CMoB: Modality Valuation via Causal Effect for Balanced Multimodal Learning

Jun Wang[1], Fuyuan Cao[1, 2], Zhixin Xue[1], Xingwang Zhao[1], Jiye Liang[1]

[1]Shanxi University [2] Shanxi Taihang Laboratory

NEURAL INFORMATION PROCESSING SYSTEMS

Modality Valuation via Causal Effect for Balanced Multimodal Learning

SHANXI UNIVERSITY 1902

Motivation:

Existing research in **Cognitive Science** [1,2] (PNAS, Nature) demonstrates that when the number of modalities within a sample space increases, the human brain can **dynamically assess** and extract discriminative feature information from heterogeneous multimodal data, continuously **refining Cognitive Processes through adaptive learning.** This aligns with the two core elements of our method:

- Modality Contribution Valuation.
- Granular Adjustment at the Sample Level.

Our research aims to establish a multimodal learning framework that **computationally formalizes this Cognitive Processes** to mitigate the modality imbalance problem.

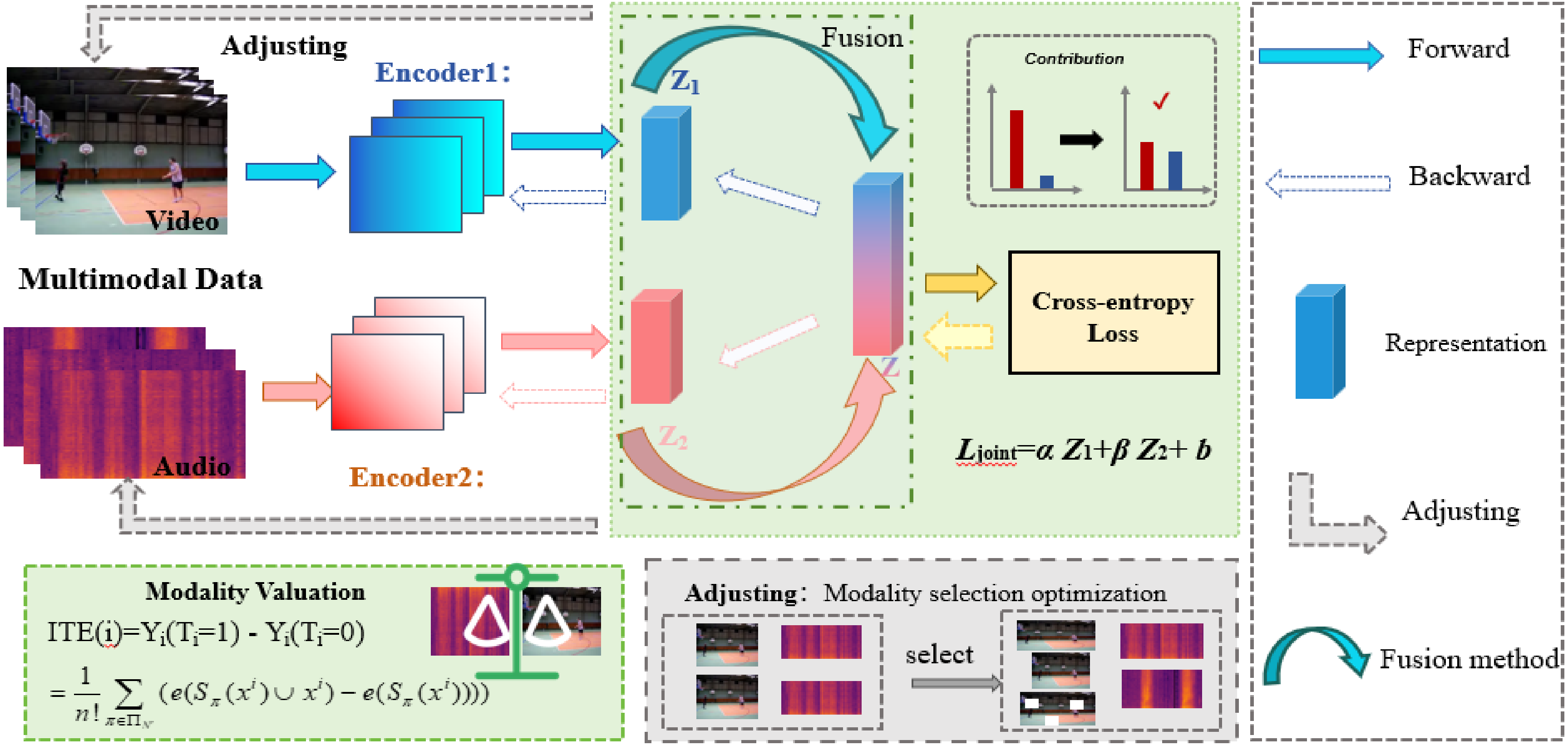
[1] Rideaux R, et al. How multisensory neurons solve causal inference[J]. **Proceedings of the National Academy of Sciences**, 2021, 118(32): e2106235118.
[2] Khilkevich A, et al. Brain-wide dynamics linking sensation to action during decision-making[J]. **Nature**, 2024, 634(8035): 890-900.

Modality Valuation via Causal Effect for Balanced Multimodal Learning

Our Proposal:

- We evaluate the importance of sample in multimodal learning by means of a benefit function designed by information uncertainty theory.
- We propose methods to quantify the degree of contribution of each modality from a causal perspective and represent the contribution of modalities in terms at the sample-level.
- We propose a modality balancing approach from a data perspective to improve the performance of multimodal learning by optimizing the selection of weak modalities from the sample level in terms of modality contribution.

The pipeline of CMoB method



Methodology

■ Sample Benefit Valuation

According to Shannon's theory of information uncertainty, "**the essence of information is to eliminate uncertainty**" brings us new thinking. In other words, adding more modalities to a sample is accompanied by enhancing more information, which will bring less uncertainty to the multimodal model. We can approximate this relationship as follow:

Proposition 1. Given a sample has M modalities denoted as $x^{(M)} = \{x^1, x^2, \cdots, x^M\}$, assume that there any two subsets $x^{(B)}, x^{(C)}$of $x^{(M)}$, and $B \subseteq C \subseteq M$, then for any multimodal classifier $\hat{F}(\cdot)$, it should be guaranteed that

$$\mathrm{Conf}(\hat{F}(x^{(B)})) < \mathrm{Conf}(\hat{F}(x^{(C)})) \leq \mathrm{Conf}(\hat{F}(x^{(M)})).$$

We defined a benefit function to evaluate the importance of the samples in the multimodal model learning process as follow:

$$B(M) = \begin{cases} |M|, & \text{if } \hat{F} = y_i \text{ and } \mathrm{Conf}(\hat{F}(x^{(C)})) \leq \mathrm{Conf}((\hat{F}(x^{(M)}) \text{ and } C \leq M, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Methodology

■ Causal-aware Quantification Method

We measure the ITE of modality j in sample i. The modality of the treatment variable is denoted as x_i^j,the control group is $S(x_i)$, and the treatment group is $S(x_i)\backslash x_i^j$ then its individual causal effect can be expressed as follows:

$$
\begin{aligned}
ITE(x_i^j) = \phi_H[x_i^j] &= \mathcal{V}(H(T)) - \mathcal{V}(H(T')) \\
&= \mathcal{V}(H(S(x_i))) - \mathcal{V}(H(S(x_i)\backslash x_i^j)) \\
&= B(\hat{F}(S(x_i))) - B(\hat{F}(S(x_i)\backslash x_i^j)) \\
&= B(\hat{F}(S(x_i))) - B(\hat{F}(S(x_i)|do(t_i = x_i^j))),
\end{aligned}
$$

When quantifying the contribution of the modality j in the sample i, we must not only calculate the effect of the intervention modality j on the output but also consider its own impact on the output. Its contribution can be expressed as follows:

$$
\begin{aligned}
\Phi(x_i^j) &= ITE(x_i^j) + \mathcal{V}(H(x_i^j) \\
&= B(\hat{F}(S(x_i))) - B(\hat{F}(S(x_i)|do(t_i = x_i^j))) + B(\hat{F}(x_i^j).
\end{aligned}
$$

NEURAL INFORMATION PROCESSING SYSTEMS

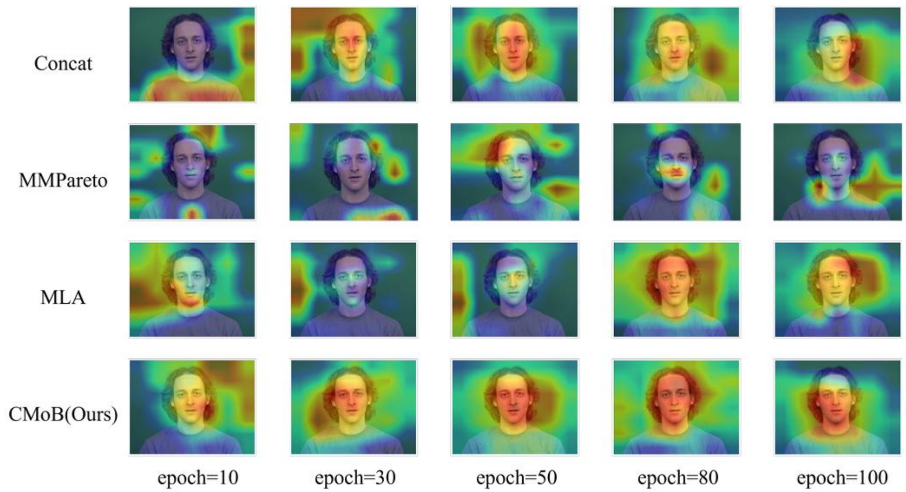
Modality Valuation via Causal Effect for Balanced Multimodal Learning

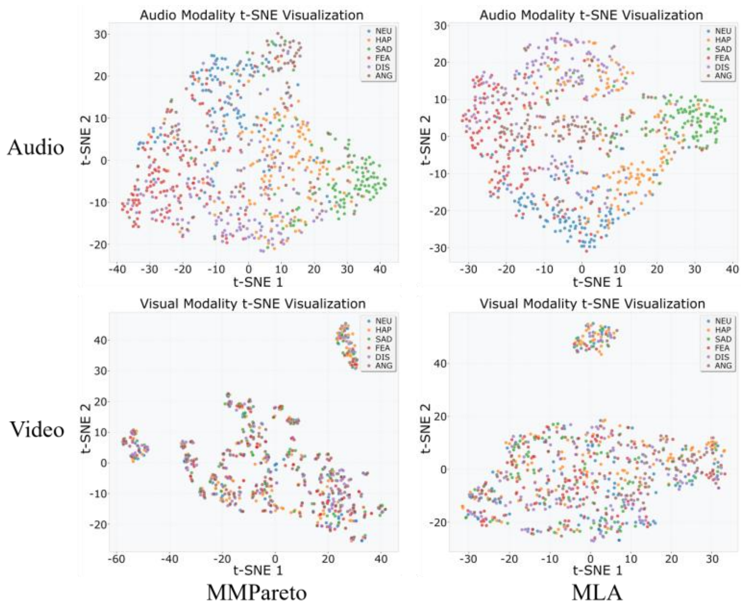
Experiments and Results:

Method	CREMA-D		KineticsSounds		UCF-101		CMU-MOSEI		NVGesture	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1
Unimodal-1	61.17	60.63	55.06	54.96	78.60	77.49	71.09	41.7	78.22	78.33
Unimodal-2	49.56	47.81	45.31	43.76	59.90	58.19	71.03	41.68	78.63	78.65
Unimodal-3	-	-	-	-	-	-	80.58	74.57	81.54	81.83
Concat	65.5	65.07	65.63	65.28	81.8	81.21	78.99	69.40	81.33	81.47
Sum	63.44	63.12	64.97	64.72	80.21	79.42	79.10	71.15	82.99	83.05
Weight	66.53	66.41	65.33	64.89	82.65	82.19	79.94	72.31	82.42	82.57
MMCosine	67.19	67.34	67.49	67.09	82.97	82.47	80.38	73.67	81.52	81.55
AGM	71.59	72.11	66.62	65.88	81.7	80.89	79.86	71.89	82.78	82.82
OGM	67.76	68.02	67.04	66.95	82.07	81.3	-	-	-	-
GBlending	71.59	71.72	68.82	66.43	85.01	84.5	79.64	73.29	82.33	82.91
PMR	67.19	67.20	67.11	66.87	81.93	81.48	-	-	-	-
MMCooperation	75.85	76.68	68.01	68.03	85.25	84.69	79.84	72.99	82.85	83.02
Relearning	71.02	71.46	65.92	65.48	82.87	82.15	78.75	70.02	82.87	82.94
MLA	79.43	79.90	69.05	68.75	85.38	84.84	78.65	70.02	83.73	83.87
MMPareto	76.87	77.35	**74.55**	**74.21**	85.3	84.89	81.18	74.64	83.82	**84.24**
CMoB	**79.75** ±0.27%	**79.98** ±0.38%	72.03 ±0.22%	71.74 ±0.32%	**86.82** ±0.27%	**86.21** ±0.34%	**81.24** ±0.19%	**74.97** ±0.26%	**84.06** ±0.14%	84.18 ±0.21%

Modality Valuation via Causal Effect for Balanced Multimodal Learning

Experiments and Results:





Modality Valuation via Causal Effect for Balanced Multimodal Learning

Summary:

- We propose a causal-aware modality validation approach for balanced multimodal learning.
- We employ intervention methods to evaluate the causal effect, quantifying changes in modality contributions at the sample level during multimodal learning.
- The fine-grained evaluation approach enables targeted optimizations across modalities at the sample level, effectively mitigating the issue of multimodal imbalance.

CMoB: Modality Valuation via Causal Effect for Balanced Multimodal Learning

Thank You !!