

Input frames 2 x 3 x 64 x 64
7 first layers of resnet-18 (pretrained, frozen weights) applied to each frame
Reshape 1 x 16384
FC 16384 \rightarrow 512
FC 512 \rightarrow 8192
Reshape 128 x 8 x 8
UpSamplingNearest(2), 3 x 3 Conv. 128 - 1 str., BN, ReLU
UpSamplingNearest(2), 3 x 3 Conv. 64 - 1 str., BN, ReLU
UpSamplingNearest(2), 3 x 3 Conv. 3 - 1 str., BN, ReLU
3 sigmoid Target mask