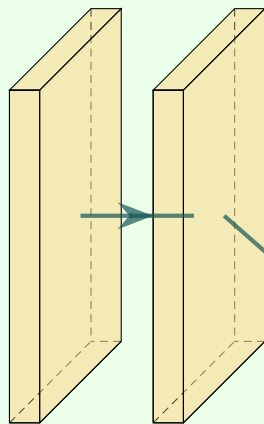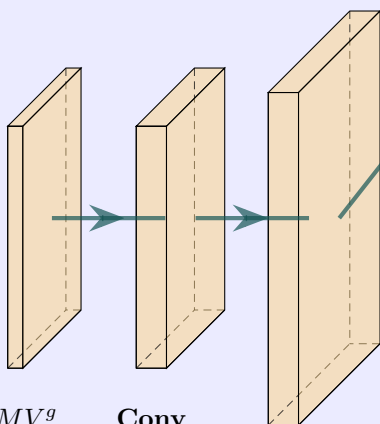**DCT Encoder**

18K–74K params (8–64 coeffs)

$\mathcal{DCT}(\Delta Y_n^g)$ **Conv**
$80{\times}80{\times}64$ $80{\times}80{\times}32$

MV Encoder

14K params

MV_n^g **Conv**
$40{\times}40{\times}2$ $40{\times}40{\times}32$

Upsample
$80{\times}80{\times}32$

Fusion & Temporal

196K params

Concat **Conv** **Conv**
$80{\times}80{\times}64$ $H{\times}W{\times}256$ $H{\times}W{\times}256$

GAP
256

256

BiLSTM
256
hid-
den

Detection
128-
dim

$\{\hat{\mathbf{b}}_i\}$
$N_{det}{\times}4$