INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 6, Issue 1)
Available online at: www.ijariit.com

Video classification detection technique for spatial and temporal motion identification using pathway system

Swati Sorte
swati.sorte24@gmail.com
G. H. Raisoni College of Engineering, Nagpur, Maharashtra, India

Dr. Prashant Sharma
pssharma2873@gmail.com
Government Polytechnic, Awasari (Kh), Taluka Ambegaon, Pune, Maharashtra, India

ABSTRACT

In Video classification, the Video scene or the tracking [1] technology, which divides the Video into semantic [4, 5, 6] sections, is an important element for adding insight and searching for metadata for Video. Data augmentation is one of the main methods of addressing the problem of learning to take few shots, but current syntheses are only tackling the detection of crime scene per image when in reality images can contain several movements with respect to crime.

Keywords— *Level features, Detector performance, Multimedia applications, Video retrieval, Anomaly dataset*

1. INTRODUCTION

Video concept detection framework requires a leveraging class of art method to enhance its beauty and efficiency in terms of detection of innovative frame extraction techniques. The videos are converted into shots and then the frames in terms of actual data in the forms of concepts to enhance the research areas to explore on high-level concepts and video extraction should include semantic differences between user views and requirements with video features at the lowest level. Thus, new semantic methods have been developed in order to reconcile the needs of users with the characteristics of the lower level. [2]Different experimental findings show predictive accuracy of more than 75% can be achieved by transferring classifiers of a concept derived from slow-fast pathways. In addition, the approach also adapts to changing the field of test data by expanding the semantic diffusion domain and calling it adaptive superimposed. Anomaly dataset experiments demonstrate that the proposed framework is both effective and efficient in improving the accuracy of semantic crime detection type of concept detection in video.

2. METHODS AND ALGORITHMS USED

In object detection frameworks, many areas can be used to extract the information. One of the recent areas where nowadays technology is making more advancements is [3] Machine Learning i.e. Deep learning. In our approach, we can use pre-trained image classification models i.e. traditional models to retrieve more similar visual features, as they are built for (like for COCO dataset models are designed and hence their performance is fairly good than others). Flow presents an object detection model wherein it uses use a single deep neural network to combine [1] spatial and motion temporal and feature extraction. The features decided are processed in the form of images and are given to the classification architecture, which results in the bounding boxes.

Applications range from tasks such as industrial image processing systems that inspect, for example, passing bottles on a production line, to researching artificial intelligence and computers or robots that can capture the world around them. [2] Computer Vision includes the core technology of automated image analysis, which is used in many areas. Machine Learning Vision usually refers to a process that combines automated image analysis architectures with other grooming methods in this field of technologies to provide automated classification in different field of industrial application.

3. LIBRARY FILES USED IN DEEP LEARNING

Video classification in deep learning is done with the help of tensor flow and Keras. A Slow fast network algorithm used for video classification is the technique wherein each individual frame of a video is independent of the other.

The assumption of various parameters to retrieve from the videos here is that subsequent frames in a video shot will have a similar semantic type of contents. The field of computer vision in machine learning aims to extract semantic knowledge for video analysis in every matter of inspection the concepts from digitized images by tackling challenges such as image classification, object detection, image segmentation, depth estimation, pose estimation, and many more. While feeding any input image, this Slow Fast model is providing to accomplish three things: object detection, object classification, and segmentation. In object concept detection, feature maps are extracted from multiple levels of the convolutional networks where different stages are pooled to a fixed representation through a mechanism.

4. ARCHITECTURE USED FOR VIDEO CLASSIFICATION

Slow Fast networks [11] is a single stream architecture that works with respect to two different frame rates. In this system, the concept of pathways is used to reflect analogy with the

biological Retinal cells i.e. Parvocellular and Magnocellular strategy. This generic architecture has a Slow and Fast pathway shown in fig 1, which are then given to the Slow Fast network, Figure 1 illustrates the concept.

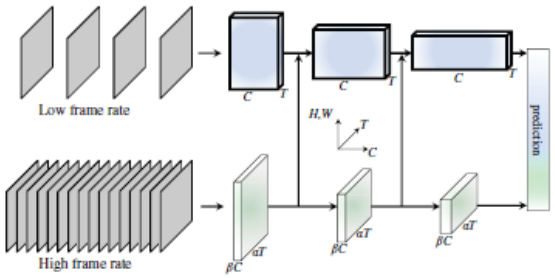

Fig. 1: Slow Fast Pathway Architecture for spatial and motion temporal

The slow pathway is a convolutional model that works on a video subdivided into shots in the form of spatiotemporal volume. The key concept in this Slow pathway has a large temporal stride r on input frames, i.e., it produces only one out of many related frames. A typical value of frame here which we have used is 16—this outstanding speed is roughly 2 frames sampled per second for 30-fps videos. The frames sampled by the Slow network architecture is r, the raw video shot length is T frames, sampled by the Slow network architecture as T, therefore the raw video shot length is T*r frames.

4.1 FAST PATHWAY

Slow pathway and Fast pathway are two convolutional models with the different properties of a high frame rate. This aims to extract a fine representation along with the temporal motion dimension. The [11] Fast pathway carries out processing with respect to the small temporal stride of T/α where $\alpha > 1$ and is the frame rate ratio between the Fast and Slow pathways. This two networks simultaneously work on the same main raw video shot data, so the Fast pathway samples αT frames, which is α times denser and carry information more regarding temporal motions than the slow pathway network. A typical value of α= 8 in this process. The α is the key ingredient of the Slow Fast network. It explicitly indicates that the two pathways work on two different temporal speeds, and thus drives the expertise of the two subnets instantiating the two pathways.

- **Fast Network method temporal motion resolution features extraction**: Fast network not only provides high resolution to the input images but also executes super higher resolution motion features throughout the network strand. In this instantiations, we have not used temporal down-sampled layers throughout the Fast network pathway, till all global satisfied layers received before video classification achieved. This all is done to maintain the fidelity criteria of the feature tensors blocks which have αT frames

- **Low channel capacity of the fast pathway network**: The Fast pathway network also distinguishes himself with other existing traditional methods wherein, it helps to reduce the channel capacity to a lower value. In a nutshell, it can be said that [6] Fast network pathway is a convolutional network and is very much analogous to the Slow pathway, but has a ratio of β (β < 1) channels as compared to the Slow pathway network. The actual value of β chosen to be 1is chosen according to the frame. This is the reason why always fast pathway network is very much computationally effective and efficient than slow pathway network. In all instantiations, the Fast pathway network takes almost ~20% of the total computation execution process. This computation as received as per the mentioned Ganglion Cells hierarchy which suggests that ~15-20% of the retinal cells in the primate visual system are M-cells which are very sensitive to fast motion but not color or spatial details present in the images.

  Also, the low channel capacity of the slow network is interpreted which has a weaker ability to represent spatial semantics data as it is only meant for spatial coordination which is sensed, extracted and separated out for information gathering related to video classification. To be more precisely and technically, Fast pathway network doesn't coordinate with the spatial dimension, so as to maintain spatial modelling network capacity to be lower than the slow pathway because of some convolutional channels of the network hierarchy. The extracted provides efficient and primitive good results for this model and method is the desired tradeoff for the Fast pathway to suppress its spatial modelling ability to determine motion-related information.

5. EXPERIMENTS: VIDEO ACTION CLASSIFICATION AND RECOGNITION

We evaluate our approach on UCF Crime video recognition datasets using standard evaluation protocols. For the action classification experiments, presented in this section we consider the Anomaly dataset comprising of Normal videos and Crime videos from which different crime scenes are detected and evaluated as per the movements for action detection and recognition we have used the challenging UCF Crime dataset [20].

6. MAIN RESULTS

From the dataset, different classification type of segregation has been done and filtered out 16 different frames for every one sec of video and this process will be repeated for a complete video. This extracted frames will be identified and used for further investigation and hence classification is carried out.

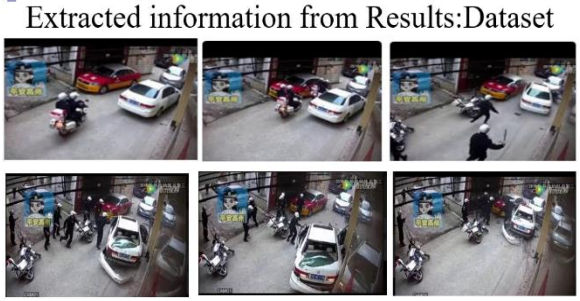

Fig. 2: Results for image retrieval showing Arrest type of Video Concept Classification

Fig. 3: Results for image retrieval showing Assault type of Video Concept Classification

Fig. 4: Results for image retrieval showing Normal type of Video Concept Classification

6.1 LOSS AND ACCURACY CALCULATION FROM ITERATIONS

The following table shows Loss and accuracy calculation for different iteration carried out while retrieving frames and hence can be said this is an efficient method to classify different type of actions.

Sr No	Iterations	Batch	Loss	Accuracy
1	0/10	437/438	1.922374	0.00
2	0/10	437/438	1.808457	49.00
3	0/10	437/438	1.743229	55.00
4	0/10	437/438	1.654741	63.00
5	0/10	437/438	1.579180	78.00
6	0/10	437/438	1.324348	79.00
7	0/10	437/438	1.056433	82.79
8	0/10	437/438	0.987698	89.16
9	0/10	437/438	0.765435	92.15
10	0/10	437/438	0.543567	93.64

Fig. 5: Table showing loss and accuracy Calculations

7. REFERENCES

[1] Panagiotis Sidiropoulos Vasileios Mezaris, "Video Tomographs and Base Detector Selection Strategy for Improving Large Scale Video Concept Detection", ISBN No.978-1-4799-3834-6/14/$31.00©2016 IEEE

[2] https://www.learnopencv.com/image-recognition-and-object-detection-part1/

[3] Gi-Hyun Na, Kyu-Sun Shim, Ki-Woong Moon, Seong G. Kong, 'Frame-Based Recovery of Corrupted Video FilesUsing Video Codec Specifications', IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 2, FEBRUARY 2014.

[4] Yong Guo, Li Chen, Zhiyong Gao, and Xiaoyun Zhang, 'Frame Rate Up-Conversion Method for

[5] Video Processing Applications', IEEE TRANSACTIONS ON BROADCASTING, VOL. 60, NO. 4, DECEMBER 2014.

[6] Chinh Dang and Hayder Radha, 'RPCA-KFE: Key Frame Extraction for Video Using Robust Principal Component Analysis', IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 11, NOVEMBER 2015.

[7] Andrei Stoian, 'Fast action localization in large scale video archives', IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY,

[8] Seokhwa Jeong, M. Younus Javed, Shaleeza Sohail, 'Multi-Frame Example-Based Super-Resolution Using Locally Directional Self-Similarity', IEEE not yet published.

[9] Jingjing Meng,'Object Instance Search in Videos via Spatio-Temporal Trajectory Discovery', IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 18, NO. 1, JANUARY 2016.

[10] https://www.intechopen.com/online-first/object-recognition-using-convolutional-neural-networks

[11] SlowFast Networks for Video Recognition Christoph Feichtenhofer Haoqi Fan Jitendra Malik Kaiming He Facebook AI Research (FAIR)

[12] C. Gu, C. Sun, D. A. Ross, C. Vondrick, C. Pantofaru, Y. Li, S. Vijayanarasimhan, G. Toderici, S. Ricco, R. Sukthankar, C. Schmid, and J. Malik. AVA: A video dataset of spatiotemporally localized atomic visual actions. In Proc. CVPR, 2018. 2, 4, 7

[13] Carlos Lopez, Nancy Arana Denial, "Image Classification Using PSO-SVM and an RGB-D Sensor", IEEE transactions on education, vol. 52, no. 3, July 2014.

[14] Maya Dawood, Cindy Cappelle, Maan E. El Najjar, Mohamad Khalil and Denis Pomorski," Harris, SIFT and SURF feature comparison for vehicle localization based on virtual 3D model and camera", 978-1-4673-2584-4/12/$31.00 ©2012 IEEE

[15] A. Murat Tekalp, 'Digital ViDeo Processing', Copyright © 2015 Pearson Education, Inc.

[16] Juan Cao, HongFang Jing, Chong-Wah Ngo 'distribution-based Concept Selection for Concept-based Video Retrieval', Copyright 2009 ACM 978-1-60558-608-3.

[17] Jianping Fan1, Hangzai Luo1, Ahmed K. Elmagarmid'Concept-Oriented Indexing of Video Databases: Towards Semantic Sensitive Retrieval and Browsing', the project is supported by National Science Foundation under 0208539-IIS and 9974255-IIS AO Foundation, HP, IBM, Intel, NCR, Telcordia and CERIAS

