

PivoTAL: Prior-Driven Supervision for Weakly-Supervised Temporal Action Localization

Mamshad Nayeem Rizve^{2*}, Gaurav Mittal^{1*}, Ye Yu¹, Matthew Hall¹,
Sandra Sajeev¹, Mubarak Shah², Mei Chen¹ (*equal contribution)



Motivation

Weakly-supervised Temporal Action Localization: Localizing actions using only *video-level* action labels. Challenging but mitigates need for dense labels.

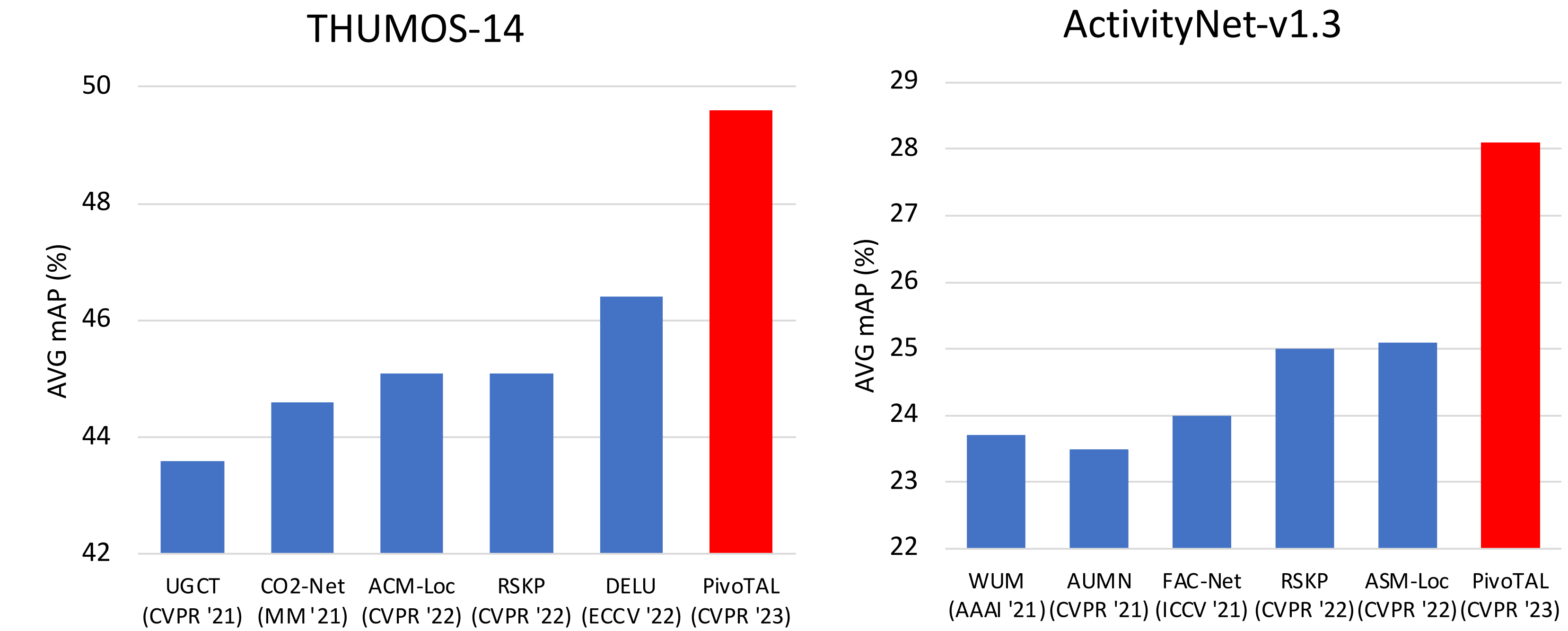
Existing WTAL:

- △ Attempts it from a *localization-by-classification* perspective
- △ No explicit notion of temporal boundaries
- △ Post-training aggregation cannot influence training

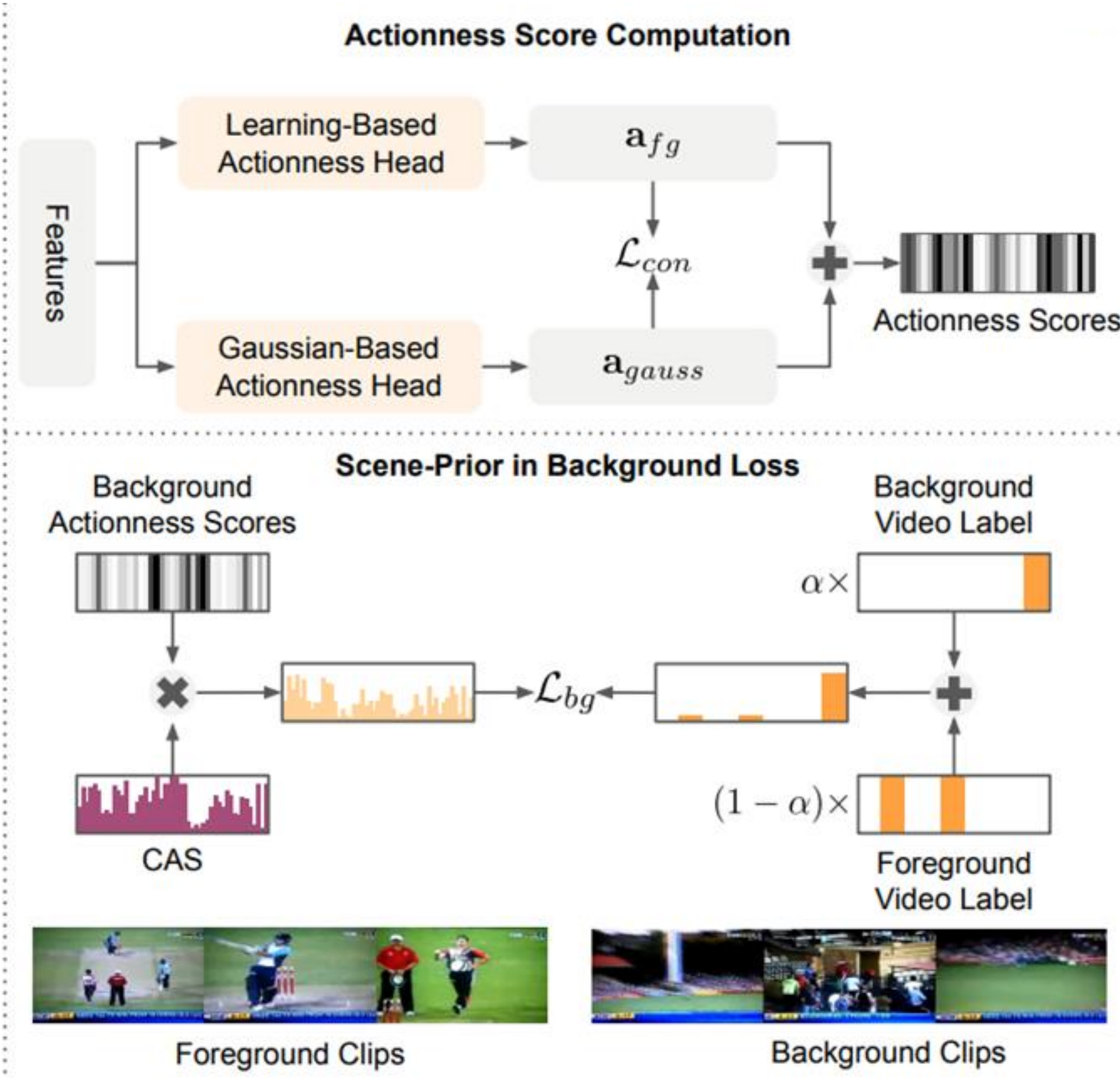
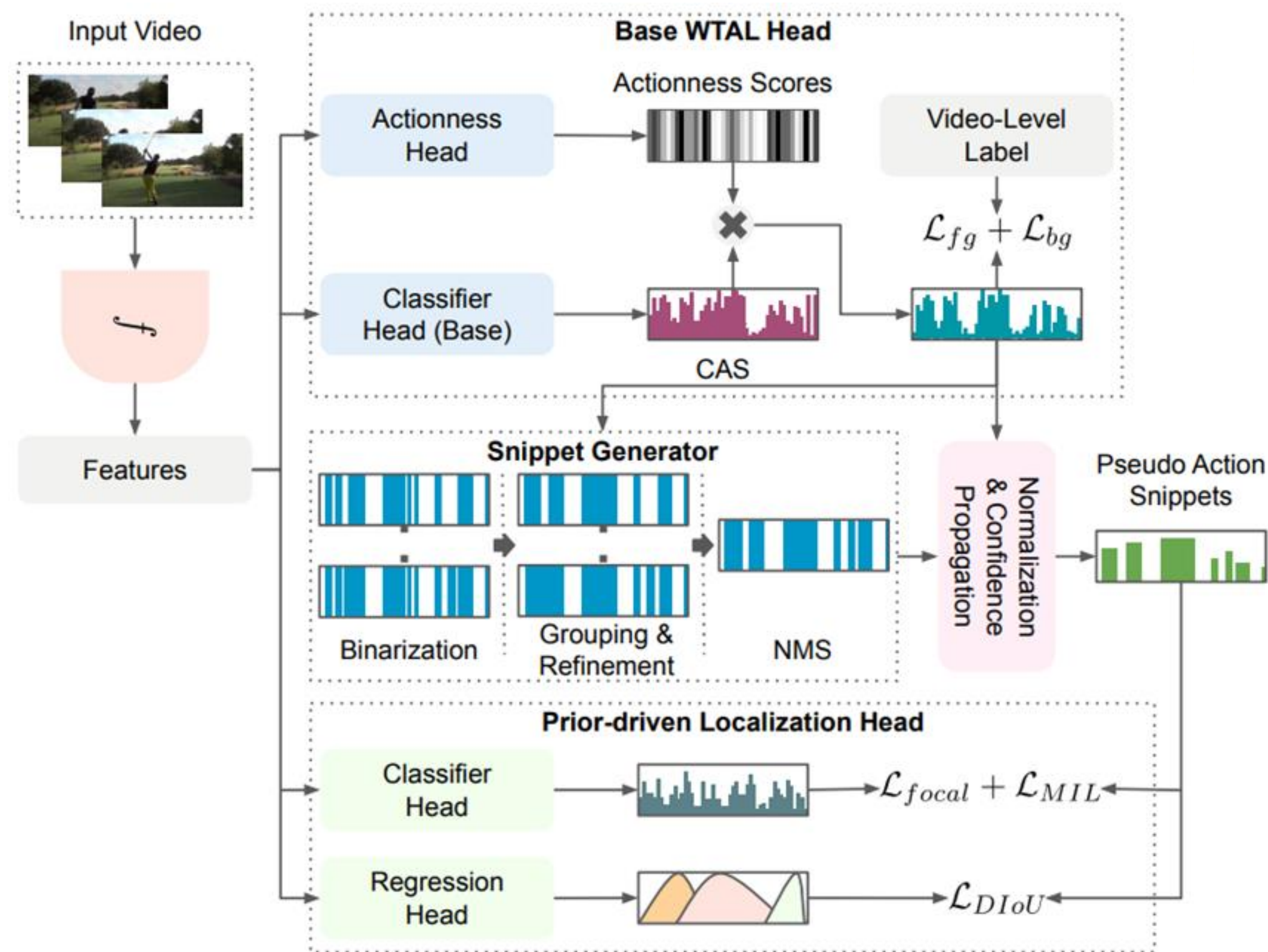
Contributions

- 👉 Solves WTAL from a *localization-by-localization* perspective
- 👉 Introduces and leverages **action-specific scene prior, action snippet generation prior, and learnable Gaussian prior**
- 👉 Generates pseudo-action snippets to localize actions directly
- 👉 Outperforms previous SOTA by **>3% average mAP**

Quantitative Results



Method



Qualitative Results

