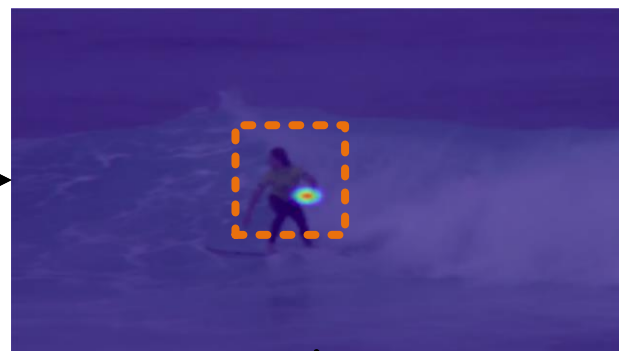


Level ( $N-2$ )



Teacher



Level ( $N-1$ )

down-sample



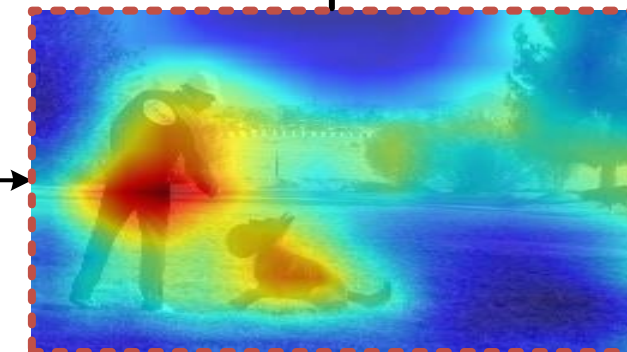
down-sample



Distillation



Student



Distillation

Target frame

Reference frame

Local correlation map

Target frame

Reference frame

Global correlation map

(a) Temporal Feature Learning

(b) Spatial Feature Learning