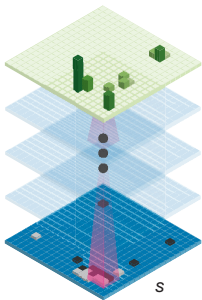


**b**

Policy network

$$p_{\sigma/\rho}(a|s)$$



Value network

$$v_{\theta}(s')$$

