

ActivityNet Challenge: Kinetics Task 2019

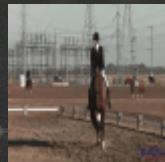
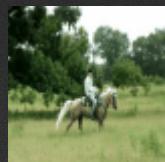
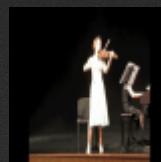
João Carreira



DeepMind

CVPR, Long Beach
17th of June, 2019

The Kinetics dataset



archery

country line dancing

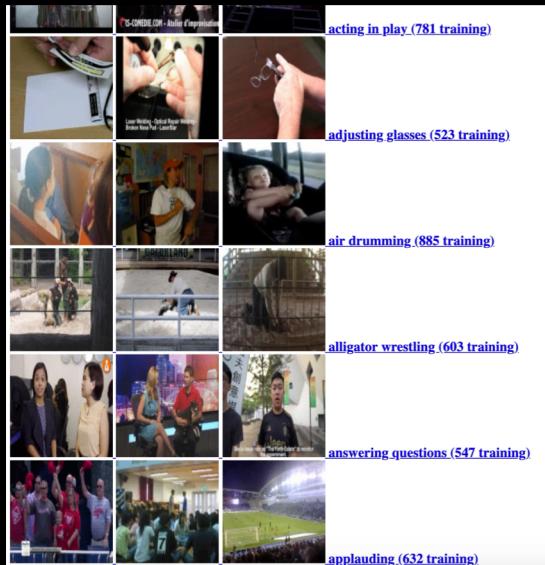
riding or walking with horse

playing violin

eating watermelon

Enabling transfer learning for native video models

Kinetics (~650k videos)
- classification



ActivityNet, Charades, AVA (~10k videos)
- (spatio)temporal localization



Kinetics-700

- 700 human action classes > 600 videos (650k total, all from unique videos)
- Human action classification (10s clips) – each clip assigned one action
- Videos from youtube searches



Joint work with Eric Noland, Chloe Hillier and Andrew Zisserman

Kinetics-700 vs previous releases

Dataset	Year	Actions	Clips	Total	Videos
HMDB-51 [15]	2011	51	min 102	6,766	3,312
UCF-101 [20]	2012	101	min 101	13,320	2,500
ActivityNet-200 [3]	2015	200	avg 141	28,108	19,994
Kinetics	2017	400	min 400	306,245	306,245

Kinetics-600	2018	600	min 600	500,000	500,000
Kinetics-700	2019	700	min 600	650,000	650,000

New in Kinetics-700: more cooking



Slicing onions



Squeezing orange



More dancing



Shoot dance



Silent disco



More face actions



Filling eyebrows



Rolling eyes



More body-only classes



Moving baby



Being in zero gravity



More interactions with objects



Looking in mirror

Walking on stilts



Dataset Collection

- 0 abseiling
- 1 laughing
- 2 swimming
- 3 shearing sheep
- 4 motorcycling
- 5 celebrating
- 6 spray painting
- 7 playing tennis
- 8 driving tractor
- 9 washing dishes
- 10 skateboarding
- 11 waxing legs

Youtube
querying

*“Drumming”
“Playing drums”
“Tocar bateria”*



Image
Classifiers



Human verification using
internal Google annotators

Evaluating Actions in Videos

Instructions

We would like to find videos that contain real humans performing actions e.g. scrubbing their face, jumping, kissing someone etc.

Please click on the most appropriate button after watching each video:

- Yes, this is a true example of the action
- No, this is not an example of the action
- You are unsure if this is an example of the action
- Replay the video
- Video does not play, does not contain a human, is an image, cartoon or a computer game.



Combine, split,
and filter
classes

Last filtering stage: manually shaping a class



Manually shaping a class

Confusions: scuba diving, feeding fish, swimming with sharks, snorkeling

Start Prev Next End Mark Page Good Mark Page Bad x Close

880 | youtube | X20 881 | youtube | X20 882 | youtube | X20 883 | youtube | X20 884 | youtube | X20 885 | youtube | X20 886 | youtube | X20 887 | youtube | X20

888 | youtube | X20 889 | youtube | X20 890 | youtube | X20 891 | youtube | X20 892 | youtube | X20 893 | youtube | X20 894 | youtube | X20 895 | youtube | X20

896 | youtube | X20 897 | youtube | X20 898 | youtube | X20 899 | youtube | X20 900 | youtube | X20 901 | youtube | X20 902 | youtube | X20 903 | youtube | X20

904 | youtube | X20 905 | youtube | X20 906 | youtube | X20 907 | youtube | X20 908 | youtube | X20 909 | youtube | X20 910 | youtube | X20 911 | youtube | X20

912 | youtube | X20 913 | youtube | X20 914 | youtube | X20 915 | youtube | X20 916 | youtube | X20 917 | youtube | X20 918 | youtube | X20 919 | youtube | X20

Start Prev Next End Mark Page Good Mark Page Bad x Close

Class: helmet diving

This year we crowdsourced this step



Ski Ballet



Evolution of data acquisition process

- Kinetics-400: search videos with text query for class name, verification with Mechanical Turk

Evolution of data acquisition process

- Kinetics-400: search videos with text query for class name, verification with Mechanical Turk
- Kinetics-600: decoupled class and query text, added concept of language
 - E.g. "folding paper" now matches against "folding paper" (en), "origami" (en), "dobrar papel" (pt)

Evolution of data acquisition process

- Kinetics-400: search videos with text query for class name, verification with Mechanical Turk
- Kinetics-600: decoupled class and query text, added concept of language
 - E.g. "folding paper" now matches against "folding paper" (en), "origami" (en), "dobrar papel" (pt)
 - Kinetics-700:
 - Added Spanish and French
 - Moved from Mechanical Turk to dedicated internal Google team
 - Crowdsourced last filtering stage

Queries in multiple languages

#	Class	In Kinetics 400	In Kinetics 600	Raw Clip Count	Filtered Clip Count	Assignee
1	acting in play	No	Yes	594	594	joaoluis
	+ Query Text	Language		Remove		
	<input checked="" type="checkbox"/> Acting in play	en ▲	Submit	-		
	<input checked="" type="checkbox"/> faire du théâtre	fr ▲	Submit	-		
	<input checked="" type="checkbox"/> fazendo teatro	pt ▲	Submit	-		
	<input checked="" type="checkbox"/> haciendo teatro	es ▲	Submit	-		
	<input checked="" type="checkbox"/> representando teatro	es ▲	Submit	-		
2	altering jeans	No	No	178	0	joaoluis
	+ Query Text	Language		Remove		
	<input checked="" type="checkbox"/> alterando jeans	es ▲	Submit	-		
	<input checked="" type="checkbox"/> alterar jeans	pt ▲	Submit	-		
	<input checked="" type="checkbox"/> altering jeans	en ▲	Submit	-		
	<input checked="" type="checkbox"/> customiser un jean	fr ▲	Submit	-		
	<input checked="" type="checkbox"/> customizar jeans	pt ▲	Submit	-		
	<input checked="" type="checkbox"/> rasgar calça jeans	pt ▲	Submit	-		

Evolution of baseline I3D performance

Version	Top-1 Acc.
Kinetics-400	68.4
Kinetics-600	71.7
Kinetics-700	57.3

Evolution of baseline I3D performance

Version	Top-1 Acc.
Kinetics-400	68.4
Kinetics-600	71.7
Kinetics-700	57.3

1. Large drop in Kinetics-700 or unnatural improvement in Kinetics-600?
2. Kinetics-600 test set leaked, so we created a new test set for Kinetics-700 from scratch → domain shift ?
3. Final clean-up now crowdsourced – different from our previous manual process → classes are less uni-modal ?

Easiest classes

Diving cliff



Climbing ice



Making latte art



Hula hooping



Hardest classes

Winking



Photobombing



Pinching



Closing door



2019th challenge

A very tight schedule:

- Released on May 29th.
- Deadline for submission June 13th
- Presentations: 17th (today)

2019th challenge

A very tight schedule:

- Released on May 29th.
- Deadline for submission June 13th
- Presentations: 17th (today)

An engineering/operational challenge as much as a computer vision one:

- Downloading 150k new videos
- Finetuning models / testing them / submitting results

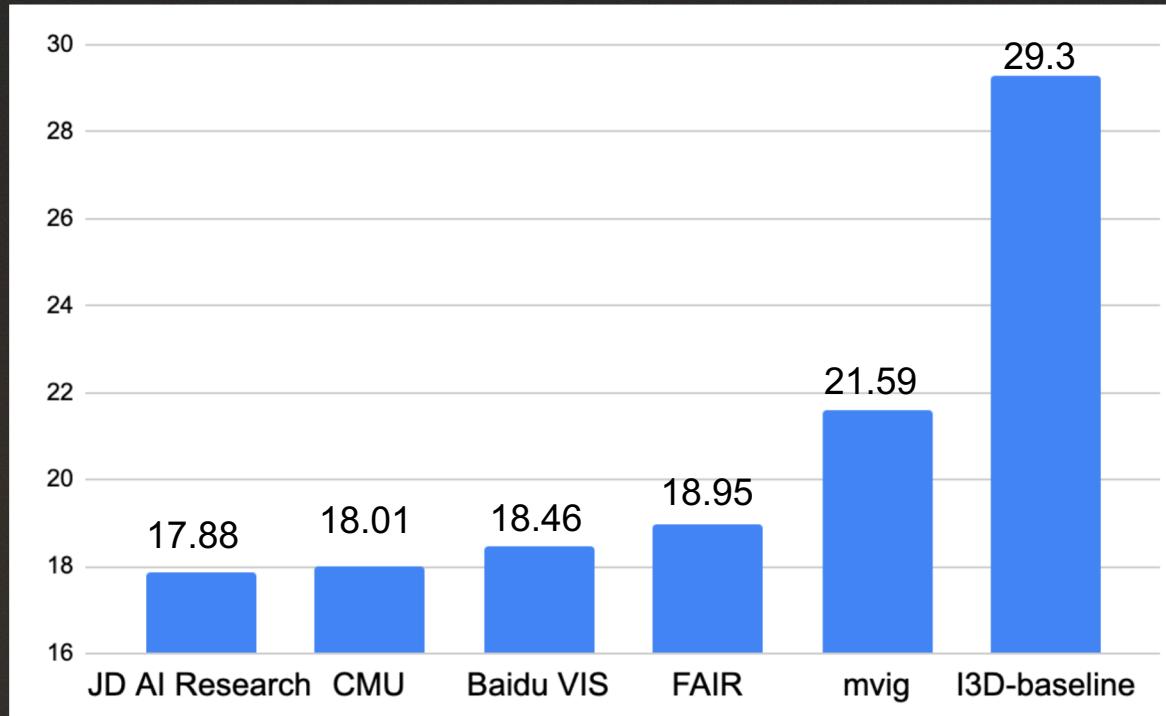
2019th challenge

- 15 teams still managed to make it!
- Evaluation metric: average of top-1 and top-5 classification error
- Examples in test set of Kinetics-700 shared with Kinetics-600 were ignored
- Baseline I3D-RGB: 29.3% error
- Top-5 methods: [17.9, 21.6]

2019th challenge: techniques

- New interesting models: Local-global diffusion networks, SlowFast networks, Temporal shift modules.
- Some top teams used RGB + Flow + Audio
- Ensembling

2019th challenge: results



2019th Kinetics challenge winner



Zhaofan Qiu, Dong Li, Yehao Li, Qi Cai, Yingwei Pan, and Ting Yao
JD AI Research, Beijing, China

panyw.ustc@gmail.com

