Motivation


Existing WTAL:

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

Incomplete action snippets and several high-confident false positives

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

Method

Ground Truth                True Positive Snippets False Positive Snippets

Quantitative Results

Qualitative Results