AG-REPA | Author Slide Upload | ICML 2026

AG-REPA: Causal Layer Selection for Representation Alignment in Audio Flow Matching

Pengfei Zhang, Tianxin Xie, Minghao Yang, Li Liu | AI Thrust, Information Hub, HKUST (Guangzhou)

Correspondence: avrillliu@hkust-gz.edu.cn | Code: github.com/zpforlove/AG-REPA

GitHub

Poster summary

Knowing is not doing: align layers that drive generation, not merely those that store teacher-aligned features.

Problem

Fixed-depth REPA can supervise layers that are representation-rich but functionally passive.

Insight: Store-Contribute Dissociation

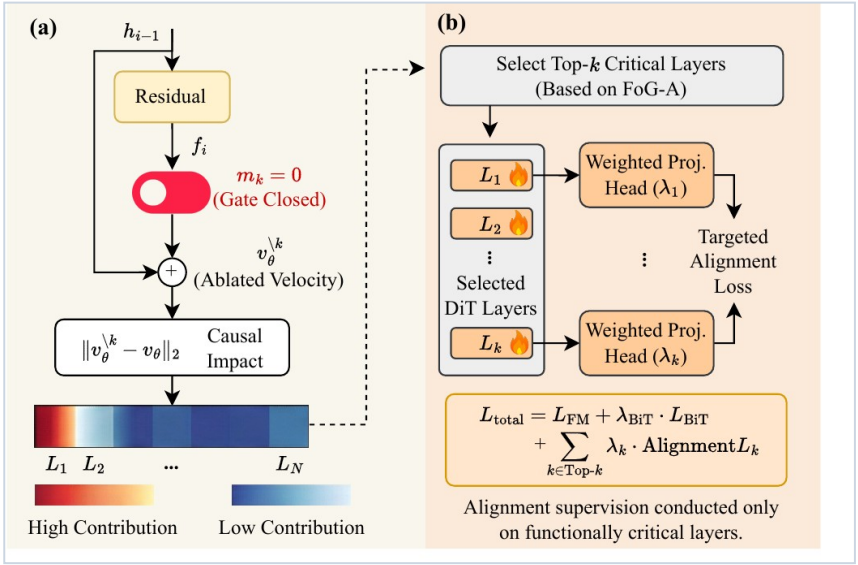
LASP identifies semantic/acoustic storage; FoG-A reveals which layers causally change the velocity field.

Method: Attribution-Guided REPA

Run forward-only gate ablation, select Top-K causal layers, and weight alignment by attribution.

FoG-A Top-3 (Config B): L1 L2 L7

Figure 3: From Causal Attribution to Optimization



Key results under unified speech + general-audio training (LibriSpeech + AudioSet)

Speech WER

5.82 -> 3.45

40.7% relative reduction vs no layer alignment

Speech FAD

1.58 -> 1.29

18% lower than best single fixed-layer REPA

Audio FAD

3.05 -> 2.56

16% lower than best single fixed-layer REPA

Human MOS

4.12 / 3.94

speech / audio MOS; strongest perceptual fidelity

Convergence

220k steps

reaches Speech FAD=1.5; about 3.3x faster than LASP selection

Keywords: Flow Matching | Audio Generation | Representation Alignment | Causal Layer Selection | FoG-A | REPA