Pose & Posture Estimation (Yoga Asanas)

Machine Intelligence

BACHELOR OF TECHNOLOGY
Department of Computer Science & Engineering

V Semester Section A

SUBMITTED BY

Batch No: 3

Student name 1: A MOHAN KUMAR
SRN: PES2UG20CS002

Student name 2: ABISHEK S
SRN: PES2UG20CS014

Student name 3: ADITHYA MATHIALAGAN
SRN: PES2UG20CS019

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)
100 Feet Ring Road, BSK III Stage, Bengaluru-560085

PAPER 1: Computer Vision Technology Based on Deep Learning

Published in: 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)

Date: 19 December 2021

Write about the research work

Jin Qiu, Jian Liu, Yunyi Shen, College of Weapon Engineering, Naval University of Engineering, Wuhan, China

With the development of artificial intelligence, computer vision technology that simulates human vision has received widespread attention. Based on the current commonly used method of computer vision technology-deep learning, this paper outlines the development of deep learning models, and determines the inflection point of the development of the introduction of convolutional neural networks. Not only that, but convolutional neural networks also combined with deep learning have developed rapidly and are widely used in the field of computer vision. Therefore, the structure and development process of the convolutional neural network are further analysed. Subsequently, the three major difficulties in the field of computer vision-image classification, object detection, and image segmentation are compared and analysed, and their differences and connection analysis are drawn. Finally, an overview of the difficulties to be optimized in computer vision, and computer vision technology It will continue to develop as a research hotspot

Techniques used

The research paper briefs about Deep learning model - Convolutional neural network.

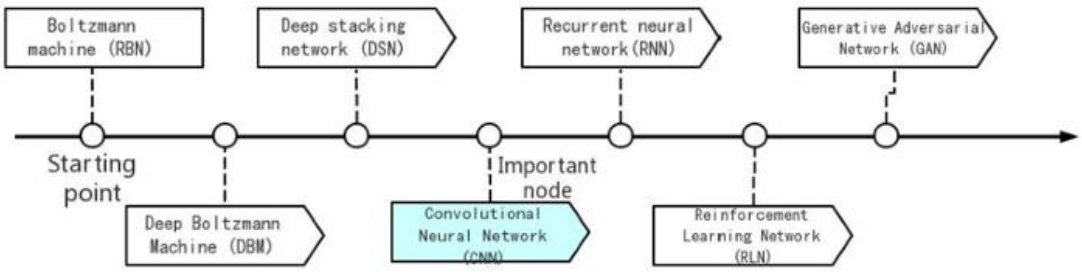


Fig. 1. The development history of deep learning models

.

A CNN comprises three main types of neural layers, namely

(i) Convolutional layers - Utilizes various kernels to convolve the whole image as well as the intermediate feature maps, generating various feature maps.

(ii) Pooling layers- Pooling layers oversee reducing the spatial dimensions (width × height) of the input volume for the next convolutional layer

(iii) Fully connected layers- Following several convolutional and pooling layers, the highlevel reasoning in the neural network is performed via fully connected layers. Neurons in a fully connected layer have full connections to all activation in the previous layer.

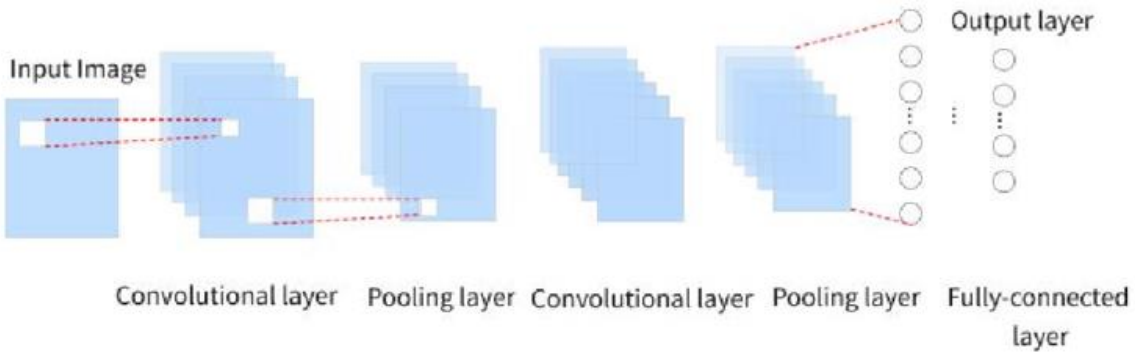


Fig. 3.CNN structure

Limitations

The difficulty of computer vision is how to improve the semantic understanding of input data and establish the mapping from pixels to semantics. Comparing the degree to which the human eye′s ability to judge objects is affected by external factors, computer vision still needs to be refined to reduce these interferences in the development of objects. From the perspective of object recognition, classification, detection, and segmentation, the following issues are all computers. The longstanding difficulties of vision:

Conclusion

In general, the purpose of computer vision technology is to allow machines to simulate human vision and distinguish objects autonomously. It is a medium for machines to communicate with the outside world. As an important research direction in the development of artificial intelligence and unmanned technology, computer vision technology has been developed and applied to many fields: face recognition, unmanned driving, medical image recognition, voice recognition and other fields have its presence

PAPER 2: A Study on Various Classifier Techniques Used in Image Processing

Published in: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)

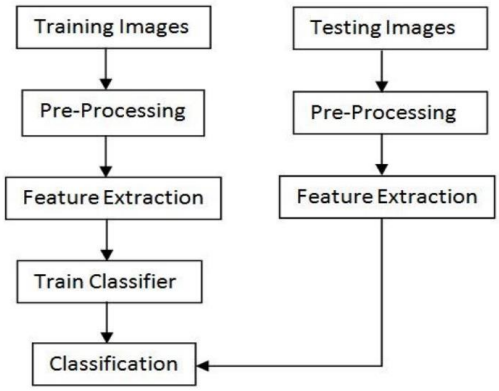
Date: 15 May 2020

Write about the research work

Arun C, Akshatha Prabhu, Mohammed Zeeshan Department of Computer Science, Amrita School of Arts and Sciences, Mysuru, India

This classification using machine techniques are essential in supermarkets essentially at the point of sale (POS). Various works have been carried out in classification methods and in recent year's classifiers such as support vector machine, K nearest neighbour, artificial and convolution neural networks have been trending. This paper presents the state of art methods in fruit classification proposed by various researchers that also hold for applications such as fruit quality assessment and harvesting.

Steps Involved



Techniques used

In this paper, they have proposed a strategy to classify fruits and vegetables using shape and region-based image classification.

For the proposed work they have gathered around 350 pictures for the dataset. Segmentation strategies are applied to the pictures. The strategies utilized are Watershed transform and Sobel edge detection methods. For deciding the exhibition of classifiers, they have taken **KNN**, **SVM**, as the classifier models.

Dataset gathered for CNN work is around 3600 pictures with 200 pictures for each organic fruit picture. A four-stage pre-processing technique is utilized here.

In this work, the creators structured a **13-layer convolution Neural Networks** for classifying fruit category based on images. CNN is utilized here for ordering the natural product classification.

The classifiers utilized are KNN, SVM, and CNN. The precision accomplished by every classifier is 48.63%, 60.5%, and 96.49% individually.

Conclusion

There are two stages in the classification process: training and testing.

In training properties of the picture features are detached, given these; a one-of-a-kind depiction of every order classification an instructional course is made.

In the testing stage, these highlights space portions are used to arrange picture highlights. In this examination, different arrangement techniques in picture training are audited.

This paper mainly focuses on fruit classification using different classifiers and different stages in image processing like pre-processing, image segmentation and feature extraction, etc. most of the research work in classification carried out in traditional machine learning and deep learning methods. These classifiers gave better outcomes in characterization.

PAPER 3: Image Classification using Convolutional Neural Network

Published in: 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)

Date of Conference: 18 February 2022

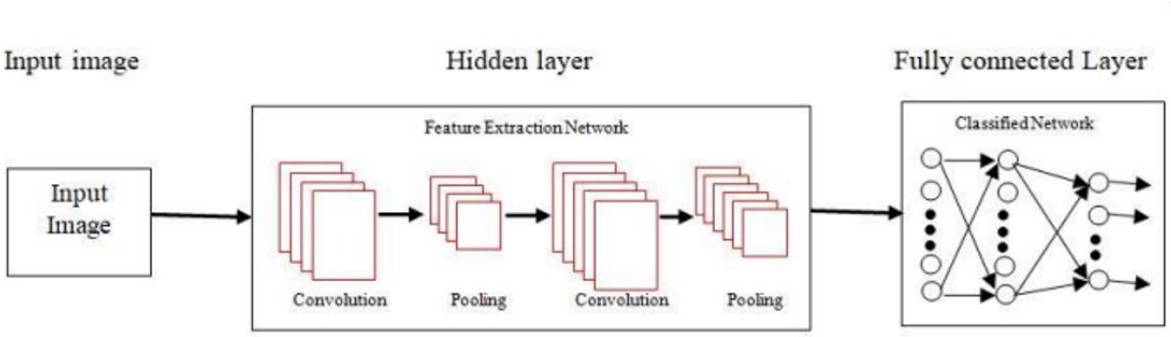
Write about the research work

G. Sathiyapriya, S. Anita Shanthi

Department of Mathematics Annamalai University Annamalai Nagar-608002 Tamilnadu, India

To enhance the performance of image classification deep features are combined with CNN and various feature extraction methods. CNN has an ANN architecture which classifies data like images, speech, and videos. In CNN simple structures of an image are detected accurately and complex features are constructed. It is used in better detection of images.

Techniques used



Procedure

The main function of classifying image is input image acceptance. The main goal is to classify the input image. It is obtained by training an ANN on thousand images of input image and enable the NN to predict the image class.

Step 1: Download the trained Dataset

Step 2: Choose three images of different categories from the downloaded data set. categories = {butterfly, flamingo, dolphin}.

Step 3: Load the dataset using 'ImageDatastore'.

Step 4: Count the number of images per category using countEachLabel'

Step 5: Determine the smallest number of images in the category

Step 6: Adjust each category to have same number of images

Step 7: Download the pretrained network, ResNet50, to visualize the first section of the network.

Step 8: Extract the training features in CNN. Weights for the convolution layers are applied to extract features of the image. $w1$ = net. Layer (2)

Step 9: Evaluate classification. Test features are passed to classify the accuracy of the trained features.

Step 10: Apply the trained classifier to one test image. Input the new image, of the above categories. It can extract image features using the CNN pixels and make a prediction of the classification.

Conclusion

In this article image classification using CNN is dealt with. Training features of the image are extracted, and weights are used to extract features of the image and categorization. Test features are passed to classify accuracy of trained features. The trained classifier is applied to one test image and a new image category can be found. **Through this paper we were able to understand how CNN works to solve Image Classification problems.**

PAPER 4: The Role of Activation Function in CNN

Published in: 2020 2nd International Conference on Information Technology and Computer Application (ITCA)

Date: 18-20 December 2020

Write about the research work

Wang Hao, Lou Yaqin, Wang Yizhou, Song Zhili, College of Computer Science and Information Engineering, Shanghai Institute of Technology ShangHai, China

The purpose of introducing activation function is to give neural network nonlinear expression ability, so that it can better fit the results, to improve the accuracy. However, different activation functions have different performance in different neural networks.

In this paper, several activation functions commonly used by researchers are compared one by one, and qualitative comparison results are given by combining with specific neural network models.

Techniques used

The Sigmoid function as a commonly used activation function in the basic neural network has a lot of room for use in logistic regression problems.

In the image classification task, since the network structure with a deeper number of layers is used, we usually use the ReLU function to improve the non-linear ability of the network and avoid the gradient disappearance or the gradient explosion after multiple iterations of the network. phenomenon.

The recent Mish activation function has good performance in many network structures, but it is not necessarily the best activation function.

In a multilayer neural network, there is a functional relationship between the output of the upper node and the input of the lower node. This function is called the activation function.

Characteristics:

The ideal activation function should have the following characteristics:

(1) Non-linearity: the derivative is not a constant. This can ensure that the multilayer network does not degenerate into a single-layer linear network.

(2) Differentiability: corresponds to the computability of the gradient in optimization.

(3) Simple: A complex activation function will reduce the calculation speed.

(4) Saturation: Saturation refers to the problem that the gradient is close to zero in certain intervals (that is, the gradient disappears), making it impossible to update the parameters.

(5) Monotonic: The sign of the derivative does not change. When the activation function is monotonic, the single-layer network can be guaranteed to be a convex function.

(6) Fewer parameters: Most activation functions have no parameters.

To compare the properties of several commonly used activation functions, we draw some of the images of the activation functions and analyse them.

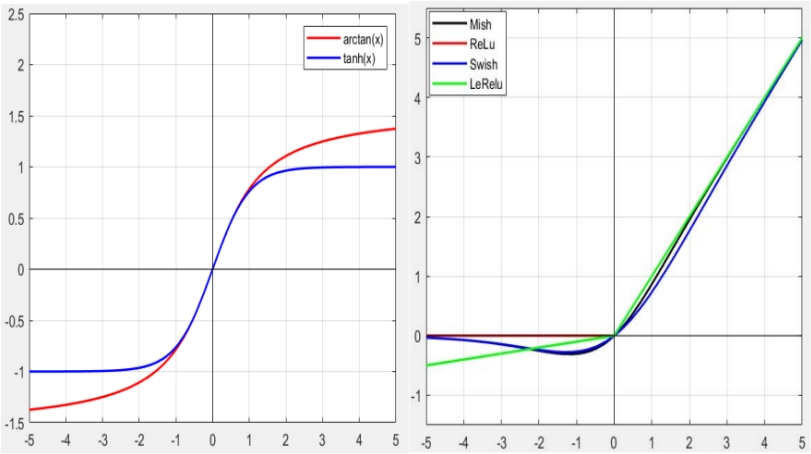


Figure 1. Function graphs of several activation functions

Conclusion

TABLE II. COMPARISON OF 8 ACTIVATION FUNCTIONS IN VGG16

epoch	ReLU	LReLU	Tanh	Swish	ELU	PReLU	Mish	Sigmoid
4000	80.387	79.879	73.871	82.087	79.260	80.273	81.790	23.505
8000	83.492	83.482	78.547	84.353	81.436	82.213	83.967	24.148
12000	86.328	86.413	80.795	86.636	84.266	86.057	86.450	33.209
16000	86.945	86.471	80.961	86.770	84.488	86.682	86.513	34.098
20000	88.596	88.464	82.200	88.321	86.044	89.397	88.326	38.171
24000	88.854	88.718	82.858	88.918	86.305	89.568	88.670	45.378
28000	89.954	90.001	84.082	89.628	87.303	90.099	89.598	46.055
32000	90.029	91.071	84.151	89.640	87.345	90.162	89.642	48.307
36000	**90.226**	90.192	84.394	89.711	87.367	90.197	89.677	57.176

PAPER 5: Comparison of Performance by Activation Functions on Deep Image Prior

Published in: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)

Date of Conference: 13 February 2019

Write about the research work

Manisha Patel, Research scholar, E. C. Department, R.K.University, Rajkot, India.

Nilesh Kalani, Director IQAC, R.K. University, Rajkot, India

In this paper, we compare the performance of activation functions on a deep image prior. The activation functions considered here are the standard rectified linear unit (ReLU), leaky rectified linear unit (Leaky ReLU), and the randomized leaky rectified linear unit (RReLU). We use these functions for denoising, super-resolution, and inpainting of the deep image prior. Our aim is to observe the effect of differences in the activation functions.

Techniques used

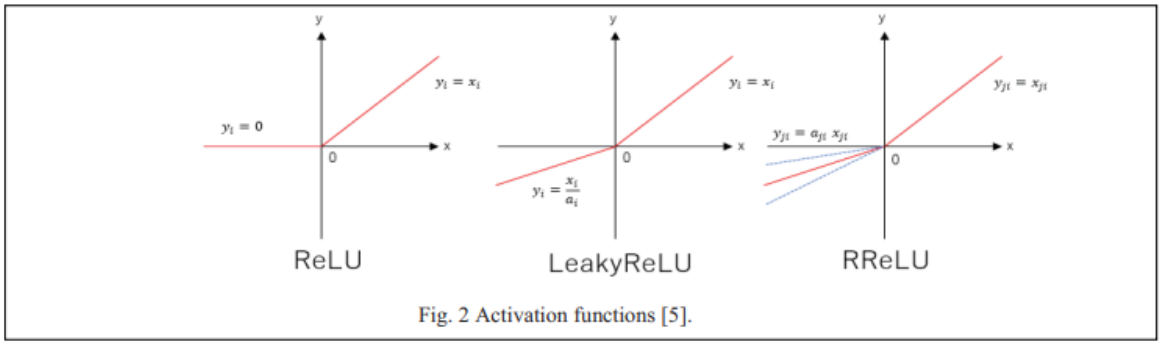


Fig. 2 Activation functions [5].

The Leaky ReLU is defined as

$$y_i = \begin{cases} x_i \ (x_i > 0) \\ \frac{x_i}{a_i} \ (x_i \leq 0) \end{cases}$$

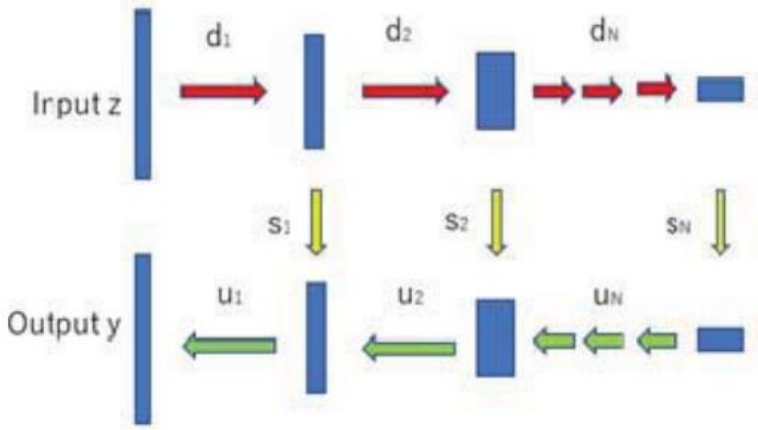
The RReLU is defined as

$$y_{ji} = \begin{cases} x_{ji} \ (x_{ji} > 0) \\ a_{ji} x_{ji} \ (x_{ji} \leq 0) \end{cases}$$

The ReLU is defined as

$$y_i = \begin{cases} x_i \ (x_i > 0) \\ 0 \ (x_i \leq 0) \end{cases}$$

Procedure:



Results:

Peak signal-to-noise ratio (PSNR) is the ratio between the maximum possible power of an image and the power of corrupting noise that affects the quality of its representation.

The higher the value of PSNR, the better will be the quality of the output image.

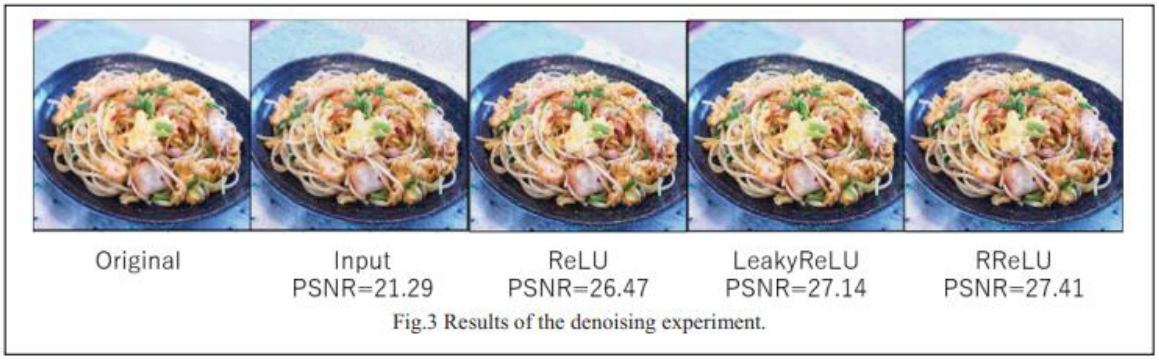


Fig.3 Results of the denoising experiment.

Conclusion:

This paper compares the performance of 3 activation functions with a deep image prior. The RRelU had the best performance in the denoising and inpainting experiments, but in the ReLU performed best in the super-resolution experiment.

PAPER 6: Hyper Parameter Optimization of Convolutional Neural Networks

Published in: 2021 International Conference on Advances in Computing and Communications (ICACC)

Date: 21 October 2021

Write about the research work

Sreekala KK, Jayakrushna Sahoo, Department of Computer Science and Engineering, Indian Institute of Information Technology Kottayam, Kerala, India.

The Convolutional Neural Networks (CNN) have recently become the most effective way of classifying images. Hyper-parameter tuning of convolutional neural network is performed by using two efficient optimization techniques as grid search and randomized search technique. The experiment is conducted by using the INbreast dataset. The results demonstrated reveals that the hyper-parameter tuning/optimization techniques along with normalization and standardization techniques is the most effective technique for the image classification.

Hyper Parameter Tuning Parameters:

Hyper parameter	Parameter	Range
Filters size	Filter-1	[16, 32, 64,]
Kernel Size	Ksize-1	[4, 5, 6]
Filters size	Filter-2	[64, 96, 128]
Kernel Size	Ksize-2	[4, 5, 6]
Filters size	Filter-3	[96, 128, 164]
Kernel Size	Ksize-3	[4, 5, 6]
Hidden Layer	Full_hidden_1	[75, 105, 125]
Hidden Layer	Full_hidden_2	[75, 105, 125]
Activation	3 layer	[relu, Irelu, elu]
Learning rate	Learning rate	[0.001, 0.0001, 0.002, 0.003]
Batch Size	Batch size	[16, 32, 64, 128, 256]
Momentum	Momentum	[0.9, 0.95, 0.97, 0.99]
Number of epoch	Epoch(max)	30
Weight decay	Weight	0.01
Dropout layer	Dropout	0.2

Procedure:

1. Grid Search (GS)

Grid Search is the most common method for exploring space configuration for hyper-parameters.GS can be regarded as an exhaustive search or brute-force technique, which assesses all the arrangements of hyper-parameters provided in the configuration grid.

The following process should then be performed manually to identify the global optimums.

1. Start with a wide screen and phase scale.

2. Reduce the search space and phase size based on the previous good hyper-parameter performance.

3. Repeat step 2 several times to achieve the maximum. GS can be mounted and paralleled quickly.

2.Random Search (RS)

RS resembles GS, however RS randomly chooses a preset number of samples among upper and lower boundaries as hyper-parametric values, instead of evaluating all values in the search space, and then trains them until the given budget is exhausted.

The key benefit of RS, since each assessment is separate, is that it is easily paralleled and allocated resources.

Unlike GS, RS samples the defined distribution with a fixed number of parameter combinations that increase system efficiency by reducing the likelihood that it will waste time in a small, poorly performing region. Furthermore, RS will detect the global optimum or the near world optimum if enough budgets are available.

Performance parameters

The developed system is assessed using evaluation metrics such as TP, FP, TN, FN, sensitivity, precision, specificity, F-Score measure, and accuracy.

Results:

Finally, both GS as well as RS have a major drawback that all assessments in their iterations are independent of previous assessments. Thus, they take a huge amount of time to assess poorly performing search areas.

Methodology	Precession (%)	Sensitivity (%)	Specificity (%)	Recall (%)	F-measure (%)	Accuracy (%)
CNN	97.46	95.85	95.96	95.32	95.21	96.01
GS-CNN	96.58	96.99	96.00	95.59	96.44	97.32
RS-CNN	97.98	97.01	96.99	97.22	97.87	98.01
GS-RS-CNN	98.52	98.78	97.64	96.09	98.21	98.66

Grid Search - (GS)

Random Search - (RS)

PAPER 7: Implementation of Machine Learning Technique for Identification of Yoga Poses

Published in: 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)

Date: 12 April 2020

Write about the research work

Yash Agrawal*, Yash Shah*, Abhishek Sharma† *†Department of Electronics and Communication Engineering the LNM Institute of Information Technology, Jaipur, Rajasthan, India

It has been observed that pose detection techniques can be used to identify the postures and to assist the people to perform yoga more accurately. Recognition of posture is a challenging task due to the lack availability of dataset and to detect posture on real-time bases. To overcome this problem a large dataset has been created which contain at least 5500 images of ten different yoga pose.

Techniques used

Algorithm draws a skeleton of a human body on the real-time bases. Angles of the joints in the human body are extracted using the Tf-pose skeleton and used them as a feature to implement various machine learning models.

80% of the dataset has been used for training purpose and 20% of the dataset has been used for testing. This dataset is tested on different Machine learning classification models and achieves an accuracy of 99.04% by using a Random Forest Classifier.

Performance parameters

Feature Extraction:

Here's an equation:

$$a^2=b^2+c^2-2bccosA$$

Where,

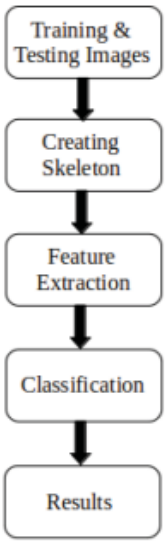
a = Distance between point p1 and p2
b = Distance between point p2 and p3
c = Distance between point p1 and p3
A = Angle made by point p2

To find the distance between two points [14]

$$a = \sqrt{(x1 - x2)2 + (y2 - y1)2}$$

Where,

(x1,y1) is the coordinate of point p1 (x2,y2) is the coordinate of point p2



: Flowchart of the procedure followed on the dataset.

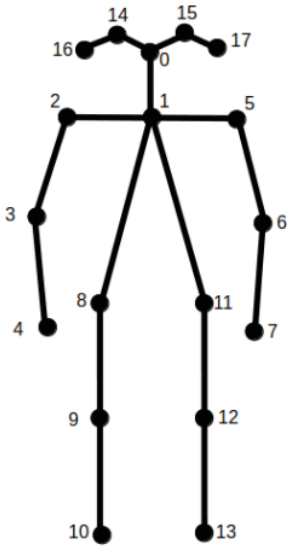


Fig 4: Stick Diagram using tf-pose Algorithm

Conclusion:

In this paper, a system is suggested that classify ten yoga poses and the dataset upholds on six classification models of machine learning. The yoga pose is detected based on the angles extracted from the Skeleton joints of TF pose estimation algorithm. 94.28% accuracy altogether was attained of all machine learning models.

Random Forest	n estimator: 30, MaxDepth: 7	0.9926
	n estimator: 30, MaxDepth: 10	0.9972
	n estimator: 30, MaxDepth: None	0.9990
SVM	Kernal Function: Linear, Loss Function: Hinge	0.8791
	Kernal Function: Polynomial, Loss Function: Hinge	0.9358
	Kernal Function: Radial Basis Function, Loss Function: Hinge	0.9871

Classifier	***Description***	***Accuracy***
Logistic Regression	IterationNumber: 1000, Solver: Newton-cg	0.8215
	IterationNumber: 1500, Solver: Newton-cg	0.8302
	IterationNumber: 2000, Solver: Newton-cg	0.8379
	IterationNumber: 2500, Solver: Newton-cg	0.8316

KNN	Neighbours: 3, DistanceWeight: Equal, Distance: Euclidean	0.9899
	Neighbours: 3, DistanceWeight: Inverse, Distance: Euclidean	0.9826
	Neighbours: 5, DistanceWeight: Equal, Distance: Euclidean	0.9853
	Neighbours: 5, DistanceWeight: Inverse, Distance: Euclidean	0.9901
	Neighbours: 7, DistanceWeight: Equal, Distance: Euclidean	0.9826
	Neighbours: 7, DistanceWeight: Inverse, Distance: Euclidean	0.9890
	Neighbours: 9, DistanceWeight: Equal, Distance: Euclidean	0.9725
	Neighbours: 9, DistanceWeight: Inverse, Distance: Euclidean	0.9891

Decision Tree	MinSampleLeaf: 1, MinSampleSplit: 2	0.9771
	MinSampleLeaf: 2, MinSampleSplit: 2	0.9670
	MinSampleLeaf: 3, MinSampleSplit: 2	0.9679
	MinSampleLeaf: 1, MinSampleSplit: 3	0.9752
	MinSampleLeaf: 2, MinSampleSplit: 3	0.9670
	MinSampleLeaf: 3, MinSampleSplit: 3	0.9761
Naive Bayes	Distribution: Normal	0.7475

PAPER 8: Deep Learning Based Yoga Pose Classification

Published in: **2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)**

DATE: 15 August 2022

Write about the research work

Shakti Kinger, Abhishek Desai, Sarvarth Patil, Hrishikesh Sinalkar, Nachiket Deore SCET, Dr.Vishwanath Karad MITWPU, Pune, Maharashtra, India

Pose identification strategies are effective in identifying posture and assisting people in performing yoga more effectively. Due to the paucity of data and real-time positioning, position detection is a difficult task. In this paper, we have done an extensive literature survey and offer a method for real-time pose estimation that uses the deep neural network model to detect and fix errors in a person's stance.

Techniques used

One of the deep learning techniques, named Transfer Learning, aims at training a neural network model on an issue that is identical to that being solved.
CNN is the most used deep neural network for image analysis:

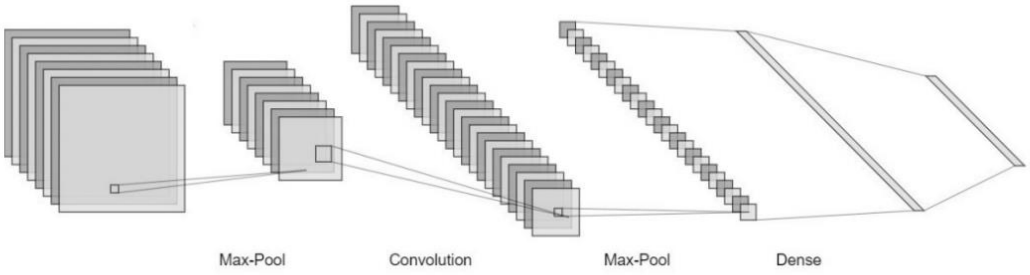
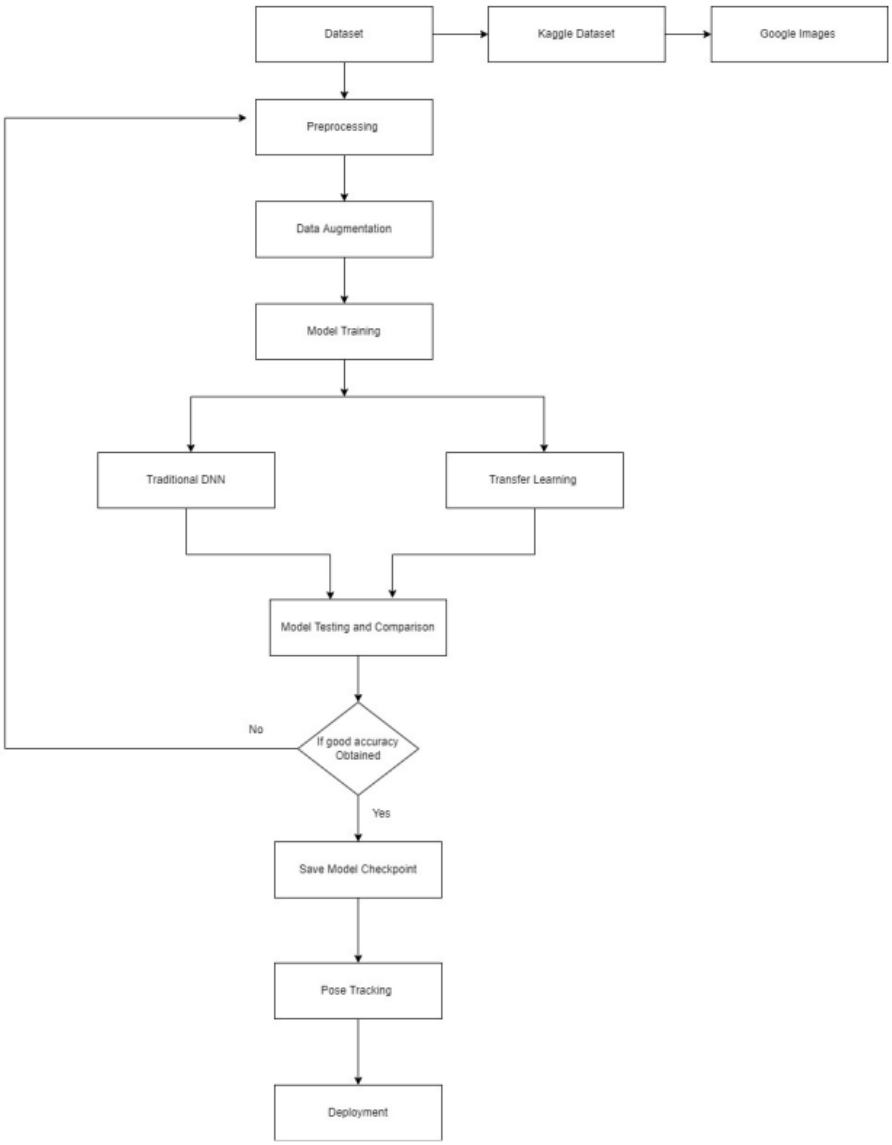


Fig. 7. CNN Structure

The general structure of CNN There are 4 main layers of CNN:
Convolutional Layer
Activation Layer
Pooling Layer
Fully Linked Layer

The first layer is a filter that operates on the matrix of filters at a time as it travels across a picture. The next layer that is the activation layer is a matrix created by the convolution layer with a lower estimate than the first one. The third layer which is the pooling layer implements the pooling strategy to encourage down scalping and thus lower's the frameworks' complexity. The output is expressed as a one-directional vector as its input. The classification layer is based on the label with the greatest chance of acceptance.

Steps Involved:



Conclusion:

Deep learning methods have been employed for pose detection. Pose estimation techniques like tf-pose-estimation, PoseNet, and OpenPose can be used. There are various challenges to be solved in the field of Pose Classification. One of them is to provide individual feedback to the user performing the pose. The methodology used in this study is based on extensive learning and then use it to classify yoga posture.

PAPER 9:

INFINITY YOGA TUTOR: YOGA POSTURE DETECTION AND CORRECTION SYSTEM

Published in: 2020 5th International Conference on Information Technology Research (ICITR)

DATE: 04 December 2020

Fazil Rishan, Binali De Silva, Sasmini Alawathugoda, Shakeel Nijabdeen, Lakmal Rupasinghe, Chethana Liyanapathirana
Department of Software Engineering Sri Lanka Institute of Information Technology Malabe, Sri Lanka

The Infinity Yoga Tutor application can capture user movements using the mobile camera, which is then streamed at a resolution of 1280 x 720 at 30 frames per second to the detection system. The system consists of two main modules, a pose estimation module which uses OpenPose to identify 25 keypoints in the human body, using the BODY_25 dataset, and a pose detection module which consists of a Deep Learning model, that uses time-distributed Convolutional Neural Networks, Long Short-Term Memory and SoftMax regression to analyze and predict user pose or asana using a sequence of frames.
This module was trained to classify 6 different asanas and the selected model which uses OpenPose for pose estimation has an accuracy of 99.91%. Finally, the system notifies the users on their performance visually in the user interface of the Mobile application.

Techniques used

The proposed Yoga Posture Detection and Correction System mainly consists of four components. They are:

• Key points Detection using OpenPose

• Key points Detection using Mask RCNN

• Higher Probability Prediction & Comparison

• Android Trainer Application

Results:

Multiple models were trained with keypoints data obtained from OpenPose and Mask RCNN pose estimation modules to find the optimum module for this system, and it was observed that the model trained with keypoints obtained from OpenPose was able to perform well overall with the least amount of delay when tested with real-time feed.

It was found that the model had a difficulty differentiating Tadasana and Vrikshana at certain times, this could be because the movements leading to both the poses are almost similar.

As mentioned above the model was able to predict with an accuracy of 99.87% for train data and 99.91% on unseen test data. Although individual models created for head, torso and legs had impressive accuracy values, the time taken to predict was higher than the other models.

Conclusion:

This proposed system can identify yoga postures using an android mobile.

Prediction module which consists of a time-distributed CNN layer that extracts spatial features and a LSTM layer which identifies spatial changes with temporal changes. The output from above layers is passed to a dense layer with an activation of SoftMax which gives a probability prediction for each class or pose. The selected model which uses OpenPose to detect keypoints, achieved an accuracy of 99.87% for dataset used for training and 99.91% for unseen test data. The integrated model also performed exceptionally well when tested in real time for all poses except for Vrikshana which had a few false positives with Tadasana.