THUMOS 2014
Action Classification Track Summary

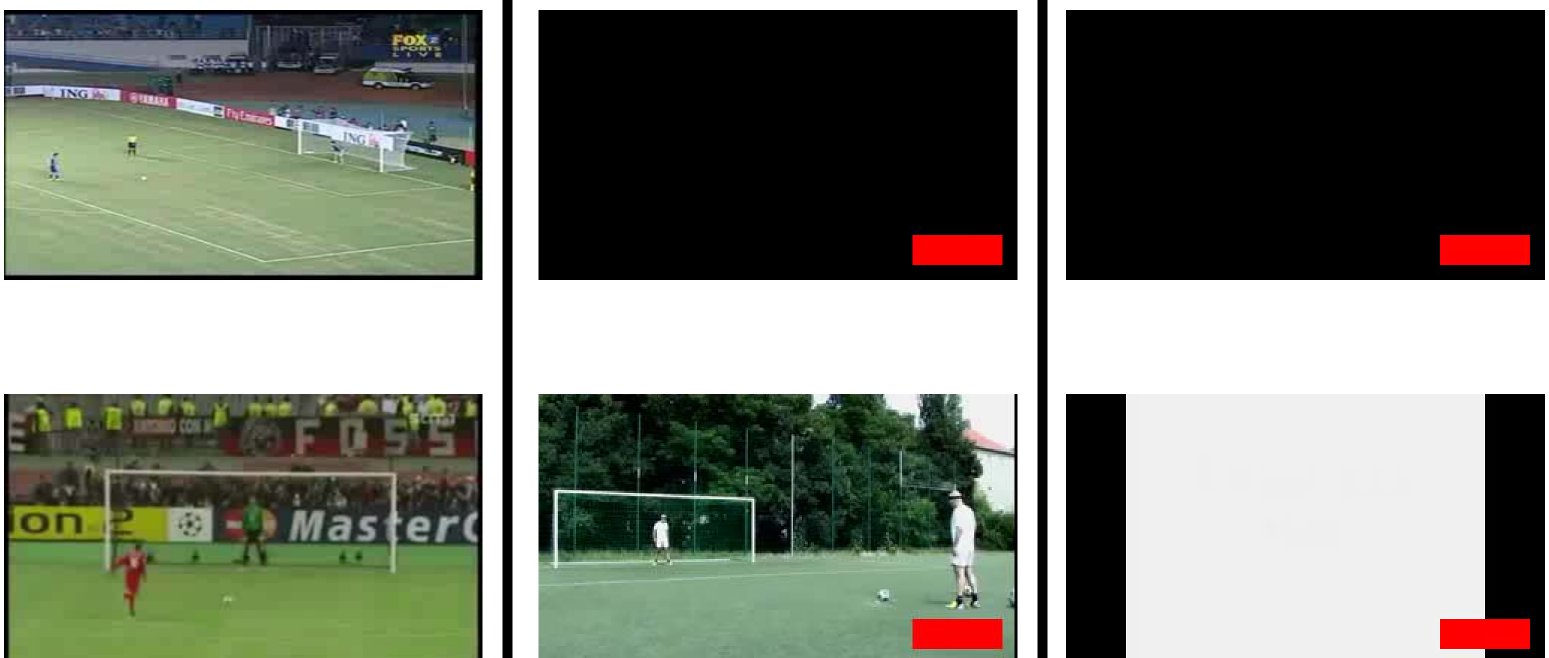
Amir R. Zamir
Yu-Gang Jiang, Jingen Liu, George Toderici, Ivan Laptev, Mubarak Shah, Rahul Sukthankar

THUMOS'14 Challenge

- Two main characteristics:
  - **Temporally untrimmed videos**
    - Crucial for making action recognition useful in practical setting.
  - **Background (null) videos**
    - Performing video level detection is essential.

Soccer Penalty

Train (trimmed) | Validation (untrimmed) | Test (untrimmed)



Cliff Diving

Train (trimmed)

Validation (untrimmed)

Test (untrimmed)

THUMOS'14 Challenge

- Two sub-competition tasks:
  - Temporal action detection
    - Identify where, in time, an action occurs
  - Action classification
    - Identify whether a video contains an action of interest or not
    - Video level
    - Evaluation metric:

$$AP(c) = \frac{\sum_{k=1}^{n}\left(P(k) \times rel(k)\right)}{\sum_{k=1}^{n} rel(k)},$$

Challenge Entries – Classification Track

- 11 Teams
- 35 Runs
- 3 submissions to both video level classification and temporal detection tasks.

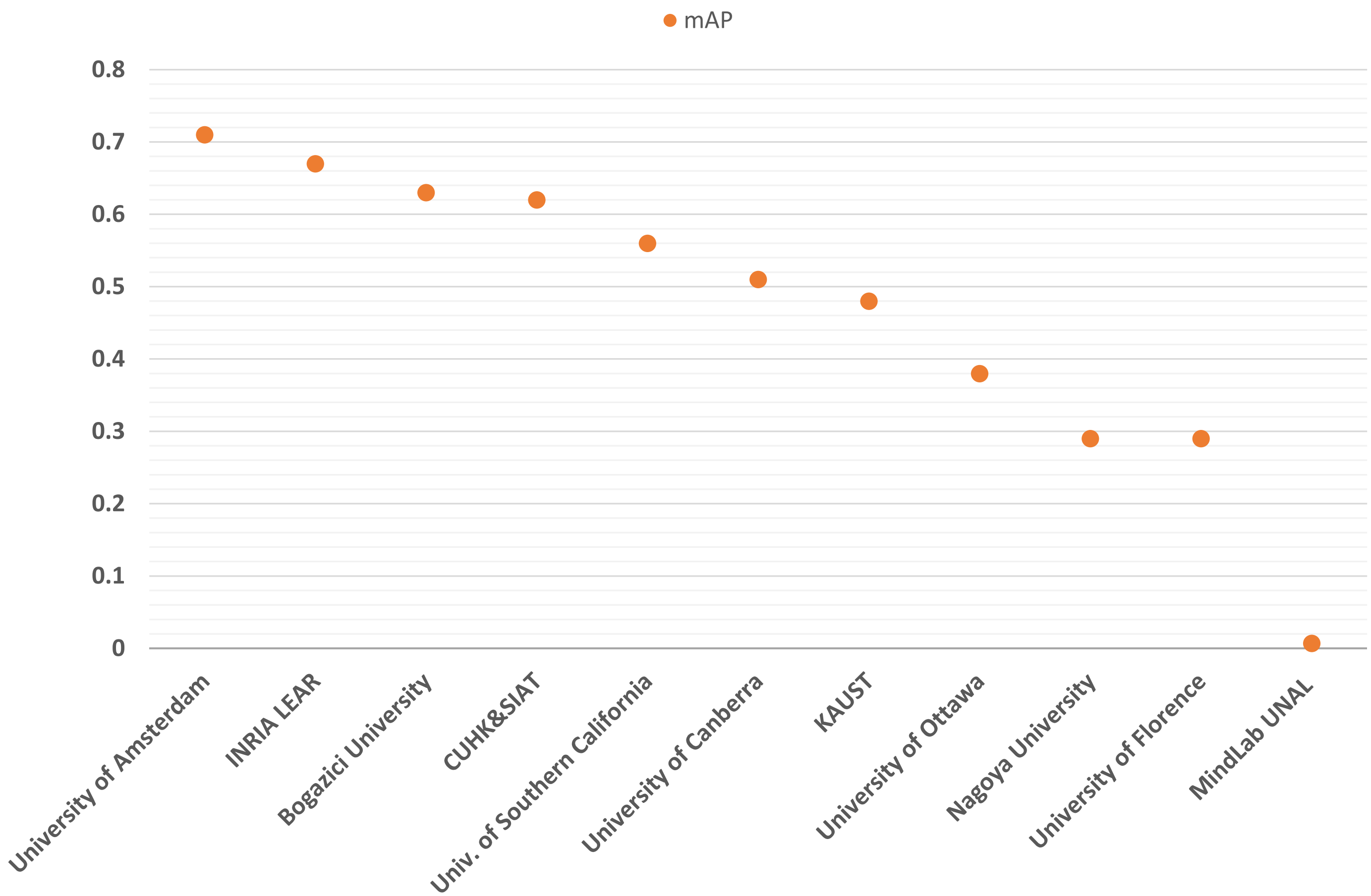
THUMOS'14 Classification Track Results

Rank	Entry	Run1	Run2	Run3	Run4	Run5
1	University of Amsterdam	0.7075	**0.7100**	0.7075	0.7076	0.6932
2	INRIA LEAR	**0.672**	0.6362	-	-	-
3	Bogazici University	0.6191	0.631	0.6171	**0.6316**	0.6173
4	CUHK&SIAT	0.6170	0.6177	0.6196	0.6174	**0.6201**
5	Univ. of Southern California	**0.5675**	-	-	-	-
6	University of Canberra	**0.5161**	-	-	-	-
7	KAUST	0.4146	**0.482**	-	-	-
8	University of Ottawa	0.3343	0.3748	0.1700	0.1401	**0.3856**
9	Nagoya University	0.2485	0.2625	0.2561	0.2941	**0.2955**
10	University of Florence	**0.2919**	0.2809	0.2679	-	-
11	MindLab UNAL	**0.0077**	-	-	-	-

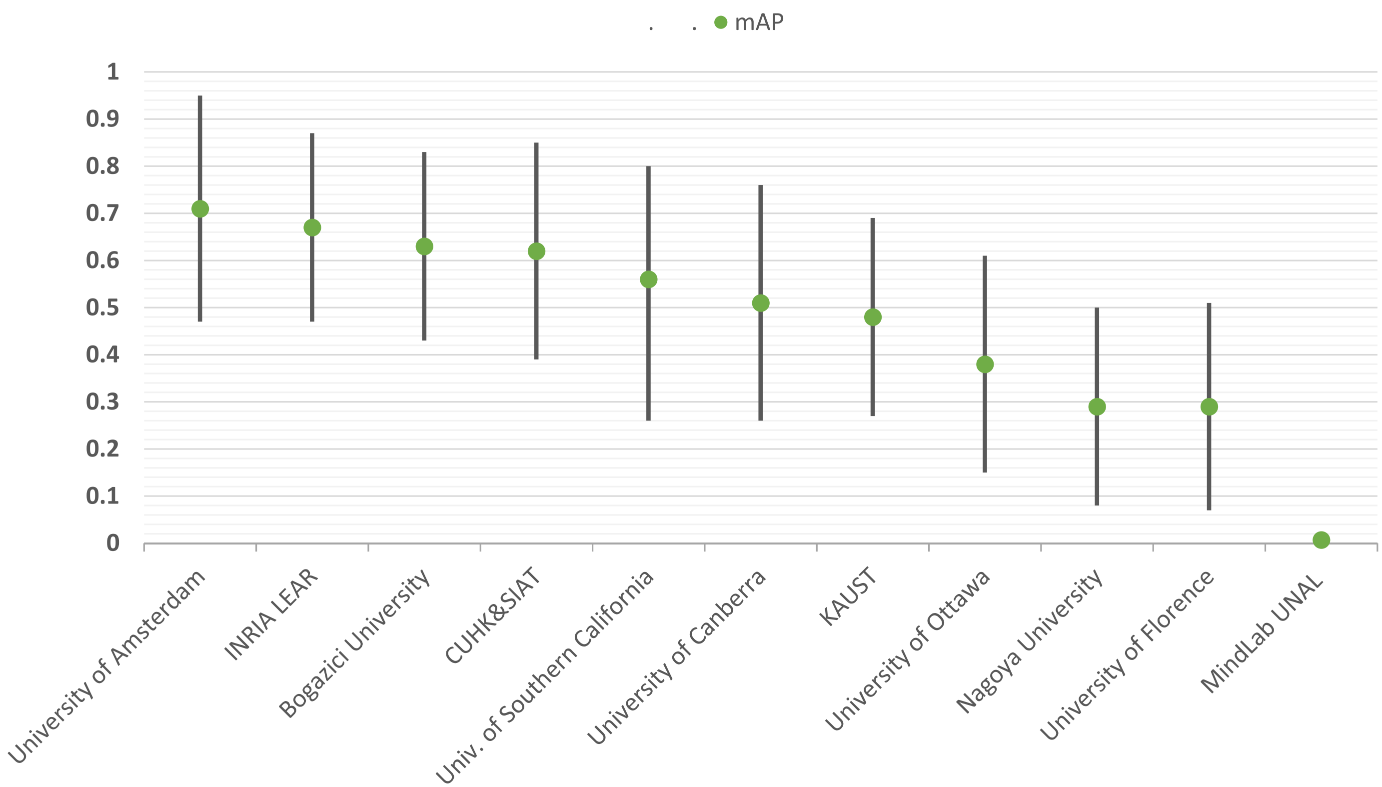
THUMOS'14 Classification Track Results

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

THUMOS'14 Classification Track Results



Class variations



Overall Classification Task Results

- Winning classes

Class Difficulty

- Curve (class vs accuracy)

Background videos

- Do not include an instance of <u>any</u> of the 101 actions.

Background videos

- Do not include <u>any</u> of the 101 actions.
- Are contextually similar to at least one class.

Positive

Background

Background videos

- Do not include any of the 101 actions.
- Are contextually similar to at least one class.

Positive

Background

- How well the entries separate:
  - A positive video from the positive videos of other classes and background videos (complete test set)
  - A positive video from the positive videos of other classes (**multi-class classification**)
  - A positive video from the background videos of the same class (**quantifies the confusion by context**)

Background videos

- How well the entries separate:
  - A positive video from the positive videos of other classes and background videos (complete test set)

Rank	Entry	Full Set
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Background videos

- How well the entries separate:
  - A positive video from the positive videos of other classes (multi-class classification)

Rank	Entry	Full Set	Positive Set
1	University of Amsterdam	0.71	0.75
2	INRIA LEAR	0.67	0.70
3	Bogazici University	0.63	0.67
4	CUHK&SIAT	0.62	0.64
5	Univ. of Southern California	0.56	0.58
6	University of Canberra	0.51	0.56
7	KAUST	0.48	0.52
8	University of Ottawa	0.38	0.41
9	Nagoya University	0.29	0.39
10	University of Florence	0.29	0.34
11	MindLab UNAL	0.007	0.01

Approaches

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Approaches

- IDTF+CNN ➔ SVM
- CNN trained on ImageNet using their own structure.

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Approaches

- IDTF + SIFT + Color Moments + CNN + Audio (MFCC+ASR).
- Feature Variety.

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Approaches

- IDTF+FV+Mid Level ➔ Extreme Learning.
- Mid-level features: face, body, eye.

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Approaches

- Sliding window: IDTF+FV➔SVM.
- CNN: Caffe fine tuned on UCF101.

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Approaches

- Sliding Window: IDTF+FV➔SVM
- Hard Negative Mining (background)

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Approaches

- IDTF+FV➔SVM.
- Background/foreground separation.
  - Separate models for each
  - Late fusion

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Approaches

- Shot boundary detection + Recover FV representation from IDTF➔SVM

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Approaches

- IDTF+FV+DeCAF CNN→SVM
- Saliency features

Rank	Entry	mAP
1	University of Amsterdam	0.71
2	INRIA LEAR	0.67
3	Bogazici University	0.63
4	CUHK&SIAT	0.62
5	Univ. of Southern California	0.56
6	University of Canberra	0.51
7	KAUST	0.48
8	University of Ottawa	0.38
9	Nagoya University	0.29
10	University of Florence	0.29
11	MindLab UNAL	0.007

Trends!

- **Dense Trajectory Features:**
  - 10 teams used DTF features (9 improved-DTF and 1 basic DTF)
- **Convolutional Neural Networks (CNN) used by 4 teams.**
- **Beyond low-level features:**
  - Mid-level body and facial features (Bogazici University)
  - Audio (INRIA LEAR)
  - Saliency features (University of Florence)
  - Shot boundary detection (Nagoya University)
- **Classification:**
  - 10 teams used SVM for classification.
  - 1 Extreme learning.
  - No fully connected CNN.
- **9 teams used Fisher Vector.**

THUMOS 2014

Action Classification Track Summary

Amir R. Zamir

Yu-Gang Jiang, Jingen Liu, George Toderici, Ivan Laptev, Mubarak Shah, Rahul Sukthankar