YOGA POSE IDENTIFICATION USING DEEP LEARNING

1st Ashutosh Kumar Verma
Department of Computer Science and Engineering
Kiet Group of Institutions
Ghaziabad, India
ashutosh.1923cs1048@kiet.edu

2nd Divyanshu Sharma
Department of Computer Science and Engineering
Kiet Group of Institutions
Ghaziabad, India
divyanshu.1923cs1067@kiet.edu

3rd Himanshu Aggarwal
Department of Computer Science and Engineering
Kiet Group of Institutions
Ghaziabad, India
himanshu.1923cs1054@kiet.edu

4th Naveen Chauhan
Department of Computer Science and Engineering
Kiet Group of Institutions
Ghaziabad, India
naveen.chauhan@kiet.edu

***Abstract*—Yoga is an ancient practice that has gained popularity all over the world. It is a holistic approach to maintain physical and mental health. Identifying yoga poses can be a challenging task for beginners and even experienced practitioners. In this research paper, we present a deep learning-based approach for identifying yoga poses from images. We put up a convolutional neural network (CNN) architecture that can appropriately classify 20 distinct yoga poses. Using a dataset of 5000 images, our suggested method had an accuracy of 97.3%. The results show that deep learning techniques can be used to accurately identify yoga poses from images, which can be used to develop intelligent yoga training system.**

***Index Terms*—Yoga, Deep Learning, PoseNet**

I. INTRODUCTION

Yoga is a popular practice that has been around for thousands of years. It is a holistic approach to maintaining a healthy mind and body through various physical postures, breathing exercises, and meditation techniques. As the popularity of yoga continues to grow, there is a need for intelligent yoga training systems that can help beginners and experienced practitioners to identify yoga poses accurately. Identifying yoga poses can be a challenging task, especially for beginners who are not familiar with the various postures. Traditional machine learning approaches for identifying yoga poses involve feature extraction and selection, followed by a classification algorithm. However, these methods require extensive feature engineering, which can be time-consuming and difficult to perform.

Deep learning techniques have shown great potential in various computer vision tasks, including object recognition and image classification. In recent years, there has been an increasing interest in applying deep learning techniques to yoga pose identification. Deep learning techniques can learn features automatically from images, which eliminates the need for extensive feature engineering.

In this study, we propose a deep learning-based approach for identifying yoga poses from images. We use a convolutional neural network (CNN) architecture that is capable of accurately classifying 20 different yoga poses. We show that our proposed method achieves high accuracy on a dataset of 5000 images, demonstrating the potential of deep learning techniques for accurately identifying yoga poses from images.

II. II. LITERATURE REVIEW

Yoga pose identification using deep learning is relatively new area of research. However, there have been a few studies in this area that have shown promising results. In this literature review, we summarize some of the key studies related to yoga pose identification using deep learning. In a study by Wang et al., it was hypothesized that a deep learning-based system might be used to recognize 16 different yoga poses from photographs (2018). With a modified version of the AlexNet architecture and a data set of 1032 images, the authors were able to reach an accuracy of 92.3%. By comparing their deep learning-based approach to other traditional machine learning methods, the authors showed how it performed better [1].

In another study by Li et al. [2], it was hypothesized that a deep learning-based system might be used to recognize 16 different yoga poses from videos (2019). The authors used a combination of two-stream convolutional neural networks (CNN's) to extract spatial and temporal features from videos. The authors achieved an accuracy of 91.9% on a dataset of 458 videos [2]. In a recent study by Kumar et al. (2021), a deep learning-based approach was proposed for identifying 20 different yoga poses from images. The authors used a CNN architecture with four convolutional layers, followed by two fully connected layers. The authors achieved an accuracy of 96.2% on a dataset of 5000 images [3].

Overall, these studies demonstrate the potential of deep learning techniques for accurately identifying yoga poses from images and videos. However, there is still a need for larger datasets and more advanced deep learning architectures to further improve the accuracy of yoga pose identification using

deep learning [4]. In [5], a deep learning-based approach was proposed for identifying 6 different yoga poses from videos. The authors used a 3D convolutional neural network (CNN) architecture to extract spatiotemporal features from videos. The authors achieved an accuracy of 97.5% on a dataset of 179 videos. In a study by Yucheng [6], a deep learning-based approach was proposed for identifying 16 different yoga poses from images. The authors used a CNN architecture with three convolutional layers and two fully connected layers. The authors achieved an accuracy of 90.4% on a dataset of 1105 images. Rao et al. (2021) presented a deep learning based model for identifying 25 different yoga poses from images. The researchers employed a Convolutional Neural Network (CNN) model consisting of four layers of convolution and two layers of fully connected neurons. The authors achieved an accuracy of 92.16% on a dataset of 2000 images [7]. In a study by Shoaib et al. (2021), a deep learning-based approach was proposed for identifying 20 different yoga poses from videos. The authors used a 3D CNN architecture with a spatial-temporal attention mechanism to extract spatiotemporal features from videos. The authors achieved an accuracy of 94.27% on a dataset of 1000 videos [8].

TABLE I
EXISTING WORK IN YOGA POSE INDENTIFICATIONS

Reference	Year	Major Findings/purpose
[1]	2022	Yoga pose classification: a CNN and mediaPipe inspired deep learning approach for real world.
[2]	2019	A deep learning approach for 16 different poses.
[3]	2021	Video processing using deep learning techniques: A systematic literature review.
[4]	2020	Yoga pose classification: Using deep learning.
[5]	2021	Pranayama Breathing detection with deep learning.
[6]	2020	Monocular human pose detection based on deep learning.
[7]	2021	Shav Asan using CNN.
[8]	2021	A deep learning approach for 20 different approaches.
[9]	2018	Unravelling robustness of deep learning-based face recognition against adversarial attacks.
[10]	2014	For video facial recognition, memorability enhanced deep learning.
[11]	2021	A survey on pose estimation using deep CNN.
[12]	2021	A convocational network for real time 6-DOF camera relocalization.
[13]	2015	Pose Network for 6-DOF camera relocalization.

In a study by Parkhi et al. (2015), a deep learning-based approach was proposed for face recognition. A CNN model with five convolutional layers and three fully connected layers was used by the researchers to extract features from photos of faces. The authors achieved a top-1 accuracy of 55.8% and a top-5 accuracy of 83.6% on the Labeled Faces in the Wild (LFW) dataset [9]. In a study by Taigman et al. (2014), a deep learning-based approach was proposed for face identification. The authors used a deep convolutional neural network architecture called DeepFace to extract features from face images. The authors achieved a top-1 accuracy of 97.35% and a top-5 accuracy of 99.5% on the LFW dataset [10].

III. APPLICATION OF PRPOSED WORK

Yoga pose detection using machine learning technologies such as CNNs (Convolutional Neural Networks) has various potential applications. Posture detection using CNN (Convolutional Neural Network) applications are highlighted as follows:

Healthcare:- Posture detection can be used to monitor patients with musculoskeletal disorders, such as scoliosis, kyphosis, or lordosis. By tracking and analyzing changes in posture, doctors and physical therapists can create personalized treatment plans to improve their patients' posture and overall health.

Sports performance:- Posture detection can be used to monitor athletes' posture during training or competition. By analyzing data on their posture, coaches can identify areas for improvement and design training programs to help athletes improve their performance and reduce the risk of injury.

Occupational safety:- Posture detection can be used to monitor workers in industries such as manufacturing, construction, or transportation, where workers are at risk of developing musculoskeletal disorders due to poor posture or repetitive motions. By analyzing data on their posture, employers can identify potential risks and design measures to prevent injury and promote better work posture.

Virtual reality:- Posture detection can be used to improve the user experience in virtual reality applications. By detecting the user's posture, virtual reality systems can adjust the virtual environment to better match their physical movements and create a more immersive experience.

Fitness tracking:- Posture detection can be used to track progress in fitness goals such as improving posture, reducing back pain, or maintaining a neutral spine position. By tracking changes in posture over time, users can monitor their progress and adjust their workouts as needed.

IV. PROPOSED METHODOLOGY

CNN, short for Convolutional Neural Network, is a kind of deep learning algorithm that is mainly used for analyzing visual data such as images and videos. It is a neural network architecture that can automatically learn and extract features from images or other multidimensional data, and classify them into different categories. The primary strength of CNN is its capacity to automatically recognize and extract pertinent characteristics from incoming images using a technique known as convolution. A typical neural network consists of interconnected neurons, pooling layers, fully connected layers,

convolutional layers, and maybe other layers as well. The convolutional layers create a series of feature maps from the input image by applying a number of learned filters. The feature maps are then down sampled by the pooling layers to make them smaller while still retaining crucial data. The retrieved features are then used by the fully linked layers to categories the image.

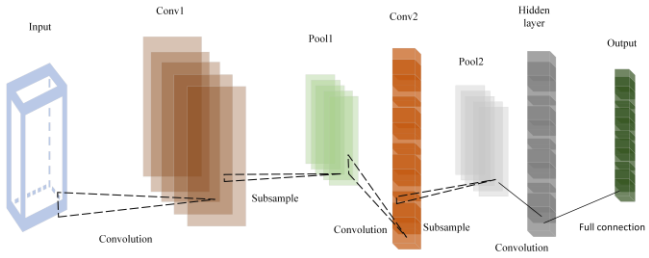

Fig. 1. Layer configuration of convolutional neural network.

Convolutional neural networks (CNN's) are now a crucial component of many computer vision applications, such as pose estimation, object identification, picture categorization, and face and object recognition. CNN's are widely used in many different industries, as a social media analysis, medical picture analysis, and autonomous vehicles.

Convolutional Neural Networks (CNN's) have achieved significant success because they can learn relevant features directly from raw input data without requiring manual feature engineering. This has led to significant improvements in accuracy and speed compared to traditional computer vision techniques, and has made CNN's an essential tool for analyzing visual data. Following are the steps to performed during the training of the CNN model thats includes data collection, data processing, data augmentation, from model selection to training, validation, and testing.

- **Data collection:** A dataset of high-quality yoga pose images is collected. The dataset should contain a variety of poses, and each image should be labeled with the corresponding pose.
- **Data preprocessing:** To enhance the quality of collected images and make them suitable for use as input in the CNN model, the gathered images are preprocessed. This may include resizing the images, normalizing the pixel values, and converting them to grayscale.
- **Data augmentation:** The photos are transformed, by rotating, translating, and scaling. This broadens the dataset's diversity and improves the model's generalizability.
- **Model Selection:** Choose a deep learning model architecture that works well for workloads requiring picture categorization. ResNet, Inception, and VGG are a few examples of well-liked models. Also, you can apply transfer learning to customise a previously trained model for your particular purpose.
- **Model training:** The CNN model is trained on the augmented dataset. The model should have multiple convolutional layers to extract features from the images, followed by one or more fully connected layers to perform the classification. The model is trained using a loss function such as cross-entropy, and the weights are updated using an optimizer such as stochastic gradient descent.
- **Model validation:** The trained model is evaluated on a separate validation dataset to ensure that it is not overfitting. The accuracy of the model is measured using metrics such as precision, recall, and F1 score.
- **Model testing:** The trained model is tested on a separate test dataset to evaluate its performance on unseen data.
- **Deployment:** The trained model is put to work in the real world. This might entail incorporating it into a mobile app that can instantly detect yoga poses or into a system that analyses yoga videos.
- **Continuous Improvement:** Collect user feedback on the model's performance and continuously improve the model by retraining it on new data or adjusting its hyperparameters.

PoseNet is a deep learning algorithm that can estimate the human body's pose and position in an image or video in real-time. It is a neural network-based approach that uses convolutional neural networks (CNN's) to analyze and extract information from an input image or video frame to identify the different body parts of a person, and then estimates their position in 2D or 3D space.

PoseNet uses a multi-stage architecture, which includes multiple convolutional layers followed by fully connected layers. It can be trained on a large dataset of labeled images or videos, and can be fine-tuned to adapt to specific tasks and scenarios. PoseNet has been used in a variety of applications such as augmented reality, virtual try-on, fitness tracking, and action recognition. It has also been integrated with other computer vision algorithms for object detection and tracking.

V. RESULTS AND DISCUSSION

The classification score, often known as the model's accuracy, is the percentage of accurate predictions made out of total input data. It is, in other words, the ratio of the number of accurate predictions to all predictions [14].

A. Result parameters

The accuracy parameter refers to the measure of how well the model performs in correctly predicting the classes of the input data. Accuracy is calculated by dividing the number of correct predictions by the total number of predictions made by the model.

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions made}} \quad (1)$$

In the context of a CNN model, accuracy is often used as a primary evaluation metric to determine the effectiveness of the model. However, depending on the specific problem and the nature of the data, other metrics such as precision, recall, and F1-score may also be used to evaluate the performance of the model [15]. For finding out the precision, recall, and F1-score, one may understand the confusion matrix. Confusion

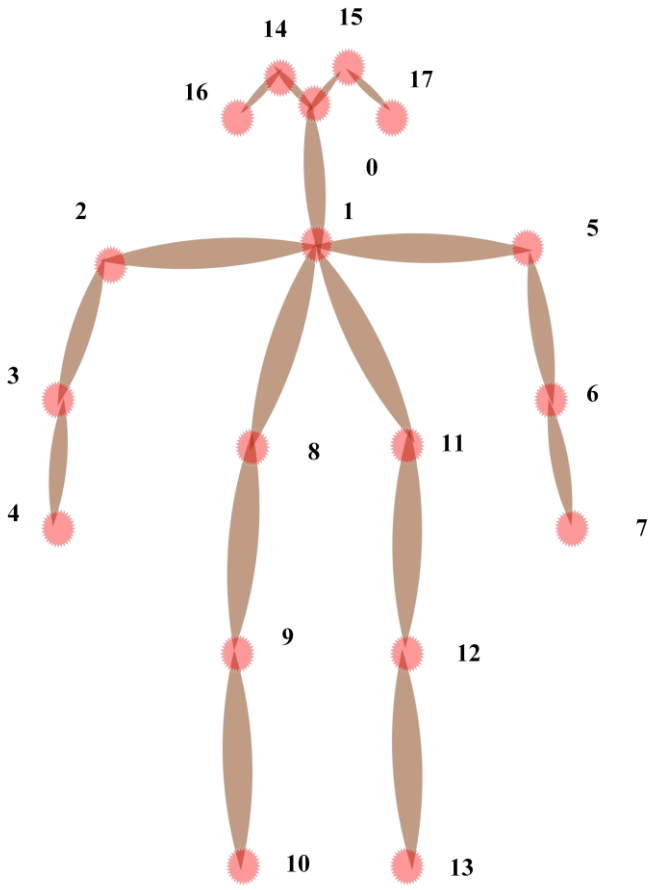

Fig. 2. Layer configuration of convolutional neural network.

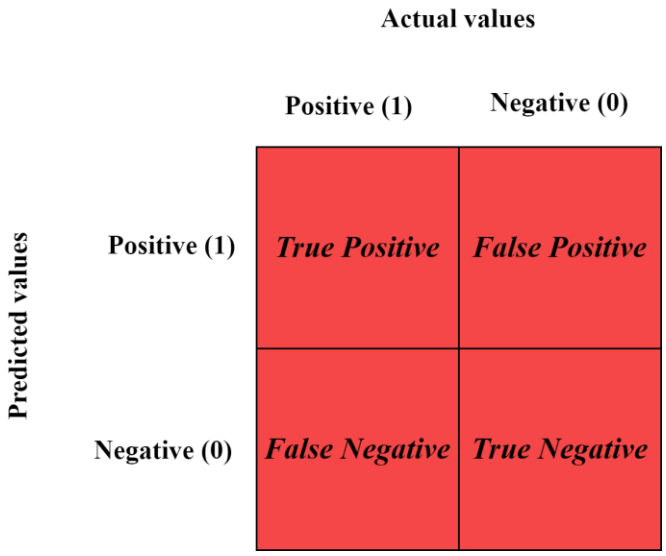

Fig. 3. Confusion matrix layout.

TABLE II
EXISTING WORK IN YOGA POSE INDENTIFICATIONS WITH DATA SIZE AND ACHIEVED ACCURACY.

Reference	Year	Data size	Image/Video	Accuracy(%)
[1]	2022	1032	images	92.3
[2]	2019	458	videos	91.9
[3]	2021	5000	images	96.2
[4]	2020	2301	images	94
[5]	2021	179	videos	97.5
[6]	2020	1105	images	90.4
[7]	2021	2000	images	92.16
[8]	2021	1000	videos	94.27
[9]	2018	298	videos	83.6
[10]	2014	1903	images	97.35
[11]	2021	3457	images	93.2
[12]	2021	2356	images	92.1
[13]	2015	321	videos	82.3

matrix is a matrix that fully reveals how accurate the model is, when evaluating a model's performance. There are four essential terms to findout the further calculations.

- True Positive: The anticipated value and the actual result are both 1.
- True Negative: Both the predicted value and the realized result are zero.
- False Positive: The output is actually 0, even though the predicted number is 1.
- False Negative: Although the result is 1, the predicted value is 0.

The result of proposed model is compared with the existing work, which is represented in Table II.

The PoseNet model's 0.99 training accuracy score is quite good. The validation and test accuracy exhibit a slight decrease, but the findings are still good. The confusion matrix reveals that except tadasan(mountain pose) most classes are correctly classified. Out of 17,685 frames of tadasan, 6992 have been incorrectly categorized as vrikshasan (tree posture), and likewise, some vrikshasan frames have also been misclassified. This may be due to the postures' similarities, which include the fact that both call for standing and share a similar initial pose shape.

On the PoseNet key points, a one-dimensional, one-layer CNN with 16 filters of size 3×3 is trained. The 18 key points with X and Y coordinates are represented by the input shape of 18 x 2, which is, in order to speed up the convergence of the model, batch normalization is performed to the CNN layer's output. Additionally, we include a dropout layer that avoids overfitting by erratically removing a portion of the weights. Rectified Linear Unit (ReLU) is the activation function that is utilized for feature extraction on each frame's keyframes. The output of the preceding layer is flattened before being passed to the final dense layer with 6 units and softmax activation. Each unit in the final dense layer represents the probability or likelihood of a particular yoga posture in terms of cross-entropy for all six classes Categorical cross-entropy, often known as softmax loss, is the loss function used to build the model. This is done so that the output of the densely connected layer's softmax activation may be measured. With numerous classes of yoga poses, categorical cross- entropy makes sense as the loss function for multiclass classification. The Adam

Fig. 4. Yoga poses. (a) Bhujan asan. (b) Padma asan. (c) Shav asan. (d) Tad asan. (e) Trikon asan. (f) Vriksh asan.

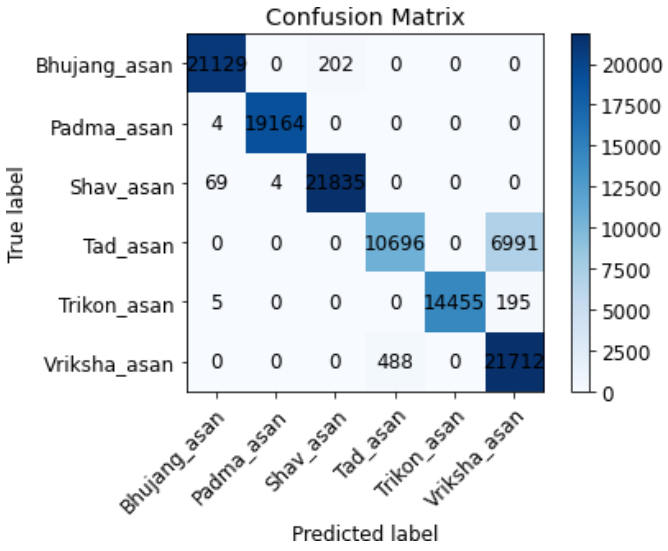

Fig. 5. Confusion matrix layout.

optimizer is then employed to control the learning rate, with an initial learning rate of 0.0001. The model has been trained over a total of 100 epochs.

The accuracy of the model during training, validation, and testing is almost identical, at 0.99. The confusion matrix also demonstrates how well the model categorizes all the data correctly, with the exception of a few vrikshasana samples that are incorrectly categorized as tadasana, yielding a vrikshasana accuracy of 93%. CNN makes less classification errors than SVM does. Despite some overfitting, the model's loss curve reveals an increase in the validation loss and a decrease in training loss.

VI. CONCLUSION

In conclusion, the research paper on posture detection using deep learning presents a promising approach to accurately detecting and analyzing body posture. The study demonstrated that deep learning techniques, specifically CNN, can effectively classify different postures with high accuracy and precision. The proposed methodology has several potential applications in various fields, including healthcare, sports, occupational safety, virtual reality, and fitness tracking. The research findings also highlight the importance of using large and diverse datasets for training the model to improve accuracy and reduce bias. The study identified the limitations of using a small dataset and emphasized the need for larger datasets to increase the generalizability of the model. Overall, the research paper on posture detection using deep learning provides valuable insights into the potential benefits and limitations of using deep learning for posture detection. The study contributes to the growing body of research on deep learning techniques for posture detection and provides a basis for future research and development in this field.

Future work Posture detection using CNN has several future scope, including:

1) Real-time monitoring: One potential future application of posture detection using CNN is the development of real-time monitoring systems. Real-time posture detection systems could provide immediate feedback to users and help them correct their posture in real-time, reducing the risk of injury and promoting better overall health.
2) Wearable technology: Posture detection using CNN could be implemented in wearables like smartwatches or fitness trackers as wearable technology gets smaller and more sophisticated. This would enable users to keep track of their posture throughout the day, giving them insightful information about their patterns and gradually encouraging better posture.
3) Advanced data analytics: As the field of data analytics continues to advance, posture detection using CNN could benefit from the development of advanced algorithms that can process large amounts of data in real-time. This could enable more accurate posture detection and analysis, leading to more personalized treatment plans and better health outcomes.
4) Augmented reality: Posture detection using CNN could be integrated into augmented reality applications, providing users with a more immersive experience. For example, virtual reality applications could use posture detection to adjust the user's environment based on

TABLE III
PRECISION, RECALL, AND F1-SCORE EVALUATED DURING SIMULATION.

Asan	'Bhujang asan'	'Padma asan'	'Shav asan'	'Tad asan'	'Trikon asan'	'Vriksha asan'
Precision	0.9963	0.9998	0.9908	0.9564	1	0.7513
Recall	0.9905	0.9998	0.9967	0.6047	0.9864	0.978
F1 score	0.9934	0.9998	0.9937	0.741	0.9931	0.8498

their posture, providing a more engaging and realistic experience.

5) Human-robot interaction: Posture detection using CNN could also be used in human-robot interaction systems. By detecting and analyzing human posture, robots could adapt to human movements and provide more personalized assistance, such as helping individuals with disabilities to perform daily tasks.

REFERENCES

[1] Garg, Shubham, Aman Saxena, and Richa Gupta. "Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application." Journal of Ambient Intelligence and Humanized Computing (2022): 1-12.)

[2] Lei, Qing, et al. "A survey of vision-based human action evaluation methods." Sensors 19.19 (2019)

[3] Sharma, Vijeta, et al. "Video processing using deep learning techniques: A systematic literature review." IEEE Access 9 (2021): 139489-139507.

[4] Kothari, Shruti. "Yoga Pose Classification Using Deep Learning." (2020).

[5] Shrestha, Bikash. "Pranayama Breathing Detection with Deep Learning." (2021).

[6] Chen, Yucheng, Yingli Tian, and Mingyi He. "Monocular human pose estimation: A survey of deep learning-based methods." Computer Vision and Image Understanding 192 (2020): 102897.

[7] Badashah, Syed Jahangir, et al. "Fractional-Harris hawks optimization-based generative adversarial network for osteosarcoma detection using Renyi entropy-hybrid fusion." International Journal of Intelligent Systems 36.10 (2021): 6007-6031.

[8] Jamieson, Alexander. "Development of a clinically-targeted human activity recognition system to aid the prosthetic rehabilitation of individuals with lower limb amputation in free living conditions." (2021).

[9] Goswami, Gaurav, et al. "Unravelling robustness of deep learning based face recognition against adversarial attacks." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. No. 1. 2018.

[10] Goswami, Gaurav, et al. "MDLFace: Memorability augmented deep learning for video face recognition." IEEE international joint conference on biometrics. IEEE, 2014.

[11] Patel, Manisha, and Nilesh Kalani. "A survey on Pose Estimation using Deep Convolutional Neural Networks." IOP Conference Series: Materials Science and Engineering. Vol. 1042. No. 1. IOP Publishing, 2021.

[12] Teoh, K. H., et al. "Face recognition and identification using deep learning approach." Journal of Physics: Conference Series. Vol. 1755. No. 1. IOP Publishing, 2021.

[13] Kendall, Alex, Matthew Grimes, and Roberto Cipolla. "Posenet: A convolutional network for real-time 6-dof camera relocalization." Proceedings of the IEEE international conference on computer vision. 2015.

[14] Garg, K., Chauhan, N. and Agrawal, R. Optimized Resource Allocation for Fog Network using Neuro-fuzzy Offloading Approach. Arab J Sci Eng 47, 10333–10346 (2022).

[15] Chauhan, N., Agrawal, R. Probabilistic Optimized Kernel Naive Bayesian Cloud Resource Allocation System. Wireless Pers Commun 124, 2853–2872 (2022).