Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.199

IJCSMC, Vol. 9, Issue. 2, February 2020, pg.49 – 55

VIDEO TAGGING USING DEEP LEARNING: A SURVEY

Dr. A. Anushya
Assistant Professor, Department of Computer Applications
anushya.alpho@gmail.com

ABSTRACT: ***This survey paper encapsulates the various deep learning algorithms employed on video data. Scarce of research has been conducted for classification, clustering and tagging the videos in social networks such as YouTube, Facebook. Instead reading we prefer to watch, hence video huddling is obligatory. The analysis and review of deep learning models for video are explained in this paper. Deep learning is merely in its infancy and, in the epochs to come, will transmute the public. Also, this article discusses the efficiency of deep learning models and concludes with thoughts on future augmentation in video clustering.***
Keywords: Video clustering, Video tagging, Deep Learning, Convolutional Neural Networks

1. INTRODUCTION

Enormous spread of smartphones and social media such as Face Book, You Tube and Instagram are considered as valuable source of Big data. Video possibly will be considered as the most important source of Big data [11]. Now YouTube activates as unique of Google subsidiaries. YouTube is an American video sharing social media with 1.9 billion users worldwide. Registration is not required to watch the most videos, but required for uploading videos, accordingly every person can rifle and watch effortlessly. Moreover the user gifted for creating playlists, liking or disliking video, comments posting. So, many people is used to watch YouTube video. 490 hours of video are uploaded to YouTube every minutes and 70% YouTube views come from mobile devices. While searching and watching in YouTube, relevant and irrelevant video will be in the queue. It will be kill our time and no necessary. Video

classification or clustering helps in multimedia contents understanding where automatic video content analysis. Deep Learning is a used for image classification with localization, object detection, object segmentation image style transfer, image colorization, image reconstruction, image super solution and image synthesis. Video is a sequence of images, so video will be segmented as a frames and classified or clustered.

Deep learning is a subfield of machine learning. While both fall under the broad category of artificial intelligence, deep learning is what powers the most human-like artificial intelligence. Deep learning systems require large amounts of data to return accurate results; accordingly, information is fed as huge data sets. When processing the data, artificial neural networks are able to classify data with the answers received from a series of binary true or false questions involving highly complex mathematical calculations. Examples of applications of deep learning describe as follows. Digital assistants like Siri, Cortana, Alexa, and Google now use deep learning for natural language processing and speech recognition. Skype translates spoken conversations in real-time. Many email platforms have become adept at identifying spam messages before they even reach the inbox. PayPal has implemented deep learning to prevent fraudulent payments. Apps like CamFind allow users to take a picture of any object and, using mobile visual search technology, discover what the object is. Google, in particular, is leveraging deep learning to deliver solutions. Google Deepmind's AlphaGo computer program recently defeated standing champions at the game of Go. DeepMind's WaveNet can generate speech mimicking human voice that sounds more natural than speech systems presently on the market. Google Translate is using deep learning and image recognition to translate voice and written languages. Google PlaNet can identify where any photo was taken. Google developed the deep learning software database, Tensorflow, to help produce Artificial Intelligence applications.

Remainder of this article is organized as follows. Machine learning vs. Deep learning is discussed in section 2. The review on video classification and video clustering are given in section 3. Section 4 suggests the proposal in video clustering of CCTV footage and section 5 concludes the paper with possible future enhancements.

2. MACHINE LEARNING VS DEEP LEARNING

Both Machine Learning and Deep Learning discovers patterns in Data, but they involve dramatically different techniques. Machine Learning and Deep Learning are both forms of Artificial Intelligence. Machine Learning algorithms are divided into supervised (training data are tagged with the class) and unsupervised (any labels that may exist are not shown to the training algorithm). Clustering, association and Dimensionality reduction are the unsupervised techniques whereas classification, regression are supervised methods. Both can handle classification where deep learning models incline to produce better fits than machine learning models. Deep learning is a form of machine learning but trained with more than one hidden layer between input and output. For example deep Neural Networks have upwards of 10 hidden layers. The most significant leap forward for neural networks happened because of the introduction of substantial amounts of labeled data with ImageNet, a database of millions of labeled images from the Internet. The cumbersome task of manually labeling images was replaced by crowdsourcing, giving networks a virtually unlimited source of training materials. In the years since technology companies have made their deep learning libraries open source. Examples include Google Tensorflow, Facebook open-source modules for Torch, Amazon DSSTNE on GitHub, and Microsoft CNTK. Data scientists prepare the inputs, selecting the variables to be used for predictive analytics. Deep learning, on the other hand, can do this job automatically. Deep learning can be considered as a subset of machine learning. It is a field that is based on learning and improving on its own by examining computer algorithms. While machine learning uses simpler concepts, deep learning works with artificial neural networks, which are designed to imitate how humans think and learn.

3. LITERATURE REVIEW ON VIDEO CLUSTERING AND VIDEO CLASSIFICATION

In literature, two types of video classification or clustering approaches. The first type contains the methods the extract global and or local feature such as number of shots, average color histogram, HOG, HOF are fed to classifier (SVM, KNN) or clustering algorithm (k-means). In the second categories the approaches based on features extracted from selected key frames. These feature are the output of certain fully connected layer in the pre-trained CNN network after application of the feed forward algorithm [11]. This section stretches review of related research carried out previously are expounded as follows.

D. Saravanan et.al., retrieved the video data by histogram clustering. Video was converted into sequences of frames. Then image matrix is used to identify the centroids to remove the duplicate frames in video, and image pixel was used to create cluster. The input frame was fragmented into columns and rows. After wards matrix cell histogram was calculated to retrieve the video [3]. Correspondingly D. Saravanan et.al., conducted an analysis of comparison result of existing three clustering algorithm BRICH, CURE, CHAMELEON and concluded as with using CHAMELEON mechanism image changed into pixel and images does not belong to the cluster. While using BRICH, labeling the grids made some error while clustering But CURE cluster is used for clustering by eliminating outliers. The centroids are clustered and dataset will be minimized [4].

Mathidle Caron et.al., presented Deep cluster to unsupervised training of convolutional neural networks on Image Net and YFCC 100M and outperformed results [8]. Gokhane Cagrici constructed a hybrid of CNN and RNN on the dataset, UCF101 which has 13320 video with 101 action and around 70-75 % accuracy attempted [5]. Travis Addair generated a classifier to label a collection on YouTube videos with up to 20 tags. By a hybrid CNN-RNN architecture for image features from video and LSTM model for label set [9].

Y.G.Jiang et.al., proposed Regularized Deep Neural Networks (DNN) to extract video, audio and trajectory features for each video and combines them using a series of deep fully connected layers [10]. Andrzej Matiolański, et.al., described classes of images with their fuzzy portraits. The fuzzy set is calculated as a preliminary result of the algorithm before the final decision or rejection that solves the problem of uncertainty at the boundaries of classes to solve the problem of knife detection in still images using MPEG-7 descriptor schemes as feature vectors. The method was experimentally validated on a dataset of over 12,000 images. The article pronounced the results of six experiments which confirm the accuracy of our method [2].

Andrej karpathy et . al., studied CNN on sports. IM dataset which consists of 1 million YouTube video with 487 classes notably CNN is a powerful architecture for learning features from weakly labeled data. Also, retraining the top 3 layers of the network on UCF 101 the video classification achieved highest transfer learning [1]. Zein Al Abidin Ibrahim, proposed a new deep learning based video representation method. The features are extracted to detect the key frames in the video. Each video is represented by deep features video matrix (DFVM). We conducted

experiments on 5261 videos from BlipTV dataset with 25 different categories, by the proposed a new method named video to vecs to get efficient result [11].

Jingyi Hou et. al., conducted experiments on HM DB51 and UCF101 datasets. HMB51 consists of 51 action categories with 6,766 realistic video clips from Internet, whereas the UCF 101 comprises of 101 action categories with 13320 realistic video clips from YouTube. Also they proposed unsupervised deep learning method for action recognition by jointly clustering and learning features using deep neural networks. Results were demonstrated the superiority of proposed model [7]. Jingya Wang et. al., used TRECVID MED 2011 dataset to evaluate the efficiency of the proposed HML-RF model for tag-based video clustering. TRECVID MED 2011 includes 2379 video samples from 15 categories such as board trick, feeding animal and soon. They aimed to group these video into the above categories using clustering by visual features. The proposed method outperforms better other methods [6].

Based on the above literature, researcher deliberates to propose and implement the video from YouTube or CCTV footage in various public place to cluster according to need. The proposal will be elucidate in next section.

4. PROPOSED METHODOLOGY

Clustering is one of the unsupervised mechanism, without labeled data. It is grouping datasets by similarities. A good clustering method will provide high quality of clusters, which should have high intra class similarity (cohesive within clusters) and law inter class similarity (Distinctive between clusters). The clustering of video may involve the following steps.

Steps:

1. Input the video
2. Segment the video into frames
3. Extract the image frame from videos.
4. Train the top layer of an Inception V3 CNN with the input images.
5. Extract sequence of images from video with a constant size and equally speed.
6. Apply clustering techniques to detect any object
7. Cluster the videos based on the image frames.

In future above steps will be implemented in PYTHON for video with the assistance of deep learning techniques by means of modification also. The following subdivision concludes this survey article and future planning.

5. CONCLUSION

This paper presents the survey of video data analysis by dint of classification and clustering using deep learning contrivance. The deep learning methods are working without human interaction and works as human. Deep learning methods are the promising techniques to handle big data especially video data. The supported literatures are presented and survey is given more idea to work with video. Further research will be aimed at assessing a set of sequential images from video, and using a combination of our method with other approaches. . Research will focus with scrutinizing video footage acquired using CCTV systems. A knife in the human hand is an example of a motion of risk. Such extracts are usually vibrant and speedy. To solve the problem of knife recognition in frames from video sequences. The tactic can be used for dissimilar pattern recognition problems with non-uniform classes where the object has a specific form, such as the knife, gun, and masks etc. in CCTV footage. Besides research anticipates to oversimplify to solve the detection problem for an extensive variety of objects in an automatic way. Another place to detect object from video is examination hall, where student can do malpractice. The items which are used for malpractice can be detected. Then, while searching and watching videos in social media, there is a queue for related video by the keywords and frequently watched video. Among these some of them are not relevant. Also, each video plays more than five minutes normally. So the most relevant video can be loaded to save our time by appropriate clustering automatically for the penetrating keywords.

REFERENCES

1. Andrej karpathy et.al., “Large scale video classification with Convolutional Neural Networks IEEE conference on comp vision and pattern RCC (CVPR) 2014.
2. Andrzej Matiolański, et.al., “CCTV object detection with fuzzy classification and image enhancement”, Multimedia Tools and Applications volume 75, pages10513–10528, 2016.
3. D. Saravanan et.al., “Video content retrieval using histogram clustering technique”, Procedia computer science, 50, pp:560-565, 2015.
4. D. Saravanan, “Performance Analysis of video data image Using Clustering Technique”, Indian Journal of Science and Technologies, volume a, Issue 10, 2016.
5. Gokhane Cagrici, “Video classification with Deep Learning”. 2019.

6. Jingya Wang et.al., "Video semantic clustering with sparse and Incomplete tag", 30th AAA1 conference on Artificial Intelligence (AAAI-16), 2016.
7. Jingyi Hou et al., "Unsupervised Deep Learning of Mid-Level Video representation for Action recognition," The thirty second AAA1 conference on Artificial Intelligence (AAA 1-18), 2018.
8. Mathidle Caron et.al., "Deep clustering for unsupervised learning of visual features', CVF open Access, 2018.
9. Travis Addair, "Deep Learning YouTube video tags",
10. Y.G.Jiang et.al., "Exploiting feature and class relationship in video categorization with regularized deep Neural Network, arXiv preprint, 2015.
11. Zein Al Abidin Ibrahim, "Video to vecs a new video representation based on deep learning techniques for video classification and clustering" SN applied sciences, 1, 560, 2019.

