Qualitative Results: SAM2 vs SAM2-Lite

Challenging scenarios: occlusions, deformations, and similar objects
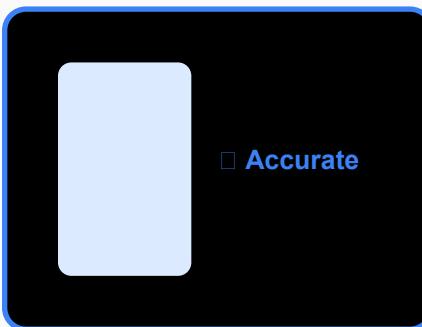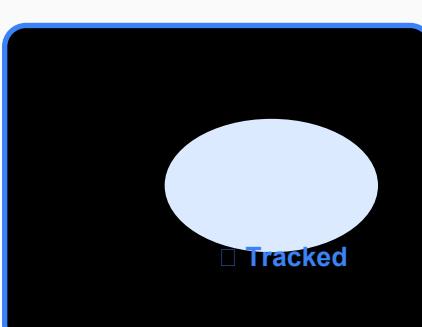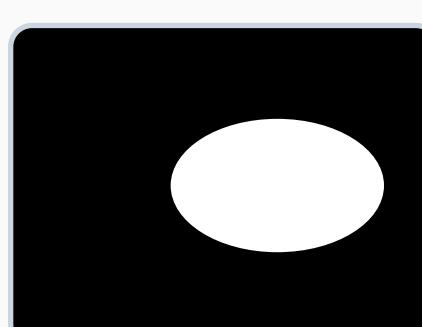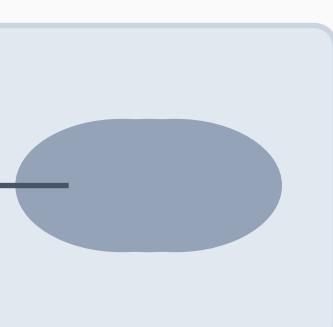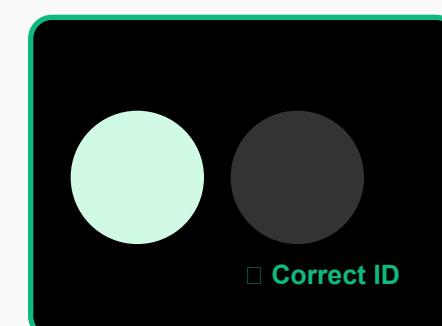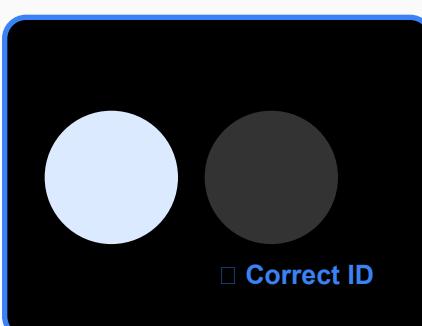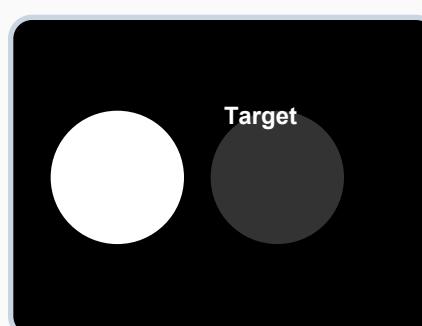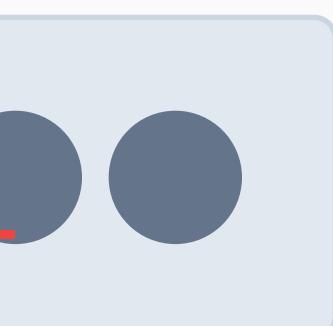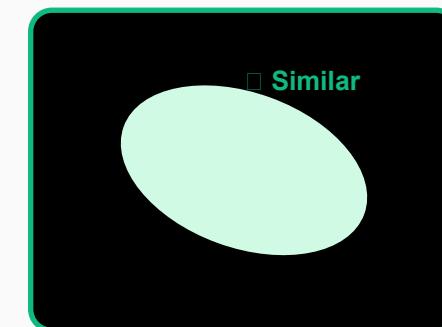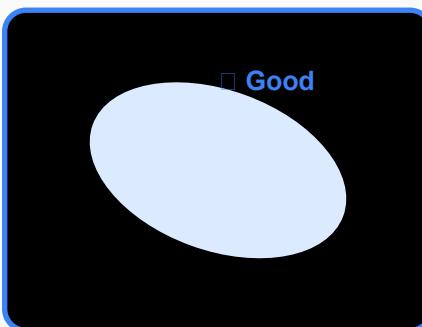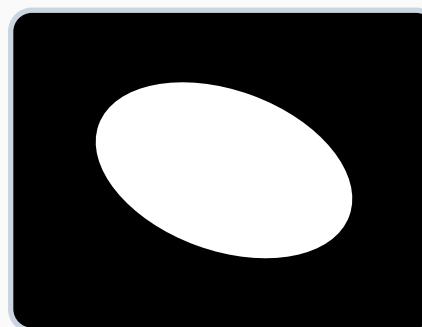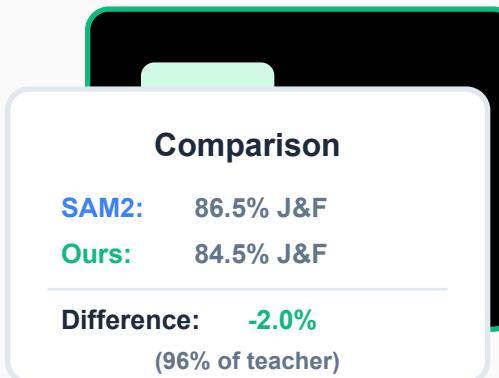
| Input Frame	Ground Truth	SAM2 (Teacher)	SAM2-Lite (Ours)

Occlusion
Person behind tree
✓ Accurate

Comparison
SAM2: 86.5% J&F
Ours: 84.5% J&F
Difference: -2.0%
(96% of teacher)

Deformation
Dancing person
✓ Good
✓ Similar

Similar Objects
Multiple similar faces
Target
✓ Correct ID
✓ Correct ID

Fast Motion
Ball in motion (blur)
✓ Tracked
✓ Tracked
(slight edge softness)

Key Observation: SAM2-Lite produces visually similar results despite using 48× fewer parameters
Minor differences in edge sharpness and very small object details | Overall tracking accuracy maintained