ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse180952020.pdf
https://doi.org/10.30534/ijatcse/2020/180952020

Calculating Screen Time of Characters in a Video Using Convolutional Neural Networks

Mohammed Inayathulla[1], Rokesh Kumar Yarava[2], Dorababu Sudarsa[3], Tanuja Betha[4]

[1]Malla Reddy Engineering College, India, inayathulla512@gmail.com
[2]Sri Satya Sai University of Technology & Medical Sciences, India, rokeshy1@gmail.com
[3]Koneru Lakshmaiah Education Foundation, India, dorababu.sudarsa@gmail.com
[4]Malla Reddy Engineering College, India, tanujabetha12345@gmail.com

ABSTRACT

In this growing world of digital technology, images and videos are one of the richest sources of data on the internet. Several technologies have emerged which help the users to deal with the image and video data. In this paper, we propose a methodology to calculate the screen time of characters appearing in the video by using the classification-count method. Supervised Learning Methodology with Convolutional Neural Network as the classifier is used in this paper. The proposed methodology is implemented on a series of Tom and Jerry videos available publicly on the internet. Experimental results show that the proposed methodology has shown good accuracy.

Key words : Classification-Count, CNN, Supervised Learning, Image classification

1. INTRODUCTION

The Screen time of a character is the total duration that involves the character in the film. It plays a vital role in determining the remuneration of the actor and for understanding the gender bias in films. Recent analysis shows that the male actors appear more often than female actors, and they share twice the screen time than females. In this paper, we implemented concepts of image classification using Convolution Neural Networks(ConvNet's). Image classification is about training the machine learning model to recognize the objects and classify them based on the previously fed images as input. It seems very useful, since, manually checking and classifying the massive number of images that are present in the world would be a strenuous task. Image classification can be implemented as a supervised learning technique or unsupervised learning technique. In a supervised learning technique, training sample classes are created by the user to train the model and to classify the new image, whereas ,in the unsupervised learning technique, a cluster of images is made by the model based on the pixels of the images. In the supervised learning technique, an array is formed having values between 0 to 1. These values are derived by normalizing the pixel values of the image that are in the range of 0 to 255. The classification is performed based on these array values. Many algorithms are used to implement image classification, but their accuracy in classifying needs consideration. To accurately classify images, the model should have the capability to understand the background of the image, camera angles, camera focus, and lighting. The model must have the ability to extract features like sharpness, color histograms, textures, and shapes that are derived from the pixels of the image. This approach is called feature engineering. These accuracy parameters are satisfied by using ConvNet's, which handles the selection of features from the image. By using ConvNet's, we can significantly reduce the time for classifying images.

The different types of tasks that are implemented using image classification are image acquisition, image enhancement, image restoration, image understanding, image recognition, and pre-processing of computer vision. Image classification plays an important role in many aspects of daily life. It is used in a variety of disciplines including fields of science and technology. It is used in medical diagnosis, where, in this case, the features can be the blood pressure, blood sugar level, lipid profile, and any other pathological data collected from the patient. The self-driving cars with image classification can recognize various objects and actions on the road. The image classification can be used in military surveillance to detect any unusual activities at the border and take automatic-decisions to save the lives of soldiers. It can be used to understand the different areas on Earth where human research is fiddly. There are many approaches to calculate the screen time of actors using image classification, but in this paper, we used ConvNet's for highly accurate results. The paper has been organized into the following sections, section II, discusses the related work done. Section III explains the proposed methodology for calculating the screen-time of an actor. Section IV discusses results found with the implementation, followed by our conclusion of using the Convolution neural network to find the screen time of an actor. Finally, we discuss future research to improve the accuracy using ConvNet's model.

2. RELATED WORK

Classification in machine learning is a supervised learning scheme. Image classification in machine learning involves preparing labelled image data and using it to train the machine for classification. Image classification is also applied on hyper spectral images. As having huge labelled data in remote sensing applications is a tough task semi supervised learning scheme using recurrent neural networks(RNN)[1] can be used. Shallow neural networks also can be used for image classification. Retinal fundus image classification using artificial neural networks (ANN)[2] was proposed by using publicly available dataset. The concept of hyper parameter tuning[7] in deep learning was introduced in image classification. Broad learning systems are special type of learning systems without deep architecture. Semi supervised learning with effective feature extraction[3] can be implemented in image classification. CNN model which is used in this paper can be applied for image classification including data augmentation[4] process for training the model. Support Vector Machine(SVM) is a popular supervised learning approach used for classification. Multiple SVM models[5] can be constructed for image classification based on AdaBoost framework. CNN's also found their applications in medical fields[6] like identifying the damaged retina of eye caused due to diabetes. Multi scale CNN [8] can be applied on fine grained classification problems by using linear scaling and random cripping methods. Auto Encoder Decoder[9] is another type of neural networks which is used for dimensionality reduction especially reducing noise in signals. Artificial Bee Colony in combination with ANN[10] also can be used for image classification. ANN with multi layer perceptron can be used as supervised classification scheme[11]. Similar image classification using ImageNet dataset[12] was proposed using deep CNN's. Pretrained deep neural networks[13] were used in pneumonia classification and unsupervised machine learning [14] techniques were used for retrieving of images of historical monuments.

3. PROPOSED METHODOLOGY

In this section, we describe the implementation for calculating the screen time of characters by using classification count method. The overview of classification count method is described in figure 1. We used convolutional neural network (ConvNet's), with which we categorized the implementation process into four stages: Data pre-processing, building CNN, training the CNN model and classification with the unseen data. We have used videos of Tom and Jerry that are available publicly on the internet. After collecting series of Tom and Jerry videos frames were extracted from these videos and manually annotated. For the data pre-processing, the video is captured, frame rate is set, and the frames are resized to 224*224*3. The formed image dataset is initially annotated with class labels and further divided as 70% for training the model and 30% for validating the model.

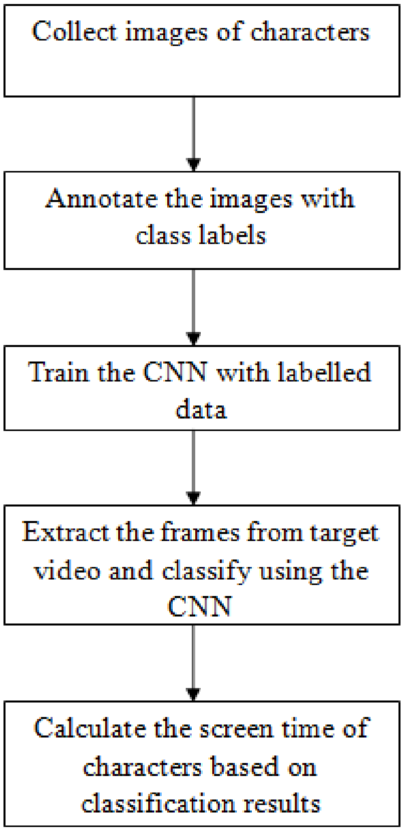
Collect images of characters
Annotate the images with class labels
Train the CNN with labelled data
Extract the frames from target video and classify using the CNN
Calculate the screen time of characters based on classification results

Figure 1: Overview of Classification Count Method

The dataset has three class labels 0 for Tom 1 for Jerry ad 2 for neither. The Convolutional neural network is built with 32 filters, and a kernel size of 3. We then used Max-Pooling strategy to normalize the input size using a pool matrix of 2*2, and with a stride of 2. Then applied these layers again, to further reduce the input size. We have applied Re-Lu activation function for the hidden layers, to increase the non-linearity of data, which eventually doesn't lead to large linear classifier. We have built the model with the training examples. In the output layer, we have used soft-max activation function to transform the values into probability distributions. We used the cross-entropy to calculate the loss, figure 2 describes the overview of CNN. After the CNN is build, from new videos frames were extracted and were applied to CNN for classification. The final classification count was used to calculate the screen time of characters.

4. RESULTS AND DISCUSSION

The proposed methodology was evaluated on 10 Tom and Jerry videos. Table 1 shows the results of classification count method of 10 videos.The use of Convolutional Neural Networks (ConvNet's) for the implementation has shown promising accuracy for calculating the screen time of characters. The screen time of characters can be used in calculating the remuneration of actors, and for analysing the gender discrimination among female and male.

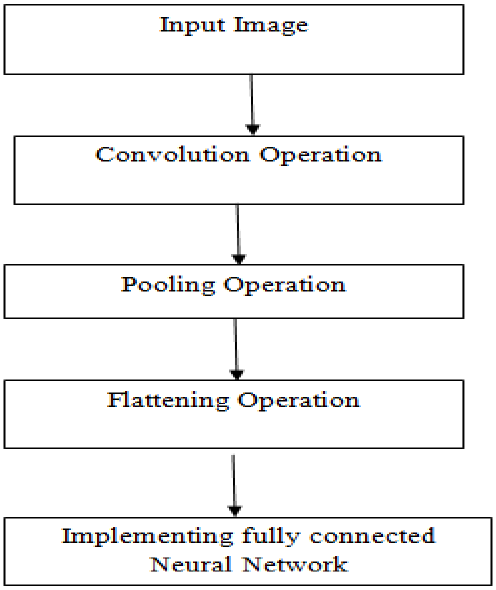

Figure 2: Overview of CNN

Table 1: Screen time of Tom and Jerry calculated using Classification Count Method

Video No	**Duration of Video**	**Screen Time of Tom**	**Screen Time of Jerry**
1	12 min 35 sec	312 sec	250 sec
2	14 min 25 sec	600 sec	120 sec
3	16 min 10 sec	712 sec	113 sec
4	8 min 40 sec	370 sec	116 sec
5	7 min 50 sec	190 sec	127 sec
6	9 min 10 sec	326 sec	139 sec
7	20 min 32 sec	895 sec	284 sec
8	13 min 25 sec	276 sec	180 sec
9	12 min 10 sec	538 sec	168 sec
10	10 min 20 sec	348 sec	215 sec

5. CONCLUSION AND FUTURE WORK

The proposed method has shown promising accuracy in calculating the screen time of characters. Unsupervised approach can be used in future in order to avoid the hectic task of manually annotating the images. The proposed approach has been applied on small size videos and hence in future lengthy videos must be used in order to evaluate the scalability and overall performance of the proposed methodology.

REFERENCES

1. Hao Wu, Saurabh Prasad, "**Semi-Supervised Deep Learning Using Pseudolabels For Hyperspectral Image Classification**", IEEE Transactions On Image Processing, Vol. 27, NO. 3, Mar 2018.
2. M. Omar, F. Khelifi, M.A. Tahir, "**Detection And Classification Of Retinal Fundus Images Exudates Using Region Based Multi Scale LBP Texture Approach**," In Proceedings of 2016 International Conference on Control, Decision and Information Technologies (CoDIT), St. Julian's, Malta, 6-8 April 2016.
3. Jianjie Zheng , Yu Yuan, Huimin Zhao, Wu Deng, "**A Novel Broad Learning Model-Based Semi-Supervised Image Classification Method**", IEEE Access 22 June 2020 pages:116756 – 116765.
4. Jinquan Li, Shanshan Xie, Zhe Chen, Hongwen Liu, Jia Kang, Zixuan Fan, Wenjie Li, "**A Shallow Convolutional Neural Network for Apple Classification**", IEEE Access 16 June 2020 pages: 111683 - 111692.
5. Huipu Xu, Haiyan Yuan, "**An SVM-Based AdaBoost Cascade Classifier for Sonar Image**", IEEE Access 23 June 2020 pages: 115857 – 115864.
6. Asra Momeni Pour, Hadi Seyedarabi, Seyed Hassan Abbasi Jahromi, Alireza Javadzadeh, "**Automatic Detection and Monitoring of Diabetic Retinopathy using Efficient Convolutional Neural Networks and Contrast Limited Adaptive Histogram Equalization**", IEEE Access 25 June 2020 pages: 136668 - 136673.
7. K. Shankar, Yizhuo Zhang, Yiwei Liu, Ling Wu, Chi-Hua Chen, "**Hyperparameter Tuning Deep Learning for Diabetic Retinopathy Fundus Image Classification**", IEEE Access 26 June 2020 pages: 118164 – 118173.

8. Chee Sun Won, "**Multi-Scale CNN for Fine-Grained Image Recognition**", IEEE Access 26 June 2020pages: 116663 - 116674.
9. Xiao-Yu Zhang, Changsheng Li, Haichao Shi, Xiaobin Zhu, Peng Li and Jing Dong, "**AdapNet: Adaptability DecomposingEncoder–Decoder Network for Weakly Supervised Action Recognitionand Localization**", IEEE Transactions On Neural Networks And Learning Systems, Dec 2019.
10. A. Kayabasi, "**An Application Of ANN Trained By ABC Algorithm For Classification Of Wheat Grains**", Int. J. Intell. Syst. Appl. Eng., vol. 1, no. 6,pp. 85–91, Mar. 2018.
11. Sabanci, A. Kayabasi, and A. Toktas, ‘‘**Computer Vision-Based Methodfor Classification Of Wheat Grains Using Artificial Neural Network**,’’J. Sci.Food Agricult., vol. 97, no. 8, pp. 2588–2593, Jun. 2017.
12. A. Krizhevsky, I. Sutskever, G.E. Hinton, "**Imagenet Classification With Deep Convolutional Neural Networks**," Communications of the ACM, vol. 60, no. 6, pp. 84-90, June 2017.
13. Julio Christian Young, Alethea Suryadibrata, "**Applicability of Various Pre-Trained Deep Convolutional Neural Networks for Pneumonia Classification based on X-Ray Images**", International Journal of AdvancedTrends in Computer Science and Engineering, Vol.9 pp.2649-2654, Jun 2020.
14. Jaimala Jha, Dr. Sarita Singh Bhaduaria, "**A Novel approach for Retrieval of HistoricalMonuments Imagesusing Visual Contents and Unsupervised Machine Learning**", International Journal of AdvancedTrends in Computer Science and Engineering, Vol.9 pp.3563-3569, Jun 2020.