Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July, 2021:1081 – 1086

Advertisement Detection in broadcasted videos using Transfer Learning and Support Vector Machine

Namrata Dave[a], Dr. Mehfuza S. Holia [b]

[a] Research Scholar, Gujarat Technological University, namrata.dave@gmail.com
[b] Assistant Professor, ,Birla Vishwakarma Mahavidyalaya, msholia@bvmengineering.ac.in

Abstract

Deep learning using convolutional neural networks is emerged as the best approach for object detection. Performance of convolutional neural network depends on its architecture as well as large dataset for training. To train the model with small dataset and get most accurate results we have adopted transfer learning approach. We have used Alexnet model to train on dataset consisting of keyframes of news and advertisement frames obtained from news video of DD Girnar, ETV Gujarati, TV9 news, Sandesh channels. Using learned weights of well-defined network trained on very large dataset, proposed work achieved very good results in related tasks by adding new dataset in training. To achieve good results, experiments were performed on different models with different layers to generate final results. In this paper, experiments performed with pretrained Alexnet model for training as well as classification is presented with the results obtained. To achieve better accuracy without compromising training time, another transfer learning approach with Alexnet as feature extractor and SVM as a classifier is proposed. To classify images, Support Vector Machine along with Bayesian optimizer is used to improve classification performance. Proposed approach gives 99.2 percent accuracy on the dataset of news video.

Keywords: Deep Learning, Support Vector Machine, Bayesian Optimizer, Advertisement Detection.

1. Introduction

In day-to-day video of news, sports, etc. major shares are comprising of advertisements which are unwanted information many times (4). Advertisement detection is having useful application in multimedia processing. Processing video to cut portions of advertisement is a tricky part in advertisement detection task. With help of deep learning approach, it can be achieved with small dataset of images for learning and training the model.

The concept of Deep transfer learning is widely used in many models. Deep transfer learning is analogous to concept of fine tuning well know training method in deep neural network. The choice of retrained layer plays a significant role in deep transfer learning concept as com-pared to conventional methods. In the conventional method, the whole network is trained again for brand sparking new categories. Although it can achieve good results in many cases, it is obviously not the best one because it is hard to know how much the previous training process helps. Besides, only retraining the last layer cannot guarantee the best results either, because categories for source training data are usually different and the semantic representations in the higher layer are quite specific to the training categories. Therefore, it is not that suitable to use the highly specific semantic representation in the new recognition task. The reason for this phenomenon lies in the semantic representation for recognition.

In convolutional neural network, higher layers of network represent high level semantic features such as objects or parts of objects whereas lower layers mostly represent features such as edges, textures, colors (3) etc. considered as low-level features. Low level features represent similar information most of the times in many systems. Compared to low level features, the high-level features are not same for different categories of objects. Due to this fact, any good scheme used for transfer learning should have the capability of learning features

which are com-mon features and tune the high-level features according to the categories of target. To achieve best results using transfer learning, we have carried our variety of experiments to get the right choice for already trained layers.

In this paper, we have proposed advertisement detection using a transfer learning scheme which used pretrained Alexnet model, svm classifier, Bayesian optimizer to achieve the better results compared to state of art work. Section 2 of paper gives brief idea about structure of video; section 3 explains some related work carried out by researchers. In section 4, Details of experiment, dataset and results are given.

2. Video Structure

Video is comprised of various parts which can be processed with different techniques to extract meaningful information related to experiment. Video can be visualized as sequence of frames. Meaningful division of video can be described as collection of scenes and shots (10). Shot is taken in single camera action. Collection of similar shots makes scene. Each shot is collection of frames. Frame of video can be used as a basic unit for processing visual information.

Figure 1 Five consecutive Frames from sequence number 671 to 675 from a news video clip

Many frames of shot can have similar information. Normally frame rate for News video of regional language like Guajarati is 25 frames per second. Video can be described as sequence of similar frames with very less change in visual information as shown in figure 1.

To represent shot with minimum redundancy, key frame can be extracted to represent shot with all necessary information. Key frame extraction can be achieved using many algorithms using different kind of visual and audio features of video (6). Also, temporal features play vital role in video processing algorithms.

3. Related Work

Commercial or advertisement detection as well as removal can be achieved for different category of broadcast videos such as news, movies, sports etc. (1)(2). Mostly visual features are used for advertisement detection from video. Also, combination of visual features as well as acoustic features are used for commercial detection from news videos (1)(3). Key frame extraction and video segmentation is important part of video processing in many applications (4)(5)(6). Methods based on difference of histogram, block level differences of frames or histogram of frames, entropy of frame etc. are used successfully by many researchers (7)(8).

Classification of advertisements from video has been achieved using various visual features such as Edge Change Ratio ECR(9), difference of frames , length of video shot(10), text location in the frame and availability of specific text bands in news video frames(6), spectral centroid, flux, roll of frequency(11), short time energy(12), zero crossing rate (ZCR) (13) etc. Audio features such as MFCC Mel Frequency Cepstral Coefficients (14) which is very much used in speech processing applications, is used to differentiate advertisements in news channel effectively (1). It is observed that when advertisement starts the audio loudness increases dramatically. This fact is used by many researchers for advertisement detection in news videos as well as in sports video.

Different classifiers like Support Vector Machine (15), Neural Network (2) etc. are used to classify advertisements from other frames. Features extracted by deep network is recently being used efficiently for feature extraction in commercial detection by researchers. Convolutional Neural Network also being used by many researchers for advertisement detection task (15).

SVMs are one of the most prevailing and robust classification and regression algorithms in various fields of application. The SVM has been playing an important role in pattern recognition. Research in some fields where SVMs do not perform well has spurred development of other applications such as SVM for large data sets, SVM for multi classification and SVM for unbalanced data sets. Further, SVM has been integrated with other optimization algorithms to improve the capability of classification and optimize parameters. (16)

Alexnet model is a well know model developed for ILSVRC 2012 (ImageNet Large Scale Visual Recognition Challenge) (7), (9), (17). The Alexnet network attained a top-5 error of 15.3%, more than 10.8 percentage points lower than that of the runner up. Primary results given in original paper was based on the fact that to achieve higher performance, model should be deep enough. Ideal of having deep model was computationally expensive. It was achieved due to the use of graphics processing units (GPUs) in training the network (17).

4. Experimental Details

The proposed method exploits the concept of transfer learning with Alexnet model for advertisement detection from news video dataset. Dataset is created with the collection of news video data of DD Girnar, ETV Gujarati, tv9 news and Sandesh broadcasted news channel of Gujarati language. Proposed method for advertisement detection is explained in figure 2. As shown in figure 2, Frames are extracted from video input. Next task of extracting key frames is performed using singular value decomposition SVD and ranking based algorithm (6). Feature extraction from key frames are performed with Alexnet model for the small size datasets followed by binary classification with SVM to obtain desired advertisement detection task. Normally in news videos, each news channel displays news in specific location as text band. This concept is used to separate news and advertisement sections successfully.

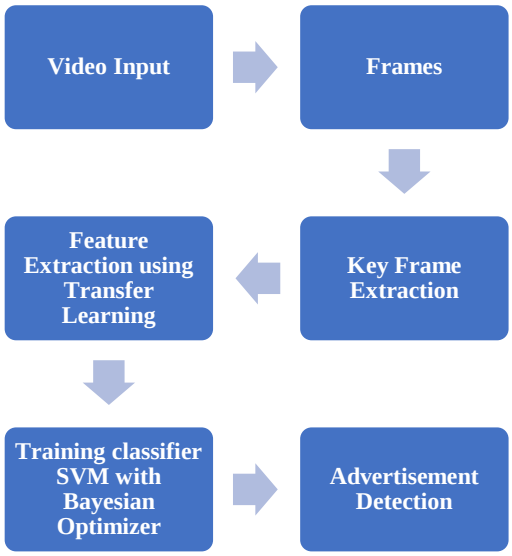


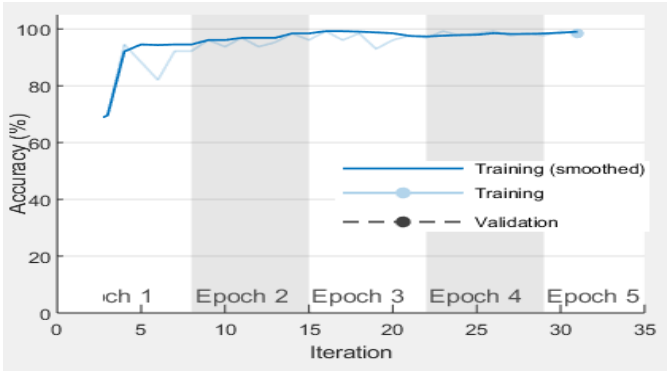
Figure 2 Proposed Method for Advertisement detection

Dataset is trained using pretrained Alexnet model with 1268 key frames of broadcasted videos of Gujarati News Channels DD Girnar, ETV Gujarati, Sandesh and TV9. Key frames are divided into two classes advertisement and news for training and testing purpose. Dataset has been divided in 75:25 proportion for training and testing of system.

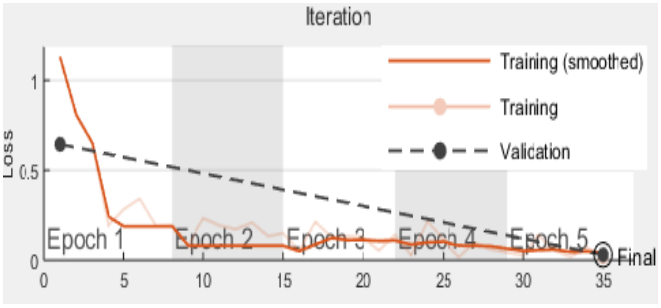
Architecture of Alexnet model has 5 convolutional layers and 3 fully connected layers. Activation Relu is applied after every layer. Dropout is applied before the first and the second fully connected year.

Transfer learning approach is applied with Alexnet network. Changes applied in last layers of network to train the model for dataset of news video frames for advertisement detection task. Although, Main idea of Alexnet is object detection, the model fits perfectly to task of advertisement classification.

In the experiments performed for transfer learning using Alexnet, the model is trained using dataset created for advertisement and news classification task. Plot of training and validation accuracy of experiments performed with Alexnet pretrained model is shown in figure 3 (a). In figure 3 (b) plot of training and validation loss is given. Proposed approach using transfer learning with Alexnet pretrained model gives validation accuracy 98.43 percent, validation loss is 0.02. Base learning rate is taken as 1.0000e-04 with 5 epochs and 50 iterations per epochs used in training.



(a)



(b)

Figure 3(a) Training and Validation Accuracy **(b)** Training and Validation Loss of proposed approach using transfer learning with Alexnet pretrained model

To improve accuracy of classification and reduce training time, another approach proposed here which uses pretrained Alexnet model as feature extractor along with support vector machine binary classifier and Bayesian optimizer to classify our data further into news and advertisement classes.

In binary classification task svm has performed well compared to other classifiers due to its kernel trick to handle nonlinear input spaces. SVM finds an optimal hyperplane which helps in classifying new data points. Due to this fact, we have applied classifier SVM with Bayesian optimizer to boost the classification performance.

Bayesian optimizer attempts to minimize a scalar objective function f(x) for x in a bounded domain. The function can be deterministic or stochastic, meaning it can return different results when evaluated at the same point x. Gaussian process model of function f(x), Bayesian update procedure for modifying the Gaussian process model at each new evaluation of function f(x) and acquisition function a(x) (based on the Gaussian process model of f) that you maximize to determine the next point x for evaluation are the key elements of Bayesian Optimization. (18)

Table 1 Confusion matrix of classification of news vs advertisements

	ADV	**NEWS**	
ADV	83 32.7 %	0 0.0 %	100 % 0.0 %
NEWS	2 0.8 %	169 66.5 %	98.8 % 1.2 %
	97.6 % 2.4 %	100 % 0.0 %	**99.2 %** 0.8 %

The model can be seen as function E_{CNN} described by equation 1, which generates f=4096 features vectors $F_{i,k}$ for video V_i from the dataset of videos.

$$F_k = E_{CNN}(F_{i,k}),\ \forall F_{i,k} \in V_i\ ,\ i = \quad 1,2,\dots,m\ ,k = ,2,..,n \qquad (1)$$

Feature Vector can be characterized as collection of features given by equation 2.

$$F_k = \left((f_k^{(1)}, f_k^{(2)},\ \dots, f_k^{(f)})^T\right) (2)$$

Figure 4 Result of classification of ADV vs NEWS.

Results of classification is shown in table 1 with use of confusion matrix. Out of total frames 32.7 percentage of advertisements frames correctly classified and 0.8 percentage of misclassification for advertisement class can be seen in figure 3. Whereas, 66.5 percentage of total frames are news frames which are correctly classified as news with no misclassification. Proposed method performs very well with accuracy of 99.2 percent reported.

In figure 4, the snapshot of results obtained by classification are shown. Frames with news stories are correctly classified and labelled as News and also Advertisement are classified correctly and labelled as ADV.

5. Conclusion& Future Work

We have achieved significant results with use of transfer learning method in advertisement detection from broadcast tv news videos. Proposed method gives accuracy of almost 99.2 percent with given experimental setup when classified using Support Vector Machine and Bayesian optimizer. Proposed work has been utilized for automatic removal of advertisements for further indexing of news key frames to achieve comparatively good results for our indexing and retrieval system. Proposed work can be investigated with a large-scale dataset and different architectures of deep neural network in future for better results and different datasets.

References

[1] A. Vyas, R. Kannao, V. Bhargava, and P. Guha(2014), "Commercial block detection in broadcast news videos," in ACM International Conference Proceeding Series, , doi: 10.1145/2683483.2683546.

[2] B. Zhang, T. Li, P. Ding, and B. Xu, "TV commercial segmentation using audiovisual features and support vector machine," in Proceedings - 2012 International Symposium on Instrumentation and Measurement, Sensor Network and Automation, IMSNA 2012, 2012, doi: 10.1109/MSNA.2012.6324579.\

[3] R. Kannao and P. Guha, "TV advertisement detection for news channels using Local Success Weighted SVM Ensemble," in 12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015, 2016, doi: 10.1109/INDICON.2015.7443801.

[4] P. M.Kamde, S. Shiravale, and S. P. Algur, "Entropy Supported Video Indexing for Content based Video Retrieval," Int. J. Comput. Appl., vol. 62, no. 17, pp. 1–6, 2013, doi: 10.5120/10169-9974.

[5] Z. Rasheed and M. Shah, "Detection and representation of scenes in videos," IEEE Trans. Multimed., 2005, doi: 10.1109/TMM.2005.858392.

[6] N. Dave, M. Holia, “Shot Boundary Detection for Gujarati News Video,” Int. J. Res. Appl. Sci. Eng. Technol., 2018, doi: 10.22214/ijraset.2018.3730.

[7] F. Garcia-Lamont, J. Cervantes, A. López, and L. Rodriguez, “Segmentation of images by color features: A survey,” Neurocomputing, 2018, doi: 10.1016/j.neucom.2018.01.091.

[8] H. Ji, D. Hooshyar, K. Kim, and H. Lim, “A semantic-based video scene segmentation using a deep neural network,” J. Inf. Sci., vol. 45, no. 6, pp. 833–844, 2019, doi: 10.1177/0165551518819964.

[9] X. S. Hua, L. Lu, and H. J. Zhang, “Robust learning-based TV commercial detection,” in IEEE International Conference on Multimedia and Expo, ICME 2005, 2005, doi: 10.1109/ICME.2005.1521382.

[10] N. Dimitrova, S. Jeannin, J. Nesvadba, T. McGee, L. Agnihotri, and G. Mekenkamp, “Real time commercial detection using MPEG features,” in Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowlwdge-based Systems (IPMU2002), 2002.

[11] N. Liu, Y. Zhao, Z. Zhu, and H. Lu, “Exploiting visual-audio-textual characteristics for automatic TV commercial block detection and segmentation,” IEEE Trans. Multimed., 2011, doi: 10.1109/TMM.2011.2160334.

[12] D. Mistry and A. Banerjee, “Comparison of Feature Detection and Matching Approaches: SIFT and SURF,” GRD Journals- Glob. Res. Dev. J. Eng., 2017.

[13] L. Zhang, Z. Zhu, and Y. Zhao, “Robust commercial detection system,” in Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, ICME 2007, 2007, doi: 10.1109/icme.2007.4284718.

[14] M. Goyani, N. Dave, and N. M. Patel, “Performance analysis of lip synchronization using LPC, MFCC and PLP speech parameters,” in Proceedings - 2010 International Conference on Computational Intelligence and Communication Networks, CICN 2010, 2010, doi: 10.1109/CICN.2010.115.

[15] M. Li, Y. Guo, and Y. Chen, “CNN-based commercial detection in TV broadcasting,” in ACM International Conference Proceeding Series, 2017, doi: 10.1145/3171592.3171619.

[16] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” Neurocomputing, 2020, doi: 10.1016/j.neucom.2019.10.118.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “2012 AlexNet,” Adv. Neural Inf. Process. Syst., 2012, doi: http://dx.doi.org/10.1016/j.protcy.2014.09.007.

[18] https://www.mathworks.com/help/stats/bayesian-optimization-algorithm.html