Deep Action:A novel action recognition using wavelet transformation

Md. Mehadi Hasan
ID: 1209020 Session: 2012-2013

Department of Computer Science & Engineering
Begum Rokeya University, Rangpur

February 27, 2021

Outline

- Introduction
- Related work
- Proposed method
- Dataset
- Experiment & result
- Conclusion
- References

Introduction 1/3

- What is action recognition
- Playing golf

Figure: Golf play

Introduction 2/3

- Wavelet transformation

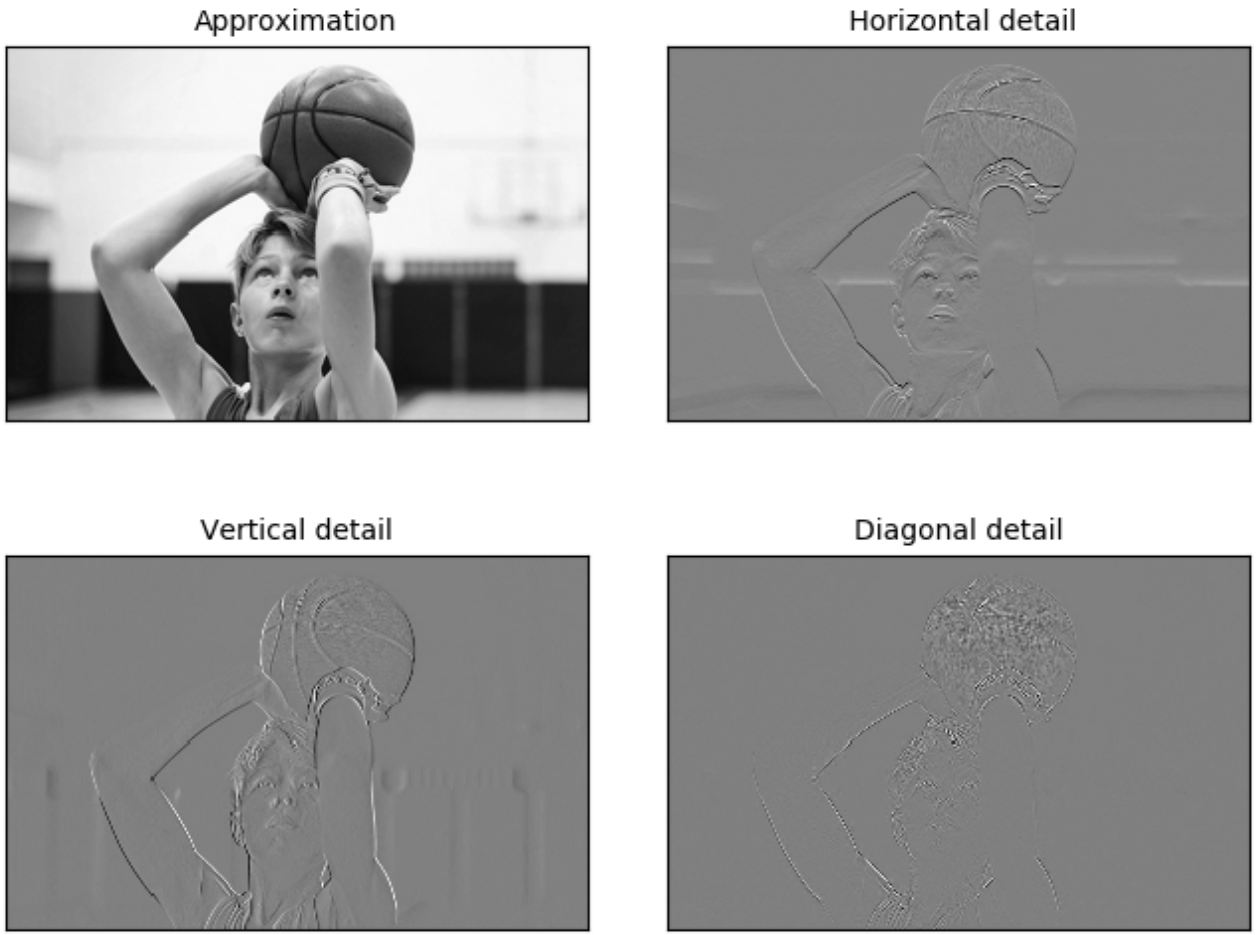

Figure: Level 1 decomposition

Introduction 3/3

- Why named deep action ?

Related work

- Two stream action recognition[2]
- Flow Net 2.0 [1]
- 3D convolutional network [3]
- Wavelet convolutional neural network [5]

Proposed method 1/5

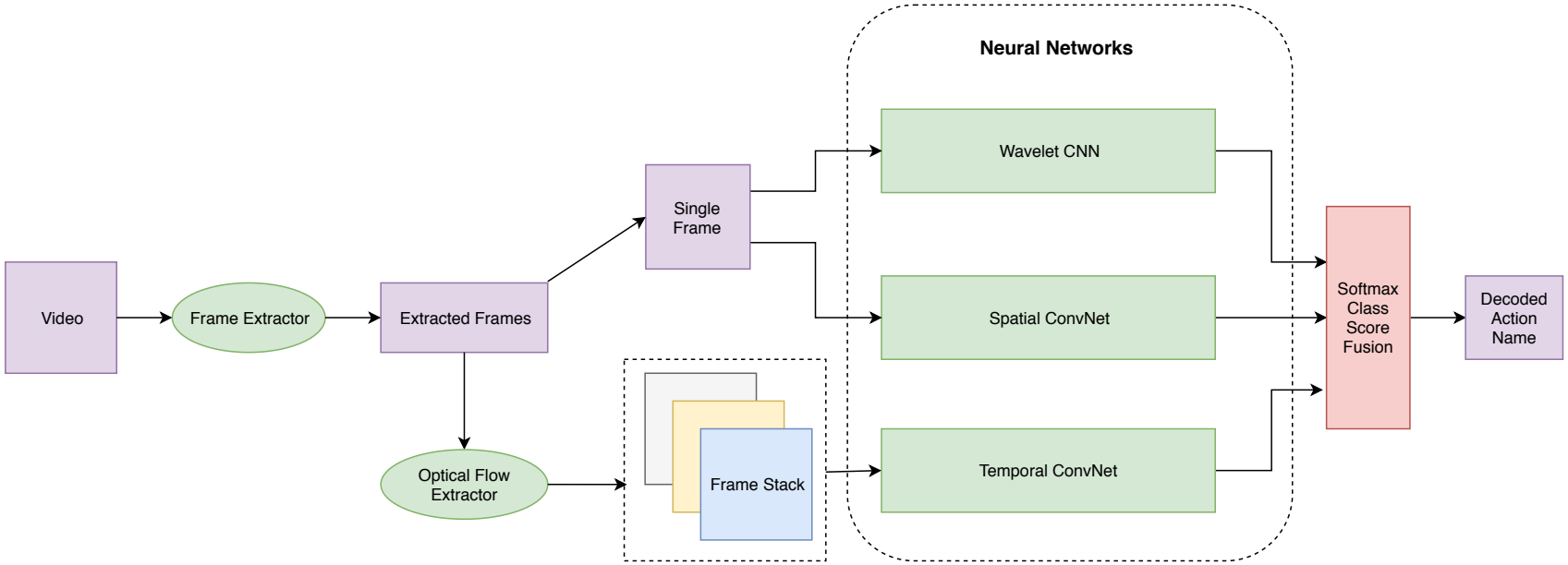

Figure: Proposed architecture

Proposed method 2/5

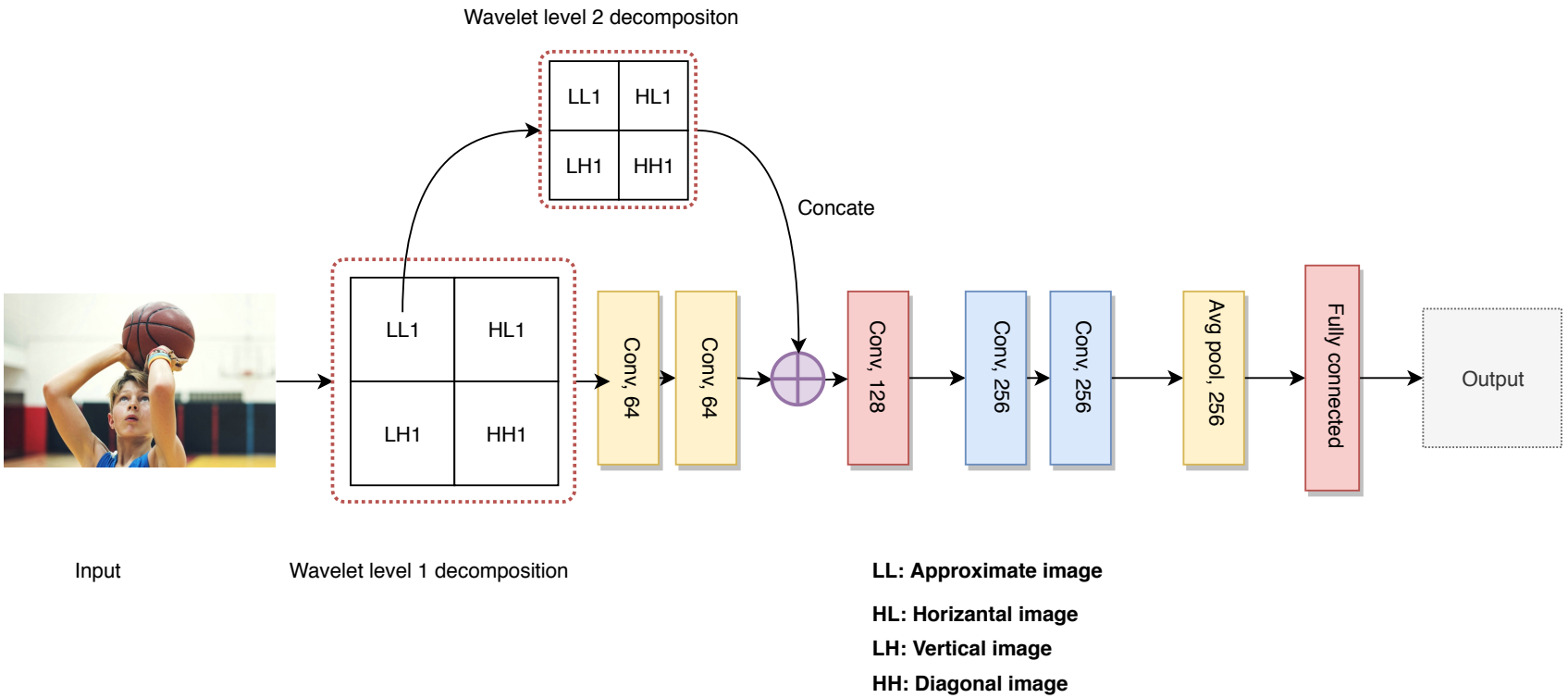

Figure: Wavelet CNN

Structure of convolution neural network

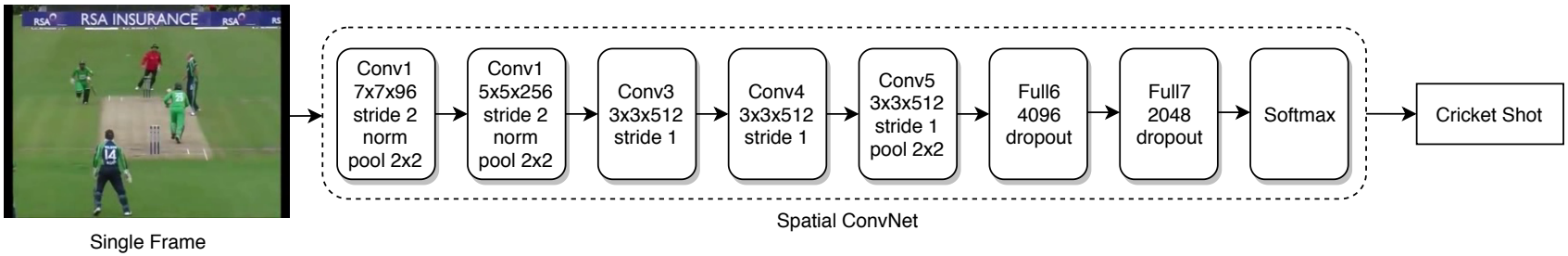

Figure: Spatial convnet

Temporal convolution neural network

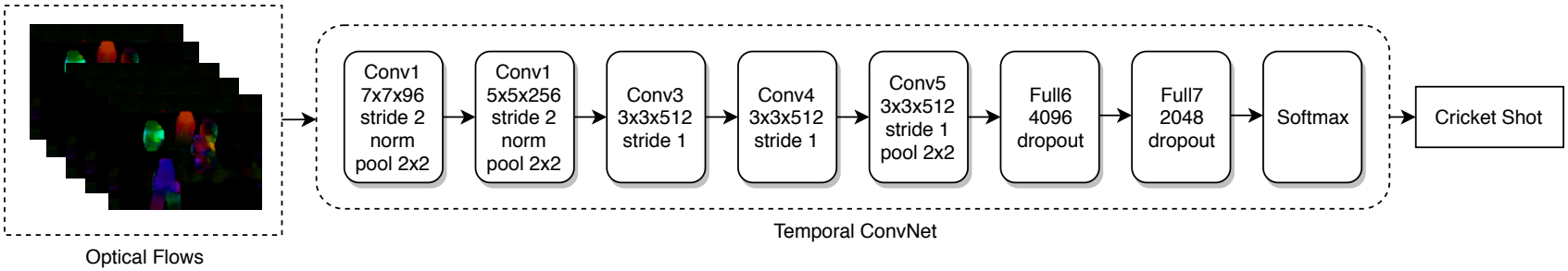

Figure: Spatial convnet

Proposed method 5/5

- What is fusion ?
- Many brain is better than one

Dataset

- UCF-101 [4]
- The action categories can be divided into five types
  - Human-Object Interaction
  - Body-Motion Only
  - Human-Human Interaction
  - Playing Musical Instruments
  - Sports.
- 101 action categories are grouped into 25 groups
- Each group can consist of 4-7 videos of an action
- Same group share some common features, such as similar background

Experiment & result 1/3

- Experimental setup
- Hardware used
  - Nvidia graphics RTX-2080Ti
  - 64GB RAM
- Software & OS
  - Ubuntu 16.04 LTS
  - Python3
  - Tensorflow - deep learning library
  - Keras - high level deep learning library
  - OpenCV - image processing library

Experiment & result 2/3

Result on wavelet convolution neural network

Training setting	Accuracy
From scratch + level of decomposition = 2	62.3%
Pre-train + level of decomposition = 3	74.6%
Pre-train + level of decomposition = 4	77.2%
Pre-train + level of decomposition = 5	**77.6%**

Table: Wavelet ConvNet accuracy on UCF-101

Experiment & result 3/3

Combine result of three stream architecture

Spatial ConvNet	Temporal ConvNet	Wavelet ConvNet	Fusion Metod	Accuracy
Pre-trained + last layer	bi-directional	Decomposition Level = 4	averaging	85.6%
Pre-trained + last layer	uni-directional	Decomposition Level = 4	averaging	85.9%
Pre-trained + last layer	**uni-directional**	**Decomposition Level = 4**	**SVM**	**92.3%**

Table: Three-stream ConvNet accuracy on UCF-101

Conclusion

- Wavelet convolution neural network used for action recognition
- Combine three different features spatial, temporal, spectral to action recognition
- We achieve **92.3%** classification accuracy
- We showed adding spectral feature to two stream architecture improve accuracy

Thank you

References

Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, Thomas Brox

FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks

12, 2016.

Simonyan, Karen and Zisserman, Andrew

Two-Stream Convolutional Networks for Action Recognition in Videos

2014.

S. Ji, W. Xu, M. Yang, and K. Yu.

3D convolutional neural networks for human action recognition.

2013

Khurram Soomro, Amir Roshan Zamir and Mubarak Shah
UCF101: A Dataset of 101 Human Action Classes From Videos in The Wild.
2012

S. Fujieda, K. Takayama, and T. Hachisuka.
Wavelet convolutional neural networks for texture classification

2017