Multi-modal Clustering

Clustering Data with Multiple Modalities

⚠ Challenge

Different feature spaces and scales across modalities

Three Clustering Approaches

Early Fusion

Concatenate features from different modalities

Late Fusion

Cluster each modality separately, then combine results

Deep Multi-modal

Learn shared representation across modalities

Image

+

Text

→

Shared Space

Joint representation

🎯 Applications

🔥 Modern Approach

CLIP-style contrastive learning for multi-modal clustering

Video Analysis (visual + audio)

Medical Imaging with Reports

Contrastive Multi-modal Learning