Metrics - How to evaluate the **realism** of a generated video?
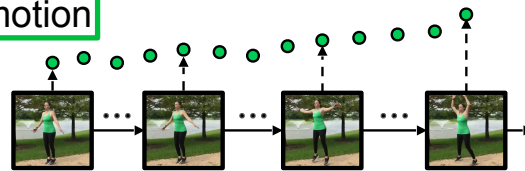
Is the action **semantically correct?**

Generated jumping jacks embedding
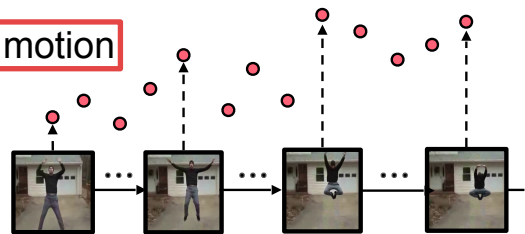
Real jumping jacks embeddings
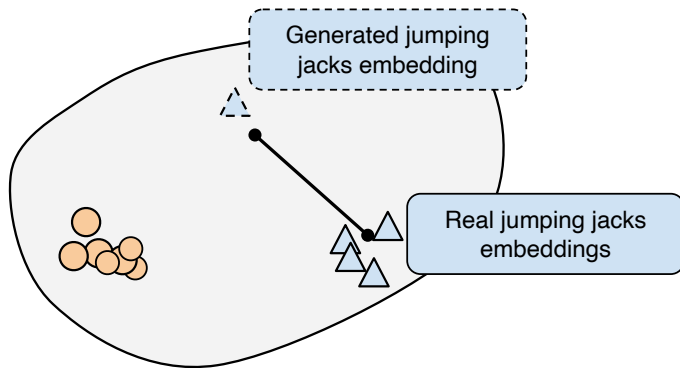
Action Consistency = Distance of generated video embedding to centroid of real video embeddings

Is the motion **temporally coherent?**

Coherent motion

Incoherent motion

Trajectory of per-frame embeddings

Temporal Coherence =
Mean distance between adjacent frame embeddings