

Perspective Shifts



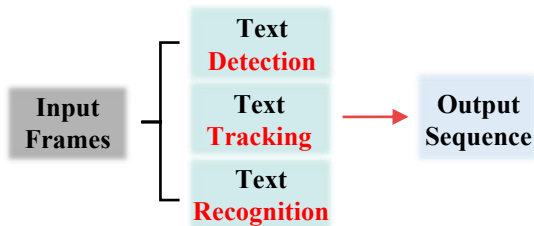
High Visual Ambiguity



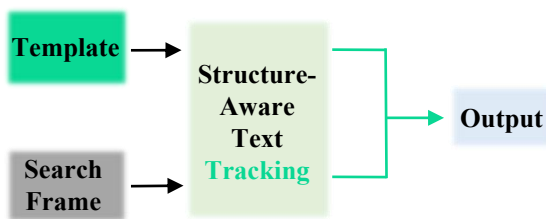
Structural Fine-granularity



(a) Typical failure cases of generic trackers highlighting the key challenges of scene text tracking. Green boxes denote ground truth, red boxes denote OTrack predictions.



(b) VTS with detection-dependent fragmented trajectories.



(c) Ours with prediction-aware and text-aware refinement.