International Journal of Current Research and Review
DOI: http://dx.doi.org/10.31782/IJCRR.2021.131123

Scopus®

Research Article

IJCRR
Section: Healthcare
ISI Impact Factor (2019-20): 1.628
IC Value (2019): 90.81
SJIF (2020) = 7.893

Sitting Posture Identifier to Overcome Health Issues

Jonathan Law Hui Hao[1], Rajasvaran Logeswaran[2], Hema Latha Krishna Nair[3]

Asia Pacific University of Technology and Innovation, Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Malaysia.

ABSTRACT

Introduction: Bad sitting posture habits cause health issues such as headaches or discomfort in the back and can lead to expensive medical expenditure for correction or cure.

Objective: This study proposes a system that can identify a person sitting posture to aid them in practising good postural habit.

Methods: The system, built using a Convolutional Neural Network, provides sitting posture feedback based on photographs taken via a developed mobile application.

Result: Upon analysis of the dataset, it was found that many factors can influence and adversely affect the results. Examples include a curved back chair may indicate good sitting posture, or perhaps an extremely thick coat may affect the labelled person sitting posture in the image.

Conclusion: Although inaccuracies may be introduced by the shape of clothes and accessories, the tool can be used as a self-improvement aid to practice good sitting posture.

Key Words: Habits, Health issues, Sitting posture, AlexNet, Tensorflow

INTRODUCTION

This study aims to inculcate cost-effective preventive measures by using intelligent systems to identify bad sitting posture. With the use of everyday devices such as a mobile phone, one can detect sitting posture habit immediately by capturing or uploading an image to the application.

Computer vision techniques are applied in this project to extract important information from the image. Techniques such as a saliency filter, edge detection or background subtraction can remove noise from the image and only return the important information of a person sitting from an image. As the input image will be sent from a mobile phone camera or webcam, the image will be a 2-dimensional image, giving no information of depth. Then a convolutional neural network is applied to extract features from the image and return a probability of a person practising a good or bad posture. [1]

LITERATURE REVIEW

There are many types of research done and techniques proposed in the field of detection using machine learning. Each technique has its pros, cons and variance in the expected result. Understanding the available options would be vital to achieving the targeted goal. A few essential components that are necessary for this project include methods of detecting posture, image pre-processing algorithms and commonly used training techniques for multi-label classification data. These are discussed below.

Concept of Posture Detection

The basic concept of posture detection is to determine a person's posture based on the collected data. The data can be in the form of visual data (e.g. image and video) or data from pressure sensors. Several methods have been used to detect human posture. Sensor-based detection was designed and tested in,[1] where the researchers placed the force sensor on the cushion, and microcontrollers and circuits were

Corresponding Author:
Hema Latha Krishna Nair, Asia Pacific University of Technology and Innovation, Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Malaysia; Email: hema.krishna@apu.edu.my

ISSN: 2231-2196 (Print) **ISSN:** 0975-5241 (Online)

Received: 05.10.2020 **Revised:** 08.12.2020 **Accepted:** 05.02.2021 **Published:** 04.06.2021

designed to retrieve the data from the person sitting against the cushion. This allowed data to be collected, analysed and categorized. Unfortunately, this method is costly as it requires physical sensors and designing of circuit boards.

Another group of researchers used flexible sensors that are inserted into clothes near core regions such as the knee and hip. [2] The method involves having the sensors transfer data through Bluetooth to their system. Based on the data that was transferred into the system, a rule-based algorithm will identify the subject's position. As Youngs, et al. recorded, the rule-based algorithm can accurately detect 88% of the 6 postures related to being in and around a bed.

Training for Posture Detection

After the data collection phase, training the model is essential for processing new information. As mentioned earlier, posture detection may be done using sensor data or images, which can be passed on as input for the training. Images will first be converted into arrays of data, then passed to a convolutional algorithm to train on.[3] Sensor data will be trained with a regular network (e.g. feedforward backpropagation). Unlike fully connected neural networks, Convolutional Neural Network neurons iterate through the image and identify features in the image. A fixed image height and width would be required for the model to extract features. Depending on the image size and colour channel, computational power will be expensive.

Feature Extraction

There are a few methods proposed by various researchers to extract features from an image. One of the methods was to detect the important information by colour.[4] However, the researcher stated that environmental lighting has a huge impact on the detection, and calibration is required frequently. The author also proposed a more sophisticated algorithm to detect features, known as Haar-like features that categorize subsections of an image based on the difference of the sum of each neighbouring pixel. The report also introduced the cascade classifier, which consists of several stages where an image will be passed through to be evaluated, and at each stage, the classifier will return a positive if the object is detected and negative otherwise. In the end, if an image that passed through all stages is positive, the algorithm will conclude that the object exists in the image.

Object Detection Based on Histogram of Oriented Gradient (HOG)

In the year 2015, the Journal of Korea Multimedia System published research that used Histogram of Oriented Gradient (HOG) to detect pedestrians in surveillance videos.[5] The distribution of intensity gradient or edge directions are what object appearance and shape in an image are based on. HOG uses a sliding window approach, where the algorithm slides across the image with a detection window. The detection window will extract the HOG and apply a classifier to determine if the window contains the object that it is looking for. This classification scheme returns a binary result indicating whether the window consists of the object (i.e. 1) or not (i.e. 0).

Another pedestrian detection system was built using Python and OpenCV's built-in HOG feature.[6] The HOG was trained using a linear support vector machine (SVM). To further improve the detection, a non-maxima suppression was used to reduce multiple overlapping detected regions into a single region. As an example, the issue arises when 2 people walking together may result in the region being suppressed into 1 region. Therefore, an overlap threshold will be passed to the non-maxima suppression algorithm to correct such situations and provide better results.

Image Enhancement Methods

Several methods can be applied to enhance or restore an image. Histogram equalization is one of the effective methods to enhance the brightness of the image.[7] This technique involves gathering the intensity value of the image pixels and then distributing the high-intensity values equally across the other pixels. This means that a dark image would be equalized to be brighter and conversely for a bright image.

Another study demonstrated that histogram equalization with the above method would not be effective in images with a large variance in intensity.[8] The default method of applying histogram equalization would affect the entire image pixel intensities as it takes every pixel in the image into its calculations.

An alternative would be to use Contrast Limited Adaptive Histogram Equalization (CLAHE), which is a form of adaptive histogram equalization method. This algorithm divides the image into regions and equalizes the pixels within each region. Therefore, the pixel intensity values are only affected by the region values, converting dark regions to bright regions and vice versa. Once the conversion is complete, a border will exist between each region due to the difference in pixel intensity, therefore bilinear interpolation could be applied to smoothen the edges. Applying CLAHE to an image result in an image with its pixel values more evenly distributed.

Multi-label Image Classification

Multi-label image classification is a type of classification where one image can have multiple classes.[9] This kind of classification is extremely useful when detecting multiple features from an image, e.g. movie genre based on a movie poster. A multi-label classification would use a softmax activation function, where the result for each label would be assigned a probability based on the weightage. On the other hand, a multi-class classification would use a sigmoid activation function, where only a binary result would be returned if the label is associated with the image. The sigmoid converts the values to 0 or 1 independent of what the other scores are, differing from softmax that takes other scores into account when assigning a probability to the labels.

A major challenge for doing multi-label classification is data imbalance, a scenario where classes appear more frequently than others, resulting in biasness when training the model.[9] Several techniques can be applied to overcome data imbalance, such as either upsampling or downsampling. Upsampling increases the dataset of classes that is less compared to others by augmenting existing images or adding new data. Downsampling decreases the dataset by dropping extra data. Both these methods would be a challenge for multi-label classification as one image may belong to 2 or more classes, hence using either method to overcome the data imbalance will affect other classes. For example, a movie poster genre may be drama and action, where dropping this data out due to excessive samples in drama class will reduce the samples for action class.

AlexNet Convolutional Neural Network (CNN)

AlexNet is a CNN architecture built by Alex Krizhevsky and was the winning entry in ILSVRC 2012.[10] AlexNet consists of 5 convolutional layers and 3 fully connected layers. With multiple convolutional layers, interesting features can be extracted from an image. Overfitting is an issue faced by AlexNet, where the model memorizes features rather than learns it, resulting in the model performing extremely well with the existing (training) dataset but fails on unseen test data.

To reduce overfitting, AlexNet architecture developers implemented a few improvements. The first method applies augmentation on the dataset. Mirroring and cropping are 2 types of augmentation that may be applied. Mirroring the image would introduce fresh data that has variance for the model to learn. Cropping an image from a larger image introduces pixels shifting, allowing the network to understand that minor shifts in pixels do not change the object in the image. By applying to crop, AlexNet managed to increase the dataset by the factor of 2048.

Another method used by AlexNet is a dropout. Dropout would drop the neuron from the network, removing it from contributing to either forward or backward propagation. As a result, input would be passed through different network architecture, allowing the weight parameters to be more robust and the model does not get fixated easily. However, dropout would increase the iterations required by a factor of 2.[10]

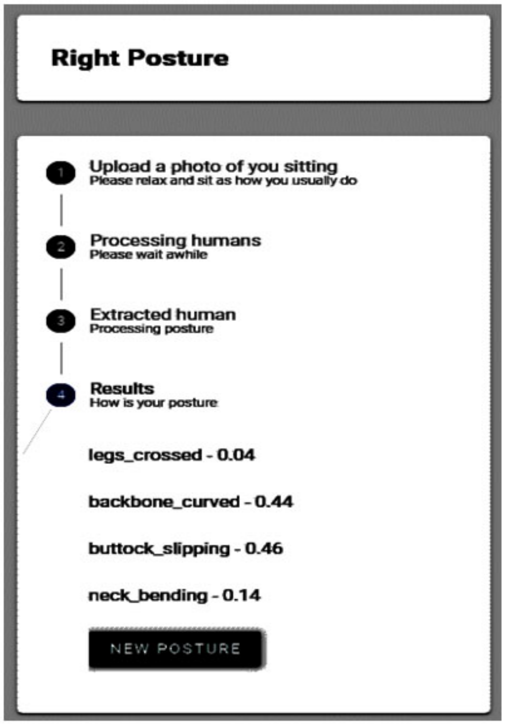
MATERIALS AND METHODS

Tensorflow, a machine learning library was chosen to train the model. It has a large online community with available support documents and examples. Compared to PyTorch, a machine learning library by Facebook, a search on Google regarding building a CNN network architecture will return top results with explanations and tutorials using Tensorflow. Free training videos provided by Tensorflow can also be found on YouTube. Keras, the front-end framework for Tensorflow can be used to accelerate development time.[11] Keras is designed with easy syntax, easy extension to other function plugins and detailed information for debugging. It is actively being developed by tech giants, which includes Google, Microsoft and Amazon.

Apart from Keras, OpenCV will be used to process the input image. OpenCV is a computer vision library that consists of many functions to extract image information and process them. As the project involves processing the image and extracting information about the subject from an image, OpenCV can simplify the process with its built-in functions.

Data gathering is essential for this project as it will serve as the training model input. The expected information from this will be a side view (i.e. profile) image of the participant sitting on a chair. To maintain privacy, the photo taken can exclude the face of the participant, as they may find it sensitive to send images of themselves. Besides allowing the participant to upload images of themselves, the observation method would be used with the participant's consent, where a photo of them sitting would be captured and saved for the model to train on. The photos that are captured and stored would not be used for any other purpose except as input to train the model. Based on the photos and a good sitting posture checklist, the sitting posture would be categorized as bad and good posture.

Figure 1 is a screenshot of the system prototype. The prototype was designed using VueJS, a Javascript framework and the prediction model was trained using Tensorflow. As seen in the figure, four main steps are displayed on the main screen. The first section allows users to upload an image or capture the image with the system. After the first step action is performed, the system will open the second section and preview the uploaded image while it extracts regions with any humans that can be identified in the image. Once the system has managed to extract at least one human, the third section will preview the extracted image while the system starts estimating the posture. Finally, once the system has estimated the values for the posture in the extracted image, a score for each label will be displayed at the bottom, as shown in Figure 1.



Right Posture

1 Upload a photo of you sitting
Please relax and sit as how you usually do

2 Processing humans
Please wait awhile

3 Extracted human
Processing posture

4 Results
How is your posture

legs_crossed - 0.04

backbone_curved - 0.44

buttock_slipping - 0.46

neck_bending - 0.14

NEW POSTURE

Figure 1: Screenshot of system prototype.

RESULTS AND DISCUSSION

Throughout the data gathering period, 90 images of people sitting were gathered and analyzed. 20 images were collected through submission via Google Form, and the remaining images were sourced from videos and posters online. Amongst the 20 images collected from Google Form, 2 of them were blurry and had to be discarded. Apart from that, participants had their image captured according to the guidelines, and therefore 18 of the images that were submitted were usable.

All the images collected were saved on Google Drive as a backup. A local copy was also kept on the machine to be used in training the model. Both sets of images were kept with proper access control permissions delegated to ensure that the images are only accessible by intended users.

A variety of sitting postures were gathered and labelled accordingly, including cases of curved back and crossed legs. Figure 2 illustrates the frequency of each label that was assigned to the data gathered. Out of the 90 samples, 31 people were identified as having a curved backbone, followed by people whose both legs were not grounded. Having legs crossed came in as the third-highest. This label could be associated with a habit that people usually practice from a young. Having bent neck came in as the second last among the other labels. From the images gathered, this feature can be found in participants using their phones or working on a computer. The last label was people with buttocks not vertically aligned with their head, hence not sitting up straight.

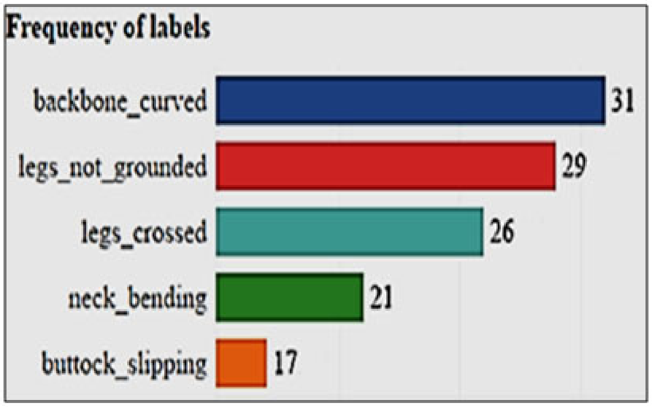


Figure 2: Frequency of labels assigned.

Environment influences of the background like outdoors and indoors, and floor types like carpet, grass or cement, vary among the images. The participants clothing varied from office attire to t-shirts, sports attire and jacket. Apart from the above variable, activities the participants were engaged in while the candid photo was taken included eating, using the phone, studying at a desk and using a computer or laptop.

A large percentage of the images were gathered from online sources such as videos from YouTube and royalty-free images. YouTube videos of chiropractors demonstrating good and bad sitting posture were captured and added into the dataset. Some videos of sitting chair advertisement were also used.

CONCLUSIONS

In developing a sitting posture identifier application, the first consideration was that there were no previous researches readily found regarding the use of camera devices to detect a person's sitting posture. Therefore, data gathering for training the model was necessary. The second factor was to develop a model that has good accuracy with a small dataset. Upon analysis of the dataset, it was found that many factors can influence and adversely affect the results. Examples include a curved back chair may indicate good sitting posture, or perhaps an extremely thick coat may affect the labelled person sitting posture in the image.

This work creates a model where it could serve as a framework or baseline for future development of projects with a similar aim. Real-time monitoring for sitting posture, or pre-assessment of a person's sitting posture could assist doctors and chiropractors in diagnosing patients quicker and more accurately than before. This work would also be beneficial to certain sectors like airlines and kindergarten. The system can be implemented in a long-haul flight, where people often suffer from back pain due to poor sitting posture. The system can be implemented in kindergarten to cultivate good sitting habits in kids.

ACKNOWLEDGEMENTS

The authors also wish to express gratitude to the management of Asia Pacific University of Technology & Innovation (APU) for their support.

Conflict of Interest

The authors involved in the current study does not declare any competing conflict of interest.

Funding and Sponsorship

No fund or sponsorship in any form was obtained from any organization for carrying out this research work

REFERENCES

1. Barba R, de Madrid ÁP, Boticario JG. Development of an inexpensive sensor network for recognition of sitting posture. Int J Distrib Sens Net 2015;11(8):969237.
2. Cha Y, Nam K, Kim D. Patient posture monitoring system based on flexible sensors. Sensors. 2017;17(3):584.
3. Suhail A, Jayabalan M, Thiruchelvam V. Convolutional Neural Network Based Object Detection: A Review. J Crit Rev 2020;7(11):2020.
4. Soo S. Object detection using Haar-cascade Classifier. Institute of Computer Science, University of Tartu. 2014;2(3):1-2.
5. Nguyen TB, Nguyen VT, Chung ST. A Real-time Pedestrian Detection based on AGMM and HOG for Embedded Surveillance. *J Korea* Multimed Soc 2015;18(11):1289-301.
6. Chavan A, Yogamani SK. Real-time DSP implementation of pedestrian detection algorithm using HOG features. In 2012 12th International Conference on ITS Telecommunications. 2012;5: 352-355.
7. Acharya T, Ray AK. Image processing: principles and applications. John Wiley & Sons; October 2005
8. Saalfeld S. Contrast Limited Adaptive Histogram Equalization. https://imagej.nih.gov/ij/plugins/clahe/index.html
9. Cevikalp H, Benligiray B, Gerek ON. Semi-supervised robust deep neural networks for multi-label image classification. Pattern Recogn 2020;100:107164.
10. Landola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv Preprint ArXiv. 2016; 16(2):73-76.
11. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neur Inform Process Syst. 2012;25:1097-105.