SpeedNet: Learning the Speediness in Videos

Sagie Benaim, Ariel Ephrat, Oran Lang, Inbar Mosseri, William T. Freeman, Michael Rubinstein, Michal Irani, Tali Dekel

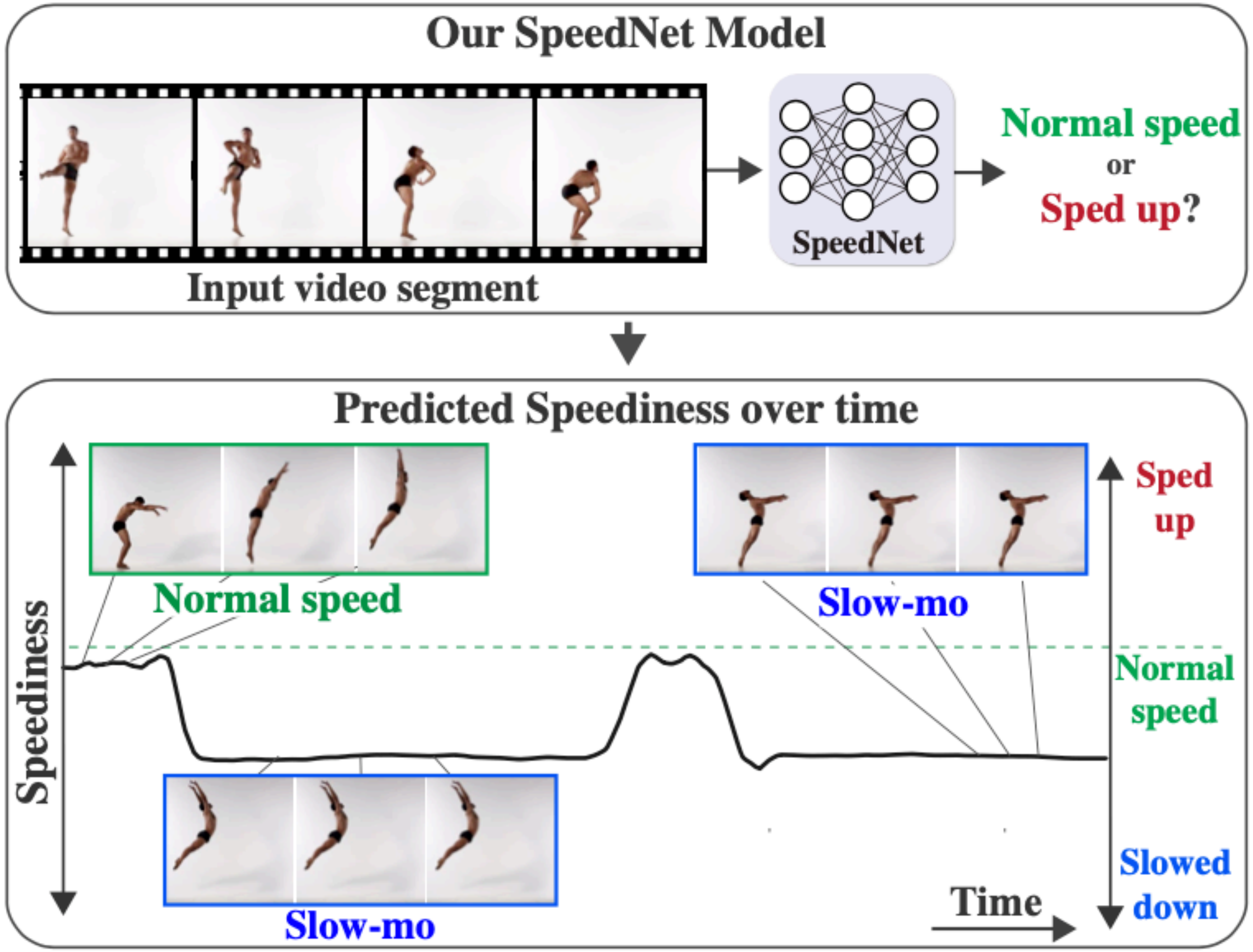
Google Research, Tel Aviv University, Weizmann Institute

CVPR 2020

Task

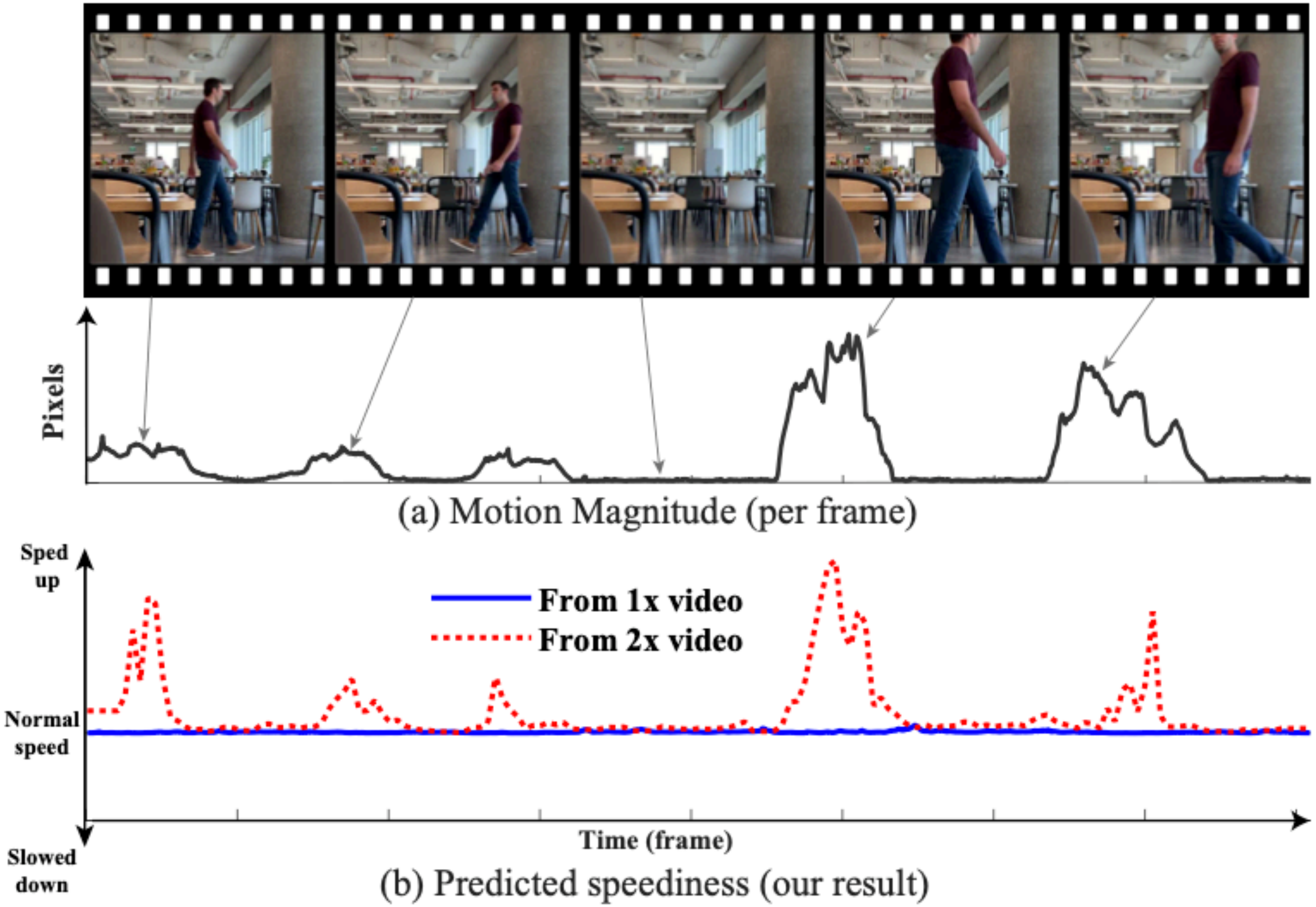
Predict the “speediness”

- Predict the “**speediness**” of moving objects in videos
- Whether the video is playing **natrually**, or **faster / slower**?
- Requires high-level reasoning:
- *The ability to correctly classify an object’s speed continues to improve even throughout adolescence*



Task

Speediness != motion magnitude



A man walking away and then towards the camera

Task

Formulation

- Binary classification: 1x / 2x

- **Input**: L frames from an L-fps video

- **Output**: 1 second or 2 seconds (*i.e.*, normal speed or sped-up).

Method

Data, supervision, and avoiding artificial cues

- **Self-supervised:** Generating normal and faster segments
- **Training dataset**: Kinetics
- **Network**: S3D-G

- *Avoid the tendency to use shortcuts—artificial cues present in the training data:*
- **Spatial augmentations:** 64x64 ~ 336x336
- **Temporal augmentations:** 1× ~ 1.2× / 1.7× ~ 2.2×
- **Same-batch training:** segment pair from the same source video in the same batch

Application

Adaptive video speedup:

- Non-uniformly change the speed of a video based on its content without corrupting its "naturalness"
- **For a test video:**
- Run the model on the video sped-up with a set of different factors.
- **V(t):** the maximum speedup factor at each time that was still classified as 'normal'
- **Optimizing for adaptive speedup:**

$$\arg\min_S E_{\text{speed}}(S, V) + \beta E_{\text{rate}}(S, R_o) + \alpha E_{\text{smooth}}(S')$$

***S**: Output speed*

***R0**: User-desired speed*

***S'**: First derivative of S*

Application

Adaptive video speedup:

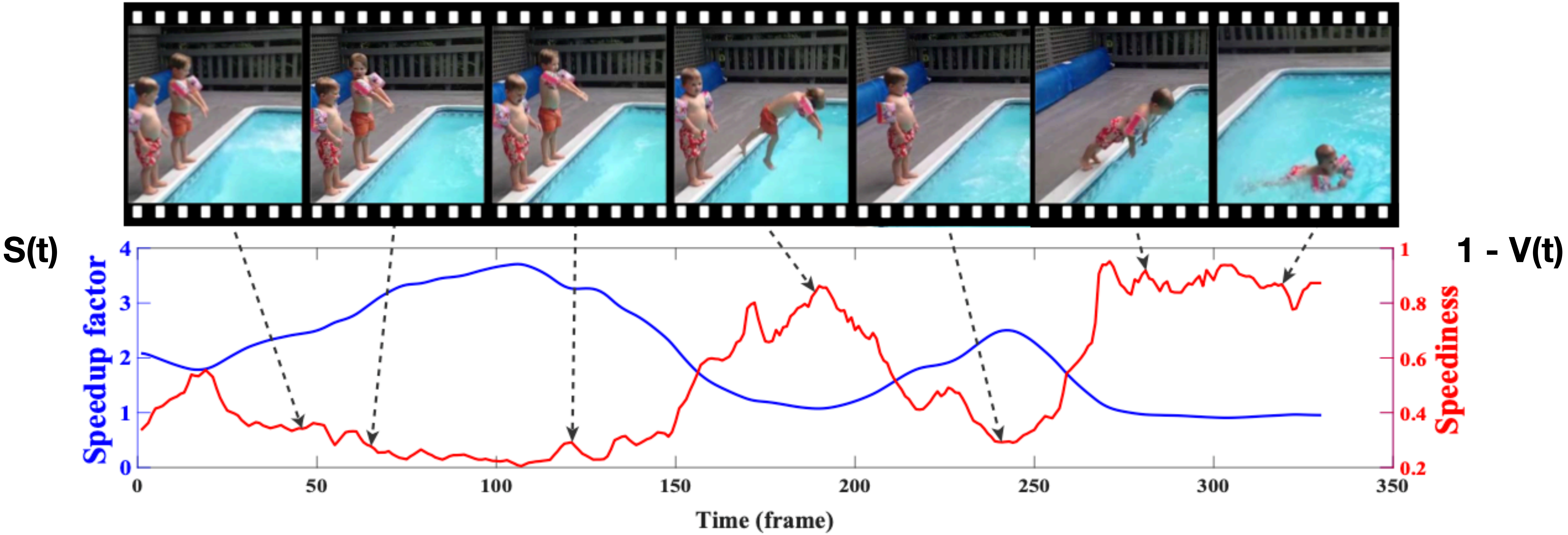
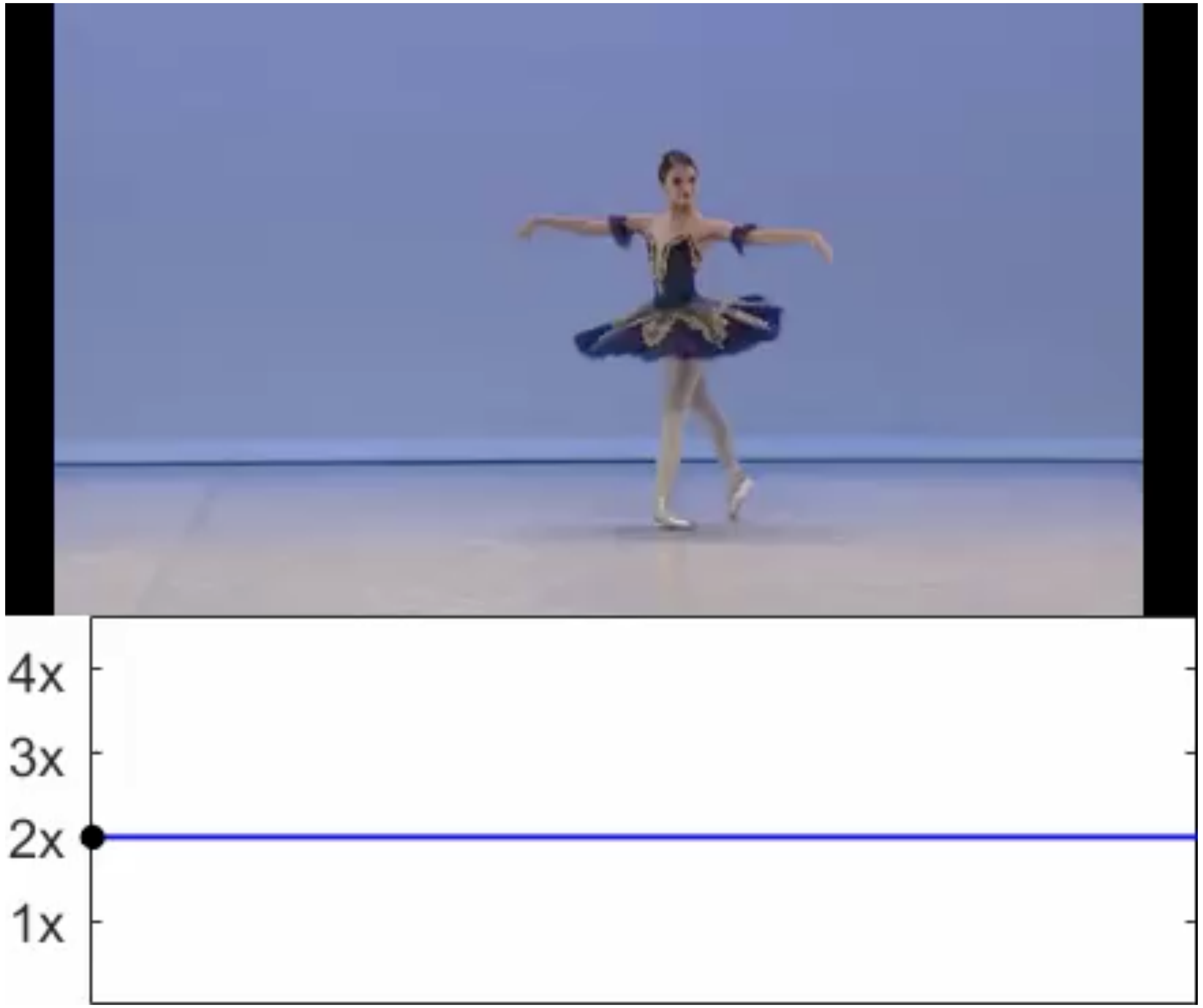
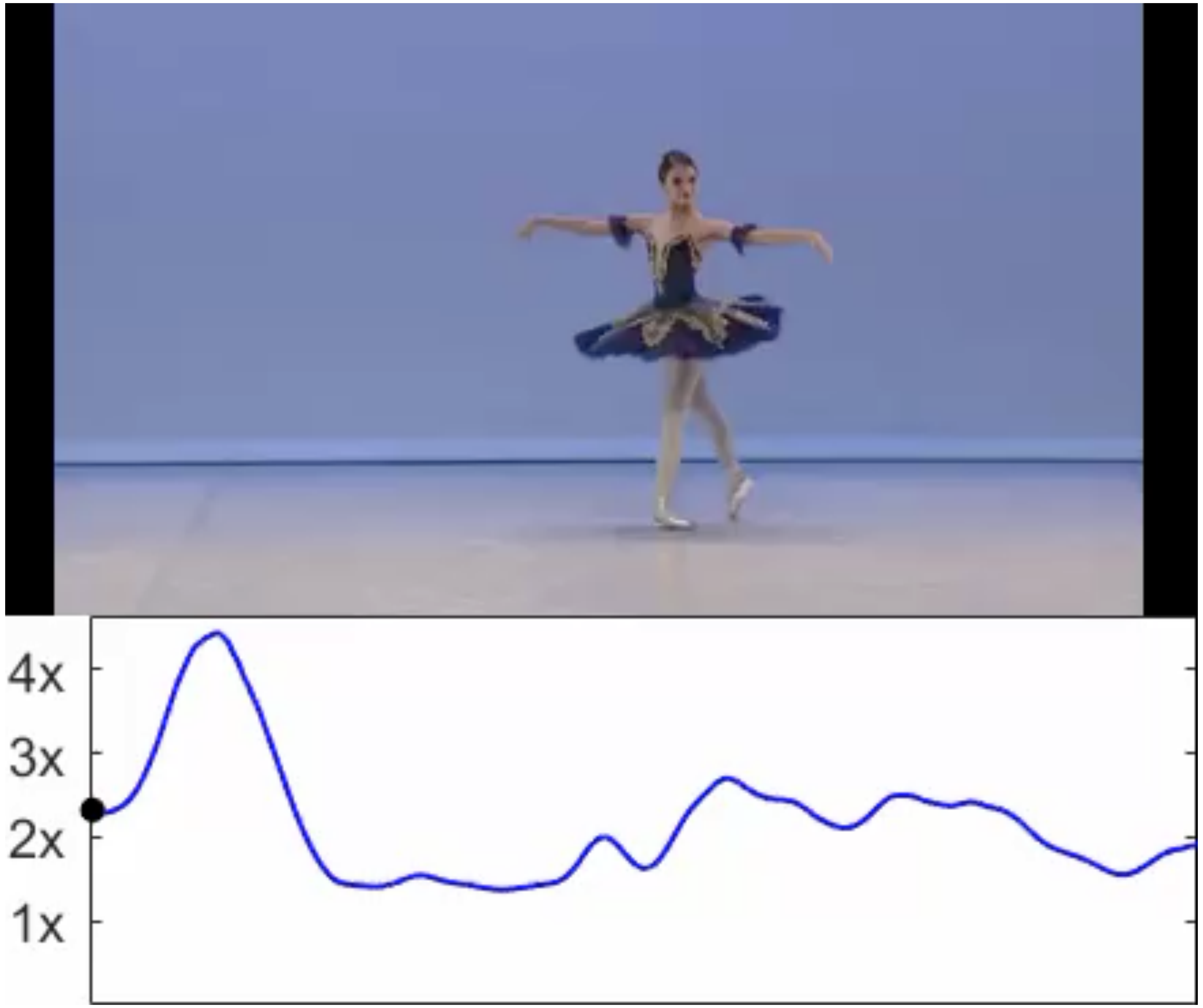


Figure 5. **Adaptive video speedup.** We apply the SpeedNet model for generating time-varying, adaptive speedup videos, based on the frames' speediness curve (Sec. 4). Here we show the speediness curve and our resulting adaptive speedup factor for a video of two kids jumping into a pool. Several selected frames are shown at the top, pointing to their corresponding times within the sequence on the predicted speediness curve.

Application

Adaptive video speedup:





Experiment

Binary classification accuracy

Model Type			Accuracy	
Batch	**Temporal**	**Spatial**	**Kinetics**	**NFS**
Yes	Yes	Yes	75.6%	73.6%
No	Yes	Yes	88.2%	59.3%
No	No	Yes	90.0%	57.7%
No	No	No	96.9%	57.4%
	Mean Flow		55.8%	55.0%

Speediness
!=
motion magnitude

Training data: Kinetics training set

Testing data: Kinetics test set & NFS

Experiment

Adaptive speedup of real-world videos (See demo)

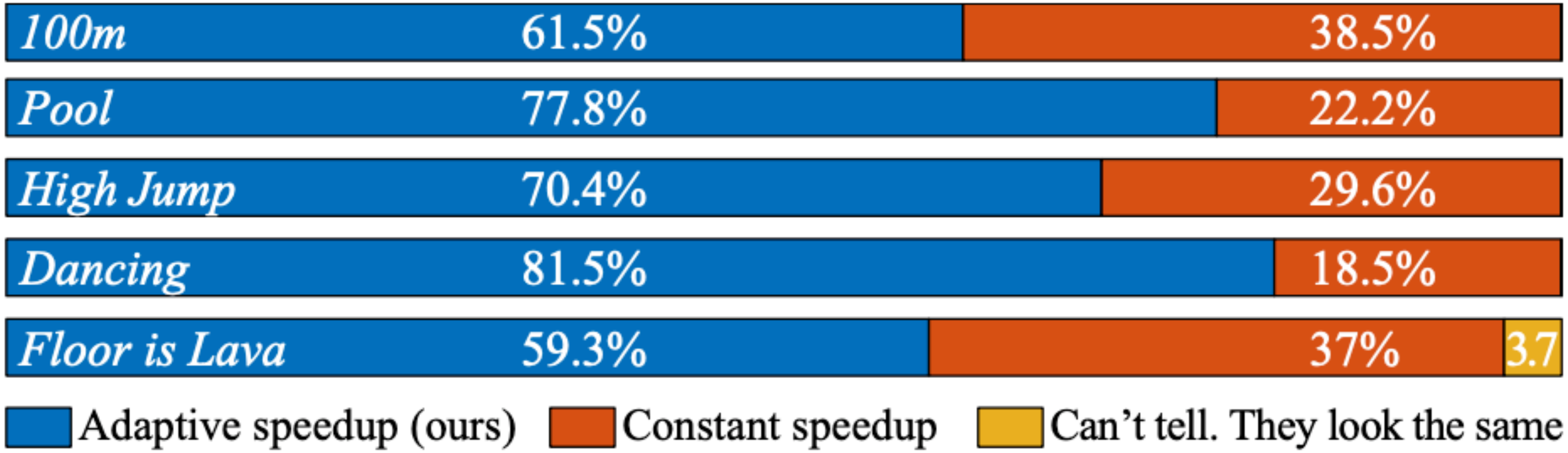


Figure 6. **Adaptive video speedup user study.** We asked 30 participants to compare our adaptive speedup results with constant uniform speedup for 5 videos (without saying which is which), and select the one they liked better. Our adaptive speedup results were consistently (and clearly) preferred over uniform speedup.

Experiment

Self-supervised action recognition

Initialization		Supervised accuracy	
Method	**Architecture**	**UCF101**	**HMDB51**
Random init	S3D-G	73.8	46.4
ImageNet inflated	S3D-G	86.6	57.7
Kinetics supervised	S3D-G	96.8	74.5
CubicPuzzle [19]	3D-ResNet18	65.8	33.7
Order [40]	R(2+1)D	72.4	30.9
DPC [13]	3D-ResNet34	75.7	35.7
AoT [38]	T-CAM	79.4	-
SpeedNet (Ours)	S3D-G	**81.1**	**48.8**
Random init	I3D	47.9	29.6
SpeedNet (Ours)	I3D	66.7	43.7

Self-Supervised on Kinetics
Then fine-tuned

Conclusion and comments

- Learn the “speediness” of moving objects in videos
- Self-supervised learning
- Effective for several tasks

- Novel task
- Useful application
- Simple method