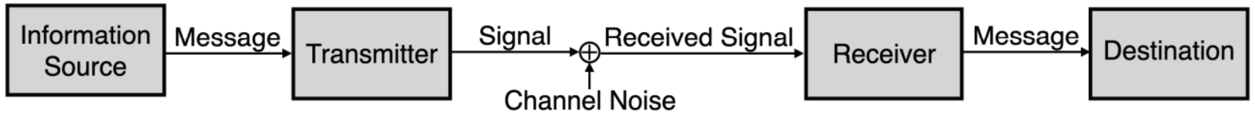
- **Research for 6G on the way**
- **Key role: intelligence transmission**
  - ➢ Semantic communications
- **Applications**
  - ➢ Machine-to-machine communications
  - ➢ Human-to-machine communications
  - ➢ Human-to-human communications
- **State-of-the-art: in its infancy**
  - ➢ Government research council
  - ➢ Telecom and AI companies



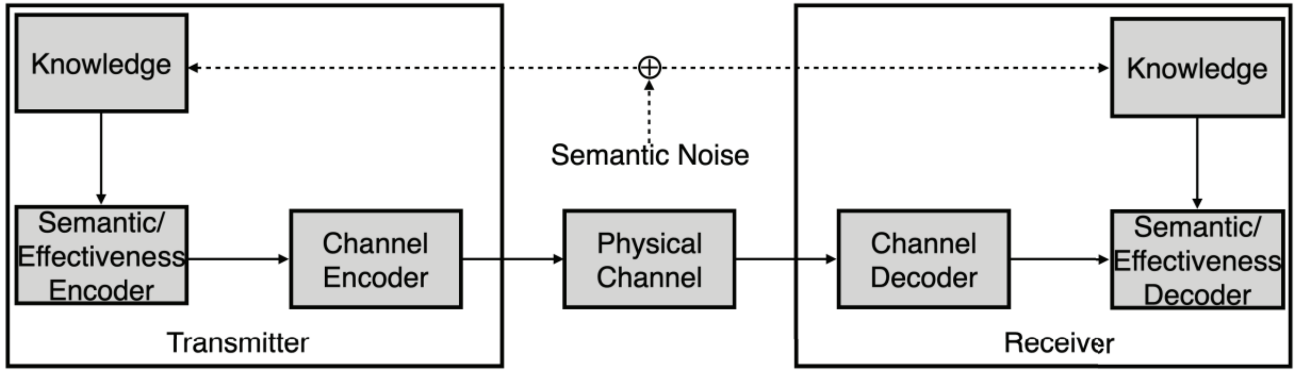
Characteristics of Semantic Communications

The semantic extraction process can filter out irrelevant image details for different tasks before transmission by performing the appropriate image processing techniques, thereby relieving the network burden without compromising the system's performance.





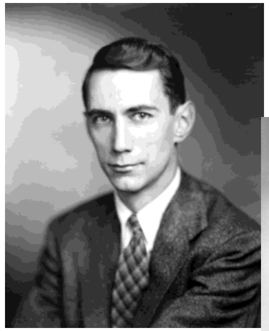
Communication System in Shannon's Theory



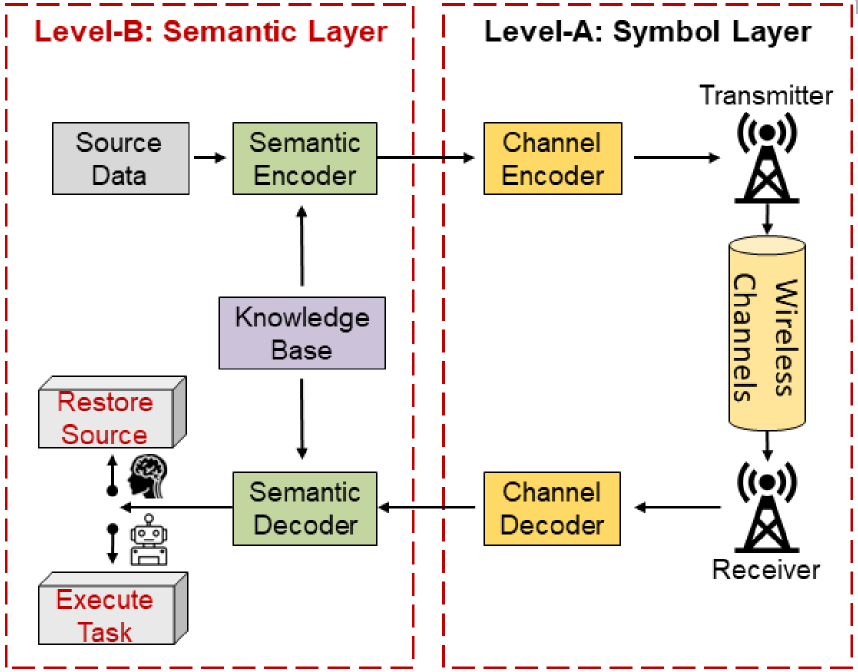
Semantic Communication System

Conventional vs. Semantic Comm

- **Shannon-Weaver three-level communications**
  - ➢**Level A: Transmission of symbols (technical problem)**
  - ➢**Level B: Semantic exchange of source information (sematic problem)**
  - ➢**Level C: Effects of semantic information exchange (effectiveness problem)**

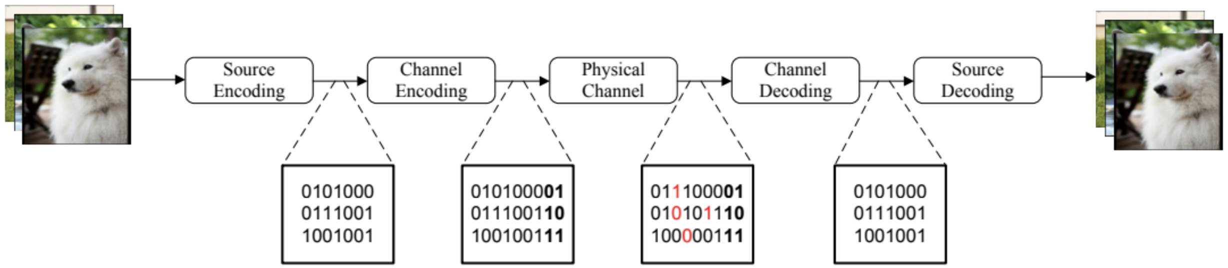


- **Semantic system architecture**
  - ➢ **Semantic knowledge base: perceive semantic features**
  - ➢ **Semantic encoder: extract semantics from input data**
  - ➢ **Semantic decoder: restore semantics per request of tasks**

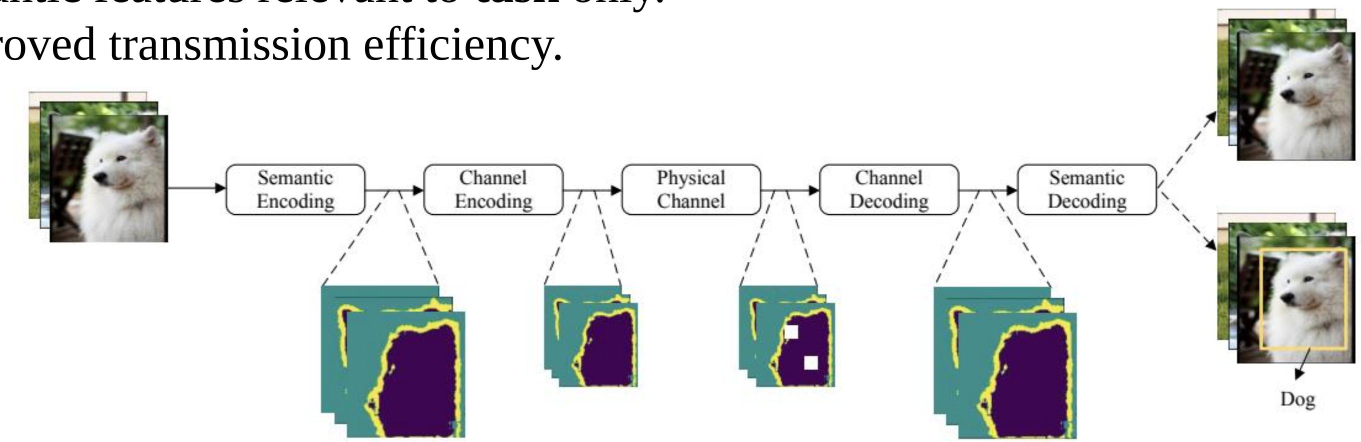


Conventional vs. Semantic Comm

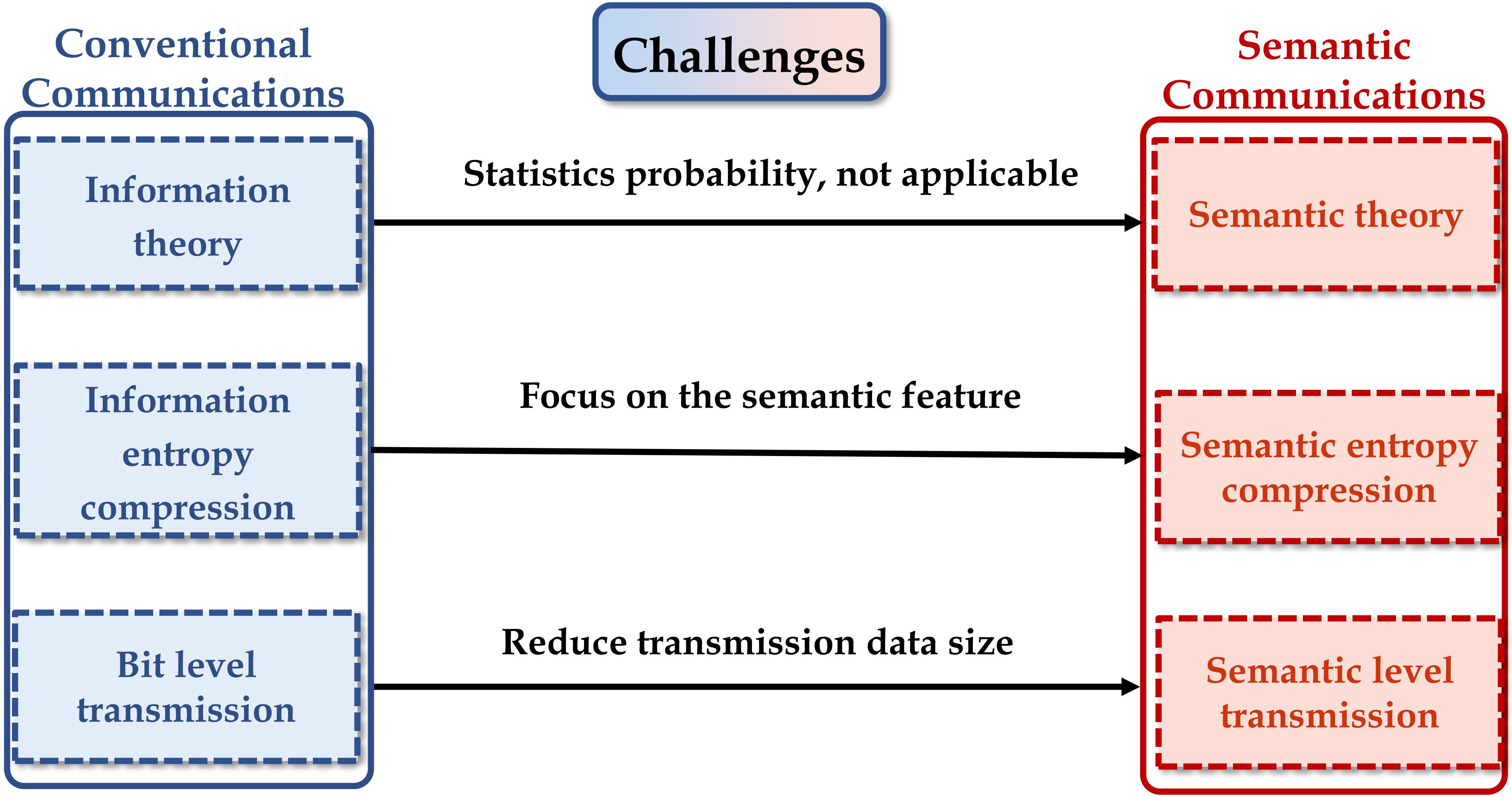
- **Conventional communications**
  - ➢ A tube for accurate transmission of symbols.
  - ➢ Regardless of content in source.



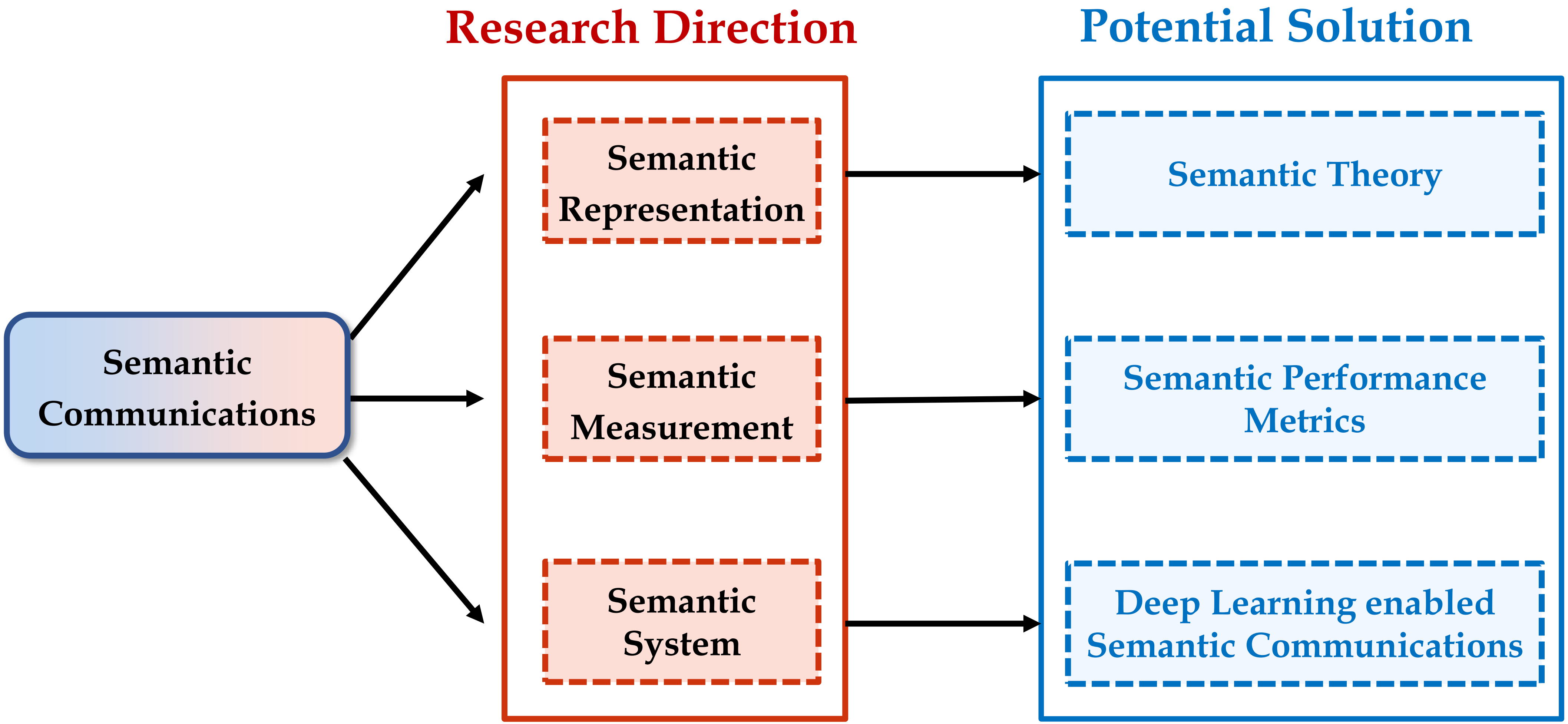
- **Semantic communications**
  - ➢ Transmitted symbols convey the desired meaning.
  - ➢ Transmitting semantic features relevant to **task** only.
  - ➢ Significantly improved transmission efficiency.



Conventional vs. Semantic Comm.



Semantic Communication Challenges



Overview: In the past

- **Initial semantic communication works**
  - Logic probability based semantic communication [1,2]
  - Word-level based semantic communication [3]
  - Cannot fully understand the meaning behind texts

- **Derive semantic capacity of a discrete memoryless channel** [2]:

$$C_s = \sup_{P(X|W)} \left\{ I(X;Y) - H(W|X) + \overline{H_s(Y)} \right\}$$

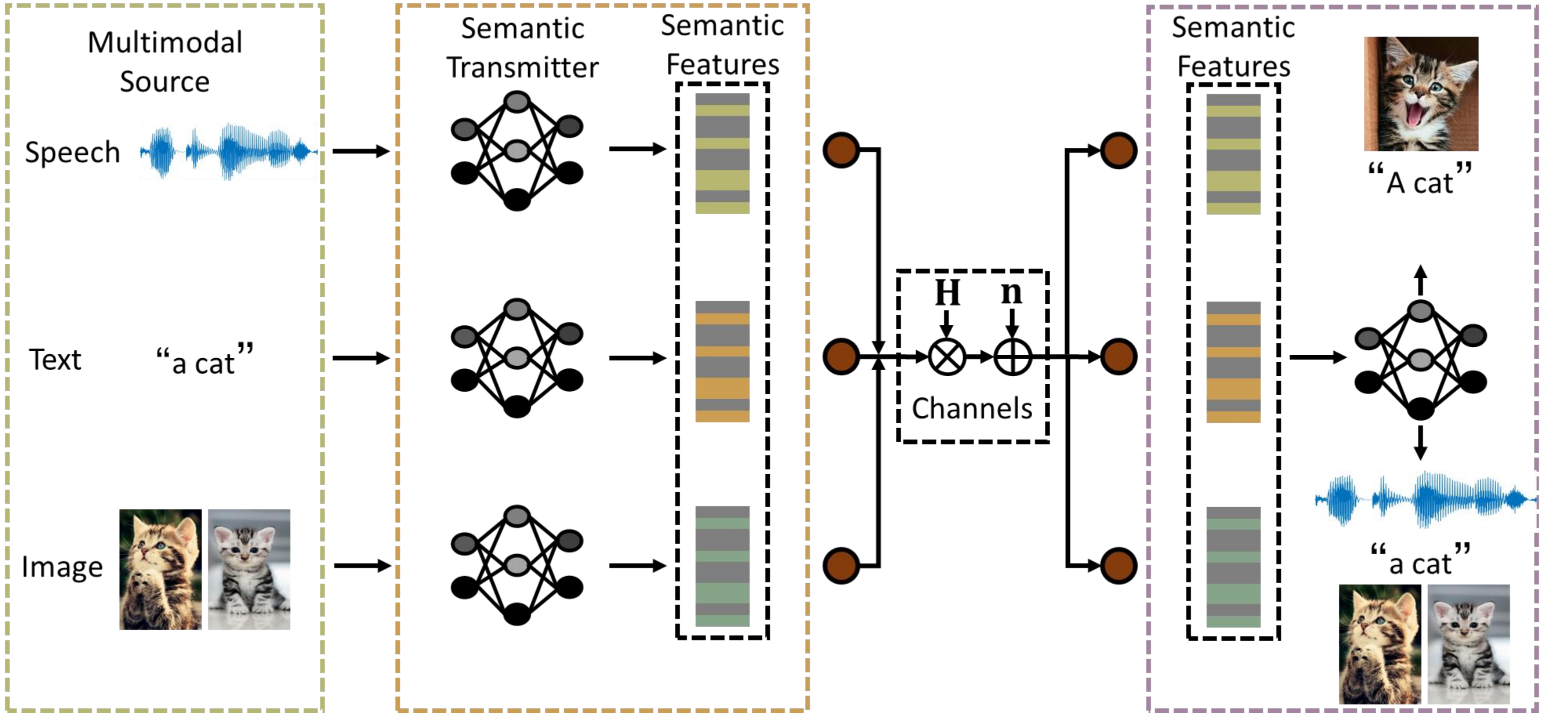
  - *I(X;Y)* is the mutual information
  - $\overline{H_s(Y)}$ is the average logical information of received messages, representing the ability to interpret received messages
  - *P(X|W)* is the conditional probabilistic distribution of a semantic coding strategy

[1] R. Carnap et al., "An outline of a theory of semantic information," Res. Lab. Electronics, Massachusetts Inst. Technol., Cambridge MA, Oct. 1952.
[2] J. Bao *et. al*, "Towards a theory of semantic communication," in *IEEE Network Science Workshop*, West Point, NY, USA, Jun. 2011.
[3] B. Guler *et. al*, "The semantic communication game," *IEEE Trans. Cogn. Comm. Networking*, vol. 4, no. 4, pp. 787– 802, Sep. 2018.

AI enabled Semantic Communications

Advantages of DL

- Learn the latent semantic information representation.
- Extract the semantic features of source data.
- Achieve the end-to-end transmission optimization to recover semantic information.
- AI (e.g. deep learning) excels at handling large, complex, and unstructured data such as images, audio, and text.



Overview: Generative AI

Landmark

In 1950, the Turing Test was proposed.

In 1967, the human-computer dialogue machine Eliza was developed.

In 2007, the first novel created by AI was published.

Microsoft

In 2012, the fully automated simultaneous interpretation system.

In 2014, Goodfellow proposed GAN.

Google

In 2017, Google proposed Transformer.

OpenAI

In 2022, OpenAI proposed ChatGPT3.5.

Characteristic

Limited by technological capabilities, GAI developed slowly.

Substantial applications emerged, but were limited by computational power and algorithms.

GAI is thriving!

Stage

Initial stage (1950s – 1990s)

Slow development stage (1990s – 2010s)

Fast development stage (2010s – now)

Overview: Generative AI

- **Advantages of AIGC**

❑ **Automation and Cost-Saving**: Saves time and resources and eliminates manual effort.

❑ **Creativity and Innovation**: Pushes the boundaries of traditional human creativity.

❑ **Customization and Personalization**: Be tailored to specific preferences or individual user data.

❑ **Multimodal and Multicultural:** Adapts to multimodal inputs and diverse cultural contexts.

AI-generated Texts

AI-generated Audio

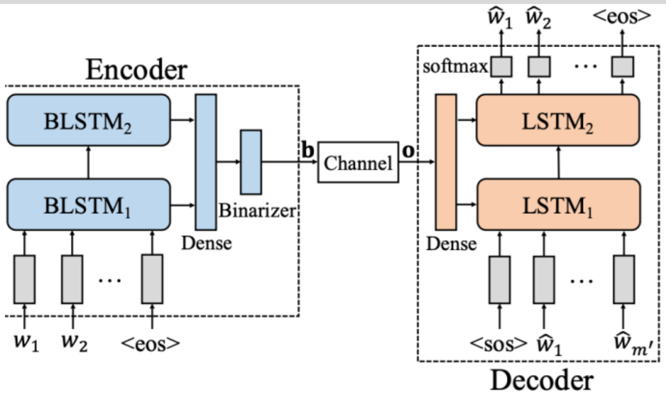
AI-generated Images

AI-generated Videos

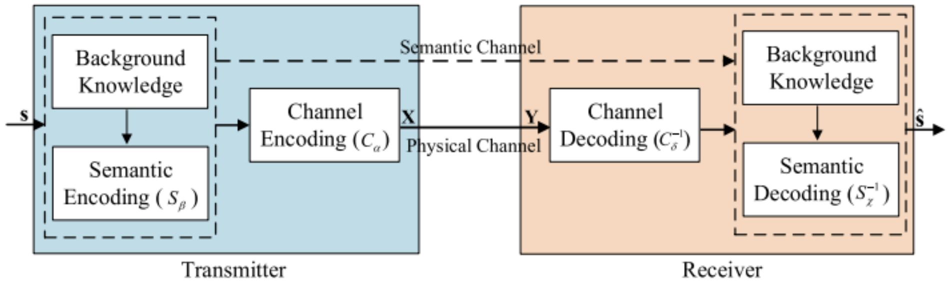
AI-generated 3D Content

Related Work: Text

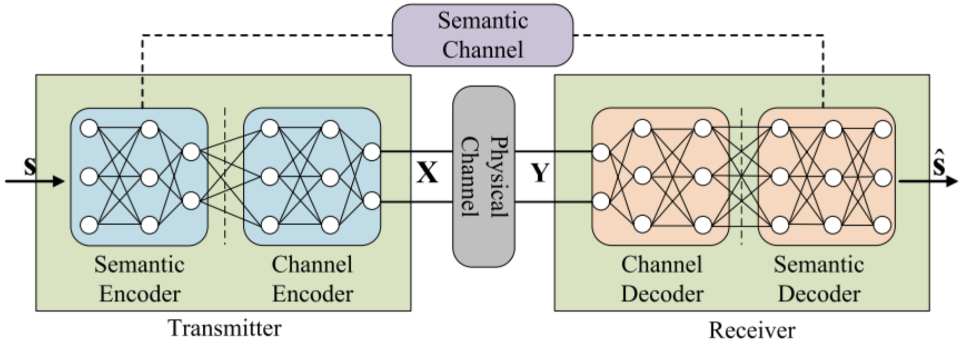
- **Deep joint source and channel coding (JSCC)** [1]
  - Recovers the text directly without performing channel and source decoding separately.



- **DeepSC** [2]
  - Clarify the concepts of semantic information and semantic error at the sentence-level for the first time.
  - Trained by maximizing the mutual information and minimizing the semantic errors.



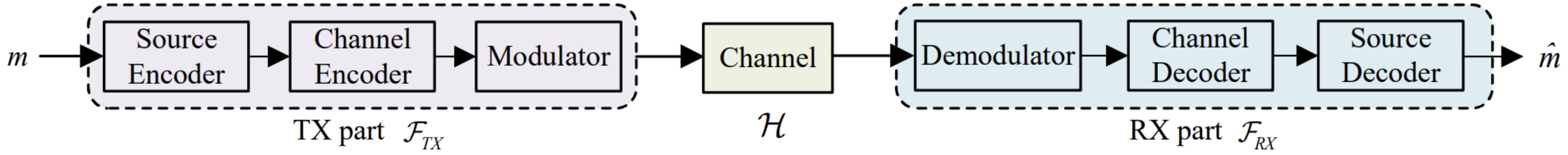
- **L-DeepSC** [3]
  - A lite DeepSC with small size and low complexity.
  - Affordable for IoT devices.



[1] N. Farsad *et. al*, "Deep learning for joint source-channel coding of text," *in Proc. IEEE ICASSP'18,* Calgary, AB, Canada, Apr. 2018, pp.2326–2330
[2] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE TSP*, Apr. 2021.
[3] H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," *IEEE JSAC*, Jan. 2021.

Related Work: Text

- **Reinforcement learning-powered semantic communication** [1]
  - Maximize the semantic similarity.
  - Use reinforcement learning to train the network.



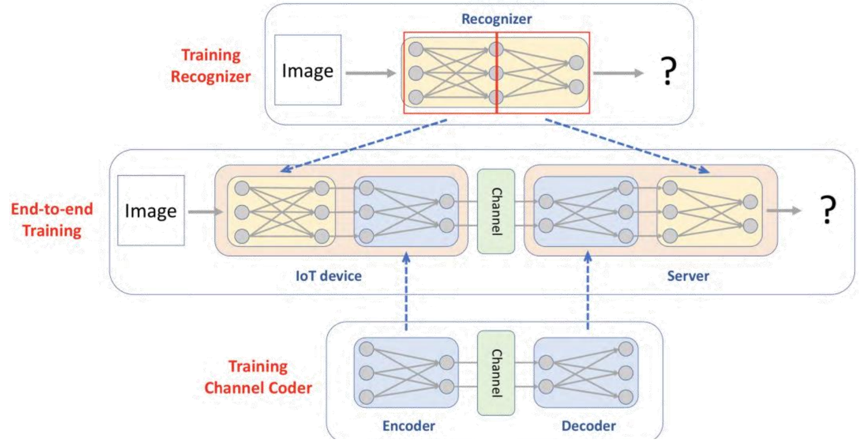
- **Sentence semantic transmission with HARQ** [2]
  - Combine semantic coding with Reed Solomon coding and HARQ, called SC-RS-HARQ, to improve the reliability of text semantic transmission.
  - Propose a similarity detection network to detect meaning error.

- **Semantic representation learning based E2E architecture** [3]
  - Capture the effects of semantic distortion.
  - Obtain performance gain for different languages.

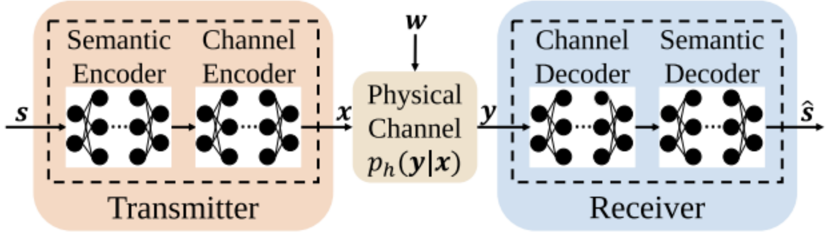
[1] K. Lu, R. Li, X. Chen, Z. Zhao, and H. Zhang, "Reinforcement learning-powered semantic communication via semantic similarity," 2021.
[2] Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep source-channel coding for sentence semantic transmission with HARQ," IEEE TCOM, 2022.
[3] M. Sana and E. C. Strinati, "Learning semantics: An opportunity for effective 6G communications," *Proc. ICC*, 2021.

Related Work: Image and Speech

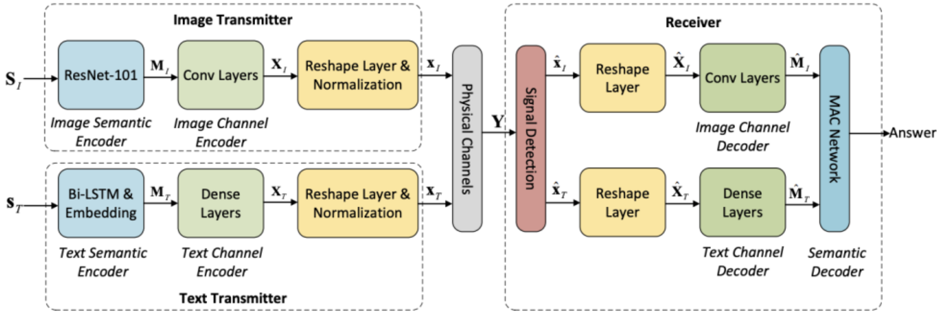
- **Transmission-recognition communication system** [1]
  - Jointly designed communication system and image classification network.
  - Achieves higher image classification accuracy than performing them separately.



- **DeepSC-S and DeepSC-ST** [2,3]
  - Joint semantic-channel coding for speech transmission.
  - Speech-to-text and speech synthesis.



- **MU-DeepSC** [4]
  - Multi-user semantic communications.
  - Multimodal data transmission.



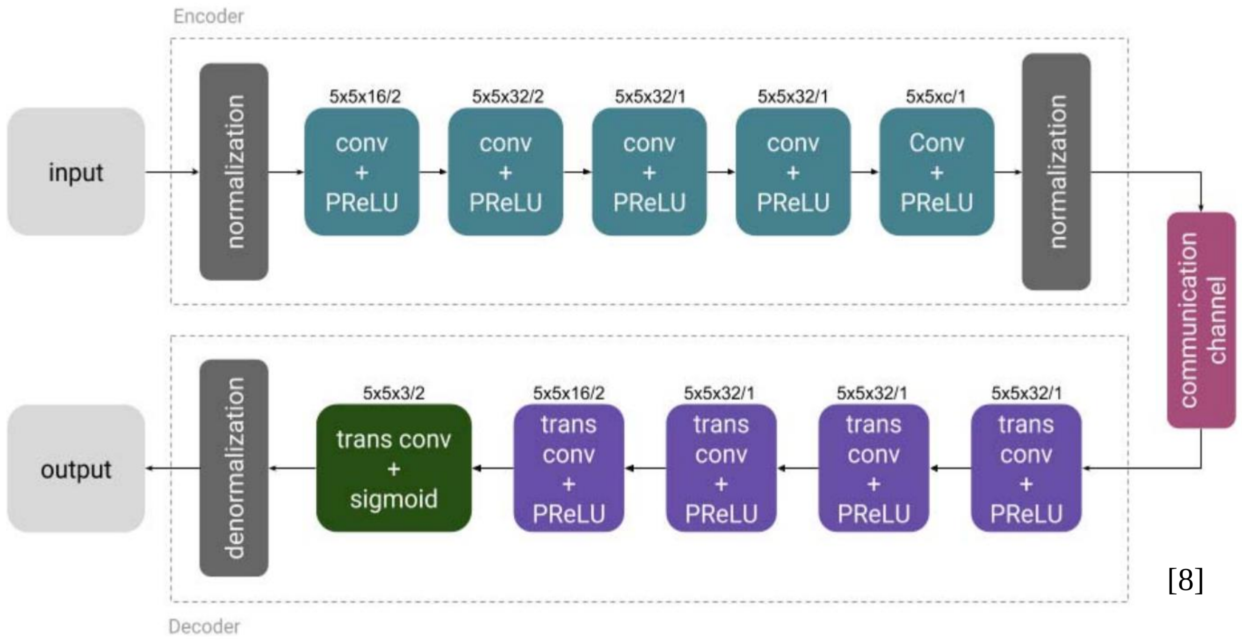
[1] C. Lee *et. al*, "Deep learning-constructed joint transmission-recognition for internet of things," *IEEE Access*, vol. 7, pp.76 547–76 561, Jun. 2019
[2] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE JSAC*, 2021
[3] Z. Weng, Z. Qin *et al*, "Deep Learning Enabled Semantic Communications with Speech Recognition and Synthesis," *IEEE TWC*, 2023
[4] H. Xie, Z. Qin, and G. Y. Li, "Task-oriented semantic communications for multimodal data", *IEEE JSAC*, 2021

Outline

- Overview: Semantic Communications and GAI
- **Research Applications**
  - **Swin-Transformer-Based Dynamic Semantic Communication for Multi-User With Different Computing Capacity**
  - An Efficient Federated Learning Framework for Training Semantic Communication Systems.
  - AI-Generated Content for SCM (AIGC-SCM)
- Demo of Generative AI Enabled Semantic Communication
- Conclusion and Future Direction

Current Work & Chanllenges

- Only consider one-to-one communication with various modalities such as image, text, and audio [1-6].



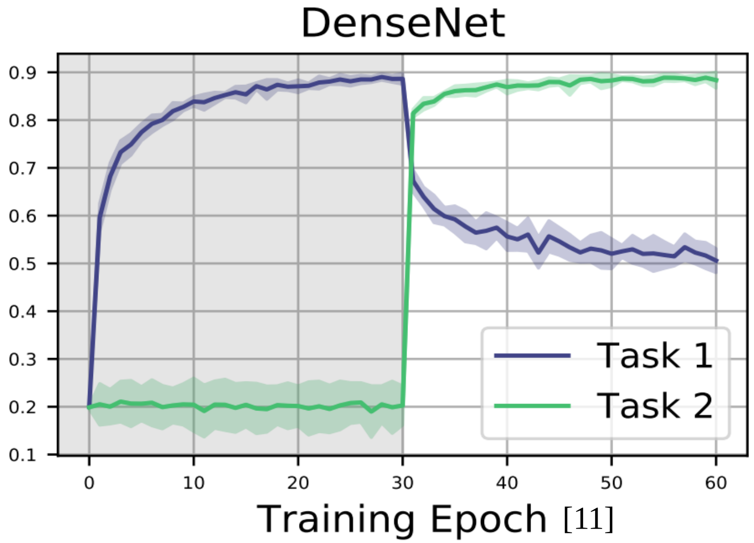
[8]

- The fixed output length of the encoder, regardless of the traffic condition, can be a waste of bandwidth resources if the traffic is in low demand.

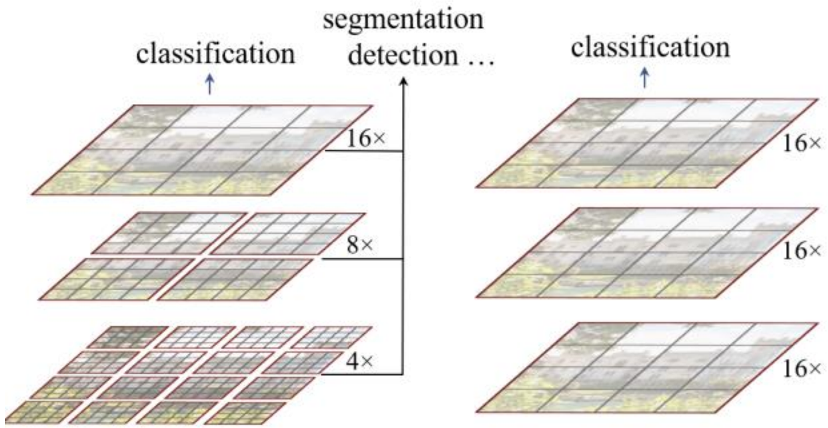
[1] N. Farsad, et al, "Deep learning for joint source-channel coding of text," in ICASSP. 2018.[2] X. Peng, et al "A robust deep learning enabled semantic communication system for text,"in GLOBECOM2022 [3] K. Yang, et al, "WITT: A wireless image transmission transformer for semantic communications," ICASSP, 2023. [4] E. Bourtsoulatze. Et al "Deep joint source-channel coding for wireless image transmission," IEEE Trans. Cogn.Commun. Netw., 2019.[5] E. Grassucci, et al "Diffusion models for audio semantic communication," in ICASSP, 2024.[6] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," IEEE J. Sel. Areas Commun.

Current Work & Chanllenges

- The training efficiency and catastrophic forgetting property of the Deep Learning network to serve multiple user-equipped different networks.



[11]



Swin-Transformer

ViT

- The absence of a centralized dataset for training semantic models.

- Vision Transformer (ViT) achieves higher performance compared to CNN network. However, ViT has quadratic computation complexity to input image size[1].

[1] Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

Swin Transformer-Based Dynamic Semantic Communication for Multi-User With Different Computing Capacity

- We employ variants of the Swin Transformer model to simulate the difference in computing capacity.
- Swin Transformer employ the self-attention computation to non-overlapping local windows and shifted window mechanism. It achieves linear computation complexity.
- We propose a dynamic compression module, increasing the length of the message when the network demand is low, and reducing in high demands.

To solve the problem of multiple users with different computing capacities in Semantic Communication (SemCom):

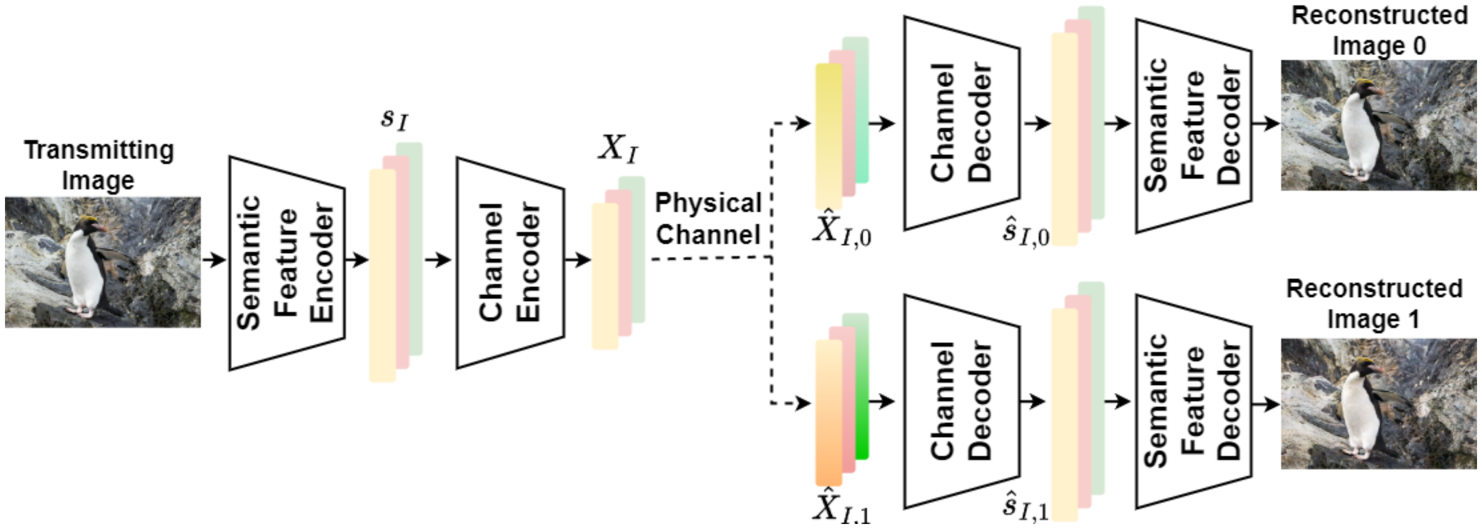
a) We propose a novel system model for BS Encoder to embed the signal accordingly to the receiver.

b) We design a new loss that consider the human visual quality instead of MSE error alone.

c) Depending on the network traffic, our model can adaptively change the compression rate of the signal.

Research Application I

- Current works[1-3] only consider one transmitter and one receiver in their proposed scenarios.



- We consider the downlink transmission from the Base Station (BS) to multiple users in SemCom.
- The proposed scenario is difficult due to the need to serve multiple users with different computing capacities.

[1] Xie, et al ," Deep Learning Enabled Semantic Communication Systems " IEEE Transactions on Signal Processing
[2] Han, et al,"Semantic-Preserved Communication System for Highly Efficient Speech Transmission," IEEE JSAC
[3] Zhang, et al "A unified multi-task semantic communication system for multimodal data, "IEEE TCOM,

Research Application I——System model

The input image is denoted as *I*, while E_α denote the source encoder of BS with the parameter set α.

$$\boldsymbol{s}_I = E_\alpha(I), \longrightarrow \boldsymbol{s}_{I,k} = E_\alpha(I|c_k),$$

We change the equation from a normal semantic encoder into a conditional semantic encoder, which embeds the receiver's index in the encoding process.

Similarly with the Channel Encoder:

$$X_I = C_\beta(\boldsymbol{s}_I) \in \mathbb{R}^k, \longrightarrow X_{I,k} = C_\beta(\boldsymbol{s}_{I,k}|g_k),$$

The received signal at the user k, under the Additive White Gaussian Noise (AWGN) is denoted as:

$$\hat{X}_{I,k} = Y_{I,k} = X_{I,k} + N_k,$$

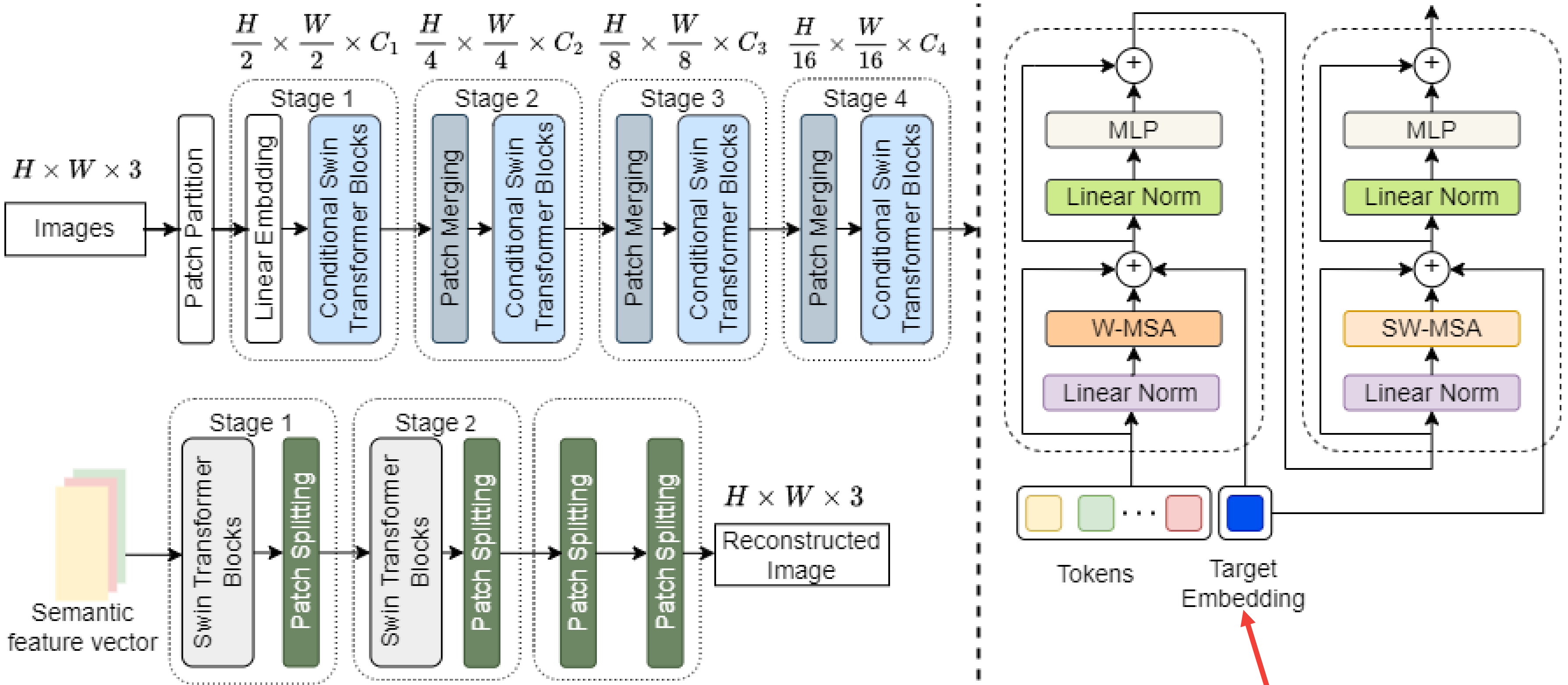
The signal can be transformed back to a similar form as:

$$\hat{X}_{I,k} = (H^H H)^{-1} H^H Y^{I,k} = X_{I,k} + \hat{N}_k,$$

Research Application I Proposed Solution

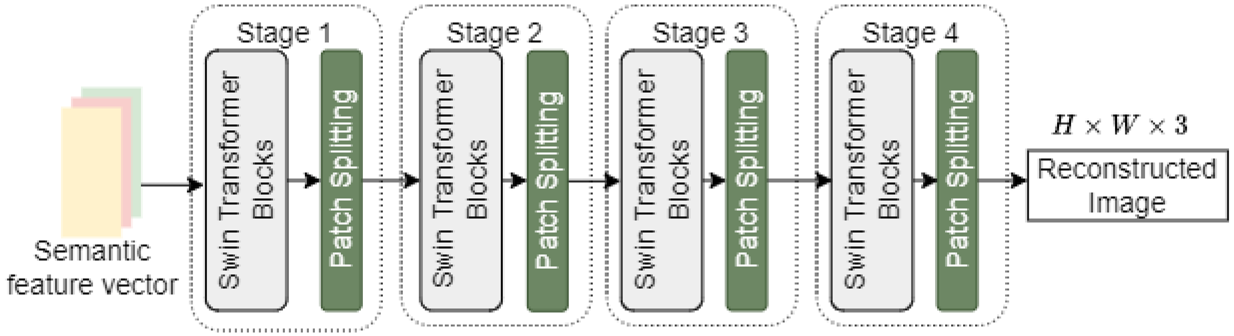
Source Encoder of the BS

Source Decoder of Low-computing User

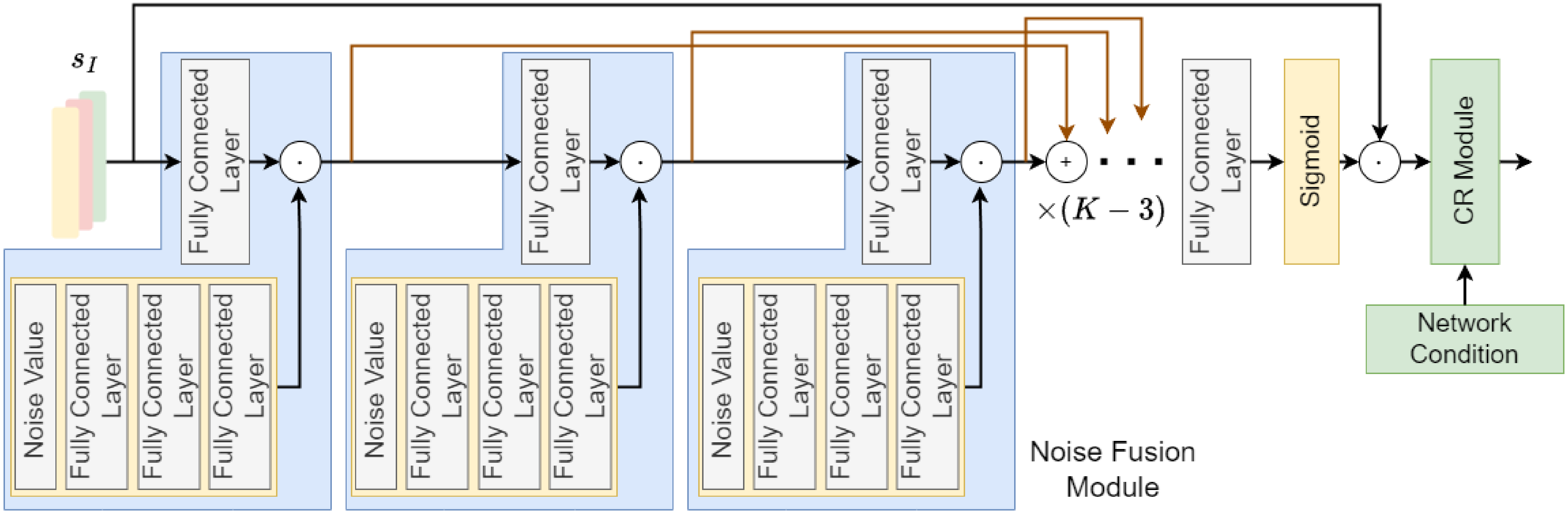


provides the information related to which receiver is being served.

Source Decoder of High-computing User



Research Application I Proposed Solution



The channel encoder is responsible for the compressed semantic features from the source encoder while considering both the noise value and the network condition.

- It is composed of K noise fusion modules and one compression rate (CR) module.
- The noise fusion module associates each semantic feature with a weight and computes the dot product between them and output from weighted noise values.
- Using the SNR information, we can derive a set of potential noise values.

Research Application I Proposed Solution

We propose a hybrid loss, which combines different loss functions:

$$\mathcal{L}^{\text{Hyb}} = \gamma \cdot \mathcal{L}^{\text{MS-SSIM}} + (1-\gamma) \cdot \mathcal{L}^{\ell_1} + \epsilon \cdot \mathcal{L}^{\ell_2},$$

- Multi-scale structural similarity index measure (MS-SSIM).
- Mean absolute error (MAE).
- Mean Square Error (MSE).
- γ *and* ϵ are coefficients of the losses to prevent any loss from dominating the objective.

With this objective, our system is able to reconstruct images with improved perceptual details compared to those obtained using the MSE loss.

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

$$\mathcal{L}^{\text{MS-SSIM}} = 1 - \prod_{j=1}^{M} \text{SSIM}_j(x,y),$$

Research Application I——Experiment Result

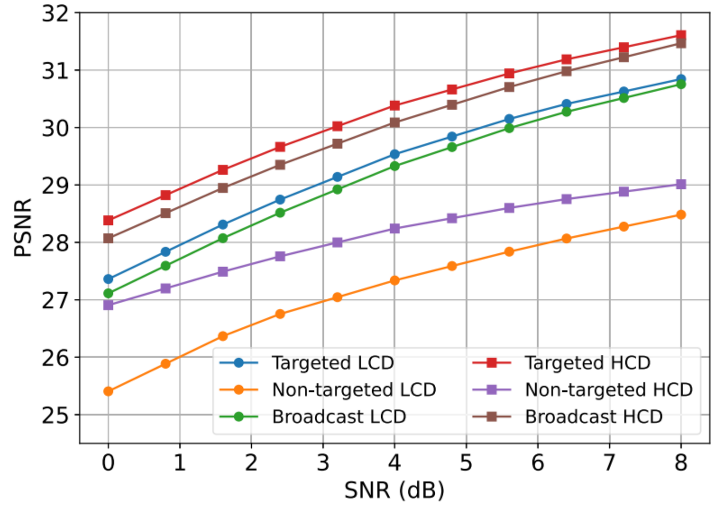


Fig. 5. PSNR results of high-computing and low-computing decoder in three scenarios: 1) the targeted decoder, 2) non-targeted decoder 3) the broadcasting case.

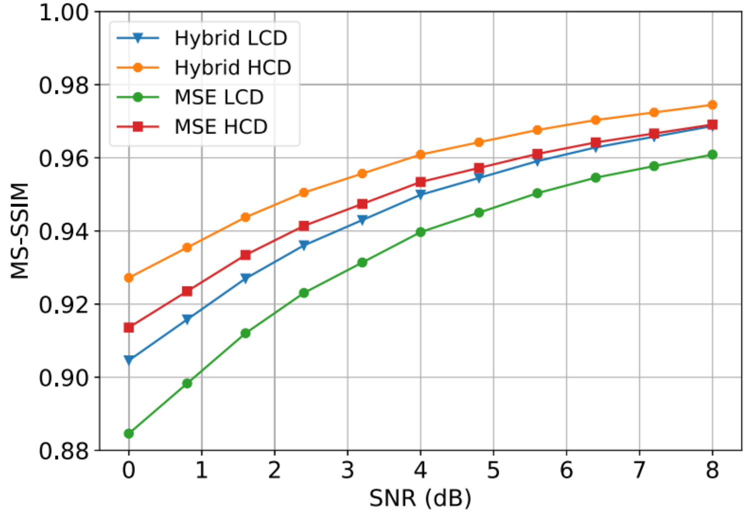


Fig. 6. Comparison in MS-SSIM between the network trained with the MSE and Hybrid losses.

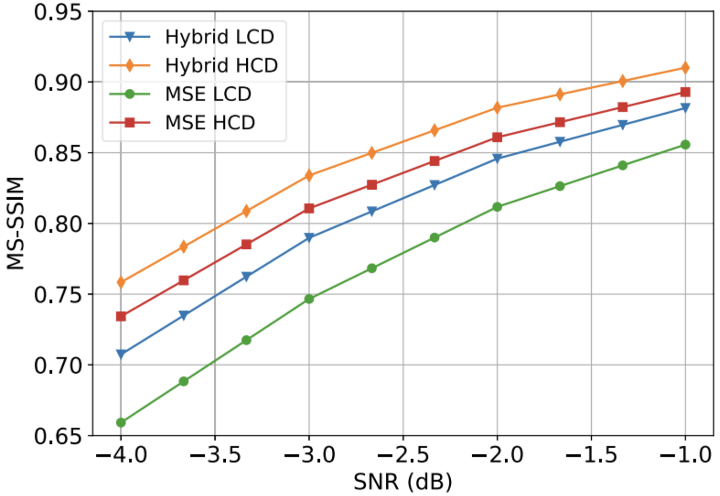


Fig. 7. Performance difference of two training losses under harsh and inexperienced physic channel conditions.

TABLE VII
PERFORMANCE CHANGE WITH DIFFERENT CR UNDER VARIOUS CHANNEL CONDITIONS

		LCD			HCD		
SNR	CR / Metric	3/64	4/64	5/64	3/64	4/64	5/64
0 dB	PSNR	26.6496 ± 0.0027	27.3811 ± 0.0044	27.8956 ± 0.0003	27.8727 ± 0.0025	28.6462 ± 0.0022	29.1580 ± 0.0040
	MS-SSIM	0.8890 ± 0.0001	0.9096 ± 0.0001	0.9223	0.9149 ± 0.0001	0.9318	0.9416
2 dB	PSNR	27.7758 ± 0.0042	28.3966 ± 0.0050	28.8191 ± 0.0014	29.0682 ± 0.0006	29.6853 ± 0.0026	30.0902 ± 0.0025
	MS-SSIM	0.9215 ± 0.0001	0.9348 ± 0.0001	0.9431 ± 0.0001	0.9411 ± 0.0001	0.9509	0.9567
4 dB	PSNR	28.7059 ± 0.0028	29.2291 ± 0.0042	29.5654 ± 0.0015	30.0122 ± 0.0011	30.5120 ± 0.0030	30.8431 ± 0.0019
	MS-SSIM	0.9416 ± 0.0001	0.9506 ± 0.0001	0.9560	0.9559	0.9624	0.9662
6 dB	PSNR	29.4620 ± 0.0014	29.8815 ± 0.0027	30.1331 ± 0.0009	30.7586 ± 0.0010	31.1535 ± 0.0024	31.4172 ± 0.0018
	MS-SSIM	0.9551	0.9609	0.9645	0.9656	0.9699	0.9724
8 dB	PSNR	30.0010 ± 0.0066	30.3356 ± 0.0092	30.5237 ± 0.0060	31.3133 ± 0.0030	31.6197 ± 0.0041	31.8254 ± 0.0018
	MS-SSIM	0.9634	0.9673	0.9697	0.9717	0.9746	0.9764

Outline

- Overview: Semantic Communications and GAI
- **Research Applications**
  - Swin-Transformer-Based Dynamic Semantic Communication for Multi-User With Different Computing Capacity
  - **An Efficient Federated Learning Framework for Training Semantic Communication Systems.**
  - AI-Generated Content for SCM (AIGC-SCM)
- Demo of Generative AI Enabled Semantic Communication
- Conclusion and Future Direction

An Efficient Federated Learning Framework for Training Semantic Communication Systems

- Limited works considered the data scattering problem for training semantic communication models, only [1] consider audio modality.
- Here, we not only provide an efficient FL algorithm to address the decentralized data issue but also reduce communication costs and achieve better performance.

To solve the data-driven problem of DL model in the SemCom system and the scattering property of data :

a) We leverage the Federated Learning (FL) algorithm to train the DL models in SemCom.

b) We propose a new approach for aggregating the global model, which is called FedLol, considering the image reconstruction task.

c) Improve the communication efficiency for FL by transmitting the model partially.

[1] H. Tong, et al “Federated learning based audio semantic communication over wireless networks,” in GLOBECOM, 2021.

Research Application II Proposed Solution

❑ Process of Federated Learning for SemCom:

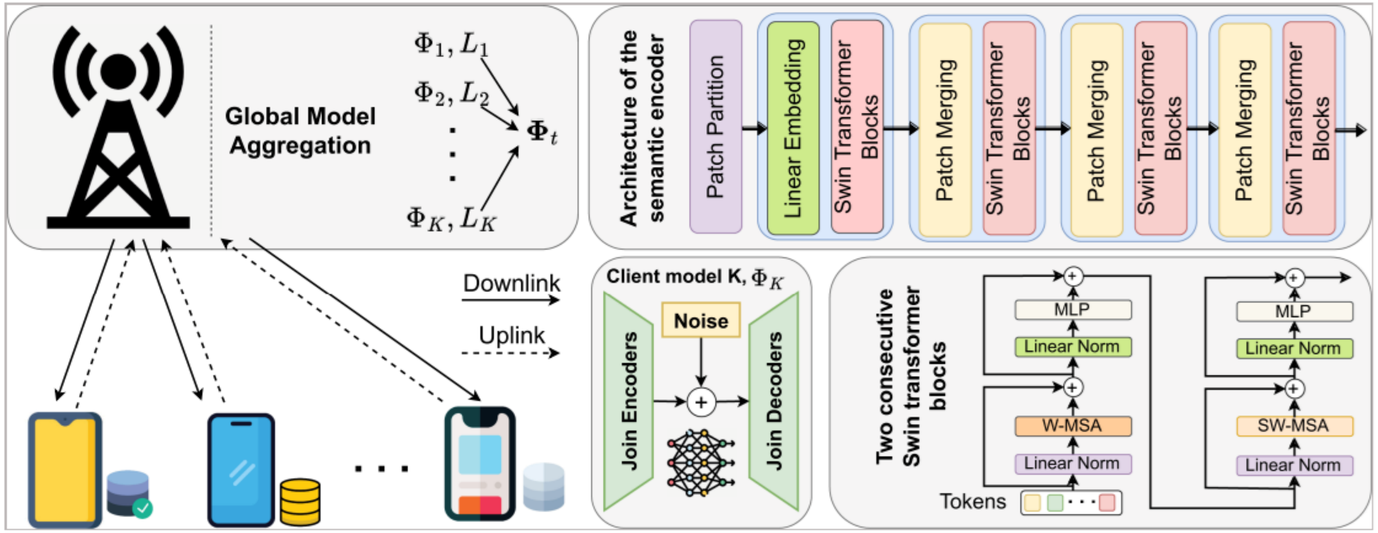
Step 1: Initializes learning models (E_{ψ}^{S} ; E_{β}^{C}; D_{γ}^{C} ; D_{σ}^{S}) at the BS and distributes them to all users.

Step 2: The user trains the model with its private data and sends back the updated models.

Step 3: The BS aggregates the global model based on the received model and continues Steps 2 & 3 until certain conditions are met.

❑ Proposal Description:

- ➢ Considering the image reconstruction task and its properties, we propose Federated Local Loss (FedLol), which determines each local model's contribution to the aggregation process based on its local loss.
- ➢ We transmitted the source encoder/decoder every global round, while the channel encoder/decoder is only updated after P global rounds. (P=5)



Research Application II Proposed Solution

I denotes the image, $\hat{I}$ denotes reconstructed image:

$$\hat{I} = D_{\sigma}^{S}(\hat{F}_I), \hat{I} \in R^{3\times H\times W}.$$

To make it simple, we use the most common loss for the task, which is MSE:

$$L(I,\hat{I}) = \mathrm{MSE}(I,\hat{I}).$$

The global model is aggregated as follows:

$$\mathbf{\Phi} = \sum_{k=1}^{K} \boldsymbol{\omega_k} \mathbf{\Phi}_k,$$

The value ω_K is calculated based on the loss of user *k*:

$$\boldsymbol{\omega_k} = \frac{1}{(K-1)} \frac{\sum_{k=1}^{K}(\mathrm{L}_k) - \mathrm{L}_k}{\sum_{k=1}^{K}(\mathrm{L}_k)},$$

Algorithm 1 Training Semantic Communication in an Efficient FL Framework: FedLol

```
Initialize: Global model Φ, number of global rounds T, local epochs R, update interval for channel encoder/decoder P.
for one global round t= 1, 2, ..., T do
    Check the current global round: if t % P == 1, send the whole model else send the semantic encoder and decoder model only.
    for each client k= 1, 2, 3..., K in parallel do
        Synchronize local model with the received model.
        while client epoch r < R do
            Train the model with local data.
            Φ_k^r ← Φ_k^{r−1} − η∇L_k.
        end while
        if t % P == 0 send the whole local model & L_k.
        else send semantic encoder/decoder models & L_k.
    end for
    Calculate w_k ∀k ∈ [1, K] as Eq. 10 and aggregating the global model with the calculated weights.
end for
Output: Global Model Φ.
```

Research Application II——Experiment Result

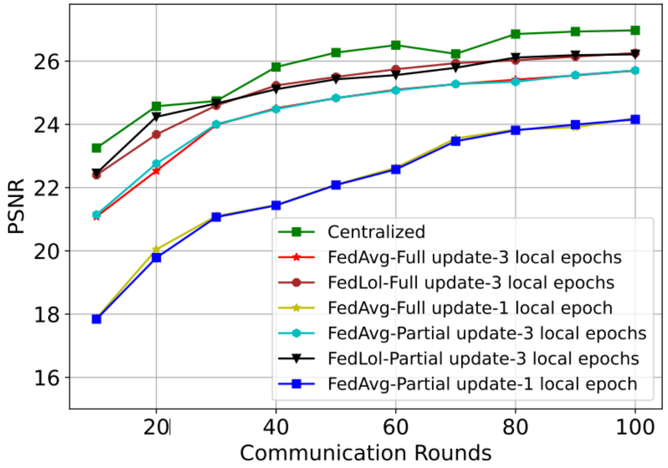


Fig. 2: The PSNR values of the proposed algorithm compared to other benchmarks, $\alpha = 1$.

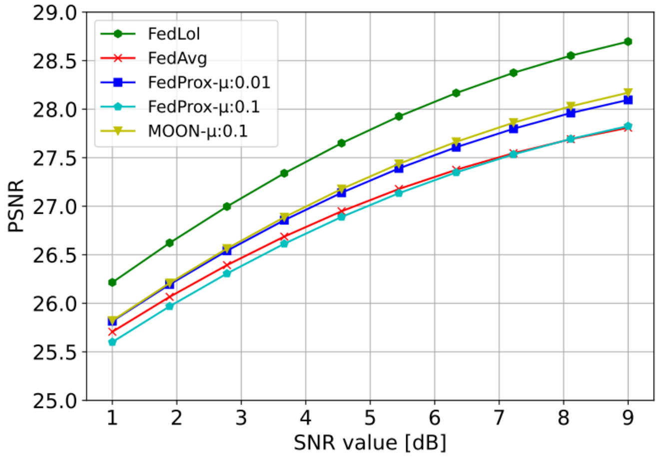


Fig. 3: The PSNR values of the proposed algorithm compared to other benchmarks, $\alpha = 1$.

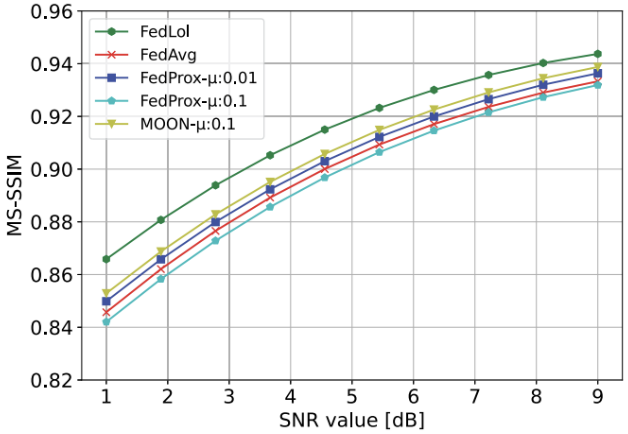


Fig. 4. The MS-SSIM values of the proposed algorithm compared to other benchmarks, $\alpha = 1$.

TABLE I
THE PSNR AND MS-SSIM RESULTS ACROSS DIVERSE NON-IID SCENARIOS

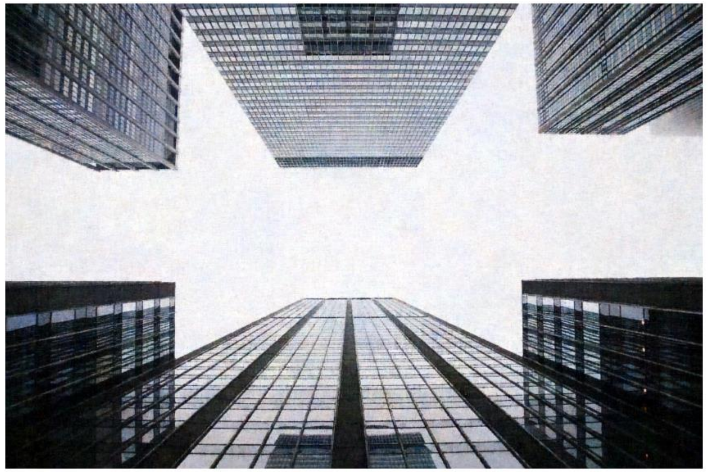
SNR = 1 dB	FedLol		MOON		FedProx		FedAvg		FedAvg (ADJSCC)	
α value	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM
α = 0.1	26.388	0.870	25.911	0.853	25.978	0.855	25.849	0.852	24.534	0.819
α = 1	26.215	0.866	25.822	0.853	25.815	0.850	25.707	0.846	24.332	0.809
α = 10	26.289	0.868	25.541	0.840	25.545	0.839	25.495	0.841	24.301	0.809
α = 10000 (IID)	26.217	0.865	25.614	0.842	25.562	0.840	25.574	0.841	24.358	0.809

Research Application II Experiment Results

(a) The original image

(b) The reconstructed image by centralized training 24.14; 0.903

(c) The reconstructed image by FedLol 23.078; 0.868

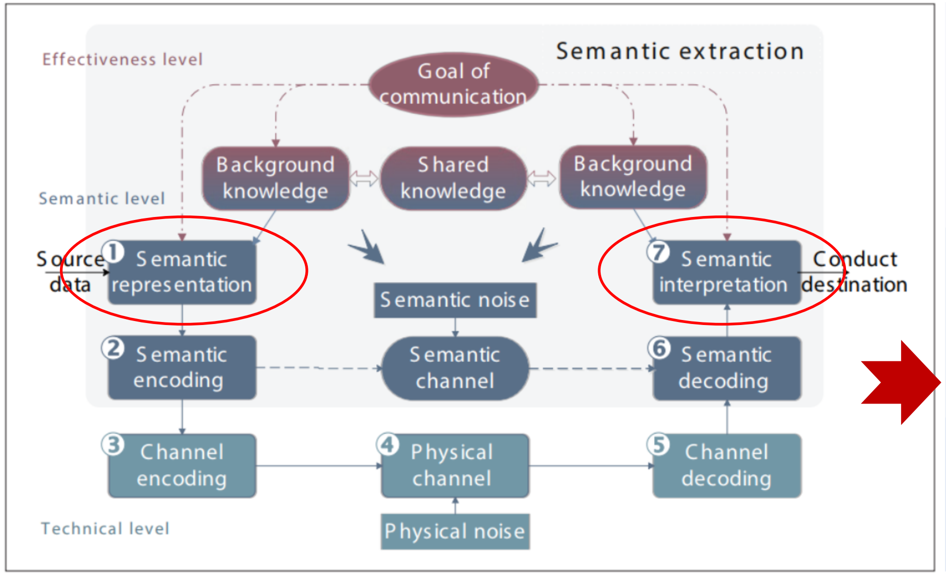


(d) The reconstructed image by MOON 22.534; 0.839

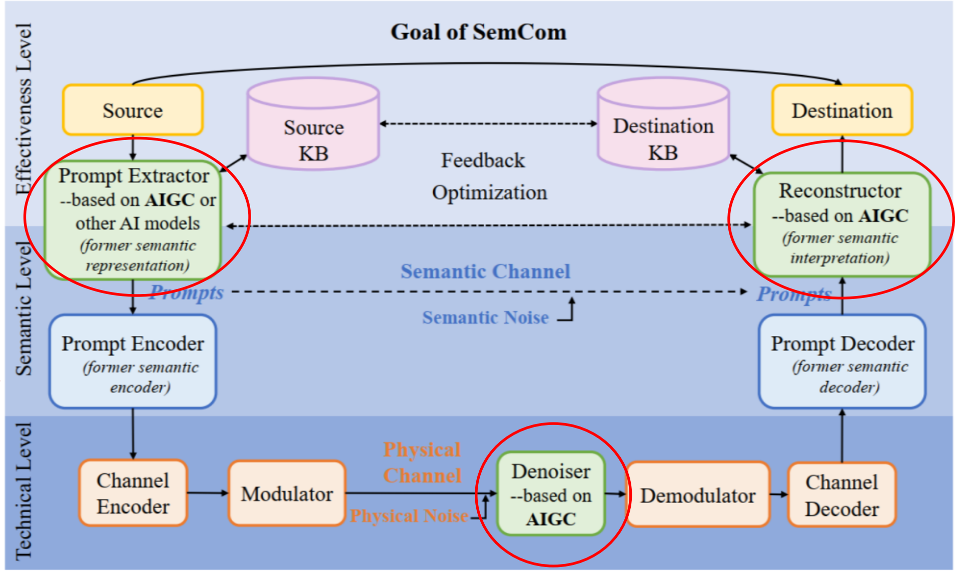
Outline

- **Overview: Semantic Communications and GAI**
- **Research Applications**
  - Swin-Transformer-Based Dynamic Semantic Communication for Multi-User With Different Computing Capacity
  - An Efficient Federated Learning Framework for Training Semantic Communication Systems.
  - **AI-Generated Content for SCM (AIGC-SCM)**
- **Demo of Generative AI Enabled Semantic Communication**
- **Conclusion and Future Direction**

Research Application III —— Motivation



The framework of conventional SCM.



The framework of AIGC-SCM.

- **Compared to DL-SCM, AIGC-SCM offers advantages in**

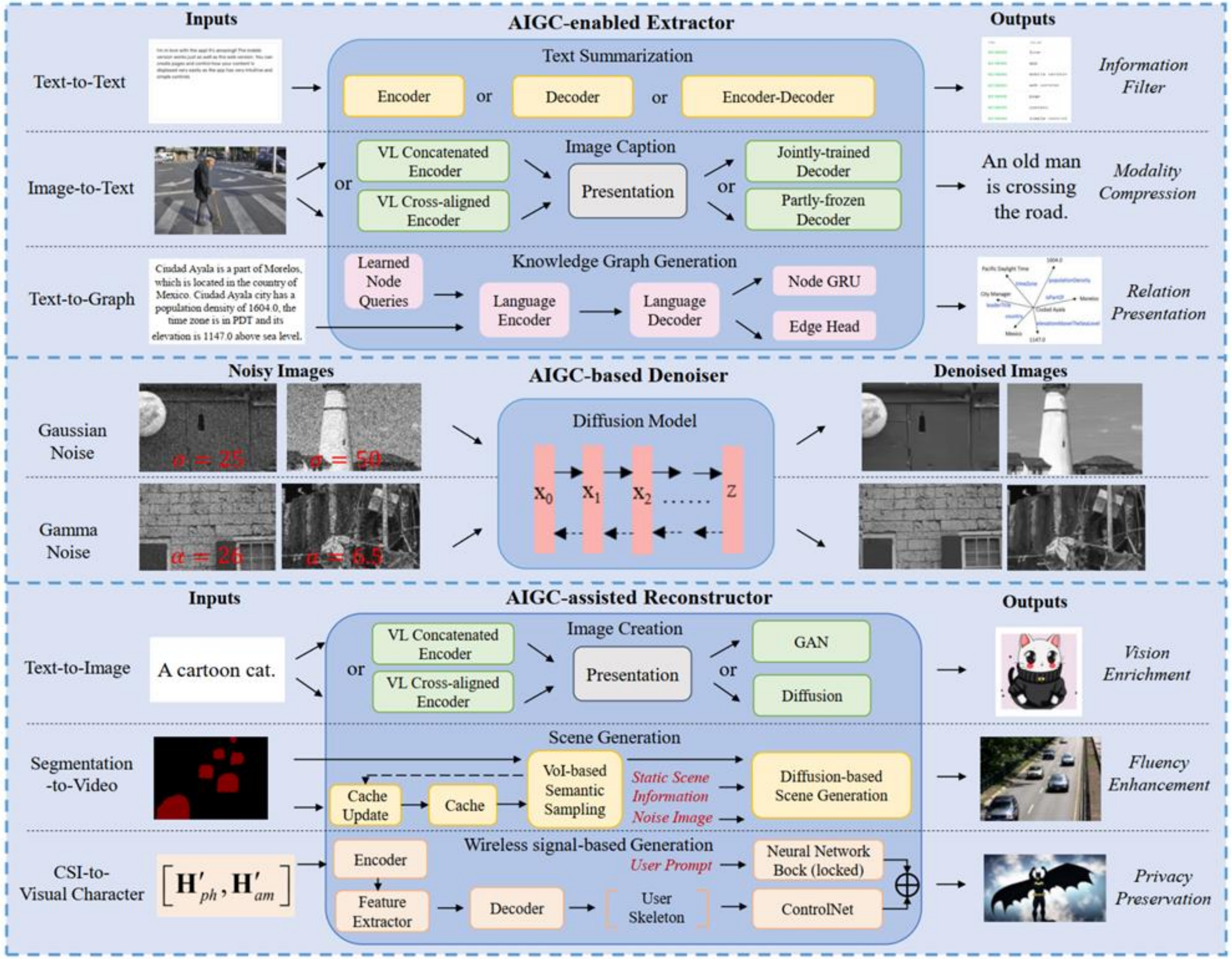
- ❑ **Easier Deployment**: It doesn't require joint codec training, making it less costly and easier to deploy than DL-SCM.
- ❑ **Broader Applicability**: Unlike DL-SCM, which demands differentiable loss functions, AIGC-SCM supports a wider range of loss function types.
- ❑ **High-fidelity Reconstruction**: Harnessing the generation capability of GAI, AIGC-SCM can reconstruct high-fidelity and semantically consistent content even when the transmitted data is highly compressed.

Research Application III——AIGC's functions

For prompt extractor: The goal of the prompt extractor is to distill semantic prompt from the source data.

For denoiser: Some AIGC models, such as diffusion models, demonstrate notable capabilities in denoising.

For reconstructor: It aims to enhance the receiver's quality of experience by generating high-dimensional semantic information based on low-dimensional prompts.



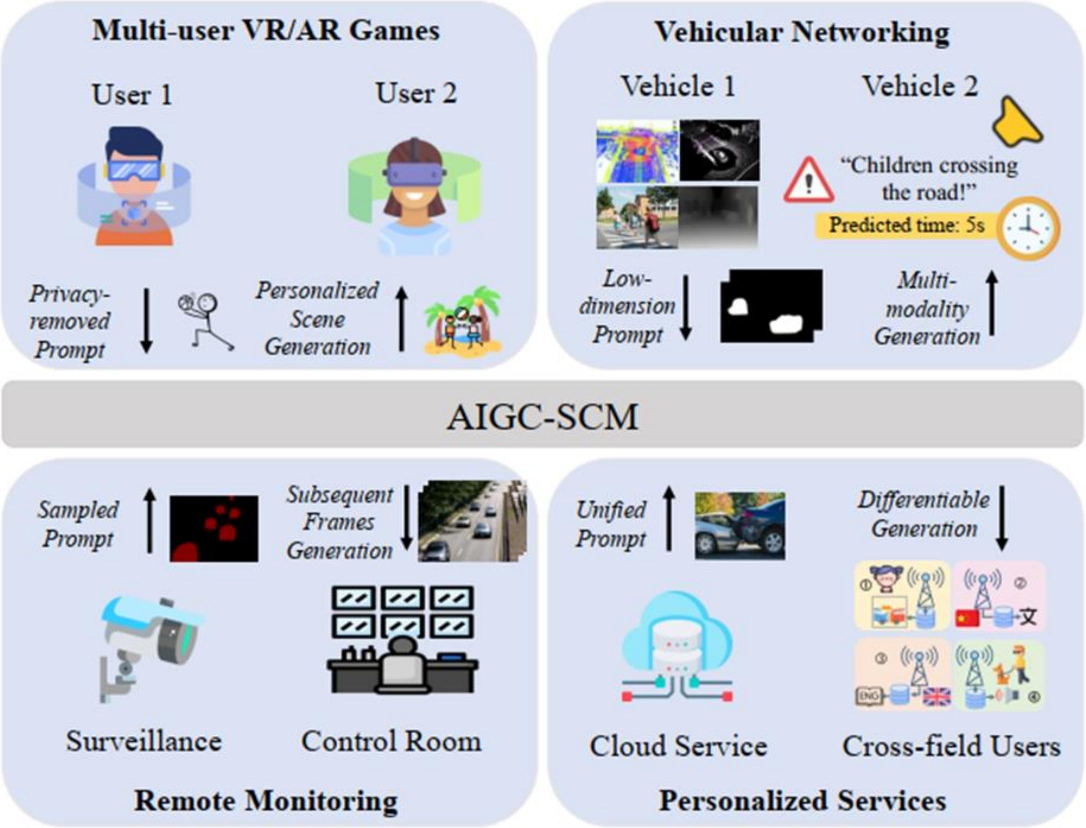
Research Application III Applications

Multi-user VR/AR Games

- AIGC-SCM can represent virtual scenes within the prompt and generate diverse and personalized content for different users.

Remote Monitoring

- In typical remote monitoring scenarios, AIGC can be applied to video prediction, thereby reducing energy consumption.



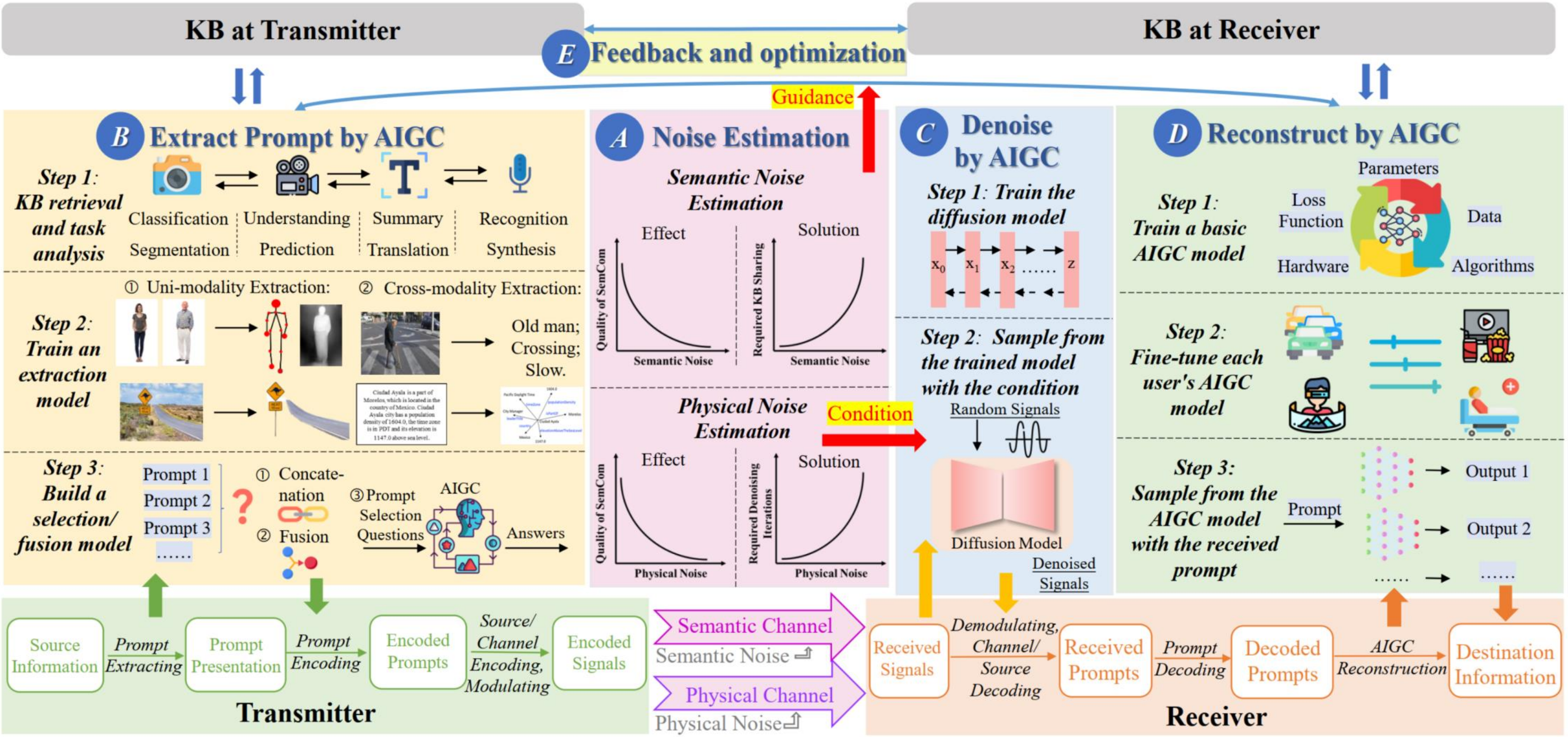
Vehicular Networking

- AIGC can be applied in prompt extraction and selection before transmission.
- At the receiver end, the reconstruction of road scenes is also possible in AIGC-SCM.

Personalized Services

- AIGC-SCM enables a single transmission to cater to multiple users with personalized needs, thereby significantly enhancing efficiency.

Research Application III —— Step of implement



The detailed presentation of AIGC-SCM, and some important procedures for implementing AIGC-SCM.

Research Application III Motivations

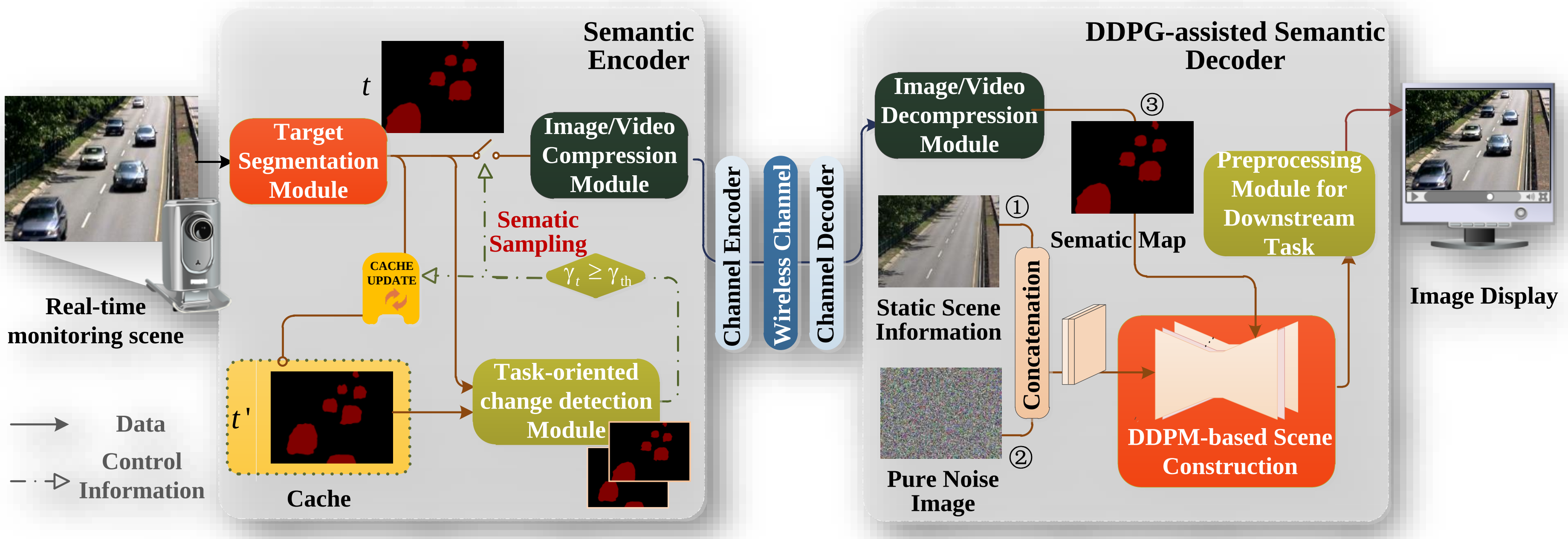
- **Motivations**

❑ **Real-time** remote monitoring challenges semantic compression for **long-term data** adopted in DL-based SemCom.

❑ Remote monitoring mainly considers **changes in the target object** and ignores dynamic background elements.

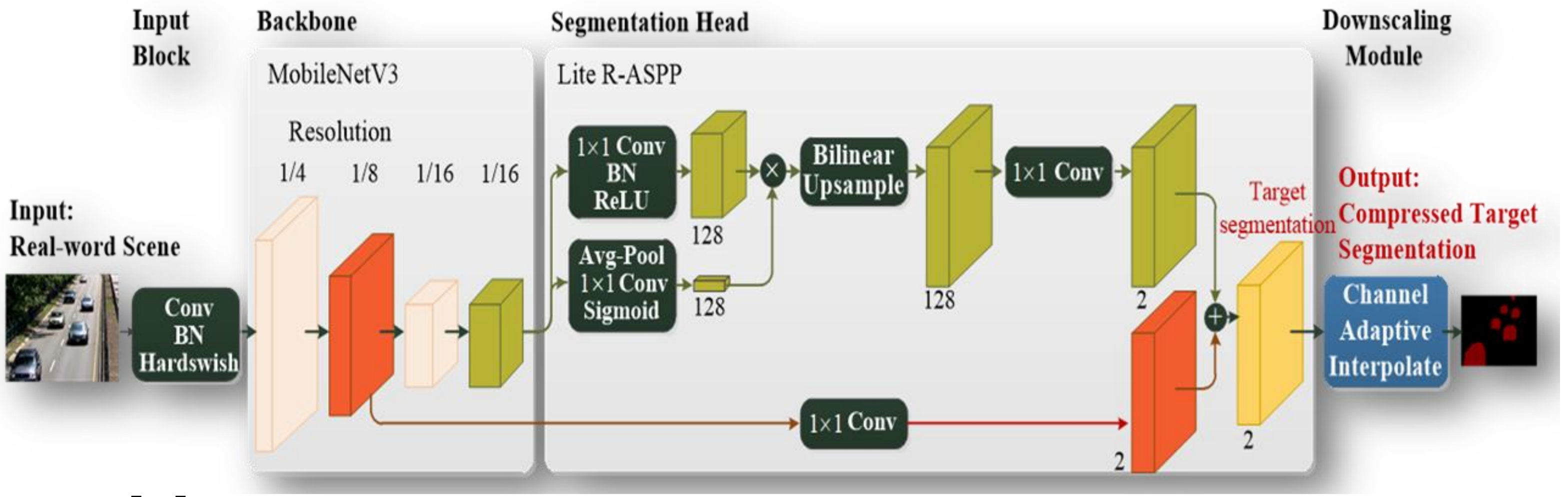
❑ DL-based visual SemCom transmission is difficult to recover **pixel-level** images due to the existence of **error floor**.

- **A change-driven modular SCM framework with semantic sampling based on Diffusion model**



Research Application III Proposed Solution

- **Target Segmentation Module**



- **VoI-based Sampling Module**

Age of information:

$$\gamma_t^{\mathrm{AoI}} = t' - t$$

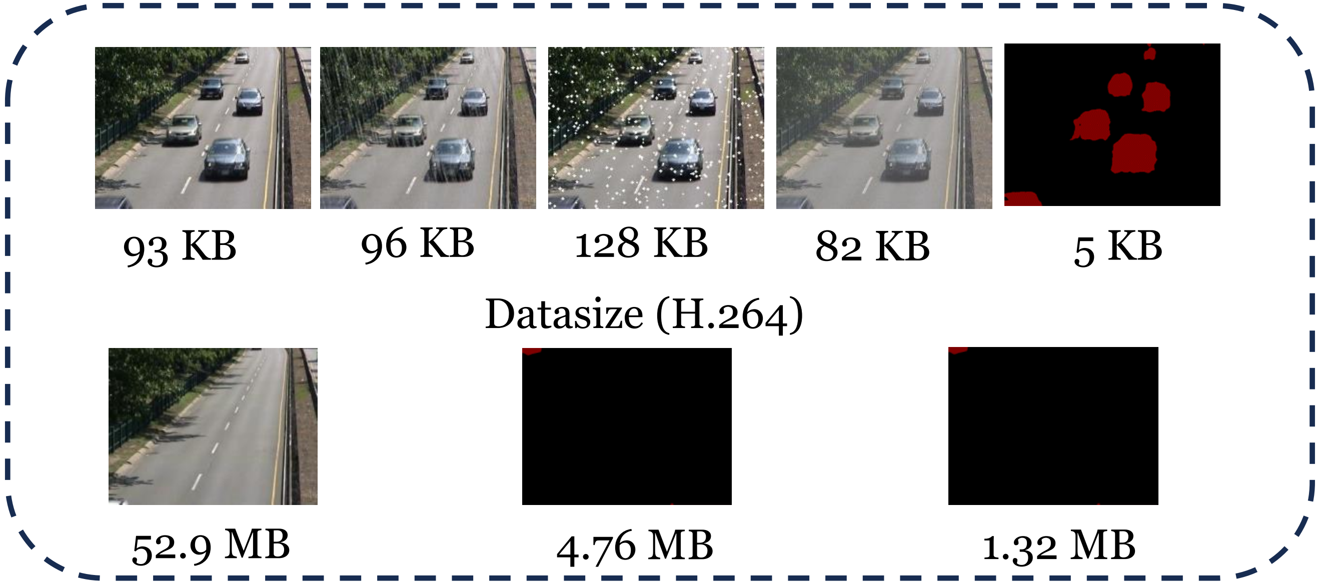
Semantic change degree:

$$\gamma_t^{\mathrm{change}} = \frac{n_t + n_{t'} - 2n_{tt'}}{n_t + n_{t'}},$$

where $\gamma_t^{\mathrm{change}} \in (0,1)$.

Value of Information:

$$\gamma_t = \tau_1\, \gamma_t^{\mathrm{AoI}} + \tau_2\, \gamma_t^{\mathrm{change}}$$



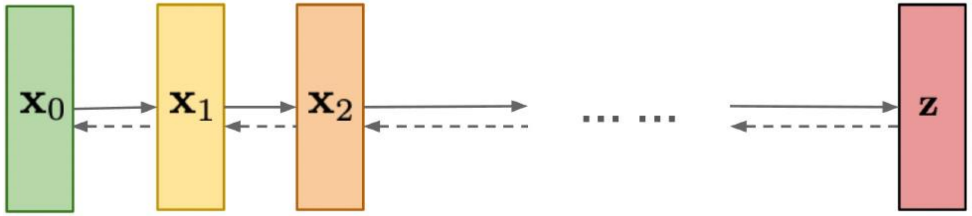
Research Application III Proposed Solution

- **DDPM-based scene construction**

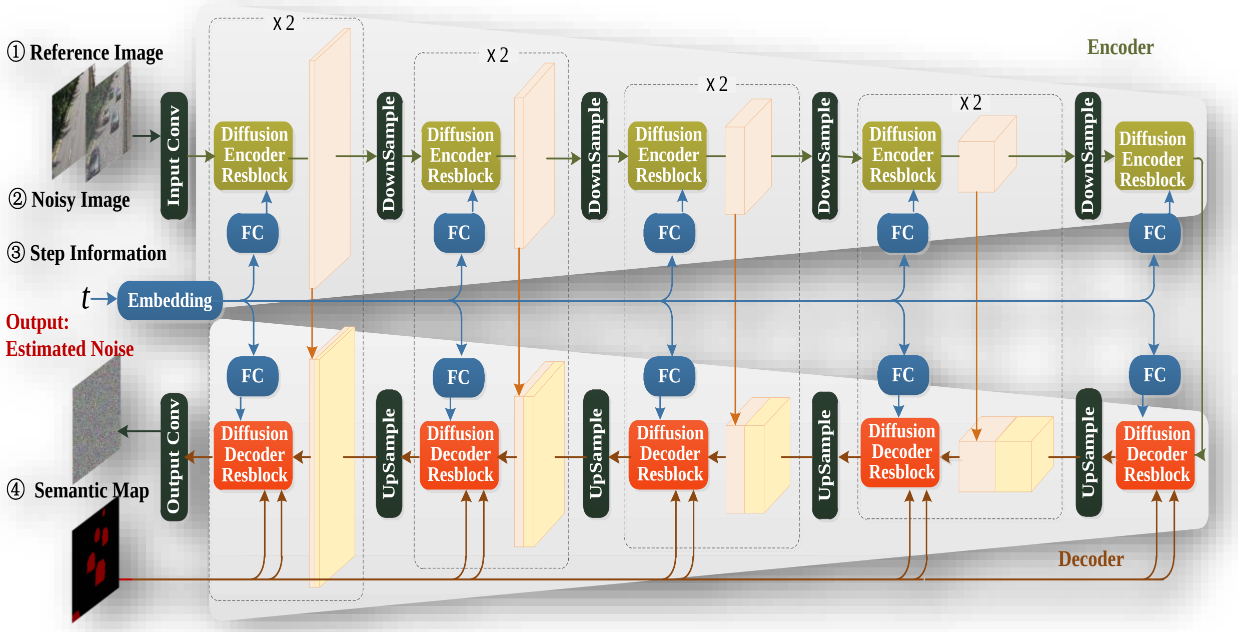
  Input of the DDPM-based scene construction:

  - Reference Image (Receiver's local information).

  - Semantic Map (Information sent by transmitter).



Gradually add noise and then reverse.



Real-world scene

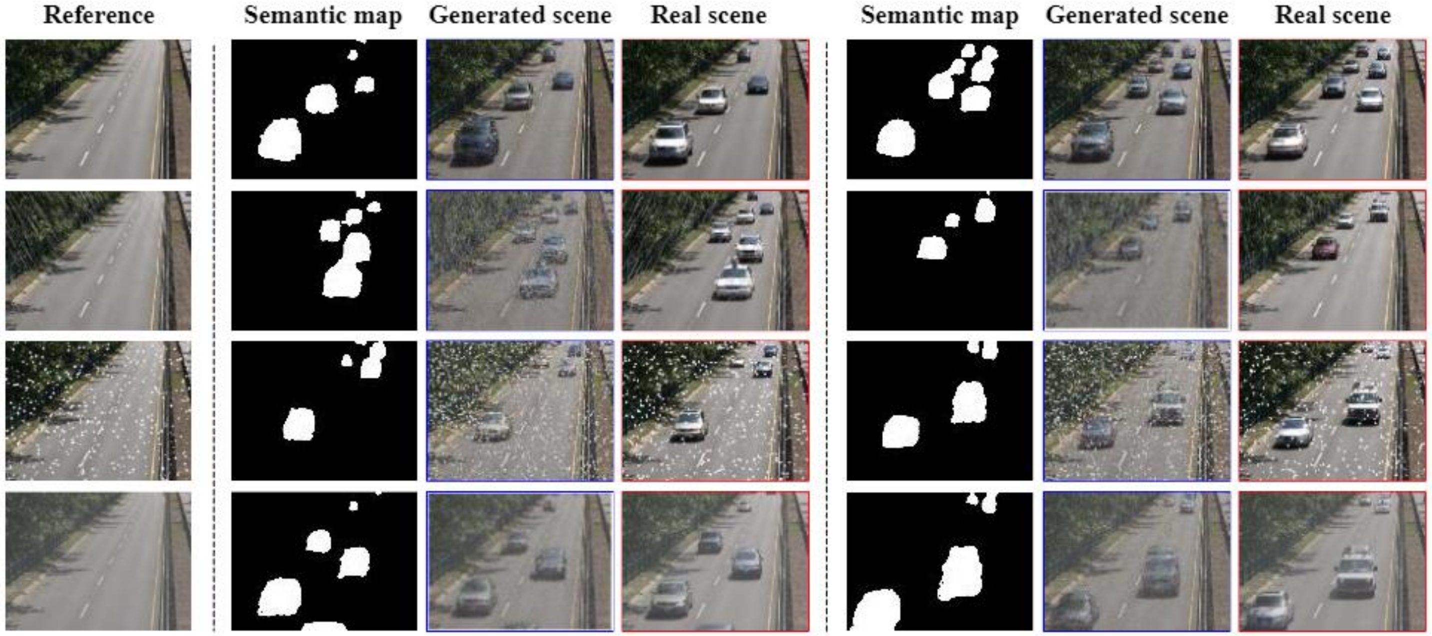
Input

x_T

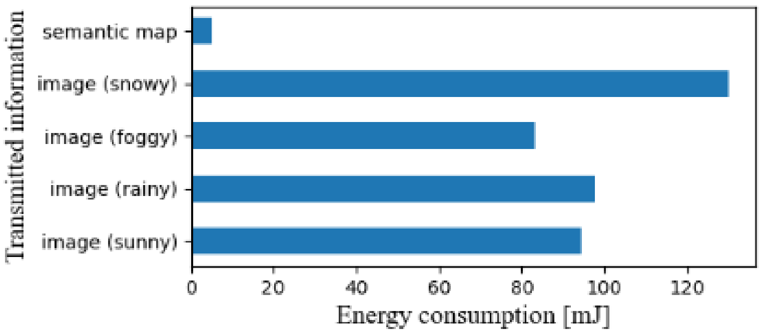
$\hat{x}_0$

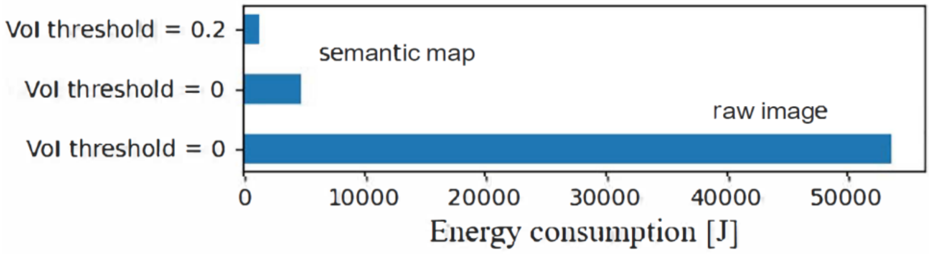
Research Application III Results

- Visual simulation results under different weather



- Comparison of energy consumption in different weather





Outline

- **Overview: Semantic Communications and GAI**
- **Research Applications**
  - Swin-Transformer-Based Dynamic Semantic Communication for Multi-User With Different Computing Capacity
  - An Efficient Federated Learning Framework for Training Semantic Communication Systems.
  - AI-Generated Content for SCM (AIGC-SCM)
- **Demo of Generative AI Enabled Semantic Communication**
- **Conclusion and Future Direction**

Demo


IEEE GlobeCom' 24 Demo
Generative AI Enabled Semantic Communication
Yinhuan Huang[1], Yun Tian[2], Weilong Chen[3], Faheem Quazi[4],
Zhijin Qin[1], Xiaoming Tao[1], Yanru Zhang[3], Yulong Feng[5], and Zhu Han[4]
[1]Tsinghua University, Beijing, China
[2]Peking University, Beijing, China
[3]University of Electronic Science and Technology of China, China
[4]The University of Houston, Houston, TX, USA
[5]ZTE Corporation, Shenzhen, China

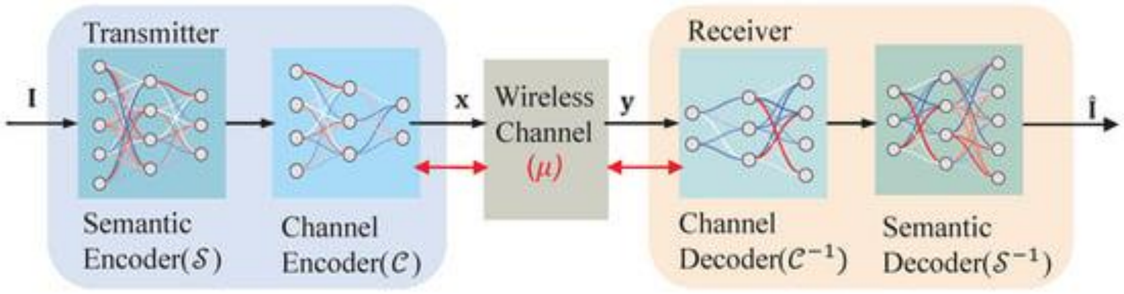
Future Work

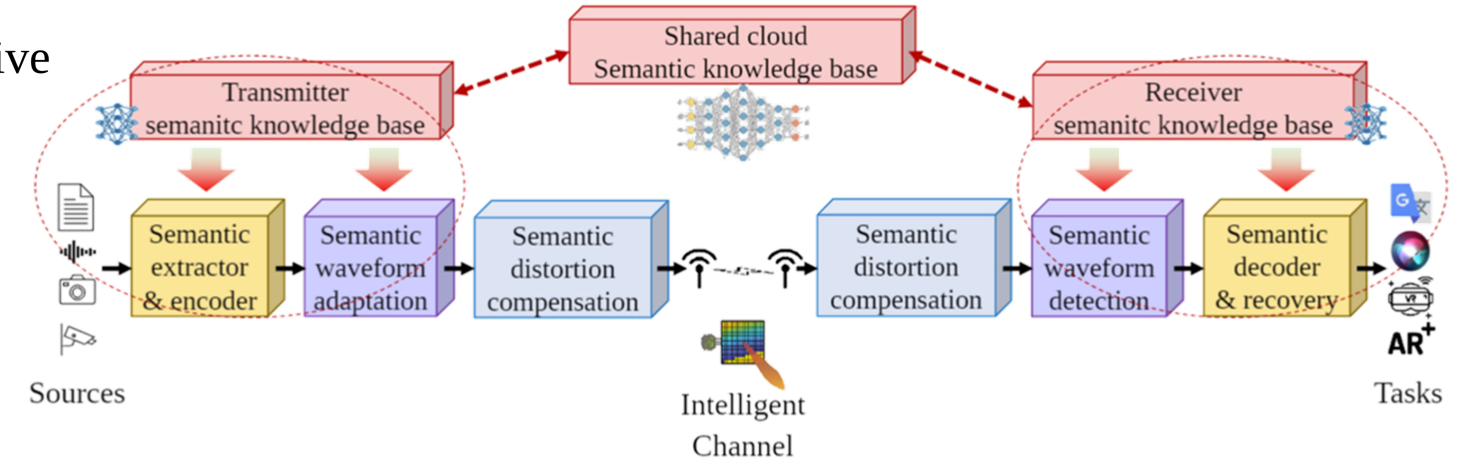
Research Challenges

- **Knowledge base**
  - General multi-scale knowledge base
  - Knowledge base update and synchronization

- **Semantic-channel coding**
  - Semantic representation
  - General transceiver structure
  - Performance metrics: subjective + objective

- **Semantic information transmission**
- **Semantic-aware network**



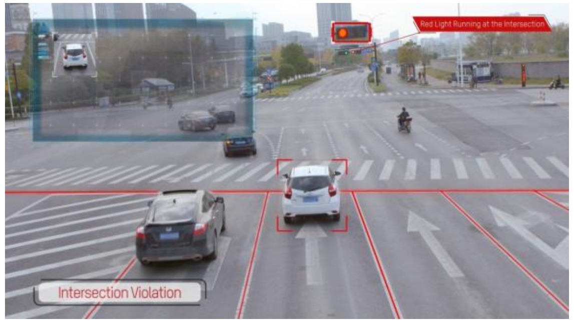


Future Work

AIGC-SCM assisted by knowledge bases in remote monitoring scenario

- **Motivations**

In certain scenarios, we may desire that the information reconstructed by AIGC-SCM at the receiving end **closely resembles the original object** as much as possible. For instance, in the Internet of Vehicles, it's preferred that the images of traffic violations captured by monitoring systems depict **the actual offending vehicle**, rather than a generated image of a different vehicle. In such cases, we can utilize a prior knowledge base to achieve this.



AIGC-SCM for collaborative perception and traffic prediction

- **Motivations**

- When a driver is approaching an intersection, receiving a forecast of the imminent road conditions becomes a piece of semantic information of paramount importance.
- The literature on collaborative perception currently considers overly simplistic channels. In reality, the channels in vehicle-to-vehicle (V2V) communication are quite complex. Signals in the V2X communication process are easily distorted by obstacles and interference.
- The data used in the V2X network is unreadable and complex. But we need to send human-friendly alerts to the drivers.

Future Work

Other modality data: wireless sensing

- For the virtual interactive game in Metaverse, GAI can generate avatars and create the corresponding scenarios according to users' requirements, thereby constructing a complete virtual world for users to explore.

- Typically, we use cameras, such as Kinect, to capture the user's image, which is then combined with the user's requirements and fed into the AIGC model to generate digital content. Nevertheless, prolonged use of the camera may raise privacy concerns even though SemCom is considered.

"Guiding AI-Generated Digital Content with Wireless Perception", IEEE Wireless Communications, 2024.

Conclusion

Generative AI enabled Semantic Communication

- ✓ Limited by Channel Capacity
- ✓ AI technique (e.g. autoencoder) does not consider channel
- ✓ Transmit info to prevent GAI hallucination at the receiver
- ✓ Subjective QoS and hard to theoretical analysis
- ✓ Many Implementation issues

Application I :

- ✓ Multi-user Scenario with Varying Computing Capacities
- ✓ Targeted Embedding Vector & Hybrid Loss
- ✓ Dynamic Channel Encoder & Loss Functions

Application II :

- ✓ Federated Learning for Semantic Communication
- ✓ Reduced Communication Overhead:

Application III :

- ✓ AIGC-SCM Architecture
- ✓ Application to Multi-Modal Systems

Question?

THANK YOU

Department of Electrical
and Computer Engineering
Cullen College of Engineering