

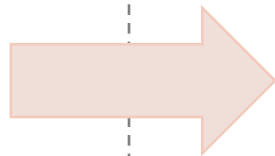
Pretraining

Downstream

Previous work

Activity Classification (eg: Kinetics-400)

- Single action per clip.
- Video-level annotations.
- + Large-scale dataset.



Ours

Weakly-guided Self-supervised Temporal Activity Detection (eg: Kinetics-400)

- + Multiple actions per frame.
- + Action segments.
- + Pseudo Frame-level annotations.
- + Large-scale dataset.



Temporal Activity Detection (eg: Charades)

- + Multiple actions per frame.
- + Action segments.
- + Frame-level annotations.
- Small-scale dataset.