# Recognize Flu-like Symptoms with Deep Learning

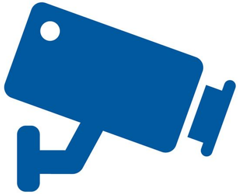
CS 231N Final Project

Alan Lou, Yechao Zhang

June 4, 2021

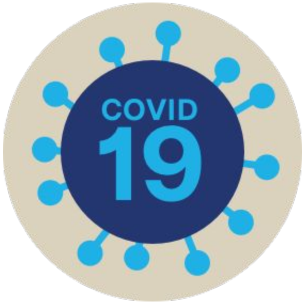
# Contents

# Introduction



- Covid-19 has reached almost every country in the world, infecting millions of people

- Video based surveillance can be used to monitor flu-like symptoms such as coughing and sneezing in densely populated areas
- Apply deep learning techniques to predict flu-like symptoms to help detect Covid-19 early and prevent further escalation

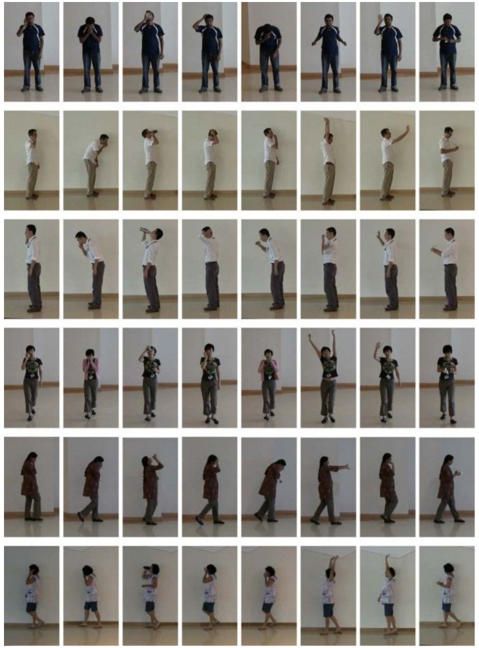



Camera Icon: cam-dex.com
Covid-19 Icon: zurich.com
Cough Icon: ocm.auburn.edu

# Dataset

## BII Sneeze-Cough Human Action Video Dataset (BIISC)

- 20 Subjects:
  - 12 Males, 8 Females
- 8 Action Types:
  - answer phone call, **cough**, drink water, scratch head, **sneeze**, stretch arms, wave hand, wipe glasses
- 3 Poses:
  - face to camera, face to the left, face to the right
- 2 Local motions:
  - stand, walk
- Horizontally flipper version generated for each video
- Total number of videos: 20 x 8 x 3 x 2 x 2 = 1920

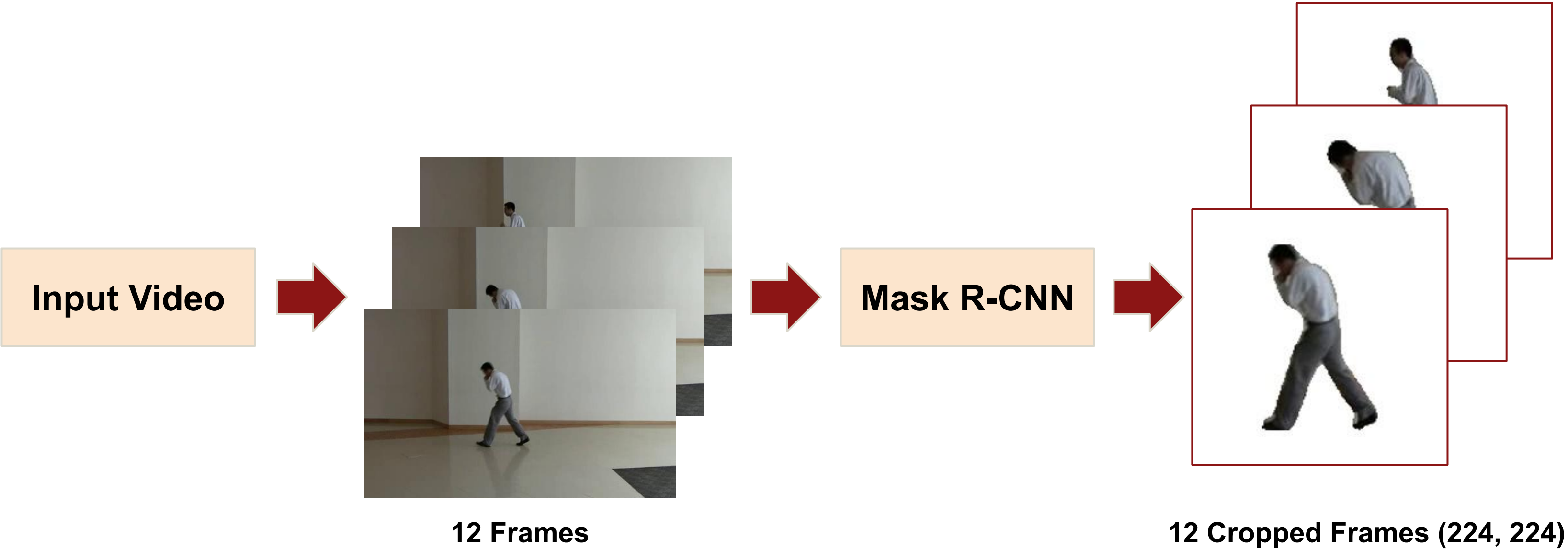


**Snapshots of Sneeze-Cough action recognition videos.** From left to right shows eight actions: answer phone call, cough, drink, scratch face, sneeze, stretch arm, wave hand and wipe glasses.

Thi, T.H., Wang, L., Ye, N. *et al.* Recognizing flu-like symptoms from videos. *BMC Bioinformatics* 15, 300 (2014).

# Data Pre-processing

- CNN-Based Models:

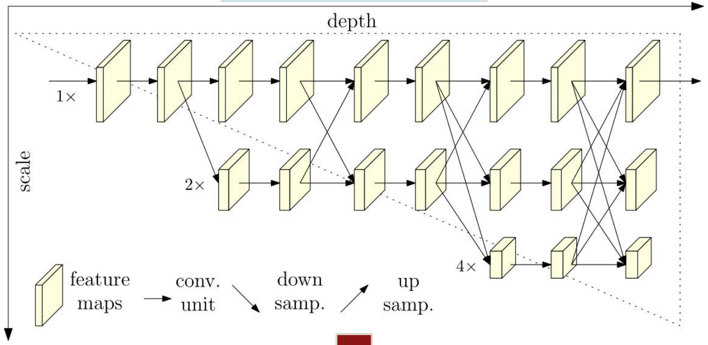


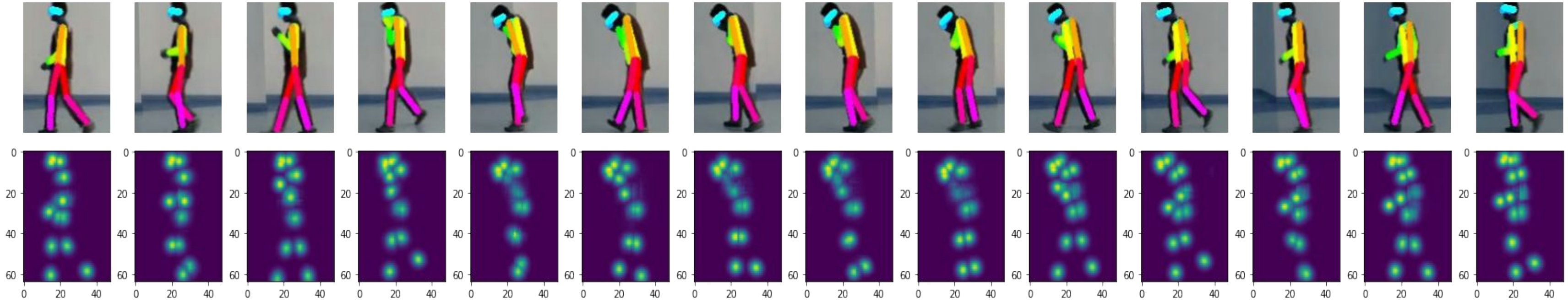
# Data Pre-processing (Cont'd)

- HRNet-Based Models:

Input Video → 60 frames (480, 290) →

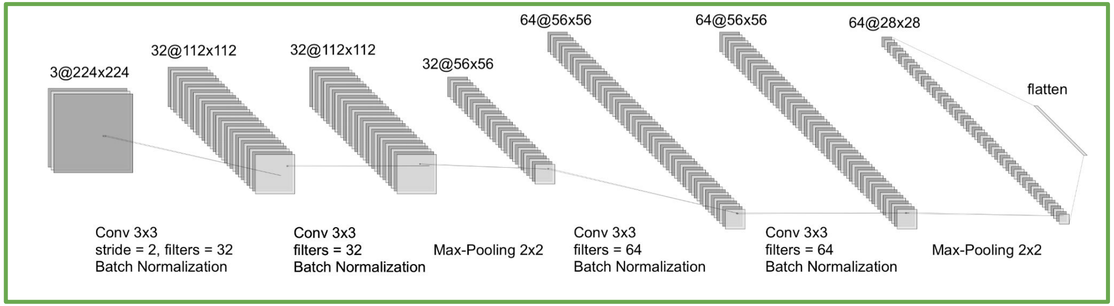
HRNet

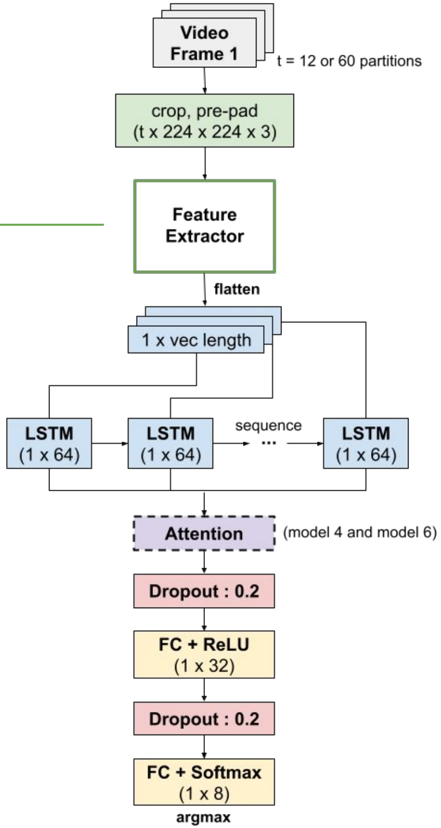


17 Heatmaps (64, 48)

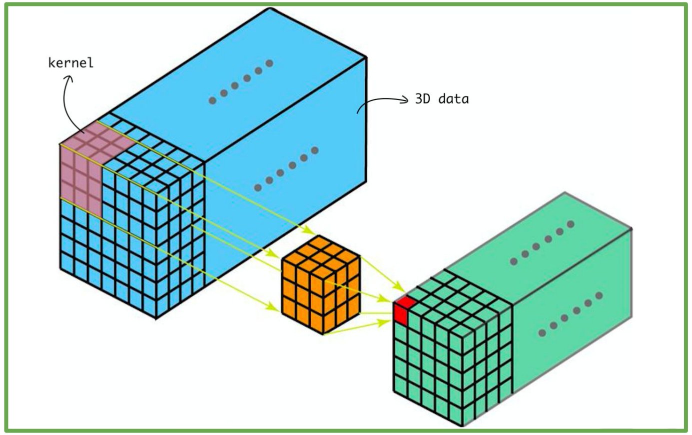
# Models - CNN + LSTM (baseline 1)




- Categorical cross-entropy loss is used.
- Trained the whole network from scratch using an Adam optimizer.

# Models - 3D-Conv (baseline 2)

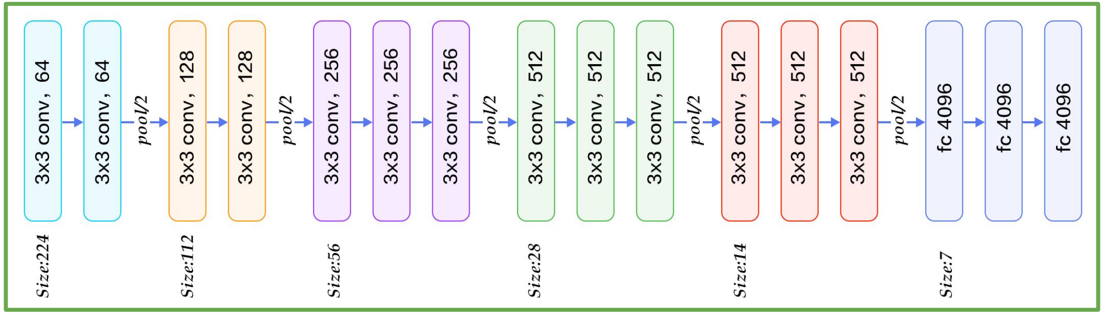


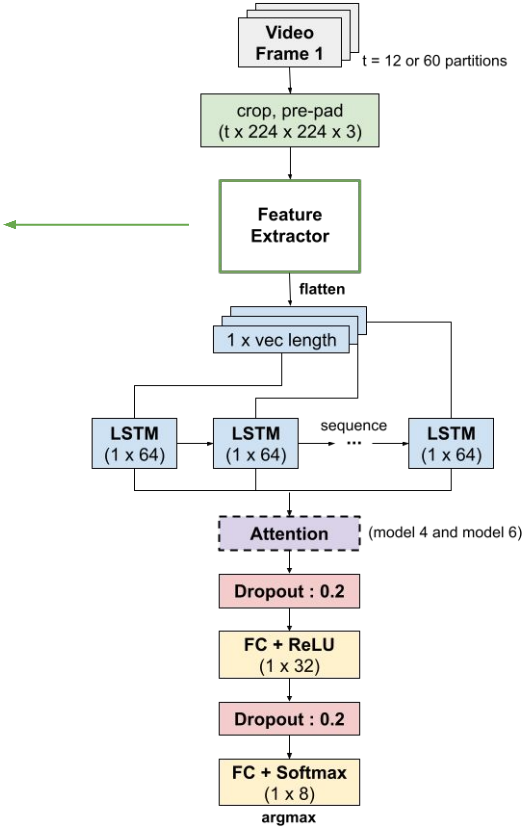

Example of a 3D convolution performed with 3D kernel and 3D data - https://towardsdatascience.com/9d8f76e29610

| Layer (type) | Output Shape | Param # |
|---|---|---|
| conv1 (Conv3D) | (None, 12, 224, 224, 32) | 2624 |
| pool1 (MaxPooling3D) | (None, 12, 112, 112, 32) | 0 |
| conv2 (Conv3D) | (None, 12, 112, 112, 64) | 55360 |
| pool2 (MaxPooling3D) | (None, 6, 56, 56, 64) | 0 |
| conv3a (Conv3D) | (None, 6, 56, 56, 128) | 221312 |
| conv3b (Conv3D) | (None, 6, 56, 56, 128) | 442496 |
| pool3 (MaxPooling3D) | (None, 3, 28, 28, 128) | 0 |
| conv4a (Conv3D) | (None, 3, 28, 28, 256) | 884992 |
| conv4b (Conv3D) | (None, 3, 28, 28, 256) | 1769728 |
| pool4 (MaxPooling3D) | (None, 1, 14, 14, 256) | 0 |
| flatten_2 (Flatten) | (None, 50176) | 0 |
| fc6 (Dense) | (None, 64) | 3211328 |
| dropout_2 (Dropout) | (None, 64) | 0 |
| dense_2 (Dense) | (None, 8) | 520 |

Total params: 6,588,360
Trainable params: 6,588,360
Non-trainable params: 0

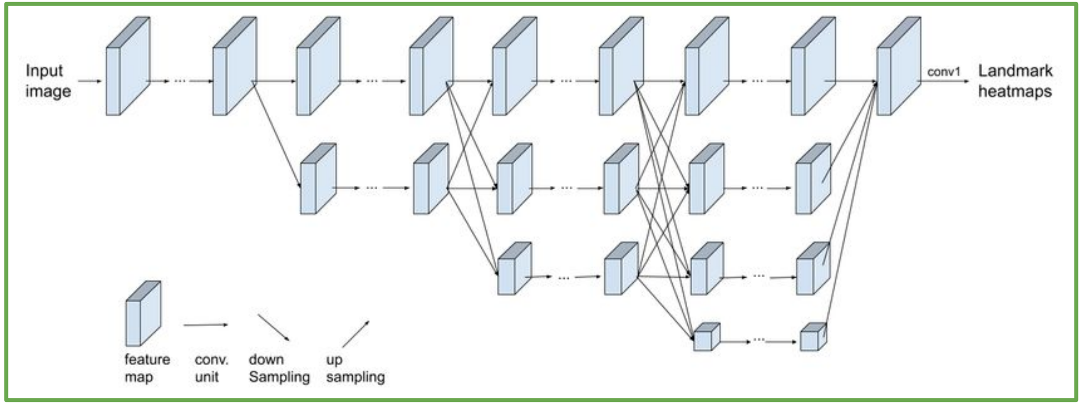
# Models - VGG-16 Features + LSTM



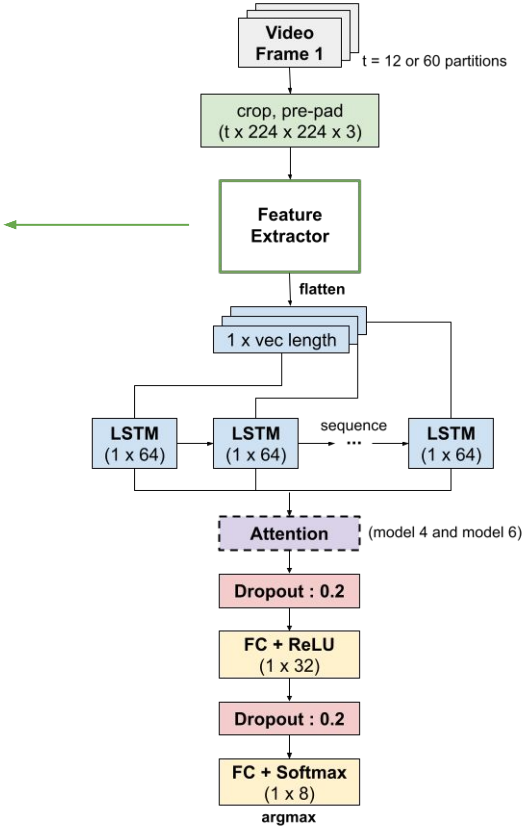




- A pre-trained VGG-16 model is used as feature extractor.
- The parameters in VGG-16 are freezed.
- Explored adding an Attention layer after the LSTM layer.

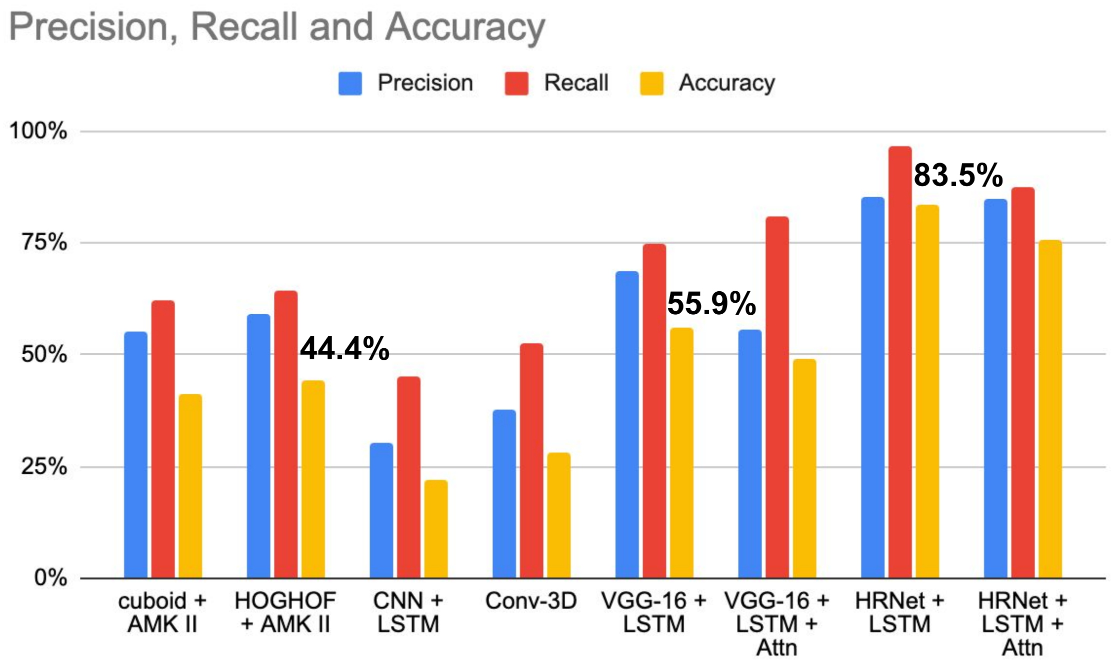
# Models - HRNet Features + LSTM




- A pre-trained HRNet model is used as feature extractor.
- High Resolution Net (HRNet) is a state of the art neural network for human pose estimation.
- Explored adding an Attention layer after the LSTM layer.



# Results




Prec. = TP / (TP + FP)
Rec. = TP / (TP + FN)
Acc. = TP / (TP + FP + FN)

# Future works

- 3D-Conv + Attention model
- Impact of extra source of information (e.g. sound)
- Other applications of the network architecture