

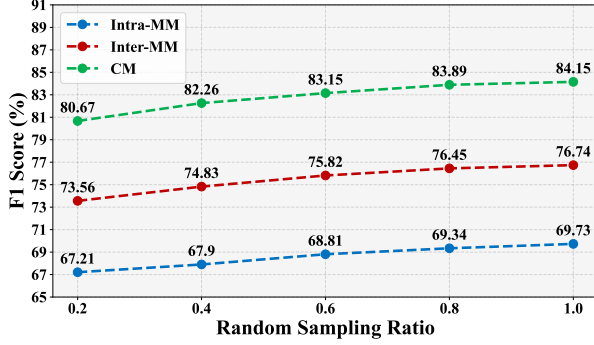
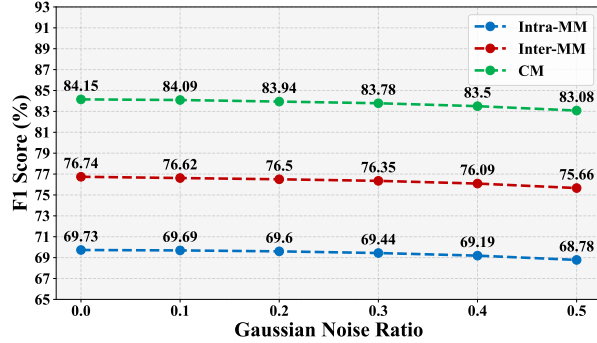
Table 1: Performance comparison among proposed framework and other baselines on different multimodal tasks.

(a) Experimental results on the UR-FUNNY dataset.

Models	Intra-MM	Inter-MM	CM
	Avg. Acc-2 ↑	Avg. Acc-2 ↑	Acc-2 ↑
DMD	53.23	55.49	71.36
SMIL	58.32	60.77	67.51
GCNet	58.90	62.24	69.89
CorrKD	60.72	63.91	70.04
ReFA (Ours)	62.90	66.54	71.12

(b) Experimental results on the MUSTARD dataset.

Models	Intra-MM	Inter-MM	CM
	Avg. Acc-2 ↑	Avg. Acc-2 ↑	Acc-2 ↑
DMD	54.82	57.67	72.89
SMIL	59.17	62.32	69.18
GCNet	60.53	64.01	71.56
CorrKD	61.49	66.02	72.35
ReFA (Ours)	63.21	67.90	73.04

**Fig. 1:** The framework performance on the MOSI dataset with different sampling ratios.**Fig. 2:** The framework performance on the MOSI dataset with different Gaussian noise ratios.**Table 2:** Ablation of different losses on the MOSI dataset.

(a) Experimental results for modality translations

Models	Intra-MM	Inter-MM	CM
	Avg. F1 ↑	Avg. F1 ↑	F1 ↑
ReFA (Full)	69.73	76.74	84.15
w/ Smooth L1 loss	69.35	76.41	83.83
w/ Huber Loss	69.51	76.58	83.98
w/ Cosine Distance	69.17	76.29	83.78

(b) Experimental results for sentiment semantic reconstruction.

Models	Intra-MM	Inter-MM	CM
	Avg. F1 ↑	Avg. F1 ↑	F1 ↑
ReFA (Full)	69.73	76.74	84.15
w/ Smooth L1 loss	69.48	76.46	83.81
w/ Huber Loss	69.56	76.59	84.02
w/ Cosine Distance	69.45	76.32	83.73

E1	True Label: <i>positive</i> Absent Modality: <i>visual</i>	
A	[Visual representation of a person's face]	
V	[Visual representation of a person's face]	
T	The animation was <u>amazing</u> .	
	Self-MM √ (positive)	CubeMLP √ (positive)
	DMD √ (positive)	MCTN √ (positive)
	TransM √ (positive)	SMIL √ (positive)
	GCNet √ (positive)	CorrKD √ (positive)
	ReFA (Ours) √ (positive)	
E2	True Label: <i>neutral</i> Absent Modality: <i>acoustic</i>	
A	[Visual representation of a person's face]	
V	[Visual representation of a person's face]	
T	But I <u>didn't</u> find it all that <u>bad</u> .	
	Self-MM × (negative)	CubeMLP × (negative)
	DMD × (positive)	MCTN × (positive)
	TransM × (positive)	SMIL × (positive)
	GCNet × (positive)	CorrKD × (positive)
	ReFA (Ours) √ (neutral)	

Fig. 3: A case study based on two samples.**Table 3:** Ablation results on the MOSEI and SIMS datasets.

Datasets	Models	Intra-MM	Inter-MM	CM
		Avg. F1 ↑	Avg. F1 ↑	F1 ↑
MOSEI	ReFA (Full)	70.48	75.63	82.93
	w/o FRF	68.23	73.56	81.40
	w/o HMI	68.59	73.89	82.02
	w/o HAL	68.45	73.70	81.66
SIMS	ReFA (Full)	71.56	77.21	83.05
	w/o FRF	69.95	75.84	81.59
	w/o HMI	69.84	75.59	82.56
	w/o HAL	69.52	75.37	82.32

Table 4: Ablation experiments of MRM on the MOSI dataset.

Models	Intra-MM	Inter-MM	CM
	Avg. F1 ↑	Avg. F1 ↑	F1 ↑
ReFA (Full)	69.73	76.74	84.15
w/o Intra-MRM	68.09	75.81	84.04
w/o Inter-MRM	68.62	74.94	83.98

Table 5: Comparison of out ReFA with the larger models.

Models	Intra-MM	Inter-MM	CM
	Avg. F1 ↑	Avg. F1 ↑	F1 ↑
VideoLLaMA	60.55	69.86	80.27
VideoLLaMA 2	63.28	71.32	81.14
ReFA (Ours)	69.73	76.74	84.15