TITLE: "Cross-Modal Knowledge Distillation for Ultra-Lightweight Edge AI: A Resource-Aware Approach"

Abstract

While edge AI continues to offer promising solutions for real-time applications, the deployment of multi-modal deep learning models remains particularly challenging considering their substantial computational and memory requirements on resource-constrained devices. This paper approaches this problem through the integration of a Resource-Aware Cross-Modal Knowledge Distillation (CMKD). Our approach introduces three key checkmarks: (1) a dynamic resource allocation mechanism that adaptively manages computational resources across different modalities based on real-time device constraints, (2) modality-specific compression techniques that optimize knowledge transfer while minimizing memory footprint, and (3) a lightweight feature alignment strategy that maintains cross-modal performance under varying resource conditions.

Literature Review

Cross-Modal Knowledge Distillation:

Cross-Modal Knowledge Distillation For Action Recognition	CMKD allows networks that have been trained on a modality like RGB videos to be adapted to recognize actions for another modality like sequences of 3D human poses. Process steps include: 1) Extract the knowledge of the trained teacher network (Source Modality) 2) Transfer it to a small ensemble of student networks (Target Modality) Loss usually used: KL-loss Loss preferred to be used: Cross entropy loss + Mutual learning
Knowledge As Priors: Cross-Modal Knowledge Generalization for Datasets Without Superior Knowledge	In case of inadequate data, one can generalize the distilled cross-modal knowledge learned from a Source dataset, which contains paired examples from both modalities to the Target dataset by modelling knowledge as priors on parameters of the Student.
EmotionKD: A Cross-Modal Knowledge Distillation	Approached emotion detection via EEG and GSR data signals. GSR is easy to access but EEG takes time to acquire.

Framework for Emotion Recognition Based on Physiological Signals	Through CMKD, fused multi-modal features can be transferred to an unimodal GSR model to improve performance. https://github.com/YuchengLiu-Alex/EmotionKD
Cross Modal Distillation for Supervision Transfer	Deals with CMKD in image data. Studied CMKD's performance with various supervision transfer methods. Could provide a framework to be followed in our paper
Links:	https://sci-hub.ru/https://ieeexplore.ieee.org/abstract/document/8802909 https://openaccess.thecvf.com/content_CVPR_2020/papers/Zhao_Knowledge_As_Priors_Cross-Modal_Knowledge_Generalization_for_Datasets_Without_Superior_CVPR_2020_paper.pdf https://openaccess.thecvf.com/content_cvpr_2016/html/Gupta_Cross_Modal_Distillation_CVPR_2016_paper.html https://dl.acm.org/doi/abs/10.1145/3581783.3612277

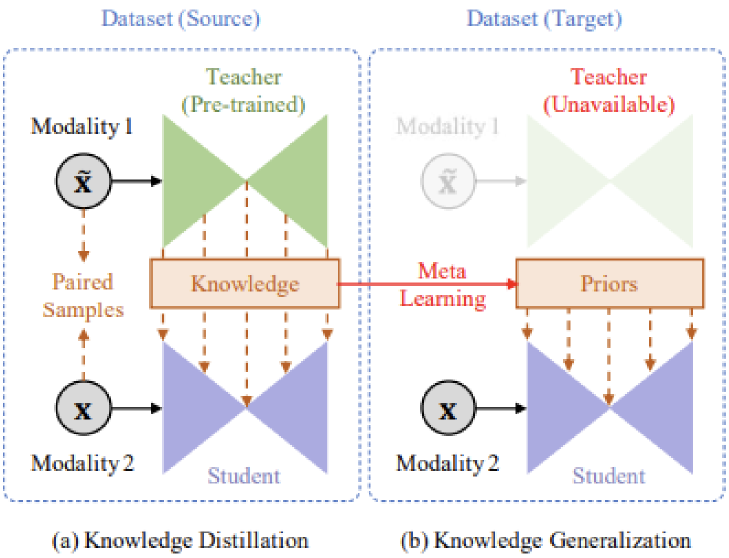

Figure 1. Cross-modal knowledge generalization. (a) Existing approaches distill cross-modal knowledge from the teacher to student in a source dataset. (b) We propose knowledge generalization which transfers learned knowledge in the source to a target dataset where the superior knowledge, *i.e.*, the teacher, is unavailable.

Project Objectives:

- Journal idea 1: Written generally and in principle
- Journal idea 2: Take a specific edge example such as atrial fibrillation detection, anything with 3D comp vision, emotion detection etc.

  https://arxiv.org/pdf/2408.04258

Propose architecture

Datasets found:

- Rendered Hand Pose Dataset: Christian Zimmermann and Thomas Brox. Learning to estimate 3D hand pose from single RGB images. In ICCV, pages 4903–4911, 2017.
- Stereo Hand Pose Tracking Benchmark: Jiawei Zhang, Jianbo Jiao, Mingliang Chen, Liangqiong Qu, Xiaobin Xu, and Qingxiong Yang. A hand pose tracking benchmark from stereo matching. In ICIP, pages 982–986, 2017
- NYUD2 dataset: N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from RGBD images. In ECCV, 2012.
- ImageNet dataset: J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei. ImageNet: A large-scale hierarchical image database. In CVPR, 2009.
- JHMDB dataset (action detection): H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black. Towards understanding action recognition. In ICCV, 2013.

Paper Structure:

- Intro
- lit review
- Current challenges with edge Ai,
- Benefits of CMKD
- Implementing CMKD
- Drawbacks of CKMD=Future Scope
- Conclusion

This work should bridge the gap between theoretical CMKD approaches and edge AI deployment, thus, offering a scalable solution for resource-constrained multi-modal systems.