Action Recognition Experiments

By Kevin Yu

Hardware Setup

Ubuntu 14.04

i7-6700HQ, 8 core

16Gb RAM

CPU only

NVIDIA Tesla K80 GPU

Which experiment to pick

Two-stream convolutional networks for action recognition in videos?

Large-scale video classification with convolutional neural networks?

Learning Spatiotemporal Features with 3D Convolutional Networks?

Installation Guides

Main Website:

https://github.com/facebook/C3D

Installation Instructions:

https://gist.github.com/arundasan91/b432cb011d1c45b65222d0fac5f9232c

http://vra.github.io/2016/03/03/c3d-use/

Legitimacy of C3D

Dataset	Sport1M	UCF101	ASLAN	YUPENN	UMD	Object
Task	action recognition	action recognition	action similarity labeling	scene classification	scene classification	object recognition
Method	[29]	[39]([25])	[31]	[9]	[9]	[32]
Result	**90.8**	75.8 (89.1)	68.7	96.2	77.7	12.0
C3D	85.2	**85.2 (90.4)**	**78.3**	**98.1**	**87.7**	**22.3**

| Method	Number of Nets	Clip hit@1	Video hit@1	Video hit@5
DeepVideo's Single-Frame + Multires [18]	3 nets	42.4	60.0	78.5
DeepVideo's Slow Fusion [18]	1 net	41.9	60.9	80.2
Convolution pooling on 120-frame clips [29]	3 net	**70.8***	**72.4**	**90.8**
C3D (trained from scratch)	1 net	44.9	60.0	84.4
C3D (fine-tuned from I380K pre-trained model)	1 net	46.1	61.1	85.2

Example Videos

Action Classification

Tested 150 clips

Action Classification Successful
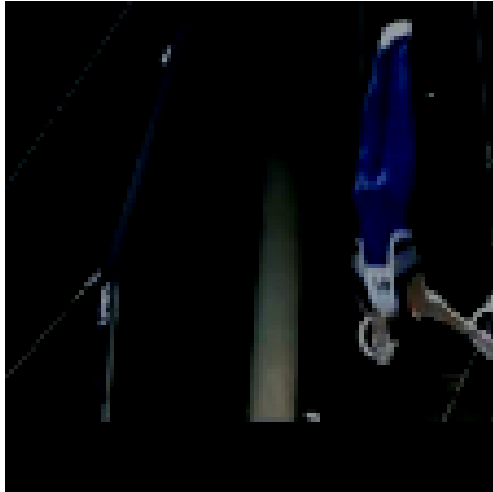
True: Still Rings

Predicted: Still Rings

True: BandMarching

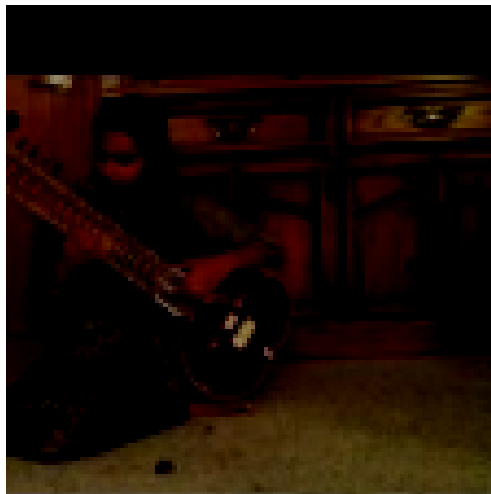Predicted: BandMarching

True: Punch

Predicted: Punch

Action Classification Unsuccessful

True: Kayaking

Predicted: Surfing

True: PlayingSitar

Predicted:
ShavingBeard

True: HammerThrow

Predicted:
PlayingSitar