convolutional layer
with non-linearities
layer $l = 1$

convolutional layer
with non-linearities
layer $l = 4$

fully connected layer
layer $l = 7$

input image
layer $l = 0$

subsampling layer
layer $l = 3$

subsampling layer
layer $l = 6$

fully connected layer
output layer $l = 8$

Figure 1: The architecture of the original convolutional neural network, as introduced by LeCun et al. (1989), alternates between convolutional layers including hyperbolic tangent non-linearities and subsampling layers. In this illustration, the convolutional layers already include non-linearities and, thus, a convolutional layer actually represents two layers. The feature maps of the final subsampling layer are then fed into the actual classifier consisting of an arbitrary number of fully connected layers. The output layer usually uses softmax activation functions.