International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

Study of SVM and CNN in Semantic Concept Detection

Jeevan J. Deshmukh[1], Nita S. Patil[2], Sudhir D. Sawarkar[3]

[1]*Department of Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai, India*
[2] *Department of Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai, India*
[3] Department of Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai, India

---

Abstract - *In Today's fast growing digital world, with very high speed internet videos are uploaded on web. It becomes need of system to access videos expeditiously and accurately. Concept detection achieves this task accurately and it is used in many applications like multimedia annotation, video summarization, annotation, indexing and retrieval. The execution of the approach lean on the selection of the low-level visible features used to represent key-frames of a shot and the selection of technique used for feature detection. The syntactic differences between low-level features that are abstracted from video and high-level human analysis of the video data are linked by Concept Detection System. In this proposed work, a set of low-level visible features are of greatly smaller size and also proposes effective union of Support Vector Machine(SVM) and Convolutional Neural Networks (CNNs) to improve concept detection, where the existing CNN toolkits can abstract frame level static descriptors. To produce the concept probabilities for a test frame, classifiers are built with help of SVM. To deal with the dataset imbalance problem, dataset is partitioned into segments and this approach is extended by making a fusion of CNN and SVM to further improve concept detection. Image or Video Frames are used to form Hue-moments and HSV histogram that yields into feature vector for classification. This paper contains mainly two parts first, the different approaches are discussed in literature survey where different techniques more or less following architecture of concept detection to reduce the semantic gap. Second, By going thoroughly with existing methods and classifiers used for concept detection author is presenting SVM and CNN as the Classifier, and the fusion of SVM and CNN will yield better results than existing ones. Two classifiers are independently trained on all sectors and fusion of two classifiers is getting executed to get desired output for the class and concept is detected. The accomplishment of the projected structure including fusion of SVM and CNN is comparable to existing approaches.*

Key Words: **Support vector machine, Video Concept Detection, Convolutional Neural Network, Key Frame Extraction, Feature Extraction**.

1. INTRODUCTION

Videos are becoming trendy for entertainment from several years. The Development of the video browsing and indexing application is growing faster as the end user feel necessity for better control over the video data. The Methods like video indexing, video browsing and video retrieval are taken into more consideration for Content-based video analysis due to rapid growth of video data. The concept probabilities for a test frame are produced by classifiers like SVM. The modern concept detection system consists of low-level feature extraction, classifier training and weight fusion. Earlier researchers focused on improving accuracy of the concept detection system using global and local features obtained from key-frame or shot of the video and various machine learning algorithm. In recent times, due to the technological advances in computing power deep learning techniques specially Convolutional Neural Network (CNN) has shown promising improvement in efficiency in various fields. CNN has the powerful ability of feature extraction and classification on large amount of data and hence widely adopted in concept detection systems. The Proposed method is to incorporate SVM and CNN for the challenging video concept detection problem due to its known ability to classify feature vector and gives efficient output. The video frames undergo different layers of the CNNs for descriptor extraction.

2. Literature Survey

Markatopoulou, Foteini [1] discussed the architecture for video concept detection that has enhanced the computational complexity as matched with typical state-of-the-art late fusion architectures.

Tong, Wenjing, et al. [2] proposed a novel video shot boundary detection method where CNN model is used to generate frames' TAGs. It is efficient to find out both CT and GT boundaries and also combines TAGs of one shot for implementation of video annotation on that shot.

Karpathy, Andrej, et al. [4] used CNN for large-scale video classification where CNN architectures are efficient to learn persuasive features from weakly-labeled data that is the best feature based methods in execution.

Krizhevsky, Alex, Ilya Sutskever [3] proposed deep Convolutional neural network and presented that deep-CNN is efficient to accomplish best output on highly challenging data sets using completely supervised learning. Xu, Zhongwen, Yi Yang [5] is first to use CNN descriptors for video representation and proposed that CNN descriptors are generated more accurately with help of suppressed concept descriptors.

Ciresan, Dan, et al [6] proposed that the DNN is generic image classifier with raw pixel intensities as inputs, without ad-hoc post-processing.

Snoek, Cees GM,[7][8] discussed to develop mapping functions from the low-level features to the high-level concepts with some machine learning techniques for concept detection or high level feature extraction. Janwe, Nitin J., and Kishor K. Bhoyar [9] proposed that in concept detection method the concept detection rate is directly controlled by semantic gap. The semantic gap is controlled by considering set of low level visual feature of very smaller size and selecting the feature-fusion methods like hybrid-fusion to improve performance of concept detection.

3. Overview of Concept Detection Method

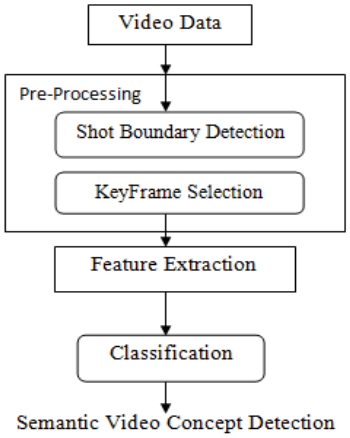

Fig.1: Architecture of Concept Detection System

Fig.1 shows the general architecture of concept detection consists of three different stages. First Preprocessing stage in which input video is divided into shots and from each shot key frames are extracted. Second step is Feature Extraction where features are extracted from key frame. Third Step is Classification step in which classifiers like SVM, CNN are used to predict the class for data.Fig.2 shows the proposed approach of system.

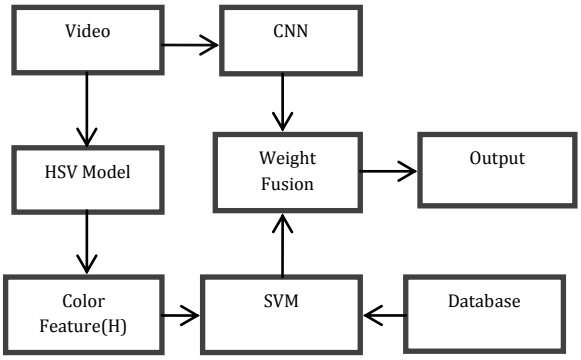

Fig.2: Proposed Methodology

3.1 Key Frame Extraction

In video processing applications, the key frame is frame of video which represents shot in the video. Video consist of Number of frames. The group of Frames will form a shot and set of shots will form a scene. The same shot contains many similar frames; therefore positive frames that perfectly revert the shot contents are selected as key-frames. Some times; shot is represented by many key-frames as per requirement. The selection of a key-frame may also base on the object or the event end user required. Whichever frame that perfectly represents the object or the event can be chosen as a key-frame.

3.2 Feature Extraction

- HSV Model

In HSV model, Hue is standard of the wavelength appeared in dominant color collected by sight while saturation is height of the size of white light mixed in hue. In Mathematical field Image is function of two dimensions which is in continuity with intensity of light in the field. In order to process image digitally through computer it must be presented with discrete values called as Image digitization in which digital form image is presented by two dimensional matrixes. Hue moments are group of 7 numbers that are calculated with help of central moments that are proportional to image transformation. It is proved that first 6 moments are proportional to translation, scale, reflection and rotation where 7^{th} moments flag changes for image reflection.

3.3 Classification

a. Support Vector Machine(SVM)

Support vector Machine is selective classifier specified by separating hyper-plane. The proper labeled training data is given and algorithm outputs with optimal hyper-plane which categorizes into two parts. In two dimensional plane cases, the hyper-plane divides space into two sets where each group lay in either side. SVM is machine learning algorithm originated from statistical learning theory. SVM is based on the risk minimization principle that is used to minimize the generalization error. Support vectors are trained data present along hyper-plane and close to class boundary. The SVM works on concept of decision planes where decision boundaries are defined by the decision planes. The classification method includes training and testing of data which consists of some data instances. The data instances present in training set have one class label i.e. target value and various features. The regularization parameter used to optimize the SVM and to avoid misclassification of training set. For large values of regularization parameter small margin hyper-plane is selected for optimization to get all training points classified correctly. In vice-versa for very small values of parameters larger margin separating hyper-plane is used as a result hyper-plane misclassifies more points. Gamma parameters value indicates how long the impact of single training set reaches. If gamma value is low then points present far away from hyper-plane are considered in calculation where as if gamma value is high then points close to hyper-plane are

considered for calculation.Fig.3 shows the working of Support vector machine. It shows hyper-plane formed in space which maximizes the margin between the two classes. The Support Vectors define the hyper-plane.

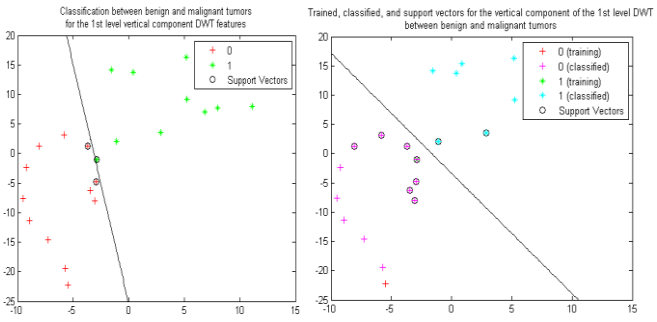

Fig.3: SVM

b. Convolutional Neural Network

Convolutional Neural Network is type of Neural Networks where features are extracted by using weights of convolution layer. In The fully connected layer the output matrix from pooling layer is converted into vector and classification function being used for classifying it into appropriate class which is determined during the training process. One can add many Convolutional layers till it satisfies. The pooling layer performs reduction of dimensionality size of image matrix. The CNN architecture consists of 7 layers. In The CNN input data structure comprise of 3dimensional array h*w*d in which h is height, w is width and d consider as number of feature channels. Height and width are considered as special dimensions for feature extraction. CNN is trained during training phase with set key frames and in test phase CNN gives desired prediction probability. In CNN, prediction calculation of class is done by last layer with the help of Softmax function. The term Fully-Connected implies that every neuron in the previous layer is connected to every neuron on the next layer. The output from the Convolutional and pooling layers represent high-level features of the input image. Figure 4 shows the basic architecture of CNN which consists of Convolution layer, pooling layer, Fully Connected layer. The convolution layer takes the input in the matrix form and also apply filter to extract features from input data. The pooling layer apply sub the sampling to matrix. The Fully connected layer is output layer where the class is predicted.

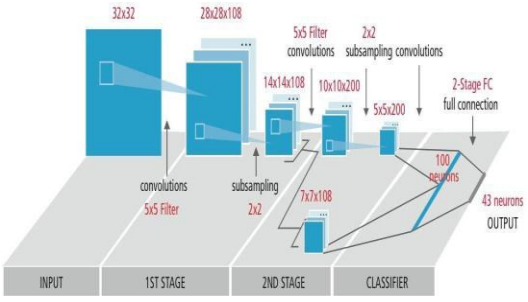

Fig.4: Basic Architecture of CNN

4. CONCLUSION

Author studied the different methods of concept detection including DCNN, SVM, CNN, multiclass SVM, CNN and FDCCM.etc. In concept detection system semantic gap directly affects on concept detection rate. If semantic gap is low then accuracy of concept is high. In order to keep semantic gap low system must have set of low level visual features with very smaller size and selection of feature fusion methods to correctly classify the features. In The method of fusion of CNN and FDCCM concept detection rate is improved but with FDCCM keeps co-occurrence data foe every foreground data with other data. Several methods are using context relationships of data to improve accuracy of concept detection but not focused on to better achieve the nature of the concept. It is observed that over the above methods in literature proposed method using combination of CNN and SVM with Hue_moments of images is more significant for concept detection. Output of CNN and SVM is combined to accurately class and concept is detected. The Proposed method is better than existing methods as fusion of CNN and SVM yield better result.

REFERENCES

[1] Markatopoulou, Foteini, Vasileios Mezaris, and Ioannis Patras. "Cascade of classifiers based on binary, non-binary and deep convolutional network descriptors for video concept detection." *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015.

[2] Tong, Wenjing, et al. "CNN-based shot boundary detection and video annotation." *2015 IEEE international symposium on broadband multimedia systems and broadcasting*. IEEE, 2015.

[3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

[4] Karpathy, Andrej, et al. "Large-scale video classification with convolutional neural networks." *Proceedinas of the IEEE conference on Computer Vision and Pattern Recognition*. 2014.

[5] Xu, Zhongwen, Yi Yang, and Alex G. Hauptmann. "A discriminative CNN video representation for event detection." *Proceedinas of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[6] Ciresan, Dan, et al. "Deep neural networks segment neuronal membranes in electron microscopy images." *Advances in neural information processing systems*. 2012.

[7] Snoek, Cees GM, and Marcel Worring. "Concept-based video retrieval." *Foundations and Trends® in Information Retrieval*2.4 (2009): 215-322.

[8] Snoek, C. G. M., et al. "MediaMill at TRECVID 2013: Searching concepts, objects, instances and events in video." *NIST TRECVID Workshop*. 2013.

[9] Ianwe, Nitin I., and Kishor K. Bhovar. "Neural network based multi-label semantic video concept detection using novel mixed-hybrid-fusion approach." *Proceedings of the 2nd International Conference on Communication and Information Processing*. ACM, 2016.