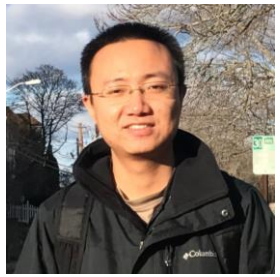
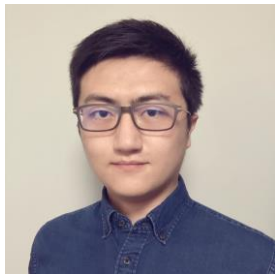
EV-Action: Electromyography-Vision Multi-Modal Action Dataset



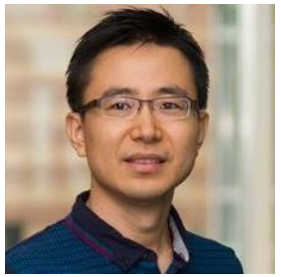
Lichen Wang
wanglichenxj@gmail.com

Bin Sun
sun.bi@northeastern.edu

Joseph Robinson
robinson.jo@northeastern.edu



Taotao Jing
scottjingtt@gmail.com



Yun Fu
yunfu@ece.neu.edu

SMILE Lab
Electrical & Computer Engineering
Northeastern University

Introduction

EV-Action Dataset:

- A large-scale multi-model human action dataset

Motivation:

- Different modalities contain extra action/motion information
- Non-visual modalities could provide more complementary and comprehensive information for deep understanding of human actions
- Electromyography (EMG) signal [1] is non-visual. It has strong connections with actions, and it has not been well-explored for human action analysis

[1] https://en.wikipedia.org/wiki/Electromyography

What is EMG?

What is EMG?

Electrical activity signal in human muscles

Modalities:

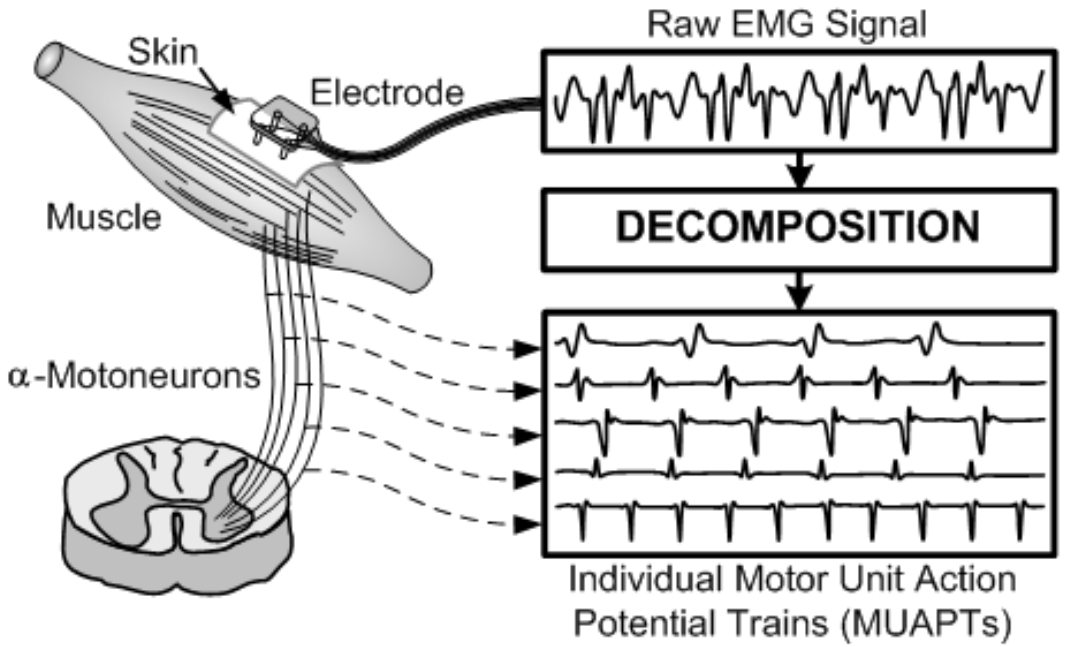
Visual:

- RGB
- Depth

Skeleton:

- Depth-based
- Tracking-based

EMG signal



Concept of EMG signal [1,2]

[1] http://www.bu.edu/ids/research-projects/muscles-alive/
[2] S.H. Nawab, S. S. Chang, C. J. De Luca, High-Yield Decomposition of Surface EMG Signals, Clinical Neurophysiology, 121: 1602-1615, Oct. 2010

Devices

Data collection devices:

- 5 modalities (RGB, depth, 2 skeleton, EMG)
- Three kinds of devices for data collection
- Specific designed data collection center

Modalities	Sensors	Resolution	Frame rate
RGB	Kinect-V2[1]	1920x1080	30
Depth		512x424	30
Skeleton-1		-	30
Skeleton-2	Vicon-T40s[2]	2336x1728	100
EMG	Delsys-Trigno[3]	-	1000

Configurations of data collection devices

Vicon Camera for motion/skeleton capturing [2]

Kinect Sensors [1]

Wireless EMG sensor [3]

Configurations of action collection devices

[1] https://developer.microsoft.com/en-us/windows/kinect/
[2] https://www.vicon.com/
[3] https://delsys.com/trigno/

Sensor placement scheme

Vicon optical marker:

- Standard scheme [1]
  - Full-body model (plug-in gait)

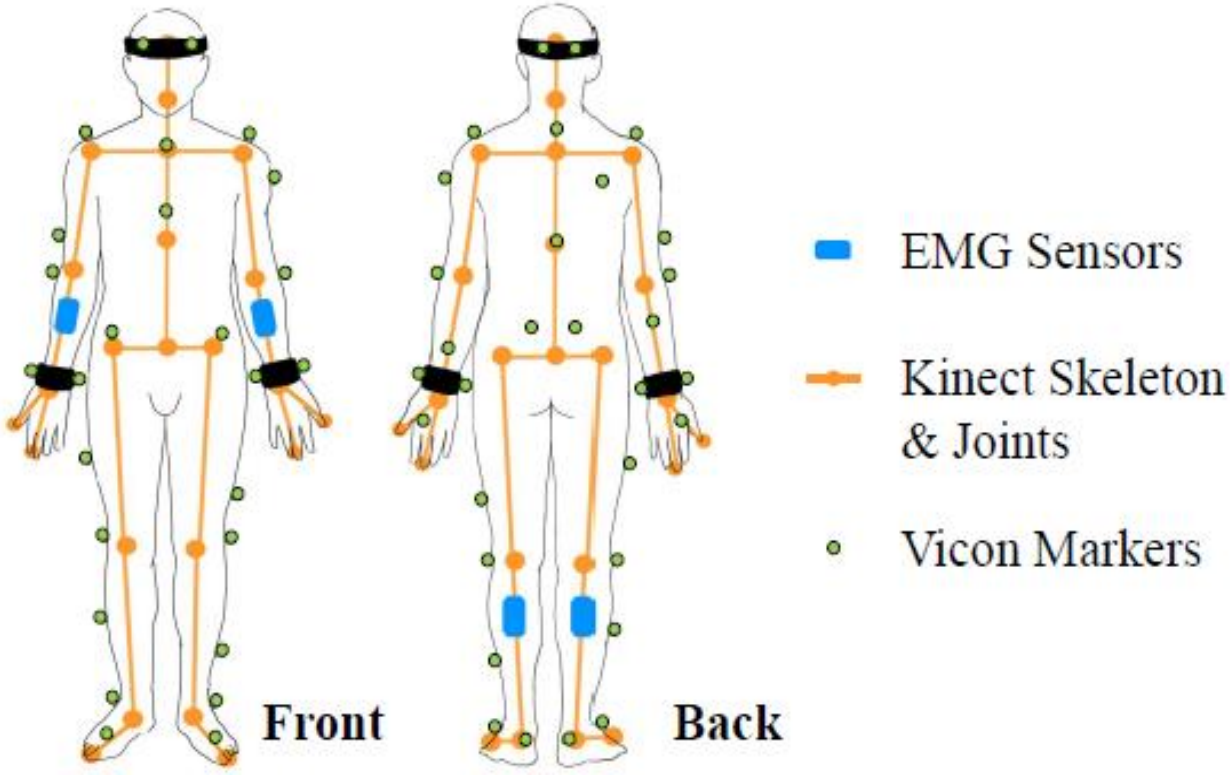
EMG sensor position:

High amplitude & low noise [2]

- Middle of left & right forearms
- Middle of left & right shank

Kinect:

- Kinect 26 joints skeleton



EMG and Vicon marker placement scheme

[1] ViconSystem: The standard Vicon full-body model (plug-in gait) marker placement scheme, 2010.

Data collection center

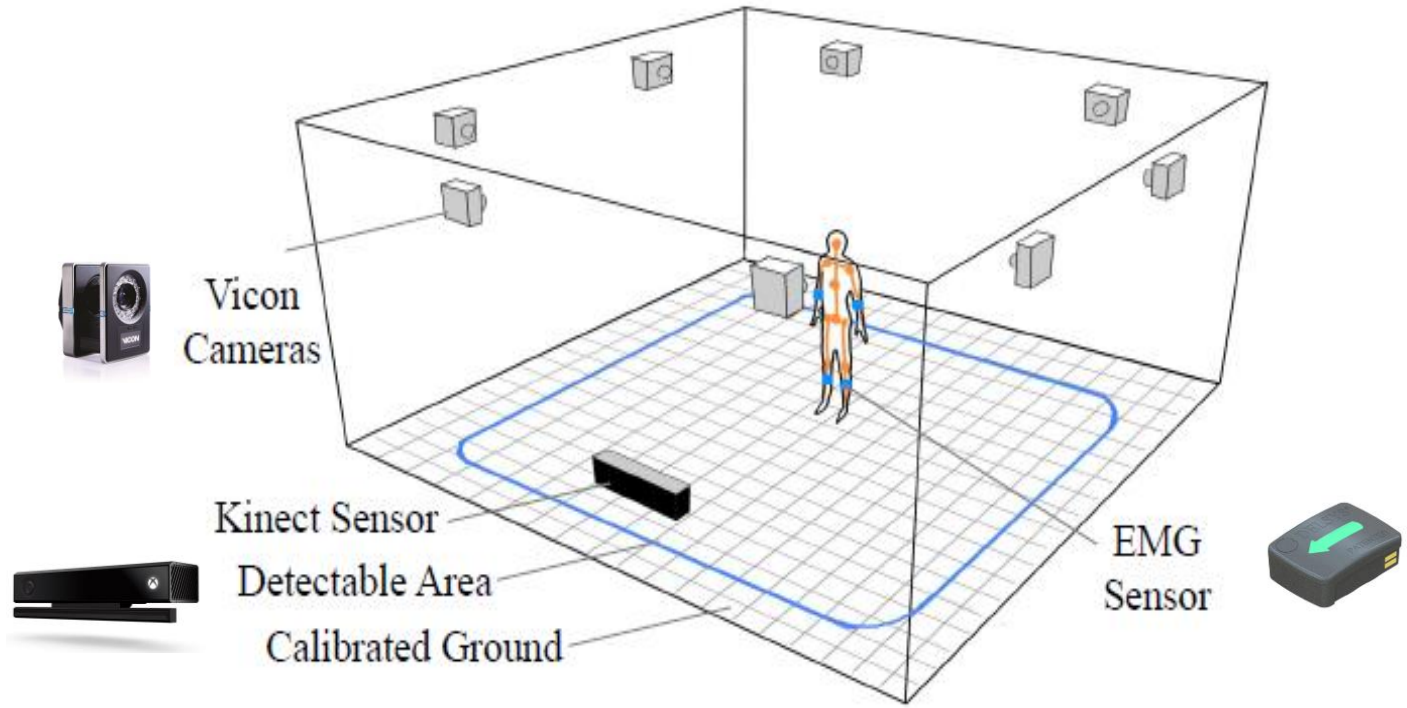
Space:

- Standard lab environment

Devices:

- 8 Vicon high speed cameras
- 4 EMG sensors
- 1 Kinect V2 sensors

RGB frame sample collected from Kinect



Setup of action collection center

Action list

Action selection:

Categories:

- Daily and common human actions for general analysis purpose

Two types:

- Single person: clear, no occlusion
- Person object: with common objects, have occlusions

10 for each type, 20 actions in total.

Single Person Actions	Person-Object Actions
Walking	Answering phone
Boxing	Checking watch
Waving hands	Standing up
Clapping hands	Sitting down
Jumping	Grabbing bag
Bending over	Throwing ball
Turning around	Drinking water
Kicking	Tying shoes
Raising hand	Reading book
Falling	Moving table

Selected action lists

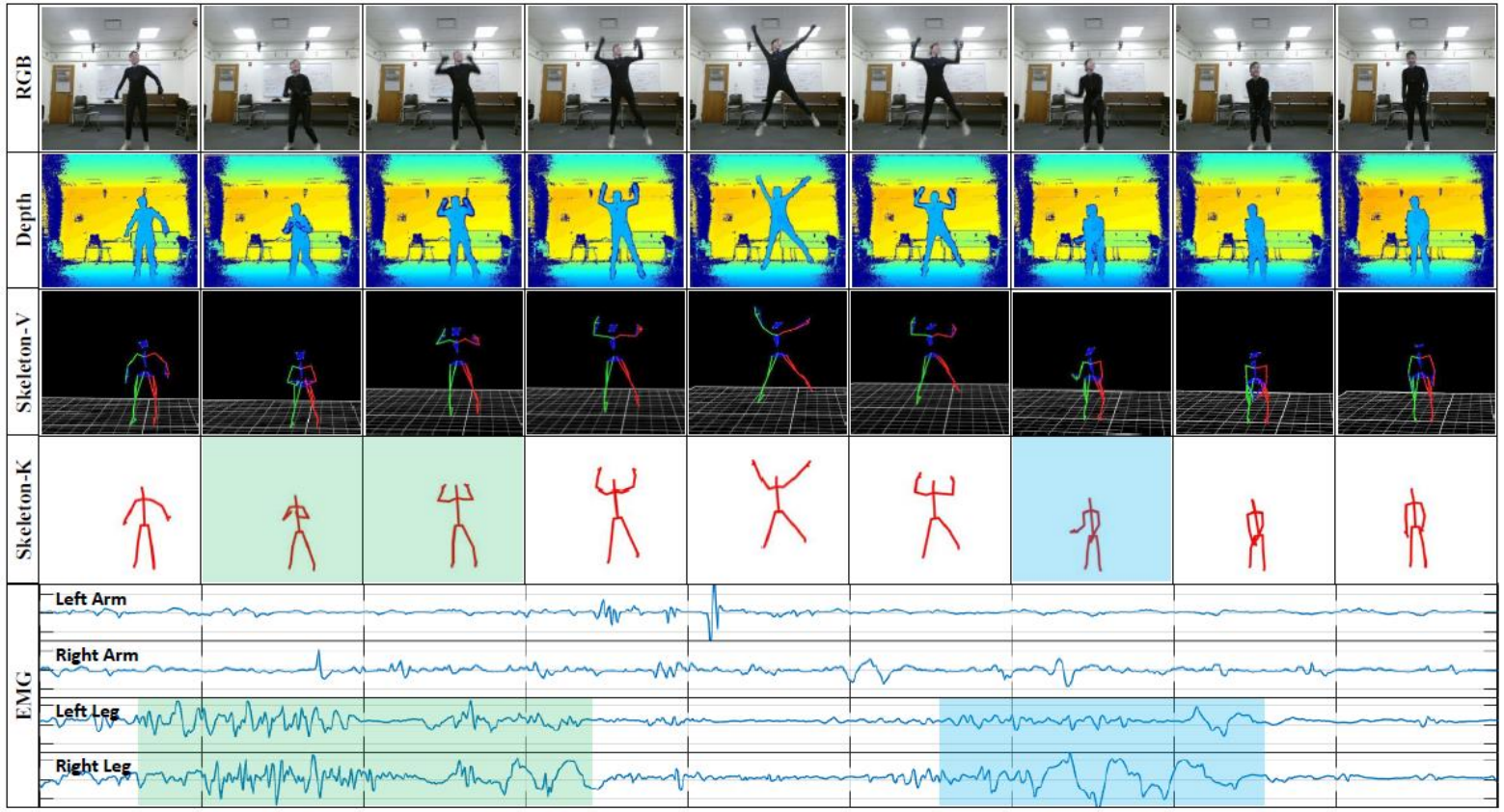
Collected dataset

Summary:

- Synchronized multi-view time-serious data
- 7000 samples from 70 subjects

Type	Number
Subjects	70
Actions	20
Repeat	5
Total samples	7000

Dataset action numbers



Frames of the EV-Action datasets with 5 modals

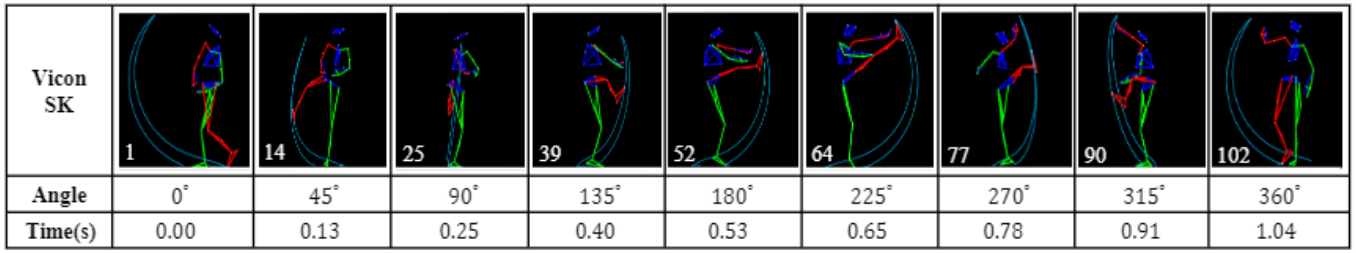
Dataset analysis

Quantitative analysis:

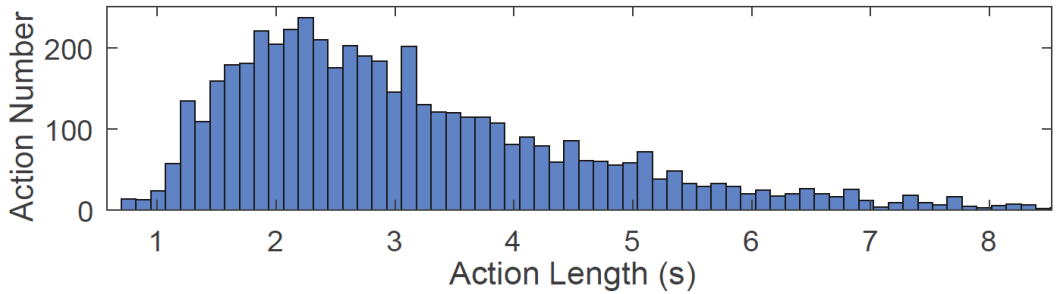
- High speed & precisions skeleton tracking performance by Vicon system

Statistical analysis:

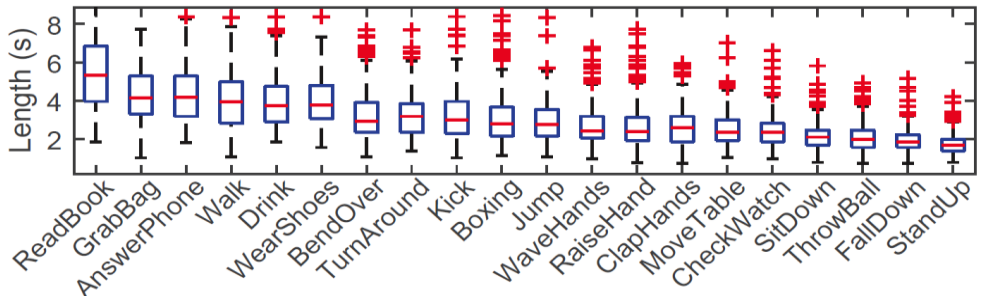
- Action length distribution
- Length variance for each action



Vicon skeleton tracking results



Action length analysis

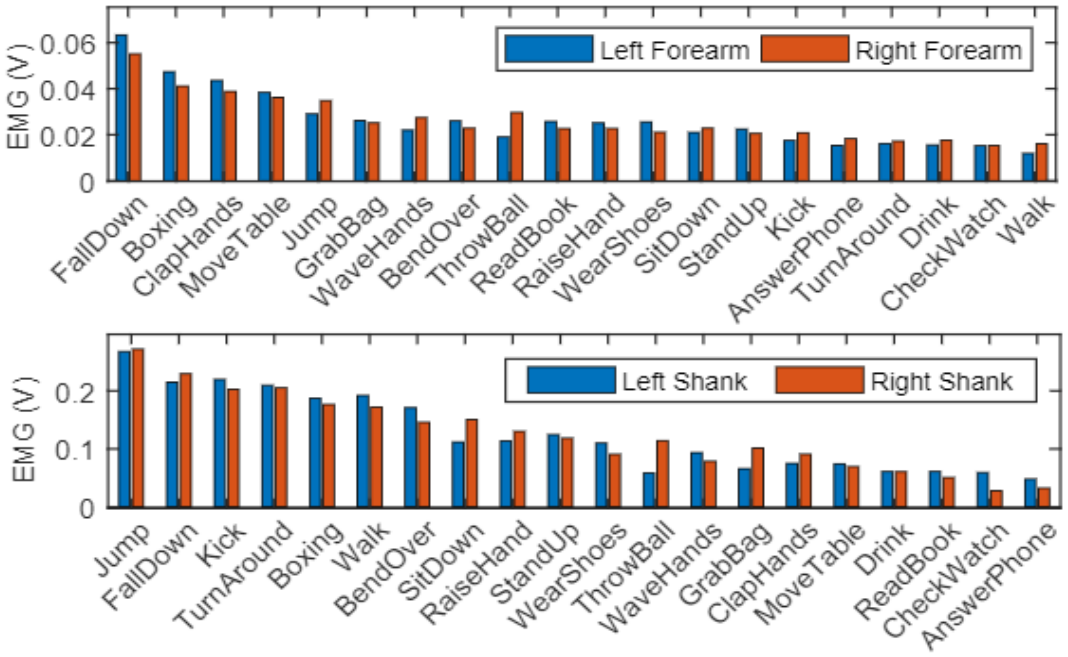


Length variance for different actions

EMG signal analysis

Unique discoveries from EMG signal

- Left/right amplitude differences across different actions
- Prior active of EMG compared with visual signals
- Lower body activity even in pure upper body actions



EMG signal analysis across different actions and different locations

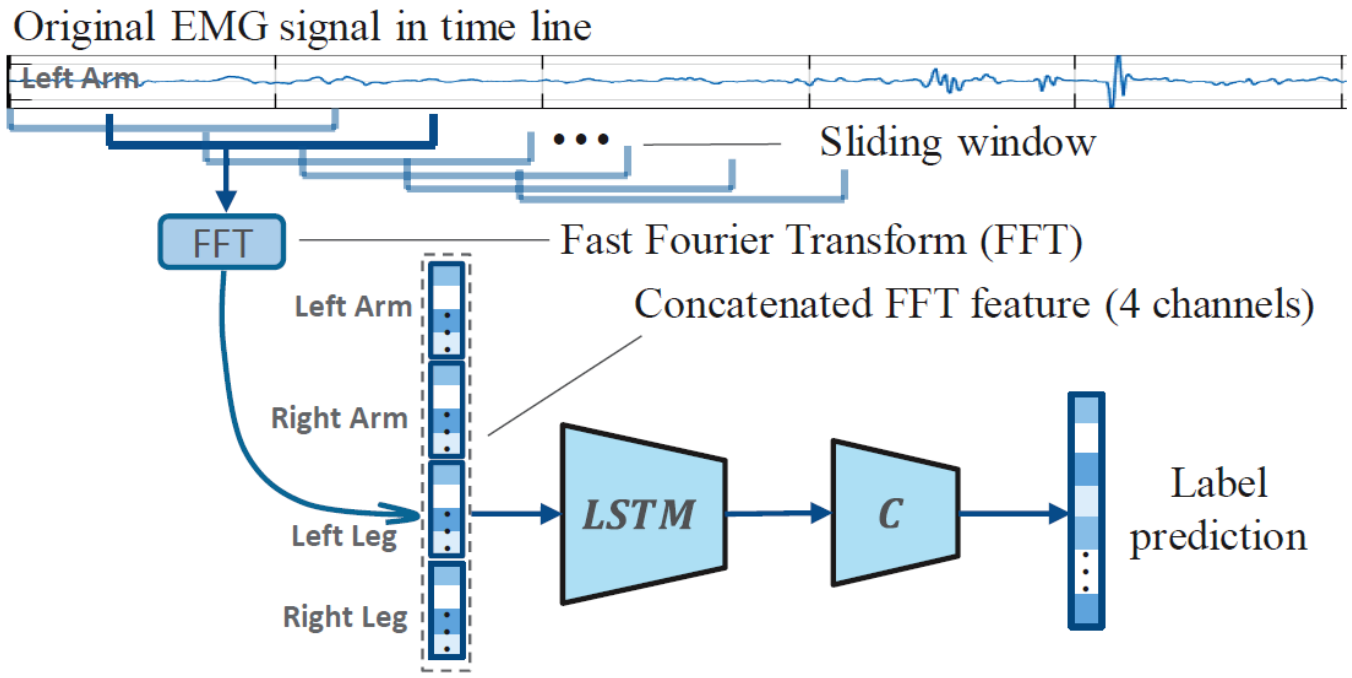
EMG-based classification

An effective framework

- Sliding window in timeline
- Fast Fourier Transform (FFT)
- LSTM feature extractor + neural network classifier

Methods	Dimension Reduction (None)	LDA	PCA
Random Forest	33.72	16.81	35.12
KNN	22.16	13.55	26.18
SVM	23.74	16.12	25.65
FFT-LSTM (Ours)	**44.13**	-	-

EMG-based classification performance



We proposed a basic framework for EMG-based classification

Multi-view classification baselines

Single-view models:

- RGB
- Depth
- Skeleton
- EMG

Multi-view models:

- Skeleton + EMG

		Single-Person										Person-Object										
		Walk	*Box*	*Wave Hand*	*Clap Hands*	*Jump*	*Bend*	*Turn Around*	*Kick*	*Raise Hand*	*Fall Down*	*Ans. Phone*	*Check Watch*	*Stand Up*	*Sit Down*	*Grab Bag*	*Throw Ball*	*Drink Water*	*Tie Shoes*	*Read Book*	*Move Table*	ACC
RGB	TSN [52]	**56.1**	**94.1**	**25.3**	**83.9**	**88.5**	**94.3**	**68.3**	**95.6**	95.1	**86.2**	**69.5**	**37.6**	**87.0**	**54.3**	**86.9**	**75.7**	**56.8**	**84.8**	**96.7**	**59.1**	**74.7**
	LRCN [12]	44.2	84.0	19.8	69.4	71.6	78.0	57.9	82.1	90.0	71.3	55.6	28.5	72.1	43.4	72.0	62.5	46.8	70.2	85.4	44.2	62.4
	VLAD [16]	47.5	91.8	21.6	75.9	78.3	85.3	63.3	89.7	**98.4**	77.9	60.7	31.1	78.8	47.5	78.7	68.3	50.8	76.7	93.4	48.3	68.2
Dep	WDMM [2]	44.3	76.3	11.4	31.4	36.5	**43.7**	**17.2**	47.4	**72.7**	**36.2**	**27.9**	**12.3**	45.1	**16.8**	**27.2**	**48.2**	**23.4**	**28.4**	42.1	13.5	35.1
	WHDMM [53]	**78.5**	**84.5**	**62.7**	**64.7**	**66.1**	12.3	17.2	**72.3**	67.9	20.1	12.5	11.7	**61.1**	10.1	16.7	22.5	17.0	11.2	**71.5**	**23.5**	**40.2**
SK-K	TCN[39]	**91.2**	82.0	**71.4**	**86.0**	**92.2**	**91.7**	**87.6**	**93.0**	**89.2**	**92.6**	**57.5**	**76.0**	**92.9**	**87.8**	**66.8**	**70.5**	**95.0**	**76.1**	**76.1**	76.4	**82.6**
	TSRNN [44]	90.0	**85.0**	70.6	81.0	91.0	90.5	86.6	91.8	86.6	91.4	56.7	75.1	91.7	86.8	66.0	69.7	93.8	75.1	65.1	**85.4**	81.5
	STGCN [56]	90.6	83.5	71.0	83.5	91.6	91.1	87.1	92.4	88.7	92.0	57.1	75.6	92.3	87.3	66.4	70.1	94.4	75.6	75.6	75.9	82.1
SK-V	TCN [39]	82.1	**77.2**	**67.2**	**87.2**	**83.8**	83.3	80.1	84.4	**81.4**	84.0	36.0	50.9	**64.3**	60.3	43.4	46.4	66.0	50.9	50.9	51.1	64.1
	TSRNN [44]	**83.0**	77.2	67.1	77.4	82.1	**84.4**	**80.5**	**84.9**	79.9	**84.1**	**38.4**	**64.1**	58.3	**64.0**	**46.3**	**49.4**	**70.1**	**54.1**	**64.1**	**64.3**	**67.5**
	STGCN[56]	57.7	53.2	45.2	53.2	58.4	58.0	55.5	58.9	56.5	59.6	36.4	48.2	58.7	55.6	42.3	44.6	60.1	45.2	25.2	54.3	50.7
EMG	LSTM-FFT	72.3	51.6	35.1	54.8	90.6	40.0	30.3	36.6	11.9	72.8	51.2	56.5	16.1	41.6	17.3	48.4	45.7	31.4	46.2	33.0	44.1
SK-K-E	TCN-RMS	91.1	83.0	**73.4**	**88.0**	93.2	**94.7**	**87.8**	91.0	**91.4**	**95.6**	60.5	**79.8**	**91.9**	88.8	**70.8**	**72.5**	94.0	74.1	**78.1**	74.4	83.6
	TCN-FFT	**92.0**	**83.7**	72.1	85.7	**94.0**	93.5	87.3	**94.8**	91.0	94.4	**60.6**	78.5	91.3	**89.6**	70.1	71.9	**94.8**	**79.5**	77.6	**77.9**	**84.0**
SK-V-E	TCN-RMS	**86.7**	**80.7**	**70.3**	**87.9**	**87.1**	**84.5**	**83.6**	**85.1**	**82.1**	83.6	**63.5**	**51.6**	64.4	**60.3**	**45.4**	46.0	**65.8**	50.5	51.2	51.1	**69.1**
	TCN-FFT	82.2	77.5	67.3	87.3	83.8	83.4	80.5	84.7	81.7	**84.5**	37.0	51.4	**64.5**	60.0	43.5	**47.4**	64.0	**53.9**	**52.9**	**51.1**	66.8

Single-modal and Multi-modal baselines

Future work

EMG exploration

- Effective EMG data processing & classification

Latent correlation

- EMG and Skeleton
- EMG and visual (RGB & depth)

Multi-view/modal learning

More areas

- Motion understanding
- Computer vision
- Biomechanics
- Interdisciplinary areas

Smile lab
Synergetic Media Learning Lab

Thank you!

Please contact: **wanglichenxj@gmail.com** for questions.

SMILE Lab
Electrical & Computer Engineering
Northeastern University