Sign Language Recognition Using MediaPipe and Opencv

Mohamed Mostafa (ID: 162021295), Martina Mamdouh (ID: 162021251), Micheal Anton (ID: 162021256)

Faculty of Computetrs and Information Assiut University

April 2024

Outline

Task Description

Our model enables real-time hand gesture recognition for American Sign Language (ASL) words using a webcam. Then, by Using MediaPipe for hand landmark detection, key points are extracted which represents hand gestures. These key points serve as features for classification. With a Convolutional Neural Network (CNN), trained on a dataset of 10 ASL words, our model achieves over 95% accuracy.

Demo

https://github.com/MohamedMostafa21/Sign-Language-Recognition-Using-MediaPipe-and-OpenCV/blob/main/30$_D$*emo*.*mp*4

Contribution

1- Using ASL vocabulary:
Instead of utilizing four random labels as the original model, we improved our dataset with 10 commonly used ASL words, this is to enhance the relevance and practicality of our model.

2- Custom dataset creation:
We created our own dataset comprising of 3006 images, precisely classified into training and testing sets with a 75:25 ratio, to ensure robust model training and evaluation and to reduce overfitting.

Contribution

3- Two hand detection:
We enabled detection of gestures from both hands at the same time, this model offers users the ability to convey multiple signs concurrently, this is to enhance the versatility and utility of the system.

4- User-friendly GUI:
We developed a simple graphical user interface (GUI) to simplify the detection for non-technical users, ensuring accessibility and usability of our model across diverse user groups.

Contribution

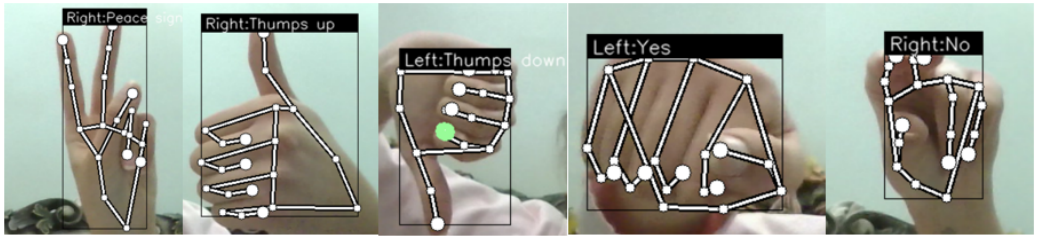
5- Dataset augmentation and hyperparameter tuning:
We expanded our dataset and explored various activation functions and training/testing ratios to optimize model accuracy, to achieve the highest performance.

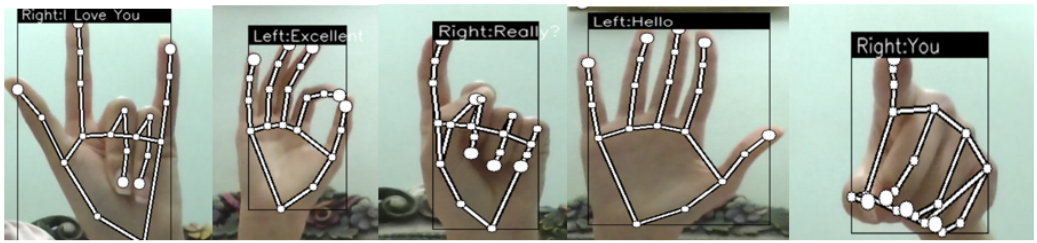
6- Adjusted some regularization techniques:
We implemented regularization techniques such as dropout and batch normalization To reduce overfitting and enhance our model generalization. This to decrease overfitting and improve our model performance.

Data

1- In this model 10 American sign language Words were used for classification. The dataset was constructed with '3306' images with a ratio of (75:25) trainging:testing.

2- We have 10 classes of words which are : Peace sign , thumps up ,thumps down ,yes,no ,excellent,i love you,hello,really,you.





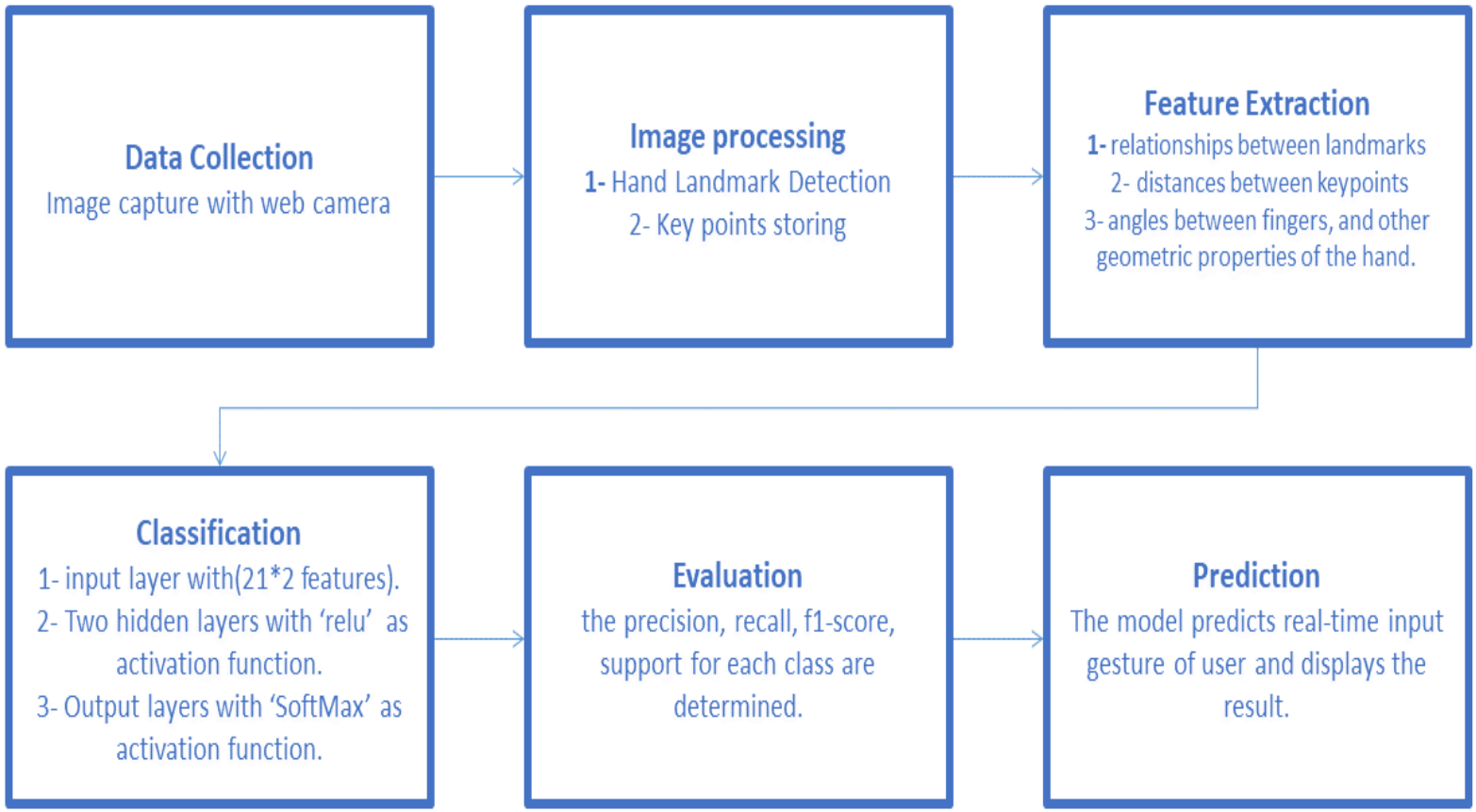
Data

3- images were saved as PNG or in JPG format.

4- In this model , The key-points were extracted using MediaPipe landmark detection model keypoints from handlandmarks on palm. those keypoints pass through the neural network architecture ,then occurred an extraction to the features and that gives us the final classification.

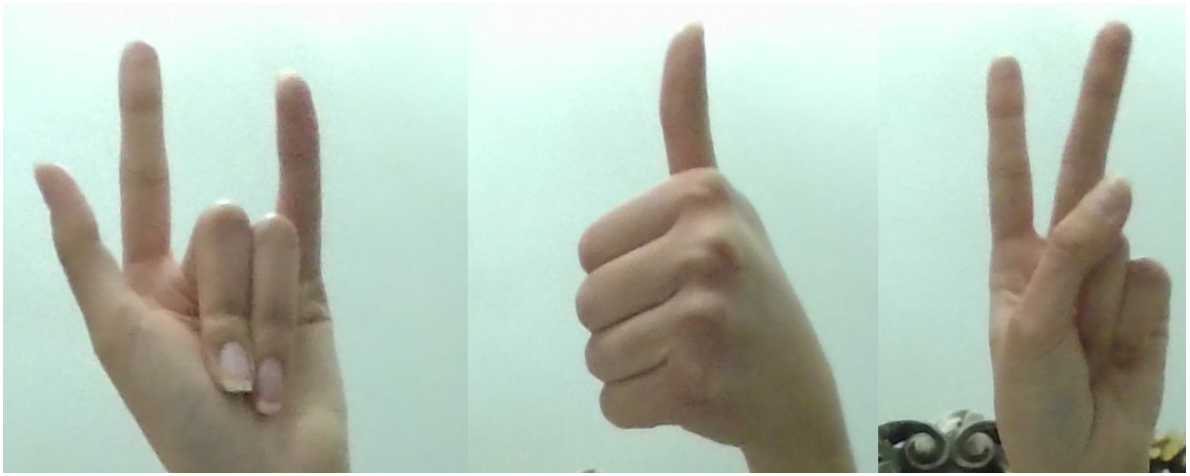
5- The webcam captures the images in the RGB colourspace.

Architecture



Methods (Data Collection)

The model captures hand signs using web cam .



Methods (image Processing)

Then uses Mediapipe's Hand landmarks Detection Model to locate hands within the frame and extract the hand landmarks

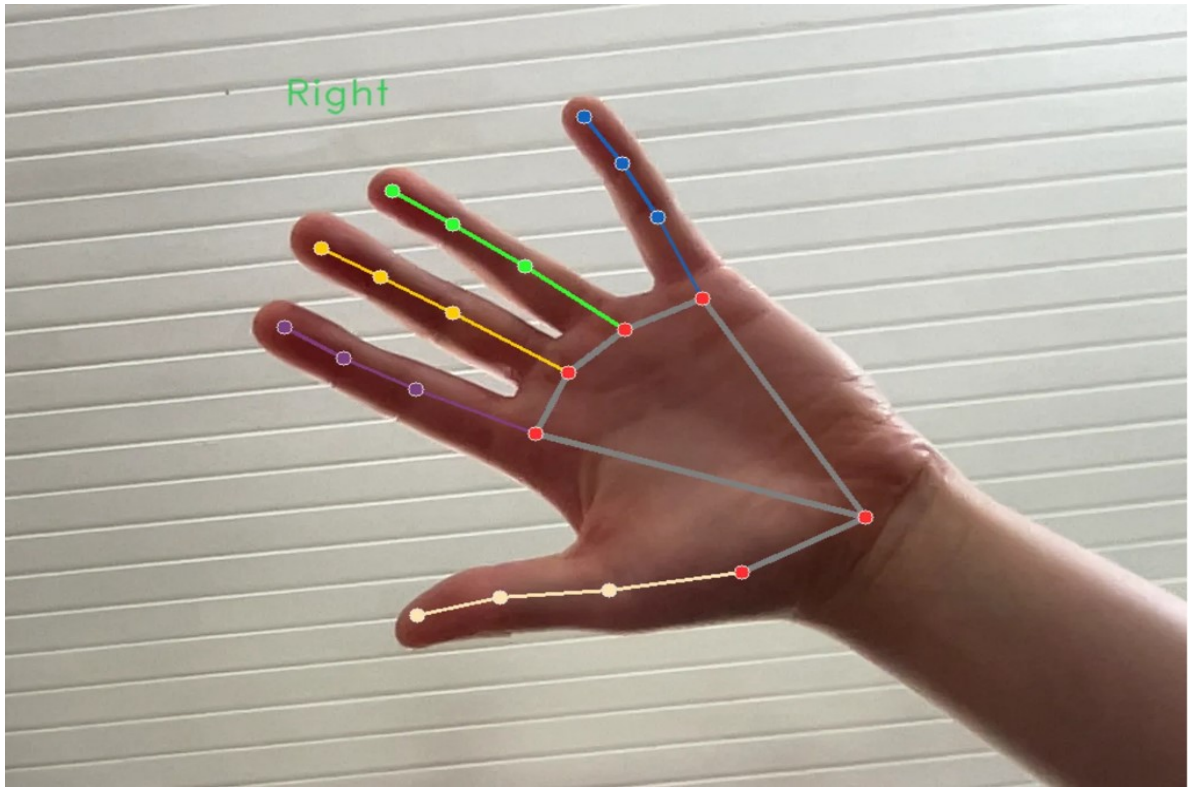
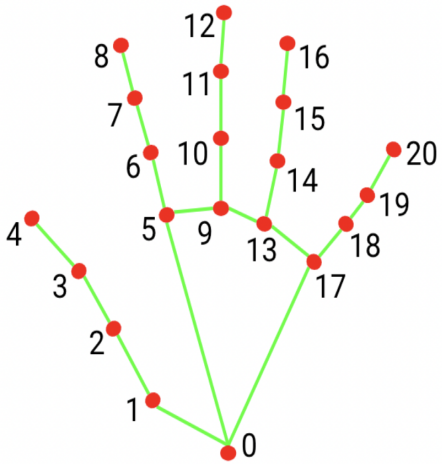


Figure: Palm Detection landmarks extraction

Methods (Feature Extraction)

In this step it extract the useful information from it like the relationship between landmarks, distance, angles, and other geometric properties of the hand



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

the 21 landmarks

Methods (Classification)

Its neural network has four layers

- Input layer with 21*2 features
- Two hidden layers with "ReLU"
- Output layer with "Softmax" as its activation function

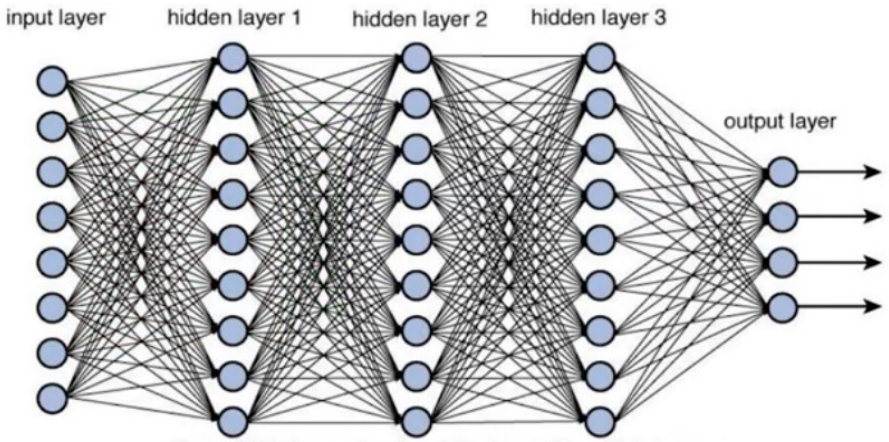


Figure 12.2 Deep network architecture with multiple layers.

Methods (Evaluation)

Here, the model gets evaluated for its precision, accuracy and the loss function is calculated

Methods (Prediction)

The model here predicts real time gestures of users and display the results

Figure: Palm Detection landmarks extraction

Results

measures	precision	recall	f1-score	support
Peace Sign	1.00	0.97	0.99	78
Thumps up	0.99	0.98	0.99	113
Thumps down	0.99	1.00	0.99	69
Yes	0.91	0.94	0.92	63
No	0.97	0.96	0.96	69
I Love You	1.00	1.00	1.00	91
Excellent	1.00	0.79	0.88	75
Really?	0.95	1.00	0.98	101
Hello	0.86	0.98	0.92	114
You	1.00	0.93	0.96	54
accuracy			0.96	827
macro avg	0.97	0.95	0.96	827
weighted avg	0.96	0.96	0.96	827