Calibrated Information Bottleneck for Trusted Multi-modal Clustering

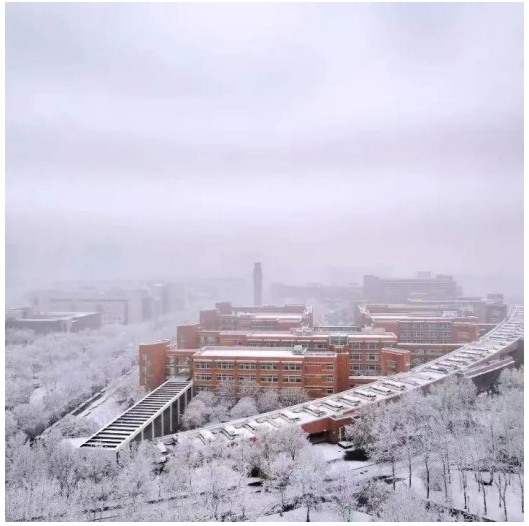
Shizhe Hu, Zhangwen Gou, Shuaiju Li, Jin Qin, Xiaoheng Jiang, Pei Lv, Mingliang Xu†

School of Computer and Artificial Intelligence

Zhengzhou University

Zhengzhou, Henan, China

Zhengzhou University (Also called “Western Park of Zhengzhou”)



Tourist Spot

Outline

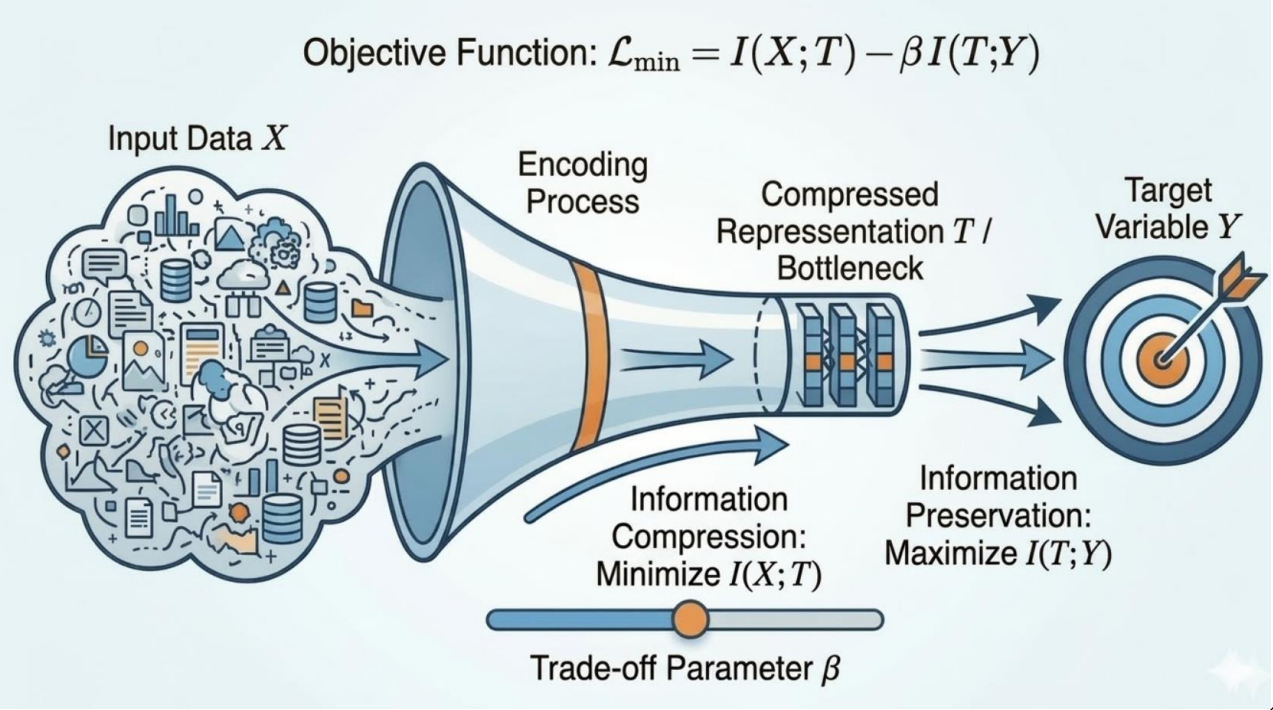
- Problem background
- Previous works
- Our proposal
- Experiments
- Conclusion

Outline

- **Problem background**
- Previous works
- Our proposal
- Experiments
- Conclusion

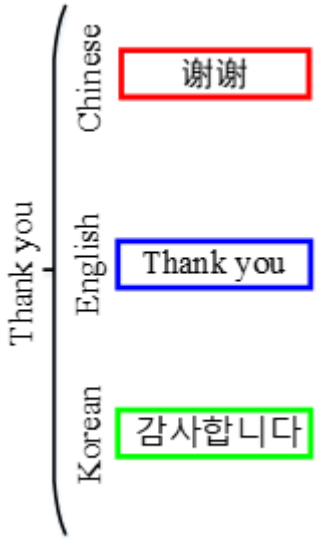
Information Bottleneck Theory

Information Bottleneck Theory provides an elegant theoretical framework for learning concise and effective data representations, centered on finding an optimal balance between data compression and information preservation.

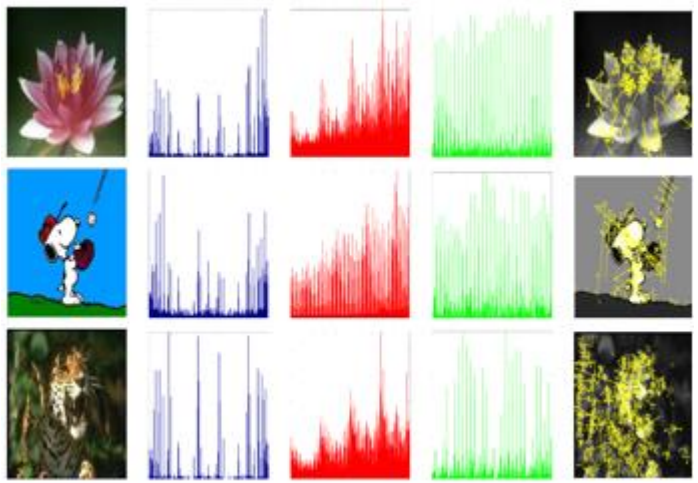


Characteristics of multi-modal datasets

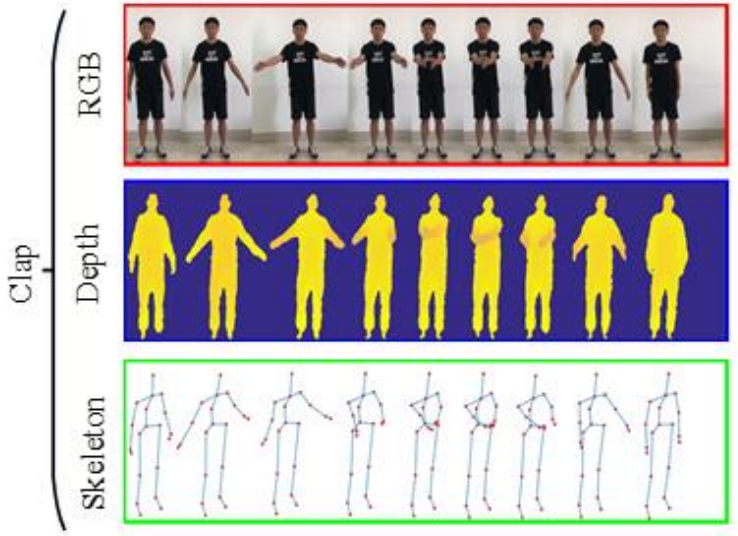
In Big Data era, many kinds of multi-modal data are emerging.



Multi-lingual Text



Multi-feature Image

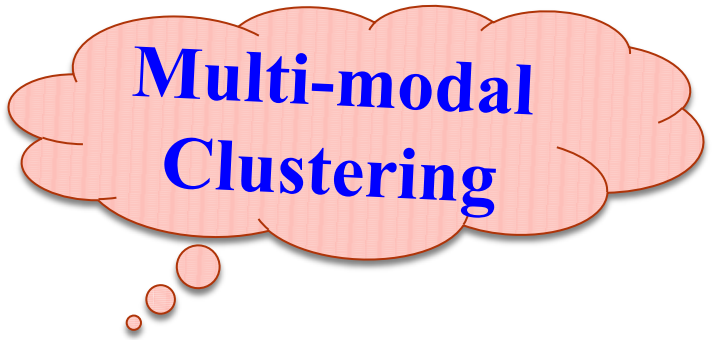


Multi-modal human action video

Property: Heterogeneous, Large-scale, Diversification, Complexity

Limitations of supervised multi-modal classification methods

1. **Time-consuming and cost-expensive for labelling;**
2. **Over-reliance on the label information of trained data;**
3. **Ignoring the characteristics of the input data itself.**



Outline

- Problem background
- **Previous works**
- Our proposal
- Experiments
- Conclusion

Previous IB-based clustering methods

- A Peer-review Look on Multi-modal Clustering: An Information Bottleneck Realization Method
- Super Deep Contrastive Information Bottleneck for Multi-modal Clustering
- Multi-aspect Self-guided Deep Information Bottleneck for Multi-modal Clustering

Limitations:

– *Assumption of accurate mutual information estimation;*

– *Difficulty in accurately calculating mutual information for high-dimensional variables;*

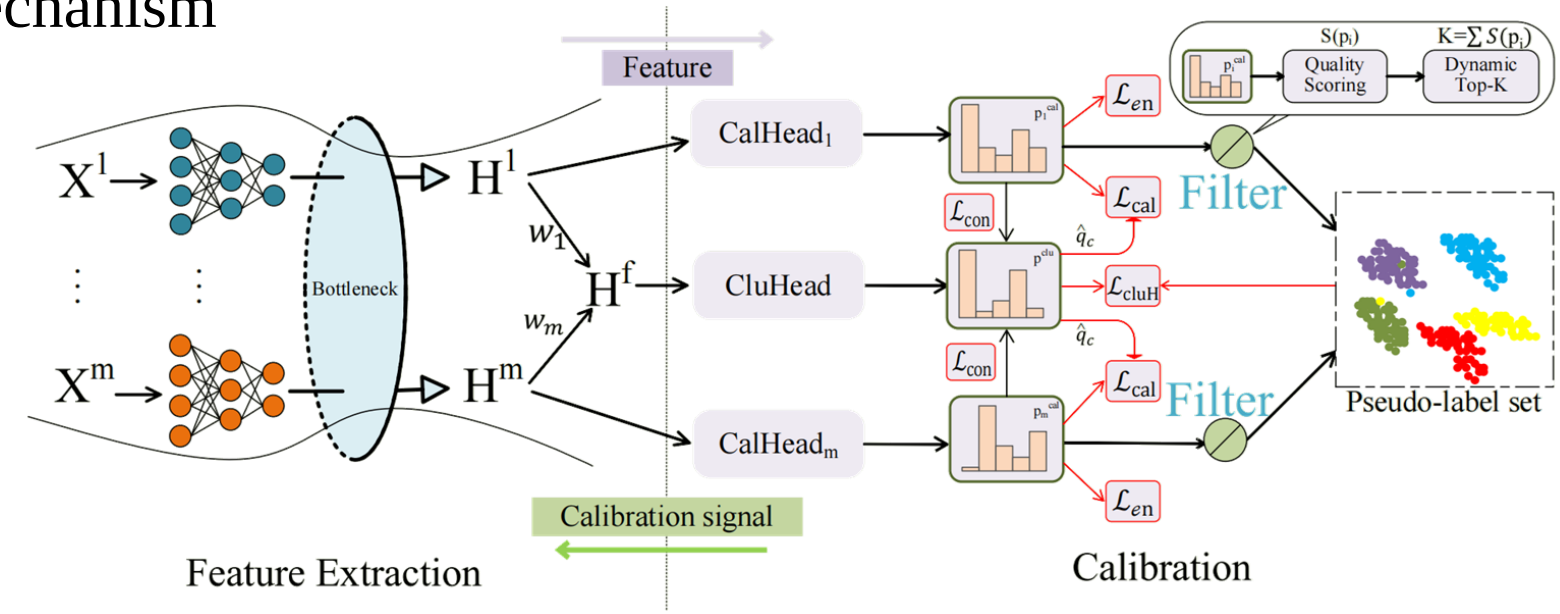
– *Lack of high-quality pseudo-label screening strategies.*

Outline

- Problem background
- Previous works
- **Our proposal**
- Experiments
- Conclusion

Our proposed method

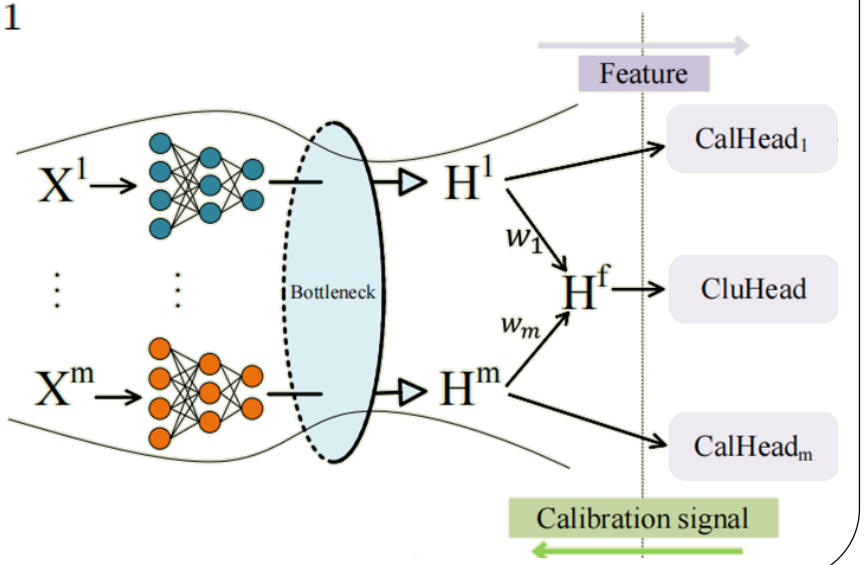
- CaLibrated Information Bottleneck for Trusted Multi-modal Clustering (CLIB) :
  - Representation Learning Based on IB
  - Calibration Heads
  - Pseudo-label Screening Mechanism
  - Cluster Head



Representation Learning Based on IB

Initially, IB is applied to extract compact and pure features from each modality. By leveraging a self-weighting manner, we then obtain the fused features. Following this, while modality-specific features are processed by multiple Calibrated Heads, the fused features are assigned to the Cluster Head for final processing.

$$\mathcal{L}_{IB} = \mathcal{L}_C - \beta\mathcal{L}_P = \sum_{i=1}^{M} I\left(X^i, H^i\right) - \sum_{i=1}^{M}\sum_{j=i+1}^{M} I\left(H^i, H^j\right) - \beta \sum_{m=1}^{M} I(H^m; H^f)$$

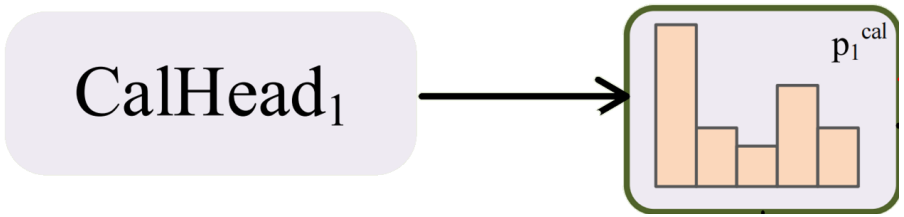


Calibration Heads

The primary task of the calibration heads is to learn a probability distribution for each single modality and to perceive the learning status of the current modality. Samples processed by the Calibration Heads are fed into a filter for screening to identify high-quality candidates for the pseudo-label set.

$$\mathcal{L}_{caliH} = -\frac{1}{B}\sum_{C}\sum_{x_i \in Q_c}\sum_{m=1}^{M}\hat{q}_c \log(p_{i,m}^{cal})$$

$$\mathcal{L}_{re} = \frac{1}{M}\sum_{m=1}^{M}\bar{p}_m^{cal}\log(\bar{p}_m^{cal})$$



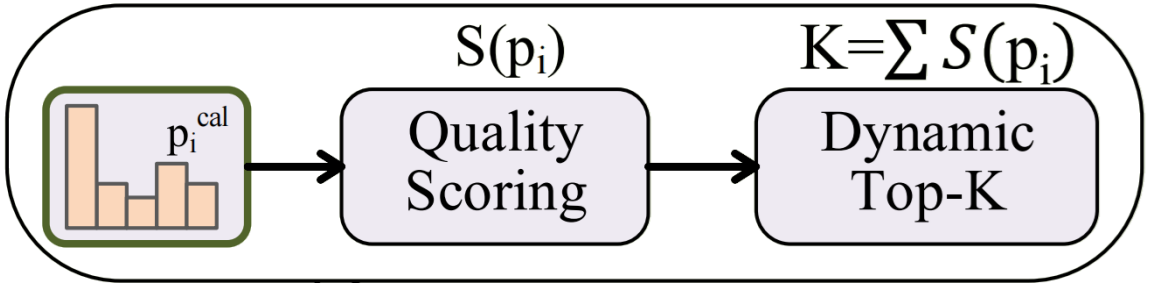
Pseudo-label Screening Mechanism

We design a dynamic screening mechanism based on information redundancy to select high-quality pseudo-labels. By filtering ambiguous samples, the model prioritizes reliable structures and avoids premature, overconfident errors on difficult data. Based on the scores from the quality function $S(\cdot)$, the top K best-performing samples within each modality are selected for the pseudo-label set.

$$S(P) = 1 - \frac{H(P)}{2 \times H_{\max}}$$

$$H(P) = -\sum_{i=1}^{c} p_i \log_b(p_i)$$

$$H_{\max} = H(P_{even}) \qquad P_{even} = (\tfrac{1}{D}, \tfrac{1}{D}, \ldots, \tfrac{1}{D})$$

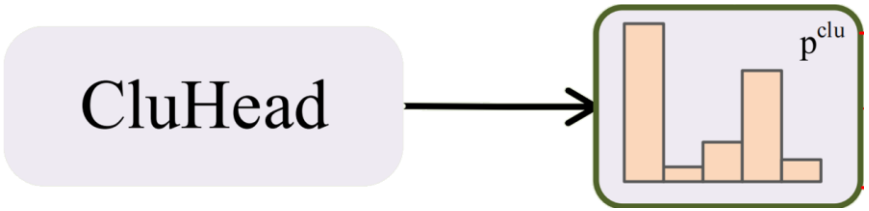


Cluster Head

The pseudo-label set is utilized for training the cluster head, where the backpropagated gradients serve to rectify the bias in Mutual Information estimation within the Information Bottleneck. By calculating the KL divergence, the model is forced to output a flatter probability distribution when modalities yield conflicting opinions, thereby honestly expressing its uncertainty.

$$\mathcal{L}_{cluH} = -\frac{1}{|S|} \sum y_{i,m} \log(p_i^{clu})$$

$$\mathcal{L}_{con} = \sum_{m=1}^{M} D_{KL}(p_{:,m}^{cal} \parallel p^{clu})$$



Objective Function

Total objective function is: $\mathcal{L}_{total} = \mathcal{L}_{IB} + \alpha \mathcal{L}_{Cal}$

$$\mathcal{L}_{IB} = \mathcal{L}_C - \beta \mathcal{L}_P = \sum_{i=1}^{M} I\left(X^i, H^i\right) - \sum_{i=1}^{M} \sum_{j=i+1}^{M} I\left(H^i, H^j\right) - \beta \sum_{m=1}^{M} I(H^m; H^f)$$

$$\mathcal{L}_{Cal} = \mathcal{L}_{caliH} + \mathcal{L}_{re} + \mathcal{L}_{cluH} + \mathcal{L}_{con}$$

α denotes the **balance parameter** between feature extraction and calibration.

β denotes the **balance parameter** trading off the information com-pression and preservation

Advantages of the CLIB

- Achieves trustworthy clustering with low ECE.
- Rectifies biased MI estimation via cross-modal calibration.
- Dynamically filters high-quality pseudo-labels.
- Self-supervision mechanism.

Outline

- Problem background
- Previous works
- Our proposal
- **Experiments**
- Conclusion

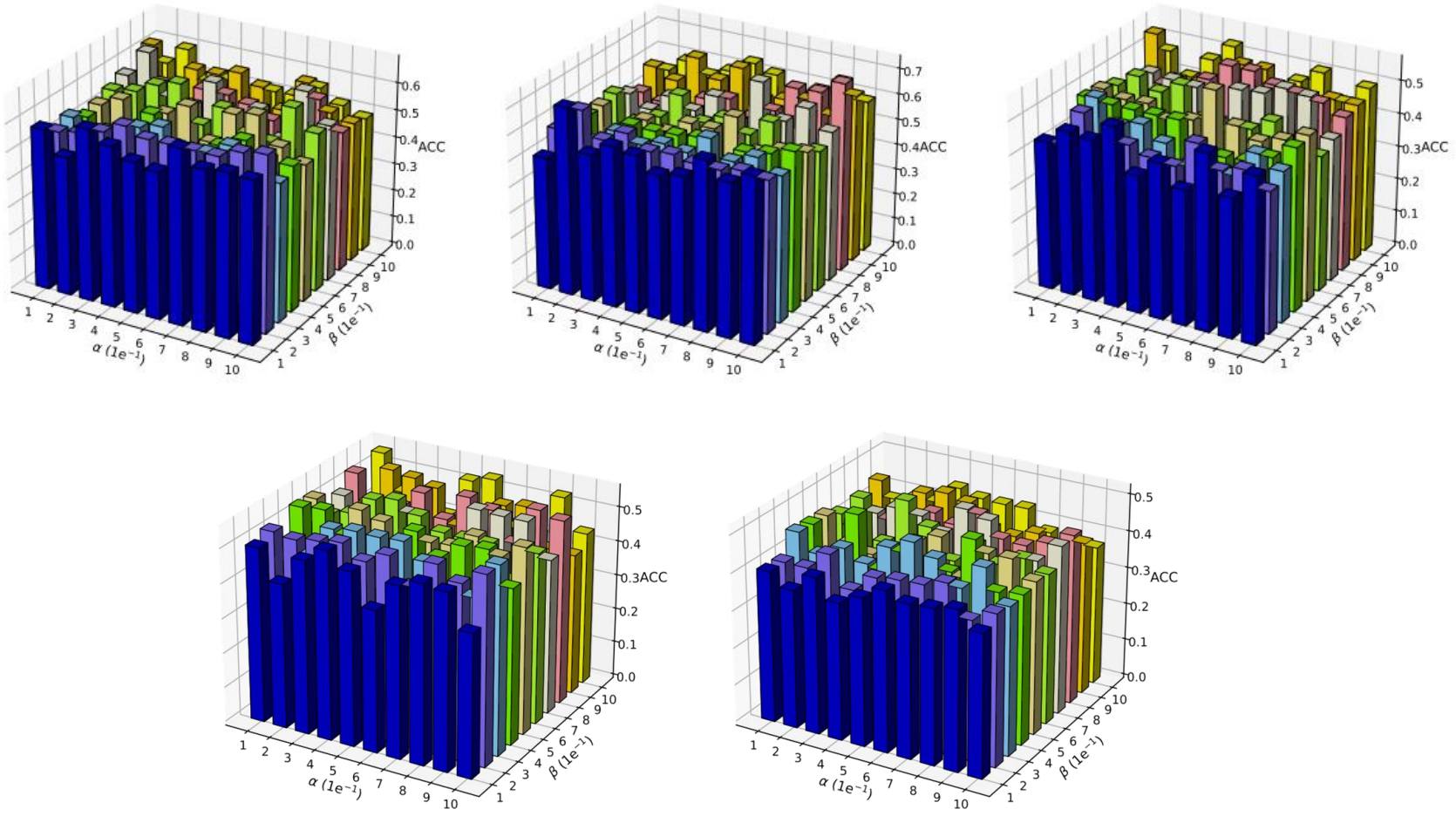
Datasets

Dataset	Modalities	Clusters	Samples
Caltech-2V	2	7	1,440
Caltech-3V	3	7	1,440
ESP-Game	2	7	11,032
IAPR	2	6	7,855
MIRFlickr	2	6	12,154

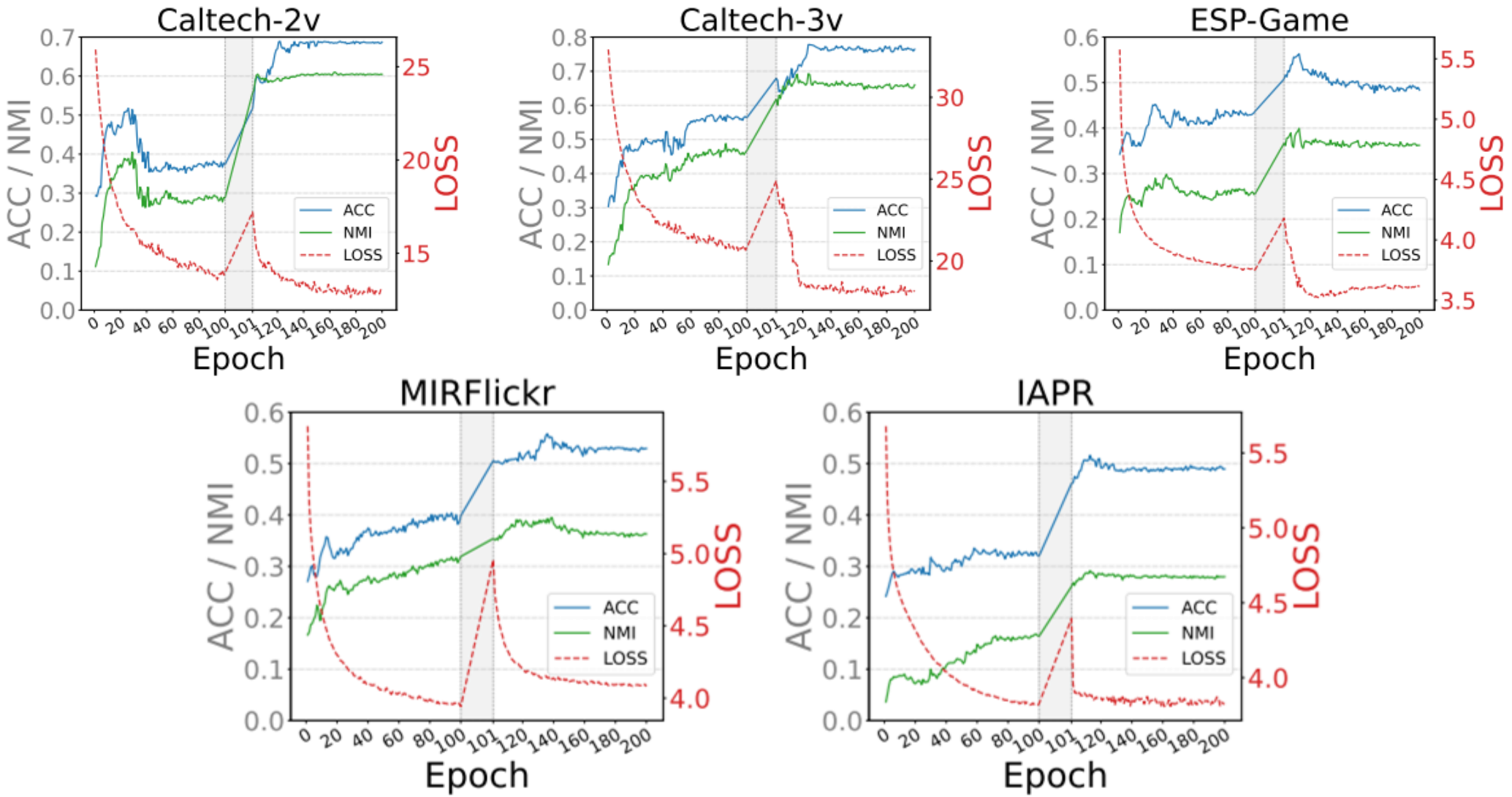
Clustering results

	Caltech-2V			Caltech-3V			ESP-Game			MIRFlickr			IAPR		
	ACC	NMI	ECE	ACC	NMI	ECE	ACC	NMI	ECE	ACC	NMI	ECE	ACC	NMI	ECE
KM	41.6	30.5	N/A	46.3	31.3	N/A	48.4	33.5	N/A	40.9	22.5	N/A	38.9	17.2	N/A
Ncuts	39.9	31.2	N/A	42.6	25.4	N/A	46.5	29.9	N/A	48.4	26.1	N/A	41.9	18.9	N/A
ALLKM	46.4	31.4	N/A	46.9	31.5	N/A	34.9	20.3	N/A	41.0	21.6	N/A	40.4	17.0	N/A
ALLNcuts	42.8	5.2	N/A	43.7	25.5	N/A	33.6	18.9	N/A	48.2	26.2	N/A	42.2	18.9	N/A
EAMC (CVPR'20)	40.3	26.6	35.7	38.9	21.4	19.6	27.1	6.5	24.1	30.5	9.1	29.7	37.1	16.4	24.8
SiMVC (CVPR'21)	51.1	36.9	30.2	56.9	50.4	21.3	35.3	16.2	25.6	45.6	26.3	35.4	42.7	18.5	47.9
CoMVC (CVPR'21)	59.2	49.2	38.7	54.1	50.4	25.6	51.8	38.2	32.5	49.3	30.6	43.2	46.7	21.5	36.8
DSMVC(CVPR'22)	57.9	49.8	31.2	65.7	54.2	26.4	32.4	27.5	31.4	48.4	29.5	31.8	38.9	15.3	24.6
MFLVC (CVPR'22)	61.5	53.6	30.3	63.1	56.6	34.2	52.1	**40.1**	21.7	53.8	32.8	19.5	47.3	22.6	23.6
DealMVC (ACM MM'23)	47.6	37.9	19.1	59.2	56.2	27.6	42.7	24.7	23.5	49.3	32.1	21.9	35	10.8	25.8
ICMVC (AAAI'24)	49.6	37.9	28.5	64.7	53.7	39.4	45.8	29.5	25.2	43.5	24.4	20.6	37.1	16.8	43.8
DIVIDE (AAAI'24)	64.1	52.9	N/A	71.6	58.5	N/A	46.5	27	N/A	52.3	33.5	N/A	45.6	23	N/A
MVCAN (CVPR'24)	46.5	37.5	23.5	70.1	62.6	24.1	50.2	35.5	33.5	50.5	31.5	26.3	34.4	17.2	22.4
ROLL (CVPR'25)	45.7	35.8	19.7	61.8	49.6	16.9	41.5	23.6	25.7	46.2	27.3	24.6	38.2	17.4	23.1
COPER (ICLR'25)	61.5	55.1	40.1	63.9	56.7	32.4	50.7	32.2	29.6	46.4	34.1	31.5	47.2	25.5	28.4
CLIB (Ours)	**68.6**	**60.5**	**16.0**	**77.8**	**69.3**	**10.9**	**56.3**	39.9	**12.1**	**55.4**	**39.5**	**10.5**	**51.6**	**29.1**	**7.8**

Parameter analysis of CLIB on five datasets



Training process of CLIB on five datasets.



The ablation results

Settings	Caltech-2V			Caltech-3V			ESP-Game			MIRFlickr			IAPR		
	ACC	NMI	ECE	ACC	NMI	ECE	ACC	NMI	ECE	ACC	NMI	ECE	ACC	NMI	ECE
$\mathcal{L}_{IB} + \mathcal{L}_{cluH}$	57.7	43.1	36.2	60.6	52.5	25.5	43.8	31.2	21.3	51.0	34.1	25.6	39.1	18.9	35.4
$\mathcal{L}_{IB} + \mathcal{L}_{cluH} + \mathcal{L}_{re}$	61.2	50.1	20.0	64.6	57.4	19.5	45.6	35.9	22.6	51.8	35.4	23.2	42.3	18.2	27.4
$\mathcal{L}_{IB} + \mathcal{L}_{cluH} + \mathcal{L}_{caliH}$	61.7	53.5	19.2	67.8	59.3	11.6	46.1	35.3	21.5	53.8	36.2	18.5	39.9	20.3	18.9
$\mathcal{L}_{IB} + \mathcal{L}_{cluH} + \mathcal{L}_{con}$	59.7	47.8	19.8	61.1	56.7	12.4	44.2	31.0	16.3	51.5	35.7	16.1	37.8	19.1	15.0
I. No warm-up for IB	59.6	57.1	19.4	61.7	58.4	15.1	48.6	34.6	21.9	52.8	38.8	27.8	40.1	22.4	32.9
II. Cali heads backprop	59.9	59.3	23.4	65.9	58.8	13.5	48.5	33.4	24.3	50.4	36.2	31.3	45.4	24.4	24.3
CLIB	**68.6**	**60.5**	**16.0**	**77.8**	**69.3**	**10.9**	**56.3**	**39.9**	**12.1**	**55.4**	**39.5**	**10.5**	**51.6**	**29.1**	**7.8**

The experiments validate the significant contribution of each component and design in the proposed CLIB to the final clustering performance, fully proving its effectiveness.

Outline

- Problem background
- Previous works
- Our proposal
- Experiments
- **Conclusion**

Summary

- Proposes CLIB, a novel multi-modal clustering method based on the IB theory;
- Overcomes the negative impact of Mutual Information (MI) estimation bias in high-dimensional data;
- Introduces a dynamic screening mechanism to learn from high-quality pseudo-labels, which in turn calibrates the IB to enhance the purity of feature extraction;
- Our approach achieves state-of-the-art performance.

Thank You!

Contact for communication:

ieshizhehu@gmail.com