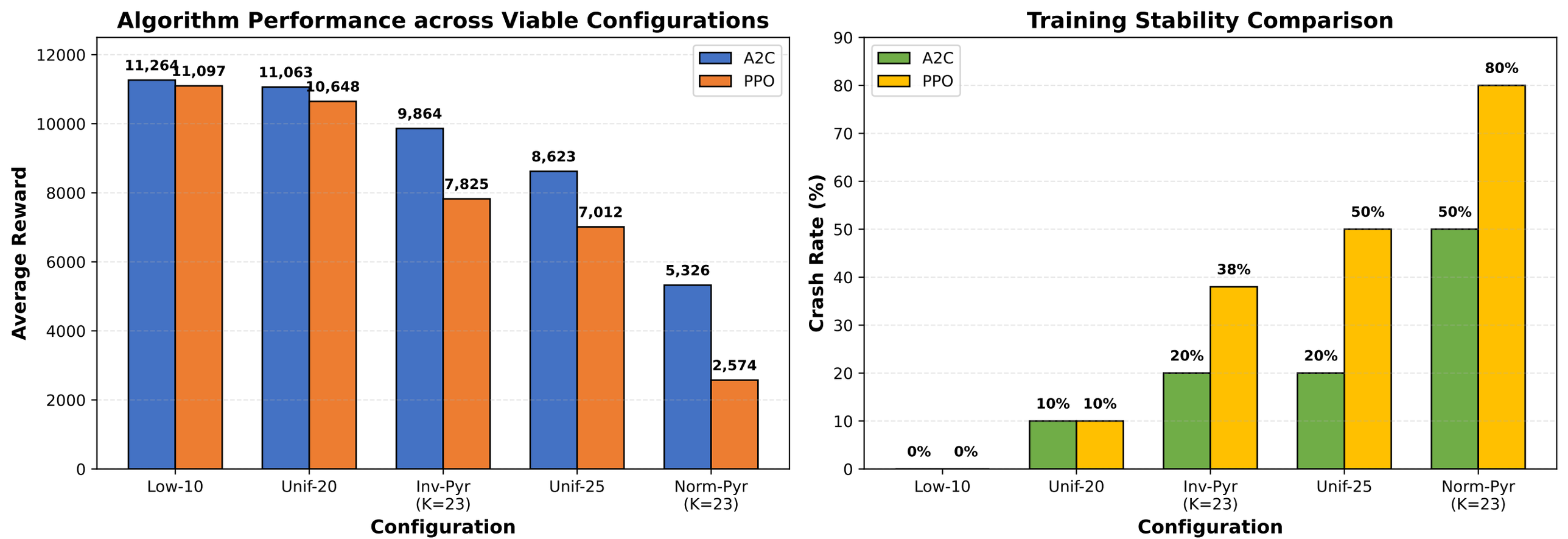
Algorithm Performance across Viable Configurations
Average Reward
Configuration
A2C
PPO
11,264
11,097
11,063
10,648
9,864
7,825
8,623
7,012
5,326
2,574
Low-10
Unif-20
Inv-Pyr (K=23)
Unif-25
Norm-Pyr (K=23)
Training Stability Comparison
Crash Rate (%)
0%
0%
10%
10%
20%
38%
20%
50%
50%
80%