BLACK: Body Language Affect Classification Kernel

Undergrad-ient Descent Expedition:
Jack (Xiang) Zhou, Karan Aujla, James Bie, Eric Gao, Insoo Rhee
Faculty of Applied Sciences
Simon Fraser University
Burnaby, Canada
xza194@sfu.ca

Abstract

This is the abstract of the paper that we are going to write

1 Introduction

Task is to assign affective label on humans, face based affective label assignment is not very good atm, use body as a supplement.

Task is difficult even for humans.

Talk about COBOL [2], work has been done before with a neurologically inspired model but not on a skeleton level

Happy is most compact and separable from other emotions in multiple findings, link that back to our own findings later on.

2 Approach

2.1 Architecture

We introduce the general architecture of the overall pipeline in this section. The novel and primary goal of the present work is to assign affective labels with learned patterns in body langauge. However, as we are only in the primary stage in this pursuit and we do not have a comprehensive method toward accounting for viewpoint invariance, using body posture as the primary is not very realistic. In the present work we use body language mainly as a supplement to existing facial affective labels if it is detected.

A snapshot is taken from an arbitrary visual stream and the resulting image is passed into OpenPose to detect all persons. For each person: (1) If a face was found, then it is passed into a face-based classifier for an evaluation of the affective rating vector based solely on face. (2) If a body was found, then it is passed into a body language-based classifier for an evaluation of the affective rating vector based solely on body. A weighted sum is computed between (1) and (2) for each person. If only one of (1) or (2) is found, then only that one is taken. A grand total is calculated by summing over every person's weighted sum of affective labels. The grand total will be passed into a softmax function to obtain a probability vector of the model prediction. A graphical depiction of this overall process is shown in Figure 1.

2.1.1 Body Language Keypoint Estimation

To extract body keypoints, we used the OpenPose system produced by the CMU Perceptual Computing Laboratory [1] [3] [4]

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

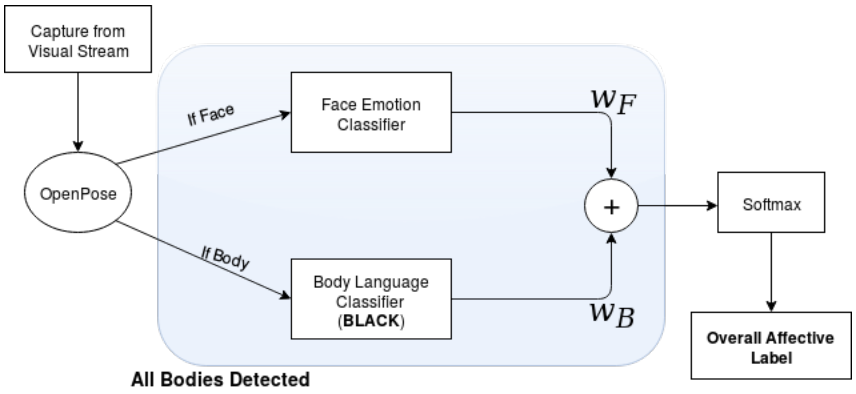

Figure 1: The overall pipeline.

Openpose is a system the detects people in an image or video and finds the estimated location of several keypoints on their body, hands, and face. It is a capable of finding the keypoints of multiple people in the same image. Openposes's basic algorithm finds each type of keypoint in all bodies at the same time and finds the connection between keypoints in a separate neural network. This program can take most image and video formats as inputs, such as jpeg, png, avi, and mp4 formats. It will output in a variety of formats depending on the flags given. For example, it can output an image with the skeletons overlaid on the people in the image, and output json files for the keypoints. In the json file, it contains the x, y coordinates of every bone detected, as well as a confidence rating for the position of each bone. Openpose can output the json in a variety of formats, such as grouping each x,y,c tuple into it's own object. For our model, openpose is configured to output a json file containing a separate object for each bone and to normalize the coordinates to [-1,1]. Then we find the angles of each bone to ensure that the input values are position invarient

2.1.2 Face Emotion Classification

aaaaaaa

2.2 Training

As the problem that the present model is trying to solve involve uncensored face along with body posture, it is particularly difficult to obtain relevant published datasets that are labelled and publicly available. Past research in this field involved manual construction of data from actors [2].

To address this issue, we manually construct a dataset of unlabelled images of humans with visible face and body posture by using authors of this paper as actors as well as sources on the internet. The face and body are both clearly visible in the training images. The image will not be given a label when it is fed into the training pipeline in a pseudounsupervised learning paradigm. We describe the training pipeline as follows:

The unlabelled image is first passed into OpenPose to obtain a set of posture keypoints (ie. a skeleton) of the human in the image, as well as a set of face keypoints. The face keypoints will be passed into an already trained face emotional classifier to give a ground truth affective label for the overall human in the image as informed by the face. The skeleton is then processed into an invariant form and used along with the affective label to train the body language affect classifier.

3 Experiments

4 Conclusion

4.1 Future Work

using a generalized linear model to deal with multiple persons instead of a simple sum

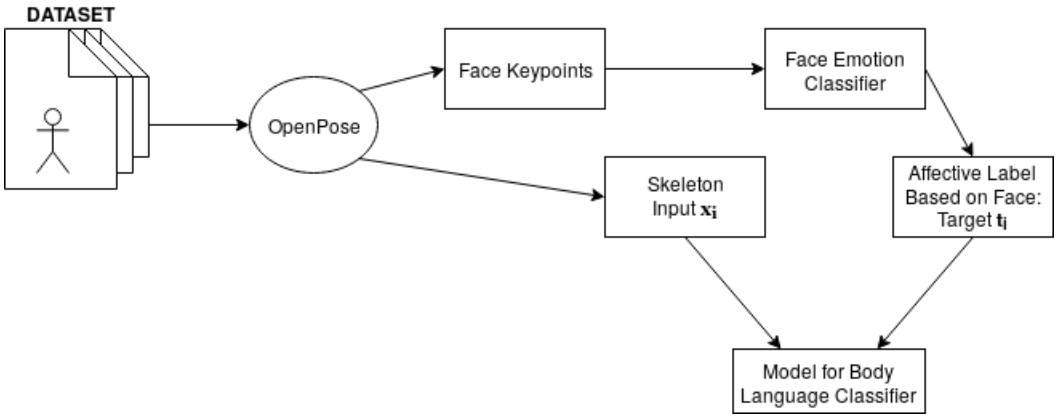

Figure 2: The training pipeline.

fitting LSTM to generalize into videos

accounting for Confidence in body keypoint measurements dynamically

5 Contributions

As with most group projects, each author of this paper contributed a considerable amount of work towards piecing together the project. Jack oversaw the project by organizing and delegating tasks for everyone as well as being the main composer of the paper and poster. Jack and James formulated the design of the overall pipeline for the models and for training the model. Karan worked with existing models for body posture keypoint extraction and implemented the top level program. Eric and James are the main contributors to data collection and piecing together a custom dataset for the project. Insoo and Jack worked on designing and implementing the model for mapping body keypoints to affective labels.

We would also like to thank Professor Angelica Lim for consultations and guidance toward the design and theoretical groundwork for the project.

References

[1] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017.

[2] Konrad Schindler, Luc Van Gool, and Beatrice de Gelder. Recognizing emotions expressed by body pose: A biologically inspired neural model. *Neural networks*, 21(9):1238–1246, 2008.

[3] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*, 2017.

[4] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *CVPR*, 2016.