International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 717-724

Human Activities Recognition Using OpenCV and Deep Learning Techniques

***Kalam Swathi*[1],**
[1]*Assistant Professor, Department of Computer Science and Engineering*
[1]*Vignan's Institute of Information Technology (A), Vishakhapatnam, AP, India,*
swathi.kalam@gmail.com,
***J.Nageswara Rao*[2],**
[2]*Sr.Assistant.Professor, Dept .of Computer Science and Engineering,*
[2]*Lakireddy Bali Reddy College of Engineering (Autonomous), Mylavaram, Krishna Dt, 521230, Andhra Pradesh, India,*
nagsmit@gmail.com,
***M.Gargi*[3],**
[3]*Assistant.Professor, Department of CSE,*
Vignan's Lara Institute of Technology & Science, A.P, India
[3]*gargilucky@gmail.com,*
***Khadri Lalitha VaniSri*[4],**
[4]*Assistant.Professor, Department of CSE,*
Vignan's Lara Institute of Technology & Science, A.P, India,
lvsri.khadari@gmail.com
***B. Shyamala*[5]**
[5]*Assistant Professor, CSE, GST, GITAM Deemed to be University, Bangalore, Karnataka, India,*
shyamala.b501@gmail.com

Abstract

Protecting privacy from hidden video is an important social issue. We need a computer vision system (for example, a robot) that can recognize human activities and provide help in our daily lives, but at the same time, make sure that it does not record videos that may violate our privacy. This article describes the basic method for resolving such conflicting tasks: Recognizing human activity when only anonymous low-resolution video (eg 16x12) is used. We will introduce deep learning, CNN, and OpenCV Paradigm, whose concept is to teach the best image conversion set to create multiple low resolution (LR) / high resolution teaching video concepts from video. Our concept is studying various types of sub-pixel conversion optimized for activity classification, which allows classroom experts to take advantage of existing high-resolution videos (such as YouTube videos) by creating some LR training videos suitable for this problem. We have verified through experiments that OpenCV in the computer paradigm can benefit from activity recognition of extreme human activity videos.

Keywords: Privacy, High/Low-Resolution, CNN, computer vision, OpenCV, DNN

1. Introductions

Deep Neural Networks (DNNs) are typically Feed Forward Networks (FFNNs) in which data streams from the information layer to the yield layer without going backward and the associations between the layers are one way which is the forward way and they never contact a center point again

By performing directed learning based on a data set that back propagates some "we want" information, the output can be obtained.
Essentially as you go to a bistro, food pros will give you musings in regards to your banquet things. FFNN works a comparative way you get certain machines when you eat, yet in the wake of eating you neglect what you eat. If the culinary ace again gives you a practically identical fixation procedure, you won't have the choice to pick the strategy for fixation of perception and you can start with no arranging, since you don't have a memory about it. Notwithstanding, human thinking can't work thusly.

AI is a subset of computerized reasoning, and “Deep learning is a significant piece of its more extensive family, including profound neural systems, profound conviction systems, and intermittent neural systems [1]”. Mainly, in profound learning, the three fundamental models of neural systems perform well on various kinds of information, for example, FFNN, RNN and CNN.
Significant learning estimations can achieve remarkable precision in PC vision endeavours, including picture request, target revelation and division.

2. Background Work

Cameras are becoming more common. Many reconnaissance cameras record people's daily behaviour in open spaces, and people use wearable cameras that record their lives (such as GoPro and Narrative Clips) to obtain countless vain records. In the same way, the robots in broad daylight are equipped with many cameras to work and communicate. At the same time, so many cameras will also cause a huge social problem: protection from harmful records. We believe that camera frames (for example, robots) can perceive important occasions and help people live day after day by understanding his / her videos, but we also believe that this will not ignore the protection of different customers. This triggered two conflicting goals. More precisely, we believe that (1) gave up using the camera frame to obtain visual information point by point. Ideally, the information contains device data, which may contain a single data (for example, the appearance of a person). At the same time, we need (2) to make the framework record point-to-point data, and this can be reasonably expected from its video so that it can understand the main content and latest developments of observation, recording and smart management.

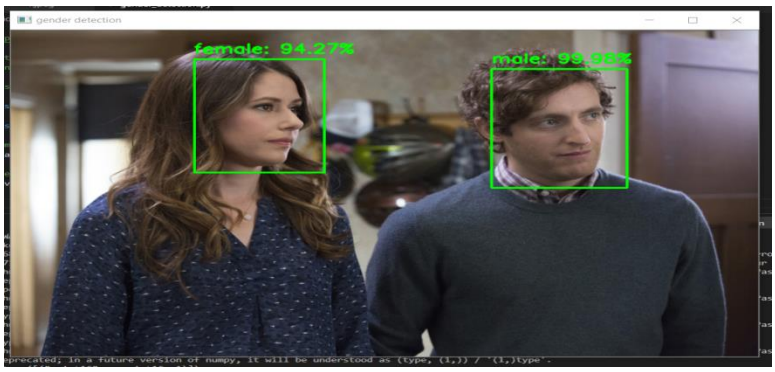
Previously, research has been conducted to meet these social needs. Temple man et al. (2014) studied the method of recognition through images taken by wearable cameras, and found places where privacy needs to be protected (such as toilets). This will cause the device to automatically shut down in sensitive locations. It can also be said that restricting the device to only process / transmit information about functions (such as HOG and CNN) instead of visual data will make it protect privacy. However, recent research on functional "visualization" (Vondrick et al., 2015) shows that it is actually possible to recover a large amount of visual information (i.e. images and videos) from these objects. In addition, all of the above methods are based on software processing of high-resolution source videos (which may already contain sensitive data), and these raw videos cannot be captured by cyber-attacks.

A more basic solution for building a confidential vision system is to use anonymous video. A typical example of anonymous video is video. (Butler et al., 2015) The direction is not to receive high-resolution video and try, but to receive anonymous video. The idea is that if we can use only such anonymous videos to develop reliable computer vision methods, then we can identify them while maintaining confidentiality. Such a concept can even make the camera

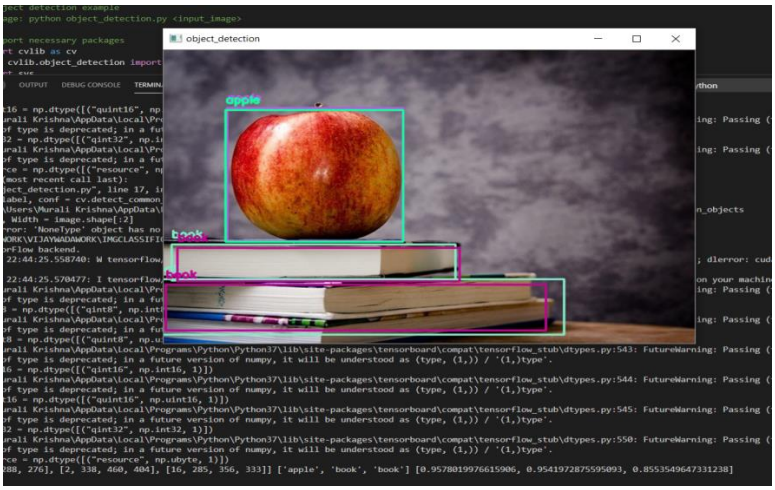
ISSN: 2233-7857 IJFGCN

choose its resolution intelligently. He only uses high-resolution cameras when necessary (for example in an emergency), which is determined based on an extreme analysis of low-resolution video.

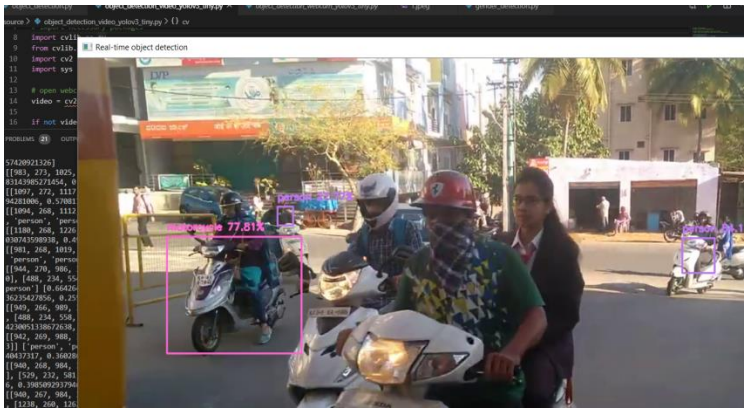
Previous attempts have been made within the framework of this paradigm (Dai et al., 2015). The traditional method is to adjust the size of the original training video to match the target resolution, so that the training video looks like a tested video.



A



B



C

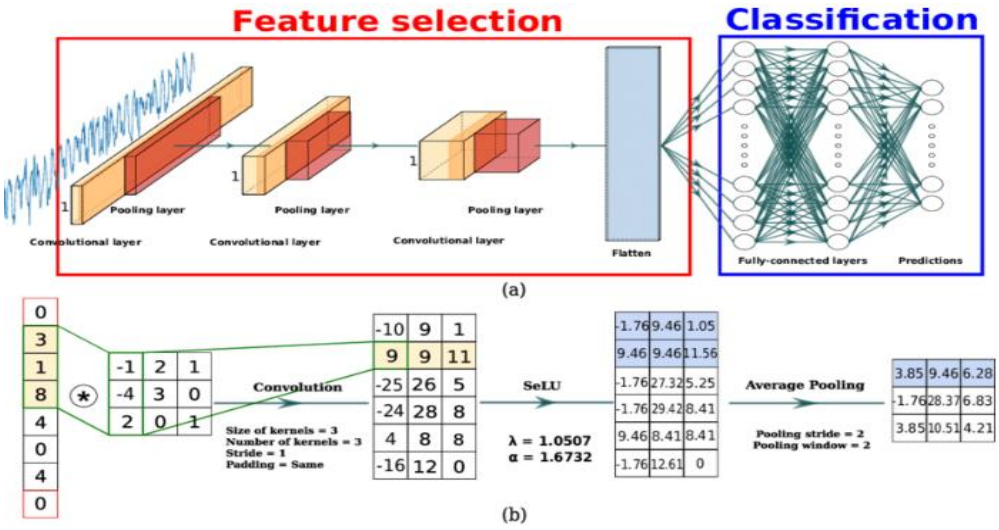
Figure 1: a.Gender Detection, b. Object Detection, c. bject_detection_video_yolov3_tiny

3. Methodology

A Convolutional Neural Network (CNN) is specific type of neural network. In this paper, we choose to simplify its presentation by considering that it can be decomposed into two parts: a

feature extraction part and a classification part Features selection aims at extracting information from the input to help the decision-making. To select features, a CNN is composed of n3 stacked convolutional blocks that correspond to n2 convolutional layers (denoted γ), an activation function (σ) and one pooling (denoted δ) layer [ON15]. This feature recognition part is plugged into the classification part of n1 Fully-Connected (FC) layers (denoted λ). Finally, we denote s the softmax layer (or prediction layer) composed of |Z| classes. To sum up, a common convolutional network can be characterized by the following formula: s ◦ [λ] n1 ◦ [δ ◦ [σ ◦ γ] n2] n3.

Convolutional layer the convolutional layer performs a series of convolutional operations on its inputs to facilitate pattern recognition (see Figure 1-b). During forward propagation, each input is convoluted with a filter (or kernel). The output of the convolution reveals temporal instants that influence the classification. These samples are called features. To build a convolutional layer, some model hyper parameters have to be configured: the length and number of kernels, pooling stride and padding. • Length of filters – Kernels are generated to identify features that could increase the efficiency of the classification. However, depending on their size, filters reveal local or global features. Smaller filters tend to identify local features while larger filters focus on global features. Figure 1-b gives an example in which the length of filters is set to 3. • Stride – Stride refers to the step between two consecutive convolutional operations. Using a small stride corresponds to the generation of an overlap between different filters while a longer stride reduces the output dimension. By default, the stride is set to 1.

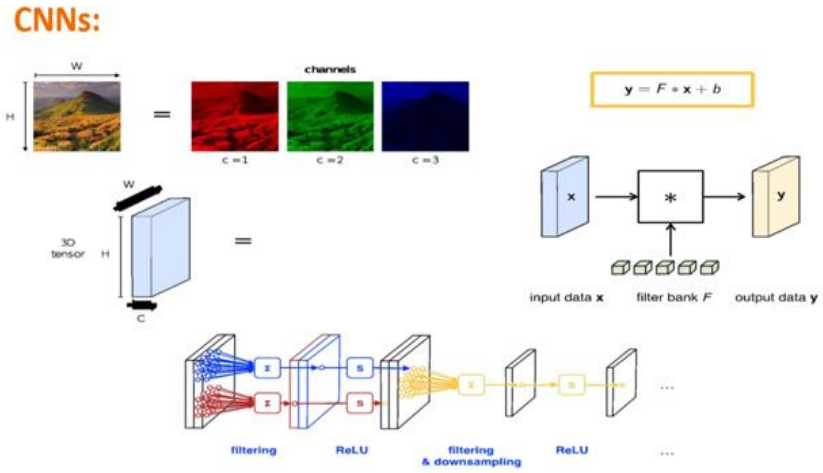


(a) -CNN System Architecture for Side Channel Attacks (Red Area: Part of the Item Selection; Blue Area: Part of the Classification). (b) Operations in Convolutional Blocks (Convolutions, Activation, Middle Pool)

Padding – Let a and b be two vectors, the dimension of the convolution between these two vectors will be dim (a ~ b) = dim (a) −dim (b) stride + 1 [GBC16] where ~ refers to the convolution operation. In some cases, a subsample may be generated. To avoid this phenomenon and to avoid losing information, we can use padding that adds a "border" to our input to ensure the same dimensions are retained after the convolutional operation. By default, two kinds of padding are used: valid padding and same padding. Valid padding means "no-padding" while same padding refers to a zero-padding (the output has the same dimension as the input) [GBC16]. Figure 1-b gives an example in which we select same padding. Indeed,

two 0 values are added at the endpoints of the vector in order to obtain an output vector of dimension 6. After each convolutional operation, an activation function (denoted σ) is applied to identify which features are relevant for the classification. As explained in [KUMH17], the scaled exponential linear unit function (SeLU) is recommended for its self-normalizing properties. The SeLU is defined as follows: selu(x) = λ, x if x > 0, α (exp(x) − 1) if x 6 0. (2) The SeLU pushes neuron activation towards zero mean and unit variance in order to prevent vanishing and exploding gradient problems.

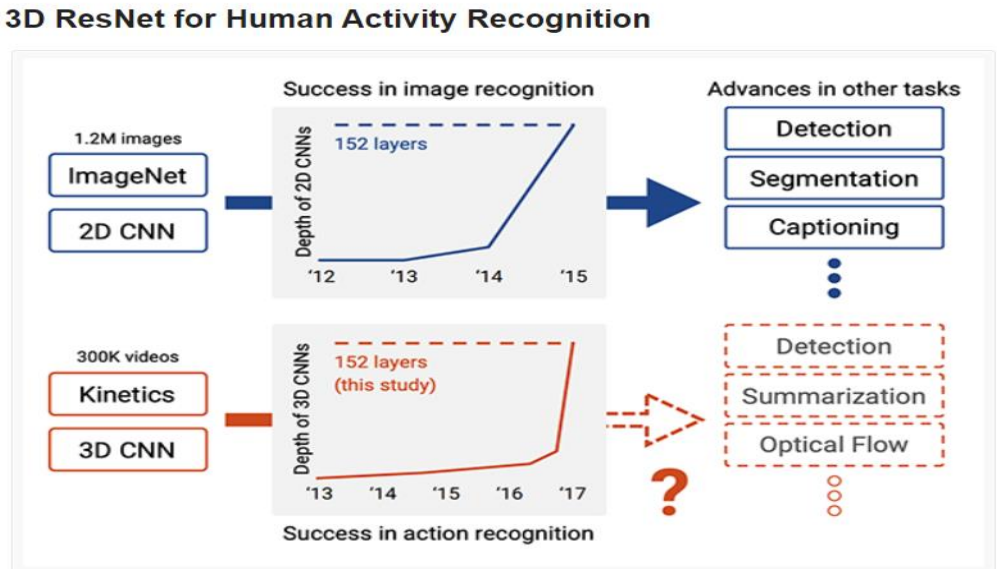
As is usually done in our model, start-up work is used. The inspiration for this assessment was to choose whether the current video dataset has enough data to prepare an important convolutional neural framework (CNN) with temporal and spatial transient three-dimensional (3D) fragments. Starting from very late, the display level of 3D CNN in the field of motion confirmation has basically improved.
In any case, until this point, customary research has just investigated moderately shallow 3D models. We have considered the engineering of different 3D CNNs, going from generally shallow 3D CNNs in current video datasets to exceptionally profound 3D CNNs. In view of the consequences of these analyses, the accompanying ends can be drawn: (I) ResNet-18 preparing will cause critical over fitting of UCF-101, HMDB-51 and dynamic systems, while elements won't. (Ii) The Kinetics dataset has enough information to prepare profound 3D CNNs, and it can prepare up to 152 ResNets layers, which is like 2D ResNets on Image Net. The normal precision of ResNeXt-101 on the active analyser arrives at 78.4%. (Iii) The straightforward 3D design of dynamic pre-preparing beats the complex 2D engineering, while the pre-prepared ResNeXt-101 accomplishes 94.5% and 70.2% on UCF-101 and HMDB-51, separately. The utilization of 2D CNNs prepared on Image Net has gained huge ground in different picture undertakings. We accept that the mix of Deep 3D CNN and Kinetics will follow the effective history of 2D CNN and Image Net and animate the advancement of video PC vision. The codes and pre-prepared models utilized in this examination are freely accessible.



4. Use Deep Learning and OpenCV to Classify Images

In this article, we will create a Python script that can be used to classify input images using the Caffe framework's OpenCV and Google Net (pre-trained on Image Net). Szeged et al. Introduced the Google Net architecture (now called "Inception" after the new

microarchitecture). OpenCV 3.3 also supports other architectures, including Alex Net, ResNet, and Squeeze Net-we will study these architectures in future blog posts for OpenCV to conduct deep learning. At the same time, let us learn how to load the trained Caffe model and use it to classify the image through OpenCV.



We will evaluate a diverse set of nonlinear and ensemble machine learning algorithms in some cases.
Nonlinear Algorithms:
k-Nearest Neighbors,
Classification and Regression Tree,
Support Vector Machine
Naive Bayes
Ensemble Algorithms
Bagged Decision Trees
Random Forest
Extra Trees
Gradient Boosting Machine

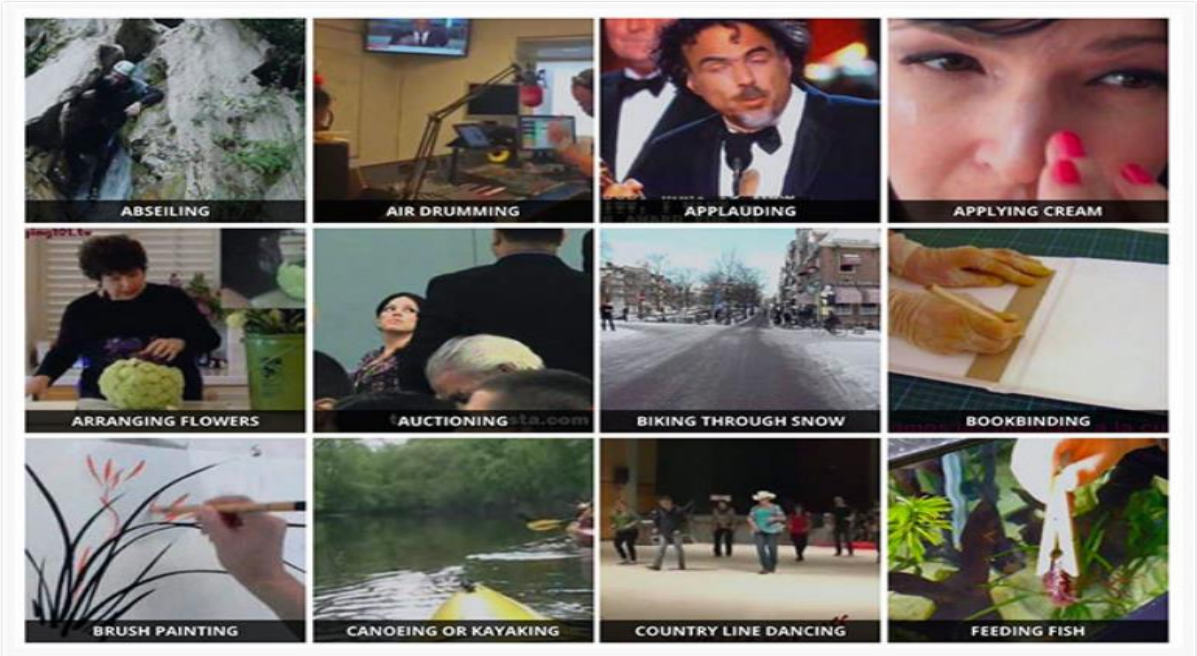
5. Experimental Results

Our human activity recognition model can identify more than 400 activities with an accuracy of 8.4-94.5% (depending on the task).
The sample activities are as follows:
Archery, wrestling, baking cookies, counting money, driving tractors, eating hot dogs, flying kites, tattoos, grooming horses, hugging, skating, juggling fire, kissing, laughing, motorcycles, news broadcasting, opening gifts, playing guitar, playing Tennis, robot dancing, sailing, diving, skiing, beer tasting, beard trimming, using a computer, washing dishes, welding, yoga, etc.
Practical applications of human activity recognition include:
Automatically classify / classify video data sets on disk. Train and monitor new employees to perform tasks correctly (for example, proper steps and procedures should be taken when

ISSN: 2233-7857 IJFGCN

making pizza, including spreading the dough, heating the oven, putting soy sauce, cheese, toppings, etc.).
Verify that food service personnel have washed their hands after going to the bathroom or handling food that may cause cross-contamination (such as chicken and salmonella). Monitor bar / restaurant customers and ensure that their services are not over-served. Usually 2D kernels are used instead of 3D kernels, which enables us to include spatiotemporal components for activity recognition. Then, we will use the OpenCV library and the Python programming language to implement two versions of human activity recognition. Summarize the results of applying human activity recognition to some example videos

Data Set:



The dataset our human activity recognition model was trained on is the Kinetic 400 Dataset/HMDA Data Set.
This dataset consists of: 400 human activity recognition classes, At least 400 video clips per class (downloaded via YouTube),A total of 300,000 videos.and various Human Reorganization Activities are shown in the below fig.

ISSN: 2233-7857 IJFGCN

a

c

b

d

6. Conclusion

We present a Human Activities Recognisation using OpenCV and deep learning techniques for improving classification performance on Human Activities Recognisation and video. We experimentally confirm its effectiveness us in different public datasets. The overall recognition was particularly successful with various datasets and our approach better than previous techniques, in future work to Enhanced Advanced Deep Learning Techniques.

REFERENCES

1 https://docs.opencv.org/2.4/modules/refman.html
2 https://www.pyimagesearch.com/2017/08/21/deep-learning-with-opencv/
3 https://opensourceforu.com/2017/11/a-quick-look-at-image-processing-with-deep-learning/
4 Shokri,R.,andShmatikov,V. 2015. Privacy-preservingdeeplearning. In ACM Conference on Computer and Communications Security (CCS).
5 Tran, L.; Kong, D.; Jin, H.; and Liu, J. 2016. Privacy-cnh: A framework to detect photo privacy with convolution neural network using hierarchical features. In AAAI.
6 https://in.mathworks.com/help/images/deep-learning.html
7 https://www.pyimagesearch.com/start-here/

ISSN: 2233-7857 IJFGCN