# Yoga Pose Classification

Anand K S
*Amrita School Of Engineering,*
Amritapuri

Bharath Prathap Nair
*Amrita School Of Engineering,*
Amritapuri

Rahan Manoj
*Amrita School Of Engineering,*
Amritapuri

***Abstract*—The project aims to classify yoga poses effectively using Convolutional Neural Network (CNN) by learning from a collection of input images. In this project a comparison of various pretrained CNN models were performed and a detail analysis of EfficientNet architecture was also carried out. Different combinations of optimizer functions and loss functions were applied to the EfficientNet architecture in order to achieve the highest accuracy of 77.941 .The project also performs a comparison of different versions of EfficientNet architecture like EfficientNet-B0, EfficientNet-B1 and EfficientNet-B2 on the dataset with various hyperparameter tunings.**

## I. Introduction

Yoga is a mind, body and spiritual enhancement practice developed traditionally in India which is now practised all around the world due to its profound benefits.The word 'yoga' comes from the Sanskrit language which literally means 'union'. Yoga is a union of mind and body, man and nature.A yogasana is any action, pose or exercise that is performed with rhythmic breathing cycles to benefit both body and mind.There are more than 100 styles of yoga which includes Iyengar yoga, Hatha yoga as well as other styles rising in popularity such as Aerial Yoga and Hot Yoga. Asanas helps to smoothen the muscles, joints, ligaments and other parts of the body. This increases blood circulation and flexibility of our body. Today yoga has evolved so much that there are over 8,400,000 yoga poses. It can be noticed that many of these poses are too complex to be captured from a single point of view and is impossible for a human being to memorize and classify each of these yoga poses. Due to this, our project was developed so as to classify yoga poses more accurately by reducing the human error when perceived through the naked human eye.

This project aims to classify Yoga Poses effectively using pretrained CNN architectures in pytorch framework. In order to achieve this, CNN was performed on architectures such as ResNet, VGG, MobileNet and EfficientNet. On training the model with these architectures, highest accuracy was observed with EfficientNet architecture. In this project we compare the classification capabilty of different variants of EfficientNet accompanied with various hyperparameter tunings. To improve the accuracy values further, data augmentation was performed on the dataset. Data Augmentation involved rotating, blurring, flipping and other manipulating techniques to increase the size of the dataset so as to provide more information during training and hence to improve accuracy.


Fig. 1. A Sample of 10 Yoga Asanas

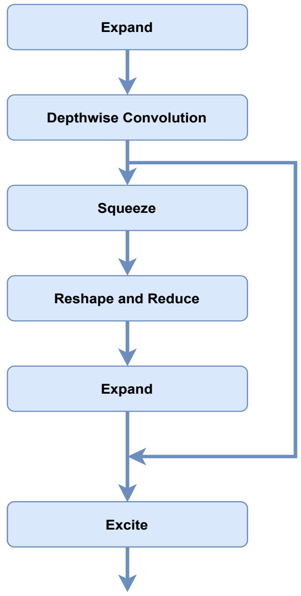


Fig. 2. Architecture of MBConv

## II. Related Works

A good amount of works have been done in the field of human pose classification and estimation. [4] has performed yoga pose estimation using keypoint detection methods such as OpenPose, PoseNet and PifPaf. It can also be found that the paper also focusses on Pose Classification using deep learninig models such as Multi-layer Perceptron(MLP), Recurrent Neural Networks(RNN) and Convolutional Neural Network(CNN). This classification was done only for 6 yoga poses/classes. [5] generated a new dataset named 'Yoga-82' and performed yoga pose classification using DenseNet Architecture. [6] performed a yoga pose detection for 6 poses using Adaboost classifier and Kinetic sensors with really good accuracies, with the help of a depth sensor based camera.

## III. Dataset

The model developed for the objective was trained using the 'Yoga Pose Image classification dataset' from Kaggle. Data augmentation was performed on the dataset at later stages.

The dataset contains 5991 images classified across 107 classes signifying 107 different yoga poses/asanas. A sample of some of these are shown in Figure 1, namely Svanasana, Padmasana, Bhujangasana, Utkata-konasanam, Marjarasanam, Trikonasana, Vriksasana, Uttanasana, Savasana and Tadasana.

## IV. Convolutional Neural Network

The Convolutional Neural Network, also known as ConvNet or CNN is a specialised deep learning model, designed to take in images as inputs and perform learning by differentiating elements of the image. The architecture of CNN is inspired by the visual cortex and can be compared to the connectivity pattern of neurons present in the human brain.A Convolutional Neural Network consists of multiple layers of artificial neurons. When an image is input into the CNN, each of these layers create activation functions that are passed on to the next layer. Each layer has its own functionality and gets complex as as it moves on. The CNN reduces the images into a form which makes it easier to process, without losing much of the critical features necessary for a good prediction.

The dataset was trained with different pretrained CNN models. The models that were considered are Resnet, VGG, MobileNet and EfficientNet. Out of these pretrained models, highest accuracy was observed while using EfficientNet architecture.

EfficientNet is a CNN based architecture introduced in 2019 by the team of Google AI. Unlike the past innovations in the field which stressed on increasing the depth, width or resolution of the model, efficientnet uses a principled/structural method to scale the parameters. Eventhough increasing individual dimensions improved performance, efficientnet used a much better method of balancing all dimensions of the model [depth, width, resolution] would lead to a better overall performance.

While scaling individual dimensions improves model performance, we observed that balancing all dimensions of the network—width, depth, and image resolution—against the available resources would best improve overall performance. This method is called compound scaling. There are different version of Efficientnet from B0 to B7, wherein the number of layers increase from B0 [237] to B7[813]. In this project, we consider EfficientNet-B0, EfficientNet-B1 and EfficientNet-B2. In each of these models, the basic building block of the architecture is mobile inverted bottleneck convolution (MBConv). Each variant of EfficientNet architecture vary with the variation in the number of MBConv blocks. The structures of MBConv and EfficientNet B0 where MBConv is the basic building block is shown in Figure 2 and 3 respectively.

## V. Data Pre-processing

A stratified split was applied on the dataset, where 80% of the data was used for training and the remaining 20% was used for validation. As the dataset consisted of images of different size, the images were resized to a uniform magnitude of 224*224 pixels in order to match the EfficientNet architecture's predefined input image size. For the network to constantly learn and update its weight, the dataset was split into batches of size 32.
In an attempt to increase the performance of our model, data augmentation was carried out. By performing data augmentation, we add slightly modified copies of already existing data or newly created synthetic data from existing data, thereby increasing the size of the dataset.

The augmentation techniques used are:

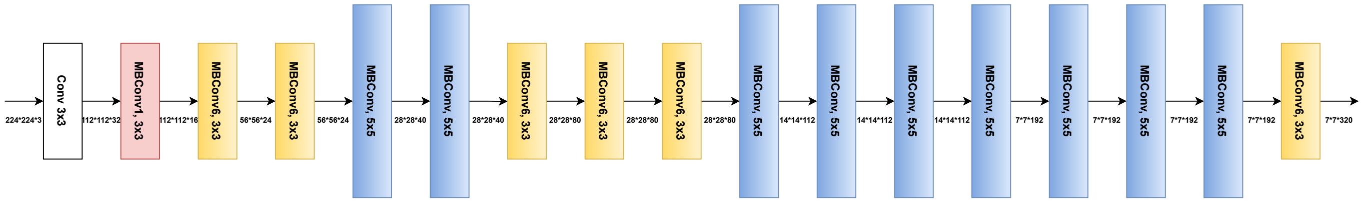


Fig. 3. Architecture of EfficientNet-B0

- Horizontal Flip : Mirrors the image around a vertical line running through its center.
- Adding Noise to Images : Noise was added by a factor of 0.01.
- Image Rotation : Rotated images by 30 degree.
- Blurring images : Blur was applied on the images by a factor 1.5 .
- Translation : Shifts the pixels of the image by specific amounts in X [20] and Y [10] directions.

## VI. Methods

The dataset was used to train the model with different pretrained CNN architectures. The accuracy values for these different architectures are tabulated below in Table 1. It can observed from the table, the highest accuracy is for EfficientNet architecture.

TABLE I
PERFORMANCE OF DIFFERENT CNN ARCHITECTURES

| Sl. No. | *Variant* | *Accuracy* |
|---|---|---|
| 1 | VGG | 66.89 |
| 2 | ResNet | 69.11 |
| 3 | MobileNet | 72.36 |
| 4 | EfficientNet | 75.40 |

Since the highest accuracy was observed in EfficientNet, now we further check for the best model among the different variant of EfficientNet architecture. Different combinations of optimiser functions and loss functions were tried out and their respective best accuracies are tabulated below in Table 2.

TABLE II
PERFORMANCE OF DIFFERENT VARIANTS OF EFFICIENTNET ARCHITECTURE

| Sl. No. | *Variant* | *Optimiser* | *Loss* | *Accuracy* |
|---|---|---|---|---|
| 1 | EfficientNet - B0 | Adadelta | NLL Loss | 77.94 |
| 2 | EfficientNet - B1 | Adagrad | Cross Entropy | 76.06 |
| 3 | EfficientNet - B2 | Adam | Cross Entropy | 77.14 |

From Table 2, it can be observed that EfficientNet B0 performs the best with Adadelta as the optimizer and Negative Log Likelihood as the loss function and provides an accuracy of 77.941. The accuracy plot for the same is plotted in Figure 4.

A sample of F1 score (weighted average of precision and recall) of 20 classes are also shown in Figure 6 .

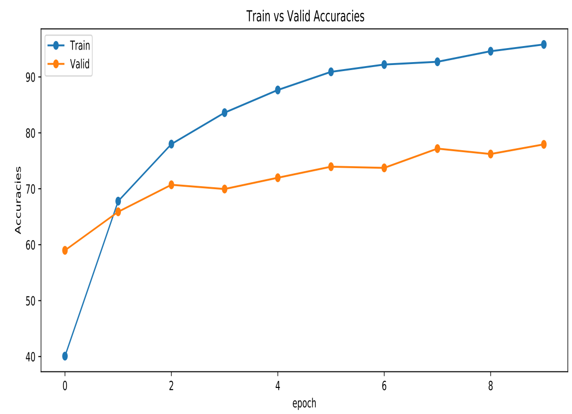


Fig. 4. Accuracy Plot for Adadelta-NLL Loss

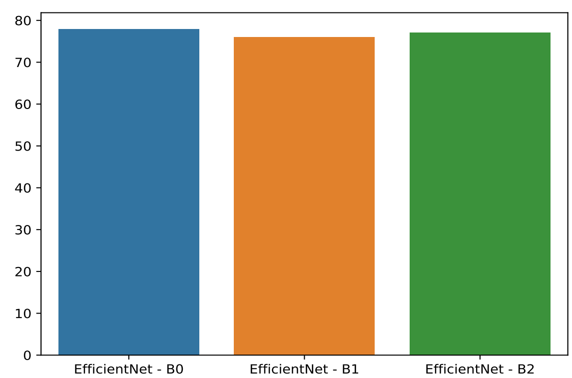


Fig. 5. Accuracy comparison of EfficientNet variants

## VII. Inference

From the initial results in Table 1 we found that EfficientNet architecture performs the best among the pretrained CNN models under consideration. Different variants of EfficientNet were tested and it was observed that EfficientNet-B0 performs the best among them with an accuracy of 77.941% for the given dataset with Adadelta as the optimiser and Negative Log Likelihood Loss as the loss function.

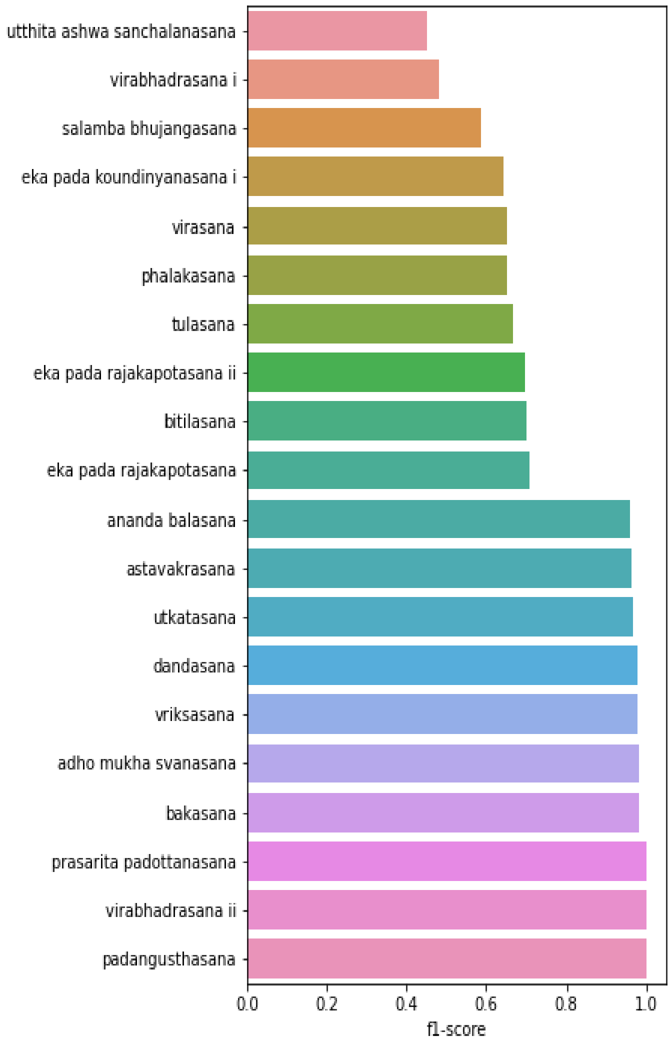


Fig. 6. F1 SCORE FOR TOP - 10 and BOTTOM-10 CLASSES

## REFERENCES


[1] Mingxing Tan, Quoc V. Le ,“EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, arXiv:1905.11946v5 [cs.LG], 11 Sep 2020

[2] Vardan Agarwal,“Complete Architectural Details of all EfficientNet Models”, Towards Data science

[3] Manisha Verma, Sudhakar Kumawat, Yuta Nakashima, Shanmuganathan Raman , “Yoga-82: A New Dataset for Fine-grained Classification of Human Poses” , Published on 22 Apr 2020

[4] Kothari, Shruti, ”Yoga Pose Classification Using Deep Learning” (2020). Master’s Projects. 932. DOI: https://doi.org/10.31979/etd.rkgu-pc9k,Included in Artificial Intelligence and Robotics Commons

[5] “Yoga-82: A New Dataset for Fine-grained Classification of Human Poses”, Manisha Verma, Sudhakar Kumawat, Yuta Nakashima, Shanmuganathan Raman,arXiv:2004.10362v1 [cs.CV] 22 Apr 2020, published in Computer Society .IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2020

[6] E. Trejo, P. Yuan, “Recognition of yoga poses through an interactive system with kinect device”, Intl. Conf. Robotics and Automation Science, 2018.

[7] Codebox image Augmentor,Github repository.

[8] Shruti Saxena, “Yoga Pose Image classification dataset”, www.kaggle.com/shrutisaxena/yoga-pose-image-classification-dataset