



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

PIVOT

Prompting for Video Continual Learning THU-PM-345



Andrés Villa



Juan León



Motasem Alfarra



**Kumail
Alhamoud**



Julio Hurtado



Fabian Caba

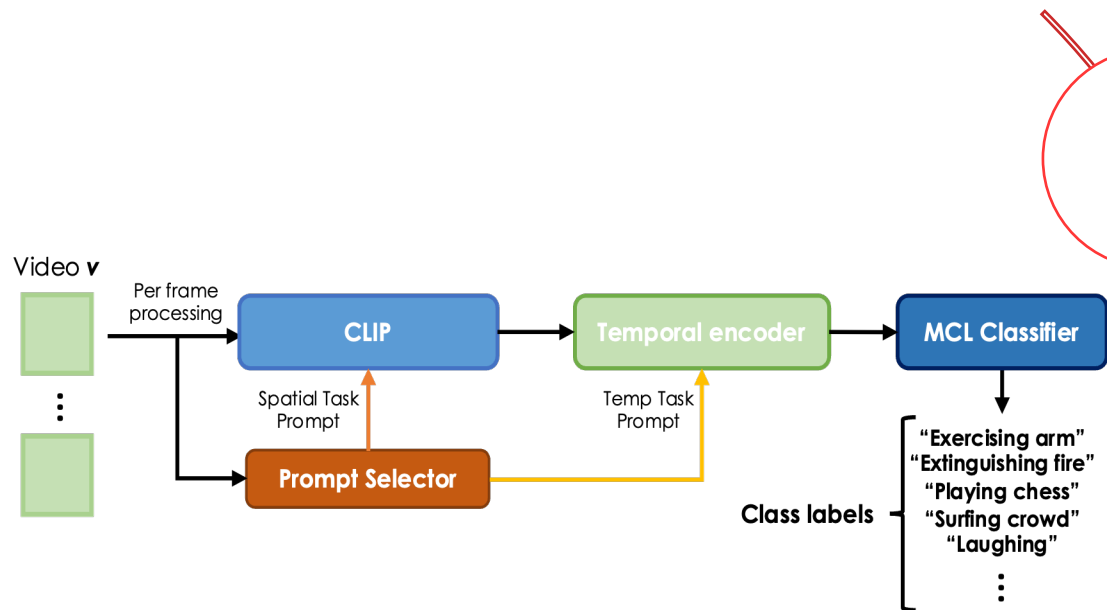


Alvaro Soto



Bernard Ghanem





PIVOT

Takeaway

1

We propose PIVOT, the first prompt-based strategy for video continual learning that leverages large-scale multimodal knowledge of CLIP.

2

It uses a transformer encoder to extract temporal information over CLIP and learnable tokens to learn and save specific information of each task.

3

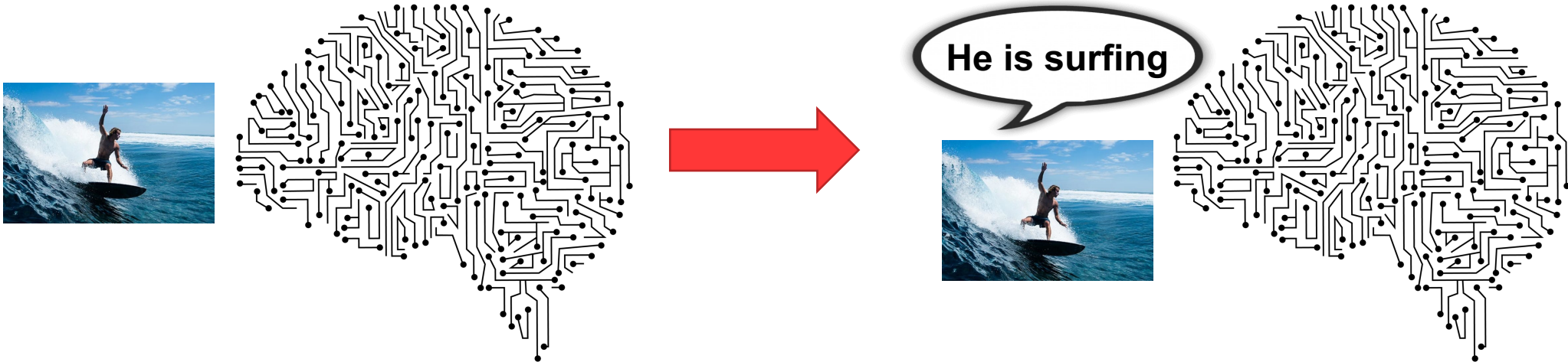
PIVOT outperforms SOTA baselines by 31%, 27%, and 17.2% in the 20-task setups of Kinetics, ActivityNet, and UCF101 in the vCLIMB.

4

We show that a multimodal classifier (Video-Text) mitigates catastrophic forgetting while greatly increasing the final average CIL accuracy.

Motivation

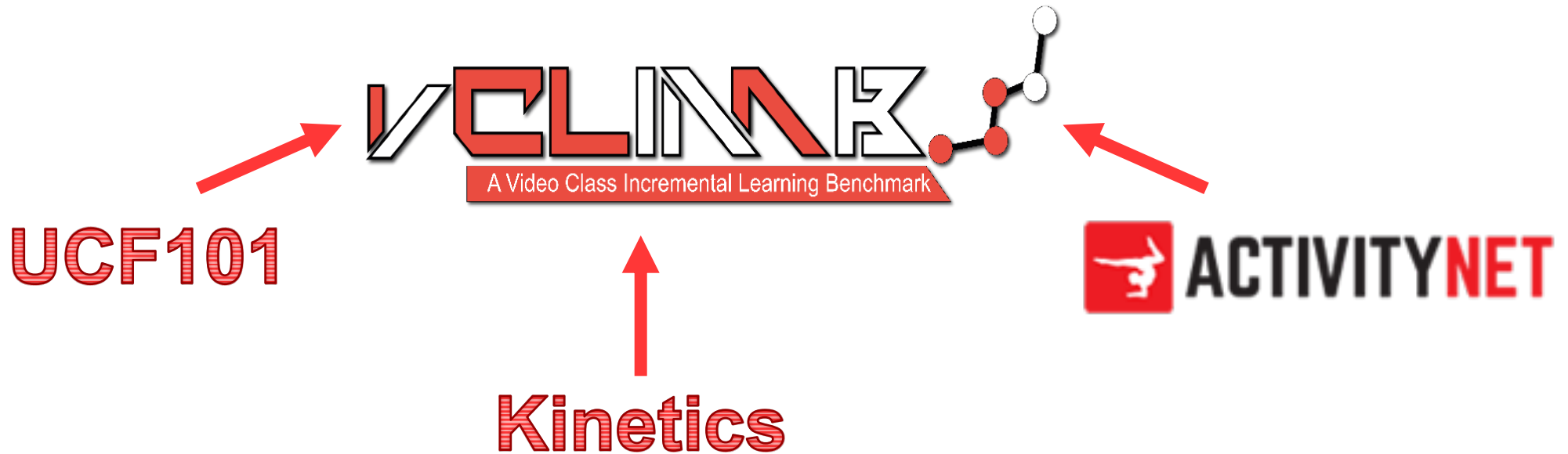
We hypothesize that careful multi-modal modeling can help the model to remember with even fewer memory samples.



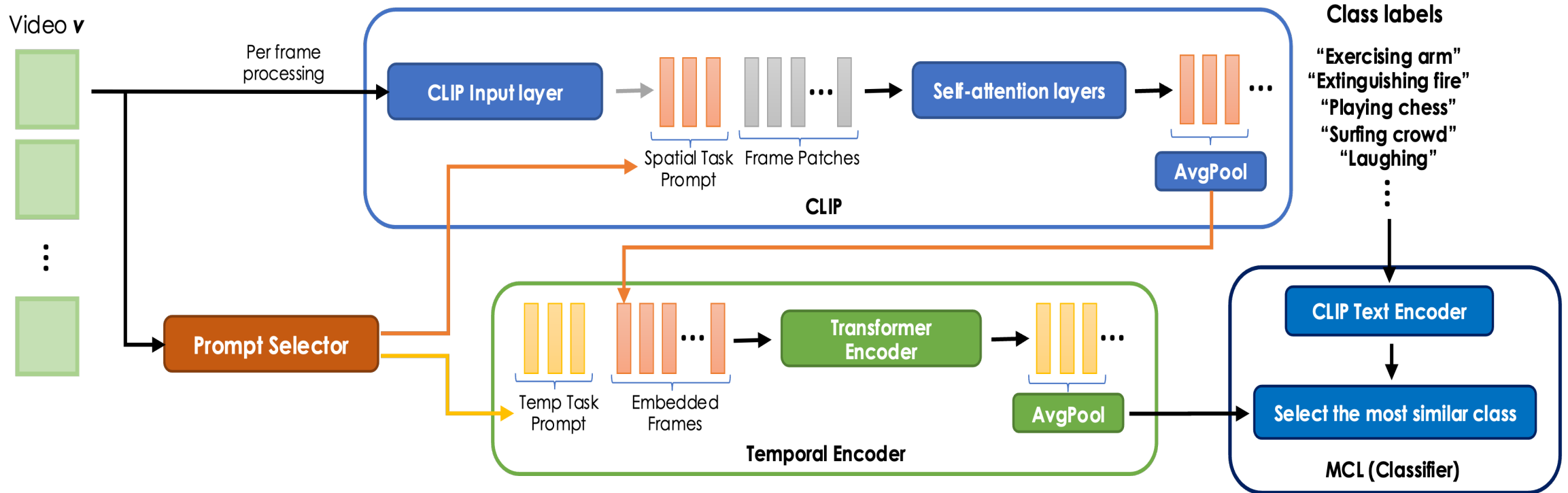
CL Benchmark



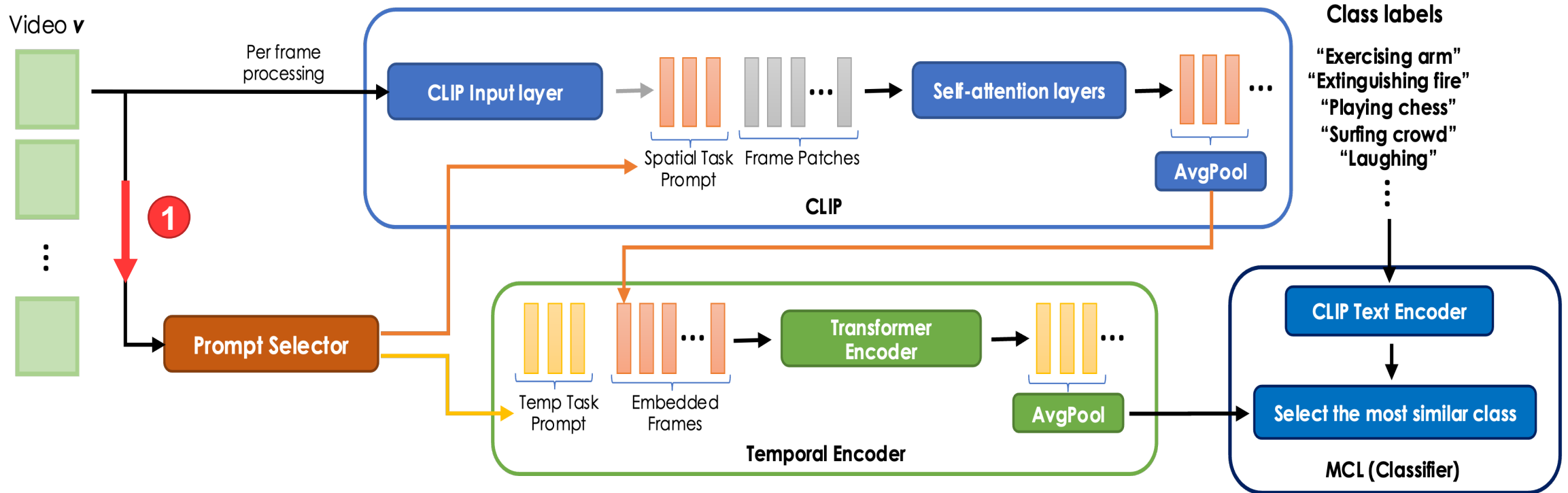
We evaluate our model in the challenging **vCLIMB** benchmark. It includes three well-known video datasets: **UCF101**, **Kinetics**, and **ActivityNet**.



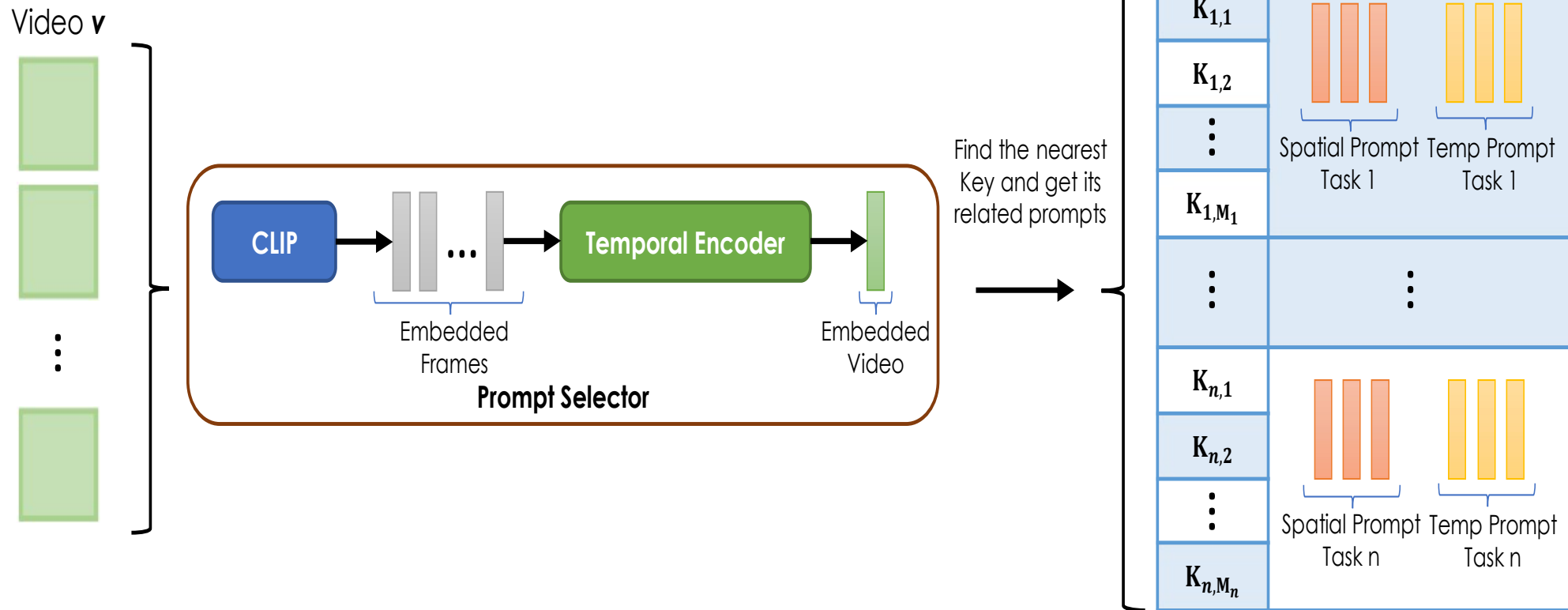
Our method



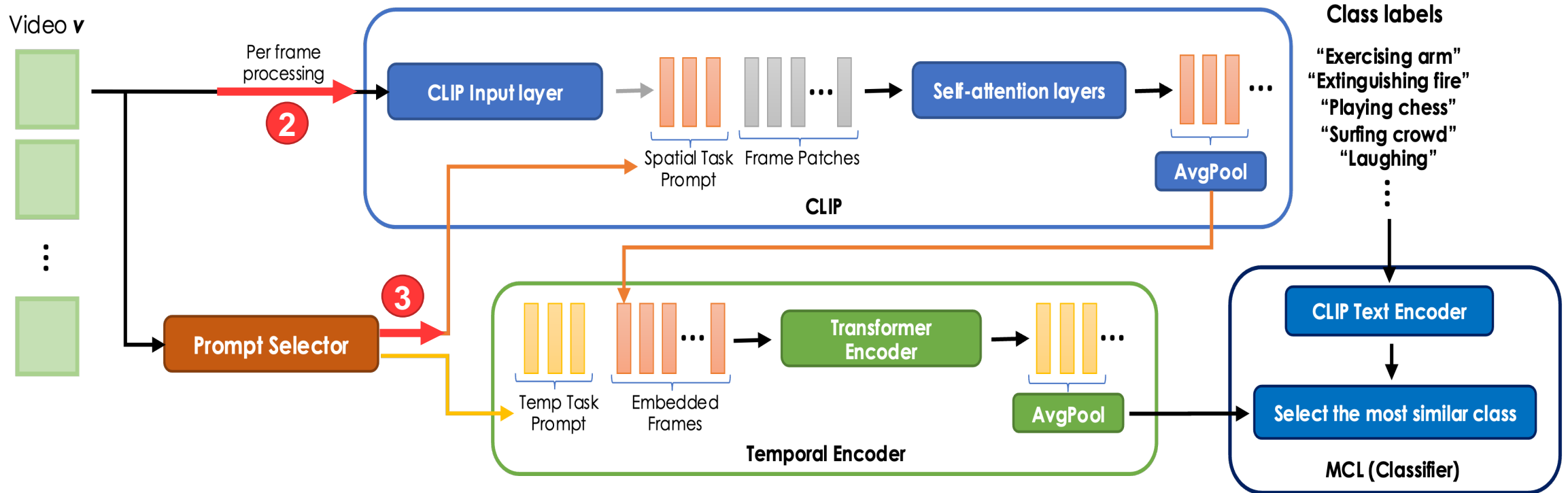
Our method



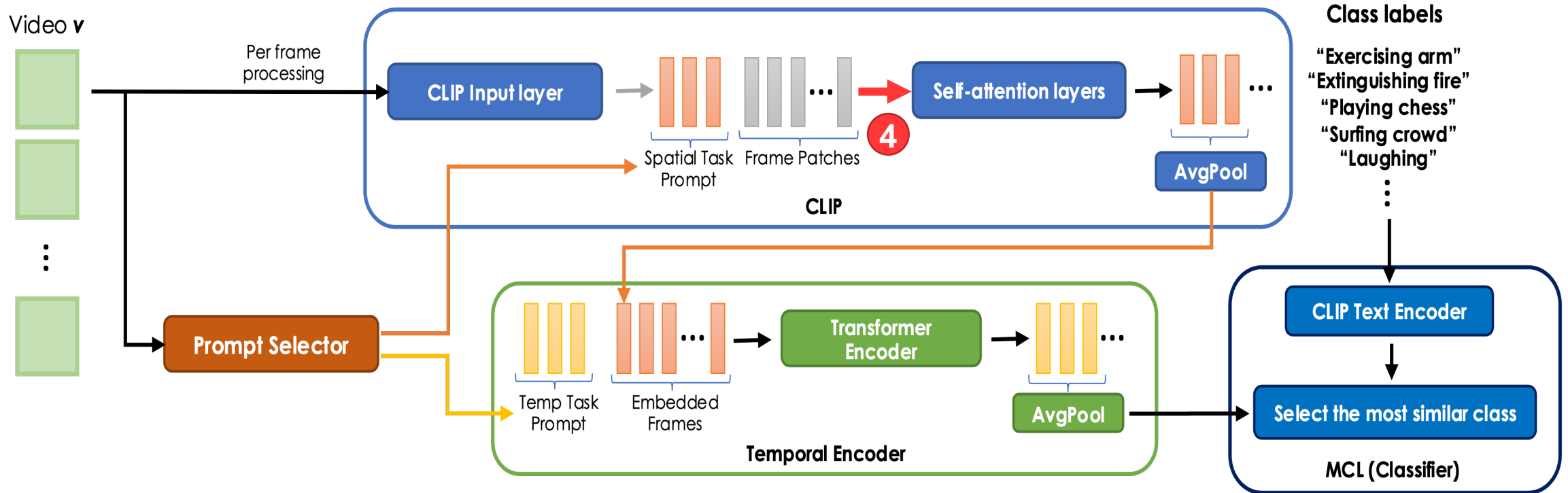
Our method



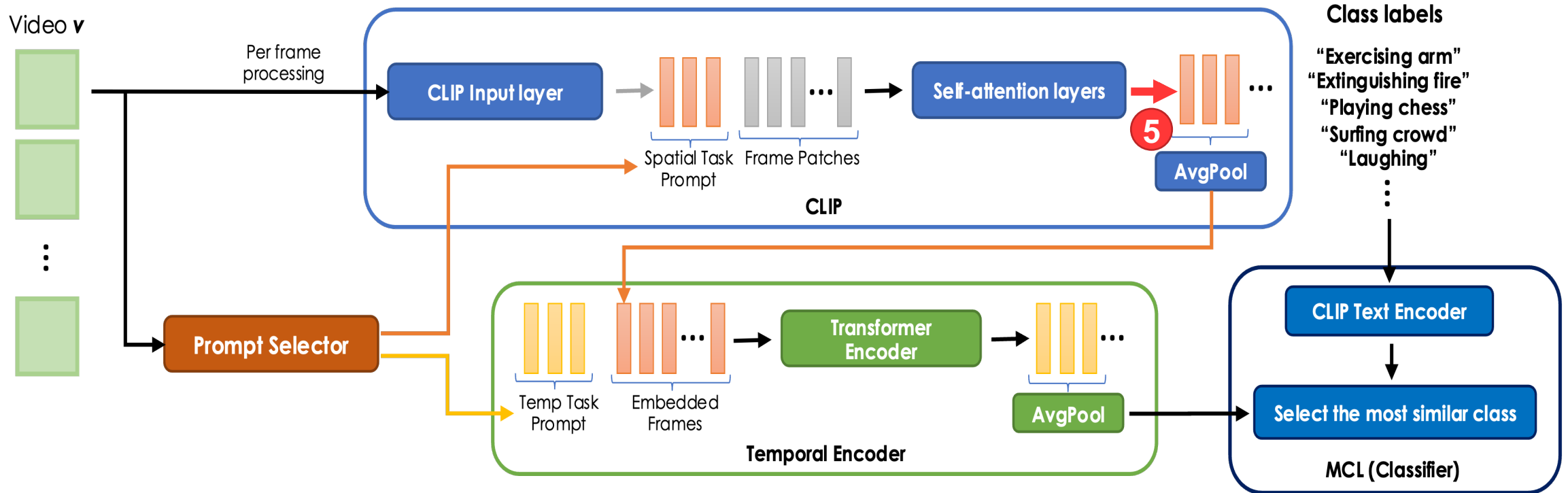
Our method



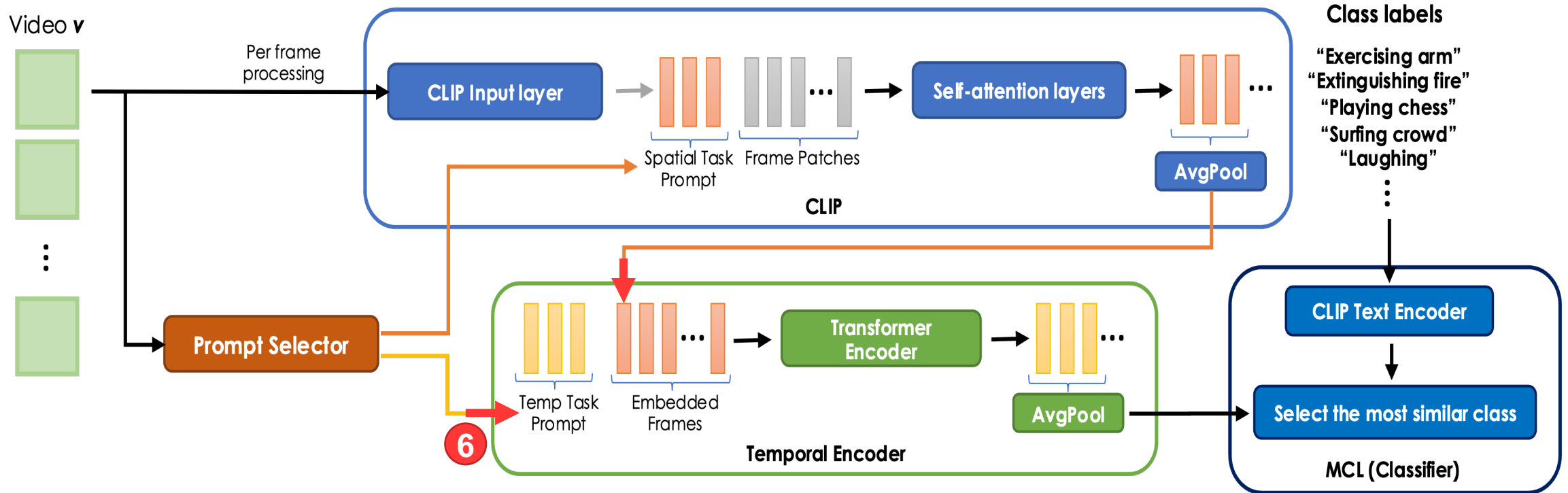
Our method



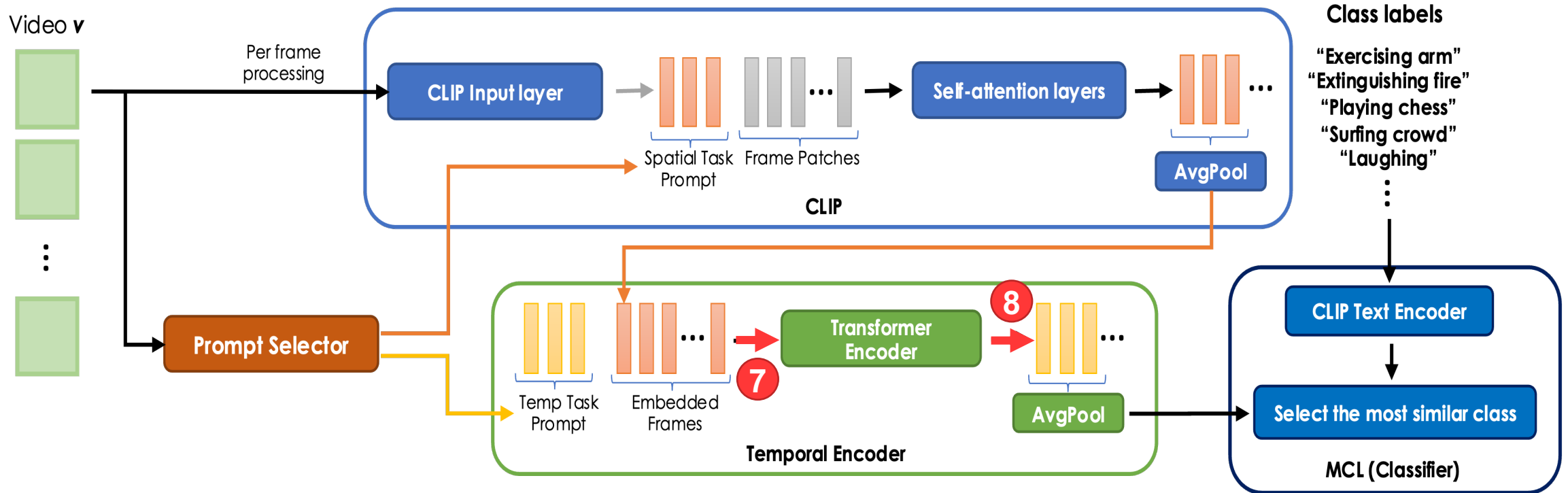
Our method



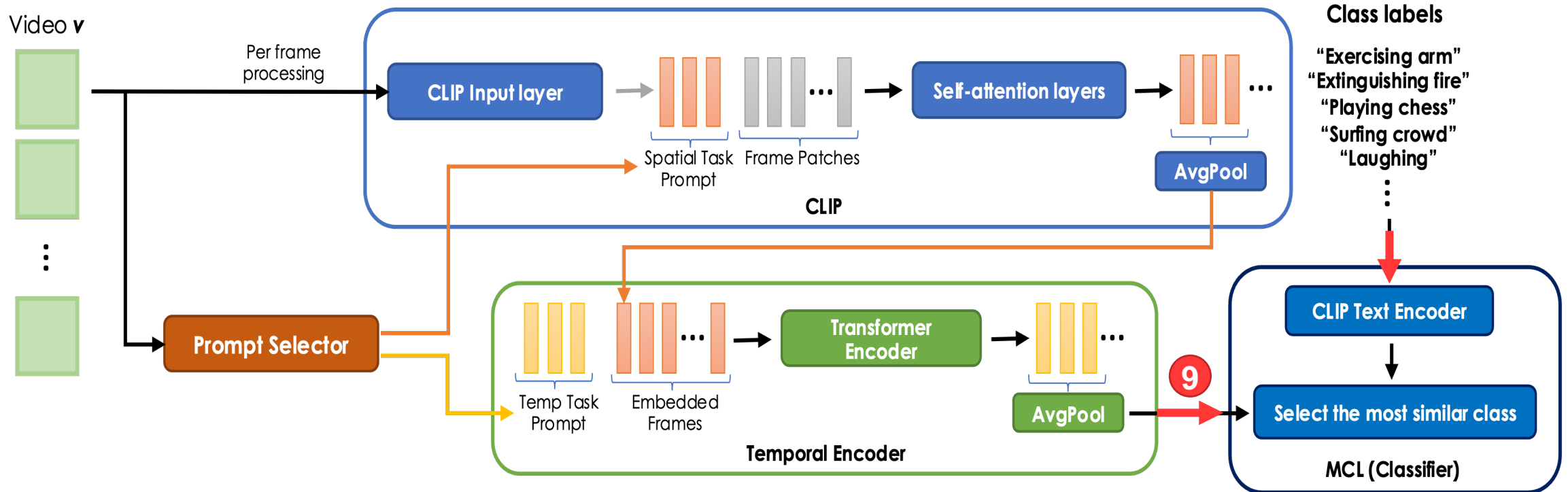
Our method



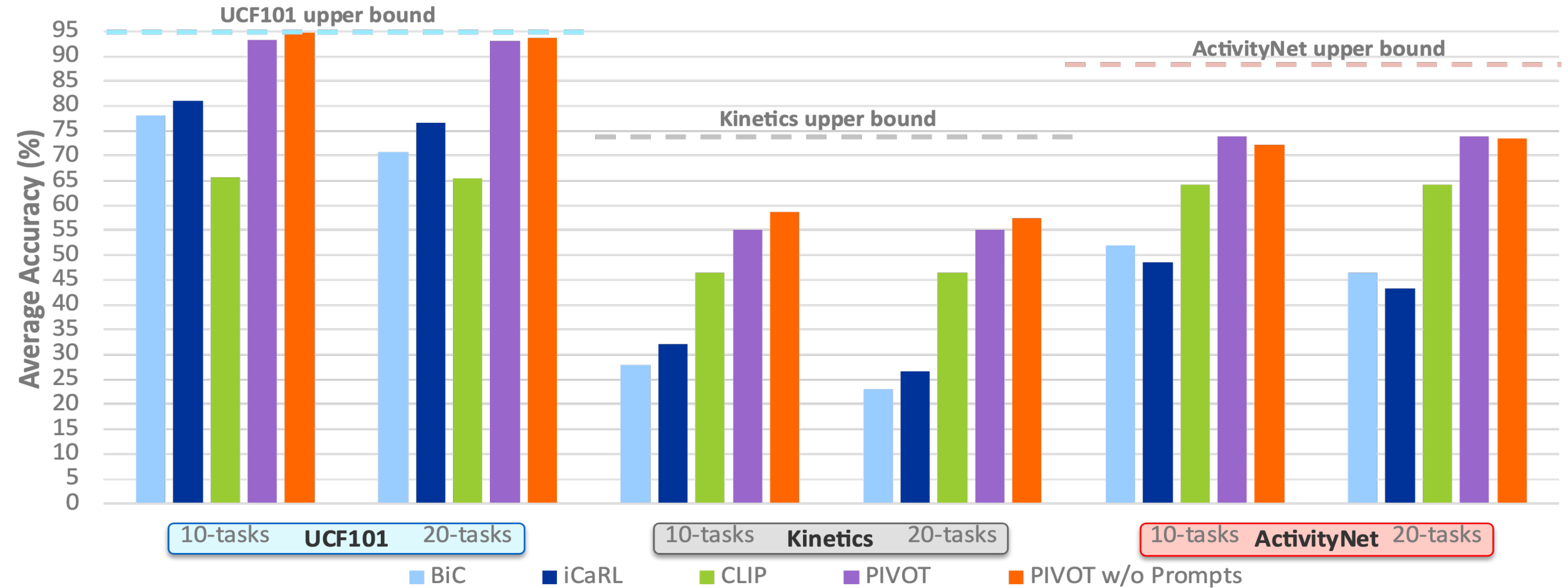
Our method



Our method



Results



Results

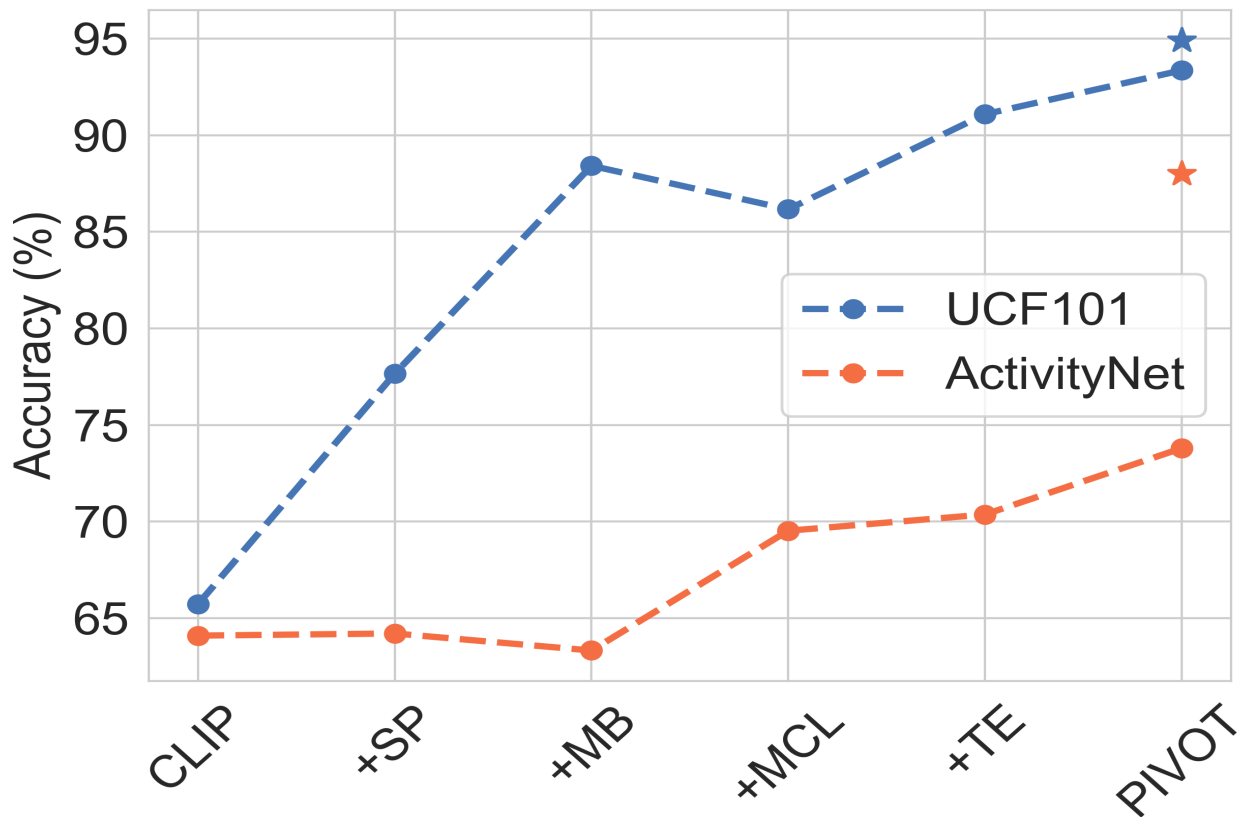


Figure 1. Performance improvement by each PIVOT Component

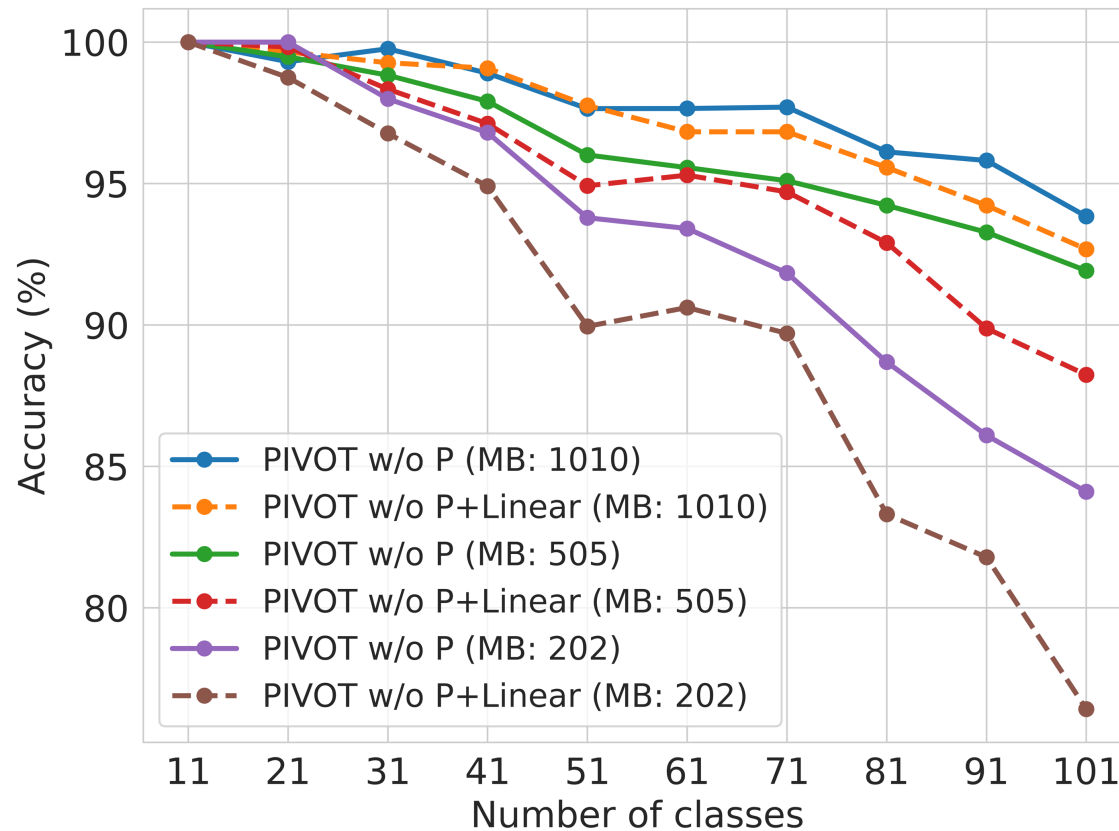


Figure 2. Multimodal Contrastive Classifier robustness to the memory size



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

PIVOT

Prompting for Video Continual Learning THU-PM-345



Andrés Villa



Juan León



Motasem Alfarra



**Kumail
Alhamoud**



Julio Hurtado



Fabian Caba



Alvaro Soto



Bernard Ghanem

