

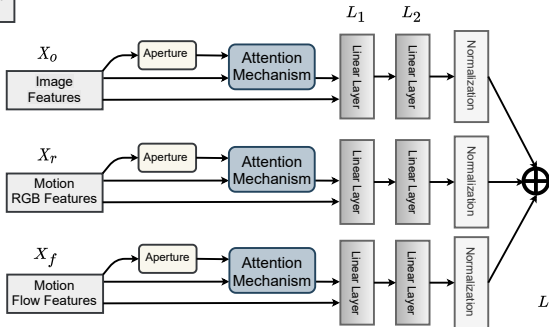
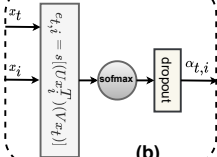
Video Frames

Extracted Visual Features

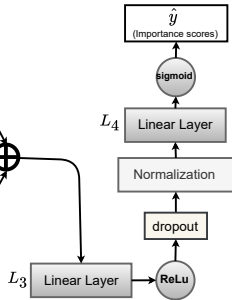
(a)

Attention Mechanism

(b)



(c)



(d)