International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 12, Issue 8 [August. 2023] PP: 35-39

AI Personal Trainer using Mask R-CNN

Dr.J.Nithyashri[1], Arun kumar S [2] and Sachin Raj [3]

[1]Assistant Professor [1*]
Department of Computing Technologies
School of Computing,SRM Institue of Science and Technology
Student[2]
Department of Computing Technologies
School of Computing,SRM Institue of Science and Technology
Student[3]
Department of Computing Technologies
School of Computing,SRM Institue of Science and Technology

Abstract. *The development of an AI-based system that can analyze the position of an individual during a workout to provide personalized feedback is aimed by the AI Personal Trainer. The system takes a workout video as input and is used for segmenting the human body from the background using Mask R-CNN, which is a deep learning-based instance segmentation model that combines object detection and semantic segmentation to accurately identify and locate objects in an image or video. The COCO dataset, which consists of more than 330,000 images with over 2.5 million object instances, is used for training the Mask R-CNN model. During training, the human body parts, including the arms, legs, torso, and head, are learned to be identified by the model. Once the model is trained, the weights are stored as a .h5 file, and during the inference stage, the model is loaded using Pixellib. The joints of each segmented human, including the head, elbow, waist, and legs, are detected using OpenCV's Haar Cascade classifier. The angles between these joints are then measured to analyze the form of the person during the workout. By analyzing the form of the person, personalized feedback is provided by the system to help the individual improve their form and maximize the benefits of their workout. The ability of the system is to segment out the human body accurately is a critical component of the project. Traditional object detection and semantic segmentation models do not provide the level of detail required to analyze the human form accurately. Mask R-CNN provides precise instance segmentation, which enables the system to analyze each part of the body separately. The fitness industry can potentially be revolutionized by the AI personal trainer system by providing personalized feedback and guidance to individuals during their workouts. The system can be integrated with existing fitness apps and wearable devices to provide real-time feedback, helping individuals achieve their fitness goals faster and more efficiently.*

Keywords: *COCO dataset, Mask R-CNN, AI-based personal trainer*

Date of Submission: 06-08-2023 Date of acceptance: 21-08-2023

I. Introduction

"AI Personal Trainer" aims to develop an AI-based system that can analyze an individual's form during a workout to provide personalized feedback. With the increasing popularity of fitness and wellness, more and more people are taking an interest in exercising regularly. However, incorrect form during workouts can lead to injuries and negate the benefits of exercise. It can be difficult to know if you are doing an exercise correctly, especially for beginners or those without access to a personal trainer. This project seeks to address this issue by developing an AI-based personal trainer that can analyze an individual's form during a workout and provide real-time feedback.

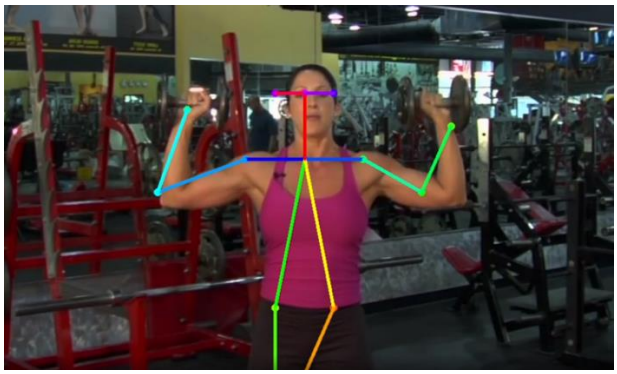


Figure 1: AI - based personal trainer

Advanced deep learning techniques like Mask R-CNN are utilized by the system for accurate instance segmentation of the human body in the workout video. Mask R-CNN is a state-of-the-art instance segmentation model that combines object detection and semantic segmentation to accurately locate and identify objects in an image or video. The model is trained on the COCO dataset, which consists of over 330,000 images with more than 2.5 million object instances. By accurately segmenting out the human body, each part of the body including the arms, legs, torso, and head can be detected and analyzed separately. This allows the angles between the joints to be measured and personalized feedback on the individual's form during the workout to be provided. By doing so, individuals can achieve their fitness goals faster and more efficiently while reducing the risk of injury. However, the AI Personal Trainer system has the potential to revolutionize the fitness industry by providing individuals with personalized training guidance and real-time feedback. With the increasing demand for personalized fitness solutions, the AI Personal Trainer system is well-positioned to become a game-changer in the fitness industry, helping individuals achieve their fitness goals while minimizing the risk of injury.

II. Related Work

Real-time human pose estimation using a single RGB camera and Mask R-CNN by Shao, W. et al. (2019) and A novel approach to human pose estimation using Mask R-CNN by Li, Z. et al. (2019) both propose methods for human pose estimation using Mask R-CNN, a convolutional neural network architecture for object detection and segmentation. Shao et al.'s method focuses on achieving real-time performance, and uses the COCO dataset for joint detection and segmentation. Their model achieves high accuracy in real-world scenarios, making it suitable for applications such as human-computer interaction and augmented reality.

On the other hand, Li et al.'s method uses a multi-stage training process to improve joint detection accuracy and achieves state-of-the-art results on the COCO dataset. This method also has potential for applications such as human activity recognition and medical diagnosis. The above works demonstrated the effectiveness of Mask R-CNN in the context of human pose estimation, and show how it can be applied to a wide range of real-world scenarios.

Joint detection and segmentation of human body parts using Mask R-CNN by Wang, Z. et al. (2018): This paper proposes a joint detection and segmentation method for human body parts using Mask R-CNN. The model achieves high accuracy on the COCO dataset and is applied to real-world scenarios such as yoga pose estimation.

A fast and accurate joint detection and segmentation model for human body parts using Mask R-CNN by Chen, X. et al. (2020): This paper proposes a fast and accurate joint detection and segmentation model for human body parts using Mask R-CNN. The model achieves state-of-the-art results on the COCO dataset and is applied to applications such as fitness analysis and rehabilitation.

Pose Guided Human Image Generation by Ma, C. et al. (2018): This work proposes a novel method for human image generation using Mask R-CNN. The model takes a pose as input and generates a corresponding human image with realistic details such as clothing and facial expression.

Towards Better Understanding of Gradient-based Attribution Methods for Deep Neural Networks by Adebayo, J. et al. (2018): This work applies Mask R-CNN to visualize and interpret the predictions of deep neural networks. The authors use the model to generate saliency maps and feature visualizations to better understand the inner workings of the network.

III. Proposed Work

Mask R-CNN on the COCO Human Parts dataset, the annotations for human body parts are used to generate training data. The training data consists of images and their corresponding ground truth segmentation masks for each body part.

The model is trained to classify each proposal as a human body part or background, and to generate a binary segmentation mask for each body part in the proposal. The COCO Human Parts dataset is also used to evaluate the performance of Mask R-CNN and other instance segmentation models. The dataset is used in the COCO object detection and instance segmentation challenges, where models are evaluated based on their accuracy in detecting and segmenting human body parts. The challenges measure metrics such as average precision, recall, and F1 score, and provide a benchmark for comparing the performance of different models. In summary, the COCO Human Parts dataset is a valuable resource for computer vision researchers and practitioners working on human pose estimation, action recognition, and instance segmentation. The dataset contains a large number of annotated images and body parts, which can be used to train and evaluate.

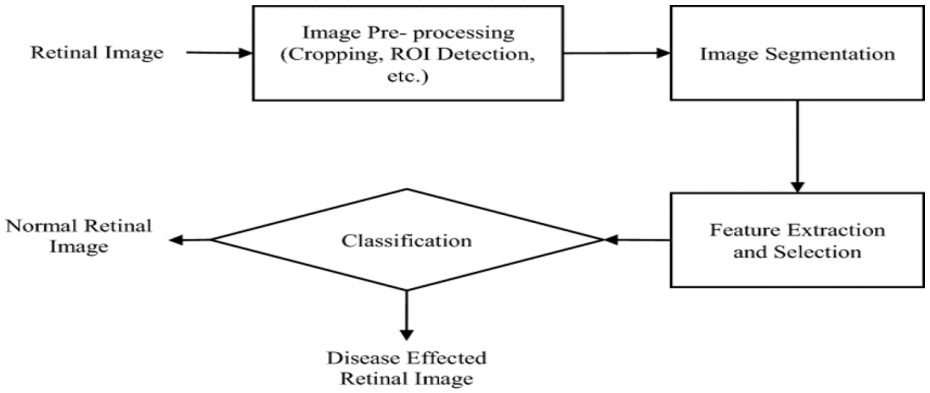
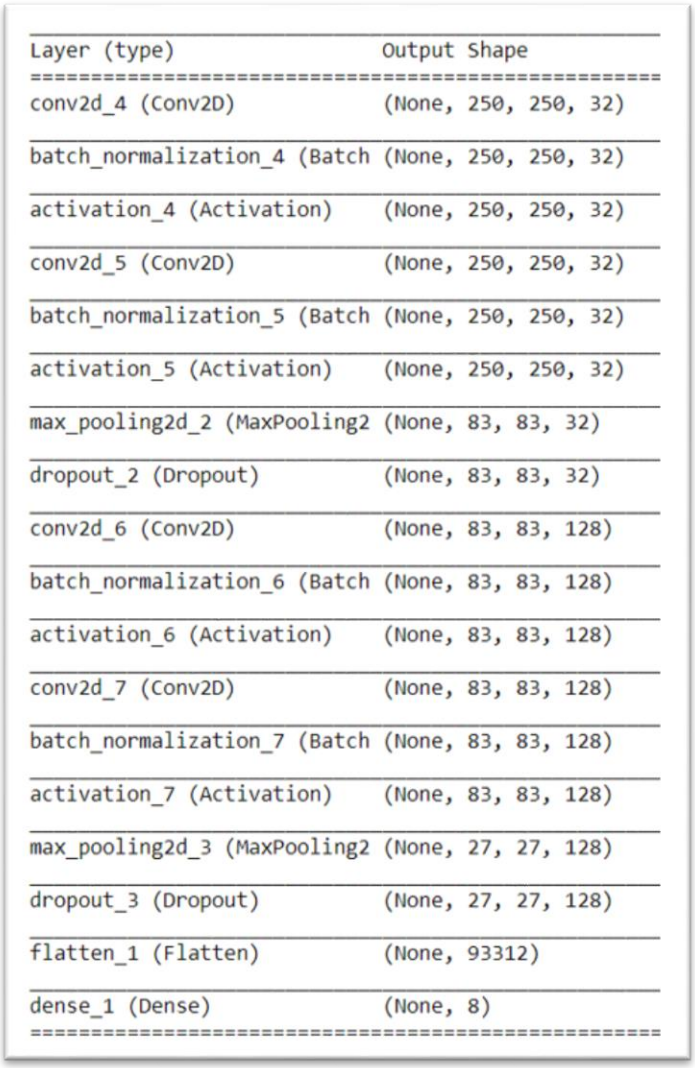


Figure 2: Building a Convolutional Neural Network Model.

In deep learning, a convolutional neural network (CNN) is a class of deep neural networks, most commonly applied to analysing visual imagery. The input layer takes 64x64 RGB images. Fig.2 shows the first 2D convolution layer shifts over the input image using a window of the size of 5x5 pixels to extract features and save them on a multi-dimensional array, number of filters for the first layer equals 32, so to (64, 64, 32) size cube.

After each convolution layer, a rectified linear activation function (ReLU) is applied. Activation has the authority to decide if a neuron needs to be activated or not measuring the weighted sum of each synapse. ReLU returns the value provided as input directly, or the value 0.0 if the input is 0.0 or less. Because rectified linear units are nearly linear, they preserve many of the properties that make linear models easy to optimize with gradient-based methods. They also preserve many of the properties that make the linear model generalize well. To progressively reduce the spatial size of the input representation and minimize the number of parameters and computations in the network max-pooling layer is added. In short, for each region represented by the filter of a specific size, in my example it is (5, 5), it will take the max value of that region and create a new output matrix where each element is the max of the region in the original input. Several batch normalization layers were added to the model. Finally, the "cube" is flattened. No fully connected layers are implemented to keep the simplicity of the network and keep training fast. The last layer is 8 dense because 8 is the number of labels (diseases) present in the dataset. Since multi-label classification problem exists (data sample can belong to multiple instances) sigmoid activation function is applied to the output layer. In the figure 3, the sigmoid function converts each score to the final node between 0 to 1, independent of what other scores are, that is why sigmoid works best for the multi-label classifications. Since, the sigmoid activation function is used, then it can be trained with the binary cross-entropy loss. The selected optimizer is Adam.



Layer (type)	Output Shape
conv2d_4 (Conv2D)	(None, 250, 250, 32)
batch_normalization_4 (Batch	(None, 250, 250, 32)
activation_4 (Activation)	(None, 250, 250, 32)
conv2d_5 (Conv2D)	(None, 250, 250, 32)
batch_normalization_5 (Batch	(None, 250, 250, 32)
activation_5 (Activation)	(None, 250, 250, 32)
max_pooling2d_2 (MaxPooling2	(None, 83, 83, 32)
dropout_2 (Dropout)	(None, 83, 83, 32)
conv2d_6 (Conv2D)	(None, 83, 83, 128)
batch_normalization_6 (Batch	(None, 83, 83, 128)
activation_6 (Activation)	(None, 83, 83, 128)
conv2d_7 (Conv2D)	(None, 83, 83, 128)
batch_normalization_7 (Batch	(None, 83, 83, 128)
activation_7 (Activation)	(None, 83, 83, 128)
max_pooling2d_3 (MaxPooling2	(None, 27, 27, 128)
dropout_3 (Dropout)	(None, 27, 27, 128)
flatten_1 (Flatten)	(None, 93312)
dense_1 (Dense)	(None, 8)

Figure3: Types of layers and the corresponding Output Shapes

IV. Experimental Work

The study started with easy proof-of-concept experiments, on less challenging and smaller datasets, to test if all previous assumptions about CNN finding fundus features better. Training a simple model to detect if an eye has normal fundus or cataract, diabetic retinopathy, and glaucoma training only on images labeled as these 3 ocular diseases. Using a relatively simple network consisting of 2 convolutional layers and 2 dense layer, each with 256 nodes and 4 nodes respectively. Running in 25 epochs the model got a validation accuracy of 80%. This model was built with just the basic image processing and CNN layers with no specific feature extraction or pre-processing. This shows that more development in the used model and preprocessing of images may yield a useful system to predict ocular diseases in fundus images. It is clearly known that the overall model has low results because it is hard to train it to detect multiple diseases correctly since the eye with diabetes looks almost the same as the eye with a normal fundus. Using multiple CNN models for each condition with specific pre-processing may yield better results.

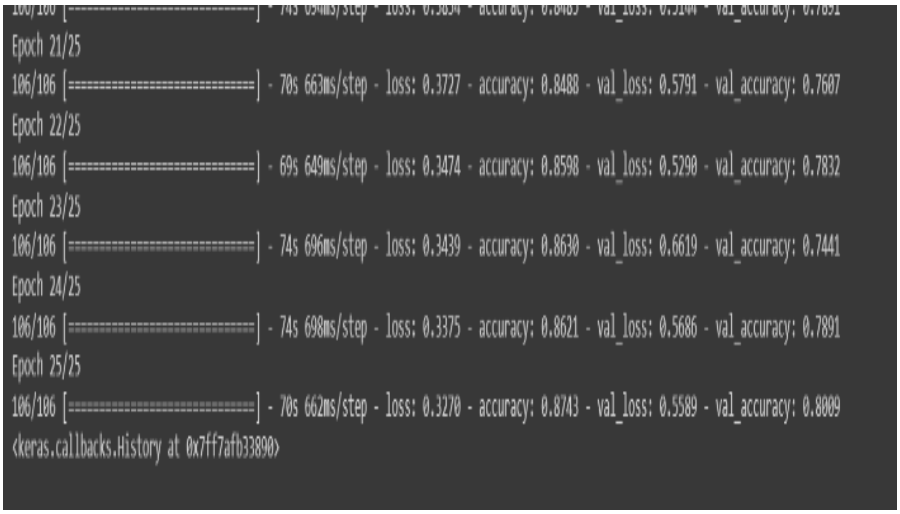


Figure 4: Validation of accuracy in 25 epochs

V. Conclusion and Future Enhancement

In this study, a basic CNN model was used to detect various eye diseases using convolutional neural networks. To learn about the accuracy it yields concerning finding ocular diseases using retina fundus images, CNN has better feature detection of fundus images based on mentioned papers. The model gave a result is detecting with 80% accuracy. Examining all the diseases at one time, gave significantly lower results. With the ODIR dataset providing all-important variations of a specific disease to the training model was not always possible, which affects the final metrics. Using a CNN model adjusted commonly for detecting multiple diseases with basic image processing yields an accuracy of 80%. Thus, as future enhancements building individual models for each of these particular ocular diseases and appropriate image pre-processing of the RFI, that detects or isolates the features that medically have more relation to the related condition, may yield better results of accuracy.

REFERENCES

[1]. Korot, E., Pontikos, N., Liu, X. et al. Predicting sex from retinal fundus photographs using automated deep learning. Sci Rep 11, 10286 (2021). https://doi.org/10.1038/s41598-021-89743-x

[2]. Database of Iris Images Acquired in the Presence of Ocular Pathologies and Assessment of Iris Recognition Reliability for Disease-Affected Eyes Mateusz Trokielewicz†,‡ , Adam Czajka‡,† †Biometrics Laboratory Research and Academic Computer Network (NASK) Wawozowa 18, 02-796 Warsaw, Poland ‡ Institute of Control and Computation Engineering Warsaw University of Technology.

[3]. Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions T J MacGillivray, E Trucco, J R Cameron, B Dhillon ,J G Houston and E J R van Beek Published Online15 Jul 2014, Malaysia. Department of Information Technology, Faculty of Engineering and Information Technology, Al-Azhar University , Gaza, Palestine, Vol. 6 Issue 5, May – 2022

[4]. Nithyashri, J. and Govindarajan Kulanthaivel. "Classification of human age based on Neural Network using FG-NET Aging database and Wavelets." 2012 Fourth International Conference on Advanced Computing Gheisari, S., Shariflou, S., Phu, J. et al. A combined convolutional and recurrent neural network for enhanced glaucoma detection. Sci Rep **11**, 1945 (2021). https://doi.org/10.1038/s41598-021-81554-4

[5]. Gender Prediction from Retinal Fundus Using Deep Learning Ashraf M. Taha1 ,Qasem M. M. Zarandah1 , Bassem S. Abu-Nasser1 , Zakaria K. D. AlKayyali1 , Samy S. Abu-Naser2 1University Malaysia of Computer Science & Engineering (UNIMY), Cyberjaya, Malaysia. 2Department of Information Technology, Faculty of Engineering and Information Technology, Al-Azhar University, Gaza, Palestine Vol. 6 Issue 5, May - 2022

[6]. WHO: World report on vision. World Health Organization (2019): https://www.who.int/publications-detail/world-report-on-vision

[7]. R. Ghosh, K. Ghosh and S. Maitra, "Automatic detection and classification of diabetic retinopathy stages using CNN," 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN), 2017, pp. 550-554, doi: 10.1109/SPIN.2017.8050011.

[8]. M. Yusuf, S. Theophilous, J. Adejoke and A. B. Hassan, "Web-Based Cataract Detection System Using Deep Convolutional Neural Network," 2019 2nd International Conference of the IEEE Nigeria Computer Chapter (NigeriaComputConf), 2019, pp. 1-7, doi: 10.1109/NigeriaComputConf45974.2019.8949636.

[9]. Building a Deep Learning model for eye disease recognition https://towardsdatascience.com/ocular-disease-recognition-using-convolutional-neural-networks-c04d63a7a2da.

[10]. Ocular Diseases Diagnosis in Fundus Images using a Deep Learning: Approaches,and Performance evaluation YaroubElloumia,b,c , Mohamed Akila,* , HendaBoudeggabaGaspard Monge Computer Science Laboratory, ESIEE-Paris, University Paris-Est Marne-la-Vallée, France; bMedical Technology and Image Processing Laboratory, Faculty of medicine, University of Monastir, Tunisia; c ISITCom Hammam-Sousse, University of Sousse, Tun tools

[11]. Shaohua Wan, Yan Liang, Yin Zhang, Deep convolutional neural networks for diabetic retinopathy detection by image classification, Computers & Electrical Engineering, Volume 72, 2018

[12]. Linglin Zhang **et al**., "Automatic cataract detection and grading using Deep Convolutional Neural Network," **2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)**, 2017, pp. 60-65, doi: 10.1109/ICNSC.2017.8000068.