



Figure 1: **Comparison of model performance across two task groups (wood and stone).** Results represent the average success rate (%) within each group. Our proposed method MoEA performs comparably to Optimus-1, and both outperform the other baselines.