HOCHSCHULE HEILBRONN

Deep Learning of Skeleton Features for Posture Recognition

Tianyu Zheng and Peter Ott

Mechatronics, Heilbronn University of Applied Sciences

Overview

Task:
- ✓ Detect different posture of old man or children in the room with Depth Camera
- ✓ Record new dataset, find an appropriate network

Approach:
- Use Nuitrack SDK to do body skeletal tracking, get body joints data;
- Input the joints to a neural network based on bidirectional LSTM (Long Short Term Memory) Network[1];
- Let the network predict the posture according to body joints;
- Design different networks and compare performance between them;
- Export the network and utilize it in a software framework.

Structure

Get video → Get joints → Process with BiLSTM

#Depth cam can detect object distance

#60 frames * 25 joints * 3 coordinates

#Bidirectional Long Short Term Memory

Fig.1 Depth Camera

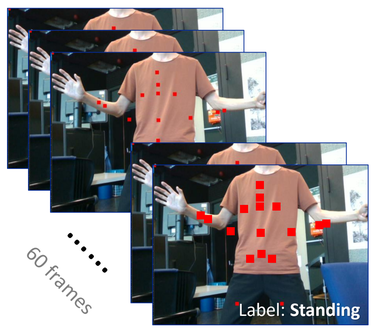


Fig.2 Grab frames

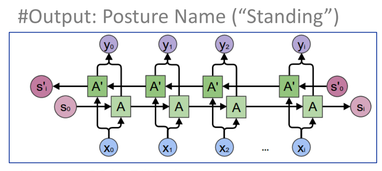


Fig.3 BiLSTM[2]

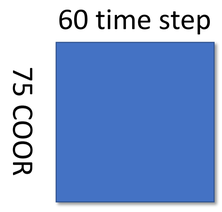
Dataset

Train:Valid = 8:2
- Train dataset:1127 samples
- Valid dataset:282 samples

Each sample corresponds to one posture, like “Standing”, “Walking” or else.

Data in each sample
60 (60 frames) * 25 (25 joints) * 3 (3 coordinates) = 4500 values

Reshape each sample
Reshape each sample to 75*60



Network

Network Structure

Index	Name	BatchNorm	Dropout Rate	Activation	Units Number
1	Dense	√		ReLU	32
2	BiLSTM		0.3		500*2
3	Dense	√	0.1	ReLU	1024
4	Dense	√	0.1	ReLU	512
5	Dense			Softmax	6 (classes)

*Trainable params: 3,178,470

Optimizer
- RMSprop, learning_rate=1e-4

Performance

Final Accuracy: Around **93%**

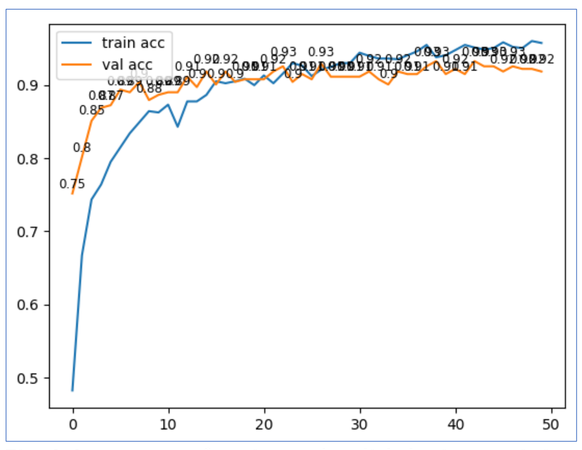


Fig.4 Accuracy of train and valid during training

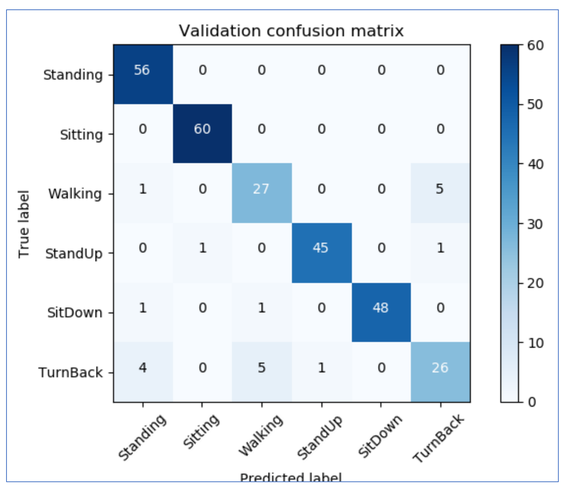


Fig.5 Confusion Matrix of Valid dataset

Conclusion

- Validation Accuracy: 93%
- Inference Speed: 0.047s/sample with GTX 1060
- In Fig5, in the end of the training, there exists slightly over-fitting
- In Fig6, “TurnBack” class sometimes seems very similar to “Standing ” or “Walking”, the dataset should be modified
- Based on current dataset, the network achieves 93% accuracy. It’s not so high. There are still a lot of things to do.

Plans

1. Record new dataset by fixing the camera in a high place
2. Try different networks and find a best one
3. Detect different postures (like falling) of old man or children

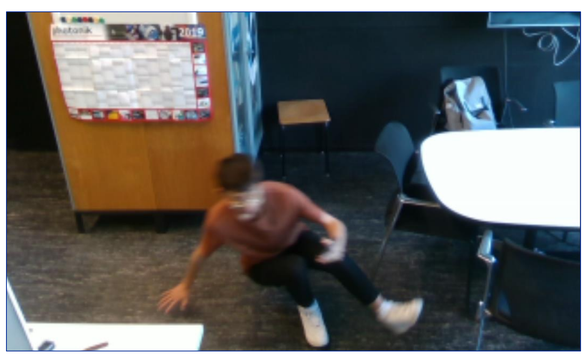


Fig.6 Falling

Reference

[1] Graves, Alex, and Jürgen Schmidhuber. "Framewise phoneme classification with bidirectional LSTM and other neural network architectures." *Neural Networks* 18.5-6 (2005): 602-610.

[2] Colah.github.io. (2019). *Neural Networks, Types, and Functional Programming -- colah's blog*. [online] Available at: http://colah.github.io/posts/2015-09-NN-Types-FP/ [Accessed 25 Jun. 2019].

tzheng@stud.hs-heilbronn.de; peter.ott@hs-heilbronn.de