Privacy-Preserving Human Activity Recognition and Resolution of the image using Deep Learning Algorithms of Convolutional neural network and Recurrent neural network.

[1]**Mr. Vijaya Kumar Kambala,** Assistant Professor, Dept. of CSE,
Research Scholar at Vellore Institute of Technology -AP, VIT-AP, Amaravathi.
Email: ***kvkumar@pvpsiddhartha.ac.in*****,*****vijaya.18phd7017@vitap.ac.in*.
Prasad V Potluri Siddhartha Institute of Technology (Autonomous), Vijayawada, A.P-520007
[1]**Dr. J. Harikiran,** Sr. Assistant Professor, *School of Computer Science and Engineering*
Email: harikiran.j@vitap.ac.in, Vellore Institute of Technology -AP, VIT-AP, Amaravathi.

Abstract

The images can be captured in different areas with some resolution and can be used and applied for various applications so, these images and video capturing of various activities need to verified and analyzed and can be categorized according to the user outputs. When dividing these activities the classification should be clear and without any conflicts. The images can be identified based on these classifications. We used CNN and RNN models to identify and classify these images of human activities. Generally the problems like over fitting and repetition of the previous data becomes conflict. To avoid these problems we used RNN algorithms to predicts image features based on the previous features, by taking the algorithm attention based models.RNN is model based of neural networks system and control the internally that match the runtime environment that direct cyclic connection to different nodes. We propose a novel base RNN(Recurrent Neural Networks) model based for various classification problems. The CNN will analysis semantic representations from images with specific representations. The RNN part models image relationship can be dependent.

Keywords: *Deep Learning, deep reinforcement learning (DRL), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN).*

I. Introduction

Deep Neural Networks is subset of neural networks are in which many number of net works can be added or taken so that taking positive sample input and can be with different outputs without with no inputs from the previous data links of networks of different layers in the way which is in the forward directions and can evaluated for better training.

By considering the outputs and trained data sets of supervised learning of different datasets taken trained on 'what we want' through back propagation.[1] Like you go to a restaurant and the chef gives you an idea about the ingredients of your meal. The work is the same way as you will have the flavor of those specific ingredients while eating but just after finishing your meal you will forget what you have eaten so like same if the data of the previously taken and trained once or multiple times the outputs can taken as various times it can be analyzed even it is of with same input. The output can be recognized if it trained multiple times[2].

II. Resolution with pixels

Effective super resolution can be set of resolutions and can be taken as minimal and training the data of images and videos from a images with resolution of various image transforms and correct make the images less resolutions for human activities recognition. Such transformations can be included for various point of process, scaling, rotation of the images, and other related images can be transforms and match the different possible cameras of video motion [3]. This resolution can be targets the outputs of the system is not supporting from get high-resolution videos and images can be evaluated and different phase due to privacy protection but has an access to a rich set of high-resolution of images and videos training videos publicly available in social media like. This type of videos can be providing the resolutions of images can be identified[4]. From the better resolution can be provided, this is a various ways of using the effective resolution formulation training, these assumption is that multiple images may contain a comparable of information to a single image. It is called 'inverse' super resolution. The effective resolution taken as the input and the output can be taken from the previous propogation.

III. Learning techniques using CNN

The training the data with the optimal set and the videos can be taken the set with v = (Z)N= -1 based on video data. The v learned from video captured from and predict to get the superior results of various transforms in any case of evaluations and can be taken uniformly and considered the same, and confirm in our experimental results and can be shown clearly in the out images[5]. In this method explains the Markov chain Monte Carlo (MCMC)-based technique to find the optimal set of transforms to provide the ideal activity classification for resolution decision boundaries. This is second method know the minimal of transform filters. The method explains, methodology of directly variable to this methods and classification functions provides us a precision solution for the problem, a fair number of MCMC sampling of repeatedly is needed for a better performance solution.

IV. Recurring neural network with deep

Deep Learning: The dataset trained and used in convolution neural network, the concept can be taken to evaluate the networks of networks in a deep. To build CNN model, gathered all the concepts needed to solve and derive the datasets in image processing. A convolution models explains the mathematical and derivative method of for two functions, and fining the third value which is the result concluded from what we derived from two functions.CNN concepts are specified and applied in several areas in medical and socio economical process management and retrieving the data and analyze and trained the datasets image processing[6] .In the area of image processing; a CNN matrix is used for image manipulation can be applied for different purposes that is blurring, sharpening, or edge detection and scaling. The final image is considered and can be set in matrix form with points from 0 to 255, based on the intensity of color in each pixel.

V. Image data pre-processing for deep learning

Deep learning concepts can be applied for neural networks with a deep hidden layers, training the large amounts data that can be taken from dataset.[3]From the dataset taken and trained for effective in getting better performance from the images and approaching human level accuracy in single predictions image capturing. The approaches of mathematical derivations are proved in years ago. Mainly the approaches have considered to develop the machine learning.1.Availability of large datasets images are there in multiple domains and b)The dataset can be trained according to deep neural network. [4]The commonly the images having parameters with how many pixels with height, width, channels, levels per pixel. Number of pre-determined solutions will be carry out when the Deep Learning techniques can be taken for solving the different layers of network and are uniform with results of image scaling and with mean standard deviation of input data taken from different networks, normalizing image inputs, and dimensionality reduction and resolution can be reduced[7].

VI. Evolutional and experimental setup

Cameras are becoming more and more ubiquitous. Millions of surveillance cameras record people's daily behaviour in different public areas and the persons who is using different handy video digital cameras designed to have life logging to get huge collections of oneself images and videos.[8] The robots in social gathering are prepared with several cameras for to do some one to one commutation. At the same time, such an abundance of cameras also causes a big social problem: protecting privacy from unwanted videos. We want the camera system to identify important things and help a person in everyday human activities of these images and videos,want this to not violate the privacy of other users. The approach tends to two conflicting goals. More precisely, this can be (1) prevent the use of a camera equipment to obtain clear visual data can contain personal data (for example, eyes or shape the head), preferably at the physical level. At the same time, To improve the equipment an can be record as much relevant data can be shown from its video, finally it can be understands the various things and current events for monitoring, logging for intellectual things[8].

Previously, studies have been conducted to meet these social requirements. Templeman(2014) studied and recognized from images and videos taken by easy carry cameras, finding places the privacy was needed to be safe(e.g., toilets). This will allow the device to turn off automatically in sensitive places. It can also be argued that restricting the device to only processing / transmitting information about functions (CNN algorithms) alternately the visual data will make it protect human activity privacy. However, recent studies of “visualization” of functions have shown that it is correct to have possible to rectify a better visual data from these objects. In addition, all of the methods described above are based on software processing of high-resolution source videos and the inability to capture these original videos with cyber-attacks. A more fundamental solution to building a confidentiality visual camera is to use anonymous videos. One shows of anonymous videos. Instead of receiving high-resolution videos and trial access to them, this direction involves only receiving anonymous videos. The idea is that if we can develop reliable approaches to computer vision that use only such anonymous videos, we can recognize them while maintaining confidentiality. Such a concept may even allow cameras that can intelligently choose their resolution; he will use high-resolution cameras only when necessary (for example, in emergency situations), which is determined based on extreme analysis of low-resolution video[9].

Attempts have previously been made within the framework of such a paradigm**.** This traditional technique is use to small size comparing with the original training images to match the final resolution, making the training videos can be shown look like the tested one.

VII. 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 convolution 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 convolution network can be characterized by the following formula: $s \circ [\lambda]\ n1 \circ [\delta \circ [\sigma \circ \gamma]\ n2]\ n3$.Convolutional layer performs a series of convolution operations on its inputs to facilitate pattern recognition (see Figure 1-b). During forward propagation, each input is convoluted with a filter (or kernel)[10]. 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.Padding – Let a and b be two vectors, the dimension of the convolution between these two vectors will be $\dim (a \sim b) = \dim(a) - \dim(b)\ \text{stride} + 1$ [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[11]. This activation function is used by default in our architectures. In this paper, it is determine whether present image and video datasets are having enough data for training the images using deep convolutional neural networks (CNNs) with spatio-temporal three proportions of kernels. The conventional neural network problems have been trained to explored relatively requiring 3D architectures. By determining the various architectures of CNNs from similar requiring analyzing net works of ones on present image video datasets. The out puts of those problems taken, the results will be finalized. (i)Training resulted are moderate for CNN-22, CNN-1-31, present net work. (ii) The trained data taken from the data set for deep machine learning[12].

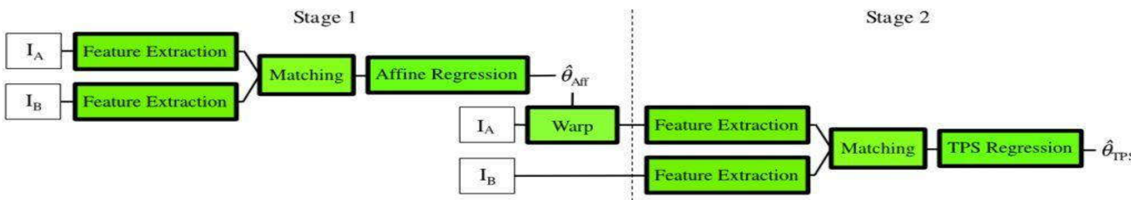

Fig 1: Convolutional neural network architecture.

VIII. Image classification using CNN and deep learning

The image classifying is done to separate according to the levels this can be achieved based on the algorithms we take. The CNN (convolutional neural network) used for classifying the images. In figure 2 shown the pooling numbered 3 and convolutional 1,2,3 shown with images with resolution 51×68×3,based on the image the different pooling taken and controlled and viewed for final output and using deep learning the pooling frames are set in order to classify using different network layers of CNN and DNN algorithms[14].

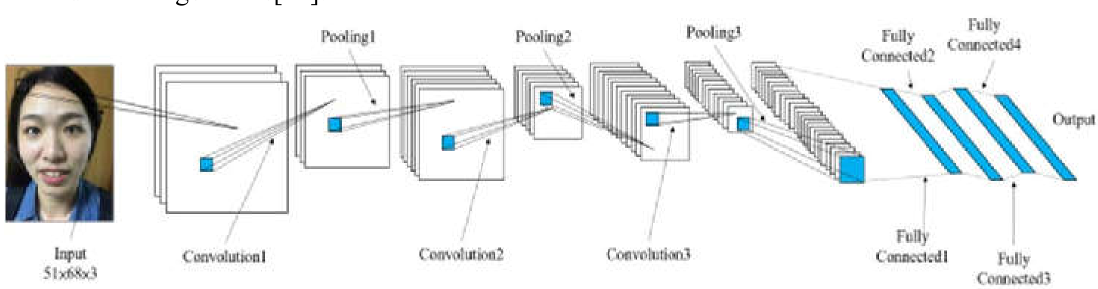

Fig 2: The image of human recognized and classified using deep neural network.

Table 1: Showing the diverse set of classification values in k entries.

Input	Layer	Mapping	Resolution	System size	Strides
1	Convolution	1	28×28	2×2	1
2	Pooling	5	32×32	5×5	0
3	Learning	8	10×10	1×1	1
4	Convolution	10	12×12	2×2	1
5	Recurrent learning	12	6×6	-	0
6	Deep learning	14	8×8	-	1
7	Convolution	16	10×10	2×2	-

IX. Experimental results

The human activity actions can be identified totally 400 things with 78.4-94.5% accuracy .Sneezing, explaining, plucking flower, walking these activities are observed in the previous experimental results. Taking the CNN (convolution neural network) and RNN (Recurrent neural network) to take sequential inputs and without any sized mention and classified using the different pooling frames shown in figure 2 and controlled final output taken. The recurrent network works with past data and considered the present samples of input influenced on the present inputs. And these RNN work on the feed forward networks and that is shown in the above table. The convolution layers specifying with different inputs 1 mapping to resolution of 28×28 with size 22 and strides of 1.Same pooling mapped with 5 with resolution of 3232 in 55 of system size. In deep learning the image is classified and mapped to 14 with resolution 1010 of 22.In RNN also the mapping of 12 and resolution of 66 with no system size. According to to these values the image classifications done and final out puts can be taken in order to preserving the various human activations. Taking the output from the experimental of average correct with specifications rate for various camera configurations using a different levels of leave person different evaluations[15].This specified calculation used for a running classification rate across different images same iterations. In each level resolutions subjects is sequentially selected to be the one image per subject. By averaging the classification rate score across all 14 test subjects we get an average classification rate for one or more iteration.

X. Changes of image resolution

The impact of spatial resolution change was evaluated for the resolutions ranging from 20×20 down to 1×1 across the cameras. The results of this method and actions on this image taken accurate of classification. Table 1.The intensity signals are same and actions will be shown non-negligible variations thus further confirming the fulfillment of the dataset.

Table 2: Shows the different changes of image resolution.

Image Quality	**classification**	**Covering the cloth of head.**	**Checking Mobile**	**Plucking flower**	**Cutting fruit**	**Group Dance**	**Studying**	**Yoga**	**Staring**
10×10	**89.40%**	66.92%	80.19%	97.56%	95.72%	85.25%	83.08%	99.83%	98.28%
6×6	**86.20%**	68.11%	74.69%	95.31%	90.17%	79.25%	65.39%	99.72%	98.58%
5×3	**54.52%**	68.42%	62.64%	88.03%	91.44%	78.89%	74.03%	99.64%	97.56%
4×4	**86.71%**	77.22%	69.64%	91.11%	89.33%	80.00%	71..56%	99.86%	97.28%
3×3	**82.71%**	70.58%	45.64%	87.94%	88.06%	76.56%	82.11%	99.64%	94.08%
1×1	**75.70%**	50.92%	35.42%	85.53%	85.69%	58.56%	79.81%	94.08%	92.94%

XI. Summary of outputs

The output values is given in table total data is provided with comparing the previous data. we given the various output values.

Table 3 : Observations of resolution standards.

.Observations	Camera values	Resolution	standards
Above average	20×20, 40 Hz, 4 cams	99.85%	2.74%
Frame-value	10×10, 2 Hz, 5 cams	96.89%	2.69%
Camera-1	20×20, 40 Hz, 1 cam	89.69%	2.33%
less Spatial values	2×2, 40 Hz, 5 cams	88.85%	2.53%
Below average	1.5×1.5, 3 Hz, 1 cam	55.55%	1.25%

XII. Conclusion

We present Human activities reorganization and deep learning techniques for improving classification performance on Human Activities reorganization and video. Based on practical outputs can be finalized the good percentage of different image datasets. Final outputs can be recognized for a quality performance with various datasets and our approach better than previous techniques, in future work to Enhanced Advanced Deep Learning RNN techniques.

XIII. Acknowledgement

Vijaya Kumar Kambala and Dr. J.Harikiran 's done the research and experimentally proved in finding the Privacy-Preserving Human Activity Recognition and Resolution of the image using Deep Learning Algorithms of Convolutional neural network and Recurrent neural network.

References

1. Boston University: Privacy-Preserving Smart-Room Analyt- ics. vip.bu.edu/projects/vsns/privacy-smartroom/, 2015.

2. https://opensourceforu.com/2017/11/a-quick-look-at-image-processing-with-deep-learning/

3. KinectSDK. www.microsoft.com/en-us/kinectforwindows/, 2015.

4. Unity - Game Engine. unity3d.com, 2015.

5. J. Aggarwal and M. Ryoo. Human activity analysis: A re- view. *ACM Comput. Surv.*, 43(3):16:1–16:43, 2011.

6. Shokri,R.,and Shmatikov,V. 2015. Privacy-preservingdeeplearning. In ACM Conference on Computer and Communications Security (CCS).

7. J. W. Davis and A. Tyagi. Minimal-latency human action recognition using reliable-inference. Image and Vision Com- puting, 24(5):455–472, 2006.

8. K. Guo, P. Ishwar, and J. Konrad. Action recognition from video using feature covariance matrices. IEEE Transactions on Image Processing, 22(6):2479–2494, 2013.

9. L. Jia and R. J. Radke. Using time-of-flight measurements for privacy-preserving tracking in a smart room. Industrial Informatics, IEEE Transactions on, 10(1):689–696, 2014.

10. B. Saghafi and D. Rajan. Human action recognition using pose-based discriminant embedding. Signal Processing: Im- age Communication, 27(1):96–111, 2012.

11. 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.

12. https://in.mathworks.com/help/images/deep-learning.html

13. https://www.pyimagesearch.com/start-here/

14. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17) Michael S. Ryoo,1,2 Brandon Rothrock,3 Charles Fleming,4 Hyun Jong Yang2,5 1Indiana University, Bloomington, IN 2EgoVid Inc.,

15. K. Vijaya kumar,Dr.J.Harikiran,.M. A. Rama Prasad,Uddagiri Sirisha, "Privacy-Preserving Human Activity Recognition and Resolution Image using Deep Learning Algorithms Spatial relationship and increasing the attribute value in OpenCV", IJAST,Vol. 29, No. 7, (2020), pp. 514-523.