



Chunk-Guided Q-Learning

Gwanwoo Song¹, Kwanyoung Park², Youngwoon Lee¹

¹Yonsei University ²UC Berkeley

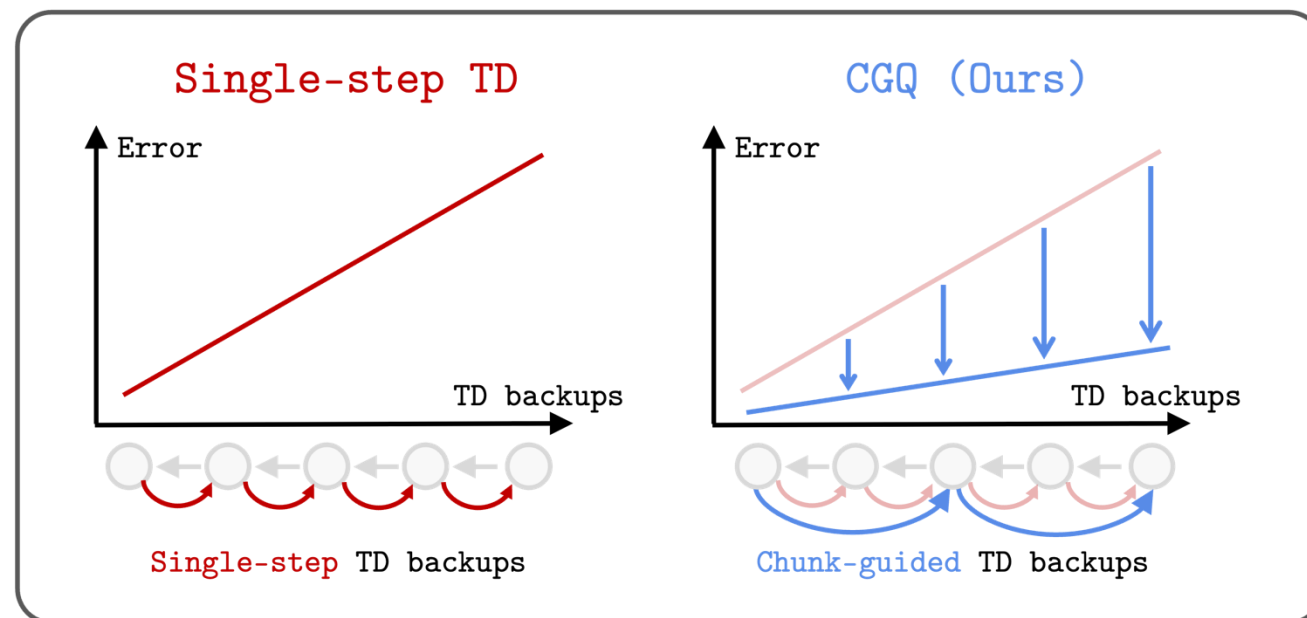


Our Goal

Improve long-horizon value learning in offline RL by mitigating error accumulation

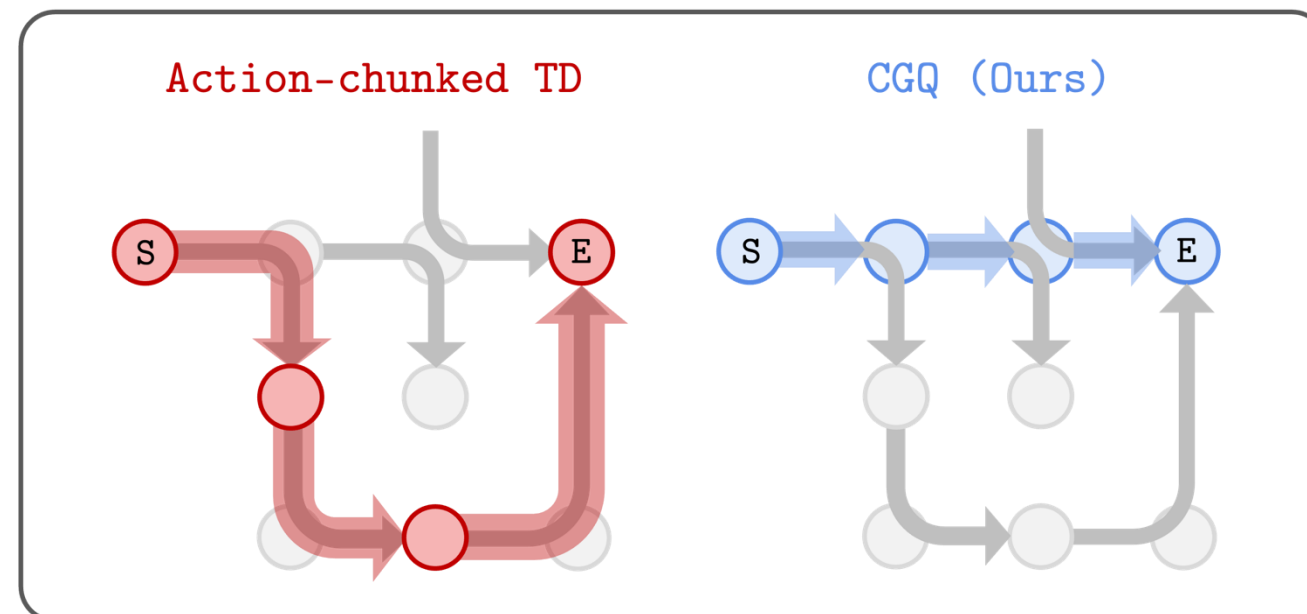
Single-step TD

Value error accumulates through bootstrapping



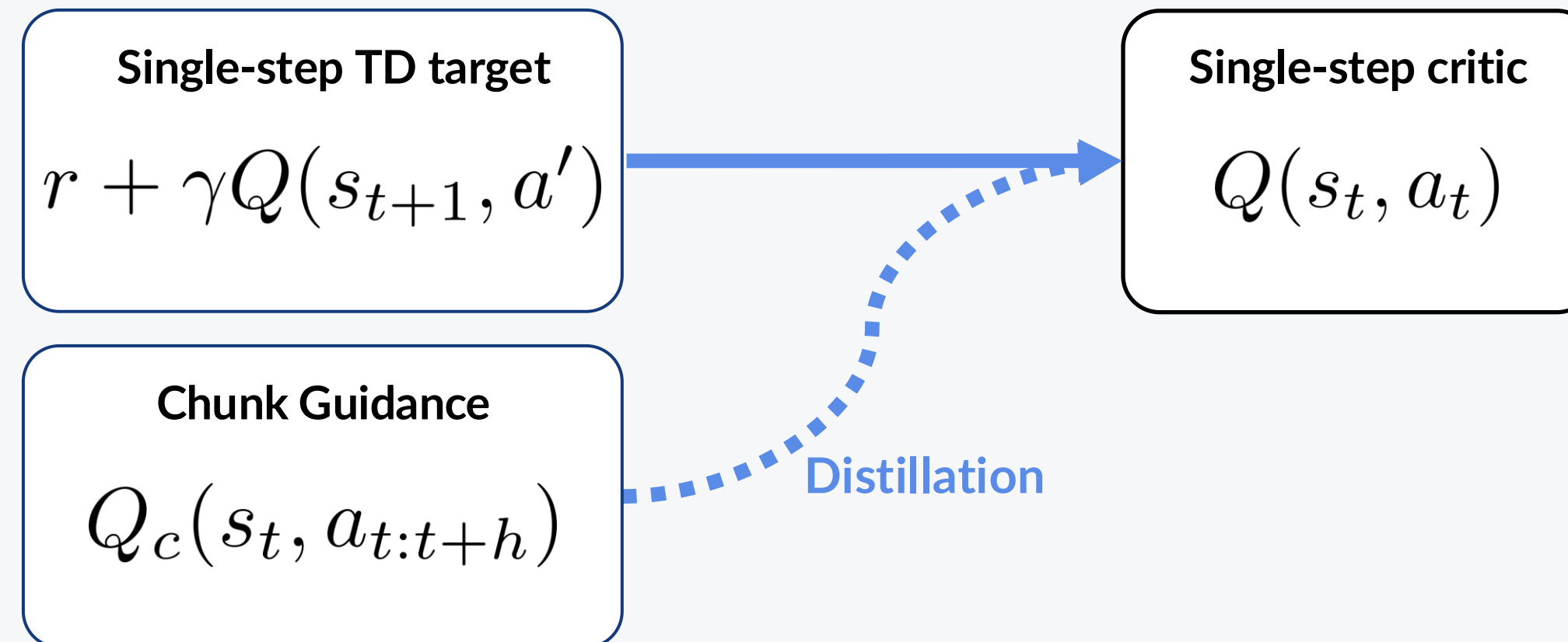
Action-chunked TD

Shorter TD backups, but suboptimal

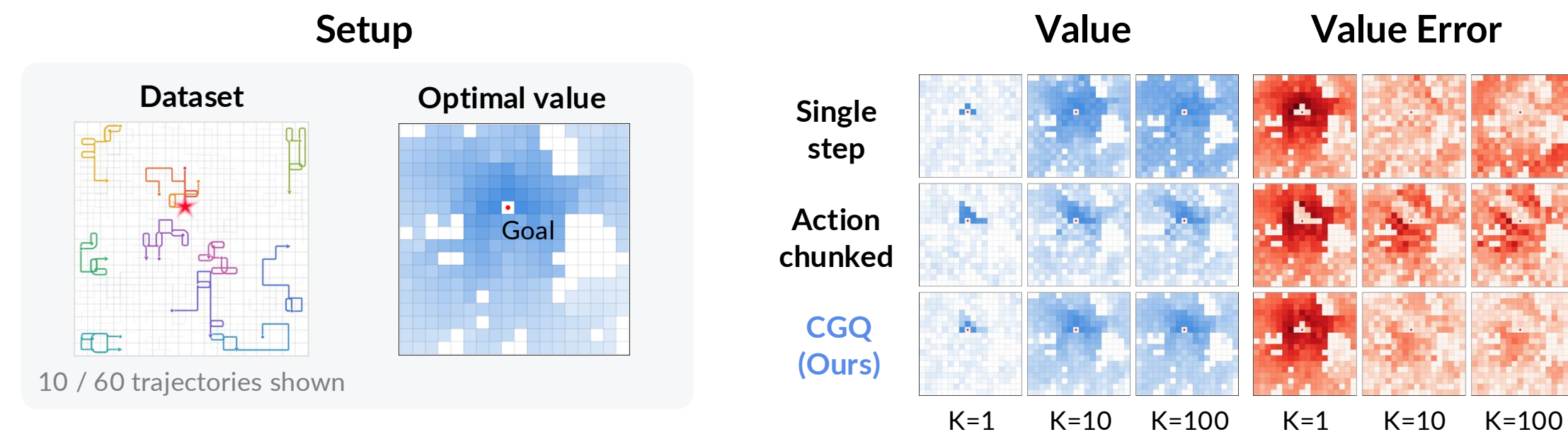


Q: How can a **single-step** critic learn stable values over **long horizons**?

A: Guiding it with an **action-chunked** critic.



Gridworld analysis



Single-step

Fine-grained value learning
Large error accumulation

Action-Chunked

Small error accumulation
Coarse value learning

CGQ

Best of Both!

Chunk-Guided Q-Learning

$$\mathcal{L}(\phi) = \mathcal{L}^{\text{TD}}(\phi) + \beta \mathbb{E}[\ell_{\tau}(Q_{\phi_c}(s, a) - Q_{\phi}(s, a))]$$

Use expectile regression to **selectively leverage coarse chunk guidance**

CGQ stabilizes long-horizon offline Q-learning by distilling coarse action-chunked value estimates into a fine-grained single-step critic

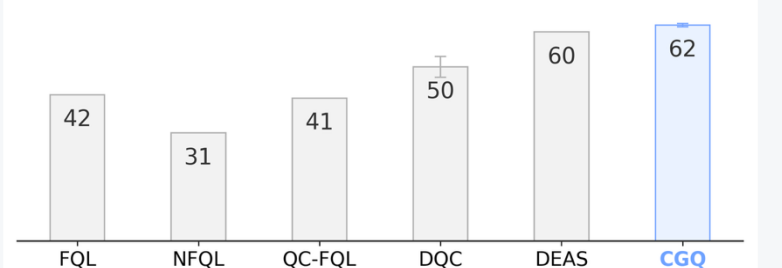
Experiments

Tasks

45 of Manipulation & Navigation

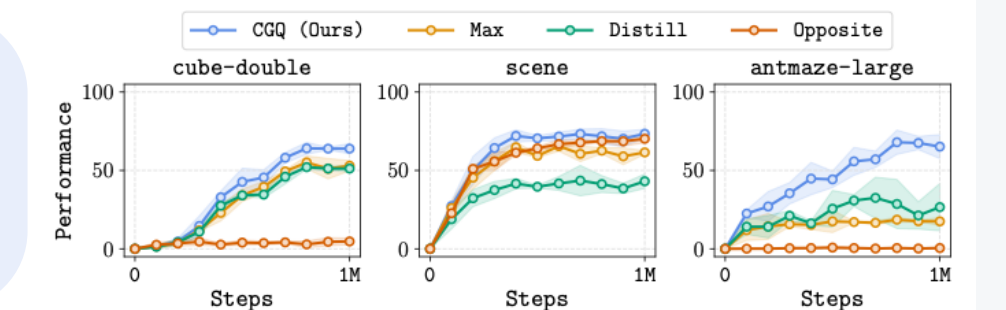


Results



Is CGQ the best critic design?

Yes, CGQ avoids overestimation, chunk only supervision, and open-loop execution.



Ablation study - Beta

Properly weighting chunk guidance improves performance

