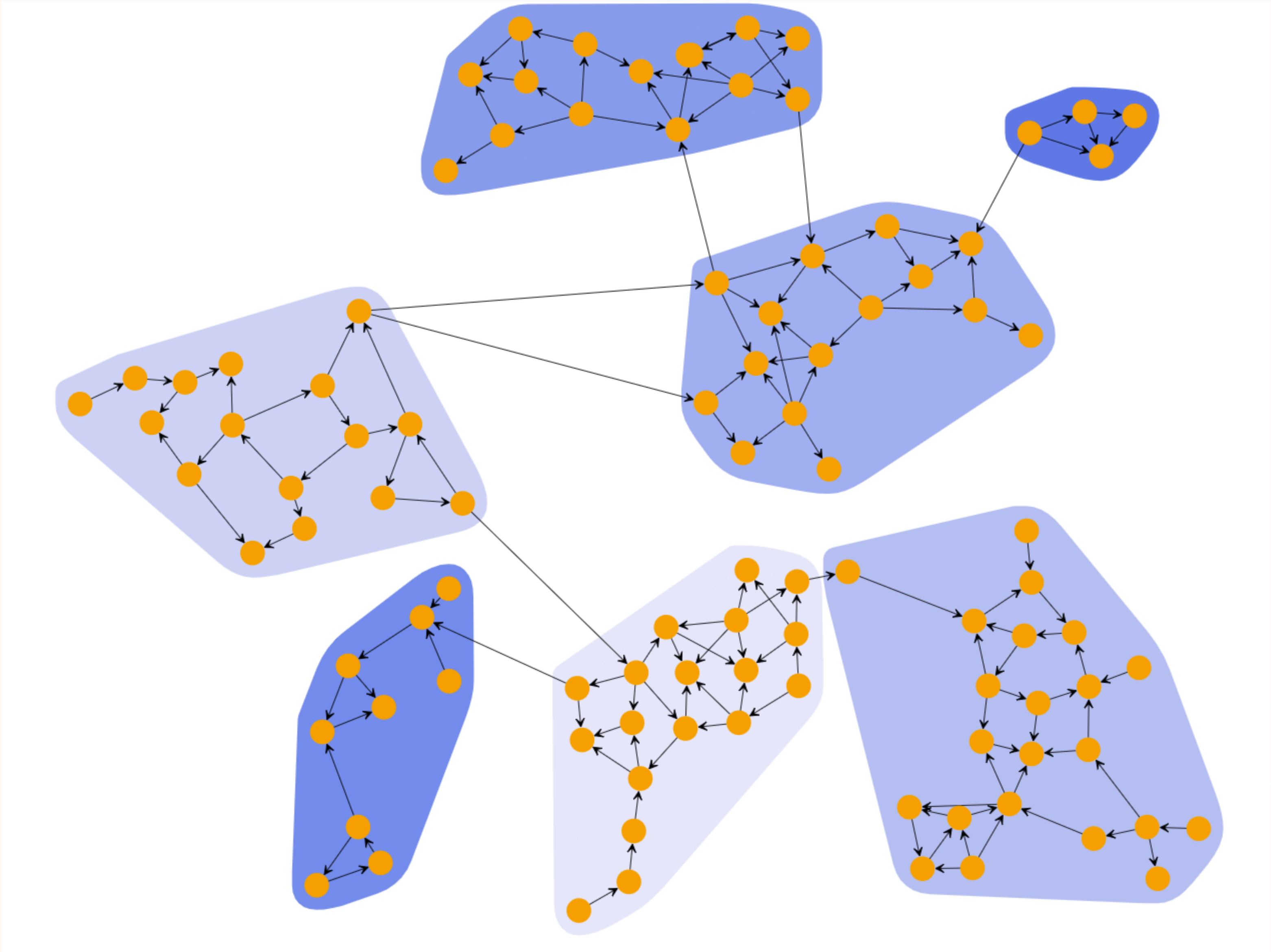
Revolutionizing Clustering with Customized Multi Modal Subspace Proxy Learning

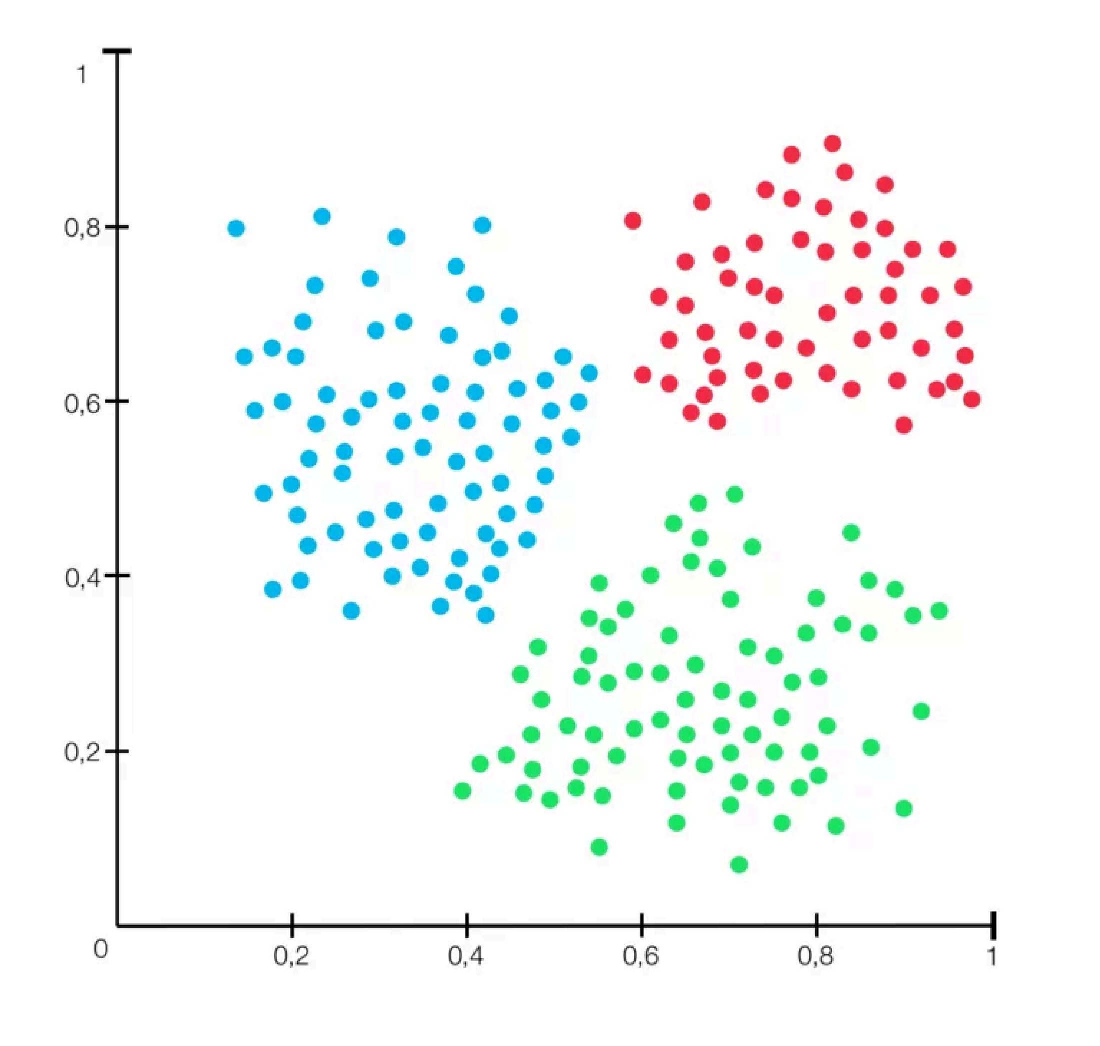


Authors: Jiawei Yao, Qi Qian, Juhua Hu

Presented by : Syeda Nida Khader

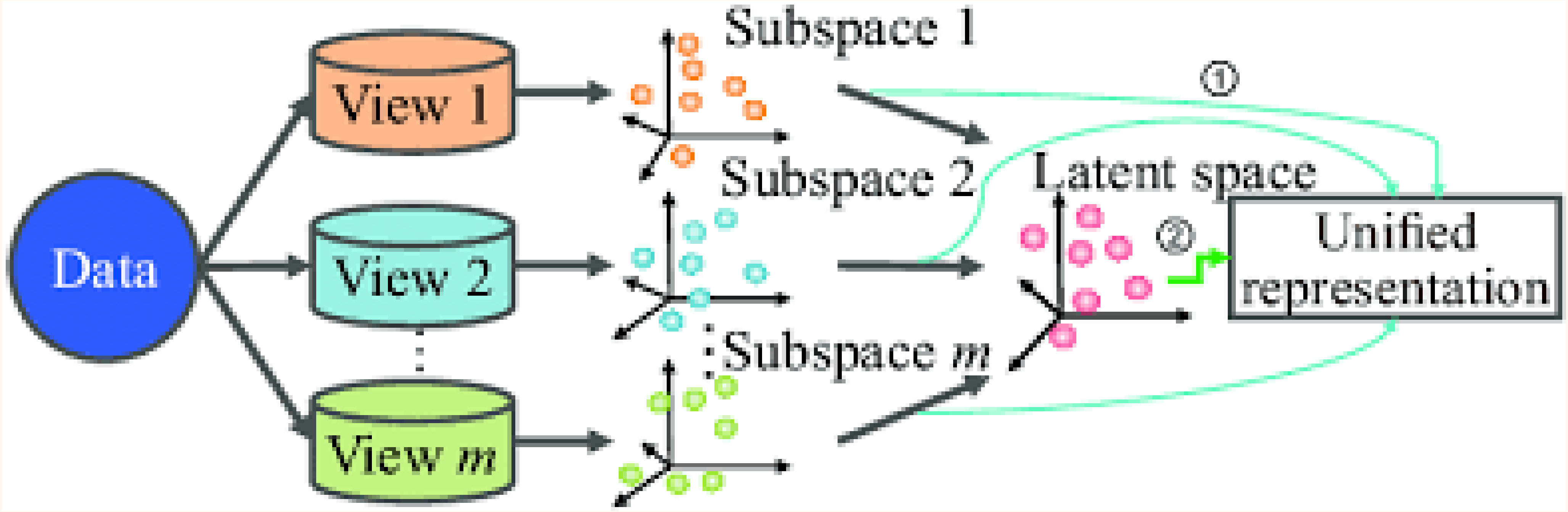
What is Clustering?

Clustering is a machine learning technique used to group data points into clusters based on their similarity. It is widely used in applications like customer segmentation, image categorization, and recommendation systems.



Challenges in traditional clustering techniques for multimodal data

- **Heterogeneity**: Multi-modal data includes different types of information (e.g., text, images, audio), which are difficult to combine effectively.
- **High Dimensionality**: Traditional methods struggle to handle the high-dimensional nature of multi-modal data.
- **Scalability**: Existing methods often fail to scale efficiently with large datasets.
- **Poor Representation**: Inadequate techniques for extracting shared representations from diverse modalities.



Importance of subspace learning in clustering

Subspace learning focuses on identifying a lower-dimensional space where meaningful patterns in multi-modal data can emerge. It is crucial for:

1. **Improved Accuracy**

   By capturing shared latent features across modalities.

2. **Reduced Complexity**

   By projecting data into lower dimensions for efficient processing.

3. **Handling Multi-Modality**

   Creating unified representations for heterogeneous data types.

Limitations of Existing Methods

1. **Poor Scalability**: Traditional clustering techniques struggle to handle large datasets, making them inefficient for real-world, high-volume data applications.
2. **Difficulty Handling Multi-Modal Data**: Most methods are designed for single-modal data, failing to integrate and process diverse data types like images, text, or audio effectively.

Objective

To develop a scalable and efficient clustering method that can seamlessly handle multi-modal data while maintaining high accuracy and computational efficiency.

Multi-Sub's Innovative Approach

1

User Interest Input

Capture high-level user concept through succinct keywords. Initiate the clustering process based on user-specific interests.

2

Subspace Proxy Generation

Utilize LLMs to generate common categories under the desired concept. Create a subspace basis for targeted representation learning.

3

Simultaneous Learning

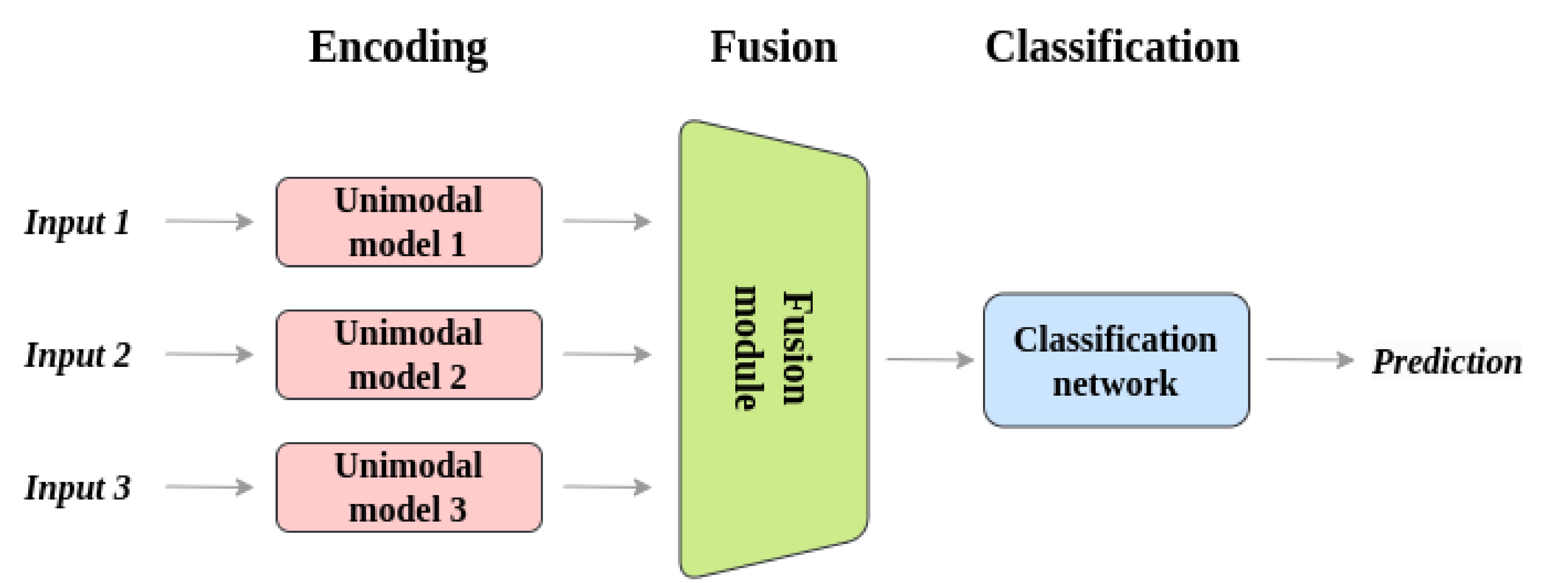
Incorporate clustering loss during subspace learning. Optimize representations and clustering results concurrently for improved performance.

Multi-Modal Integration

Textual Processing

Extracting meaningful features from textual data using advanced NLP techniques.

Examples: embeddings, semantic analysis.



Visual Analysis

Leveraging computer vision to process image data for clustering.

Examples: ResNet-based feature extraction.

Multi-Modal Fusion

Combining textual, visual, and other modalities into a cohesive framework.

Enables richer and more precise clustering outcomes.

Multiple Clustering Applications

1

2

3

4

Personalized Recommendations

Multi-Sub can analyze user preferences and browsing history to provide personalized recommendations for products, content, or services.

Customer Segmentation

Businesses can utilize Multi-Sub to segment customers based on their demographics, purchasing behavior, and online activity.

Healthcare Diagnosis

Multi-Sub can analyze patient data to assist in disease diagnosis, identify patient subgroups, and personalize treatment plans.

Financial Risk Assessment

Multi-Sub can be used to analyze financial data for risk assessment, fraud detection, and investment optimization.

Key Contributions

Introduction of CMMSPL: A groundbreaking approach tailored for clustering multi-modal data with high precision and efficiency.

Innovative Integration: Combines the ability to handle multi-modal data with a scalable and adaptive clustering mechanism.

Superior Performance: Demonstrates significantly improved clustering accuracy and robustness when compared to traditional baseline models.

Background

1 Multi-Modal Data

Multi-modal data combines different data types, such as images, text, and other formats, often requiring specialized techniques to process and integrate. Example: An image-text dataset where captions and visuals must be analyzed together for meaningful insights.

2 Subspace learning

A technique to identify meaningful low-dimensional representations of data while preserving its structure. Particularly useful for clustering high-dimensional and multi-modal data.

3 Proxy-Based Clustering

Clustering method that leverages representative "proxies" or anchor points for scalable and efficient clustering. Reduces computational complexity by clustering around these proxies rather than individual data points.

Proposed Method

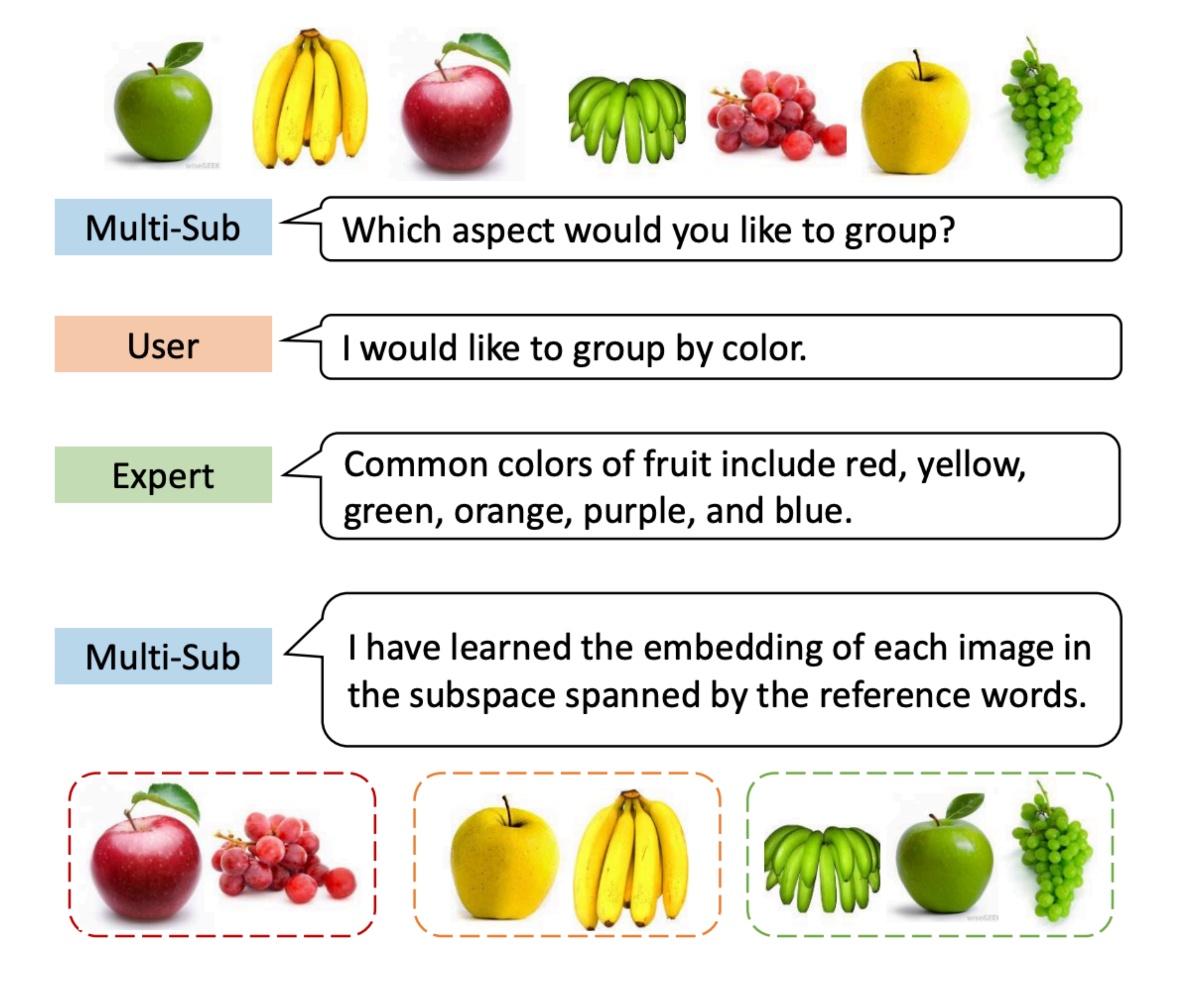
Overview of CMMSPL (Customized Multi-Modal Subspace Proxy Learning):

Multi-Modal Feature Extraction:
Extracts meaningful features from diverse data modalities (e.g., text, images, audio).
Ensures that unique characteristics of each modality are preserved during processing.

Subspace Learning:
Identifies shared low-dimensional representations across modalities.
Captures underlying correlations between modalities to facilitate better clustering.

Proxy-Based Clustering:
Uses representative proxies (anchor points) to group similar data points efficiently.
Reduces computational overhead by avoiding direct pairwise comparisons.

Focus on Scalability and Efficiency:
Designed to handle large datasets with multiple modalities.
Optimizes performance for real-world, high-dimensional clustering tasks.



Methodology

1

Data Preprocessing

- Normalize and clean data from multiple modalities to ensure consistency and quality.
- Align data formats (e.g., converting text to embeddings or resizing images) for integration

2

Multi-Modal Feature Learning

- Extract meaningful features from each modality (e.g., text embeddings, visual features using deep neural networks).
- Maintain modality-specific characteristics during feature extraction

3

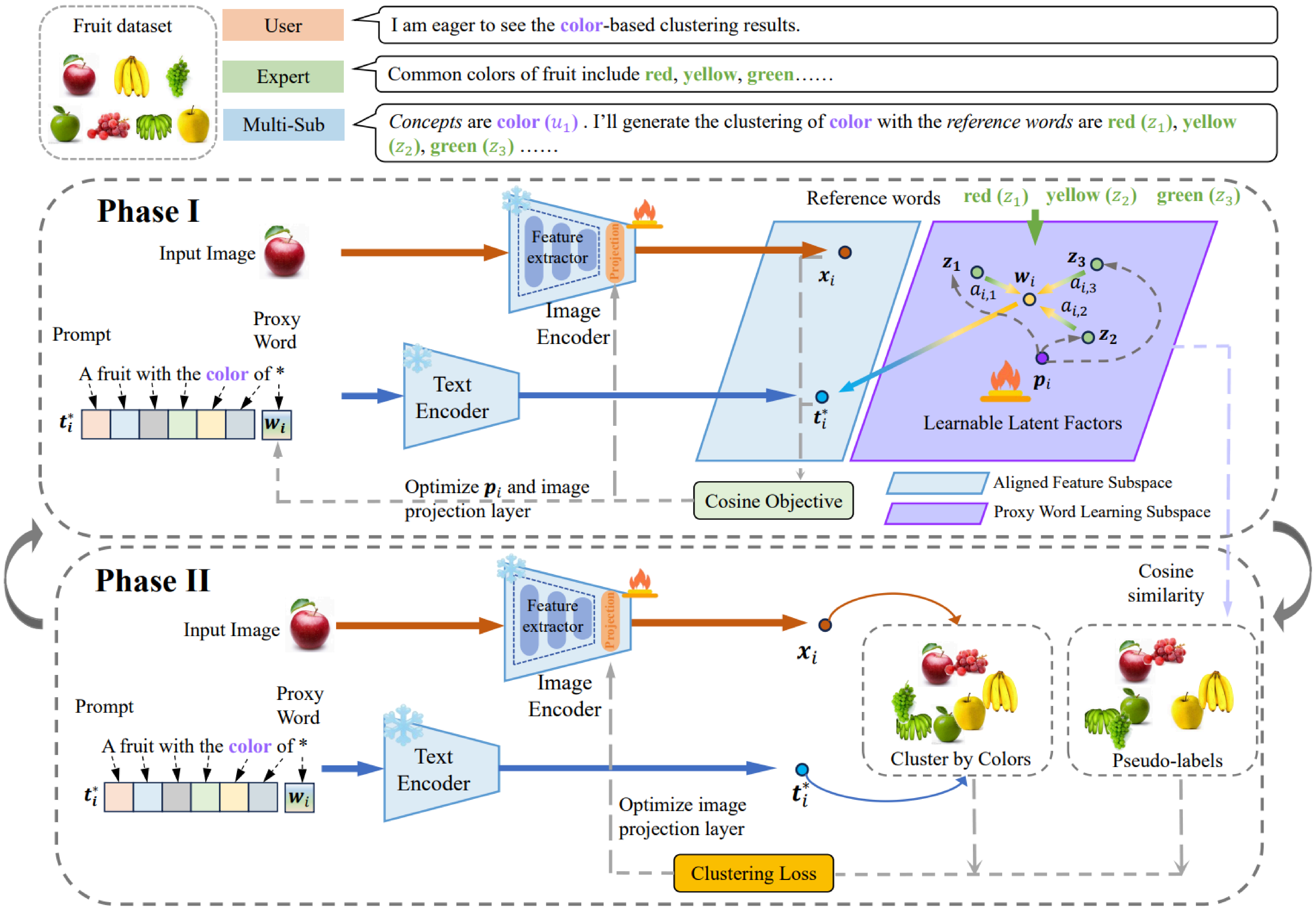
Proxy Generation

- Generate representative proxies for data clusters to reduce computational complexity.
- Proxies act as anchor points, facilitating efficient clustering.

4

Subspace Clustering

- Learn shared low-dimensional subspaces for combined modalities.
- Group data points into clusters based on their proximity to proxies in the subspace.



Fruit dataset
User
I am eager to see the color-based clustering results.
Expert
Common colors of fruit include red, yellow, green……
Multi-Sub
Concepts are color (u_1) . I'll generate the clustering of color with the reference words are red (z_1), yellow (z_2), green (z_3) ……
Phase I
Input Image
Feature extractor
Projection
Image Encoder
Reference words
red (z_1) yellow (z_2) green (z_3)
x_i
z_1 w_i z_3 $a_{i,1}$ $a_{i,3}$ $a_{i,2}$ z_2 p_i
Prompt
Proxy Word
A fruit with the color of *
t_i^* w_i
Text Encoder
t_i^*
Learnable Latent Factors
Optimize p_i and image projection layer
Cosine Objective
Aligned Feature Subspace
Proxy Word Learning Subspace
Phase II
Input Image
Feature extractor
Projection
Image Encoder
x_i
Cosine similarity
Prompt
Proxy Word
A fruit with the color of *
t_i^* w_i
Text Encoder
t_i^*
Cluster by Colors
Pseudo-labels
Optimize image projection layer
Clustering Loss

Conclusion

Multi-Sub represents a significant leap forward in personalized clustering. It utilizes LLMs for targeted subspace learning, enabling more accurate and relevant clustering results.

This innovative approach addresses the limitations of traditional clustering methods, making it ideal for applications requiring user-specific insights.

Multi-Sub's ability to handle multi-modal data and adapt to individual preferences opens new possibilities in personalized recommendations, customer segmentation, and beyond.

Summary Findings

- CMMSPL Outperforms Traditional Methods: Demonstrated superior accuracy and robustness compared to existing clustering approaches.
- Scalable and Efficient for Multi-Modal Clustering: Handles large-scale, multi-modal datasets effectively while maintaining computational efficiency.