

**Video file (e.g. mp4, avi)**



*sample frames (e.g. 1 per second)*

Frame 1

Frame 2

Frame 3

...

Frame N



*embed each frame (e.g. ResNet)*

[2048 vals]

[2048 vals]

[2048 vals]

...

[2048 vals]



*aggregate (mean across frames)*

**Video embedding: [2048 values]**

*(2048 is ResNet50; other models vary)*