|  |
| --- |
| **CHAPTER -1**  **INTRODUCTION**  Surveillance security is a laborious and time-consuming activity. In this tutorial, we will create a system to automate the work of evaluating video surveillance. We will evaluate the camera footage in real time to identify any unusual activity such as violence or theft.  There is a lot of study going on in the market concerning video surveillance, and the function of CCTV videos has risen. CCTV cameras are strategically positioned across the area to provide monitoring and security.  Deep learning algorithms for deep surveillance have advanced during the previous decade. These developments have demonstrated an important trend in deep surveillance and promise a significant increase in efficiency. Deep surveillance is commonly used for theft detection, violence detection, and detecting the possibility of an explosion.  **CHAPTER - 2**  **PROBLEM DEFINITION**  Smart Video Surveillance  **2.1 Project Objective**   * To train an autoencoder for abnormal event detection. * To train the autoencoder on normal videos. * To identify the abnormal events based on the euclidean distance of the custom video feed and the * frames predicted by the autoencoder. * To automate the task of analyzing video surveillance.   **2.2 Proposed Methodology**  IMG_256  Figure 2.1 Block Diagram Of Smart Video Surveillance  Deep neural networks have usually been used for computer vision, picture classification, and object recognition applications. In this research, we must expand deep neural networks to three dimensions in order to learn spatiotemporal aspects of the video feed.  We will propose a spatiotemporal autoencoder based on a 3D convolution network for this video surveillance project. The encoder collects spatial and temporal information, which is subsequently used by the decoder to reassemble the frames. By estimating the reconstruction loss using the Euclidean distance between the original and reconstructed batches, aberrant occurrences are found.  **3D Convolution Encoder**  Convolutional Autoencoders are a kind of Convolutional Neural Network used for unsupervised learning of convolution filters. They are commonly used in image reconstruction to decrease reconstruction mistakes by learning the best filters.  **3D Deconvolution Decoder**  For many applications and network topologies, we frequently wish to execute changes that are the inverse of a typical convolution, i.e. we want to do up-sampling. Creating high-resolution pictures and translating low dimensional feature maps to high-dimensional space, as in auto-encoder or semantic segmentation, are two examples.  Up-sampling was traditionally accomplished through the use of interpolation techniques or the creation of rules by hand. Modern designs, such as neural networks, on the other hand, allow the network to learn the appropriate transformation on its own, without the need for human involvement. To do this, we may employ the transposed convolution.   * We will use spatial temporal encoders to identify abnormal activities.   **Spatio-temporal Autoencoder**  Modeling aberrant spatiotemporal occurrences is difficult since data from abnormal activities are few throughout the course of a surveillance stream. We tackle this problem by employing a normalcy modeling technique, in which abnormalities are recognized as departures from normal patterns. To that aim, we present a residual spatiotemporal autoencoder that can be trained end-to-end to detect anomalies in surveillance footage. Irregularities are discovered utilizing the reconstruction loss, in which normal frames are recreated well with a cheap reconstruction cost, while aberrant frames are identified as such  **CHAPTER – 3**  **THE DATASETS FOR ABNORMAL EVENT DETECTION**  **CUHK Avenue Dataset**-  This dataset contains 16 training videos and 21 testing videos with a total of 47 abnormal events, including throwing objects, loitering and running. The size of people may change because of the camera position and angle. The video contains 30652 frames in total.  **UCSD pedestrian Dataset-**  This dataset contains videos with pedestrians. It includes groups of people walking towards, away, and parallel to the camera.  The abnormal event includes:   * Non-pedestrian entities * Anomalous pedestrian motion patterns   **CHAPTER – 4**  **ALGORITHMS AND LIBRARIES USED**  **4.1 ALGORITHMS USED**  **4.1.1 CNN (Convolution Neural Network)**  A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.  A Convolutional neural network (CNN) is a neural network that has one or more convolutional layers and are used mainly for image processing, classification, segmentation and also for other auto correlated data. A convolution is essentially sliding a filter over the input.  CNN is a type of neural network model which allows us to extract higher representations for the image content. Unlike the classical image recognition where you define the image features yourself, CNN takes the image's raw pixel data, trains the model, then extracts the features automatically for better classification.    **4.1.2 DNN (Deep Neural Network)**  A deep neural network (DNN) is an [artificial neural network](https://en.wikipedia.org/wiki/Artificial_neural_network) (ANN) with multiple layers between the input and output layers. There are different types of neural networks but they always consist of the same components: neurons, synapses, weights, biases, and functions  It is a collection of neurons organized in a sequence of multiple layers, where neurons receive as input the neuron activations from the previous layer, and perform a simple computation (e.g. a weighted sum of the input followed by a nonlinear activation).  Deep neural network with some level of complexity, usually at least two layers, qualifies as a deep neural network (DNN), or deep net for short. Deep nets process data in complex ways by employing sophisticated math modeling. ... A model is a single model that makes predictions about something.  **4.2 LIBRARIES UESD**   * **Keras**   It is an open-source software library that provides a Python interface for artificial neural networks.   * **Numpy**   It is a general-purpose array-processing package which provides a high-performance multidimensional array object.   * **CV2**   It is a Python library to solve computer vision problems.   * **Imutils**   A series of convenience functions to make basic image processing functions such as translation, rotation, resizing, skeletonization, displaying etc.   * **PIL**   Python Imaging Library is a free and open-source additional library for the Python programming language that adds support for opening, manipulating, and saving many different image file formats.  **CHAPTER – 5**  **WORK FLOW**  **Step-1 Imports**   * import numpy as np * import glob * import os * import cv2 * from keras.layers import Conv3D,ConvLSTM2D,Conv3DTranspose * from keras.models import Sequential * from keras.callbacks import ModelCheckpoint, EarlyStopping * import imutils * from keras.models import load\_model * import argparse * from PIL import Image   **Step-2 Initialize Directory Path**  store\_image=[]  train\_path='./train'  fps=5  train\_videos=os.listdir('train\_path')  train\_images\_path=train\_path+'/frames'  os.makedirs(train\_images\_path)  This code will initialize a directory named train where we will store our video Frames.  **Step-3 Store Video Frames**  Process frames using CV2  Define Function  ‘store\_inarray’  Define Function  ‘store\_inarray’      Add this processed Frame to store\_image array    **Step-4 Extract Frame from video and Call store function**  After execution all frames present in training will be processed  Pass these frames in store\_inarray  funtion  Extract frames from Training Dataset  **Step-5 Store the store\_image list in a numpy file**    1- Create a file numpy file named training.  2- Extension of the file will be .npy.  3- Store data of store\_image array to training.npy file.    training.py  **Step-6 Create spatial autoencoder architecture**  **Spatio-temporal Autoencoder**- This is an encoder which utilizes deep neural networks to learn video representation automatically and extracts features from both spatial and temporal dimensions by performing 3-dimensional convolutions.  **Spatial dimension-** A measure of spatial extent, especially width, height, or length.  **Temporal dimension-** A temporal dimension, or time dimension, is a dimension of time. A temporal dimension is one way to measure physical change.  **Step-8 Train the autoencoder**   1. It’s time to train our autoencoder on training data. 2. After training save the model with name saved\_model.h5.   **Step-7 Our Model is Ready to Use**   1. Import the model named saved\_model.h5. 2. Apply this model on any sample video. 3. It will alert a message whenever any Abnormal event will happen.   **CHAPTER – 6**  **SOFTWARE REQUIREMENTS**  Video recognition requires intensive processing by your computer. The minimum requirements for a single IP camera at QVGA are:   * Windows 7, 8, 10, and Server 2012 and newer releases. * 2GHz Pentium 4 processor or higher * 2GB RAM or higher * Minimum screen resolution of 1024 x 768 * At least 50GB of free disk space per camera is required * An internet connection for software download, updates, and activation   **CHAPTER – 7**  **MODULE DESCRIPTION**  Real-time crowd anomaly detection using spatiotemporal texture modeling. Spatiotemporal texture is made up of spatiotemporal slices and spatiotemporal volumes. Wavelet transformations are used to extract information from these slices. To discriminate between normal and pathological behaviour, texture patterns are subjected to a Gaussian approximation model.  Deep convolutional framework for abnormal behavior detection in a smart surveillance system includes three sections.   * Human subject detection and discrimination * A posture classification module * An abnormal behavior detection module   A research of deep convolutional auto encoders for video anomaly detection [12] suggests a structure that is a hybrid of auto encoders and CNN. An auto encoder consists of an encoder and a decoder. Convolutional and pooling layers are used in the encoder, and deconvolutional and unpooling layers are used in the decoder. The design enables the use of low-level frames in conjunction with high-level aesthetic and motion characteristics. Reconstruction mistakes are used to express anomaly scores.  Going deeper with convolutions implies that it is possible to outperform standard neural networks. By incorporating sparsity into the architecture, fully linked layers are substituted by sparse ones. The study proposes dimensionality reduction to assist minimize the growing demand for computing resources. Computing reductions begins with 1x1 convolutions and progresses to 5x5 convolutions. The technique makes no mention of the execution time. In addition, no conclusion can be drawn regarding the crowd size that the technique can handle satisfactorily.  Deep learning for visual understanding: a review, a study of the core deep learning models. CNN, RBM, Autoencoder, and Sparse coding were the models and techniques discussed. The study also discusses the disadvantages of deep learning models, such as people's inability to grasp the underlying theory.  The approaches discussed in the preceding sections are effective for automated feature creation. All approaches are effective in dealing with individual and group entities of limited size.  The majority of issues in the actual world originate from crowds. The strategies listed above are ineffective in dealing with crowd scenarios.  **CHAPTER – 8**  **APPLICATIONS AND LIMITATIONS**  **8.1 Applications**  The contexts identified are listed as application areas. Major part in existing work provides solutions specifically based on the context.  1. Traffic signals and main junctions  2. Residential areas  3. Crowd pulling meetings  4. Festivals as part of religious institutions  5. Inside office buildings  Among the listed contexts crowd analysis is the most difficult part. All type of actions, behaviour and movement are needed to be identified.  **8.2 Limitations**   * Time complexity * Bad weather conditions * Real world dynamics * Occulsions * Overlapping of projects   Existing techniques dealt with the issues separately. There is no technique that addresses all of the objectives as features in a single proposal.  To handle successful intelligent crowd video analysis in real time, the technique should be capable of addressing all of these issues. Traditional approaches are incapable of producing effective economic solutions in a timely manner.  The availability of high-performance computer resources, such as GPUs, enables the development of deep learning-based solutions for rapid data processing. Existing deep learning architectures or models can be merged by including desirable characteristics and eliminating undesirable ones.  **REFERENCES**  [1] Bouachir W, Gouiaa R, Li B, Noumeir R. Intelligent video surveillance for real-time detection of suicide attempts. Pattern Recogn Lett. 2018;110:1–7 (ISSN 0167-8655).  [2] Khan MUK, Park H, Kyung C. Rejecting motion outliers for efficient crowd anomaly detection. IEEE Trans Inf Forensics Secur. 2019;14(2):541–56  [3] [Jesús Benito-Picazo](https://www.researchgate.net/profile/Jesus-Benito-Picazo), [Enrique Domínguez](https://www.researchgate.net/profile/Enrique-Dominguez-2), [Esteban J. Palomo](https://www.researchgate.net/profile/Esteban-Palomo), [Ezequiel López-Rubio](https://www.researchgate.net/profile/Ezequiel-Lopez-Rubio). Deep learning-based video surveillance system managed by low cost hardware and panoramic cameras.  [4] [Muhammad Ishtiaq](https://www.researchgate.net/profile/Muhammad-Ishtiaq-19), [Sultan Almotiri](https://www.researchgate.net/profile/Sultan-Almotiri), [Rashid Amin](https://www.researchgate.net/profile/Rashid-Amin-4). Deep Learning based Intelligent Surveillance System. [International Journal of Advanced Computer Science and Applications](https://www.researchgate.net/journal/International-Journal-of-Advanced-Computer-Science-and-Applications-2156-5570)  [5] Author: Lizzi Goldmeier**.** The top security surveillance challenges and how to solve them  [6] [Antonio Carlos Cob-Parro](https://www.ncbi.nlm.nih.gov/pubmed/?term=Cob-Parro%20AC%5BAuthor%5D&cauthor=true&cauthor_uid=33922548), [Cristina Losada-Gutiérrez](https://www.ncbi.nlm.nih.gov/pubmed/?term=Losada-Guti%26%23x000e9%3Brrez%20C%5BAuthor%5D&cauthor=true&cauthor_uid=33922548), [Marta Marrón-Romera](https://www.ncbi.nlm.nih.gov/pubmed/?term=Marr%26%23x000f3%3Bn-Romera%20M%5BAuthor%5D&cauthor=true&cauthor_uid=33922548), [Alfredo Gardel-Vicente](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gardel-Vicente%20A%5BAuthor%5D&cauthor=true&cauthor_uid=33922548), and [Ignacio Bravo-Muñoz](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bravo-Mu%26%23x000f1%3Boz%20I%5BAuthor%5D&cauthor=true&cauthor_uid=33922548). Smart Video Surveillance System Based on Edge Computing.  [7] [Arun Hampapur](https://www.researchgate.net/profile/Arun-Hampapur), [Leslie Brown](https://www.researchgate.net/scientific-contributions/Leslie-Brown-2080507033), [Jonathan Connell](https://www.researchgate.net/profile/Jonathan-Connell),[Ahmet Ekin](https://www.researchgate.net/profile/Ahmet-Ekin). Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. [IEEE Signal Processing Magazine](https://www.researchgate.net/journal/IEEE-Signal-Processing-Magazine-1558-0792).  [8] Arunnehru J, Chamundeeswari G, Prasanna Bharathi S. Human action recognition using 3D convolutional neural networks with 3D motion cuboids in surveillance videos. Procedia Comput Sci. 2018;133:471–7 (ISSN 1877-0509).  [9] Xu M, Fang H, Lv P, Cui L, Zhang S, Zhou B. D-stc: deep learning with spatio-temporal constraints for train drivers detection from videos. Pattern Recogn Lett. 2017. <https://doi.org/10.1016/j.patrec.2017.09.040> (ISSN 0167-8655).  [10] Wang J, Xu Z. Spatio-temporal texture modelling for real-time crowd anomaly detection. Comput Vis Image Underst. 2016;144:177–87 (ISSN 1077-3142). |