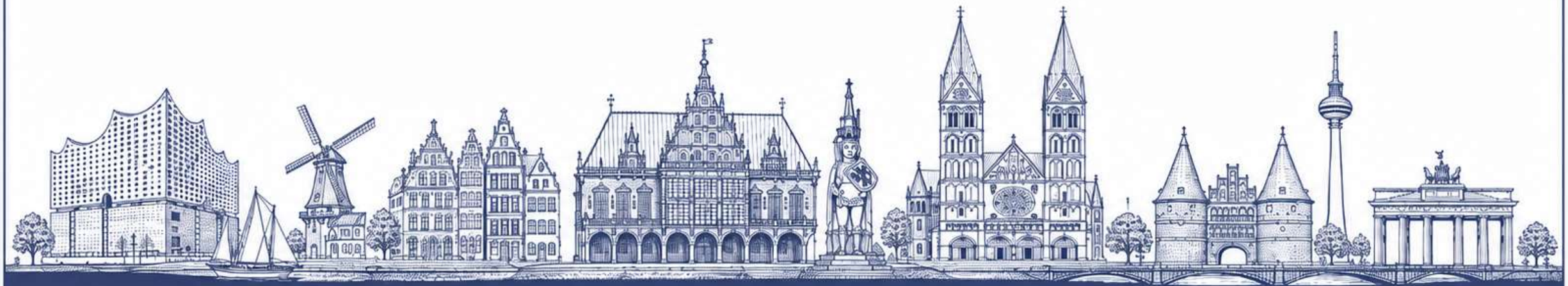
IJCAI ECAI BREMEN 2026

Bridge: A Cross-Modal Learning Framework for Unified Semantic Representation in Noisy Communication

Liang Chen, Yanze Huang, Limei Lin*, Xiaoding Wang, Wei Lou, Jie Wu and Sun-Yuan Hsieh

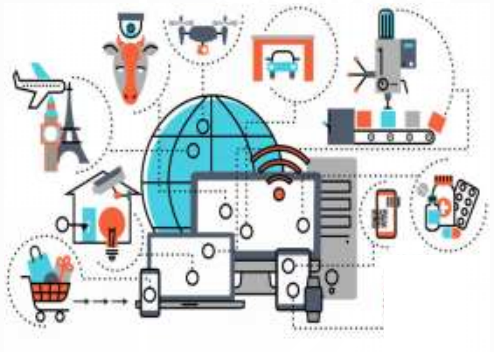


IJCAI–ECAI 2026

15-21 August 2026 Bremen,Germany

CONTENTS

Semantic Communication: From Bits to Meaning



Why Semantic Communication?

- Traditional communication transmits raw bits, causing high redundancy.
- Semantic communication focuses on meaning, improving transmission efficiency.

Why Multimodal Semantic Communication?

- Complementary semantics from text, audio, and video.
- Richer understanding through cross-modal information.
- Robust communication in complex real-world scenarios.

Typical Application Scenarios:

Intelligent Web Services | **Telemedicine** | **Internet of Things** | **Vehicle-to-Everything** | **Digital Twins & Smart Sensing**

Motivation

Challenges in Multimodal Semantic Communication:

- Heterogeneous modalities exhibit large distribution and semantic gaps.
- Channel noise and fading may cause semantic drift and misalignment.
- Accurate alignment and robust transmission remain difficult under bandwidth constraints.

Limitations of Existing Methods:

- Existing methods often struggle with cross-modal alignment.
- Shared semantics can be mixed with modality-specific information.
- Most alignment methods assume relatively stable communication conditions.

Motivation for this research:

How can we learn a unified semantic representation that remains consistent across modalities and robust under noisy channels?

Key Contributions of the Paper

- **Bridge Framework:**

  We propose Bridge, a unified cross-modal framework for robust semantic communication across text, audio, and video.

- **Disentangled Representation Learning:**

  Bridge separates modality-specific and modality-invariant features to preserve unique information while learning cleaner shared semantics.

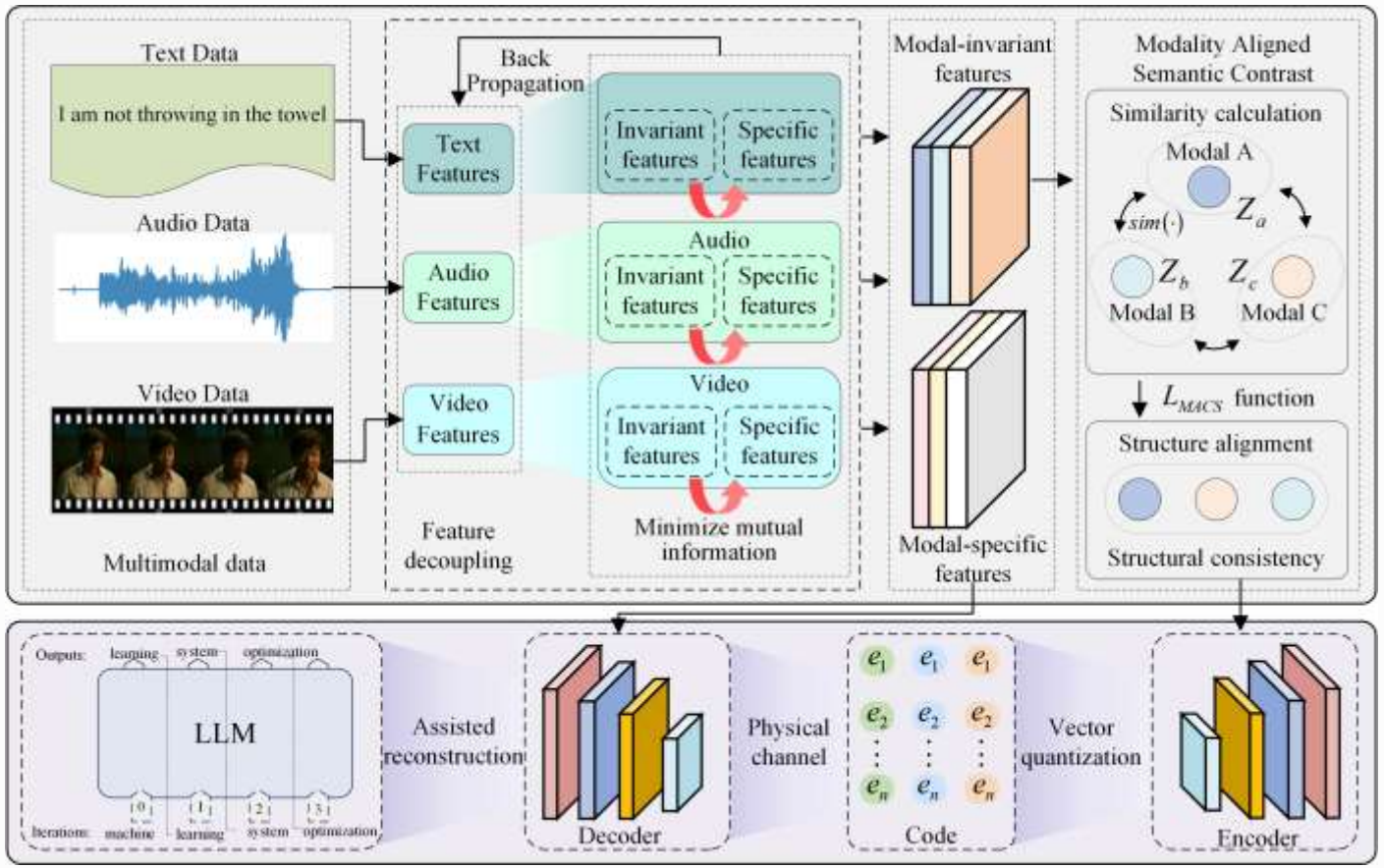
- **Adaptive Cross-Modal Alignment:**

  A modality-aware alignment mechanism maps heterogeneous features into a shared semantic space and improves robustness under noise and fading.

- **Experimental Results:**

  Experiments on multiple multimodal benchmarks demonstrate strong robustness, especially under low-SNR conditions, with about 16.3% improvement in image reconstruction.

Overview of the Bridge Framework



The figure illustrates the key components: SDMI, MACS, vector quantization, and foundation-model-assisted reconstruction.

We propose Bridge, a unified cross-modal learning framework for semantic communication that integrates text, audio, and video into a shared semantic space while preserving modality-specific information.

Bridge disentangles modality-specific and modality-invariant features, aligns shared semantics through MACS, and employs vector quantization for compact and robust transmission over noisy channels.

Method

1. Specific-Invariant Dependency Minimization (SDMI)
Disentangle each modality into specific and invariant features, and minimize their dependency to obtain purified shared semantics.

$$h^m = \mathrm{MLP}^m_{\mathrm{spec}}(\phi_m(x^m)) + \epsilon^m, \qquad z^m = g_{\mathrm{shared}}(\phi_m(x^m))$$

$$\mathcal{L}_{\mathrm{SDMI}} = \mathbb{E}_{p(z_m, z_{\mathrm{inv}})}[T_\theta] - \log \mathbb{E}_{p(z_m)p(z_{\mathrm{inv}})}[\exp(T_\theta)]$$

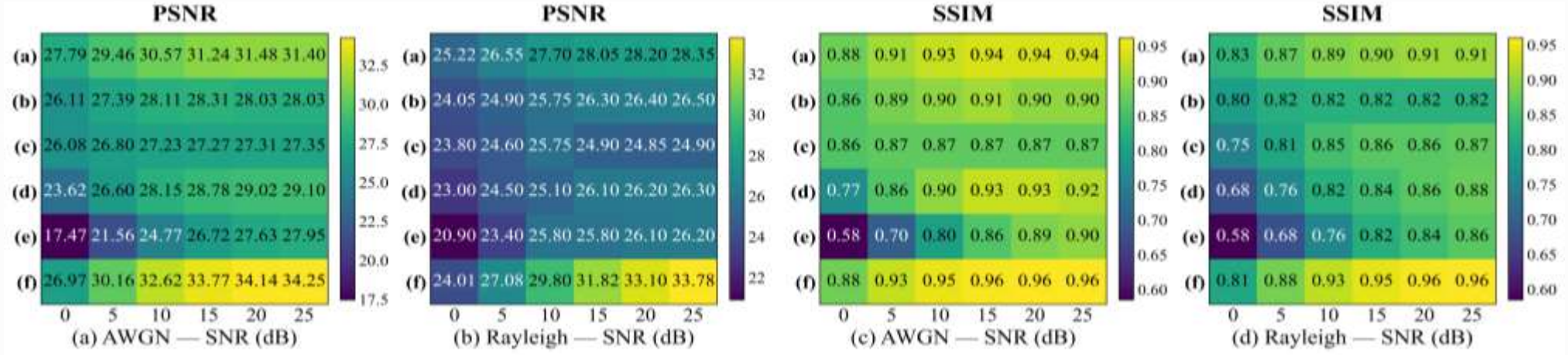
2. Modality-Aware Cross-Modal Semantics (MACS)
Perform pairwise-gated contrastive alignment to emphasize reliable matches and suppress noisy cross-modal pairs.

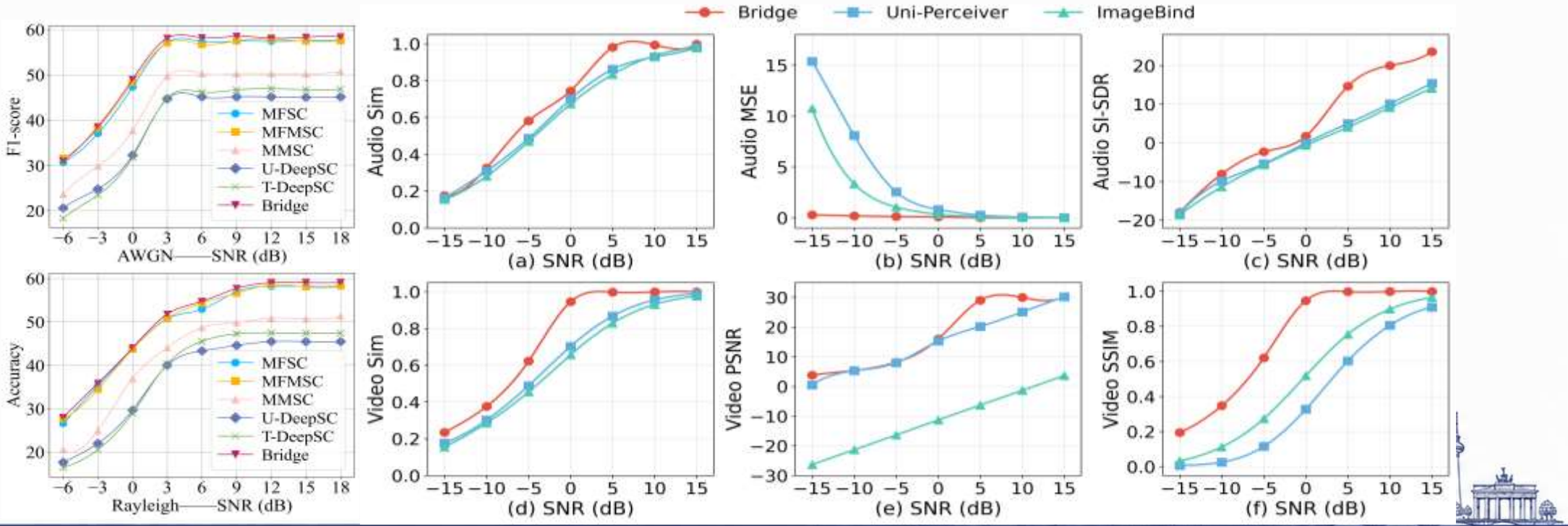
$$g_{t,j} = \sigma\left(\mathrm{MLP}_{\mathrm{gate}}([\phi_a(z^a_t) \| \phi_b(z^b_j)])\right)$$

$$\mathcal{L}_{\mathrm{MACS}} = -\frac{1}{N}\sum_{t-1}^{N} \log \frac{\exp(g_{t,t}s_{t,t})}{\sum_{j-1}^{N}\exp(g_{t,j}s_{t,j})}$$

Key idea: SDMI purifies shared semantics, while MACS enables robust cross-modal alignment.

Experiments & Results





Conclusion

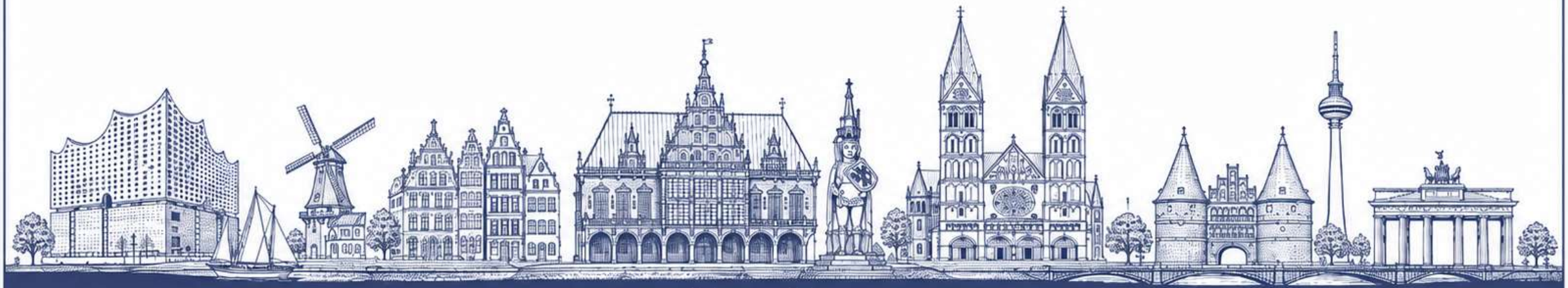
- We propose Bridge, a unified cross-modal framework that learns shared semantic representations across text, audio, and video.
- Bridge combines specific-invariant disentanglement and modality-aware alignment to preserve semantic consistency under noisy channels.
- Extensive experiments demonstrate superior reconstruction quality, task performance, and robustness, especially in low-SNR conditions.

Future Work

- Extend Bridge to more heterogeneous modalities and complex real-world communication scenarios.
- Explore more efficient and adaptive semantic transmission under dynamic channel conditions.

IJCAI ECAI BREMEN 2026

Thanks!



IJCAI–ECAI 2026

15-21 August 2026 Bremen,Germany