HAVANA: Hierarchical stochastic neighbor embedding for Accelerated Video ANnotAtions

TUDelft   Alexandru Bobe and Jan C. van Gemert   TUDelft
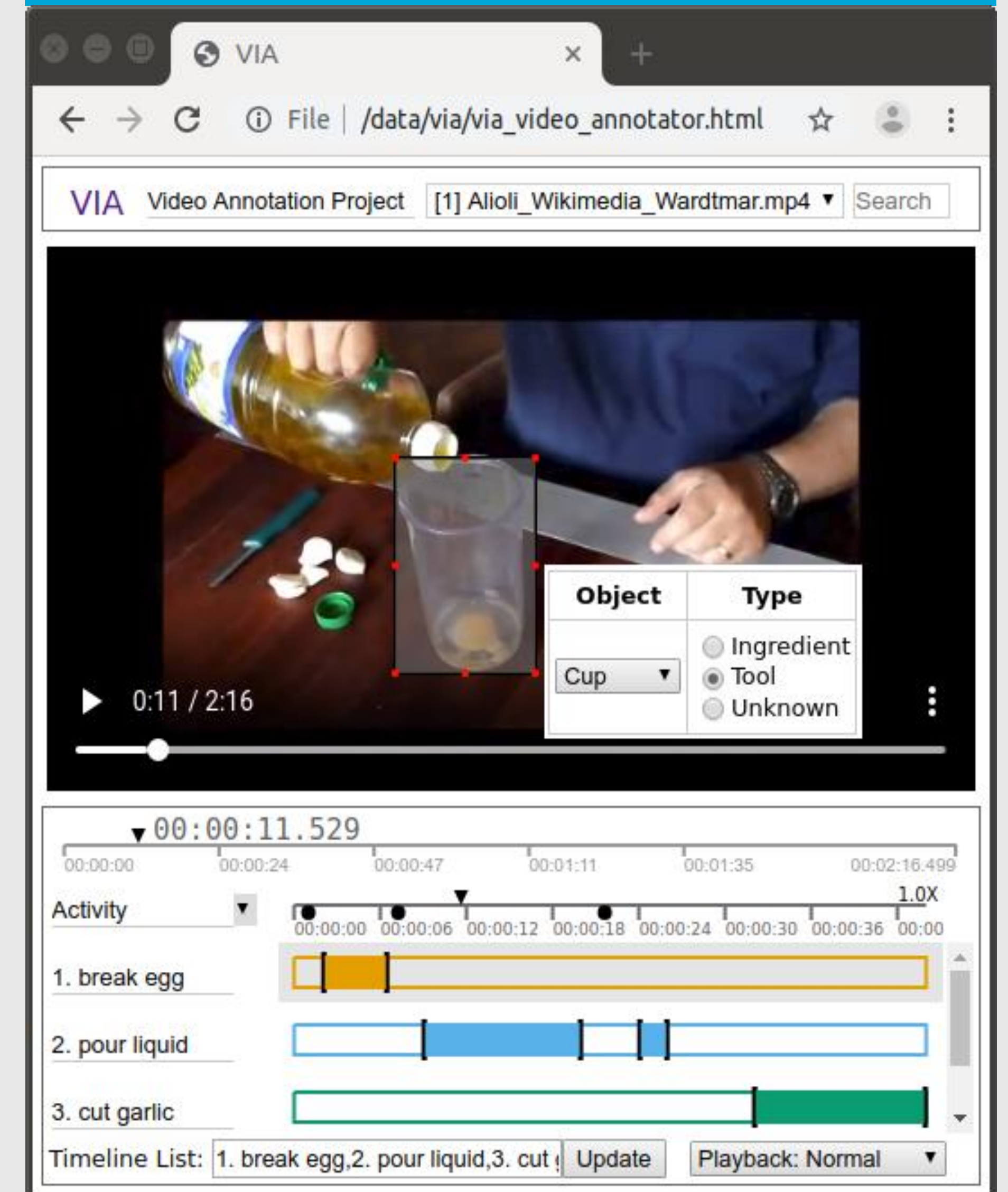
1. Motivation

- **Temporal video annotation** is expensive
- Similarities between videos used to speed up the **annotation process**
- Watching each video is not necessary
- Feature extraction and dimensionality reduction in **our pipeline**
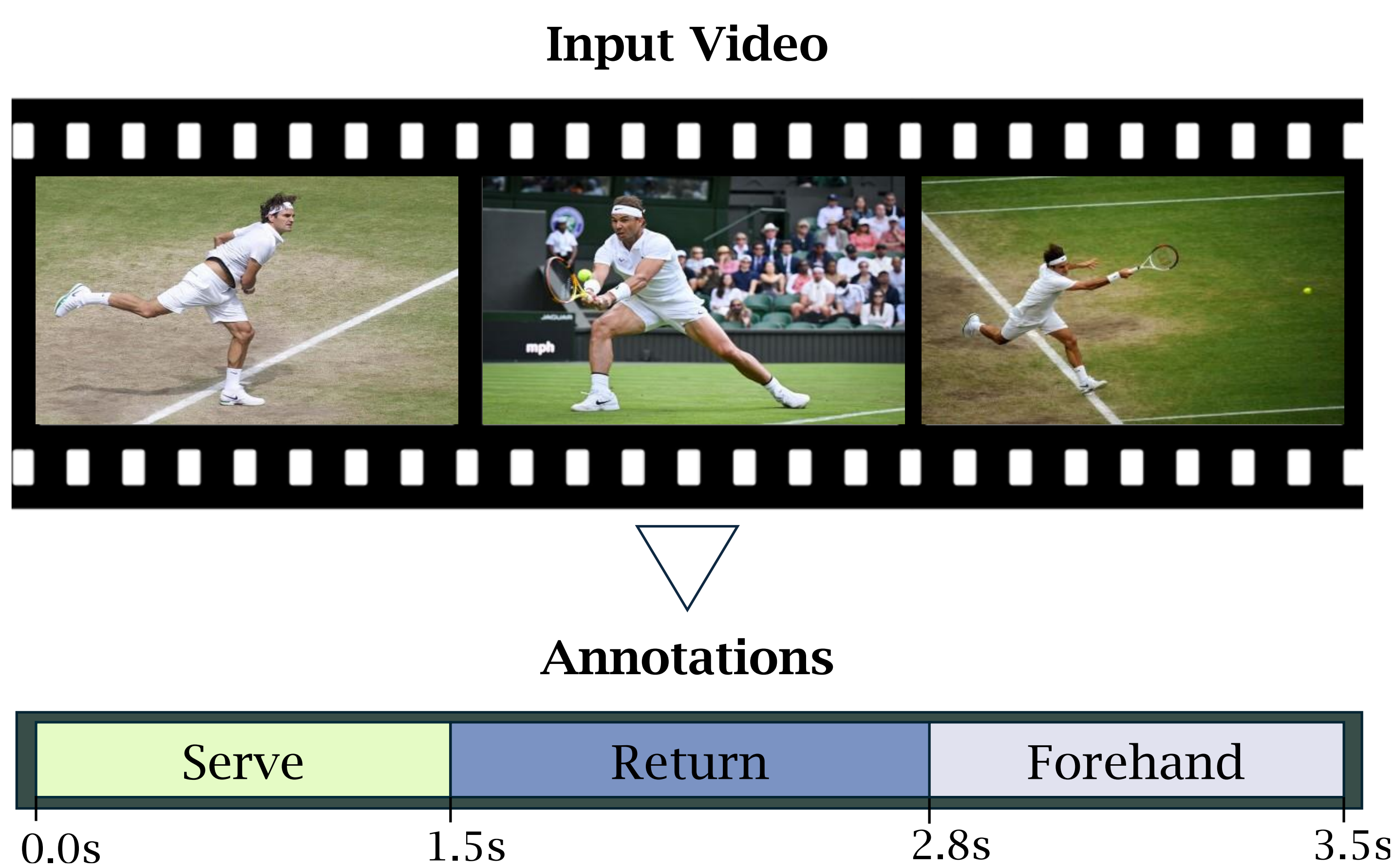- Suitable for large amounts of videos
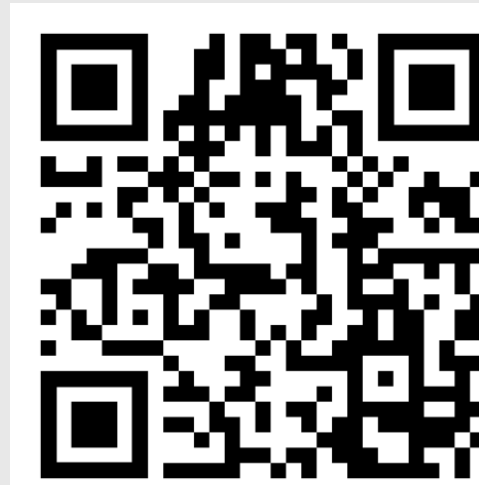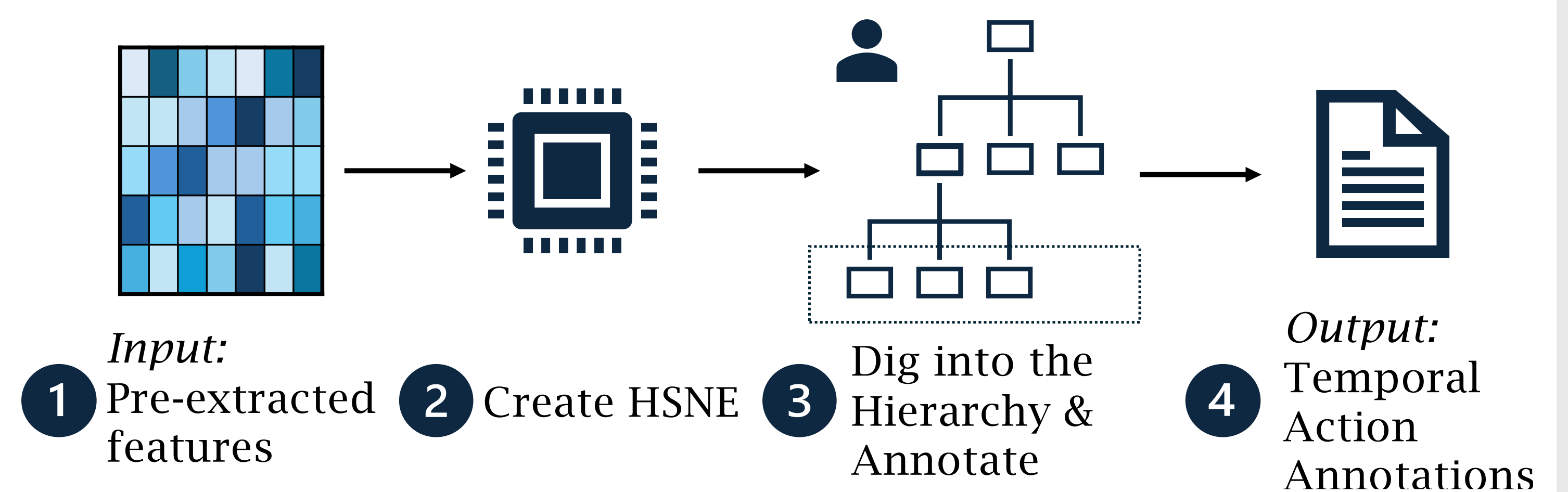
2. Hierarchical SNE

Landmarks LS — Scale S (displayed first)

S-2 Scales

Landmarks L1 — Scale 1 (dataset)
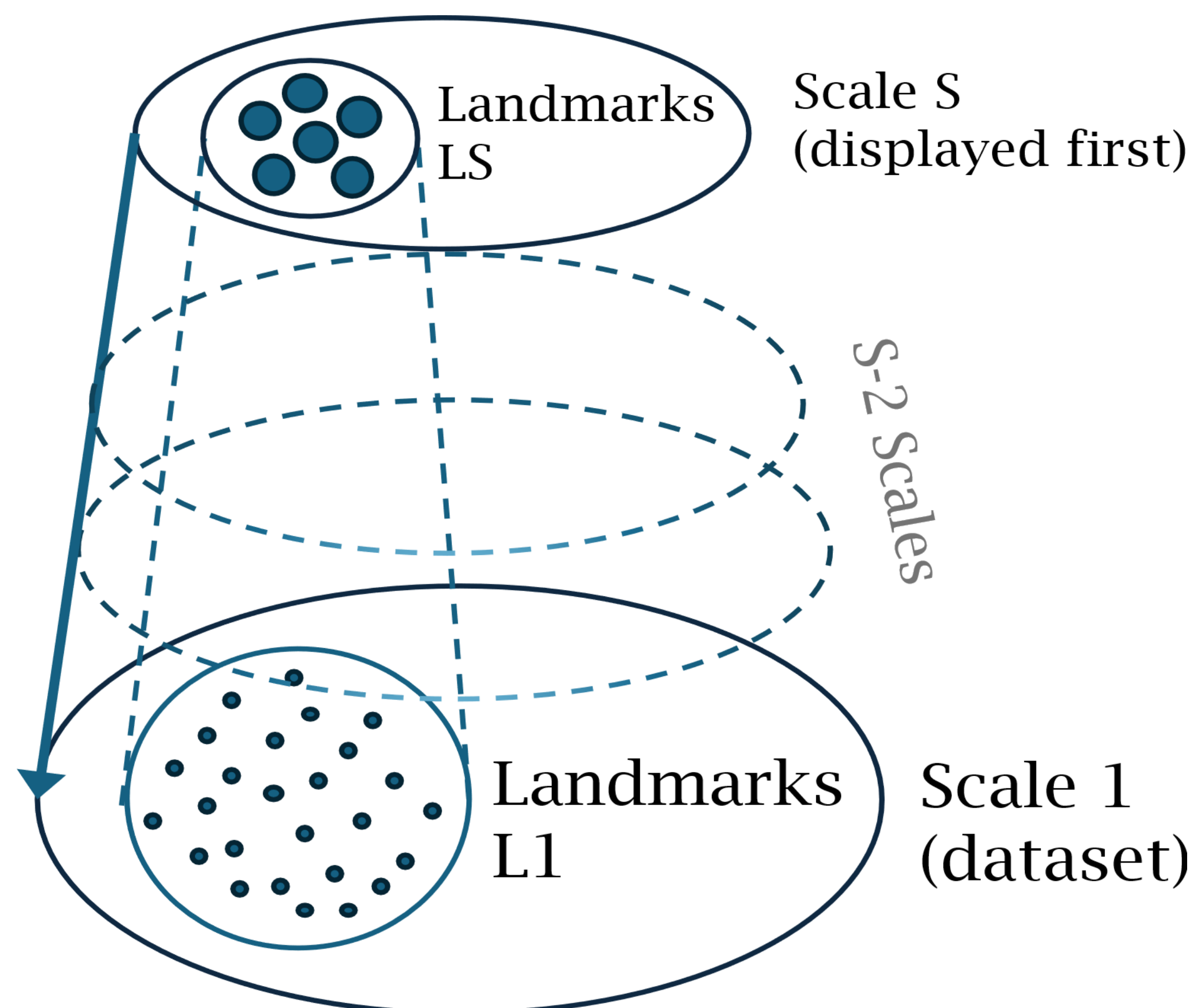
4. Before - Linear

3. Temporal Video Annotation

Input Video

Annotations

Serve	Return	Forehand

0.0s   1.5s   2.8s   3.5s

5. Our pipeline

Input: ① Pre-extracted features   ② Create HSNE   ③ Dig into the Hierarchy & Annotate   ④ *Output:* Temporal Action Annotations

SCAN FOR THE CODE!

7. THUMOS14 Test Set

Estimated Effort Required for Annotating THUMOS14

- Our pipeline
- Traditional Method (Linear)
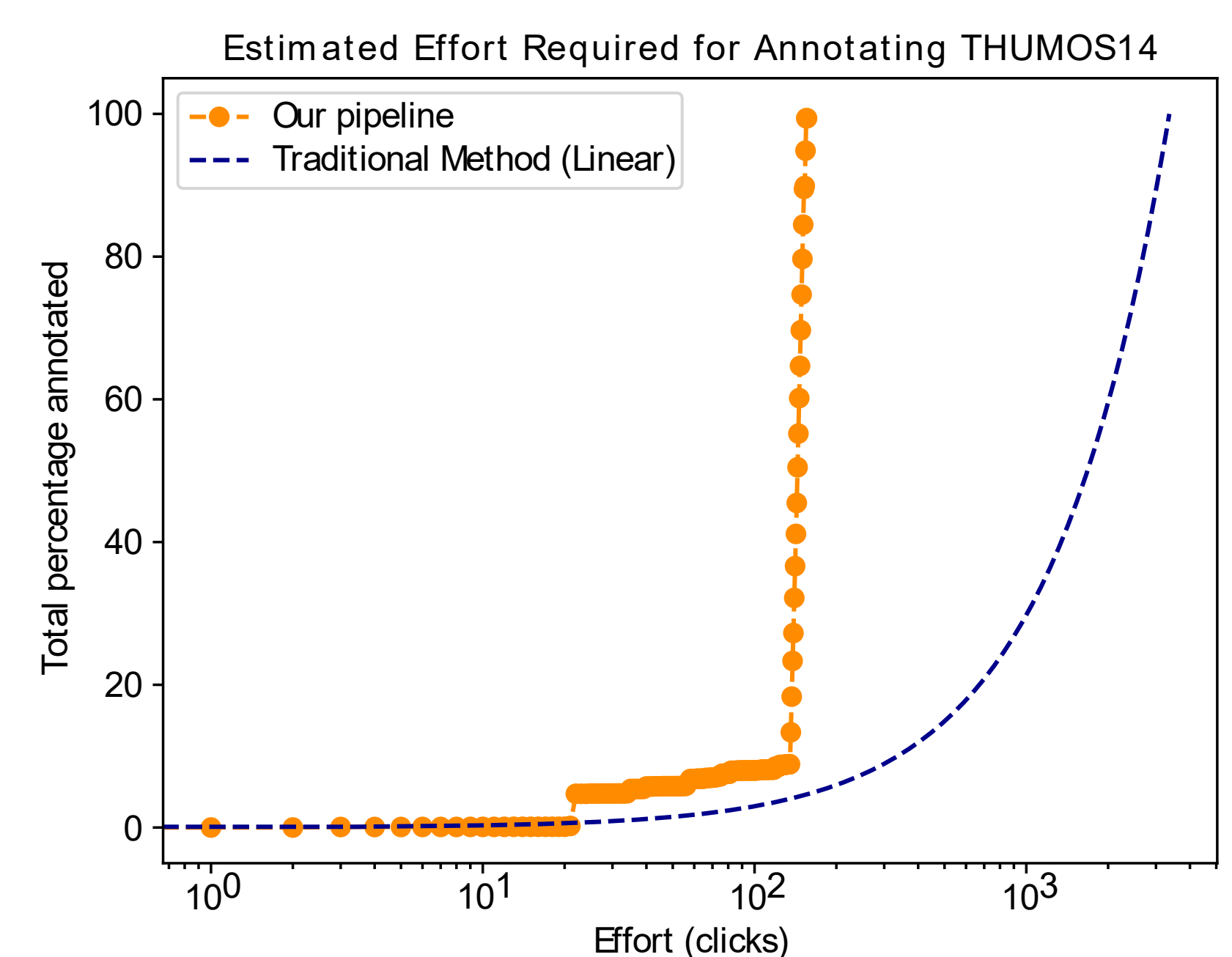
- 10x improvement in terms of *clicks* for annotating the THUMOS14 test set

6. Software Visualization

Analysis[id=1, num points=1937, scale=1]

Class: segment_gt_label

Group	Color
Background	
BaseballPitch	
BasketballDunk	
Billiards	
CleanAndJerk	
CricketBowling	
CricketShot	
Diving	
FrisbeeCatch	
GolfSwing	
HammerThrow	
HighJump	
JavelinThrow	
LongJump	
PoleVault	
Shotput	

Landmarks selected: 92   New Analysis

Iteration: Completed
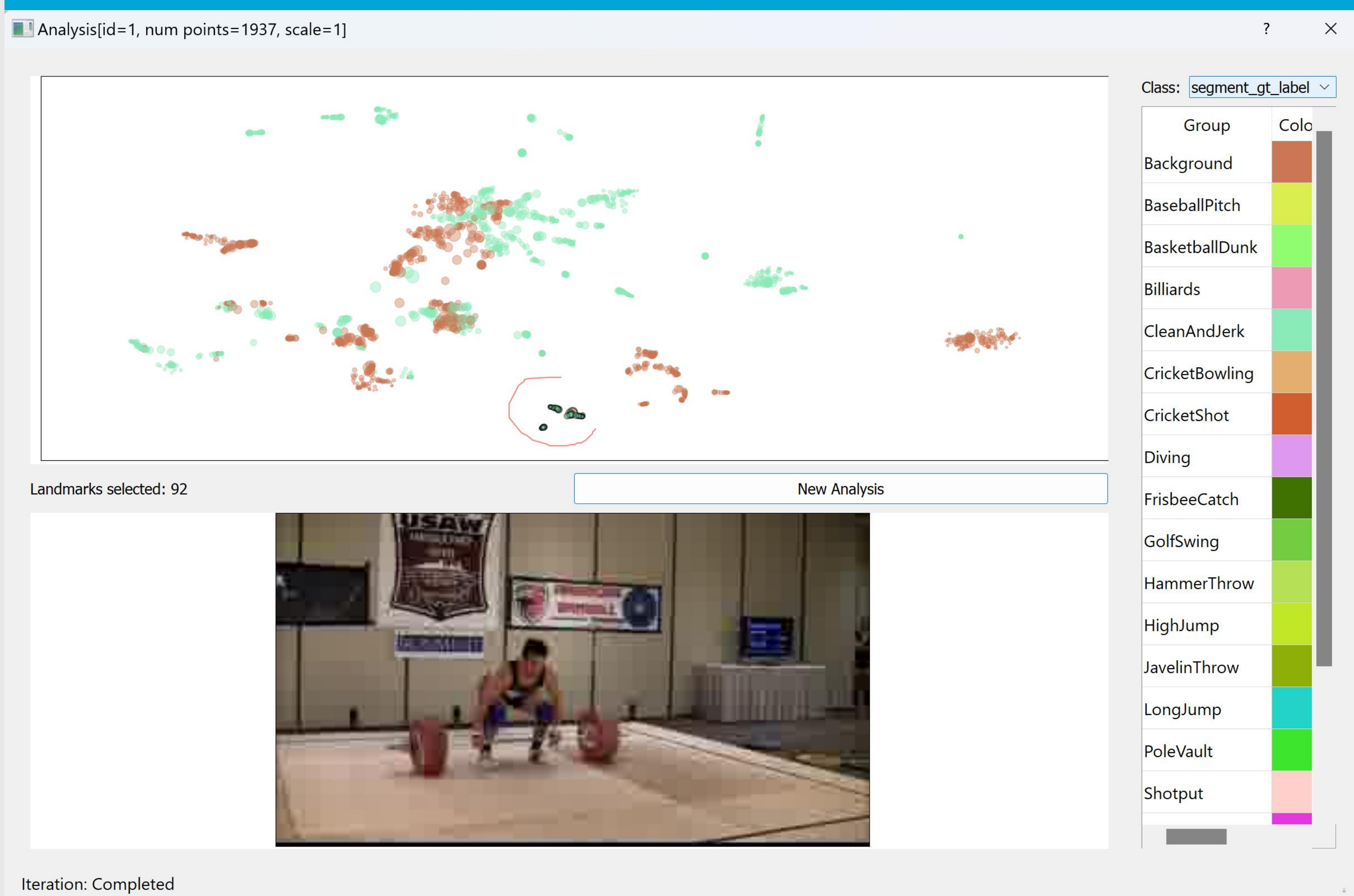
8. Conclusion

- Better than linear video annotation
- Pipeline works with multiple datasets and parameters
- Parts of the annotation can be automated to save more effort