SSRG International Journal of Electronics and Communication Engineering Volume 6 Issue 2, 5-8, February 2019
ISSN: 2348 - 8549 / doi:10.14445/23488549/IJECE-V6I2P102

Violence Detection System using Convolution Neural Network

Goutham Sakthivinayagam[1], Raveena Easawarakumar[1], Alagappan Arunachalam[1] and M. Pandi [2]
[1]*Student, Dr. Mahalingam College of Engineering and Technology,*
[2]*Assistant Professor (SG), Dr. Mahalingam College of Engineering and Technology,*
Udumalai Road, Annamalai Nagar, Makkinampatti, Pollachi, Tamil Nadu 642003

Abstract

The demand for automatic action recognition systems has increased due to a rapid increase in the number of video surveillance cameras installed in cities and towns. The main purpose of the algorithm is used to generate an alarm in case of abnormal activities and to assist human operators and for offline inspection. A challenge is to develop intelligent video systems capable of automatically analyzing and detecting the violence that occurred in the scene. This work describes and evaluates the uses of Convolution neural networks to identify the violent content from video scenes. Also, it demonstrates the results and effectiveness of the proposed method when applied to our datasets. The result shows that the proposed system is more efficient and more accurate. This system helps the police to identify the criminals much faster. It may increase the chances of the criminals being caught.

Keywords - *Violence detection, Neural networks, Convolution neural networks, Crime, Video surveillance.*

I. INTRODUCTION

The greatest concern of government and large private organizations is to keep cities and towns safe against violent actions and security breaches. These cameras are under the supervise of humans; we humans are always known for our careless errors and mistakes during an operation. So certain suspicious activities cannot be identified by the human effortlessly.

Hence there is always a demand for computer vision technology to analyze and detect the action that is occurred in the scene. In the early days, to analyze a picture is considered a very high computing problem because the hardware for the computer is in a very infant stage. Due to computing and storage limitations, there is not quite good many software, but theoretically, the algorithm is proved and explained decades ago. The bottleneck problem is eliminated due to the availability of high-performance hardware and also robust programming languages.

In this method, we have suggested using a deep learning approach, such as a convolution neural network with a binary classification of either violence and non-violence. Why we use this approach? Convents have outperformed all other machine learning approaches in the IMAGE NET challenge. Furthermore, a video is a collection of single framed images that are moving in a fast phase. This approach starts with dissecting the frame from the image and classifies it, whether it is violence or non-violence.

II. BASIC CNN COMPONENTS

Convolutional neural network layer types mainly include three types, namely Convolutional layer, pooling layer, and fully-connected layer. Fig. 1 shows the architecture of LeNet-5[1], which is introduced by Yann LeCun.

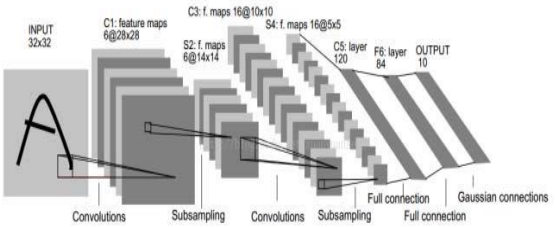

Figure 1 The architecture of LeNet-5[1] network (adapted form[1])

A. Convolution Layer

The convolutional layer is the core part of the Convolutional neural network, which has local connections and weights of shared characteristics. The Convolutional layer aims to learn feature representations of the inputs. As shown above, the Convolutional layer is consisting of several feature maps. Each neuron of the same feature map is used to extract local characteristics of different positions in the former layer, but for single neurons, its extraction is local characteristics of the same positions in a former different feature map. In order to obtain a new feature, the input feature maps are first convolved with a learned kernel, and then the results are passed into a nonlinear activation function. We will get different feature maps by applying different kernels. The typical activation functions are sigmoid, tanh, and Relu[15].

B. Pooling Layer

The sampling process is equivalent to fuzzy filtering. The pooling layer has the effect of the second feature extraction; it can reduce the dimensions of the feature maps and increase the robustness of feature extraction. It is usually placed between two Convolutional layers. The size of feature maps in the pooling layer is determined according to the moving step of kernels. The typical pooling operations are average pooling [16] and max-pooling [17]. We can extract the high-level characteristics of inputs by stacking several Convolutional layers and pooling layer.

C. Fully-connected Layer

In general, the classifier of a Convolutional neural network is one or more fully-connected layers. They take all neurons in the previous layer and connect them to every single neuron of the current layer. There is no spatial information preserved in fully-connected layers. The last fully-connected layer is followed by an output layer. For classification tasks, sigmoid regression is commonly used because of its generating a well-performed probability distribution of the outputs. Another commonly used method is SVM, which can be combined with CNNs to solve different classification tasks [18].

III. CONVOLUTION NEURAL NETWORK FOR VIOLENCE DETECTION

Here we use the depth Convolutional neural network; it is obvious that the expression ability of the network is enhanced with the depth of the network. The time complexity of the same two network structures, the deeper the performance of the network will have a relatively improved. However, the network is not as deep as possible. As the depth of the network increases, the memory consumption will be more and more, and the network performance may not improve.

So based on this idea, we have developed a network that can able to perform well on the single image classification as well as a serial of image classification at the same time. Here, our convolutional neural network analyses are mainly used for video sequence prediction; if we deconstruct a video, is It compiled with frames and frames are images. A video is an illusion of continuous moving images. The idea is if we classify the images in a frame we can able to classify the screen of a video and sequence in a video.

The following shows the data flow of the neural network structure(Our network is based on our own Dataset). Our simple Convolutional neural network introduces relu[15] and dropout[19]. Since it is a binary classification, so it is better to use the standard sigmoid function. In order to classify the images in scene we use 0 and 1 output classes as violence and non-violence scenes. In addition, Relu[15] is just a non-linearity that is applied to the neural network and, the formula is g (x) = max (0, x). The formula lets the value equal to 0 if the output value is less than 0; otherwise, keep the original values. This is a simple and brutal way to force some data to 0, but the practice has proved that the network is fully trained with a moderate sparse. Moreover, the visualization effect of training is similar to that of the traditional method, which indicates that relu[15] has the ability to guide the sparse. Training Convolutional neural network, when the iteration times increase, there will be a good fit training set, but the degree of fitting to the verification set is very poor. We follow a certain probability on the weight of the parameters of random sampling, selection of updates in the training process. Dropout[19] is the model training in the random network layer hidden layer node weights do not work, but to retain its weight, but not to be updated. The whole idea was implemented in TensorFlow, which is a library for machine learning and deep learning. The following shows the entire data flow graph created by TensorFlow.

A. Data flow Diagram of Convolution Layer 1

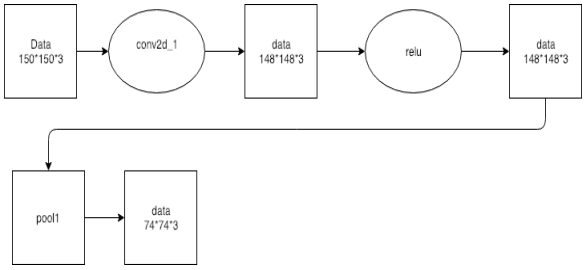

Figure 2 Data Flow Diagram of Convolutional Layer1

B. Data flow Diagram of Convolution Layer 2

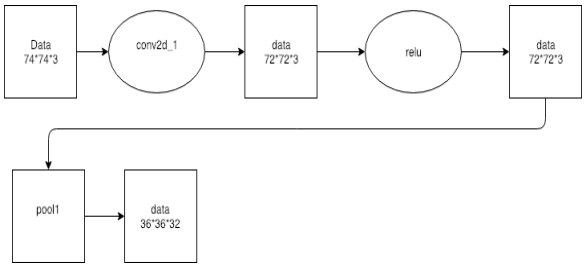

Figure 3 Data Flow Diagram of Convolutional Layer 2

C. Data flow Diagram of Convolution Layer 3

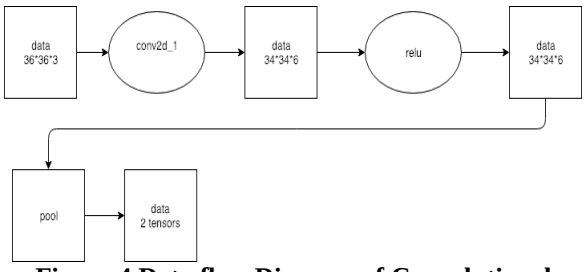

Figure 4 Data flow Diagram of Convolutional Layer 3

D. Data flow Diagram of Dense layer 1 and flatten

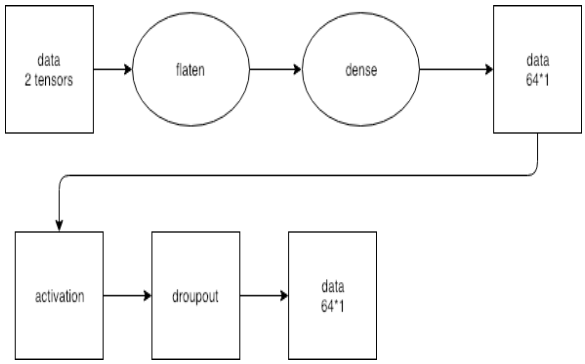

Figure 5 Data Flow Diagram of Dense layer 1 and flatten

E. Data flow Diagram of Dense layer 2

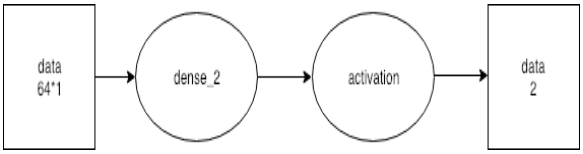

Figure 6 Data flow Diagram of Dense layer 2

F. Graph Created by Tensorflow

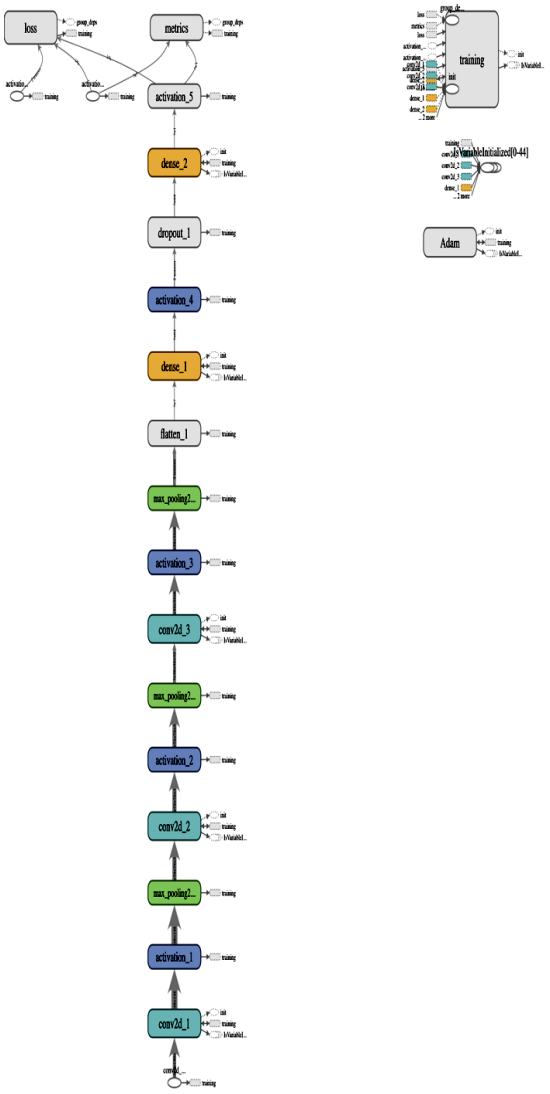

Figure 7 Graph Created by Tensorflow

IV. EXPERIMENT RESULT AND ANALYSIS

A. Learning Rate and Algorithm of Solving the Optimal Parameters

The weights and bias of the convolutional neural network need to be solved by the gradient descent algorithm. In the repeated process, the network is trained for 50 epochs for the learning rate, and weight needs to be adjusted, and the adjustment strategy has different choices. In the following, we give information about the training and validation data of our own Dataset. The trained data are images dissected from multiple videos.

To handle the overfitting and underfitting problem in training, we have used data augmentation and normalization of the image in order to fit in the neural network properly; here, every generated image is 150x150x3 in dimension as input to the neural network. When the neural network is in the predicting phase, it again normalizes the image changes its width and height to 150x150.

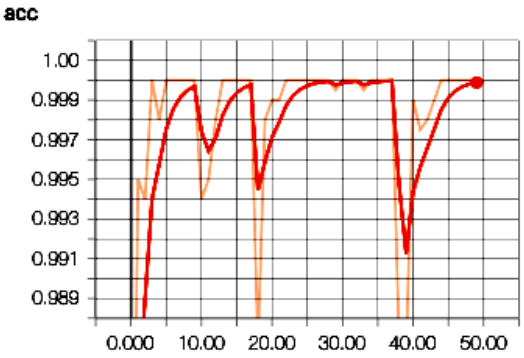

Figure 8 Training Accuracy

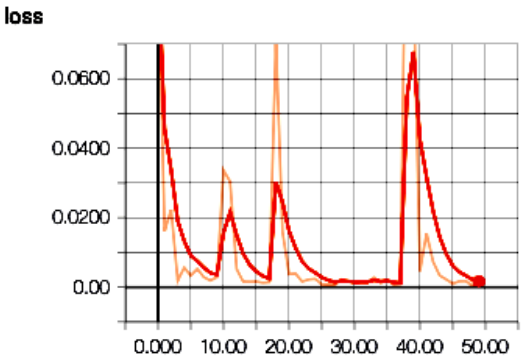

Figure 9 Training Loss

As we can see from the figure above, In the time of training the network with the increase of the number of iteration, the recognition rate also increases.

B. Classification and evalution

We have specifically chosen images from movies and boxing. Furthermore, we have tested with some of the scenes from the movie the result was astonishing; it correctly predicts violent and non-violent image from the given inputs. A Sample violent image is given below.

Figure 10 Sample image of our Dataset

Here is an example of a violent image that is given as input. The network now processes the image by forcing it to all the layers of our neural network. Furthermore, finally, it predicts the output. In Some cases, the network may not be able to identify the image as violent that is because either the image may be too small or it can not find the feature is required to classify the image. So far, the neural network can able to predict up to 95% accuracy from the given input.

C. The result of Our Dataset

Our Dataset consists of 2400 color and grayscale images, which is split as 2000 image as train fit and 400 as test fit. The Classification result is as follows:

TABLE I. RESULT OF OUR OWN DATASET

Method	***Accuracy Rate***
Our CNN for Violence Detection	95%
Fast violence Detection in video[21]	90%
Violent activity detection with transfer learning method[22]	94.40%
Detection of Violent Events in Video Sequences based on census Transform Histogram[23]	92.79%

Compared with the existing methods, our method produces more accurate. We also verify that the experiment is performed on CPU only.

V. SUMMARY

In this paper, we proposed a simple Convolutional neural network on image classification for violence detection. This simple convolutional neural network imposes less computational cost and more accuracy for specific violence classification in scenes. On the basis of the convolutional neural network, we also analyzed different methods of violence detection method, but every method is either a machine learning method or naïve method, but the deep learning approach is the optimal method for video and image classification.

REFERENCES

[1] LecunY, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[2] Cun Y L, Boser B, Denker J S, et al. Handwritten digit recognition with a back-propagation network[C] Advances in Neural Information Processing Systems. Morgan Kaufmann Publishers Inc. 1990:465.

[3] Hecht-Nielsen R. Theory of the backpropagation neural network[M] Neural networks for perception (Vol. 2). Harcourt Brace & Co. 1992:593-605 vol.1.

[4] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2):2012.

[5] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in ECCV, 2014.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in ICLR, 2015.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with Convolutionals," Co RR, vol. abs/1409.4842, 2014.

[8] Hinton G E. Deep belief networks[J]. Scholarpedia, 2009, 4(5): 5947.

[9] Scott rozelle, et.al. The depth of the study review [J]. Computer application research, 2012, 29 (8) : 2806-2810.

[10] lili guo, shifei ding. Deep learning research progress [J]. 2015.

[11] FanYaQin Wang Binghao, et.al. Depth study domestic research review [J]. China distance education, (6) : 2015-27 to 33.

[12] Liu Jinfeng. A concise and effective to accelerate the Convolutional of the neural network method [J]. Science, technology and engineering, 2014 (33) : 240-244.

[13] Markoff J. How many computers to identify a cat? 16,000[J]. New York Times, 2012.

[14] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013, 35(8): 1798-1828.

[15] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in ICML, 2010, pp. 807–814.

[16] T. Wang, D. Wu, A. Coates, and A. Ng, "End-to-end text recognition with Convolutionalal neural networks," in International Conference on Pattern Recognition (ICPR), 2012, pp. 3304–3308.

[17] Y. Boureau, J. Ponce, and Y. Le Cun, "A theoretical analysis of feature pooling in visual recognition," in ICML, 2010, pp. 111–118.

[18] Y.Tang,"Deeplearningusinglinearsupportvectormachines,"ar Xiv prprint arXiv:1306.0239, 2013.

[19] S.WangandC.Manning,"Fastdropouttraining,"inICML,2011

[20] Wan, L., et al. Regularization of neural networks using dropconnect. in Proceedings of the 30th International Conference on Machine Learning (ICML-13). 2013.

[21] Oscar Deniz, Ismael Serrano, Gloria Bueno and Tae-Kyun Kim. Fast Violence Detection in Video, 2017 Electronics letter 20th july.

[22] A.S.Keceli and A. Kaya: Violent activity detection with transfer learning method. ar Xiv preprint ar Xiv:1505.03229, 2015.

[23] Felipe de Souza and Helio Pedrini ,Detection of Violent Events in Video Sequences based on Census Transform Histogram ,2017 30th SIBGRAPI Conference on Graphics, Patterns and Images.

[24] R. Vasudevan, "Neural Networks and Web Mining" SSRG International Journal of Electronics and Communication Engineering 1.1 (2014): 9-14.