Grasp: Refining Semantic Graphs into Purified Knowledge for Cross-Modal Communication

Liang Chen, Xiaoding Wang, Limei Lin*, Dajing Wang, Zhiquan Liu, Jie Wu

The Web Conference 2026

29 June - 3 July 2026 Dubai, United Arab Emirates

CONTENTS

Background & Motivation

1

ONE

Semantic Communication: towards Knowledge-Centric Cross-Modal Communication

Typical Application Scenarios:

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

Communication Challenges in the Multimodal Web Era:

- Explosive growth of video, audio, and text databrings high transmission cost.
- Traditional communication focuses on raw bitreconstruction, causing redundancy and lowsemantic efficiency.

Why Cross-Modal Communication is More Challenging:

- Modalities are often asynchronous, incomplete,and affected by channel noise.
- Surface-level feature transmission may failunder low SNR or missing modalities.

Motivation

Challenges in Real-World Cross-Modal Communication:

- Surface-level features are easily corrupted by noise,packet loss, and low SNR.
- Video, audio, and text are often asynchronous,causing semantic misalignment.
- Missing modalities and distribution shifts reduce semantic fidelity and robustness.

Limitations of Existing Methods:

- Most methods focus on feature alignment or reconstruction rather than core knowledge.
- Shared semantics are entangled with modality-specific factors.
- They lack structured, schedulable semantic units for efficient transmission and recovery.

Motivation for this research:

There is an urgent need for a knowledge-centric semantic communication framework that can organize semantics, purify modality-invariant knowledge, and support robust transmission and recovery under realistic conditions.

Key Contributions of the Paper

- **Grasp Framework:**

  Grasp organizes multimodal content as semantic graphs and extracts prioritized transmissibleknowledge units.

- **Knowledge Purification::**

  A conditional mutual-information objective enables three-way disentanglement for robust tri-modal alignment.

- **Unified End-to-End Pipeline:**

  The framework unifies semantic modeling, purification, shared knowledge coding, and attention-based recovery.

- **Experimental Results:**

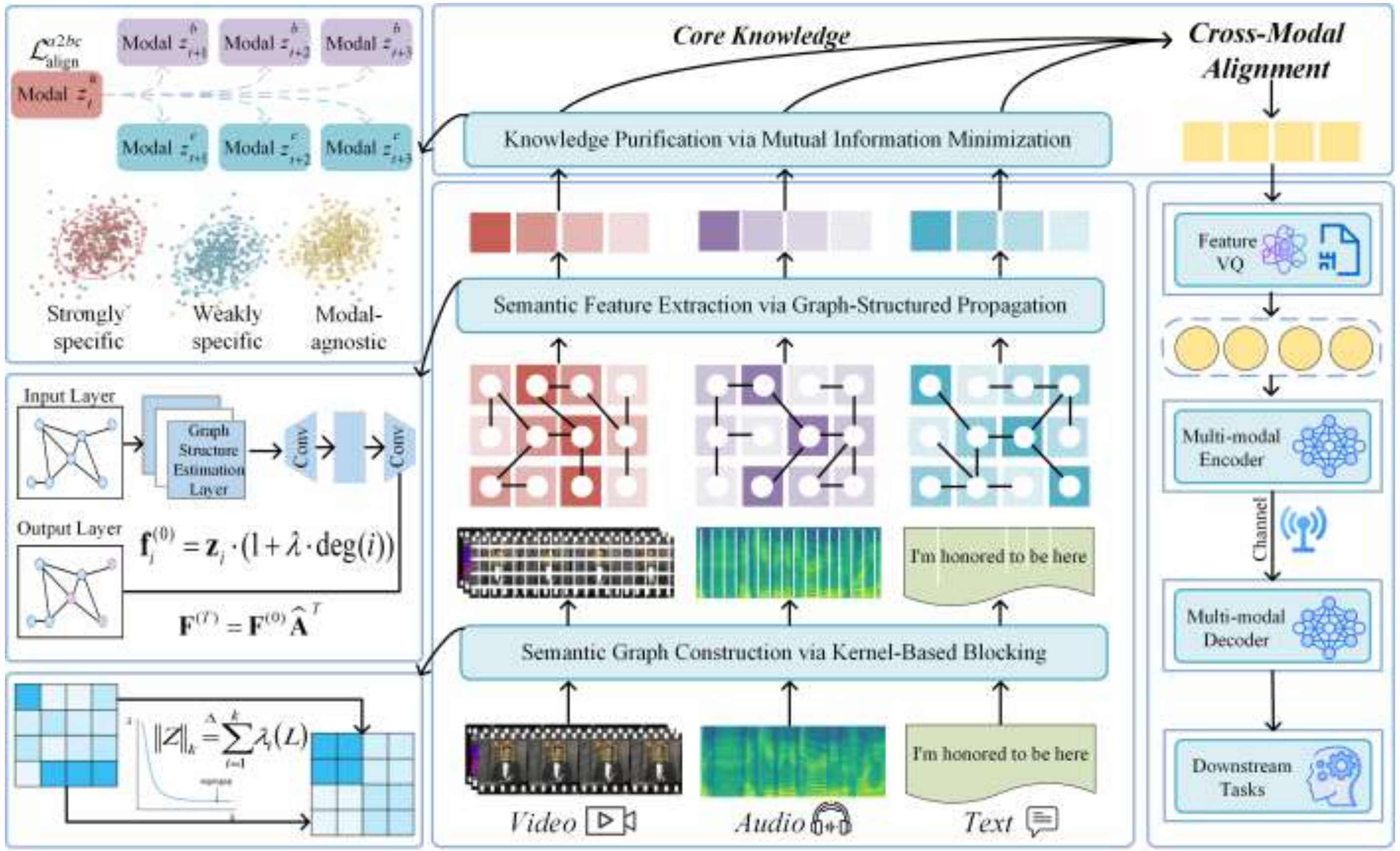
  Experiments show better semantic quality vs. bit cost, low latency, and effective component-wise ablation results.

Grasp Framework Design

2

TWO

Overview of the Grasp Framework

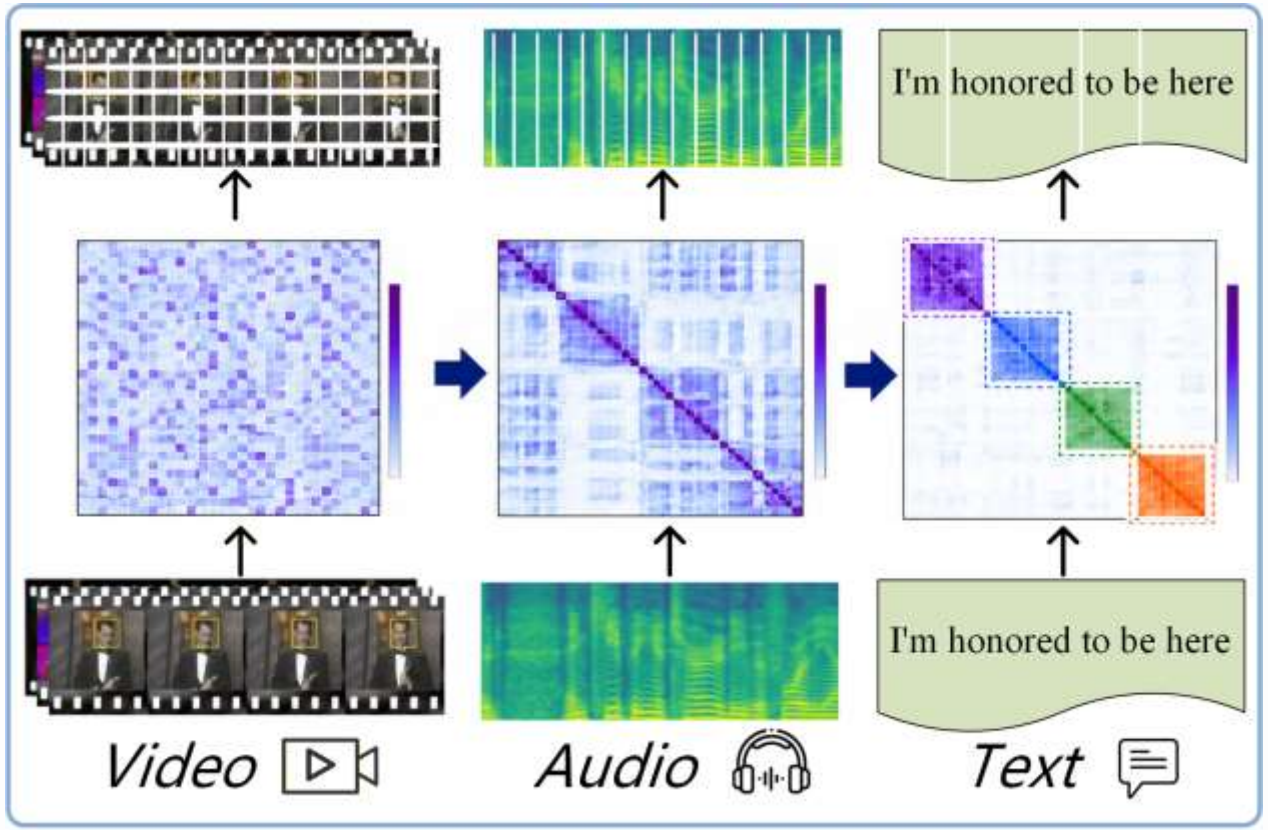


The figure illustrates key components: KBB, GSP, MICA, shared knowledge coding, and robust semantic recovery.

We propose Grasp, a knowledge-centric framework for cross-modal communication that organizes multimodal inputs into semantic graphs, purifies modality-invariant knowledge, and transmits compact semantic units efficiently.

The framework integrates semantic graph construction, graph-structured feature propagation, mutual-information-based knowledge purification, cross-modal alignment, and shared vector-quantized knowledge coding.

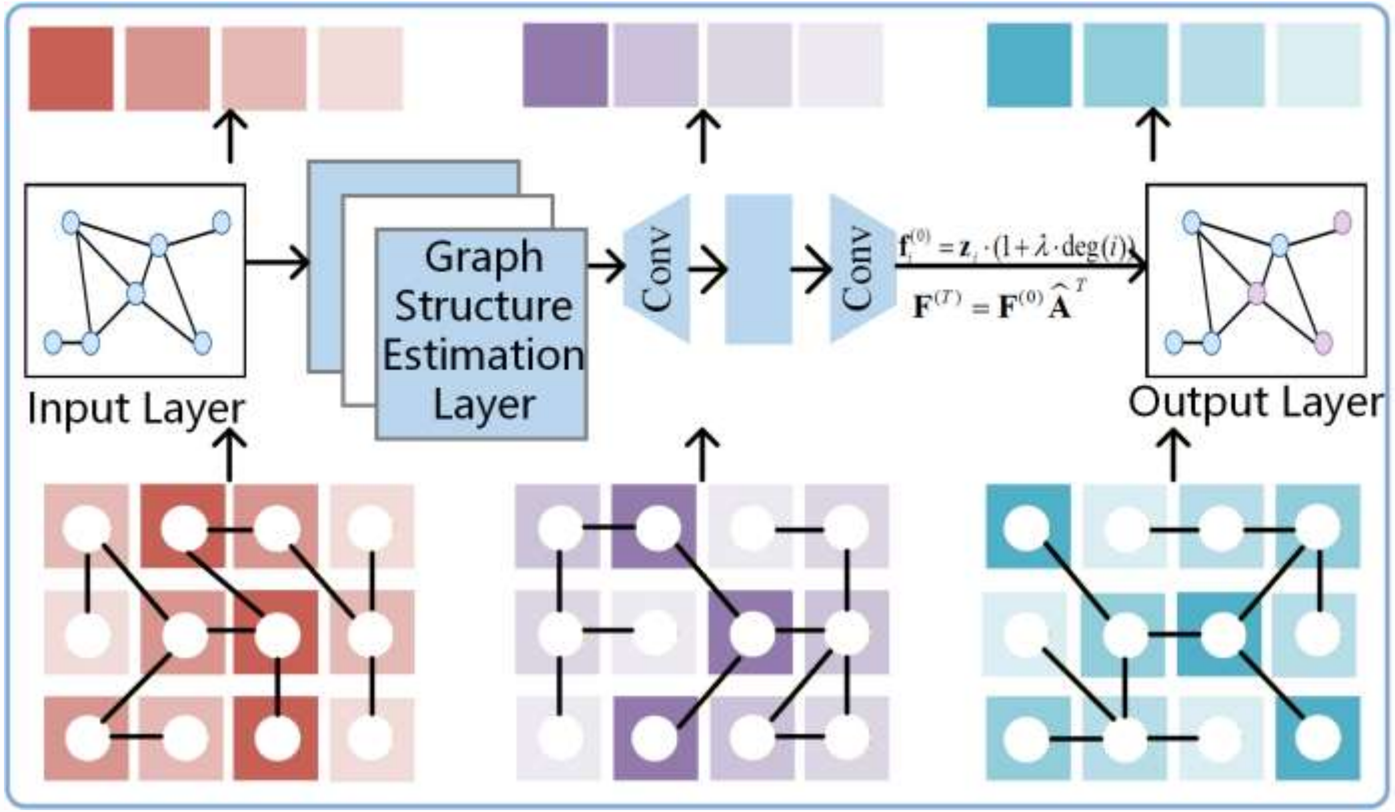
Semantic Graph Construction via Kernel-Based Blocking



- KBB determines how to divide raw multimodal streams into semantic blocks.
- It learns semantic relations among candidate segments through kernel self-representation.
- The learned affinity gradually forms a block-diagonal structure, where each block corresponds to a coherent semantic unit.
- These semantic blocks serve as the basic nodes for subsequent semantic graph construction and knowledge purification.

Key idea: KBB decides semantic blocks by discovering block-diagonal structures from kernel-space relations.

Semantic Feature Extraction via Graph-Structured Propagation



Purpose of GSP

- Estimate semantic relations among blocks.
- Propagate features through the semantic graph.
- Enhance each block with neighborhood context.
- Generate importance-aware semantic representations.

Key idea: GSP uses graph propagation to enrich semantic blocks with relational context.

Knowledge Purification via Mutual Information Minimization

1. Component Refinement

MICA further refines the routed components into purified core, specific, and weak/noisy representations.

$$x_i^m = g_{\text{inv}}(\mathbf{x}_i'^m), \quad y_i^m = g_{\text{str}}(\mathbf{y}_i'^m), \quad z_i^m = g_{\text{weak}}(\mathbf{z}_i'^m)$$

2. Conditional MI Minimization

Minimize conditional dependency to purify modality-invariant knowledge.

$$\begin{aligned} I_{\text{cmi}}(x; y \mid z) = {} & \mathbb{E}_{p(x,y,z)}\left[\log q_\theta(y \mid x, z)\right] \\ & - \mathbb{E}_{p(z)\,p(x|z)\,p(y|z)}\left[\log q_\theta(y \mid x, z)\right] \end{aligned}$$

3. Temporal Cross-Modal Alignment

Align one anchor modality with the other two modalities over time.

$$\mathcal{L}_{\text{align}}^{a2bc} = \frac{-1}{K} \sum_{k=1}^{K} \log \left[\frac{\exp\left(z_{t+k}^b W_k^a h_t^a + z_{t+k}^c W_k^a h_t^a\right)}{\sum_{(z_j^b, z_j^c) \in Z_{bc}(t,k)} \exp\left(z_j^b W_k^a h_t^a + z_j^c W_k^a h_t^a\right)} \right]$$

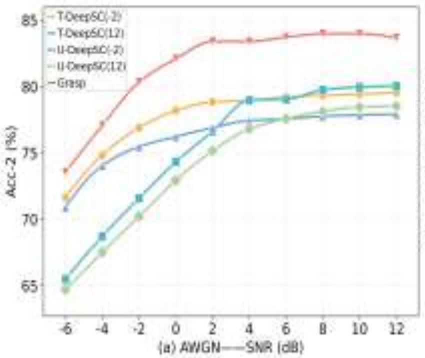
Key idea: MICA purifies core knowledge and aligns tri-modal semantics.

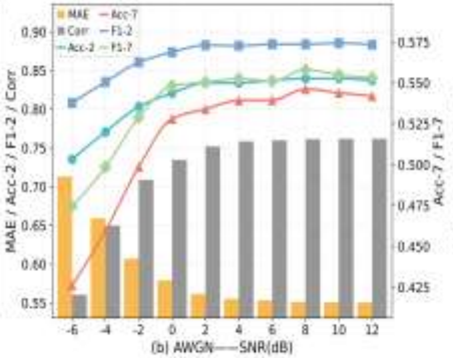
Experiments & Results

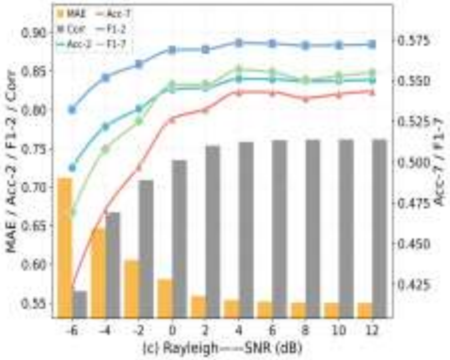
F 3 U

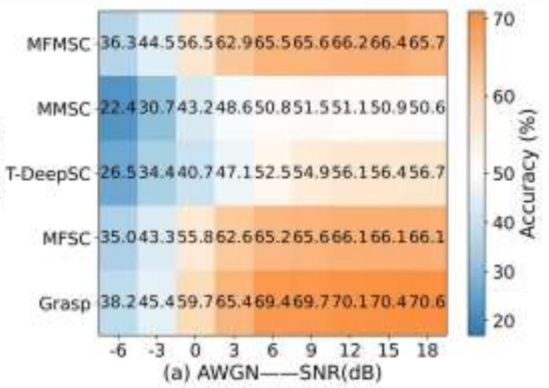
R

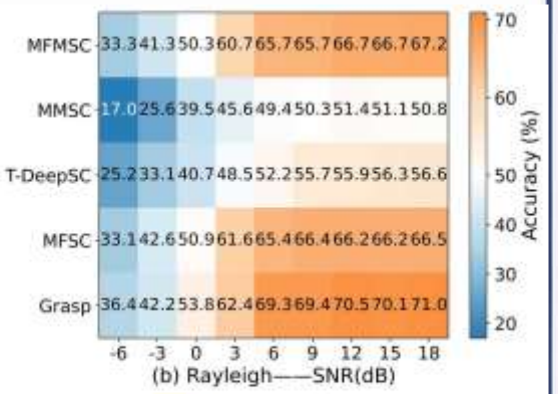
Experiments & Results

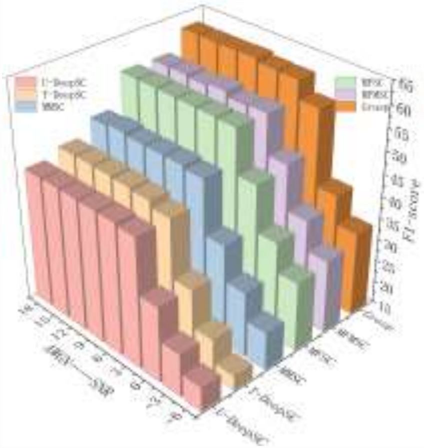


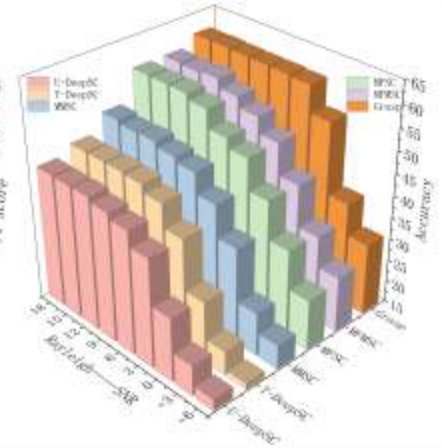


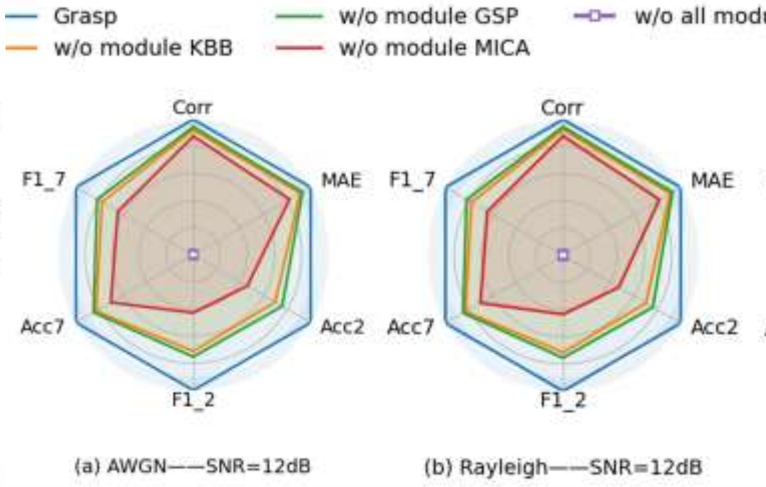


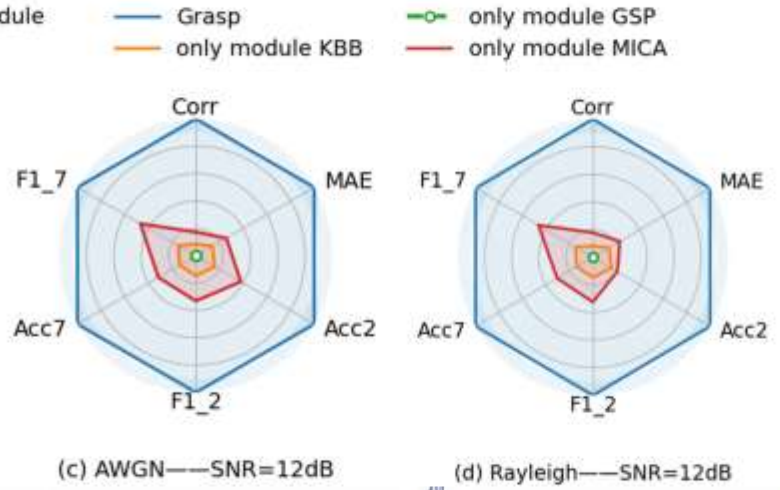












Conclusion & Future Work

4

FIVE

Conclusion

- We propose Grasp structures tri-modal inputs as semantic graphs and transmits purified, modality-invariant knowledge.
- It disentangles strongly related, weakly related, and irrelevant components for robust tri-modal alignment.
- Experiments show consistent improvements in fidelity, robustness, and downstream performance.

Future Work

- Extend to richer modalities such as depth, LiDAR, and physiological signals.
- Enhance robustness and security under adversarial and real-world deployment settings.

Thanks!

The Web Conference 2026

29 June - 3 July 2026 Dubai, United Arab Emirates