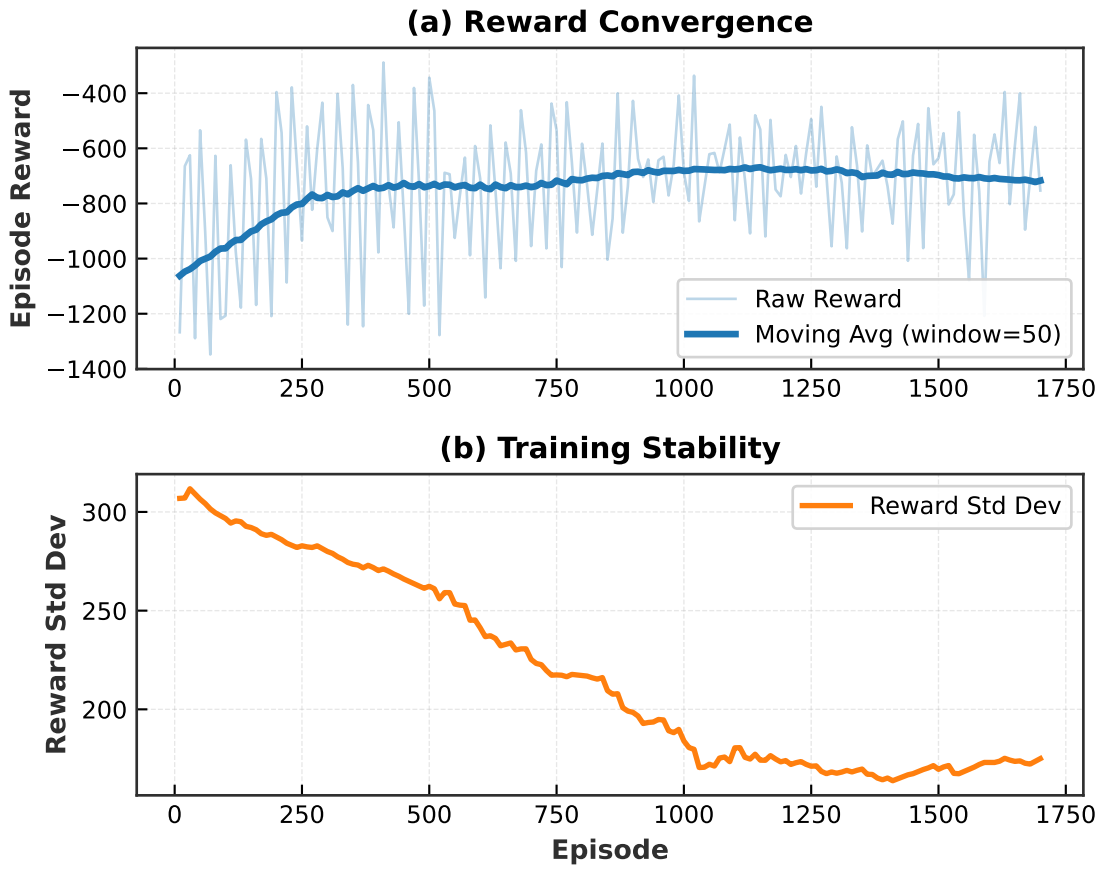
(a) Reward Convergence
Episode Reward
−400
−600
−800
−1000
−1200
−1400
0
250
500
750
1000
1250
1500
1750
Raw Reward
Moving Avg (window=50)
(b) Training Stability
Reward Std Dev
300
250
200
Reward Std Dev
0
250
500
750
1000
1250
1500
1750
Episode