International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT2062134

Video Classification Using Deep Learning

Dr. Sheshang Degadwala[1], Harsh Parekh[2], Nirav Ghodadra[2], Harsh Chauhan[2], Mashkoor Hussaini[2]
[1]Associate Professor, Computer Department, Sigma Institute of Engineering, Vadodara, Gujarat, India
[2]U.G. Scholar, Sigma Institute of Engineering, Sigma Institute of Engineering, Vadodara, Gujarat, India

ABSTRACT

Video order has been comprehensively investigated in PC vision in view of its wide spread applications. In any case, it remains a surprising task by virtue of the mind boggling troubles in fruitful segment extraction and successful arrangement with high-dimensional video depictions. Video groupings present uncommon irregularity as a result of monster scope changes, point of view assortment, and camera development, which pose fantastic challenges for both video depictions and characterization. With the phenomenal accomplishment of significant learning, convolutional neural frameworks (CNNs) and their 3-D varieties have been considered in the video territory for an immense grouping of order assignments. Video depictions have accepted a basically noteworthy activity in video examination, which authentically impact a complete execution of video characterization. Both the spatial and brief information should be gotten and encoded for extensive and educational depiction of video progressions. Significant Learning feature level blend plans have exceptional capacity of video depictions for improving the introduction of video grouping. We have driven wide assessment on four going after for video arrangement including human movement affirmation and dynamic scene grouping.

Keywords : Video Event, Classification, Median channel, Histogram Equalization, CNN and RCNN.

I. INTRODUCTION

Video classification has been broadly investigated in PC vision because of its spread widely which is use in numerous significant applications, for example, human acknowledgement activity and dynamic scene arrangement, etc. Be that as it may, video arrangements present exceptional vacillation because of enormous scale changes, perspective variety, and camera movement, which posture incredible difficulties for both video portrayals and classification. With the exceptional accomplishment of profound learning, Regional convolutional neural systems (RCNNs) and their 3-D variations have been contemplated in the video area for an enormous assortment of classification errands. It is exceptionally wanted to have a start to finish learning methods that can build up successful video portrayals while all the while leading productive video classification [3].

Deep Feature level combination plans have extraordinary ability of video portrayals for improving the exhibition of video classification. Troupe learning has indicated extraordinary ability of taking care of high-dimensional highlights to direct exact and effective classification assignments. By working with profound learning models, gathering learning can to a great degree improve the presentation [2].

Figure 1: Video Classes

II. RELATED WORK

In [1] Eng-Jon Ong, Sameed Husain, Mikel Bober-Irizar, Miroslaw Bober depicts Deep Architectures and Ensembles for Semantic Video Classification. In that they utilized PC vision, Artificial Neural Networks, Machine Learning Algorithms. They proposed a leftover design based DNN for video classification, with cutting edge classification execution at altogether decreased multifaceted nature. They additionally proposes four new ways to deal with fair variety driven multi-net troupe. However, they found that on a people level, each DNN accomplishes generally a similar scope of GAP20 exactness from 82% to 83%.

In [2] S. Jothi Shri, S. Jothilakshmi portrays Crowd Video Event Classification Using Convolution Neural Network. In that they utilized Deep Learning, CNN, SVM and Deep Neural Network. They proposed swarm occasion classification in video is a significant and testing task in PC vision based frameworks. It perceives an enormous number of video occasion. The exemplification of profound learning in a video occasion classification determines amazing and recognizes highlights depictions. Lamentably, the VGG16 model gives the 82% outcome and it demonstrates conflicting precision in consistent methodologies.

In [4] Hongji Seong, Junhyuk Hyun, Suhyeon Lee, Suhan Woo, Hyunbae Chang and Euntai Kim portrays New component level Video Classification through Temporal Attention Model and utilized untrimmed video classification, Convolution Neural Network, transient consideration, fleeting cushioning, information growth. They proposed coView 2018 is another test which goes for concurrent scene and activity acknowledgment for untrimmed video. They centers around the examination motel transient area and the worldly consideration model. They recorded 95.53% exactness in the scene and 87.17%, all things considered, using transitory thought model, nonzero padding and data development. The essential obstacle is that, coView 2018 dataset doesn't give enough data to use a capricious designing.

In [5] Haiman Tian, yudongTao, Samira pouyanfar, Shu ching chen, Mei-ling Shyu portrays Multi-model profound portrayal learning for video classification and utilized multi-model profound learning, move learning, multi organize combination, calamity the board framework. They proposed new multi-model profound taking in structure for occasion location from videos for utilizing late advances in profound neural system. In any case, this mixes doesn't for the most part improve the general precision and mAP execution which show why an unrivaled blend procedure is required for such complex multi-model datasets.

In [6] Aaron Chadha, Alhabib Abbas and Yiannis Andreopoulos depicts Video Classification with CNNs: Using the codec as a spatio-worldly movement sensor and utilized Video coding, classification, profound learning. They research Video Classification by means of a two stream Convolution Neural Network (CNN) structure that straightforwardly ingests data removed from compacted video bit-stream. Anyway they saw that the worldly and spatial streams shows unmistakable predispositions as far as class, contingent upon the idea of movement.

In [9] zhenqi xu, Jiani Hu, Weihong Deng portrays Recurrent Convolution Neural Network For Video Classification utilized Video Classification, Deep Learning, Action Recognition and Recurrent Convolution Neural Network. They introduced another profound learning design called Recurrent Convolution Neural Network which joins convolution and intermittent connections for video classification errands. One downside of their framework is that it can't demonstrate long time relations from videos.

In [12] Liuwang Wang, Huaping Liu, Fuchun Sun portrays Dynamic Texture Video Classification Using

Extreme Machine Learning utilized Extreme Machine Learning, partiality proliferation, dynamic surface. They proposed another way to deal with handle the dynamic surface acknowledgment issue. What's more, they moreover moved closer on the DynTex dataset, and besides exhibited the practicality of the proposed approach. However, their objectives is quiet low similarly as there is simply single occasion per class and deficient classes are available for practical grouping reason.

In [13] Mohammad Tavakolian and Abdenour Hadid depicts Deep Discriminative Model For Video Classification utilized Deep learning, Video based classification, Heterogeneous Deep Discriminative Model (HDDM). They displayed another profound learning approach for video based scene classification. The exhibition of proposed framework is broadly assessed on two activity acknowledgment datasets and three unique surface and dynamic scene datasets. Tragically, In request to dispose of the repetitive data in video arrangements and stay away from the solid connections between adjoining outlines and spoke to them scantily utilizing meager cubic even example.

III. METHODOLOGY

A. Data Analyze:

Video: Dataset of different games.
Real Time: Camera Capture

B. Pre-Process

[1] Histogram Equalized:

Histogram Equalization is a PC picture planning strategy used to improve separate in pictures. It accomplishes this by sufficiently spreading out the most normal power regards, for instance slackening up the force extent of the image. This strategy generally extends the overall separation of pictures when its usable data is addressed by close unpredictability regards. This considers districts of lower close by contrast to build a higher multifaceted nature.

Figure 2: Histogram Example [2]

A concealing histogram of an image addresses the amount of pixels in each kind of concealing part. Histogram evening out can't be applied freely to the Red, Green and Blue portions of the image as it prompts electrifying changes in the image's concealing parity. In any case, if the image is first changed over to another concealing space, as HSL/HSV concealing space, by then the figuring can be applied to the luminance or worth direct without realizing changes to the tint and inundation of the image.

[2] Median Filtering:
The middle channel is one kind of nonlinear channel. It is exceptionally viable at evacuating drive clamor, the "pepper and salt" commotion, in a picture. The standard of the center channel is to override the dull level of each pixel by the center of the diminish levels in a region of the pixels, instead of using the ordinary movement.For middle separating, we indicate the portion size, list the pixel esteems secured by the bit, and decide the middle level.

Figure 3: Median Filtering Example [3]

On the off chance that the part covers a considerably number of pixels, the normal of two middle qualities is utilized. Before starting middle separating, zeroes must be cushioned around the line edge and the section edge. Henceforth, edge mutilation is presented at picture limit.

C. Object Detection

[1] Background subtraction:

A foundation deduction estimation is first applied to each video packaging to find the territories of interest (ROIs). A CNN request is then finished to orchestrate the gained ROIs into one of the predefined classes. Our system much lessens the figuring multifaceted nature interestingly with other article area computations. For the examinations, new datasets are created by recording back doors and play zones, places where infringement are most likely going to occur. Differing picture sizes and exploratory settings are attempted to build up the best classifier for perceiving people.

Figure 4: Background Subtraction [3]

[2] Temporal Difference:

Passing differentiation (TD) learning insinuates a class of sans model help learning methods which take in by bootstrapping from the present check of the value limit. These systems test from nature, like Monte Carlo methodologies, and perform refreshes reliant on current assessments, like powerful programming techniques.

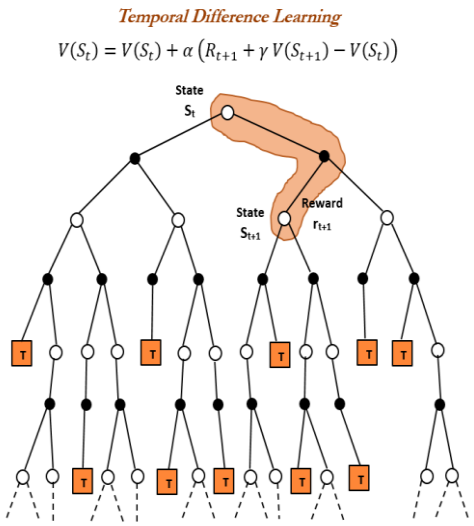

Figure 5: Temporal Difference [5]

D. Deep Learning

[1] CNN:

In significant getting data of, a convolutional neural system (CNN, or ConvNet) is a class of significant neural frameworks, most outrageous usually realized to researching visual imagery. CNNs are regularized versions of multilayer perceptron's. Multilayer perceptron's as a rule recommend totally related frameworks, that is, every neuron in one layer is trapped to all neurons inside the accompanying layer. The "totally connectedness" of those frameworks makes them in danger to over fitting assurances. Typical strategies for regularization consolidate including two or three kind of significance estimation of burdens to the setback work. Regardless, CNNs take an uncommon philosophy towards regularization: they take expansion of the different leveled plan in bits of knowledge and gather progressively obfuscated styles the usage of humbler and simpler styles. Thusly, on the size of connectedness and multifaceted nature, CNNs are on the decrease outrageous.

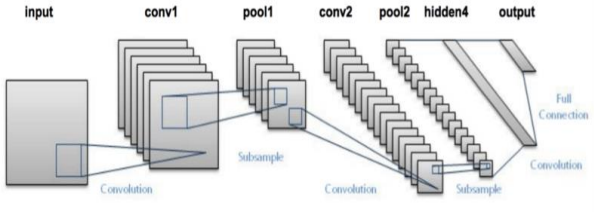

Figure 6: CNN [7]

[2] RCNN:

Computer vision is an interdisciplinary field that has been increasing enormous measures of footing in the ongoing years(since CNN) and self-driving vehicles have become the dominant focal point. Another vital piece of PC vision is object discovery. Item identification helps in the present estimation, vehicle recognition, observation, and so forth. The distinction between object recognition calculations and characterization calculations is that in discovery calculations, we attempt to draw a jumping box around the object important to find it inside the picture. Additionally, you may not really draw only one bounding box in an article recognition case, there could be many jumping boxes speaking to various objects of enthusiasm inside the picture and you would not know what number of previously.

Faster R-CNN has two systems: area proposition organize (RPN) for creating district recommendations and a system using these proposition to recognize objects. The essential various here with Fast R-CNN is that the later uses explicit chase to deliver district proposition. The time cost of making area proposition is much humbler in RPN than explicit request, when RPN shares the most figuring with the article acknowledgment arrange. Rapidly, RPN positions area boxes (called hooks) and proposes the ones without a doubt containing things.

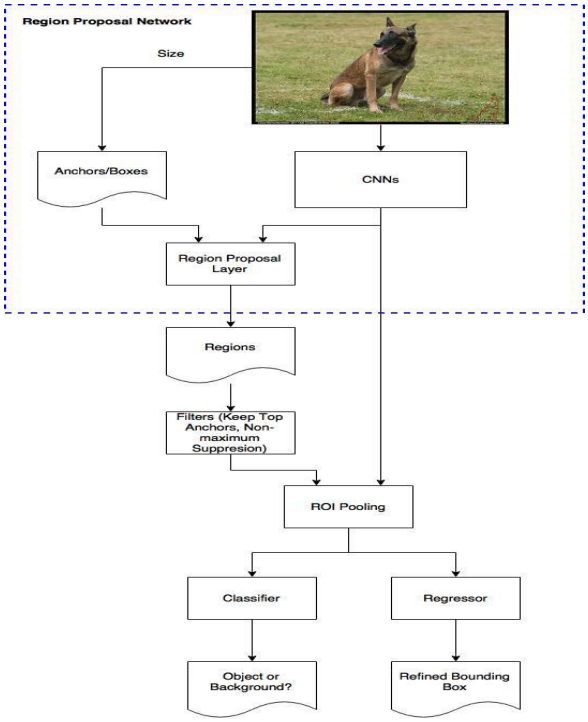

Figure 7: RCNN [5]

IV. Comparative Study

Table 1: Pre-Processing

Pre-processing	Advantages	Disadvantages
Median Filtering	It can preserve sharp features. More strong than mean separating. Single very unrepresentative pixel in neighbour does not affect mean.	Hard to treat systematically the impact of a middle channel. There is no error propagation.
Histogram Equalization	It helps to adjust image intensity to enhance image contrast. It helps to increase global contrast.	It is indiscriminate. It may increase the contrast of background noise.

Table 2: Object Detection

Object Detection	Advantages	Disadvantages
Background Subtraction	It works very accurately. Once in a while camera may move.	It is having a restricted applicability. Foreground objects may not contain the background color.
Temporal Feature	It learns to enhance each feature via temporal relationships among objects in a frame.	Sometimes not give accurate result as it is not robust.

Table 3: Deep Learning

Deep Learning	Advantages	Disadvantages
Convolutional Neural Network(CNN)	Accurate in image recognition problems. Helps to contribute to improve the efficiency of deep neural networks	High computational cost. They use to need a lot training data
Region Convolutional Neural Network(R-CNN)	Helps to bypass the problem of selecting huge number of regions. Interference is minimized.	It is difficult to detect small objects. It only predicts a label, not a segmentation box.

V. Proposed System

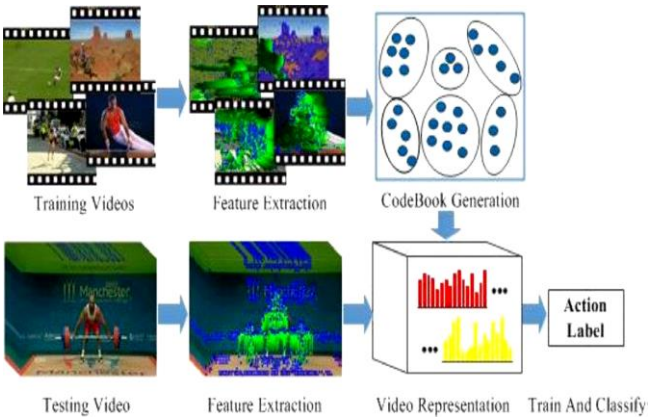

Figure 8: Proposed System

Here, as shown in the figure initially it will take whole video as an input. Then after, it will convert the video into the frames and it will be further extract the features of that given frames. Moreover, it will generate codebook on the basis of the features. After that this will create the graphical representation from the past results and then it will compare these two graphs. At last, after classification from the graph representation it will generate the label of that activity.

VI. Results and Analysis

Figure 9: Result of activity football

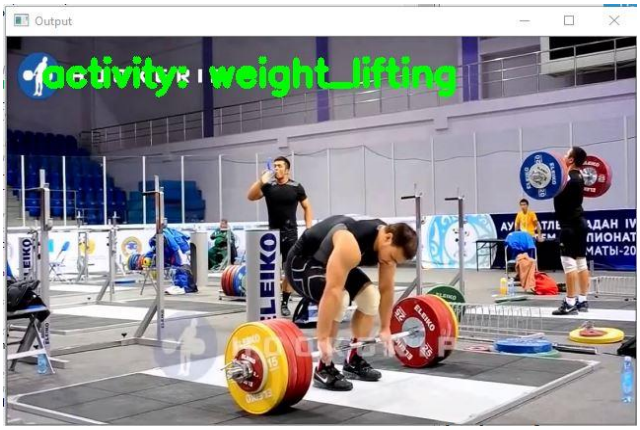

Figure 10: Result of activity weight-lifting

Figure 11: Result of activity tennis

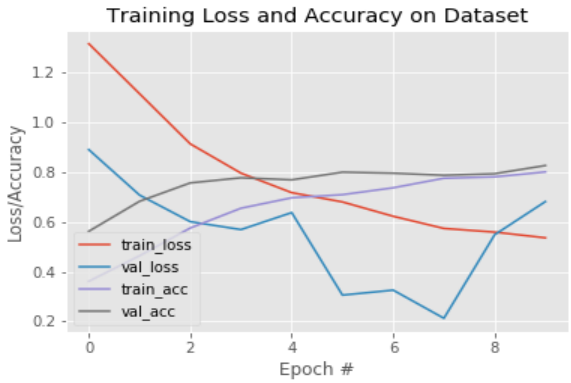

Figure 12: Result of 10 Epoch

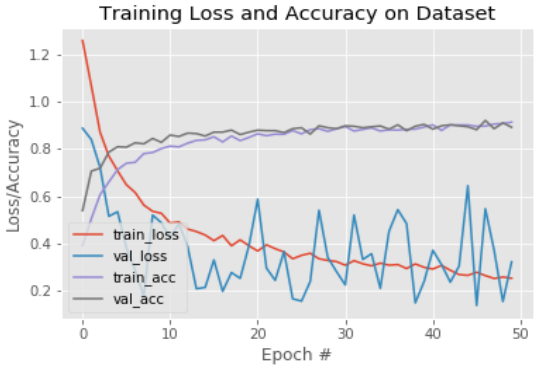

Figure 13: Result of 50 Epoch

VII. CONCLUSION

In this project, we have introduced a start to finish profound learning system, the RCNN, for video arrangement. In light of the spatial-transient highlights extricated by heterogeneous profound CNNs, we propose a group learning module dependent on irregular projections to take a shot at the highest point of CNNs. Irregular projections diminish high-dimensional video portrayals into a lot of low-dimensional subspaces dependent on which an outfit of classifier is found out for expectation. The underlying outcomes from the classifiers are additionally encoded by a recently proposed layer, which is trailed by a completely associated layer to consolidate all the classifiers to create the last grouping outcomes examination. Our methodology use the qualities of profound CNNs for include extraction and learning for effective grouping.

VIII. REFERENCES

[1]. J. Zheng, X. Cao, S. Member, B. Zhang, X. Zhen, and X. Su, "Deep Ensemble Machine for Video Classification," IEEE Trans. Neural Networks Learn. Syst., vol. PP, pp. 1–13, 2018.

[2]. E. Ong, S. Husain, M. Bober-irizar, M. Bober, and C. V Oct, "Deep Architectures and Ensembles for Semantic Video Classification," vol. 14, no. 8, pp. 1–15, 2018.

[3]. S. J. Shri and S. Jothilakshmi, "Jo urn," Comput. Commun., 2019.

[4]. K. Zhao et al., "Research on video classification method of key pollution sources based on deep learning," J. Vis. Commun. Image Represent., 2019.

[5]. S. Lee, "New Feature-level Video Classification via Temporal Attention Model," pp. 31–34, 2018.

[6]. H. Tian, Y. Tao, S. Pouyanfar, and S. C. M. Shyu, "Multimodal deep representation learning for video classification," 2018.

[7]. A. Chadha, A. Abbas, Y. Andreopoulos, and S. Member, "Video Classification With CNNs: Using The Codec As A Spatio-Temporal Activity Sensor," vol. 8215, no. c, pp. 1–11, 2017.

[8]. E. Ergün, F. Gürkan, O. Kaplan, and B. Günsel, "Derinlikli Öğrenme ile Video Aktivite Sınıflandırma Video Action Classification by Deep Learning," pp. 0–3, 2017.

[9]. A. Burney, "Crowd Video Classification using Convolutional Neural Networks," 2016.

[10]. X. C. Road and H. District, "RECURRENT CONVOLUTIONAL NEURAL NETWORK FOR VIDEO CLASSIFICATION Zhenqi Xu , Jiani Hu , Weihong Deng Beijing University of Posts and Telecommunications," no. 10.

[11]. N. Najva and E. B. K, "SIFT and Tensor Based Object Detection and Classification in Videos Using Deep Neural Networks," Procedia - Procedia Comput. Sci., vol. 93, no. September, pp. 351–358, 2016.

[12]. J. Liu, C. Chen, Y. Zhu, W. Liu, and D. N. Metaxas, "Video Classification via Weakly Supervised Sequence Modeling," Comput. Vis. Image Underst., vol. 000, pp. 1–9, 2015.

[13]. L. Wang, H. Liu, and F. Sun, "Using Extreme Learning Machine," vol. 2, pp. 41–50, 2015.

[14]. M. T. B and A. Hadid, Deep Discriminative Model. Springer International Publishing, 2018.

[15]. L. Wang, H. Liu, and F. Sun, "Author's Accepted Manuscript Dynamic Texture Video Classification Using Extreme Learning Machine," Neurocomputing, 2015.

[16]. V. Suresh, C. K. Mohan, R. K. Swamy, and B. Yegnanarayana, "Content-Based Video Classification Using Support Vector Machines," pp. 726–731, 2004.

Cite this article as :

Dr. Sheshang Degadwala, Harsh Parekh, Nirav Ghodadra, Harsh Chauhan, Mashkoor Hussaini, "Video Classification Using Deep Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 406-413, March-April 2020. Available at doi : https://doi.org/10.32628/CSEIT2062134
Journal URL : http://ijsrcseit.com/CSEIT2062134