Introduction to digital imaging
final project report

Insulting hand gesture detector

Nuttapol Onton 6330166021
Natchapon Chitturongaporn 6330148721

1. Introduction

Social media platforms have become integral to global communication, providing individuals with the ability to express themselves, share experiences, and connect with others. Unfortunately, the anonymity offered by these platforms has given rise to a concerning trend of inappropriate and offensive content, often conveyed through visual cues such as hand gestures. Understanding the importance of maintaining a positive and inclusive digital space, our project aims to address the issue of insulting gestures by using digital image processing techniques.

In this project, our primary focus is on detecting and censoring the middle finger gesture, a widely recognized and offensive hand gesture, in real-time during social media interactions. To achieve this, we employ a targeted approach using pre-trained models and real-time censorship mechanisms.

2. Theory and concept

2.1 Mediapipe

Mediapipe is an open-source framework developed by Google that provides developers with a pre-built and customizable machine learning pipeline for visual perception tasks. These tasks include hand tracking, face detection, pose estimation, object recognition, and more. MediaPipe simplifies the development of applications involving computer vision and machine learning by offering ready-made tools and components. It is particularly useful for augmented reality, virtual reality, and robotics applications, allowing developers to focus on specific use cases without having to build models from scratch. The framework supports both desktop and mobile platforms.

The Mediapipe pipeline is as follows.

1. Input Modules : Receives raw data, like videos frames or images, as the initial input of the pipeline
2. Preprocessing Model : Applies necessary steps to prepare the input data for subsequent stages.
3. Model Modules : Encompasses machine learning models specialized for distinct tasks such as face detection, hand tracking, or pose estimation.
4. Postprocessing Module: Refines the output from the models, processes the results, and readies the final output for application use.
5. Output Module: Supplies the ultimate output of the pipeline, which could be visual representations, data structures, or information pertinent to the application.

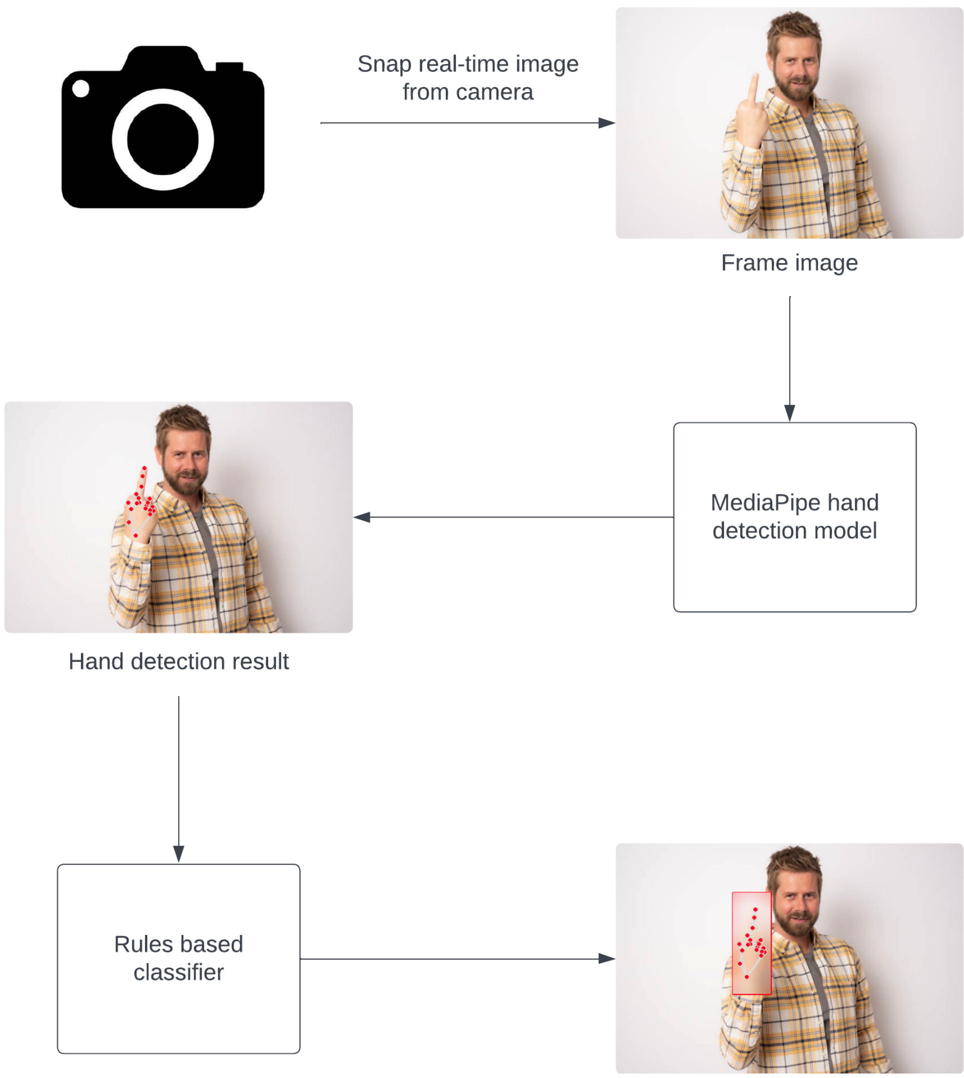
2.2 Rule-based machine learning

Rule-based machine learning involves making decisions using a set of predefined rules, typically expressed in "if-then" statements. These rules are often crafted by domain experts or derived from data through rule induction. Such systems are transparent and interpretable, as decisions can be traced back to specific rules, making them suitable for scenarios where understanding the decision-making process is important. Rule-based models, like decision trees, are examples of this approach. However, they may face challenges with complex relationships or large datasets, and more advanced machine learning models may be more effective in such cases.

2.3 Gaussian blur

Gaussian blur is based on the mathematical concept of a Gaussian function, also known as a normal distribution. The Gaussian function is a bell-shaped curve that describes the distribution of values in a dataset. In image processing, this function is used to create a convolution kernel, which is then applied to the image to achieve the blurring effect.

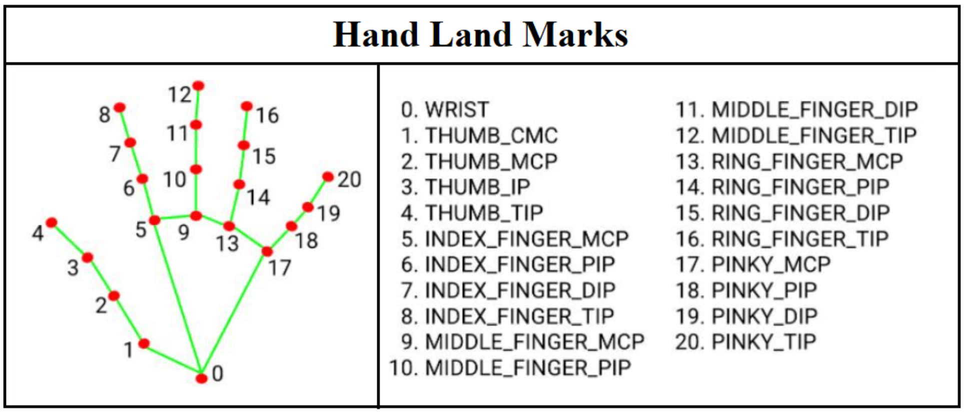
3. Implementation



To implement the Insulting Hand Gesture Detector, we propose a two-fold approach that combines the accuracy of the MediaPipe hand detection model with a rule-based classification system to identify middle finger hand gestures.

1. Hand detection

First, we use MediaPipe hand detection, we identify hands and capture the precise finger positions. This information is then integrated into a rules-based classifier in the subsequent stage.



2. Rules-based classification

Applying a rules-based classification approach involves determining whether each image represents a middle finger hand gesture by analyzing the finger positions obtained from hand detection.

Rules:

Let finger_length = | TIP - MCP |
and std_length = middle_length / 2

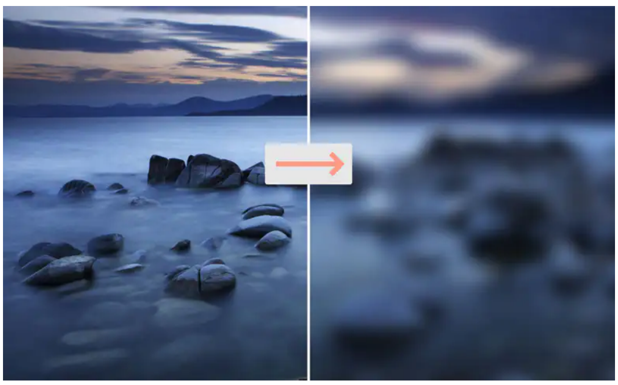
If std_length > index_length and std_length > ring length and std_length > pinky_length

Description:

The provided rules involve defining a variable, finger_length, as the absolute difference between the lengths of the fingertip (TIP) and the metacarpophalangeal joint (MCP) for a given finger. Additionally, another variable, std_length, is established, representing half the length of the middle finger. The condition specified in the rules checks if std_length is greater than the lengths of the index, ring, and pinky fingers. If this condition is met, it implies that the standard length is comparatively larger than the lengths of these fingers, suggesting a potential configuration consistent with a specific hand gesture, such as the middle finger gesture.

3. Gaussian blur

Apply Gaussian blur to images featuring a middle finger hand gesture as identified through the classification process.



4. Evaluation

4.1 Testing data

For the evaluation phase, we employed datasets from two distinct sources to ensure a comprehensive assessment of our model's performance:

1. Hand Dataset on Kaggle
2. HOD Benchmark Dataset on GitHub

To create a representative testing subset, we strategically sampled 150 images containing middle finger hand gestures and an additional set of 150 images devoid of middle finger hand gestures.

We separated datasets into 3 kinds of testing data

1. Normal case (50 insulting hand gesture + 50 non-insulting hand gesture)
2. Hard case (100 insulting hand gesture + 100 non-insulting hand gesture)
3. All case (150 insulting hand gesture + 150 non-insulting hand gesture)

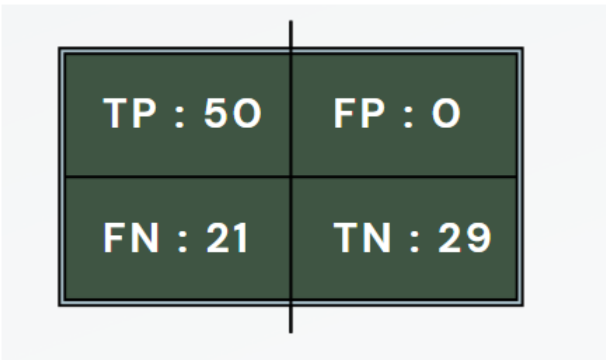
4.2 Result

4.2.1 Accuracy

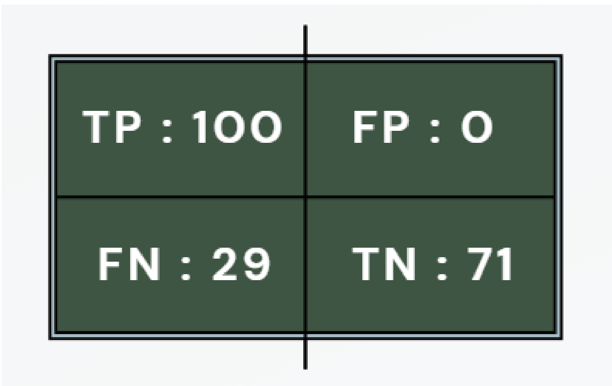
- Normal Case : 0.855
- Hard Case : 0.79
- All Case : 0.833

4.2.2 Confusion matrix

- Normal Case :



- Hard Case :



- All Case :

TP : 150	FP : 0
FN : 50	TN : 100

4.2.3 Classification Report

- Normal Case :

```
Classification Report:
              precision    recall  f1-score   support

           0       0.78      1.00      0.87       100
           1       1.00      0.71      0.83       100

    accuracy                           0.85       200
   macro avg       0.89      0.85      0.85       200
weighted avg       0.89      0.85      0.85       200
```

- Hard Case :

```
Classification Report:
              precision    recall  f1-score   support

           0       0.70      1.00      0.83        50
           1       1.00      0.58      0.73        50

    accuracy                           0.79       100
   macro avg       0.85      0.79      0.78       100
weighted avg       0.85      0.79      0.78       100
```

- All Case :

```
Classification Report:
              precision    recall  f1-score   support

           0       0.75      1.00      0.86       150
           1       1.00      0.67      0.80       150

    accuracy                           0.83       300
   macro avg       0.88      0.83      0.83       300
weighted avg       0.88      0.83      0.83       300
```

4.3 Analysis

In the evaluation of this model's performance, we identified challenges related to side-view hand gestures, blurred vision, glove-wearing, and hand gesture recognition in Black individuals. Consequently, the model exhibits limitations in addressing these challenges.




However, upon examining the confusion matrix, it is noteworthy that there are no false positives. This implies that the model has not incorrectly classified any instances of the negative class as positive. This absence of false positives can be considered a positive aspect, suggesting that the model performs well in terms of not missing positive cases.

In conclusion, this model demonstrates good performance in detecting the middle finger hand gesture. It could be effectively utilized for detecting such gestures in social media and implementing censorship measures. Consequently, this model can play a crucial role in filtering out some toxic hand gestures on social media platforms. This helps a social community be more positive and respectful.

5. Comparation

We have another possible way to detect the middle finger gesture, such as training a new object detection model and using images of the middle finger gesture as training data. With this solution, there is potential for improved accuracy, but, of course, we need to train a new model and collect more data for training.

On the other hand, the solution we implemented in this project to detect the middle finger gesture doesn't require training a new model, we use the pre-trained model from MediaPipe and apply rules for detection. Moreover, the accuracy achieved with this solution is acceptable, making it a worthwhile and straightforward approach.