ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Anomaly Detection using 3D Convolution Neural Network

Niket Tangadpalliwar, Shubham Thapa, Vinit Rege, Kuldeep Singh, Prof. Priya U. Thakare

Department of Computer Science, Sinhgad Institute of Technology and Science, Pune, India

ABSTRACT: The main aim of the project is to provide a platform/system where anomalies can be detect and other events using image analysis system with the help of digital image processing technique and understand the circumstances and if required contact or alert the administrator so that the administrator can relay information to the sub-units present in the area so that some action can be taken. During the last decade, there has been significant progress in both the theoretical aspects and the applications of neural networks on the image analysis, and processing, has been made. The constant development of computers has led to several major enhancements in terms of data processing. The computational power has gone exponentially up, doing millions of calculations per second. Along with the elaboration of computational force, data.Main working modules include feature extraction module which is able to use and then compares different learning techniques, such as Artificial Neural Networks, with old-fashioned engineering tools like SIFT for feature extraction. Then classification module containing models like Support Vector Machines, k-NN, decision trees for classification. The system is built to utilise most of the computational power. Moreover, it is based on database centric approach in order to decrease computational cost and join different modules.This project aims to present a image/video analysis system using digital image processing techniques with the help of neural networks which will be a part of machine learning and will be used to analyses and distinguish images. We will also be using a big data platform to store and process large amount of data that will be generated during the implementation and execution of the project. An IOT platform will also be used to relay the data generated by the implementation and execution of image analysis system using digital image processing techniques with the help of neural networks.

KEYWORDS: Feature Extraction, Image Classification, SIFT, PCA-SIFT, F-SIFT, SURF, FAST, Convolution Neural Network (CNN), CNN training.

I. INTRODUCTION

In the past few decades artificial intelligence (AI) and machine learning has advanced rapidly. As the computers have grown more powerful day by day, the computation power has sped up, doing millions of operations per seconds, however to accompany this data has also grown at rapid rate too .This advancement has opened large potential for areas such as Visual-Image processing, and the identification and classification of data. This has huge usage and benefits with the currentlygrowing data rates and simplifying and solving of bigdata problems. So with image learning getting increasing popular this has led to a large number of on-going researchImage processing is a technique in which images are processed using various mathematical operations by using any form of signal processing for which the input might be an image or a series of images or a video such as a photograph. It might be video frame and the output of image processing may be either an image or a set of characteristics related to that image. Image analysis isa technique used for extraction of meaningful information from number of images mainly from the digital images by using any digital image processing techniques. In machine learning, artificial neural networks are a group of models inspired by biological neural networks (the central nervous systems of animals/humans, in particular the brain) which are used to estimate or approximate the functions that can depend on a large number of inputs and are generally unknown. The use of artificial neural network models is that they can be used to infer a function from a number of observations. Internet of things is nothing butthe network of physical devices or vehicles or buildings and other items embedded with electronics,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

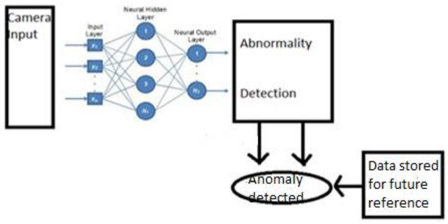
(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

software, sensors, actuators, and network connectivity that enable these objects to collect and exchange small or large amount of data. IOT is one of the emerging and upcoming fields in R&D and has vast application.In this project, we will be using mage/video analysis system using digital image processing techniques with the help of neural networks which will store its data in a big data platform and relay information to various sub units using IOT platform.

II. BASIC ARCHITECTURE



Various cases & accidents in India remain undetected even existence of cctv surveillances because lack of accuracy in manual detection.Various different factors affect the accuracy in manual detection like rolling eyes through various screens may cause less concentration. There is need of a proper platform which will bring the focus directly to the abnormal activity so as to make it possible to act rationally. Use of neural networks to make use of machine learning so that surveillance cameras can recognize anomaly and bring the administrator to notice. After confirmation from administrator relay information to the various subunits in the vicinity to make sure that proper action is taken with IOT.As the Pune city is soon going to become a smart city, this would bring lot of efficiency and less reaction time to various such accidents and abnormalities. As it is machine learning administrator don't have to update the system frequently which makes the system autonomous and less prone to human defined errors.

III. 3D CONVOLUTIONAL NEURAL NETWORKS

2D convolution process is performed at the convolutional layers in which features can be extracted from local neighbourhood on feature maps in the previous layer. Then an additive bias is applied after which the result is passed through a sigmoid function. Formally, the value of unit at position (x, y) in the jth feature map in the ith layer, denoted as v_{ij}^{xy} , is given by

$$v_{ij}^{xy} = \tanh\left(b_{ij} + \sum_{m} \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{ijm}^{pq} v_{(i-1)m}^{(x+p)(y+q)}\right)$$

here tanh(·) is the hyperbolic tangent function, b_{ij} is the bias for this feature map, m indexes over the set of feature maps in the (i − 1)th layer connectedto the current feature map, w_{ijk}^{pq} is the value at the position (p, q) of the kernel connected to the kthfea-ture map, and P_i and Q_i are the height and width of the kernel, respectively.The resolution of the feature maps in the subsampling layers is reduced by pooling over local neighborhood on the feature maps in the previous layer, which in turn increases invariance to distortions on the inputs.CNN architecture can be constructed with the help of stacking multiple layers of convolution and subsampling in

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

an alternating fashion. The parameters of CNN, such as the bias b_{ij} and the kernel weight w_{ijk}^{pq} , are usually trained using either super-vised or unsupervised approaches

IV. 3D CONVOLUTION

Convolutions can be applied on the 2D feature maps which help to compute features from the spatial dimensions only. When this is applied to video analysis problems, it is desirable to capture the motion information en-coded in multiple contiguous frames. We propose to perform 3D convolutions in the convolution stages of CNNs which would compute features from both spatial and temporal dimensions. The 3D convolution can be achieved by convolving a 3D kernel to the cube formed by stacking multiple contiguous frames together. By this construction, the feature maps in the convolution layer is connected to multiple contiguous frames in the

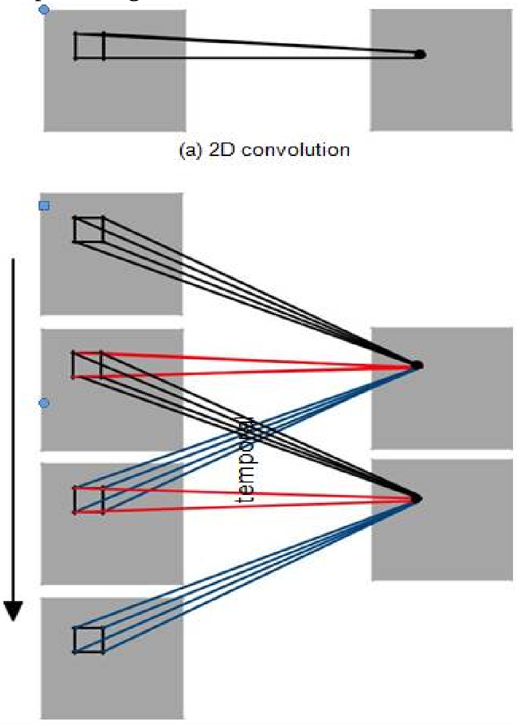


Figure 1.**Comparison of 2D (a) and 3D (b) convolutions.**

In (b) the size of the convolution kernel in the temporal dimension is 3, and the sets of connections are color-coded so that the shared weights are in the same color. In 3D convolution there is the same 3D kernel which is applied to overlapping 3D cubes in the input video to extract motion features.

Formally, the value at position (x, y, z) on the jthfea-ture map in the ith layer is given by

$$v_{ij}^{xyz} = \tanh\left(b_{ij} + \sum_{m}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right)$$

WhereR_i is the size of the 3D kernel including the temporal dimension, w_{ijm}^{pqr} is nothing but the (p, q, r)th value of the kernel connected to the mth feature map in the previ-ous layer. A comparison of 2D and 3D convolutions is given in Figure 1.In a 3D convolutional kernel only one type of features can be extracted from the frame cube, since the kernel weights has to be replicated across the entire cube. A general design principle of CNNs is that the number of feature maps should be increased in late layers by generating multiple types of features from the samecan be applied to con-

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

tiguous frames to extract multiple features. As in Figure 1, the sets of connections are color-coded so that the shared weights are in the same color. Note that all the 6 sets of connections do not share weights, resulting in two dfferent feature maps on the right.set of lower-level feature maps. Similar to the case of 2D convolution, this can be achieved by applying multiple 3D convolutions with distinct kernels to the same location in the previous layer (Figure 2).

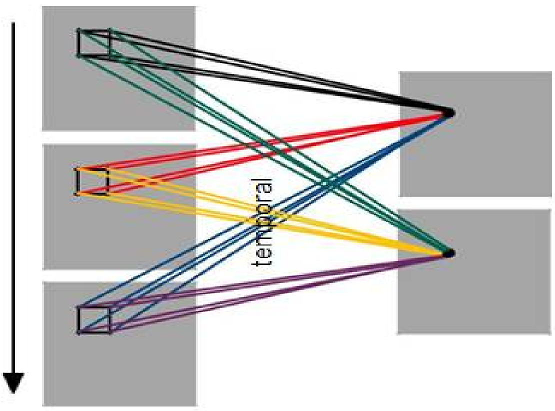


Figure 2. Extraction of multiple features from contiguous frames. Multiple 3D convolutions

V. 3D CNN ARCHITECTURE

With the help of 3D convolution described above, a large number ofCNN architectures can be devised. In this, we can describe a 3D CNN architecture that we have developed for detecting anomalies in traffic using various data set. We consider there are 7 frames of size 60×60 centered on the current frame as inputs to the 3D CNN model. We first apply a set of hardwired kernels to generate multiple channels of information from the input frames. The gray channel contains the gray pixel values of the 5 input frames. The feature maps in the gradient-x and gradient-y channels are obtained by computing gradients along the horizontal and ver-tical directions, respectively, on each of the 5 input frames, and the optflow-x and optflow-y channels con-tain the optical flow fields, along the horizontal and vertical directions, respectively, computed from adja-cent input frames. This hardwired layer is used to en-code our prior knowledge on features, and this scheme usually leads to better performance as compared to random initialization.

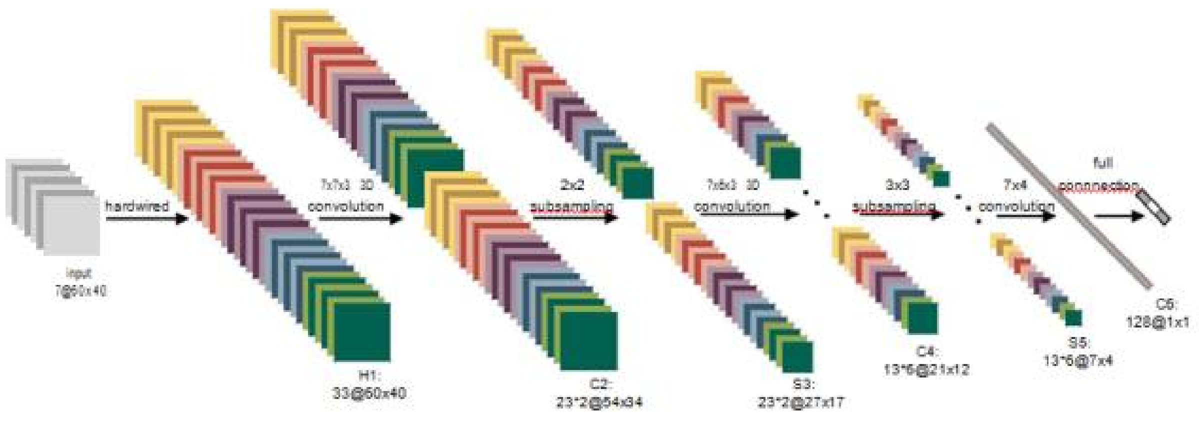


Figure 3. 3d CNN Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

We then apply 3D convolutions with a kernel size of 5 × 5× 5 (5 × 5 in the spatial dimension and 5 in the temporal dimension) on each of the 5 channels separately. To increase the number of feature maps, two sets of different convolutions are applied at each location, resulting in 2 sets of feature maps in the C2 layer each consisting of 23 feature maps. This layer contains 1,480 trainable parameters. In the subsequent subsampling layer S3, we apply 3 × 3 subsampling on each of the feature maps in the C2 layer, which leads to the same number of feature maps with reduced spatial resolution. The next convolution layer C4 is obtained by applying 3D convolution with a kernel size of 5*5*5 on each of the 1channels in the two sets of feature maps separately. To increase the number of feature maps, we apply 2 convolutions with different kernels at each location, leading to 6 distinct sets of feature maps in the C4 layer each containing feature maps. This layer contains trainable parameters. The next layer S5 is obtained by applying 3×3 subsampling on each feature maps in the C4 layer, which leads to the same number of feature maps with reduced spatial resolution. At this stage, the size of the temporal dimension is already relatively small (3 for gray), so we perform convolution only in the spatial dimension at this layer.By the multiple layers of convolution and subsampling. The 7 input frames have been converted into a 128D feature vector capturing the motion information in the input frames. The output layer consists of the same number of units as the number of actions, and each unit is fully connected to each of the 128 units in the C6 layer. In this design we essentially apply a linear classifier on the 128D feature vector for action classification. For an action recognition problem with 2 classes. The last dense layer consist of 2 neurons which is used to identify the class of input video.

VI. CONCLUSION

From the consideration of all above points we conclude that the main aim of the project is to provide a platform/system where anomalies can be detect and other events using image analysis system with the help of digital image processing technique and understand the circumstances and if required contact or alert the administrator so that the administrator can relay information to the sub-units present in the area so that some action can be taken. During the last decade, there has been significant progress in both the theoretical aspects and the applications of neural networks on the image analysis, and processing, has been made. The constant development of computers has led to several major enhancements in terms of data processing. The computational power has gone exponentially up, doing millions of calculations per second. Along with the elaboration of computational force, data.Main working modules include feature extraction module which is able to use and then compares different learning techniques, such as Artificial Neural Networks, with old-fashioned engineering tools like SIFT for feature extraction. Then classification module containing models like Support Vector Machines, k-NN, decision trees for classification. The system is built to utilise most of the computational power. Moreover, it is based on database centric approach in order to decrease computational cost and join different modules. This project aims to present a image/video analysis system using digital image processing techniques with the help of neural networks which will be a part of machine learning and will be used to analyses and distinguish images. We will also be using a big data platform to store and process large amount of data that will be generated during the implementation and execution of the project. An IOT platform will also be used to relay the data generated by the implementation and execution of image analysis system using digital image processing techniques with the help of neural networks

REFERENCES

1) "MLP Neural Network Based Face Recognition System Using Constructive Training algorithm ", HayetBoughrara , Mohamed Chtourou , Chokri Ben Amar , 978-1-4673-1520-3/12/$31.00 ©2012 IEEE
2)"Internet of Things for Smart Cities ", Andrea Zanella, Senior Member, IEEE, Nicola Bui, Angelo Castellani, Lorenzo Vangelista, Senior Member, IEEE, and Michele Zorzi, Fellow, 2327-4662 © 2014IEEE
3)"Real-time camera anomaly detection for real-world video surveillance",yuan-kaiwang, ching-tang fan, ke-yucheng, peter shaohuadeng,978-1-4577-0308-9/11/$26.00 ©2011 IEEE
4)"A parallel computing platform for training large scale neural networks", RongGu, FuraoShen, YihuaHuang,National Key Laboratory for Novel Software Technology Nanjing University, Nanjing, China 210093, 978-1-4799-1293-3/13/$31.00 ©2013 IEEE
5)C. E. A. Mulligan and M. Olsson, "Architectural implications of smart city business models: An evolutionary perspective," IEEE Commun. Mag.,vol. 51, no. 6, pp. 80–85, Jun. 2013.