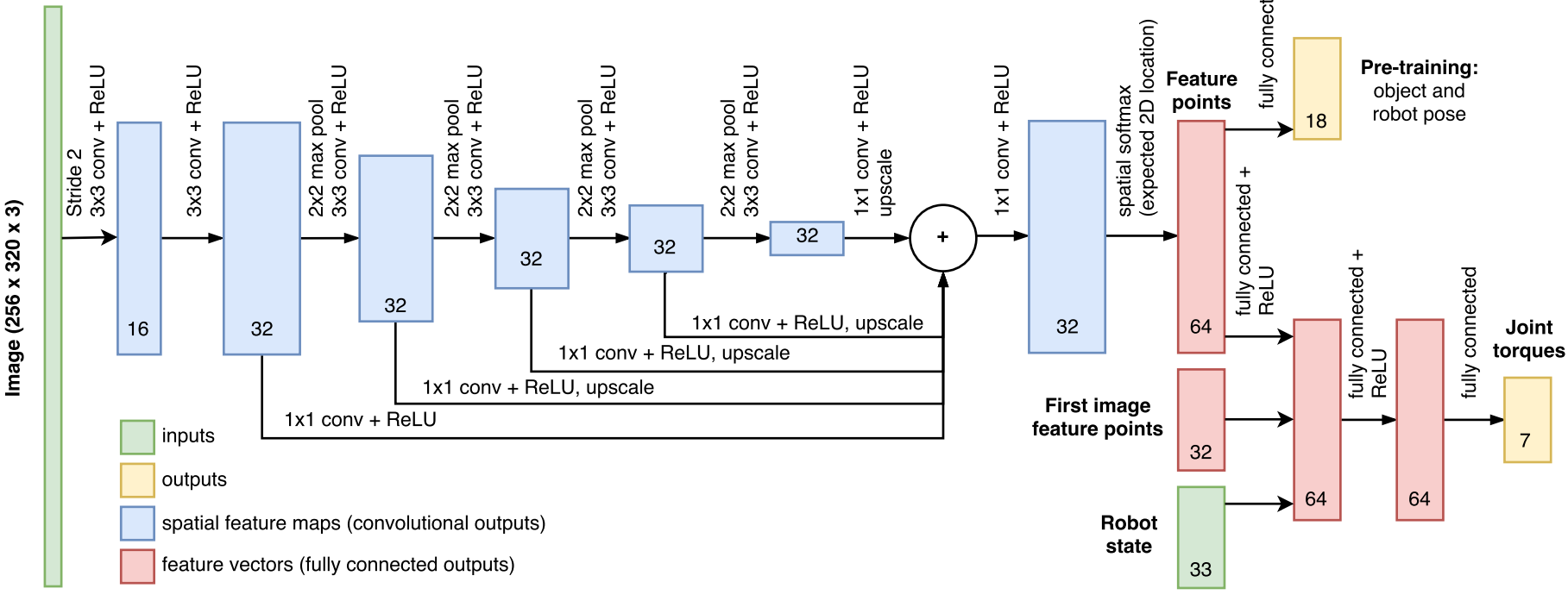
Image (256 x 320 x 3)
Stride 2
3x3 conv + ReLU
16
3x3 conv + ReLU
32
2x2 max pool
3x3 conv + ReLU
32
2x2 max pool
3x3 conv + ReLU
32
2x2 max pool
3x3 conv + ReLU
32
2x2 max pool
3x3 conv + ReLU
32
1x1 conv + ReLU
upscale
+
1x1 conv + ReLU
32
spatial softmax
(expected 2D location)
Feature points
64
fully connected
18
Pre-training:
object and
robot pose
fully connected +
ReLU
64
fully connected +
ReLU
64
fully connected
Joint torques
7
First image
feature points
32
Robot state
33
1x1 conv + ReLU, upscale
1x1 conv + ReLU, upscale
1x1 conv + ReLU, upscale
1x1 conv + ReLU
inputs
outputs
spatial feature maps (convolutional outputs)
feature vectors (fully connected outputs)