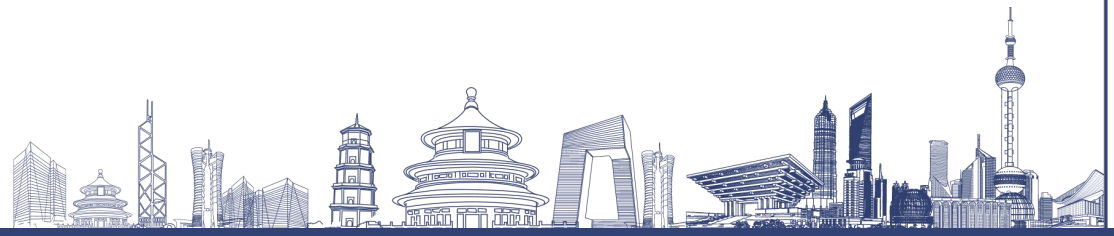
RepObE: Representation Learning-Enhanced Obfuscation Encryption Modular Semantic Task Framework

✪Liemi Lin*, Jinpeng Xu, Xiaoding Wang, Liang Chen, Sun-Yuan Hsieh, Jie Wu

34th International Joint Conference on Artificial Intelligence (IJCAI 2025)

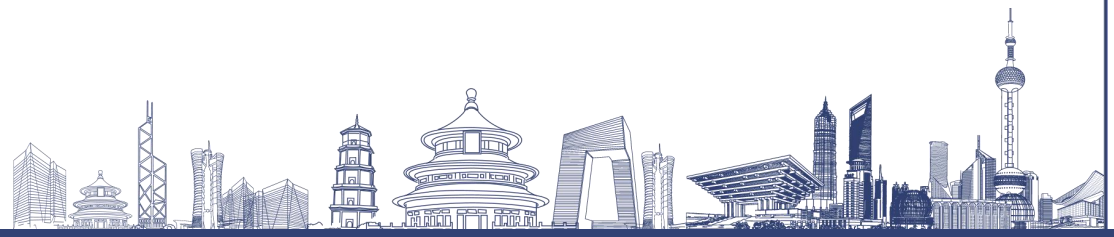
CONTENTS



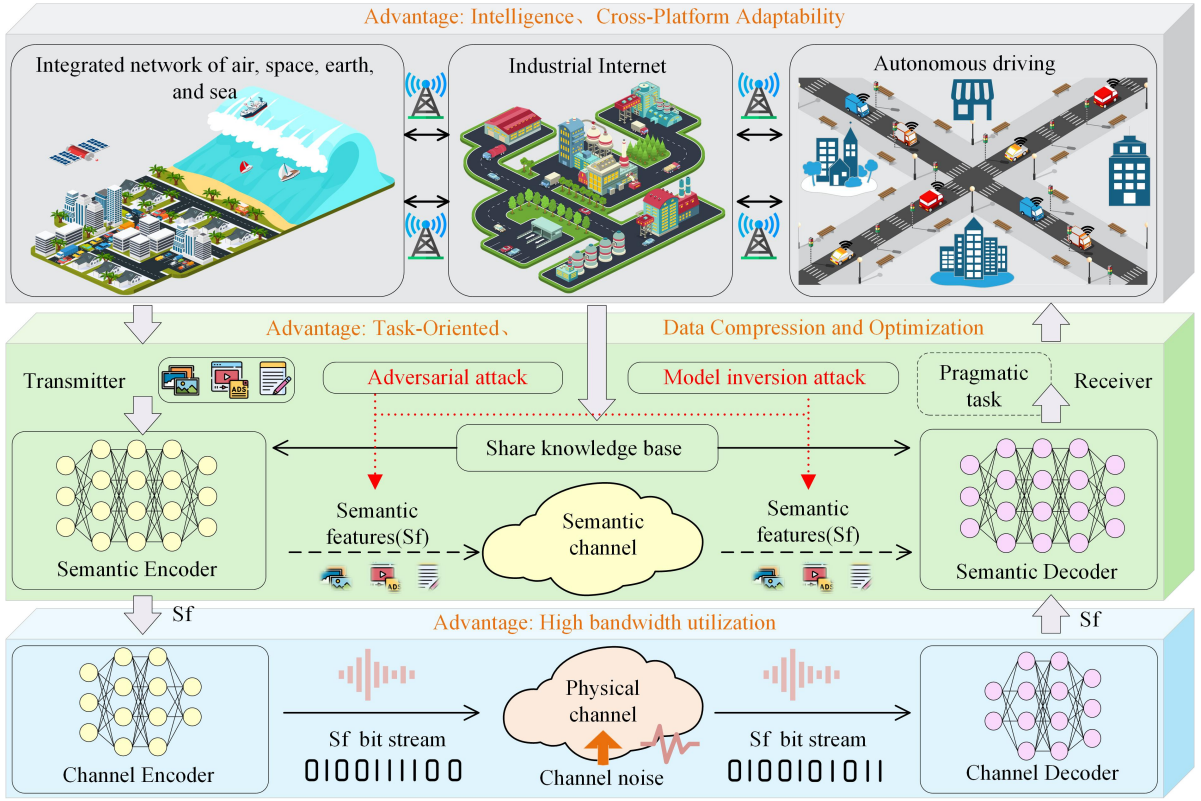
Background & Motivation

1

ONE



Semantic Communication: towards a new paradigm for intelligent communication



Communication Challenges in the 6G Era:

- High bandwidth, low latency, and dramatic increase in demand for intelligent processing.
- Traditional communication only transmits bits, low efficiency and high redundancy.

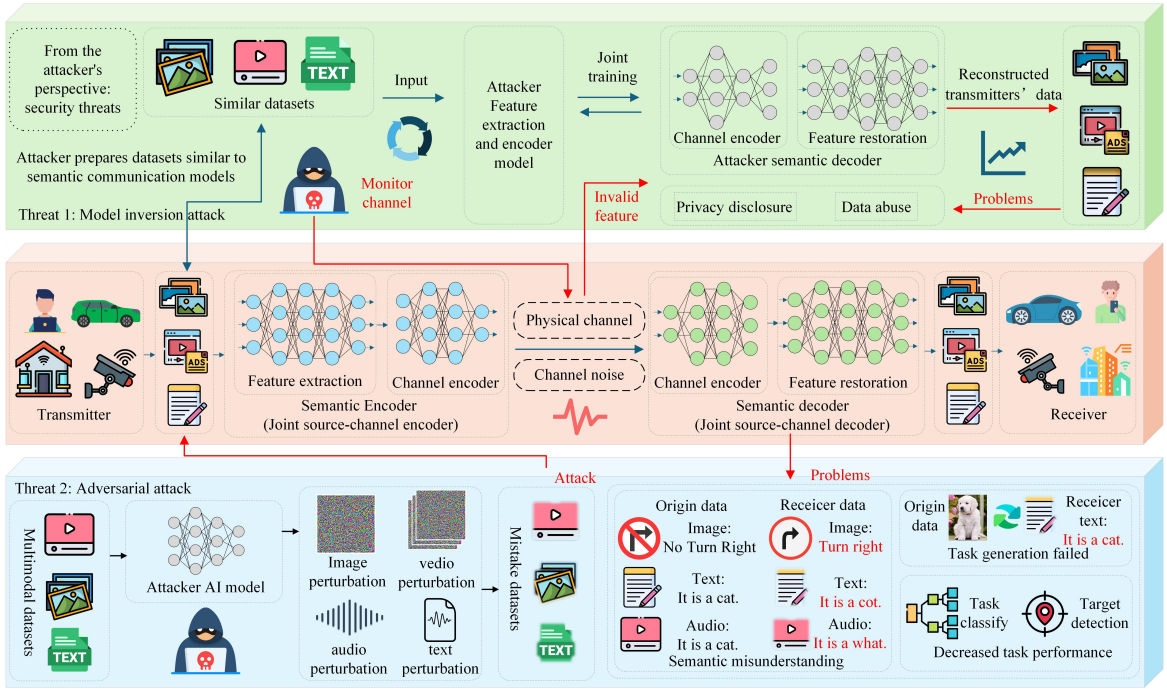
Core concept of Semantic Communication:

- Emphasize the delivery of “meaning” rather than “raw data”.
- Improve communication efficiency and reduce redundancy load.

Typical Application Scenarios:

Industrial Internet, Drones, Digital twin, Telematics and so on.

Motivation



New types of security threats:

- Deeper data processing, resulting in a more complex risk of privacy leakage.
- Information extraction process exposed to attack surface.

Limitations of existing approaches:

- Most focus on "data privacy" (e.g., federated learning, perturbation).
- Ignore "model security" and "feature space attacks".

Motivation for this research:

There is an urgent need for a semantic communication scheme that can simultaneously protect the privacy of semantic features and enhance the robustness of the system.

Key Contributions of the Paper

- **RepObE Framework:**
  Dynamic encryption during semantic extraction and transmission for strong privacy protection.

- **Adversarial Robustness:**
  Prototype-based adversarial training improves system resilience.

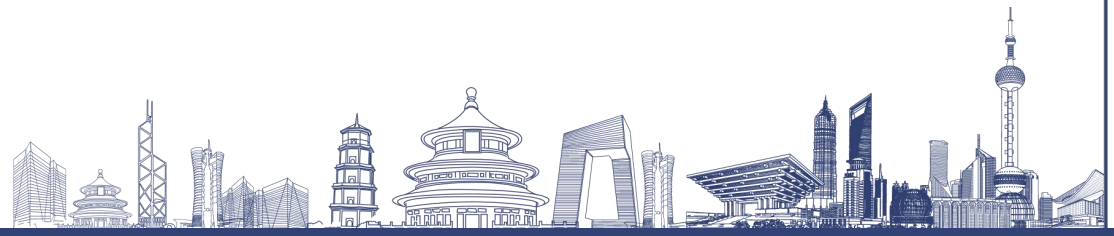
- **Reliable Communication:**
  Fusion of obfuscation and robust learning ensures stable semantic communication.

- **Experimental Results:**
  >2% gain against model inversion (MNIST), 3–5% gain against adversarial attacks (CIFAR-10).

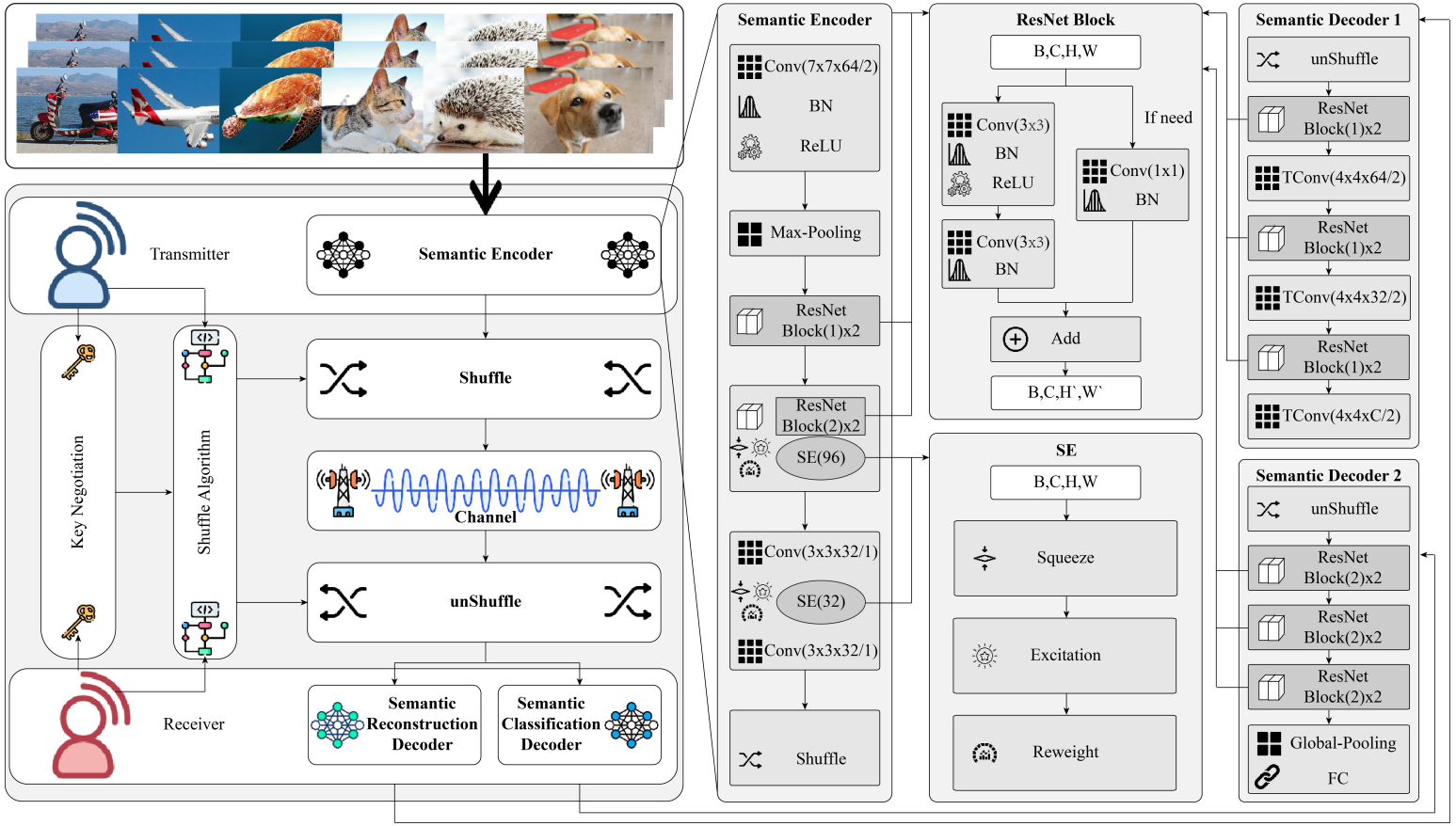
RepObE Framework Design

2

TWO



Overview of the RepObE Framework

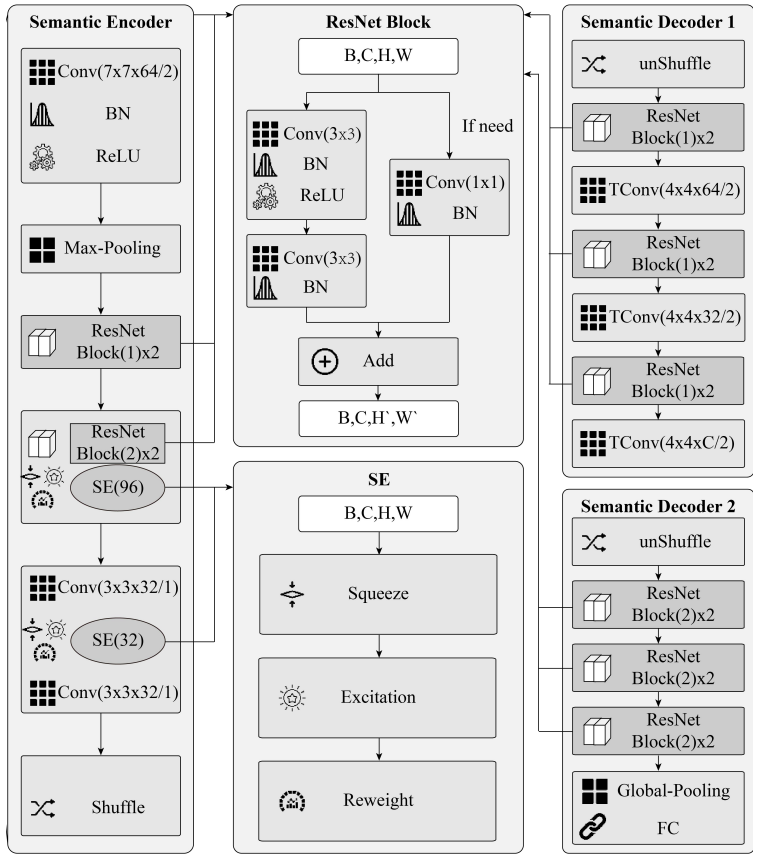


We propose RepObE, a framework combining semantic representation learning, dynamic encryption, and modular task processing.

It aims to enhance privacy and robustness in semantic communication.

The figure illustrates key components: semantic encoder, decoder, shuffle modules and secure transmission.

Semantic Encoder and Decoder Design



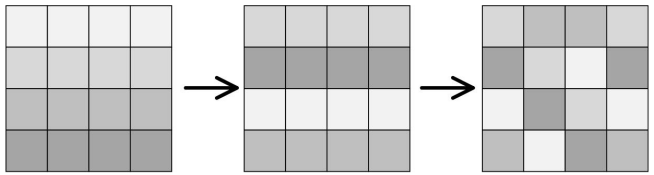
The encoder uses convolutional and residual blocks to extract semantic features, producing a $4 \times 28 \times 28$ feature map.

Two types of decoders are used:

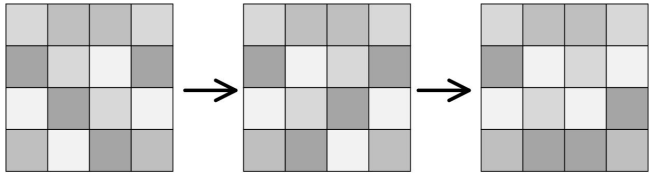
- The reconstruction decoder uses transposed convolutions to restore the original image.
- The classification decoder applies deep residual blocks and a fully connected layer for a 10-class prediction.

SE attention modules enhance the encoder's representational capacity.

Shuffle Module



Row confusion and offset



Column confusion and offset

The encryption consists of: row confusion, column offset, column confusion, and row offset.

The sender uses the key $K = (K_r, K_c, r, c)$ for 4-step obfuscation:

1. Row Confusion using K_r
2. Column Offsetbased on c
3. Column Confusion using K_c
4. Row Offset based on r

The final feature F_{en} is transmitted securely over the channel.

The receiver performs 4 decryption steps:

1. Inverse Row Offset
2. Inverse Column Confusion
3. Inverse Column Offset
4. Inverse Row Confusion

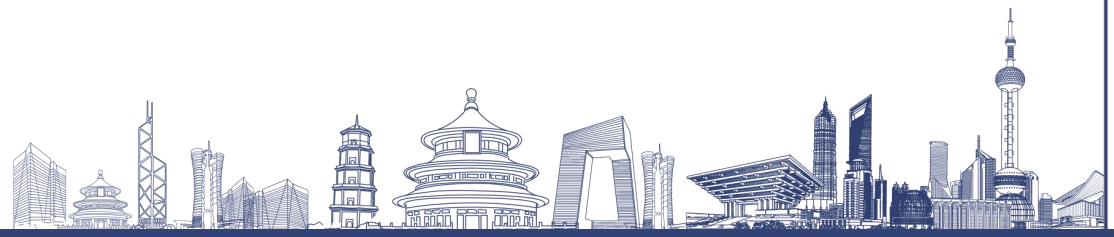
The recovered feature map F_{de} matches the original feature.

Total key space: $|K| = H' \cdot W' \cdot H \cdot W$

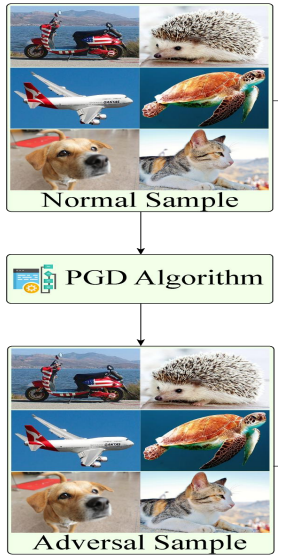
Adversarial Training

3

THREE



Adversarial Training Strategy for Robustness



Adversarial examples are generated using PGD to simulate worst-case perturbations.

Adversarial sample generation (maximizing KL divergence):

$$\max_{\|I_{adv} - I_{nat}\|} D_{KL}(M_\theta(I_{nat}) \| M_\theta(I_{adv}))$$

Training Targets:

$$\min_\theta \left[L_{CE}(I_{nat}) + \epsilon \cdot D_{KL}(I_{nat} \| I_{adv})\right]$$

Introduce contrastive learning loss to make I_{adv} and I_{nat} representations similar, while other samples repel each other.

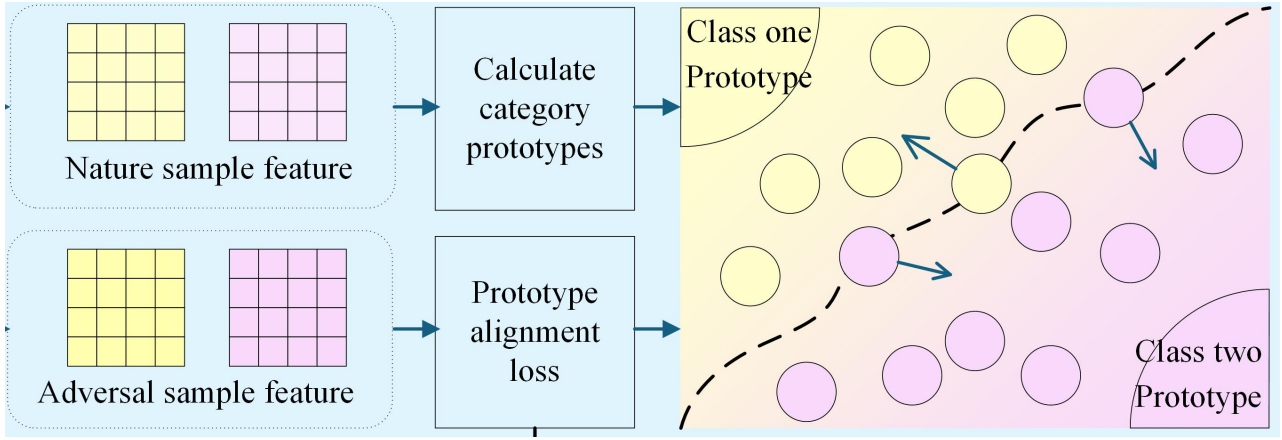
Comparative loss expression:

$$L_{contrast} = -\log \frac{exp(sim(z_{adv}, z_{nat})/\lambda)}{\sum_i exp(sim(z_{adv}, z_i)/\lambda)}$$

Final training loss combines classification, KL, and contrastive terms:

$$L_{class} = L_{contrast} + L_{CE} + \epsilon \cdot \widehat{\omega} \cdot D_{KL}$$

Prototype Alignment



Semantic prototypes p_y are computed per class from clean features.

Distance to prototype indicates adversarial difficulty.

Introducing weight function $w(I_{adv})$: The more difficult the sample, the greater the weight:

$$w(I_{adv}) = 1 - exp\left(-\frac{\|M_\theta(I_{adv}) - p_y\|}{\tau}\right)$$

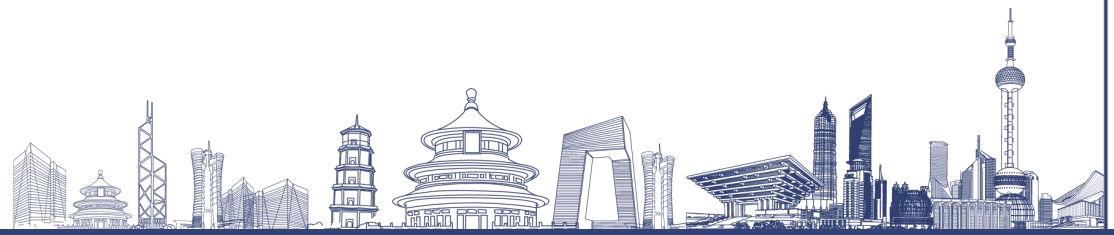
Normalizing:

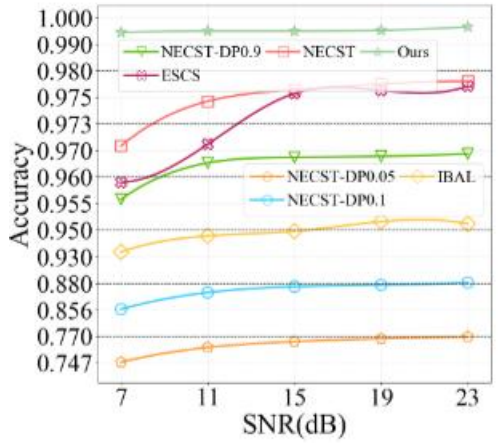
$$\widehat{w}(I_{adv}) = \frac{N \cdot w(I_{adv})}{\sum_{n=1}^{N} w(I_{adv})}$$

Align adversarial samples with their corresponding class prototypes to reduce offset differences.

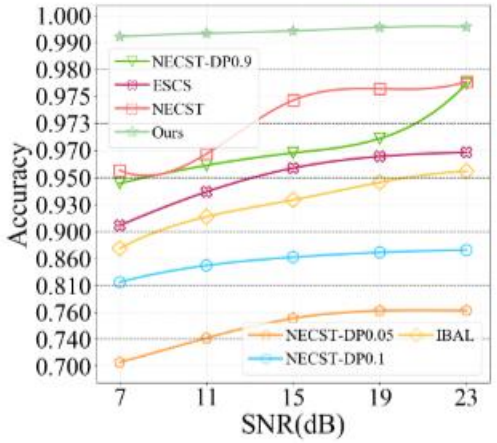
Experiments & Results

FOUR
4

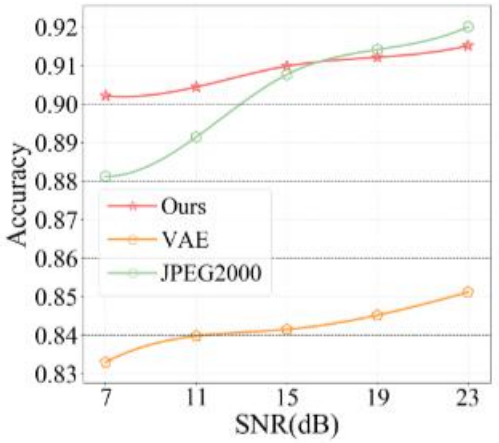




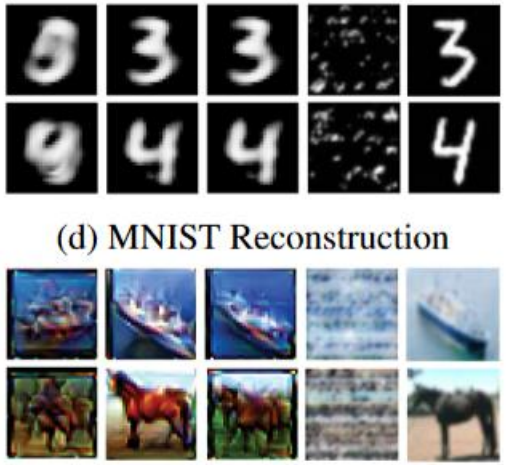
(a) MNIST under AGWN



(b) MNIST under Rayleigh



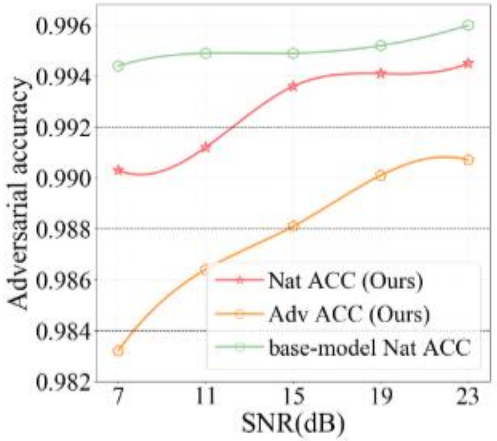
(c) CIFAR-10 under Rayleigh



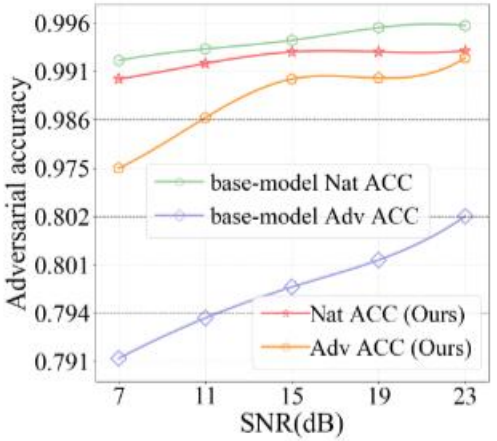
(d) MNIST Reconstruction

(e) CIFAR-10 Reconstruction

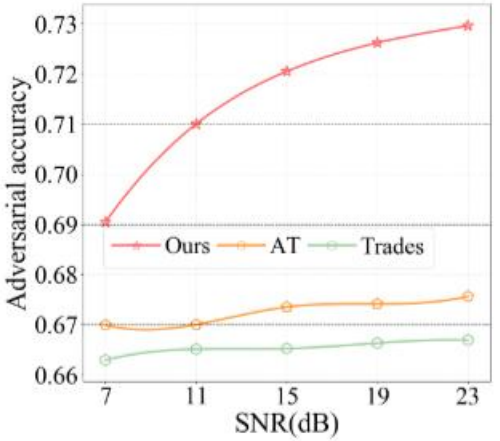
Figure 3: Experimental study on classification performance of MNIST and CIFAR-10 datasets under different SNRs and channels.



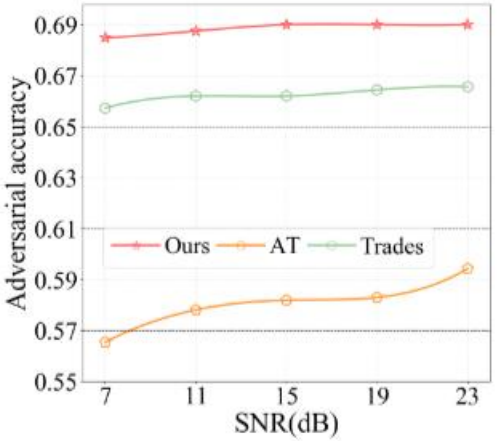
(a) MNIST under AGWN



(b) MNIST under Rayleigh



(c) CIFAR-10 under AGWN



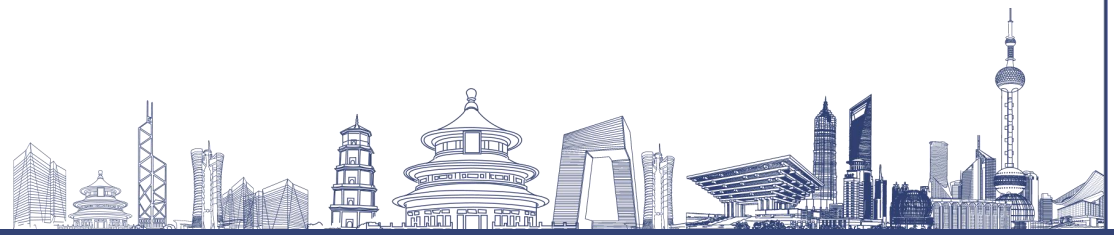
(d) CIFAR-10 under Rayleigh

Figure 4: Adversarial attack experiments on MNIST and CIFAR-10 datasets under different SNRs and channels.

Conclusion & Future Work

5

FIVE

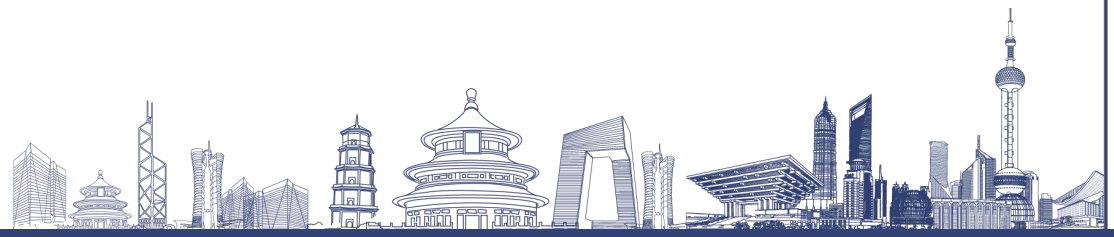


Summary

- We propose RepObE, a secure semantic communication framework with dynamic encryption during semantic extraction and transmission to prevent data reconstruction.
- The integrated representation learning and prototype alignment adversarial training enhances semantic robustness.
- Experiments show >2% improvement in inversion defense and 3%–5% gain in adversarial robustness, with strong visual obfuscation against attackers.

Future Work

- Extend RepObE to multi-modal semantic tasks.
- Investigate lightweight deployment for edge or constrained devices.



Thanks!

the 34th International Joint Conference on Artificial Intelligence (IJCAI 2025)