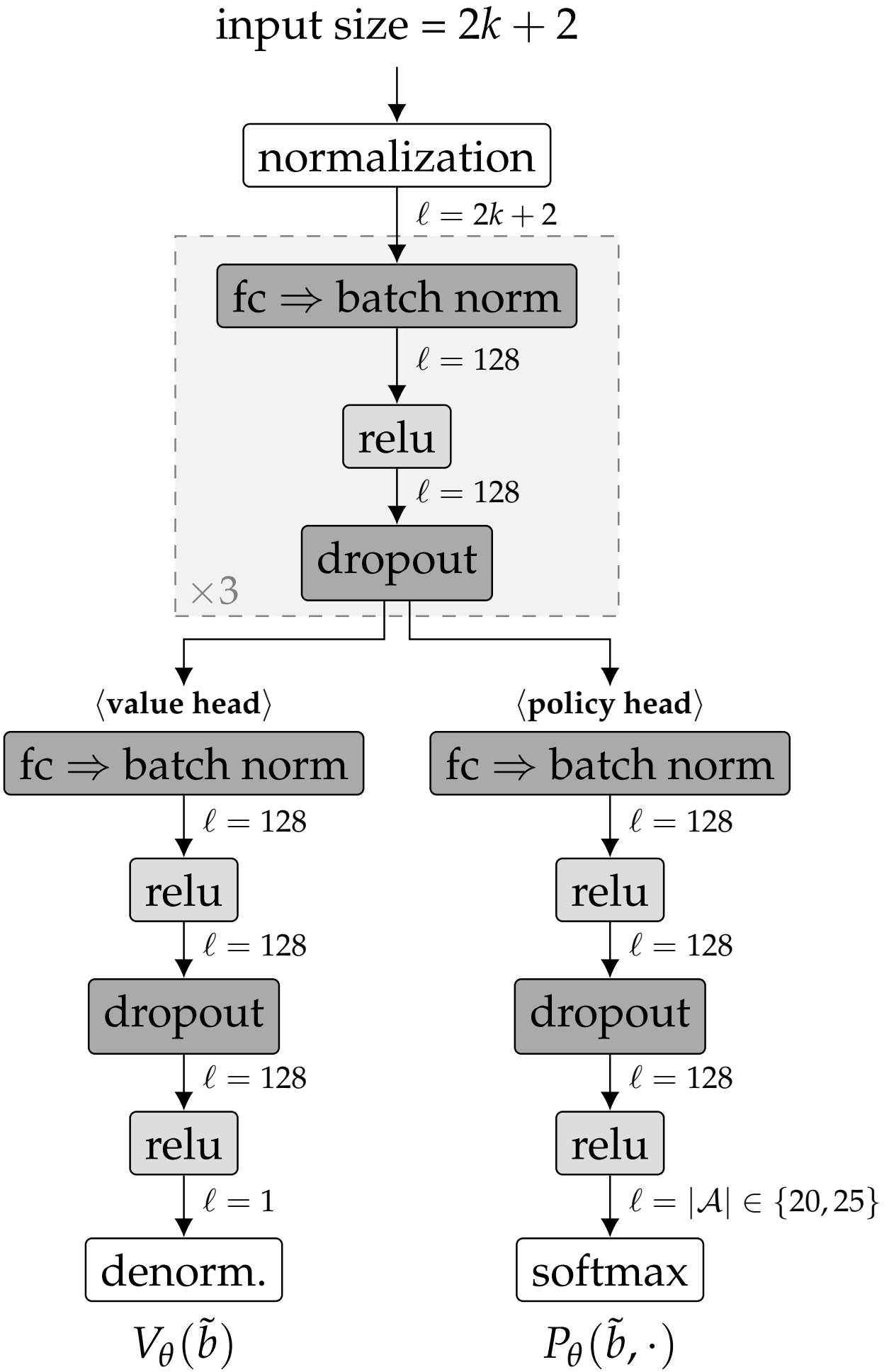
input size = $2k + 2$
normalization
$\ell = 2k + 2$
fc $\Rightarrow$ batch norm
$\ell = 128$
relu
$\ell = 128$
dropout
$\times 3$
⟨value head⟩
fc $\Rightarrow$ batch norm
$\ell = 128$
relu
$\ell = 128$
dropout
$\ell = 128$
relu
$\ell = 1$
denorm.
$V_\theta(\tilde{b})$
⟨policy head⟩
fc $\Rightarrow$ batch norm
$\ell = 128$
relu
$\ell = 128$
dropout
$\ell = 128$
relu
$\ell = |\mathcal{A}| \in \{20, 25\}$
softmax
$P_\theta(\tilde{b}, \cdot)$