Human Body Poses Recognition Using Neural Networks with Data Augmentation

Ahmad al-Qerem[1] Arwa Alahmad [2]
[1]Zarqa University, Jordan, ahmad_qerm@zu.edu.jo
[2] PSUT, Jordan, arwa-alahmad@hotmail.com

ABSTRACT

The continuous development of technology of IT enabled computers to see and learn. There are many viable applications for computer learning and vision to solve new tasks. In this paper, we propose a framework, able of automatically perceiving the human body poses from a single image, acquired by a traditional low-cost camera. Our methodology exploiting the vision of computers features and neural networks to detect and recognize a human from an image. The processing start with detection human in image and then extracting the silhouette from an image then using a neural network to recognize body poses based on silhouettes that extracted. To classified the detected silhouettes with body poses, the neural network was trained with dataset of preprocessed images contains silhouettes and labelled using 0, 1 for standing and walking poses respectively labelled , according to our outcomes that resulted in an accuracy over 93% with the best neural network model. The proposed approach offers promising effects with accepted accuracy. This knowledge will help greatly in decision-making and provide independent vehicles with valuable information for their artificial intelligence to process and make decisions as regards the movement of pedestrians.

Key words: Computer vision, machine learning, human detection, silhouette extraction, human body pose classification

1. INTRODUCTION

Recognition of pedestrian has been achieved great progress in line with the progress of deep learning Which has been exploited in several areas to achieve satisfactory results[1] This issue is very essential for future intelligent systems and autonomous vehicles, detection of pedestrian not only important but also recognize wherein poses the pedestrian is such as standing, walking, running, etc Since this valuable information help greatly in decision-making further, Hence our contributions for this work, Building framework that can automatically detect human body pose from single image that taken with a low-cost camera by utilizing multiple methods. showing that chaining more than one techniques together to create a framework for learning classifying neural networks can achieve good accuracy by given reasonable dataset for training, also development of scientific methods that solve human body pose detection problem and therefore solve many real-world problems. The remainder of this paper is organized as follows : Section 2 is the Background and Related Work that gives an overview of popular algorithms utilized in image processing and human detection. Section 3 presents the proposed approach for human body poses recognition .Section 4 presents the experimentation results finally conclusion and future work.

2. RELATED WORK

This section discusses the previous work conducted with the goal to detect and recognition of human body pose. Many researching over the year have been innovating different applications to people detection and recognition [2] [3], also, the focus was on recognizing human body poses such as walking or standing. Model was presented to recognition human body poses and lateral speed in [4] Another approach presented in [5] based on HOG and SVM to performing a simple Classification of human body pose. In addition to using the SVM to recognition ,recently many papers presented machine learning and deep learning methods to the computer vision problems and for the recognition. where the authors in [6] presented two methods for human body pose recognition based on deep learning using a convolutional neural network (CNN) and optimized neural network with similarity evaluation between dimensions for mapping human body poses to the right estimation. Human body pose recognition task solving by deep learning in [7] using binarized normed gradient features to extract objectiveness using saliency map with CNN that learnt the hierarchies of the features and made predictions based on learnt.

A. Computer vision

The Computer vision continues growth and development and exploitations in useful application Over the years , The term "computer vision" was originally formulated to describe the overall goal of enabling computers with connected cam- eras to see images intelligently and to recognize the same things as human do, it deals with the extraction of significant facts from the contents of images. Many applications founded in various fields that applicable the concept of computer vision and are now part of our daily lives , There are many real-life application that one of the most essential is object detection[8]. In [9] the authors present a novel detection method for automatic defective apple using computer vision system combining with automatic lightness correction

B. Human detection

Human detection is one of the great importance problems of computer vision, is often described as a complicated task to be solved. Building and training model based on human characteristics like dimensions and silhouette is considered the solution to identify human in an image. Primarily, the most of all strategies are based on appearance characteristics[10][11].The main common thing among previous work is the diversity in the angles taken into consideration as a focal point of the research. model used is the based for different methods depend on appearance. According to the work conducted in [12] The simplest of methods presented that only deal with a region or shape, More complex model was launched in [11]

C. Machine learning

Machine Learning techniques that used in computer vision is specifically fall in old categories that use SVM or RF methods and newer neural network methods

3. PROPOSED APPROACH

In this Section, the proposed approach for human body poses recognition is discussed. The main block of this proposal involves training a Convolutional Neural Network on classifying human body pose based on silhouette. Detecting human and Extracting human silhouette from image being employed as well. We use Convolutional Neural Network CNN with GRAZ-01 date set to automatic human body poses recognition.
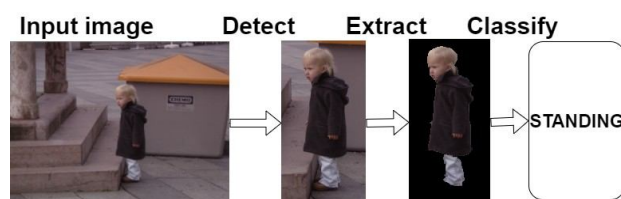

Figure 1: Methodology of the proposed approach

The proposed framework can be divided into three main parts as following:
• Detecting human : To detect human from images we using pertained HOG and linear SVM model that deployed through an OpenCV package.
• Extracting human silhouette: GrabCut algorithm implemented in Python used to extract foreground in images and cut out a human silhouette from the image.
• Classifying human body pose: convolutional neural net- work is created using TensorFlow with Keras library to training and testing neural networks to classify human body poses with acceptable accuracy.

The first and second part recognize and cut out silhouettes from images, which creates input training data for the third part which is classification.

A. Detection

We depend on OpenCV package, that includes with pre-trained HOG and linear SVM method . HOG is a feature descriptor used to detect objects in computer vision and image processing was suggested in [13] HOG is proven efficient in capturing and detect objects. Using a linear SVM with HOG image descriptor could be train highly accurate object classifiers. OpenCV provides pertained linear SVM that has been trained with HOG descriptors to detect humans from images. The bounding boxes is the result found using HOG detection will be used in next step for extracting the silhouettes of the humans.

B. Extraction

The GrabCut algorithm is an extension of the graph-cut algorithm that proposes a more powerful and iterative version of the optimization. Hence, in our case we used GrabCut algorithm [14] for human silhouette extraction where the rate of its results is good. After detection phase, Once we detect that the human in image, we can be divided image into two parts foreground and background. The foreground is represented by the bounding box that get it from the human detection phase. The background is represented by outside part of the bounding box. GrabCut utilizes the region of interest rectangle coordinates x1, y1, x2, y2 and the times to run the algorithm the 10 iterations is used to getting best results without giving away too much performance. GrabCut compares the colors and structure of background to try to remove background elements from foreground and keep on human silhouette only onto the image to work with.

C. Classification

The process of detecting of the human and extracting the silhouette from the images are applied to dataset in previous two phase. In this papers, two poses are classified standing and walking. The convolutional neural network is a sequential network containing multiple layers which is the best suited for image processing tasks so we use it to evaluate

our model to tackle more accuracy. We resize input silhouette images to a size of 64x64 pixels and make the RGB images into grey-scale images, data augmentation is used to improve our model to achieve more accuracy by increasing dataset. The process of building a Convolutional Neural Network involves convolution, pooling, flattening and Full connection steps where convolution on the training images perform by first convolutional layer, pooling operation perform by a pooling layer using a max-pooling function. Max-pooling is used because for each region of interest, the maximum pixel value is needed, flattened from the 2-dimensional array into a 1-dimensional array perform by flattening layer.

D. Data Augmentation

Getting a significant increase in validation accuracy when we training deep networks to classify images we augment the data to prepare the training set as well as the test set for each individual silhouette by randomly transform the images in certain ways using different methods of cropping, resizing. The following processes is applied to data augmentation:

- Flipping - horizontally with 50% probability
- Cropping image randomly
- Blurring - adding random blur factor between 0 and 0.5 with 50% probability.

4. EXPERMENTAL RESULT

An experimental testing environment is established to evaluate our methodology .In this section we describe the experiments performed to evaluate our methodology .We use classic popular people dataset GRAZ-01 [15] to apply the methods. GRAZ-01 Is a popular data set consisting of images often used in research papers dealing with human detection. The data set contains a diversity of mixed people with various situations that providing a good state of training and testing of computer vision algorithms. Each image in the data set goes through the first two steps: detecting and extracting a human silhouette. at the end the processed dataset that consist of human silhouettes images used for training and testing using multiple neural networks where the output silhouettes images classified manually standing or walking based on the human poses. Around 226 silhouettes used to training and testing the neural networks to classify two poses standing and walking. The silhouettes dataset respectively contains 125 images of standing pose and 101 images of walking pose, 80/20 ratio used to split training and test datasets where 226 images are divided 180 images for training and 46 images for testing the neural network. In general, dataset of 226 images is too small to be able to correctly train a neural network, data augmentation is used to make the dataset more diverse. The data augmentation applied using

Various processes such as horizontally flipping, cropping image randomly, Blurring image ,contrast strengthening or weakening in image, Gaussian noise by applying only noise per pixel or applying noise per pixel and channel, Making image darker or brighter by random factor, Change image color and image affine transformations using scaling , zooming, translating, moving or rotating.

A. Result

To achieve different results and holding comparisons, multiple neural network have been tested for that. Two datasets were used for testing, 226 images of raw dataset and 2260 images of augmented dataset. The simple Neural Network with hidden layer and CNN with added convolutional layer to simple Neural Network are the neural network structures were used for testing , also we used the CNN to detection human poses without the previous detection and extraction phases just use one method to compare with other result to prove that chaining multiple methods together can achieve good accuracy with given reasonable dataset. .

The results appear neural network that training and testing with augmented data performs better than simple neural network with raw data set, upon validation the accuracy difference does not range greater than 5% between these models. The CNN model outperforms other two models when trained with augmented data as we can see in table 1. Upon validation the CNN achieve an accuracy of 93-97% according to epoch's number. Raw data with simple neural network and Augmented with simple neural network don't manage to get over 90% even after 25 epochs of training. Difference between accuracy of simple neural network and convolutional neural network is at some points as high as 10%.The results also show that the use of one method just use CNN to detection human poses achieve an accuracy between 52% and 86% with raw data and achieve an accuracy between 54% and 79% with data augmentation , much less than the use of chaining multiple methods together.

Table 1: TRAINING RESULTS USING RAW DATASET AND AUGMENTED ATASET

Models	Raw DataSet	Augmented dataSet
Simple NN	81-83%.	85-86%.
CNN	-	93-97%
Only CNN Method	52-86%	54-79%

5. CONCLUSION

In this paper, we presented approach and techniques based a neural network with class-based data augmentation for detecting and recognizing human body poses. The proposed methodology aims to provide efficient detection and recognition approach capable of dealing with detection of the human, extraction of the human silhouette and classification and recognition of the human body pose, it

produces accuracy over 90% results by exploiting the features of pre- trained SVM for human detection, GrabCut algorithm for human silhouette extraction and using a CNN that trained with an augmented dataset consisting of 2260 silhouettes

In the future, the intelligence is increased by more training data feeding into the system. so we will increase the size of the dataset and classifications that will inevitably result in human body pose recognition accuracy. The ultimate goal is to enhance the automation of human body poses recognition detection using multi techniques and use Deep learning algorithms that is more efficient and have a larger area in the future.

REFERENCES

1. Angelova, A., Krizhevsky, A., Vanhoucke, V., Ogale, A., Ferguson, D. (2015). Real-time pedestrian detection with deep network cascades.
https://doi.org/10.5244/C.29.32

2. Zhao, L., Thorpe, C. (1999). Stereo-and neural network-based pedes- trian detection. In Proceedings 199 IEEE/IEEJ/JSAI International Con- ference on Intelligent Transportation Systems (Cat. No. 99TH8383) (pp. 298-303). IEEE.

3. Cheng, S. Y., Trivedi, M. M. (2006). Turn-intent analysis using body pose for intelligent driver assistance. IEEE Pervasive Computing, 5(4), 28-37.
https://doi.org/10.1109/MPRV.2006.88

4. Hariyono, J., Jo, K. H. (2017). Detection of pedestrian crossing road: A study on pedestrian pose recognition. Neuro computing, 234, 144-153.
https://doi.org/10.1016/j.neucom.2016.12.050

5. Gepperth, A., Ortiz, M. G., Heisele, B. (2013, October). Real- time pedestrian detection and pose classification on a GPU. In 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013) (pp. 348-353). IEEE.
https://doi.org/10.1109/ITSC.2013.6728256

6. Jammalamadaka, N., Zisserman, A., Jawahar, C. V. (2017). Human pose search using deep networks. Image and Vision Computing, 59, 31-43.
https://doi.org/10.1016/j.imavis.2016.12.002

7. Xia, D. X., Su, S. Z., Geng, L. C., Wu, G. X., Li, S. Z. (2017). Learning rich features from objectness estimation for human lying- pose detection. Multimedia Systems, 23(4), 515-526.
https://doi.org/10.1007/s00530-016-0518-5

8. Arzoo Lakhwani, Krupali Shah, Ankita Vaghela, Disha Panchal,and Saurabh Rathod. Review on basics of computer vision and itsapplica- tions. Journal of Computational Biology, 6(2):3340, 2017

9. Zhang, B., Huang, W., Gong, L., Li, J., Zhao, C., Liu, C., Huang, D.(2015). Computer vision detection of defective apples using au- tomaticlightness correction and weighted RVM classifier. Journal of Food Engineering, 146, 143-151

10. Vezzani, R., Cucchiara, R. (2008, September). Annotation collec- tionand online performance evaluation for video surveillance: The visorproject. In 2008 IEEE Fifth International Conference on Ad- vancedVideo and Signal Based Surveillance (pp. 227-234).
https://doi.org/10.1109/AVSS.2008.31

11. Wu, B., Nevatia, R. (2005, October). Detection of multiple, par- tiallyoccluded humans in a single image by bayesian combination of edgeletpart detectors. In Tenth IEEE International Conference on ComputerVision (ICCV05) Volume 1 (Vol. 1, pp. 90-97).

12. Xu, F., Fujimura, K. (2003, July). Human detection using depth andgray images. In Proceedings of the IEEE Conference on Advanced- Video and Signal Based Surveillance, 2003. (pp. 115-121).

13. Dalal, N., Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In international Conference on computer vision Pattern Recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE Computer Society.

14. Rother, C., Kolmogorov, V., Blake, A. (2004, August). Grabcut: Interactive foreground extraction using iterated graph cuts. In ACM transactions on graphics (TOG) (Vol. 23, No. 3, pp. 309-314). ACM.
https://doi.org/10.1145/1015706.1015720

15. Graz-01 persons dataset. http://www- old.emt.tugraz.at/ pinz/data/GRAZ01/persons.zip.Accessed $2017 - 12 - 01$.