Hand Gesture Recognition

Abisek R K, 21Q050004
Tejpal, 21Q050008
Naveen, 213050052
Mohan Rajasekhar, 213050060

Task Description

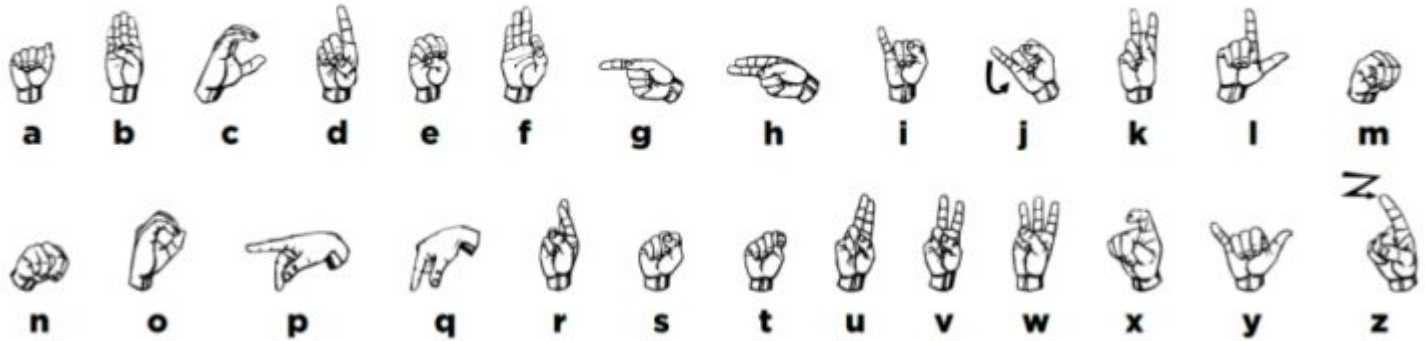
- Hand Gesture Recognition is very important for human-computer interaction in various applications of AR/VR, Robotics etc.
- A large portion of previous research requires specialized hardware like depth sensors and are not light-weight enough to run on real-time.
- We developed a real time application to recognize the Hand Gesture and convert it into text interactively.
- Our application is able to recognize new Hand Gestures along with Default gestures.
- For Default gestures, we tried testing our method on ASL alphabet letters.

Dataset

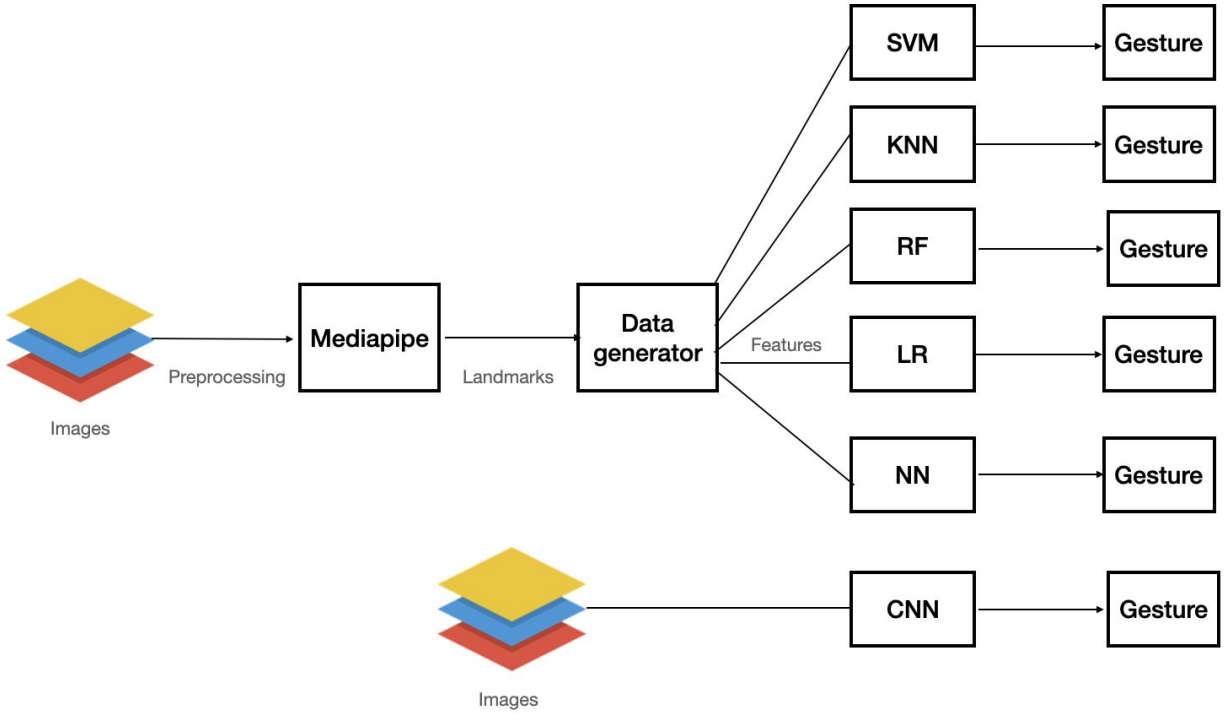
- ASL Alphabet

  https://www.kaggle.com/grassknoted/asl-alphabet

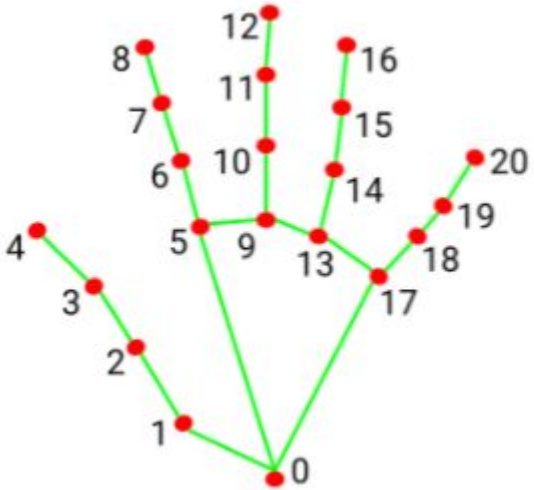
- Collection of American Sign Language(ASL) alphabets images which was separated into folders based on alphabets.
- Each alphabet consists of 3k color images.
- Resolution : 200 X 200



Approach


SVM
Gesture
KNN
Gesture
RF
Gesture
LR
Gesture
NN
Gesture
CNN
Gesture
Images
Preprocessing
Mediapipe
Landmarks
Data generator
Features
Images

Approach - Mediapipe Landmarks



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

https://google.github.io/mediapipe/solutions/hands.html

Dataset Preprocessing

- Removing images from the dataset whenever landmarks are not detected.
- For custom gesture recognition the images are captured only when the landmarks are detected.
- Dropping the z coordinates from each feature/landmark.

Approach - Default Gesture

- Preprocessed images are passed through “Mediapipe” library to obtain landmarks.
- 21 Landmarks are captured for each image.
- The cartesian coordinates of these landmarks are identified and used as training data.
- ML/DL models are trained on image features represented by their landmark coordinates.
- The user will be able to input a gesture through webcam, which will be passed through the mediapipe and ML models layers, eventually predicting the text (or voice) equivalent to the input gesture.

Approach - Custom Gesture

- User will have an option to add custom gesture on the go.
- User can define a custom gesture and input a few samples of the gesture.
- From the samples received as input, important landmark coordinates will be identified and added to the existing training data.
- Our light-weight ML models will be retrained on this new data within minutes.
- The user can then use these new gestures as if they were already present before.

Results

ML Model	Micro-Average f1-score	Macro-Average f1-score	Weighted Average f1-score
Logistic Regression	0.960130	0.956708	0.960227
KNN	0.932603	0.931085	0.933061
SVM	0.986196	0.984910	0.986231
Random Forest	0.979050	0.977642	0.979032
Neural Networks	0.984166	0.982567	0.984185
CNN	Train Accuracy: `0.9915` Test Accuracy: `0.9839`		

Work Distribution

- **Data collection, preprocessing, generation + Mediapipe** : Team effort
- **Training using Machine Learning models** : Models were divided between different team members.
- **OpenCV and Front End UI** : Naveen and Tejpal
- **Custom gesture generation** : Rajasekhar and Abisek

Source Code + Demo

- **Link to source code :** https://github.com/naveen-badathala/CS725-2021-Hand_Gesture_Recognition.git
- **Link to demo :** https://drive.google.com/file/d/1iqJGpvdR-TV2sI9IWbKrx9mZS9vaur4q/view?usp=sharing
- **Acknowledgement :**
  - https://github.com/sid-1998/Sign-Language-Recognition/blob/master/Gesture_Recognize_sign.py (UI reference)

Thank You

References

- Mediapipe Reference: https://arxiv.org/pdf/2006.10214.pdf
- https://google.github.io/mediapipe/solutions/hands.html
- https://techtutorialsx.com/2021/04/10/python-hand-landmark-estimation/
- https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn
- https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python
- https://machinelearningmastery.com/neural-network-models-for-combined-classification-and-regression/
- https://www.kaggle.com/yasinsoylu123/asl-recognition-keras-cnn/notebook
- https://www.datacamp.com/community/tutorials/random-forests-classifier-python
- https://towardsdatascience.com/logistic-regression-using-python-sklearn-numpy-mnist-handwriting-recognition-matplotlib-a6b31e2b166a
- https://github.com/sid-1998/Sign-Language-Recognition/blob/master/Gesture_Recognize_sign.py
- https://stackoverflow.com/questions/32609098/how-to-fast-change-image-brightness-with-python-opencv
- https://stackoverflow.com/questions/66876906/create-a-rectangle-around-all-the-points-returned-from-mediapipe-hand-landmark-d