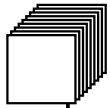


Video frames  
16 frames (112x112)



1x7x7 conv, 45, stride 1,2,2

batch normalization

ReLU

3x1x1 conv, 64, stride 1,1,1

batch normalization

ReLU

R2+1D Block [64, 64]

R2+1D Block [64, 64]

R2+1D Block [64, 64]

R2+1D Block [64, 128]

R2+1D Block [128, 128]

R2+1D Block [128, 128]

R2+1D Block [128, 128]

R2+1D Block [128, 256]

R2+1D Block [256, 256]

R2+1D Block [256, 256]

R2+1D Block [256, 256]

R2+1D Block [256, 256]

R2+1D Block [256, 256]

R2+1D Block [256, 512]

R2+1D Block [512, 512]

R2+1D Block [512, 512]

1x7x7 avg-pool, stride 1,1,1

Predictions

**R(2+1)D Model**