Algorithm 1 Actor-Critic Algorithm(On-Policy)

Require:

Learning rate for Actor α_θ,
Learning rate for Critic α_ω,
Initial parameters θ, ω,
Number of iterations N

$\mathcal{D} \leftarrow \emptyset$
for iteration = 1 to N **do**
  Run policy π_θ for T steps to collect data $\mathcal{D}$.
  $\mathcal{L}_{\text{Critic}} \leftarrow$ Comput Critic Loss$(\mathcal{D}, \mathrm{V}_\omega)$
  $\omega \leftarrow \omega + \alpha_\omega \nabla_\omega \mathcal{L}_{\text{Critic}}$
  $\mathcal{L}_{\text{Actor}} \leftarrow$ Compute Actor Loss$(\mathcal{D}, \pi_\theta)$
  $\theta \leftarrow \theta + \alpha_\theta \nabla_\theta \mathcal{L}_{\text{Actor}}$
end for