Report

Gesture recognition task:

The Gesture Recognizer uses a model bundle with two pre-packaged model bundles: a hand landmark model bundle and a gesture classification model bundle. The landmark model detects the presence of hands and hand geometry, and the gesture recognition model recognizes gestures based on hand geometry.

- For hand gesture recognition the number of fingers present in the hand gesture is calculated by CNN, using fault spots. The acquired gesture is passed via a 3-dimensional Convolutional Neural Network. CNN is used in succession to recognize the current gesture.

Here are some algorithms used for hand gesture recognition

- **K-Nearest Neighbour (KNN)**

  A supervised machine-learning algorithm that uses Euclidean distance to find the nearest neighbour. KNN is used for classification, where a data point's classification is determined by how its neighbour is classified.

  KNN is an algorithm used to predict the hand gestures displayed on the camera. The reason we chose MediaPipe to detect hands is that it is easy to use and can provide output in the form of hand landmark coordinates.

  The KNN algorithm can compete with the most accurate models because it makes highly accurate predictions. Therefore, we can use the KNN algorithm for applications that require high accuracy but that do not require a human-readable model. The quality of the predictions depends on the distance measure. Hence, KNN is a good algorithm for gesture recognition.

- **Convolutional Neural Network (CNN)**

  A 3-dimensional Convolutional Neural Network that calculates the number of fingers present in the hand gesture. CNN is used in succession to recognize the current gesture.

  We use the CNN approach to understand hand gestures. The CNN algorithm is used to classify an image based on various characteristics and make it possible to differentiate it from its respective classes. By passing, the input images across different layers, the CNN technique works.

We use CNNs (Convolutional Neural Networks) in image processing because they can effectively extract features from images and learn to recognize patterns, making them well-suited for tasks such as object detection, image segmentation, and classification.

- **Support Vector Machine(SVM)**

  The SVM algorithm's objective is to establish the optimum decision boundary or line that can divide n-dimensional space into classes so that we may quickly classify fresh data points in the future. For hand gesture recognition through SVM, first hand motion is acquired using a web camera, during processing.

  SVM is considered one of the best algorithms because it can handle high-dimensional data, is effective in cases with limited training samples, and can handle non-linear classification using kernel functions.

- **CNN-based model**

  A model that uses a depth image-based segmentation method to extract the gesture region. After feature extraction using a CNN, final recognition is done using the SVM method.

Comparisons Between the Models

❖ CNN is better than SVM

Convolutional Neural Networks (CNNs) are typically better than Support Vector Machines (SVMs) for image classification because they are able to learn more complex features from images. CNNs are specifically designed to extract features from images, while SVMs are more general-purpose classifiers.

Clearly, the CNN outperformed the SVM classifier in terms of testing accuracy. In comparing the overall correctacies of the CNN and SVM classifiers, CNN was determined to have a static-significant advantage over SVM when the pixel-based reflectance samples used, without the segmentation size.

❖ SVM is better than KNN

The SVM is extremely fast, classifying 12-megapixel aerial images in roughly ten seconds as opposed to the kNN which takes anywhere from forty to fifty seconds to classify the same image.

❖ CNN is better than KNN

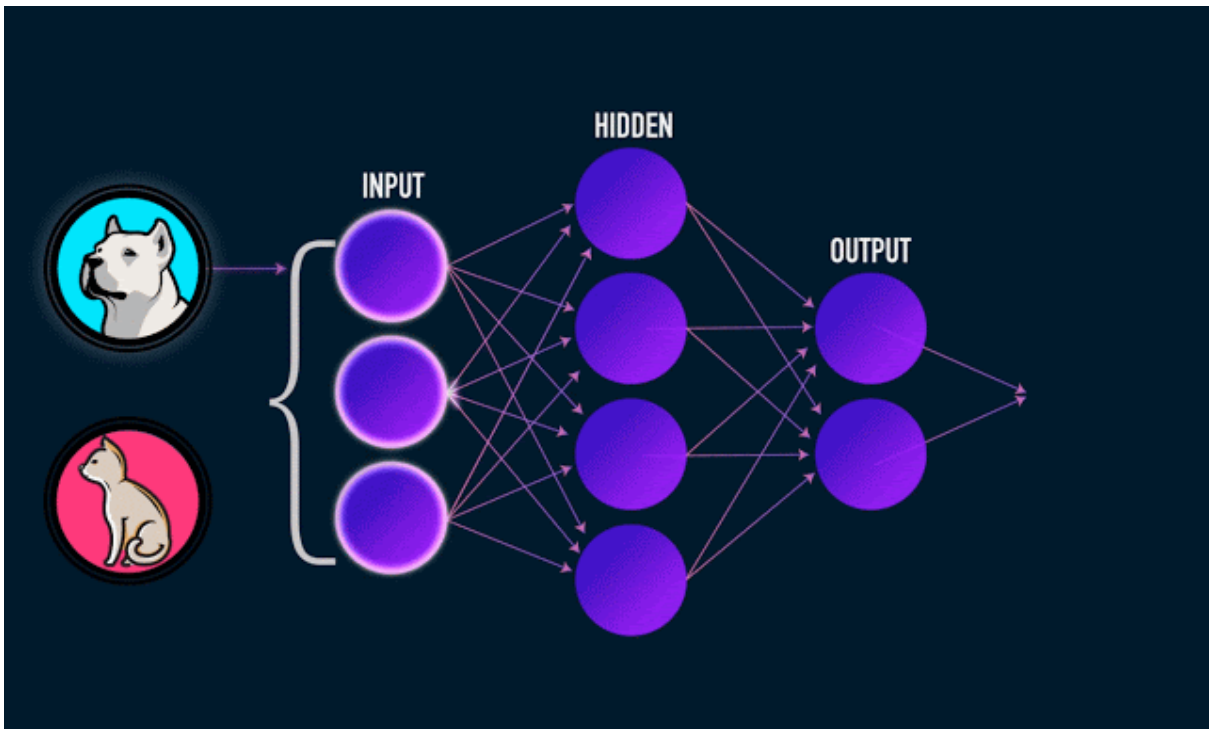
The accuracy of the CNN method is 94%, while the accuracy of the KNN(3) method is 89%. The F1-score value for the CNN method is 0.94 and the KNN(3) method is 0.89. The CNN allows the model to produce an average precision of 87.7%, the accuracy of 86.89%, recall of 86.89%, and F1-score of 86.33%.

Conclusion – CNN is better than both SVM and KNN (CNN >> SVM >> KNN).

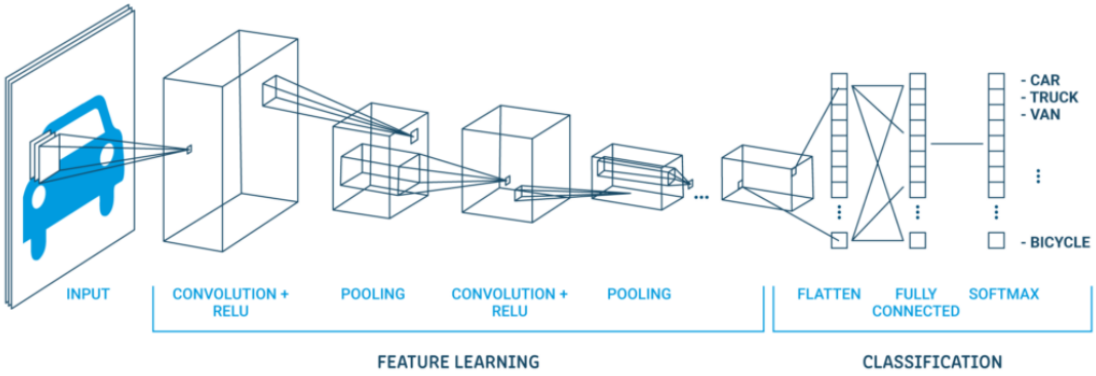
Implementation of CNN Algorithm

The Convolutional Neural Network (CNN or ConvNet) is a subtype of Neural Networks that is mainly used for applications in image and speech recognition. Its built-in convolutional layer reduces the high dimensionality of images without losing its information. That is why CNNs are especially suited for this use case.

Image classification is the task of taking an input image and outputting a class or a probability of classes that best describes the image. In CNN, we take an image as an input, assign importance to its various aspects/features in the image and be able to differentiate one from another. The pre-processing required in CNN is much lesser as compared to other classification algorithms.



A classical CNN architecture looks like as shown below:



Convolutional Neural Networks specialized for applications in image & video recognition. CNN is mainly used in image analysis tasks like Image recognition, Object detection & Segmentation.

There are three types of layers in Convolutional Neural Networks:

1) **Convolutional Layer:** In a typical neural network each input neuron is connected to the next hidden layer. In CNN, only a small region of the input layer neurons connect to the neuron hidden layer.

2) **Pooling Layer:** The pooling layer is used to reduce the dimensionality of the feature map. There will be multiple activation & pooling layers inside the hidden layer of the CNN.

3) **Fully-Connected layer:** Fully Connected Layers form the last few layers in the network. The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer.