

Towards the Causal Complete Cause of Multi-Modal Representation Learning

王婧瑶*, 赵思雨*, 强文文, 李江梦, 郑昌文, 孙富春, 熊辉

International Conference on Machine Learning

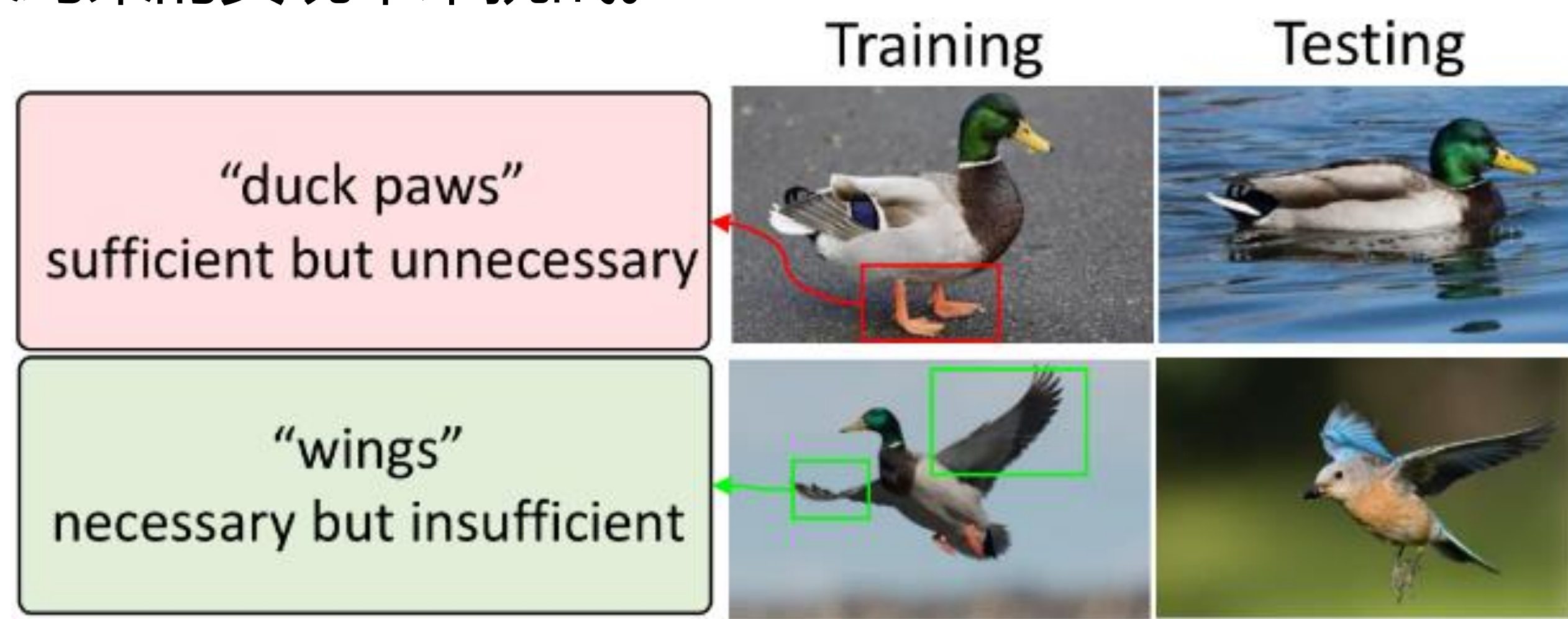
王婧瑶 Jingyao_wang0728@163.com

概述

多模态表征应在因果意义上同时满足充分性与必要性。为解决虚假关联与模态冲突问题，我们提出因果完备原因 (C^3)，并在放宽外生性与单调性假设下引入工具变量以实现其可识别性。基于此，设计双分支网络估计 C^3 风险，并通过正则化策略优化表征学习。理论与实验证明该方法有效提升多模态模型的因果鲁棒性与下游性能。

动机与分析

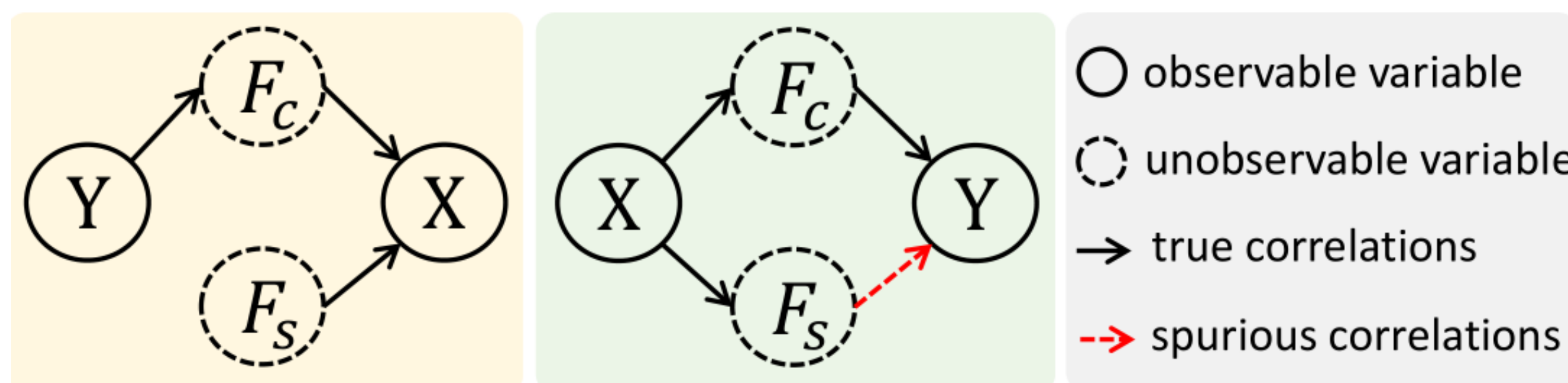
人类感知依赖多模态融合，多模态学习 (MML) 由此提出，以提取共享语义或保留模态特异性。然而，这些方法常忽视因果充分性与必要性，导致表征含虚假或冗余信息。仅满足充分性会损害泛化，强调必要性则易受背景干扰。我们提出真正有效的多模态表征应具备因果完备性，即同时满足因果充分性与必要性。然而，现实中的模态冲突与非线性交互违背外生性与单调性假设，给因果约束的实现带来挑战。



图一 因果充分但不必要特征可以是 "duck paws"，当样本中出现 "duck paws" 时可以得到标签 "duck"，但在 "duck swim on the lake" 的示例中没有 "duck paws" (Line 1)；因果必要但不充分特征可以是 "wings"，因为标签 "duck" 对应样本一定有 "wings"，但有 "wings" 不一定得到 "duck"，e.g., "bird" (Line 2)。

方法

为解决外生性与单调性假设受限的问题，本文放宽相关约束，提出因果完备原因 (C^3)，以衡量多模态表征的因果充分性与必要性。我们通过干预表征并观测标签变化，定义 C^3 并分析其可识别性，结合工具变量实现在弱假设下的无偏估计。进一步构建孪生网络度量 C^3 风险：真实分支剔除虚假关联，虚假分支生成反事实样本。理论证明该方法可靠，并据此提出 C^3 正则化策略，促进因果完备表征学习。



图二 结构因果图。左侧表示MML样本X是如何基于因果生成机制生成的；右侧则与实际的MML过程保持一致，其中因子通常是耦合的，用于预测Y。学习到的 MML 表示可能分为四种情况：(i) 充分且必要：包含所有；(ii) 必要但不充分：包含部分；(iii) 充分但不必要：包含所有带有的；(iv) 不充分且不必要：仅包含。

实验

为验证方法的有效性与鲁棒性，本文在六个多模态数据集和十五种以上基线上进行实验。结果表明，在高斯与文本噪声干扰下， C^3R 显著提升了各类基线模型的平均与最坏情况精度，表现出良好稳定性。在BraTS数据集中，面对15种模态缺失组合， C^3R 不仅提升了整体分割精度，还有效缩小了不同模态间的性能差距，验证了其在模态缺失下的因果鲁棒性。

Method	NYU Depth V2				SUN RGB-D				FOOD 101				MVSA			
	(0,Avg.)	(0,Worst.)	(10,Avg.)	(10,Worst.)	(0,Avg.)	(0,Worst.)	(10,Avg.)	(10,Worst.)	(0,Avg.)	(0,Worst.)	(10,Avg.)	(10,Worst.)	(0,Avg.)	(0,Worst.)	(10,Avg.)	(10,Worst.)
CLIP (Sun et al., 2023)	69.32	68.29	51.67	48.54	56.24	54.73	35.65	32.76	85.24	84.20	52.12	49.31	62.48	61.22	31.64	28.27
ALIGN (Jia et al., 2021)	66.43	64.33	45.24	42.42	57.32	56.26	38.43	35.13	86.14	85.00	53.21	50.85	63.25	62.69	30.55	26.44
MaPLE (Khatuk, et al., 2023)	71.26	69.27	52.98	48.73	62.44	61.76	34.51	30.29	90.40	86.28	53.16	40.21	77.43	75.36	43.72	38.82
GoOp (Jia et al., 2022a)	67.48	66.94	49.43	45.62	58.36	56.31	39.67	35.43	88.33	85.10	55.24	51.01	74.26	73.61	42.58	37.29
VPT (Jia et al., 2022a)	62.16	61.21	41.05	37.81	54.72	53.92	33.48	29.81	83.89	82.00	51.44	49.01	65.87	64.98	32.79	29.21
Late fusion (Wang et al., 2016)	69.14	68.35	51.99	44.95	62.09	60.55	47.33	44.60	90.69	90.58	58.00	55.77	76.88	74.76	55.16	47.78
ConcatMML (Zhang et al., 2021)	70.30	69.42	53.20	47.71	61.90	61.19	45.64	42.95	89.43	88.79	56.02	54.33	75.42	75.33	53.42	50.47
AlignMML (Wang et al., 2016)	70.31	68.50	51.74	44.19	61.12	60.12	44.19	38.12	88.26	88.11	55.47	52.76	74.91	72.97	52.71	47.03
ConcatFlow (Zhang et al., 2023c)	49.64	48.66	31.43	29.87	41.25	40.54	26.76	24.27	70.77	70.68	35.68	34.92	64.09	62.04	45.40	40.95
ConcatBERT (Zhang et al., 2023c)	70.56	69.83	44.52	43.29	59.76	58.92	45.85	41.76	88.20	87.81	49.86	47.79	65.59	64.74	46.12	41.81
MMTM (Jozic et al., 2020)	71.04	70.18	52.28	46.18	61.72	60.94	46.03	44.28	89.75	89.43	57.91	54.98	74.24	73.55	54.63	49.72
TMC (Han et al., 2020)	71.06	69.57	53.36	49.23	60.68	60.31	45.66	41.60	89.86	89.80	61.37	61.10	74.88	71.10	60.36	53.37
LCKD (Wang et al., 2023b)	68.01	66.15	42.31	40.56	56.43	56.32	43.21	42.43	85.32	84.26	47.43	44.22	62.44	62.27	43.52	38.63
UniCODE (Xia et al., 2023a)	70.12	68.74	44.78	42.79	59.21	58.55	46.32	42.21	88.39	87.21	51.28	47.95	66.97	65.94	48.34	42.95
SimMMDG+ C^3R	71.34	70.29	45.67	44.83	60.54	60.31	47.86	45.79	89.57	88.43	52.55	50.31	67.08	66.35	49.52	44.01
MMBT (Kielic et al., 2019)	67.00	65.84	49.59	47.24	56.91	56.18	43.28	39.46	91.52	91.38	56.75	56.21	78.50	78.04	55.35	52.22
QMF (Zhang et al., 2023c)	70.09	68.81	55.60	51.07	62.09	61.30	48.58	47.50	92.92	92.72	62.21	61.76	78.07	76.30	61.28	57.61
CLIP+ C^3R	76.54	75.12	56.73	52.90	62.31	58.71	41.59	37.52	92.93	91.80	59.77	57.54	69.61	68.64	39.58	35.89
MaPLE+ C^3R	77.07	74.45	58.94	55.95	66.21	65.51	40.12	37.34	94.38	93.51	60.63	46.07	81.19	81.51	49.32	45.98
Late fusion+ C^3R	73.26	71.62	57.21	50.98	64.84	63.25	53.35	50.43	94.09	92.24	65.27	59.02	83.77	79.79	62.14	52.50
LCKD+ C^3R	77.14	75.12	50.11	47.98	60.97	60.14	47.23	46.21	90.89	90.14	54.48	51.16	66.78	65.67	49.28	42.84
SimMMDG+ C^3R	75.32	74.61	49.99	47.22	65.50	64.58	52.69	51.70	92.24	91.14	57.32	53.56	73.62	71.01	51.65	51.07
MMBT+ C^3R	73.74	71.82	54.35	52.57	61.47	59.99	48.42	46.07	94.25	93.90	60.41	60.11	82.76	81.64	62.12	58.93
QMF+ C^3R	77.58	74.95	59.72	59.18	67.35	65.84	52.26	51.28	94.87	93.79	66.45	63.69	83.13	81.98	66.66	64.51

表一 当 50% 样本受高斯噪声影响时的性能比较。“(N, Avg.)”和“(N, Worst.)”分别表示平均准确率和最差准确率。最佳结果以粗体突出显示。

Modalities	Enhancing Tumour					Tumour Core					Whole Tumour										
	F1	T1	T2	HMS	HVED	RSeg	mmFm	LCKD	LCKD+ C^3R	HMS	HVED	RSeg	mmFm	LCKD	LCKD+ C^3R	HMS	HVED	RSeg	mmFm	LCKD	LCKD+ C^3R
•••••	11.78	23.80	25.69	39.33	45.48	49.81	(+4.33)	26.06	57.90	53.57	61.21	72.01	76.65	(+4.64)	52.48	84.39	85.69	86.10	89.45	91.62	(+2.17)
•••••	10.16	8.60	17.29	32.53	43.22	49.13	(+6.01)	37.39	33.90	47.90	56.55	66.58	72.18	(+5.60)	57.62	49.51	70.11	67.52	76.48	82.39	(+5.91)
•••••	62.02	57.64	67.07	72.60	75.65	80.50	(+4.85)	65.29	59.59	76.83	75.41	83.02	88.06	(+5.04)	61.53	53.62	73.31	72.22	77.23	81.93	(+4.70)
•••••	25.63	22.82	28.97	43.05	47.19	54.13	(+6.94)	57.20	54.67	57.49	64.20	70.17	77.32	(+7.15)	80.96	79.83	82.24	81.15	84.37	90.78	(+6.41)
•••••	10.71	27.96	32.13	42.96	48.30	54.16	(+4.86)	41.12	61.14	60.68	65.91	74.58	79.83	(+5.25)	64.62	85.71	88.24	81.15	89.97	93.63	(+3.66)
•••••	66.10	68.36	70.30	75.07	78.75	82.98	(+4.23)	71.49	75.07	80.62	77.88	85.67	89.74	(+4.07)	68.99	85.93	88.51	87.06	90.47	93.91	(+3.44)
•••••	30.22	32.31	33.84	47.52	49.01	56.12	(+7.11)	57.68	62.70	61.16	69.75	75.41	82.57	(+7.16)	68.95	87.58	88.28	87.59	90.39	95.48	(+5.09)
•••••	66.22	61.11	69.06	74.04	76.09	81.76	(+5.67)	72.46	67.55	78.72	78.59	82.49	88.32	(+5.83)	82.47	64.22	77.18	74.42	80.10	87.03	(+6.96)
•••••	32.39	24.29	32.01	44.99	50.09	56.03	(+5.94)	60.92	56.26	62.19	69.42	72.75	78.78	(+6.03)	82.41	81.56	84.78	82.20	86.05	92.33	(+6.28)
•••••	67.83	67.83	69.71	74.51	76.01	83.97	(+7.96)	76.64	73.92	80.20	78.61	84.85	93.57	(+8.72)	82.48	81.32	85.19	82.99	86.49	94.40	(+7.91)
•••••	68.54	68.60	70.78	75.47	77.78	82.94	(+5.06)	76.01	77.05	81.06	79.80	85.24	90.46	(+5.22)	72.31	86.72	88.73	87.33	90.50	95.23	(+5.01)
•••••	31.07	32.34	36.41	47.70	49.96	56.25	(+6.29)	60.32	63.14	64.38	71.52	76.68	82.69	(+6.01)	83.43	86.72	88.81	87.75	90.46	96.23	(+5.77)
•••••	68.72	68.93	70.88	75.67	77.48	83.90	(+6.42)	77.53	76.75	80.72	79.55	85.56	92.39	(+6.83)	83.85	88.09	89.27	88.14	90.90	96.78	(+5.88)
•••••	69.92	67.75	70.10	74.75	77.60	82.54	(+4.94)	78.96	75.28	80.33	80.39	84.02	89.43	(+5.41)	83.94	82.32	86.01	82.71	86.73	91.73	(+5.00)
•••••	70.24	69.03	71.13	77.61	79.33	86.36	(+7.03)	79.48	77.71	80.86	85.78	85.31	91.43	(+6.12)	84.74	88.46	89.45	89.64	90.84	95.41	(+4.57)

表二在 BraTS 上缺失模态时的表现。括号“()”表示引入 C^3R 后的效果变化。“•”和“◦”表示对应模态在测试中的可用和缺失情况。最佳结果以粗体突出显示。