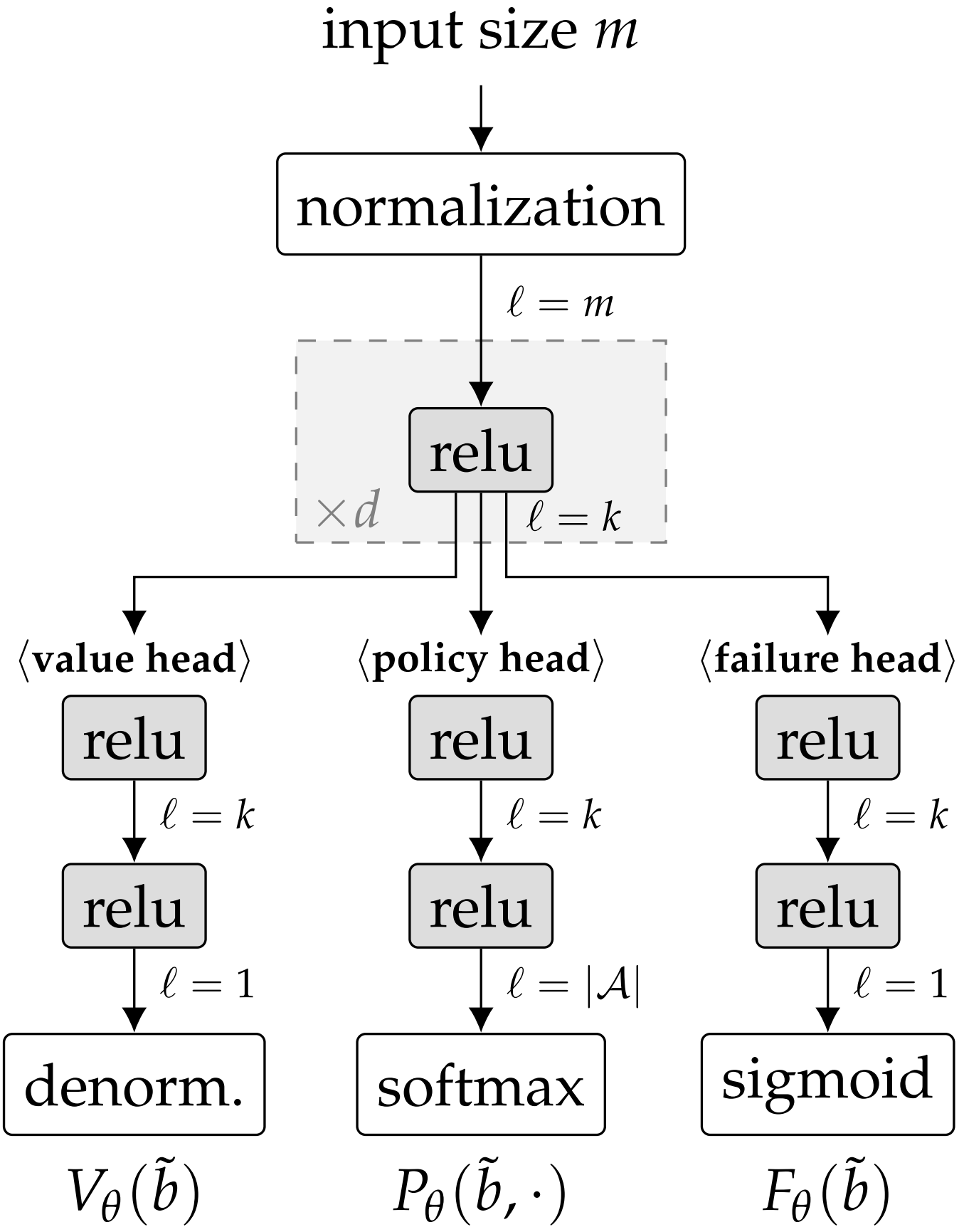
input size m
normalization
$\ell = m$
relu
$\times d$
$\ell = k$
⟨value head⟩
⟨policy head⟩
⟨failure head⟩
relu
$\ell = k$
relu
$\ell = 1$
denorm.
$V_\theta(\tilde{b})$
relu
$\ell = k$
relu
$\ell = |\mathcal{A}|$
softmax
$P_\theta(\tilde{b}, \cdot)$
relu
$\ell = k$
relu
$\ell = 1$
sigmoid
$F_\theta(\tilde{b})$