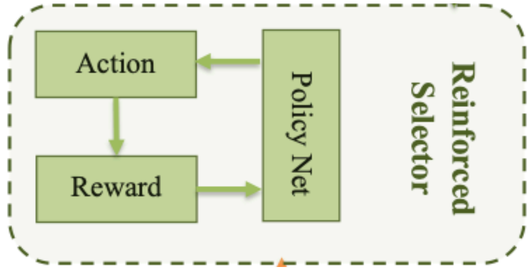
Methodology

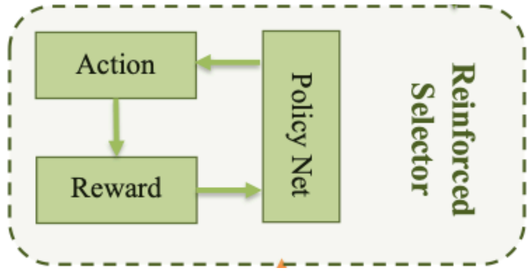
Data Selection via Reinforcement Learning - *Reward*



- Use performance changes of detection model $D_n\left(\cdot;\theta_n\right)$ as the reward function
- Given $\tilde{X}^{(k)} = \{x_1^{(k)}, x_2^{(k)}, \cdots, x_B^{(k)}\}$, the actions of retaining or removing are made based on the probability output from the policy network
  - To evaluate the performance changes, need to set a baseline accuracy acc
  - Calculate acc with $D_n\left(\cdot;\theta_n\right)$ on validation dataset
  - Then new accuracy acc_k can obtained with the retrained model
- Finally, the reward R_k for k-th bag data $\left\{x_i^{(k)}\right\}_{i=1}^{B}$: $R_k = acc_k - acc$

Methodology

Data Selection via Reinforcement Learning - *Reward*



- For k-th bag data $\left\{ x_i^{(k)} \right\}_{i=1}^{B}$, aim to maximize the expected total reward
- Since the scale of R_k is small use the summation of reward to define the objective function:
  - $J\left(\theta_s\right) = \sum_{i=1}^{B} \pi_{\theta_s}\left(s_i^{(k)}, a_i^{(k)}\right) R_k$