

input size = 2

normalization

$\ell = 2$

relu

$\ell = 64$

dropout

$\times 3$

$\langle \text{value head} \rangle$

relu

$\ell = 64$

dropout

$\ell = 64$

relu

$\ell = 1$

denorm.

$V_{\theta}(\tilde{b})$

$\langle \text{policy head} \rangle$

relu

$\ell = 64$

dropout

$\ell = 64$

relu

$\ell = |\mathcal{A}| = 3$

softmax

$P_{\theta}(\tilde{b}, \cdot)$