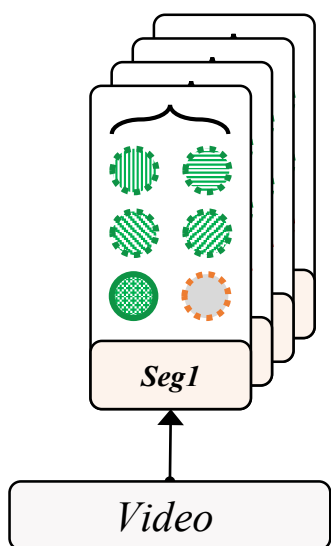
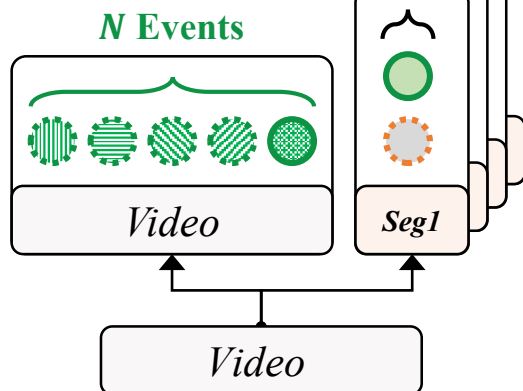


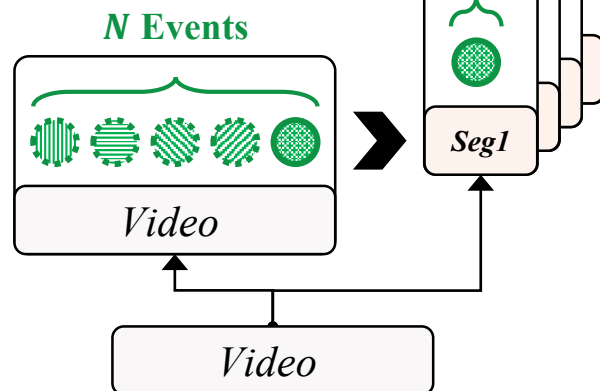
N Events + BG



Event or BG



i -th Event or Not



(a) Single Branch
 $N + 1$ classes

(b) Double Branches
 N classes, 2 classes

(c) Event-Aware Double Branches
 N classes, $N \times 2$ classes

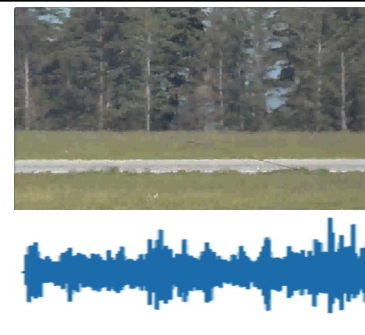
(d) Visual-Dominated

Event: airplane

Visual labels: 1 1 0

Audio: noisy overall

labels: 1 1 0



(e) Audio-Dominated

Event: man speaking

Visual: slight change

Audio labels: 0 1 0

Overall labels: 0 1 0



(f) Modality-Balanced

Event: car racing

Visual labels: 1 1 0

Audio labels: 1 1 0

Overall labels: 1 1 0

