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