A. Artifact Appendix

A.1 Abstract

This appendix describes the python implementation of the paper "Unsupervised Method for Video Action Segmentation Through Spatio-Temporal and Positional-Encoded Embeddings".

A.2 Artifact check-list (meta-information)

- **Algorithm:** ID3, Slowfast, FINCH, K-means
- **Transformations:** Temporal Embeddings, Positional Encoding
- **Data set:** Breakfast, INRIA
- **Run-time environment:** Python, Conda
- **Hardware:** 6 cores i7 2.60 GHz CPU, RTX-2070 Max-Q Design GPU
- **Execution:** Conda virtual env, Jupyter notebook
- **Metrics:** MoF, IoU
- **Experiments:** Action segmentation, positional encoding, temporal window
- **Code licenses (if publicly available)?:** MIT License

A.3 Description

A.3.1 How delivered

This project is delivered via the git repository[1]. The experiments run in a Conda environment, and the experiment's codes are available as ipynb files.

A.3.2 Hardware dependencies

A CPU with 32GB of RAM and a GPU with at least 8GB of RAM is recommended.

A.3.3 Software dependencies

Python version 3.8, Conda, pip, video_features, decord, and pyTorch

A.3.4 Data sets

Breakfast dataset[2], and INRIA Instructionals dataset[3].

A.4 Installation

1. Install **Conda**[4];
2. Install **video_features**[5];
3. Install **decord**[6] for efficient video reading;
4. Run > conda env create -f environment.yml

A.5 Experiment workflow

1. Execute the extract_features.ipynb file to extract video features using I3D and Slowfast;
2. Run evaluation_inria.ipynb file to execute the experiment on the INRIA dataset;
3. Run evaluation_breakfast.ipynb file to execute the experiment on the Breakfast dataset.

A.6 Evaluation and expected result

Expected results for the experiment are in Tables 1 and 2 for the Breakfast and INRIA datasets respectively. Expected Results for the experiment with temporal window lengths variation are displayed in Table 3 for the Breakfast dataset using combinations of Slowfast, FINCH and Postional Encoding. Table 4 is similar, but for the I3D, FINCH and Positional Encoding combinations using the same dataset. The INRIA results for the Slowfast features behave very similarly to the Breakfast dataset, and are depicted in Tables 5 and 6 for the Slowfast and I3D feature extraction methods respectively.

Table 1. Experiment result with the Breakfast dataset

#	Method	MoF	IoU
01	Slowfast-32+KMeans	56.5	33.8
02	I3D-10+FINCH	55.33	27.83
03	I3D-10+KMeans+PE	54.7	29.4
04	I3D-10+KMeans	54.2	29.4
05	Slowfast-32+FINCH	53.9	27.5
06	I3D-10+FINCH+PE	53.89	29.68
07	Slowfast-32+FINCH+PE	53.2	40.4
08	Slowfast-32+KMeans+PE	45.1	43.4

Table 2. Experiment result with the INRIA dataset

#	Method	MoF	F1-Score
01	I3D-10+FINCH	49.85	43.42
02	I3D-10+FINCH+PE	47.25	43.22
03	Slowfast-32+FINCH	45.83	40.27
04	Slowfast-32+FINCH+PE	45.47	40.17
05	Slowfast-32+KMeans	45.4	39.89
06	Slowfast-32+KMeans+PE	44.93	39.71
07	I3D-10+KMeans	44.69	39.23
08	I3D-10+KMeans+PE	41.22	36.72

[1] https://github.com/gpmarques/unsup_action_seg_st_pe_embed

[2] https://serre-lab.clps.brown.edu/resource/breakfast-actions-dataset/

[3] https://www.di.ens.fr/willow/research/instructionvideos/data_new.tar.gz

[4] https://docs.conda.io/en/latest/miniconda.html

[5] https://github.com/v-iashin/video_features

[6] https://github.com/dmlc/decord

Table 3. Temporal window length variation with Slowfast, FINCH and Positional Encoding combinations for the Breakfast dataset

#	Method	MoF	IoU
01	Slowfast-72+FINCH+PE	57.38	32.17
02	Slowfast-64+FINCH+PE	57.11	32.75
03	Slowfast-48+FINCH+PE	56.72	36.03
04	Slowfast-128+FINCH+PE	55.86	25.21
05	Slowfast-40+FINCH+PE	54.75	36.94
06	Slowfast-48+FINCH	54.07	27.05
07	Slowfast-72+FINCH	53.92	25.24
08	Slowfast-32+FINCH	53.90	27.5
09	Slowfast-32+FINCH+PE	53.20	40.40
10	Slowfast-128+FINCH	51.2	20.42
11	Slowfast-40+FINCH	50.71	24.84

Table 4. Temporal window length variation with I3D, FINCH and Positional Encoding combinations for the Breakfast dataset

#	Method	MoF	IoU
01	I3D-64+FINCH	57.99	29.61
02	I3D-48+FINCH+PE	57.75	34.3
03	I3D-40+FINCH+PE	57.73	33.32
04	I3D-64+FINCH+PE	57.49	35.52
05	I3D-32+FINCH+PE	57.47	32.98
06	I3D-48+FINCH	57.42	28.87
07	I3D-24+FINCH+PE	56.33	32.09
08	I3D-40+FINCH	56.07	29.25
09	I3D-16+FINCH	55.45	30.49
10	I3D-32+FINCH	55.4	28.6
11	I3D-10+FINCH	55.33	27.83
12	I3D-24+FINCH	54.49	27.8
13	I3D-16+FINCH	54.28	27.67
14	I3D-10+FINCH+PE	53.89	29.68

Table 5. Temporal window length variation with Slowfast, FINCH and Positional Encoding combinations for the IRIA dataset

#	Method	MoF	IoU
01	Slowfast-128+FINCH+PE	53.89	46.32
02	Slowfast-64+FINCH+PE	52.91	44.81
03	Slowfast-48+FINCH+PE	51.56	44.48
04	Slowfast-40+FINCH+PE	51.45	45.91
05	Slowfast-48+FINCH	51.34	44.17
06	Slowfast-64+FINCH	50.09	43.75
07	Slowfast-40+FINCH	49.53	43.74
08	Slowfast-128+FINCH	48.47	42.26
09	Slowfast-32+FINCH	45.83	40.27
10	Slowfast-32+FINCH+PE	45.47	40.17

Table 6. Temporal window length variation with I3D, FINCH and Positional Encoding combinations for the IRIA dataset

#	Method	MoF	IoU
01	I3D-64+FINCH+PE	53.35	45.97
02	I3D-48+FINCH	53.08	44.37
03	I3D-40+FINCH	52.86	44.42
04	I3D-64+FINCH	52.74	44.56
05	I3D-48+FINCH+PE	52.51	45.79
06	I3D-16+FINCH	50.91	43.95
07	I3D-40+FINCH+PE	50.77	44.54
08	I3D-10+FINCH	49.85	43.42
09	I3D-16+FINCH+PE	49.24	44.56
10	I3D-32+FINCH	48.88	42.27
11	I3D-32+FINCH+PE	48.49	42.07
12	I3D-24+FINCH	48.31	41.9
13	I3D-10+FINCH+PE	47.25	43.22
14	I3D-24+FINCH+PE	46.76	41.5