

Advanced Machine Learning

Final Project Presentation — MobileViTs for Sign Language Recognition



SAPIENZA
UNIVERSITÀ DI ROMA

Dario Loi — 1940849

M.Sc. in Computer Science,
Sapienza, University of Rome.

A. Y. 2024 – 2025

Task and Motivation

- **Task:** Continuous Sign Language Recognition (CSLR).
- **Dataset:** RWTH-PHOENIX-Weather 2014.

We believe CSLR to be an impactful task if solved efficiently enough to be used in real-time applications. Currently, it has been approached with either CNNs[3, 1] that don't capture strong temporal dependencies, or expensive LLMs[2].

Models, Tools, Novelty

We want to apply MobileViTs[4] to the task to exploit exploit temporal dependencies in video sequences while maintaining a lightweight model.

Current MobileViT implementations do not make full use of optimized PyTorch attention implementations. We believe that we can provide a model that reduces FLOPs while achieving SOTA performance on CSLR. ¹


¹<https://github.com/jaiwei98/mobile-vit-pytorch> manually implements attention instead of using flash attention.

Metrics & Benchmarks

We will compare our model to current state-of-the-art models on the RWTH-PHOENIX-Weather 2014 dataset in terms of the Word Error Rate (WER) metric

$$WER = \frac{\text{Substitutions} + \text{Deletions} + \text{Insertions}}{N}$$

We will also compare the number of FLOPs used by CNN state of the art models[3, 1] and an open-source MobileViT implementation to our model.

-  J. Ahn, Y. Jang, and J. S. Chung.
SlowFast Network for Continuous Sign Language Recognition,
Sept. 2023.
[arXiv:2309.12304 \[cs\]](#).
-  J. Gong, L. G. Foo, Y. He, H. Rahmani, and J. Liu.
LLMs are Good Sign Language Translators, Apr. 2024.
[arXiv:2404.00925 \[cs\]](#).
-  L. Hu, L. Gao, Z. Liu, and W. Feng.
Continuous Sign Language Recognition with Correlation
Network, Mar. 2023.
[arXiv:2303.03202](#).
-  S. Mehta and M. Rastegari.
MobileViT: Light-weight, General-purpose, and Mobile-friendly
Vision Transformer, Mar. 2022.
[arXiv:2110.02178 \[cs\]](#).