Policy network — Value network

$$p_{\sigma|\rho}(a|s) \qquad v_\theta(s')$$

$s \qquad s'$

Move Picker

Position Evaluator