# A Experimental details

## A.1 Label Smoothing

We denote a convolutional layer by W×W×N-S, where W is the width of the convolution, N the number of filter maps, and S the stride. Our architecture is: 3×3×32-1 → BatchNorm → 3×3×32-1 → BatchNorm → MaxPooling (2×2) → Dropout (0.2) → 3×3×64-1 → BatchNorm → 3×3×64-1 → BatchNorm → MaxPooling (2×2) → Dropout (0.3) → 3×3×128-1 → BatchNorm → 3×3×128-1 → BatchNorm → MaxPooling (2×2) → Dropout (0.4) → 10 fully-connected units. We use weight decay of 0.0001 in the final fully-connected layer, which, together with dropout and batch normalization, was switched on and off in the different runs of our ablation study in Table 1. Our models were trained for 500 epochs using a minibatch size of 128 and the Adam optimizer with a learning rate of $10^{-3}$. The label smoothing hyperparameter $\varepsilon$ was set to 0.1 as per [22].

Note however, that we were unable to replicate the results of [22] exactly, as they did not share their code, nor did they describe their architecture in full.

## A.2 Actor-Mimic Network

Our architecture is 8×8×32-4 → 4×4×64-2 → 3×3×64-1 → 7×7×1024-1 → 512 fully-connected units → 6 fully connected units (corresponding to 6 possible actions). We used the Adam optimizer with a learning rate of $10^{-5}$, and a minibatch size of 32.