HUMAN ACTIVITY RECOGNITION

Deepthi Nayak*

Student, Department of MCA, NMAM Institute of Technology, Karkala, India

*Corresponding author: nayakdeepthi20@gmail.com

Abstract –*Action detection is still closely related to the computer vision problems of pose estimation, frame tagging in movies, object identification, picture retrieval and video activity recognition. Recognizing someone's behavior or conduct from just one frame is the goal of this assignment. For motionless photos, there are no spatial-temporal properties, which makes the issue more difficult. In contrast to action detection in movies, a similary established field of study where they are utilised, still image action detection is a much more recent field. The current work only considers object-related activities. Complex actions can be dissected into their constituent elements thanks to semantics. Each of these factors' contributions to the recognition of actions are carefully investigated.*

Keywords *– Faster RCNN, image processing, object detection, tensorflow, deep learning*

---

i. Introduction

In the subject of still-based image recognition, it is extensively known and studied. Due to the recent rise in the number of images uploaded on social networks, it has drawn a lot of attention. Activity detection in still images is still a difficult topic to solve given that motion cannot effectively measured from a still image and because spatiotemporal information can't be applied to characterise the action. Although it is simpler and more logical to detect activity in moving images, it is nevertheless possible and extremely useful to detect it in static images.

A range of action types can be effectively depicted in a single image (without movement or a video signal), and these actions can be appropriately understood based on human sight. One frame is enough to adequately capture these action types' actions. This knowledge supports the creation of computer-based techniques for the automatic analysis and recognition of activity in still photos. The primary difficulties of perception of activity in still images include loss of spatiotemporal characteristics, backdrop clutter, high intra-class variance and low inter-class variance among certain action classes, change in background lighting, and variation in human attitude. The spatiotemporal properties of actions in videos are the most important aspect to describe. Images lose the temporal information, which makes describing an activity much more challenging.

One frame is enough to adequately capture these action types' actions. This knowledge supports the creation of computer-based techniques

for the automatic analysis and recognition of activity in still photos. The primary difficulties to identifying activity in static photos include loss of spatiotemporal characteristics, backdrop clutter, high intra-class variance and low inter-class variance among certain action classes, change in background lighting, and variation in human attitude. The spatiotemporal properties of actions in videos are the most important aspect to describe. Images lose the temporal information, which makes describing an activity much more challenging. Object recognition, on the other hand, is a prerequisite for action acknowledgment. When doing still image-based action detection, store-bought object detectors include those frequently employed to record the names of all the item classes seen in the image. Then, a feature is created using a model of their co-occurrence. Action recognition in still images is a closely linked task to scene analysis and pose estimation in computer vision. Numerous articles have used the subject's pose estimation and the action's setting as inputs to action recognition. Pose and scene details are used to train the action recognition model with some degree of accuracy. Using the trained still picture action recognition model as an input, a more accurate posture estimate and scene understanding model is then developed. One job is then used to increase the efficacy of another, creating a cycle. Using the trained still picture action recognition model as an input, a more accurate posture estimate and scene understanding model is then developed. One job is then used to increase the efficacy of another, creating a cycle.

ii. Literature survey

- The most often used high-level indicators for detecting human action in still images are human body parts, action-related objects, interactions between humans and other objects, and the full context or surrounds. Numerous behavioural characteristics are described by these measurements. [1]
- Wang et al. used the rough overall outline of the human body in the photograph. The form was represented by a group of edge points gathered with the aid of a clever edge detector [2]. To organise and categorise photographs into distinct groups, the shape is employed as a feature [3].
- Recognizing activities also requires the ability to read body language. Ikizler et al"Conditional .'s Random Field (CRF)" was utilised to create deformable models utilising the body posture that was derived from images using edges and region data.
- Yao et al. [5] used a variation of random forests to locate the helpful, discriminative patches from the human body area for action recognition. Another graphical depiction of essential patch information is a saliency map [6].
- • Yao et al. [7] made use of a partial model made up of objects and human postures. The objects that are connected, for instance, are either person- or scene-related and show a person riding a bike. The qualities are verbal justifications for how people behave. The artwork consists of both human and inanimate object

positions. Modeling activities in still images are built using attributes and components [8].

- Le et al. partition input photos into recognisable objects in [9], after which they apply a language model to identify every possible action that might be taken with the objects when they are employed in different combinations. Some techniques define the co-occurrence of items to produce actions, while others blend scene data with object data for activity recognition. For the purpose of action recognition, Desai et al. [10] made use of contextual information, such as the item arrangement their discriminative models [11] identified.
- Since digital images are the proverbial "water from the same pool," computer vision techniques and digital image processing techniques inevitably overlap[13].
- A digital picture is processed to create a different, desired image by applying techniques like noise reduction, detail augmentation, or filtering. For instance, imaging methods may be used to improve a blurry image of a car registration plate to create a clear picture of the same, allowing the police to identify the automobile's owner[16]. Contrarily, computer vision uses the same digital imaging methods to process a digital picture in order to analyse and comprehend what it represents[14].
- According to the majority of the studied literature, HAR is used in healthcare systems that are placed in residential settings, hospitals, and rehabilitation facilities. HAR is frequently used to track the activities of elderly patients residing in rehabilitation facilities for the management and prevention of chronic diseases [15].

iii. Methodologies

A unique dataset with five classes is created since the current inquiry is only interested in behaviours involving objects. Images for this dataset came from a variety of sources. Photos for a few action classes were directly chosen from datasets like the Stanford 40 Action dataset [12] and the Willow dataset [13] when there was only one person doing an action with an object. The Google search engine also provided some additional images. The 200 images in the special collection depict five various motions, including standing, sitting, sleeping, ascending stairs, and descending.

Fig1: custom data-set collection

The data sets are pre-processed and normalised in fig 2 before being labelled with the "labelImg" tool in accordance

with their categorization. Following feature extraction, the images go through training where they are classified based on their features. Testing is conducted after training is finished, and the output is then shown using the anaconda prompt.

Google Colab and Anaconda IDE were employed in the development of this system. One of the best IDEs for creating systems like these is Google Colab's Anaconda. To write and run code in a single window, use this tool. Python has been used to create this project. Python has been utilised because it is a fairly adaptable language with a wide range of library support, which makes programming easier.

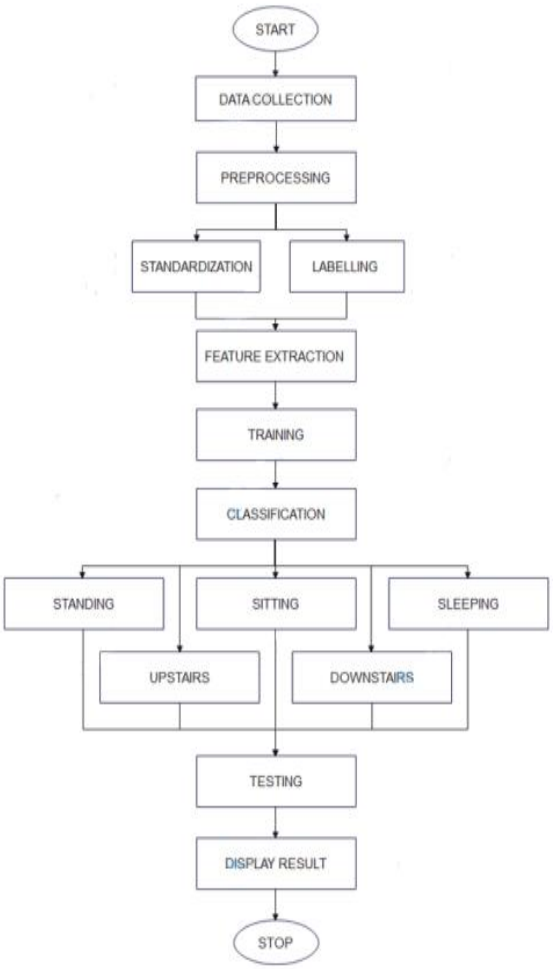

Fig 2: flow diagram

iv. Proposed solution

Working of RCNN:

Step 1: After running a region proposal method similar to a selective search on the input image, we receive 2,000 candidate region proposals in the image that need to be evaluated.

Step 2: We will warp each potential region into a set size, say, because region suggestions can have varying sizes and aspect ratios (224x224)

Step3: We will individually process each warped image region via a Convolutional Neural Network (CNN), and the CNN will produce a classification score for each of these regions.

We contrast these bounding boxes using the Intersection over Union metric (IOU).

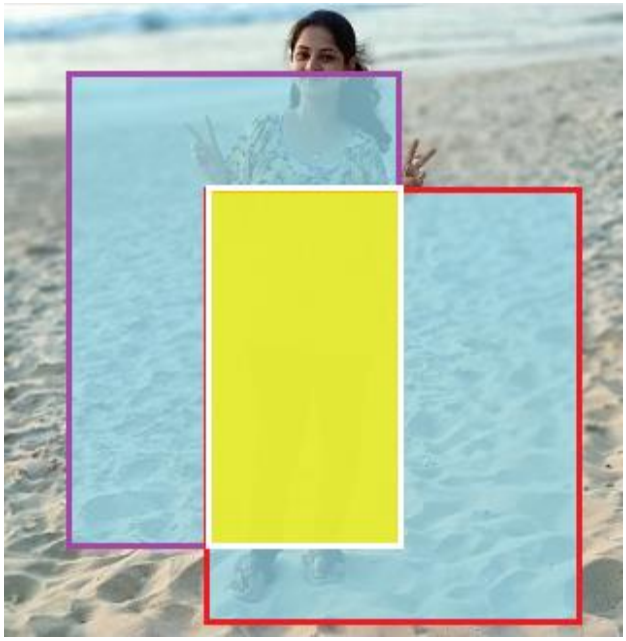

Fig3: Intersection over union

Intersection over union(IOU)
= size of ▭ / size of ▭

If IOU >= 0.5, it is "correct."

Area of Intersection = IOU (Area of Union)

The measure of overlap between these two bounding boxes in Fig. 3 is called IOU.

IOU < 0.5 is considered "BAD"

IOU > 0.5, we deem it deplorable.

IOU>0.7 is considered "GOOD"

IOU > 0.9 is considered "ALMOST PERFECT."

v. Output

The evaluation of each combination of component attributes for classifying actions is shown in the table below. These classifications are reported to be accurate.

Fig4 . Downstairs detection

Fig5 . Sitting detection

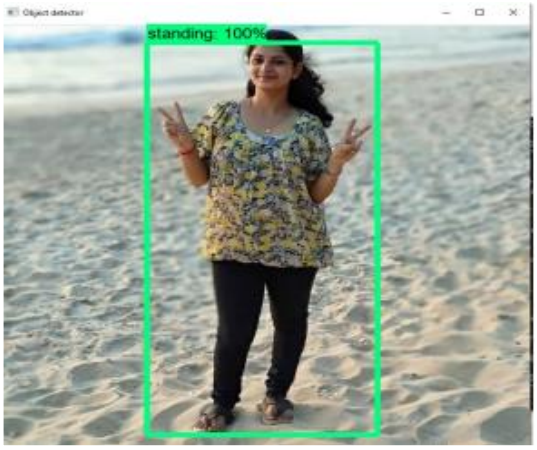

Fig6 . Standing detection

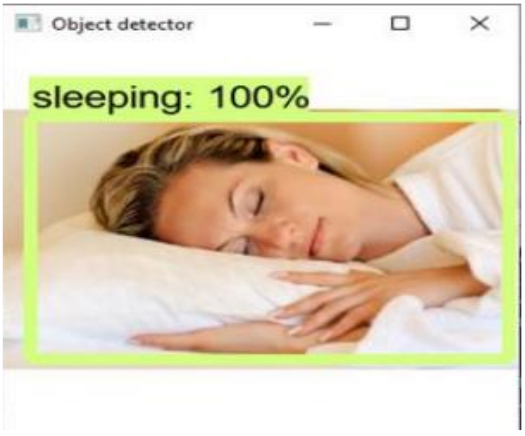

Fig7 . Sleeping detection

Fig8 . Upstairs detection

vi. Conclusion and Future scope

Recognition of human activity has numerous uses in human survey systems and medical research. We created a recognition system for five human activities in this project: standing, sitting, sleeping, going upstairs, and going downstairs. Mobile devices and other social media platforms are used to gather the information or photographs. They are then trained to produce reliable findings after being pre-processed, labelled, and classified according to their category. Here, a faster version of RCNN has been employed as the algorithm.

Future implementation may include the following:

- Putting in place a real-time system and employing mobile devices

References

1. Anca Ralescu and Koji Miyajima. The representation and identification of basic spatial relations in 2D segmented pictures. 225–236 Fuzzy Sets and Systems, 65(2-3):1994,
2. 2. Greg Mori, Yang Wang, Hao Jiang, Mark S. Drew, and Ze-Nian Li. Action class discovery without supervision. Volume 2, pages 1654–1661 of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, 2006.
3. John Canny, an edge detection method using computing. IEEE Transactions on pattern analysis and machine intelligence, vol. 6, no. 6, 1986, pp. 679–698.
4. Pinar Duygulu, Selen Pehlivan, R. Gokberk Cinbis, and Nazli Ikizler. identifying actions in still pictures. Pages 1-4 of the 2008 19th International Conference on Pattern Recognition. IEEE, 2008.
5. Li Fei-Fei, Aditya Khosla, and Bangpeng Yao. utilising a combination of randomness and discrimination to classify images finely. Pages 1577–1584 of CVPR 2011. IEEE, 2011.Gaurav Sharma, Fred´ eric Jurie, and Cordelia Schmid. Dis- ´ criminative spatial saliency for image classification. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 3506–3513. IEEE, 2012.
6. Leonidas Guibas, Andy Lai Lin, Bangpeng Yao, Xiaoye Jiang, and Li Fei-Fei. Recognizing human action through studying the components and features of action. Pages 1331–1338 in: 2011 International Conference on Computer Vision. IEEE, 2011.
7. Alice Lai and Guodong Guo. a study on recognising human action from still

images. 2014;47(10):3343–3361. Pattern Recognition

8. Jasper Uijlings, Raffaella Bernardi, and Dieu Thu Le. using language models to detect hidden activities 2013; pages 231-238 in Proceedings of the Third ACM Conference on International Conference on Multimedia Retrieval.
9. Charless Fowlkes, Deva Ramanan, and Chaitanya Desai. Discriminative models for interactions between static humans and objects. Pages 9–16 in Computer Vision and Pattern Recognition-Workshops, 2010 IEEE Computer Society Conference. IEEE, 2010.
10. Charless C. Fowkes, Deva Ramanan, and Chaitanya Desai. Models that discriminate for multi-class object arrangement. 95(1):1–12 in International Journal of Computer Vision, 2011.
11. Joint pose estimation and action identification in picture graphs. Kumar Raja, Ivan Laptev, Patrick Perez, and Lionel Oisel. Pages 25–28 of 2011's 18th IEEE International Conference on Image Processing. IEEE, 2011.
12. J. Sivic, I. Laptev, and V. Delaitre. Identifying human actions in still images: a comparison of part-based versus bag-of-features representations. Available at http://www.di.ens.fr/willow/research/stillactions/, this version was modified in 2010.
13. L. Wang, Y. Qiao, and X. Tang. Action recognition with trajectorypooled deep-convolutional descriptors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 4305–4314, 2015.
14. un X, Chen C, Manjunath BS , ―Probabilistic motion para-meter models for human activity recognition,‖ In: Proceedings of 16th international conference on pattern recognition, pp 443–450.
15. U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in Int. Workshop on Wearable and Implantable Body Sensor Networks, (Washington, DC, USA), IEEE Computer Society, 2006.
16. M. Berchtold, M. Budde, H. Schmidtke, and M. Beigl, "An extensible modular recognition concept that makes activity recognition practical," in Advances in Artificial Intelligence, Lecture Notes in Computer Science, pp. 400–409, Springer Berlin / Heidelberg, 2010.