



**Supplemental Figure 1:** Illustration of architecture of the image feature extractor. "n $\times$ n conv, m, /k" represents a convolutional layer with n by n filter size, m channels, and k stride step. "Avg pool, 1/k" represents an average pooling layer with pooling size k by k. The extractor had the same structure as the first 5 layers of the ResNet18 network.