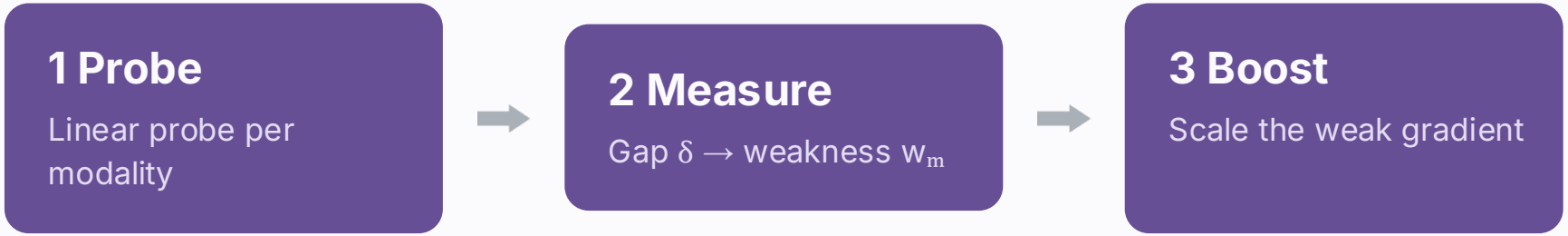
NEURIPS 2026

Probe-Guided Gradient Boosting for Multimodal Learning

Vu Vo Quoc, Haytham M. Fayek, Thuy Nguyen

RMIT University

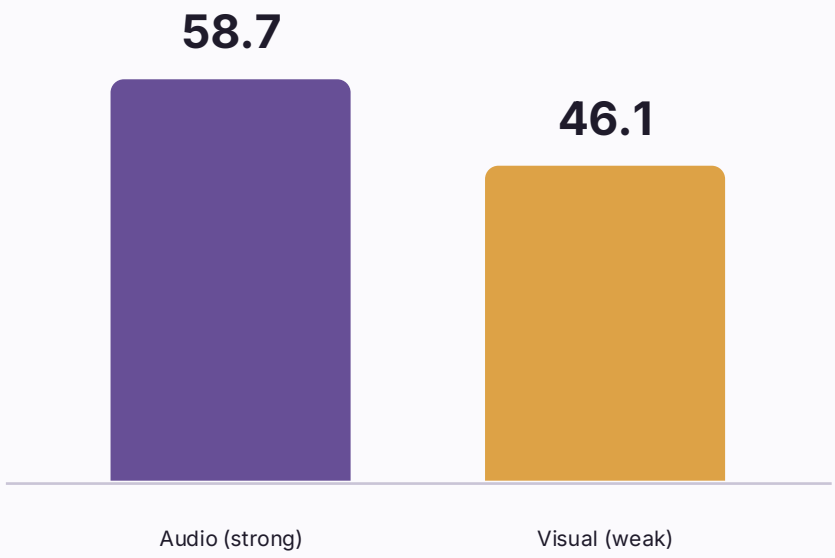


github.com/VuVoCodes/PGGB

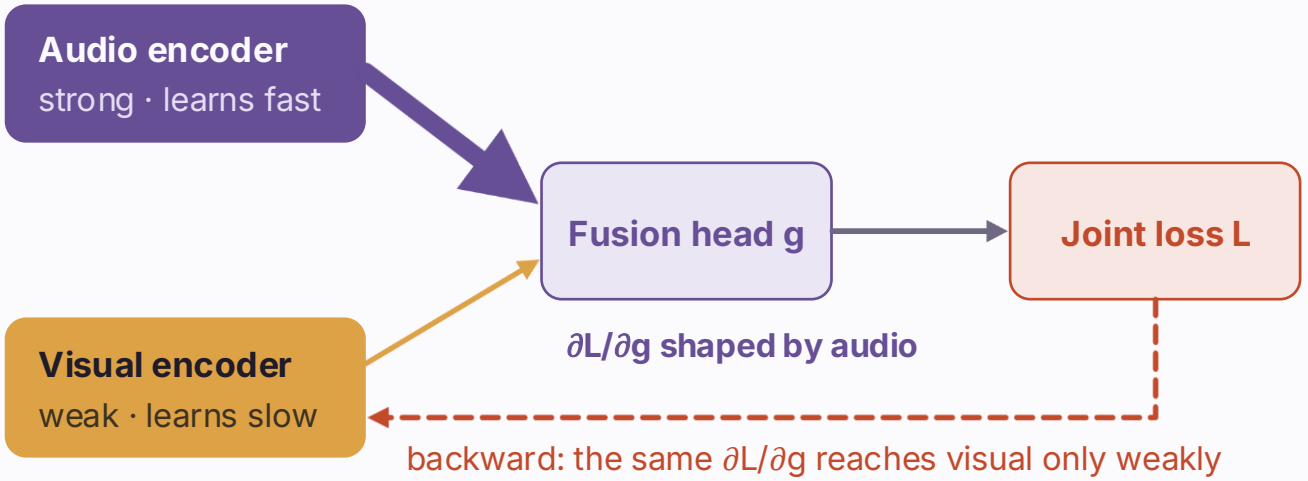
THE PROBLEM

Joint training starves the weak modality

CREMA-D, joint training · linear-probe accuracy (%)



12.5 pp utilization gap under joint training



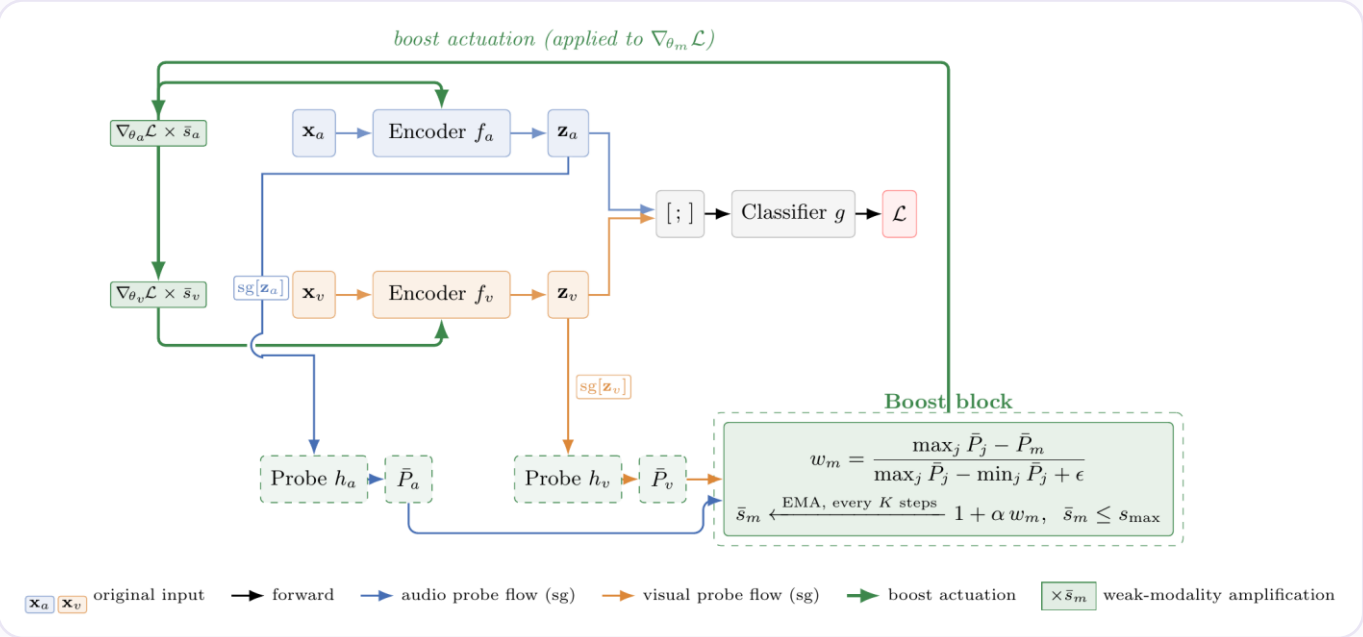
Schematic · solid = forward, width ≈ each modality's pull on the output · dashed = backward gradient

Audio **dominates the fusion loss**. Most fixes throttle audio; boosting visual with a loss-based signal inherits this same bias.

RMIT UNIVERSITY

OUR PROPOSAL: PGGB

Measure outside the loss, then **boost the weak**



1 **Relative weakness from probes**

$$w_m = \frac{\max_j \bar{P}_j - \bar{P}_m}{\max_j \bar{P}_j - \min_j \bar{P}_j + \epsilon}$$

2 **Bounded scale, at most 2×**

$$s_m = \min\left(1 + \alpha \cdot w_m,\; s_{\max}\right)$$

3 **Smooth, then scale the weak gradient**

$$\frac{\partial \mathcal{L}}{\partial \theta_m} \leftarrow \bar{s}_m \cdot \frac{\partial \mathcal{L}}{\partial \theta_m}$$

Probes never backpropagate into encoders

Bounded and self-attenuating

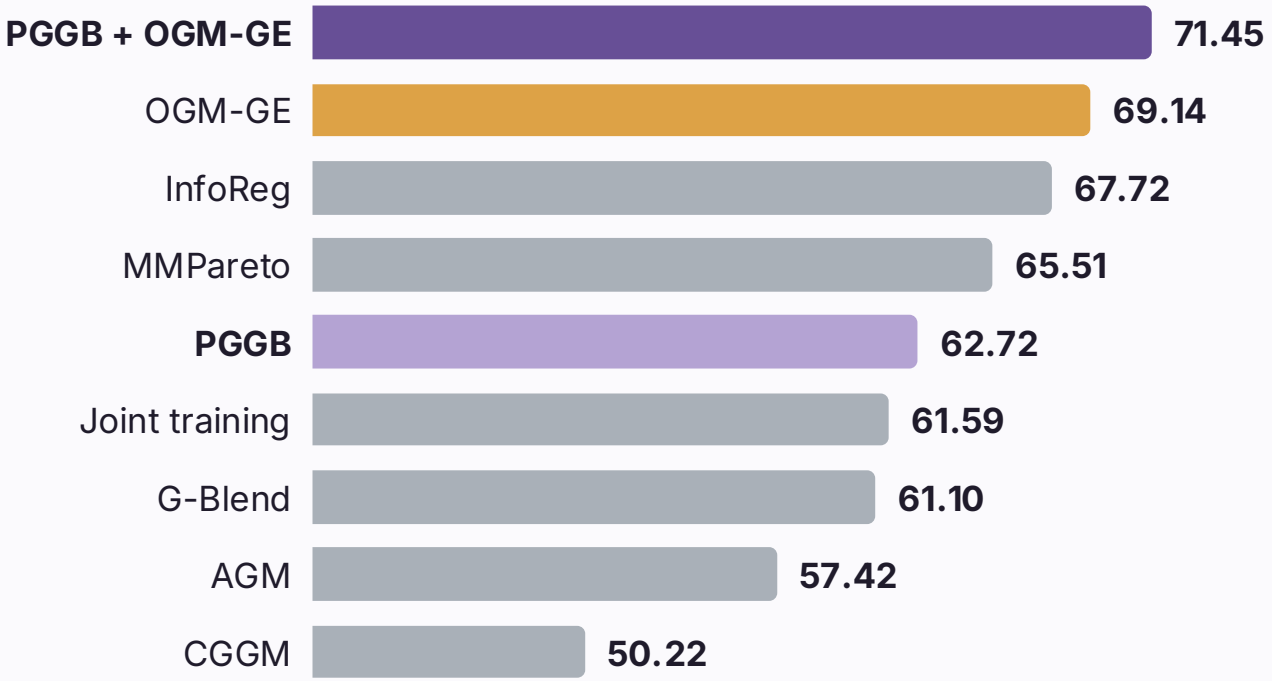
Composes with throttling · ~1% overhead

RMIT UNIVERSITY

RESULTS

PGGB + OGM-GE reaches 71.45% on CREMA-D

CREMA-D accuracy (%), 3-frame, mean of 5 seeds · bars start at 40%



71.45%

CREMA-D accuracy, ours

+2.31 pp

over OGM-GE, the best baseline: the gain from the boost

5.4×

smaller audio–visual probe gap (12.5 → 2.3 pp)

RMIT UNIVERSITY

RESULTS

All eight benchmarks

Method	CREMA-D	AVE	KS	MOSI	Twitter15	Sarcasm	CH-SIMS	BraTS
Joint training	61.59	86.54	79.05	72.42	66.60	82.40	70.42	85.77
OGM-GE	69.14	86.96	77.25	**72.68**	66.33	81.81	**72.47**	86.21
MMPareto	65.51	86.37	78.21	**72.68**	66.81	82.27	70.20	—
AGM	57.42	84.42	77.85	72.50	66.79	82.32	69.28	—
G-Blend	61.10	87.09	77.75	72.36	66.66	82.35	70.15	—
CGGM	50.22	76.72	73.18	59.45	62.41	80.71	68.05	81.13
PGGB	62.72	**87.41**	**79.17**	71.89	**66.85**	**82.44**	69.80	85.98
PGGB + OGM-GE	**71.45**	87.23	77.33	72.60	66.59	81.76	72.43	**86.49**

Accuracy %, Dice % for BraTS 2021 · mean of 5 seeds · **bold** = best per column

4 / 4 low-imbalance sets: PGGB alone is best

≈ best text sentiment: within seed std

86.49% BraTS Dice, +0.72 pp vs. joint

Lift the weak modality, keep the strong

CONCLUSION

Lifts weak modalities without hurting strong ones

Best alone on all 4 low-imbalance benchmarks

Composes with throttling: +2.31 pp on CREMA-D

FUTURE WORK

Adaptive composition: throttle only when the probed gap warrants it

Optimizer-aware boost: make the scaling survive Adam

New tasks: beyond classification and segmentation

Caveat: throttling can hurt from-scratch, low-data training (AVE −4.56 pp)

github.com/VuVoCodes/PGGB

Thank you!

Come find us at the poster