

# Multimodal Data

Combines information from multiple modalities simultaneously



Video with Audio



Image with Captions



Sensor with Location



## Benefits

- Complementary information beyond single modality
- Improved robustness: other modalities compensate for noise
- Richer understanding of data



## Challenges

- Alignment across modalities
- Synchronization of data streams
- Fusion of different modalities

## Requirements

Cross-Modal Learning

Joint Representations

## Applications

Video Understanding

Visual Question Answering

Robotics