Experiment No. 1
Review of Deep Learning techniques
Date of Performance: 22/08/2023
Date of Submission: 05/09/2023

Paper – 01: Video Anomaly Detection Using Pre-Trained Deep Convolutional Neural Nets and Context Mining

Problem Statement:

The problem statement of the paper is to address the challenge of detecting anomalies in video streams using pre-trained models and context mining. In modern surveillance and security applications, the need to efficiently and accurately identify anomalies within video data has become paramount. Anomalies can range from suspicious activities, such as unauthorized access or vandalism, to critical events like accidents or intrusions. Traditional methods of video anomaly detection often struggle to handle the complexity and variability of real-world scenarios, leading to high false-positive rates and inadequate anomaly recognition. The authors aim to improve the accuracy and performance of anomaly detection by incorporating contextual information and leveraging pre-trained CNN models. They propose a system that combines features derived from these models to analyze video streams and identify abnormal behaviors. The goal is to create an effective and interpretable method for detecting anomalies in surveillance videos.

Solution:

The system improves anomaly detection performance by incorporating contextual information and leveraging pre-trained deep learning models. It combines features derived from these models to analyze video streams and identify abnormal behaviors. By integrating multiple pre-trained models and deriving contextual properties from the high-level features, the system enhances the accuracy and efficiency of anomaly detection. The approach also addresses the issue of model size reduction, making it suitable for resource-constrained devices. Additionally, the system aims to create interpretable models by avoiding the use of black-box models, allowing for better understanding and interpretation of the model's decisions.

Technologies:

The technologies used in this document include distributed cameras, related drivers, embedded computer vision tasks, pre-trained deep convolutional neural networks (CNNs), denoising auto encoder, and context mining. The hardware layer consists of distributed cameras and drivers that transfer raw video streams into the system's software. The middle layer includes the embedded computer vision tasks that process the video streams and generate structural data output. The system leverages pre-trained CNN models to extract high-level features from the video streams. Context mining techniques are used to derive contextual properties from the high-level features. Additionally, the system incorporates a denoising auto encoder for efficient and accurate anomaly detection

Dataset:

They showed the anomaly detection result on the UCSD Ped1 and Ped 2 datasets. The Ped1 dataset has 34 training videos and 36 testing videos. Each video consists of 200 frames with 238 × 158 pixels at 30 FPS. The Ped2 dataset has 16 training videos and 12 testing videos. The video frame numbers of the Ped2 dataset are ranging from 120 to 180 frames with 360 × 240 pixels. The training video only includes pedestrians. Both Ped1 and Ped2 provide complete frame-level abnormal labels and partial pixel-level abnormal labels. In this experiment, we only consider the frame-level samples since our work mainly considers the contextual features. The abnormal event includes unexpected entities (bicycle, skateboard, motorcycle, etc.), irregular trajectory (deviate from the major moving direction), and entering the prohibitive region (walking on the grass).

Conclusion:

In this work, they have presented a novel design of a video anomaly surveillance system that is based on the high-level features from the pre-trained models and using denoising auto encoder to detect anomalous video events. Two UCSD pedestrian datasets are used to evaluate our approach and to compare it with several state-of-the-art methods. Our experimental results show that contextual features improve model performance. Moreover, our proposed model achieves comparable results while significantly reducing the model complexity and computational overhead of our model. Furthermore, the results produced by our method are easily interpretable. Our approach is not developed to replace state-of-the-art approaches; instead, it offers a better understanding of how pre-trained CNNs can be used for video anomaly detection and provides an alternative approach, especially when training data is not available for large models.

Paper 2- Deep Learning-Based Fault Localization in Video Networks Using Only Client-Side QoE.

Problem Statement:

The problem addressed in the paper is the detection and isolation of network faults in video streaming services. The authors aim to develop a method that can accurately detect and localize faults using only client-side Quality of Experience (QoE) metrics. They highlight the limitations of existing approaches, such as the need for full access to the end-to-end path and reliance on QoE prediction rather than direct measurement. The proposed method aims to overcome these shortcomings and provide an effective solution for fault detection and isolation in video networks.

Solution:

Based on the research paper, the proposed solution is a deep learning-based method for detecting and isolating network faults in video streaming services. The solution utilizes client-side Quality of Experience (QoE) metrics to accurately detect and localize faults. The authors collected a dataset from an actual video streaming testbed and trained two artificial neural networks, namely MLP and LSTM, to perform the fault detection and isolation tasks. The results of the experiments show a higher fault isolation accuracy for the proposed methods compared to existing approaches.

Technologies:

The technologies mentioned in the paper include:

1. Multi-media systems: The paper focuses on network operations and management in the context of multimedia systems, specifically video streaming services.
2. Deep learning: The proposed solution utilizes deep learning techniques, specifically artificial neural networks (MLP and LSTM), to detect and isolate network faults in video streaming services.
3. Quality of Experience (QoE) metrics: The solution relies on client-side QoE metrics, such as video quality, buffering time, and playback interruptions, to accurately detect and localize network faults.
4. Artificial Neural Networks (ANNs): The authors train two types of ANNs, MLP and LSTM, to perform fault detection and isolation tasks based on the client-side QoE metrics.
5. Video streaming testbed: The authors collected a dataset from an actual video streaming testbed, which serves as the basis for training and evaluating the proposed solution.
6. Fault detection: The solution aims to detect network faults in real-time by analyzing the client-side QoE metrics and identifying patterns indicative of network issues.
7. Fault isolation: Once a fault is detected, the solution employs the LSTM network to isolate the fault by determining its specific location within the network.

Overall, the paper combines the use of deep learning techniques, client-side QoE metrics, and artificial neural networks to develop an effective method for detecting and isolating network faults in video streaming services.

Dataset:

The dataset mentioned in the research paper is collected from an actual video streaming testbed. It is used to train and evaluate the proposed solution for network fault detection and isolation in video streaming services. The dataset consists of client-side Quality of Experience (QoE) metrics, which are measurements that reflect the end-users' experience while streaming videos. These metrics include video quality, buffering time, and playback interruptions, among others. The QoE metrics are collected from different clients, capturing a range of scenarios and network conditions.

The dataset is divided into three subsets: training, validation, and testing. The paper states that 81% of the dataset is used for training, 9% for validation, and 10% for testing. Specifically, the test set includes 236 videos, which corresponds to 1652 samples. These videos are not used in the training and validation phases, ensuring an unbiased evaluation of the proposed solution's performance.
The dimensions of each video's samples in the dataset are similar to what is presented in Figure 3 of the paper. This means that each video sample consists of seven time steps and six features at each time step. The input training/validation data for the LSTM model has three dimensions: 2119 video samples, seven time steps each, and six features for each time step. For the MLP model, the time steps and features are vectorized.

The dataset plays a crucial role in training the artificial neural networks (MLP and LSTM) used in the proposed solution. By leveraging this dataset, the networks learn the patterns and relationships between the client-side QoE metrics and the occurrence and location of network faults.
Overall, the dataset collected from the video streaming testbed provides the necessary real-world data to train and evaluate the proposed solution for network fault detection and isolation in video streaming services.

Conclusion:

In conclusion, the research paper presents a deep learning-based solution for network fault detection and isolation in video streaming services. The proposed method utilizes client-side Quality of Experience (QoE) metrics and artificial neural networks (MLP and LSTM) to accurately detect and localize network faults.The paper highlights the limitations of existing approaches and addresses them by leveraging client-side QoE metrics, which directly reflect the end-users' experience. The authors collected a dataset from an actual video streaming testbed, consisting of various QoE metrics such as video quality, buffering time, and playback interruptions.The MLP network is trained for fault detection, while the LSTM network is trained for fault isolation. These networks learn the patterns and relationships between the QoE metrics and the occurrence and location of network faults. The proposed solution aims to improve fault detection and isolation accuracy by utilizing deep learning techniques to capture complex patterns and dependencies in the data.The results presented in the paper demonstrate the effectiveness of the proposed solution, with higher fault isolation accuracy compared to existing approaches. This indicates that the method can provide more accurate and reliable fault detection and localization in video streaming networks.Overall, the research paper offers a promising approach to address the challenges of network fault detection and isolation in video streaming services. By utilizing client-side QoE metrics and deep learning techniques, the proposed solution shows potential for enhancing the performance and reliability of video streaming networks.

Paper 3- A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos

Problem Statement:

The problem statement for the research paper is to develop a deep learning-based approach for the detection and classification of inappropriate content in YouTube videos. The goal is to create a model that can accurately identify and categorize videos containing inappropriate or unsafe content, providing a solution for video sharing platforms to remove such content or implement parental control measures. The proposed approach utilizes a combination of an EfficientNet-B7 CNN model and a bidirectional long short-term memory (BiLSTM) network, along with an attention mechanism, to achieve high accuracy in video classification.

Solution:

The solution proposed in the research paper is a deep learning-based approach for the detection and classification of inappropriate content in YouTube videos. The researchers developed a model that combines an EfficientNet-B7 convolutional neural network (CNN) model with a bidirectional long short-term memory (BiLSTM) network. This model is trained to accurately identify and categorize videos containing inappropriate or unsafe content. The proposed approach achieved a high accuracy of 95.66% in video classification, outperforming traditional machine learning classifiers. The solution can be used by video sharing platforms to automatically filter and remove inappropriate content or implement parental control measures.

Technologies:

The research paper utilizes deep learning techniques, specifically a combination of an EfficientNet-B7 convolutional neural network (CNN) model and a bidirectional long short-term memory (BiLSTM) network. These deep learning models are commonly implemented using frameworks such as TensorFlow or PyTorch. Additionally, the paper mentions the use of an attention mechanism in the network architecture. Overall, the research paper leverages deep learning methodologies and potentially popular deep learning frameworks to address the problem of detecting and classifying inappropriate content in YouTube videos.

Dataset:

YouTube has a huge collection of videos and metadata of videos (i.e., likes, dislikes, view count, comments, etc.) that. Google released the YouTube-8M benchmark dataset of more than 8 million video IDs with corresponding labels from 4716 classes. Apart from it, there exists other video benchmarks of specific categories like sports (Sports-1M [29], UCF-101 [66]), action recognition (HMDB51 [67], Kinetics [68]), face recognition (YTF [69], YouTube Celebrities [70]), sentiment analysis([71]), and video captioning (MSVD [72], MSR-VTT [73]). However, none of these existing benchmarks aims for the proposed video classification problem. The datasets closely related to our problem are the NPDI cartoon dataset [50],

the Elsagate dataset [61], and the dataset of Singh et al. [62]. Comparatively, the NPDI dataset is the smallest with 900 images only and is not suitable to perform our deep learning based video classification task. The Elsagate dataset is a publicly available dataset of cartoon videos from sensitive and non-sensitive classes. In this dataset, whole video is considered either safe or unsafe where the clean frames of video are also labeled as unsafe. Secondly, it lacks the complex behaviors of sensitivity content. The videos in this dataset are targeted for toddlers.

Conclusion:

In this paper, a novel deep learning-based framework is proposed for child inappropriate video content detection and classification. Transfer learning using EfficientNet-B7 architecture is employed to extract the features of videos. The extracted video features are processed through the BiLSTM network, where the model learns the effective video representations and performs multiclass video classi fication. All evaluation experiments are performed by using a manually annotated cartoon video dataset of 111,156 video clips collected from YouTube. The evaluation results indicated that proposed framework of EfficientNet-BiLSTM (with hidden units = 128) exhibits higher performance (accuracy = 95.66%) than other experimented models includ ing EfficientNet-FC, EfficientNet-SVM, EfficientNet-KNN, EfficientNet-Random Forest, and EfficientNet-BiLSTM with attention mechanism-based models (with hidden units = 64, 128, 256, and 512). Moreover, the performance comparison with existing state-of-the-art models also demonstrated that our BiLSTM-based framework surpassed other existing models and methods by achieving the highest recall score of 92.22%.

Analysis Table:

Aspect \ Paper	Paper 1	Paper 2	Paper 3
Advantages	-Incorporates contextual information for improved anomaly detection. - Leverages pre-trained deep learning models for feature extraction.	-Deep learning-based fault localization method: The document proposes a novel approach using deep learning techniques for fault localization in video networks.	- Integrates an ImageNet pre-trained CNN model (EfficientNet-B7) for effective video descriptor extraction. - Incorporates an attention mechanism to apply attention probability distribution in the network.
Disadvantages	-The document does not provide a detailed evaluation of the proposed approach.	-Limited scope: The document focuses specifically on fault localization in video networks, which may not be applicable to other domains.	- Does not provide detailed information on the dataset used for training and evaluation.
Performance	- Achieves comparable performance to state-of-the-art methods.	-High accuracy: The proposed method achieves high accuracy in detecting and localizing faults.	- Achieves an accuracy of 95.66% in detecting and classifying inappropriate video content.
Complexity	-Relatively low model complexity.	- Training data collection: Collecting a dataset from an actual video streaming testbed may require significant resources and effort.	- The proposed model consists of nine layers with varying output sizes and learnable parameters - Trained with 152 million parameters (number of neurons) that are updated during the backpropagation process.
Dataset	-UCSD Ped1 and Ped 2 datasets.The Ped1 dataset has 34 training videos and 36 testing videos.. The Ped2 dataset has 16 training videos and 12 testing videos.	-YouTube-8M benchmark dataset of more than 8 million video IDs with corresponding labels from 4716 classes.	-The dataset used in the document consists of videos collected from an actual video streaming testbed.The test set includes 236 videos, which are not used in training and validation. -The dataset is used to train and evaluate the performance of the MLP and LSTM models for fault localization in video networks.