CS 231N

Recognize Flu-like Symptoms with Deep Learning

Alan Lou
Stanford University
alanlou@stanford.edu

Yechao Zhang
Stanford University
yechaoz@stanford.edu

1 Problem Description

The novel coronavirus disease 2019 (COVID-19) pandemic has reached almost every country in the world, infecting millions of people and plunging the global economy into recession as governments imposed tight restrictions to tackle the spread of virus. Despite their wide availability now, Covid tests and vaccines are usually inaccessible at the early stage of the outbreaks, especially for unknown viruses. Early detection of the outbreaks therefore plays a vital role in preventing rapid escalation. Vision-based surveillance such as monitoring and analyzing video footage from densely populated area to predict flu-like symptoms can be utilized to detect respiratory viral infections early.

The previous work [3] predicts flu-like symptoms from videos with Action Matching Kernels (AMKs) and Support Vector Machine classifier (SVM). For this project, we are going to improve the classification with adoption of the recent deep learning models leveraged for computer vision.

2 Dataset

We plan to use the BII Sneeze-Cough Human Action Video Dataset (BIISC) [2] created by the authors of the paper mentioned above [3]. Below is a summary of the dataset statistics:

- Number of subjects: 20, M/F:12/8
- Number of action types: 8, answer phone call, cough, drink water, scratch head, sneeze, stretch arms, wave hand, wipe glasses
- Number of poses: 3, face to the camera, face to the left, face to the right
- Number of locomotion types: 2, standing, walking
- An extra horizontally flipped version has been synthesized for each of the videos.
- Number of videos in total: 20x8x3x2x2=1920

3 Methods

Recognizing flu-like symptoms from videos is an action recognition problem. Convolutional Neural Network (CNN) and Long short-term memory (LSTM) have shown to be successful in the field of action recognition [4]. Attention-based LSTM models have also gained interests recently [1].

In this paper, we propose a flu-like symptoms recognition framework by utilizing frame level deep features of the CNN and processing them through recurrent neural networks like LSTM and its variations. This will help recognize complex frame to frame hidden sequential patterns in the features. Additionally, we will explore if incorporating attention into the LSTM network help improve the model performance.

We are planning to implement different combinations of the CNN and LSTM models, which include but not limited to:

Figure 1: From top to down shows eight actions: answer phone call, cough, drink, scratch head, sneeze, stretch arm, wave hand and wipe glasses. From left to right shows six pose-and-view variations: stand-front, stand-left, stand-right, walk-front, walk-left, and walk-right.

- CNN + LSTM
- Deep Layered CNN + LSTM
- Pretrained VGG16 + LSTM
- FineTuned VGG16 + LSTM
- FineTuned VGG16 + LSTM + Attention

The inputs of our model are video frames and the outputs are action labels corresponding to the videos.

4 Evaluation

The goal of our model is to recognize flu-like symptoms (sneeze and cough) among a variety of actions, which makes it a binary classification task. Since the class distributions are imbalanced, we are going to use F1 score to evaluate the model performance quantitatively.

The previous work [3] proposed an accuracy measure of $\frac{TP}{TP+FP+FN}$ for the binary classification, which can be regarded as a lower-bounding summary of the (precision, recall) pair. We will calculate this measure for our models as well and use it to compare against the previous work.

References

[1] Shikhar Sharma, Ryan Kiros, and Ruslan Salakhutdinov. "Action Recognition using Visual Attention". In: *CoRR* abs/1511.04119 (2015). arXiv: 1511.04119. URL: http://arxiv.org/abs/1511.04119.

[2] Tuan Hue Thi et al. *Recognizing flu-like symptoms from videos*. 2014. URL: https://web.bii.a-star.edu.sg/~chengli/FluRecognition.htm.

[3] Tuan Hue Thi et al. "Recognizing flu-like symptoms from videos". In: *BMC Bioinformatics* 15.1 (2014), p. 300. DOI: 10.1186/1471-2105-15-300. URL: https://doi.org/10.1186/1471-2105-15-300.

[4] A. Ullah et al. "Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features". In: *IEEE Access* 6 (2018), pp. 1155–1166. DOI: 10.1109/ACCESS.2017.2778011.