# Exercise Pose Prediction using Convolutional Neural Network (CNN) and Residual Networks (Resnet)

Priyanka P
*Department of Information Technology*
*Kumaraguru College of Technology*
Coimbatore, India
priyanka.18it@kct.ac.in

Guruvishnu M
*Department of Information Technology*
*Kumaraguru College of Technology*
Coimbatore, India
guruvishnu.18it@kct.ac.in

Madhumitha S
*Department of Information Technology*
*Kumaraguru College of Technology*
Coimbatore, India
madhumitha.18it@kct.ac.in

Kavitha S
Assistant Professor
*Department of Information Technology*
*Kumaraguru College of Technology*
Coimbatore, India
kavitha.s.it@kct.ac.in

Saroja M N
*Assistant Professor*
*uGDX Institute of Technology*
Hyderabad, India
mnsaroja88@gmail.com

***Abstract*— Image classification is broadly used in almost all the fields. It can be used in medical, military, surveillance and many other fields. In this paper, we carried out image classification for four classes of exercise poses. Out of all available methods for image classification, we chose neural networks for classification. The CNN and ResNet50 algorithms were implemented and results were included in this paper. This can be applied in various exercise related fields like exercise monitoring, exercise posture correctness, virtual exercise training, etc.**

***Keywords*— *image classification, neural network, CNN, ResNet50***

## I. Introduction

We live in the age where data is supreme. Because the Internet of Things (IoT) and Artificial Intelligence (AI) have become ubiquitous platforms, we now have huge quantities of data. Speech, text, image, or a composite of these can be used to act for data. Images, either in the sort of photos or videos, make up an appreciable aggregate of global data production. Image classification is a complicated process that relies on a number of factors. Image classification is the process of identifying and labelling pixels in an image or matrices within an image using particular rules. The capacity of machines to discover hidden knowledge from a dataset of organized and unorganized examples is used in machine learning for picture detection. Deep learning is the most widely used machine learning approach, and it involves a system with a huge number of hidden layers.

In this paper, we discuss exercise pose prediction for 4 different exercise poses such as squats, lunges, plank, glute bridge. We have predicted the exercise poses by the use of the Convolutional Neural Networks (CNN) and also by applying the Residual Network (Resnet 50) algorithm. This exercise prediction can have its applications in exercise recommendation systems, exercise monitoring systems, etc.

## II. Related Works

Mikołajczyk, A., et al (2018) [1] takes us to the style and texture shifting technologies that mix the information of a foremost image with the aspects of another to create new images of high recognition rate. This strategy is commonly used to combine artistic styles in order to generate new aesthetically appealing artworks. This approach can be used in a variety of sectors, particularly in photographs from the same class but with somewhat different types for example, skin moles with malignant/benign types. In this work, they made use of the VGG16 and Convolutional Neural Network algorithms.

Wang, D., et al (2019) [2], The properties of the network structure used to reduce reliance on dataset quantity and increase adaptability to differences between datasets. A number of network models with useful reusable properties were chosen and trained. The pre-training prototype was used to obtain transfer learning for the structure after a comparative analysis of the collected data network architectures. Each network model's properties were also modified to meet this objective. The effectiveness of the neural networks, as well as the influence of changes between datasets, were assessed using two publicly accessible OCT datasets. Five neural network topologies were tested and analyzed, as well as three neural networks and SVM algorithms to better empirical analyses. These properties of feature reuse have been shown to boost the performance of neural network models and their flexibility to variations between datasets.

Lee, H., et al (2017) [3], a unique deep convolutional neural network (CNN) for hyperspectral image categorization that is deeper and wider than other available deep networks was used. A multi-scale convolutional bit stream is used as an early element of the proposed CNN pathway to leverage the spatio-spectral data jointly. The suggested technique surpasses the existing state-of-the-art in categorization, according to the findings. The study described is among the paramount to triumphantly employ an extremely deep fully convolutional neural network for hyperspectral categorization. This has two advantages in terms of testing the proposed network: i) using this division to assess the system network can verify their participation, which is building a broader network with a limited number of training samples, and ii) using this division to evaluate the proposed network can provide solid results of fairly decent benchmarks, such as RBF-SVM and the shallower CNN.

Han, D., et al (2018) [4], CNN transfer learning and web data enrichment were combined in a novel two-phase technique. The suggested methodology will be an excellent tool for addressing practice issues involving the use of deep CNNs on short datasets. For its high speed and durability, they choose Caffe to create these networks. Transfer learning eliminates the need for a huge amount of training samples and allows for the creation of a strong and resilient predictor. When the training data is restricted, it is quite useful; additionally, the Bayesian optimization approach was successfully used to modify the hyper-parameters for system fine-tuning. The augmentation process was time-consuming, which is a flaw.

Pan, B., et al (2017) [5], proposed a new deep-learning approach that relies on the rolling guidance filter (RGF) and the vertex component analysis network (R-VCANet), which can exhibit positive prediction efficiency with very few learning data than existing deep-learning-based techniques. Although IFRF and E-ICA-RGF perform similarly to R-VCANet, there is still a 1.5 percent difference. All of the other evaluated approaches, with the exception of R-VCANet, have exhibited less than 85% accuracy in one or more classes. The Hughes effect may be caused by the enormous number of spectral bands in HSI, as well as the high dimensionality of HSI. As a result, employing spectral fingerprints explicitly for HSI categorization may not be appropriate. Due to the complicated properties of HSI data, classification remains a challenge.

Zhao, W., et al (2016) [6] offers a spectral–spatial feature-based classification (SSFC) structure that extracts spectral and spatial features using a combination of feature reduction and deep learning algorithms. Meanwhile, a convolutional neural network is used to detect spatial-related properties at increased rates autonomously. Then, by piling spectral and spatial characteristics together, the fusion feature extraction is done. Finally, the picture categorization is trained using the multiple-feature-based classifier. Urban planning, environmental management, agricultural assessment, and mineral identification all use hyperspectral images (HSIs). With a limited number of training examples, these applications frequently need the detection of the category of each pixel. When working with pictures with such great spectral/spatial resolution, however, greater interpretation exactness is difficult to achieve.

Wei, Y., et al (2015) [7], Hypotheses-CNN-Pooling (HCP) is an adaptable deep CNN architecture that takes an infinite amount of object sequence hypotheses as input information, connects each hypothesis to a mutual CNN, and then aggregates the CNN yield findings from various hypotheses with max pooling to construct the utmost multi-label assumptions. Nevertheless, because the bulk of real-world photos comprise elements from various segments, multi-label image categorization is a more comprehensive and realistic challenge. Variable kinds of elements are situated at multiple roles with various levels and attitudes in a typical multi-label photograph. Furthermore, the various compositions and interactions between items in multi-label photos, such as partial visibility and occlusion, add to the problem's complexity, necessitating more labeled data to cover the various scenarios.

Mahajan, A., et al (2019) [8] categorizes hundreds of hyperspectral photos into eight unique classifications, the system employs a deeper convolutional neural network. We used picture features taken from a pre-educated Representational deep Neural network (RESNET) to educate machine learning algorithms. A classifier that uses a support vector machine (SVM). Thousands of photos are required to train categorical image classification. In addition, the system requires more time to extract characteristics and classify them.

## III. CONVOLUTIONAL NEURAL NETWORK

A convolutional neural network (CNN, or ConvNet) is a type of Artificial Neural Network (ANN) which is employed in deep learning to analyze visual representations. Only a few of the technologies include visual data recognition, recommendation systems, image categorization, image segmentation, brain-computer interactions, natural language processing, medical image analysis, and economic time - series data. CNNs demand far too little pre-processing in comparison to certain other picture categorization algorithms. This means that, in comparison to conventional methodologies, the network evolves to optimize the filters (or kernels) through automated learning. The fact that feature extraction does not rely on past knowledge or human interaction is a key advantage.

## IV. RESIDUAL NETWORK

ResNet-50 is a deep convolutional neural network with 50 layers. You may run a pre-trained version of the structure on the Imagenet dataset, which has been prepared on over a million images. The technology can categorise images into 1000 different classification tasks, such as keyboards, mouse, pencils, and other animals. As an outcome, the structure has gained knowledge of a number of ample feature representations for a wide range of images. The image input size for the system is 224 224 pixels.

## V. PROPOSED SYSTEM

### *A. Dataset*

Instead of working with the existing dataset, we have manually created a dataset with a total of 160 images. We have four different classes viz, squat, lunge, plank, glute bridge. Each class accounts for 40 images.

### *B. Data Preprocessing*

The data preprocessing consists of the following steps:

1. Since our data is categorical, one hot encoder is used to convert data into integer values which in turn is represented in binary vectors. In our work, 0 acts for Squat, 1 acts for Plank, 2 indicates Lunge and 3 indicates Glute Bridge.
2. All the images are resized into 224 x 224 pixels.

### *C. Train-Test Split*

A total of 80% of the images are chosen for the training sample, while the rest 20% are chosen for the testing data.

### *D. CNN*

1. Add the initial layer (Convolution 2D): In the convolution 2*2 filter matrix, we utilize 32 output filters that combine to the input RGB size picture 32*32 and apply activation Relu.
2. Layer of Batch Normalization is added. It is the method of assigning more layers to a deep neural network in order to make it faster and more reliable. On the intake of a former layer, the new layer performs the standardization and normalization methods.
3. Apply (MaxPooling2D), Processing, Hidden Layer 1 to all the images. Step 1 and 2 are repeated again.
4. The matrix is flattened into a single array after using flattening.
5. Establishing a full connection - 128 final output layers, activation=Relu & Dense layer, SoftMax activation.

### *E. ResNet50*

1. The pretrained model of ResNet is executed with ImageNet as weights.
2. The average pooling was chosen. Average Pooling is a down sampling process for feature maps that estimates the overall average for regions and uses it to create a pooled feature map.
3. Flattening the matrix makes it into a linear array.
4. Establishing a full connection - 128 final output layers, activation=Relu & Dense layer, SoftMax activation

TABLE I. COMPARISON

| Metrics | *CNN* | *ResNet50* |
|---|---|---|
| Average accuracy (in %) | 67 | 89 |
| Average loss (in %) | 12 | 28 |
| No. of errors | >25 | 7 |

## VI. CONCLUSION

Out of 160 images from the dataset, a sample of 2 images, each from different classes, were displayed for the result of both Convolutional Neural Network (CNN) and Residual Neural Network (ResNet 50).

### *A. CNN*


Fig. 1. Glute Bridge Sample Result

### *B. ResNet50*

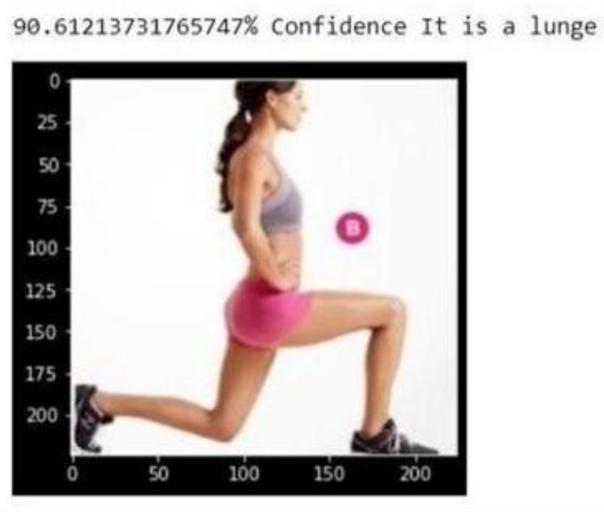


Fig. 2. Lunge Sample Result

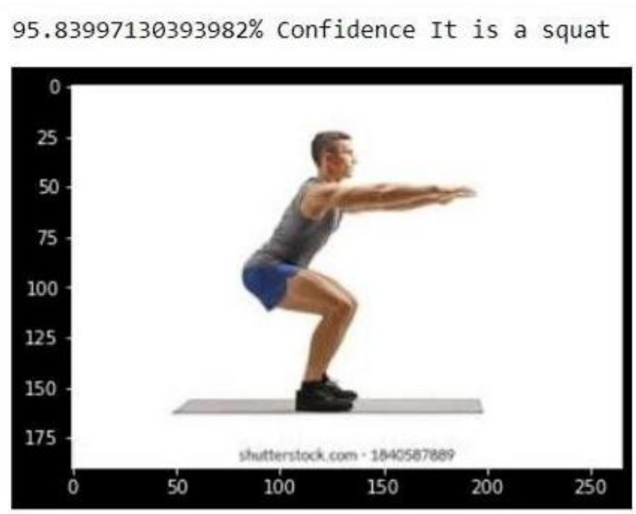


Fig. 3. Squat Sample Result

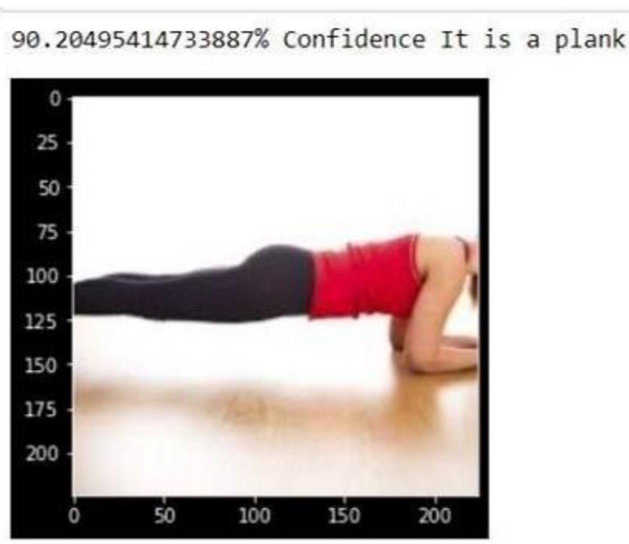


Fig. 4. Plank Sample Result

## VII. FUTURE SCOPE

The image classification carried out in this paper can be extended to video classification. The video classification of exercise poses can be used in various applications like exercise monitoring and exercise recommendation systems, virtual trainers and so on. The problem in the available applications is that those can be utilized only for the specified exercise practice. By extending the number of exercise classes used, the user can have all the exercise poses cumulatively at one place.

## REFERENCES

[1] Mikołajczyk, A., & Grochowski, M. (2018, May). Data augmentation for improving deep learning in image classification problem. In 2018 international interdisciplinary PhD workshop (IIPhDW) (pp. 117-122). IEEE.

[2] Wang, D., & Wang, L. (2019). On OCT image classification via deep learning. IEEE Photonics Journal, 11(5), 1-14.

[3] Lee, H., & Kwon, H. (2017). Going deeper with contextual CNN for hyperspectral image classification. IEEE Transactions on Image Processing, 26(10), 4843-4855.

[4] Han, D., Liu, Q., & Fan, W. (2018). A new image classification method using CNN transfer learning and web data augmentation. Expert Systems with Applications, 95, 43-56.

[5] Pan, B., Shi, Z., & Xu, X. (2017). R-VCANet: A new deep-learning-based hyperspectral image classification method. IEEE Journal of selected topics in applied earth observations and remote sensing, 10(5), 1975-1986.

[6] Zhao, W., & Du, S. (2016). Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing, 54(8), 4544-4554.

[7] Wei, Y., Xia, W., Lin, M., Huang, J., Ni, B., Dong, J., ... & Yan, S. (2015). HCP: A flexible CNN framework for multi-label image classification. IEEE transactions on pattern analysis and machine intelligence, 38(9), 1901-1907.

[8] Mahajan, A., & Chaudhary, S. (2019, June). Categorical image classification based on representational deep network (RESNET). In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 327-330). IEEE.