Explainable AI with a Blockchain-Enabled Public Safety Solution

BTP GROUP-42

2020BCS0057-PALASH RAWAT
2020BCS0030-DOMMATI MANIKANTA
2020BCS0082-ADITYA SHITALE
2020BCS0124-RAUNAK KUMAR

GUIDED BY,
DR. EBIN DENI RAJ

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
KOTTAYAM

Introduction

- ## Brief introduction
  - This project seeks to ensure the safety and protection of the general population. Specifically, we are developing a model that will detect anomalies in cctv footage using videos or images to identify violations.
  - The model will then send this information via blockchain to the nearby police station. Most studies so far have focused on just one thing, such as developing a knife or gun detection. Detection also requires high-resolution videos or photos to train the model, but that won't always be the case, thus our project will detect anomalies in all areas of operation.
- ## Significance of the project
  - **Public Safety:** Ensuring the safety of the general population is a fundamental responsibility of any society. Your project directly contributes to this by using technology to identify potential threats or violations in public spaces.
  - **Protection of Privacy:** While the project's primary goal is to identify potential threats, it's important to ensure that privacy rights are respected. Implementing safeguards to protect the privacy of individuals in public spaces is crucial.

Novelty

1. **Explainability:** In addition to identifying anomalies our model also provides insights into the process by which it derived the specific outcomes from the provided datasets.
2. **Comprehensive Anomaly Detection:** Unlike many existing projects that focus on specific types of anomalies (like weapons detection), our project aims to detect a wide range of violations.
3. **Versatility in Data Input:** Acknowledging the limitations of high-resolution footage, this project aims to work with a variety of input data, including both images and videos.
4. **Blockchain Integration for Data Security:** The incorporation of blockchain technology adds an additional layer of security to the project. It ensures that the data remains tamper-proof and is securely transmitted to the appropriate authorities.
5. **Potential for Scalability and Replicability:** there is considerable potential for it to be scaled to various environments and seamlessly integrated with existing surveillance systems. This could lead to a significant enhancement of public safety measures on a broader scale.

Problem Statement

The project aims to address the critical challenge of enhancing public safety through the development of an integrated solution for intrusion and anomaly detection in CCTV surveillance, coupled with the use of blockchain technology to safeguard the integrity and authenticity of surveillance data.

The primary issues to be resolved include:

1. Explainability of current models.
2. Data Tampering and Privacy Risks.
3. Ineffective Surveillance Systems.
4. Data Integrity

Previous phase Summary

1. **Dataset Acquisition:**We successfully located and acquired the UCF datasets, which serve as crucial resources for training and testing our anomaly detection models. These datasets provide diverse and realistic scenarios for evaluating the efficacy of our algorithms.

2. **Anomaly Detection Model Development:**Our team dedicated substantial effort towards developing an anomaly detection model tailored to the characteristics of the UCF datasets.

3. **Motion Detection:**Motion detection is a fundamental component of our anomaly detection system. We implemented advanced motion detection algorithms to effectively capture and analyze motion patterns in the video streams.

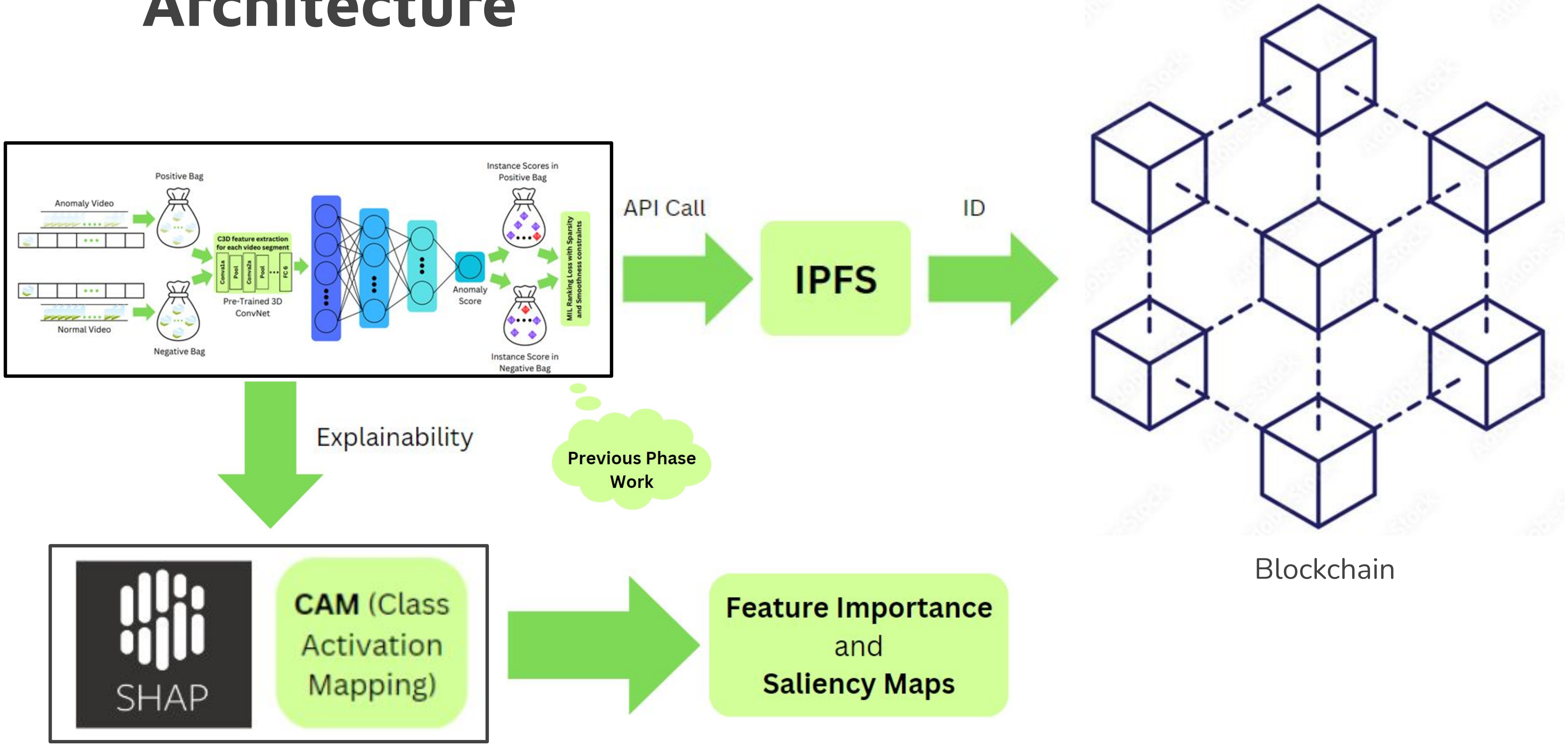
4. **Data Preprocessing:**Data preprocessing plays a crucial role in optimizing the performance of our anomaly detection pipeline. We implemented sophisticated preprocessing techniques to enhance the quality and relevance of the input data, thereby improving the overall efficiency and accuracy of our models.

1. **Introduced SHAP library for model explainability**: Successfully integrated the SHAP (SHapley Additive exPlanations) library into the project's workflow to enhance the explainability aspect of the machine learning model. SHAP provides insights into feature importance and contribution to individual predictions, helping stakeholders understand the model's decision-making process in a more interpretable manner. This step significantly improved transparency and trust in the model's outputs.

2. **Introduced IPFS system in blockchain for data storage:**
    - **InterPlanetary File System (IPFS) integration:** The team introduced IPFS within the blockchain network to enhance data storage capabilities. IPFS is a distributed system for storing and accessing files, utilizing a content-addressable system that ensures data integrity and availability.
    - **Implementation benefits:**Decentralized data storage: IPFS allowed the project to store model outputs and related data in a decentralized manner across multiple nodes, improving fault tolerance and resilience against data loss.
    - **Efficient data retrieval:** IPFS's content-addressable system enables efficient data retrieval by referencing data through its unique hash, reducing latency and improving overall system performance.

Indian Institute of Information Technology Kottayam

Architecture



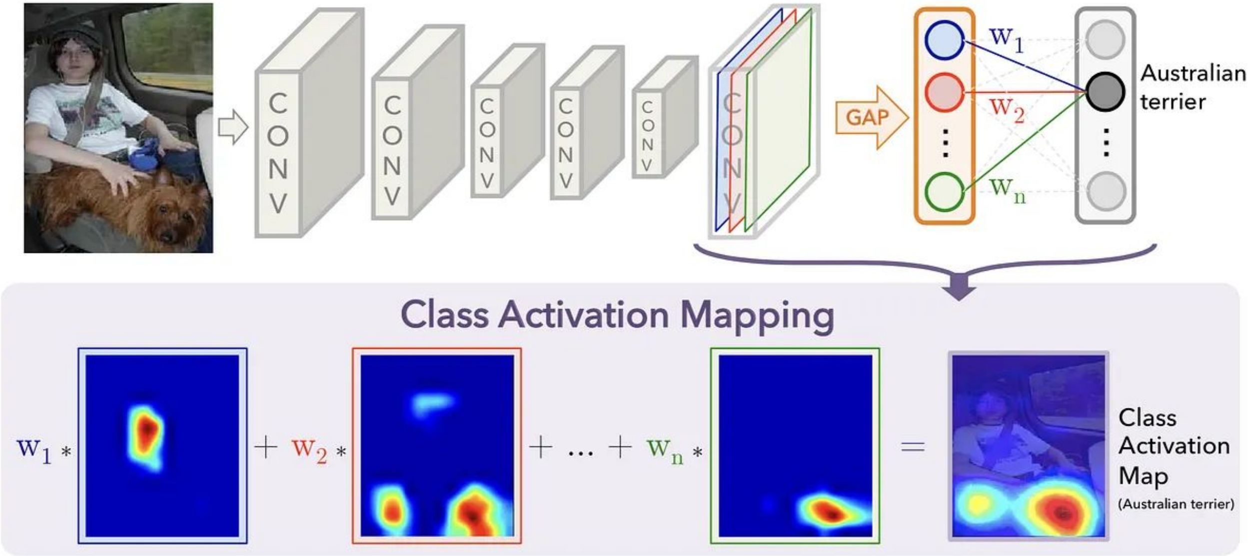
CAM

In order to build trust in machine learning models and move towards their integration into our everyday lives, we need to make “transparent” models that could explain why they predict what they predict.

Class Activation Mapping is an explanation method used for CNNs.

In these networks, the stack of fully connected layers at the very end of the model has been replaced by a layer named *Global Average Pooling (GAP)*. GAP simply averages the activations of each feature map and concatenates these averages and outputs them as a vector.

Then, a weighted sum of this vector is fed to the final softmax loss layer. Using this architecture, we can highlight the important regions of the image by projecting back the weights of the output on the convolutional feature maps.



Class Activation Mapping. The figure is borrowed from [Zhou et al. 2016].

Grad-CAM

Grad-CAM is a more versatile version of CAM that can produce visual explanations for any arbitrary CNN, even if the network contains a stack of fully connected layers too.

We let the gradients of any target concept score (logits for any class of interest such as 'human's face') flow into the final convolutional layer.

We can then compute an importance score based on the gradients and produce a coarse localization map highlighting the important regions in the image for predicting that concept.

More formally, at first, the gradient of the logits of the class c w.r.t the activations maps of the final convolutional layer is computed and then the gradients are averaged across each feature map to give us an importance score.

Guided Grad-CAM was also proposed by combining guided-backpropagation with Grad-CAM (which only produce coarse-grained visualizations).

Grad-CAM++

It is a better alternative to Grad-CAM that according to its authors can provide better visual explanations for network predictions, and does a better job at locating objects even for occurrences of multiple object instances in a single image.

Grad-CAM++ is practically equivalent to a very simple variation of Grad-CAM in which gradients are replaced with positive gradients.

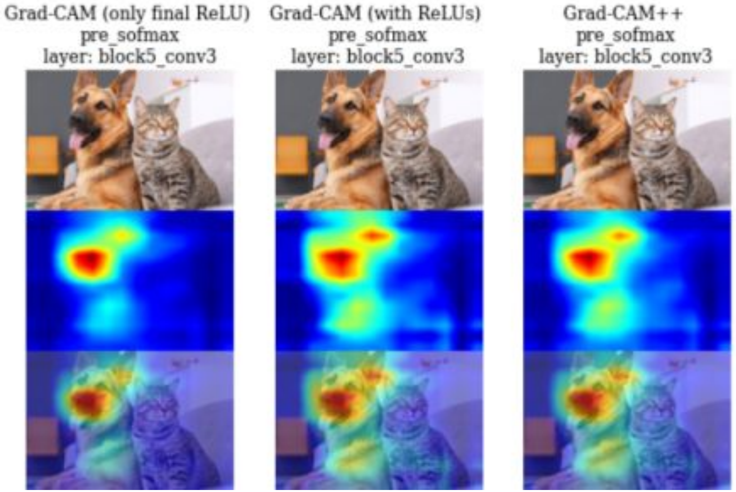


Figure borrowed from https://arxiv.org/pdf/2205.10838v1.pdf

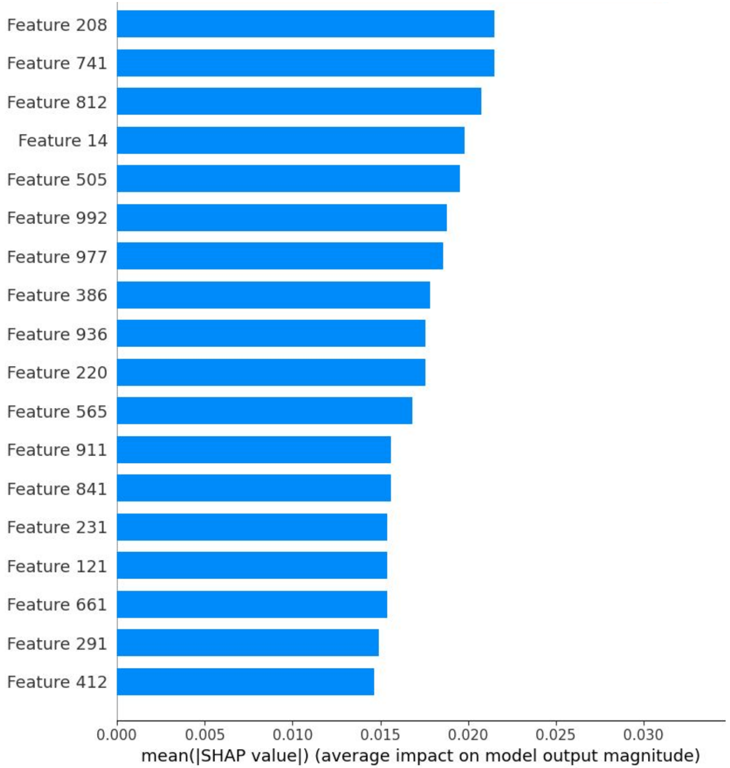
SHAP and its result

1. Creating and Fitting the Anomaly Detection Model:
    a. model = IsolationForest(random_state=0): Creates an Isolation Forest model object with a specified random state.
    b. model.fit(input_data): Fits the Isolation Forest model with the input data to learn the patterns of normal and anomalous behavior.
2. Initializing SHAP Explainer:
    a. explainer = shap.Explainer(model.predict, input_data): Initializes a SHAP explainer object using the Isolation Forest model's predict function and the input data. This step prepares the explainer to compute SHAP values for interpretability.
3. Computing SHAP Values:
    a. shap_values = explainer.shap_values(input_data): Computes SHAP values for the input data using the initialized explainer. SHAP values represent the impact of each feature on the model's predictions.
4. Visualizing SHAP Values:
    a. shap.summary_plot(shap_values, input_data, plot_type='bar'): Creates a summary plot of SHAP values using a force plot. This plot helps visualize the contributions of different features to the model's predictions.
    b. plt.show(): Displays the SHAP summary plot using Matplotlib.

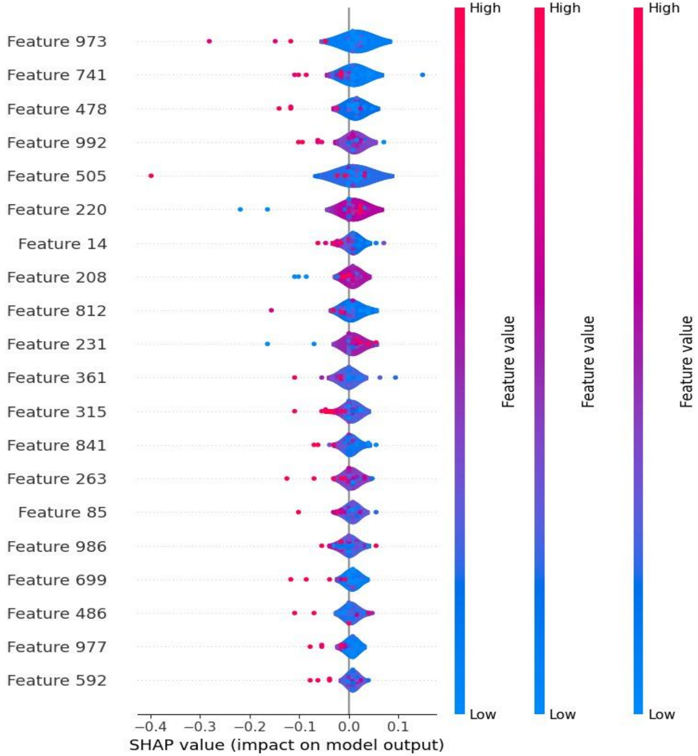
SHAP enhances the interpretability and granularity of our model. The code provided generates a graph that illustrates the output of this enhancement. In this graph, each feature corresponds to a frame of the video, showcasing which frame had the most significant contribution to the specific output predicted by the model. This visualization using SHAP clarifies how individual frames influence the model's decision-making process.

```
💡 Click here to ask Blackbox to help you code faster
import shap
import numpy as np
import matplotlib.pyplot as plt
from sklearn.ensemble import IsolationForest  # Example anomaly detection model

# Load the input data from the .npy file
input_data = np.load('/content/BTP-42/workspace/DATA/UCF-Crime/all_flows/Abuse/Abuse001_x264.mp4.npy')  # Replace with your .npy file path

# Create and fit the anomaly detection model (Isolation Forest)
model = IsolationForest(random_state=0)
model.fit(input_data)  # Fit the model with the input data

# Initialize a SHAP explainer with the model's predict function
explainer = shap.Explainer(model.predict, input_data)

# Compute SHAP values for the input data
shap_values = explainer.shap_values(input_data)

# Visualize the SHAP values (force plot)
shap.summary_plot(shap_values, input_data, plot_type='bar')
plt.show()
```

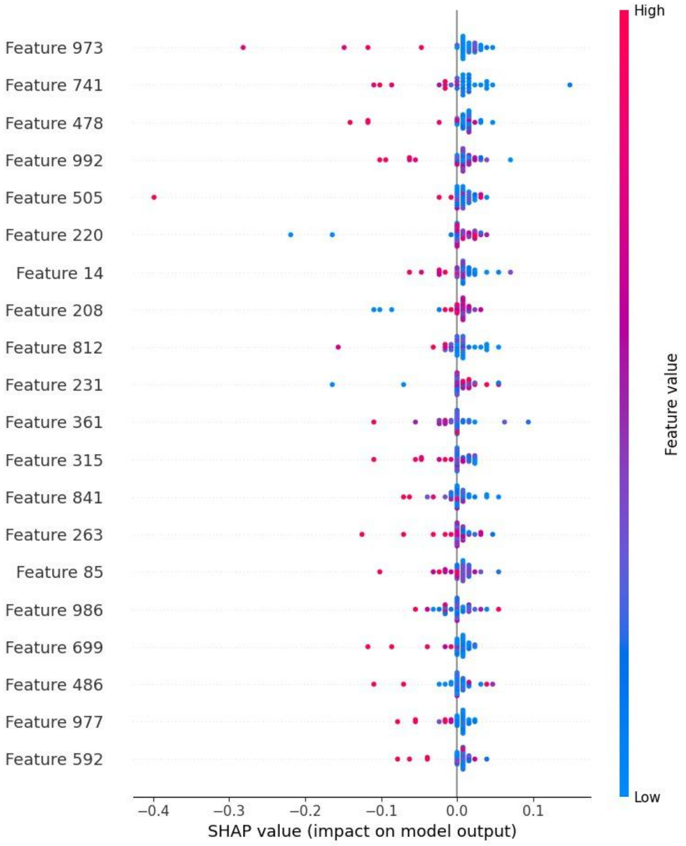
SHAP GRAPH

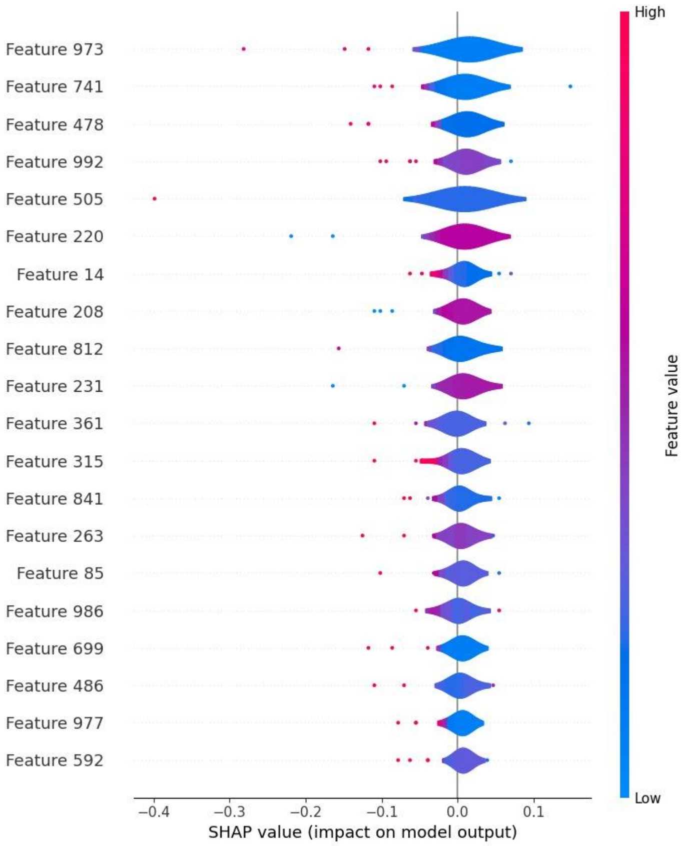


Feature 208
Feature 741
Feature 812
Feature 14
Feature 505
Feature 992
Feature 977
Feature 386
Feature 936
Feature 220
Feature 565
Feature 911
Feature 841
Feature 231
Feature 121
Feature 661
Feature 291
Feature 412
0.000 0.005 0.010 0.015 0.020 0.025 0.030
mean(|SHAP value|) (average impact on model output magnitude)



Feature 973
Feature 741
Feature 478
Feature 992
Feature 505
Feature 220
Feature 14
Feature 208
Feature 812
Feature 231
Feature 361
Feature 315
Feature 841
Feature 263
Feature 85
Feature 986
Feature 699
Feature 486
Feature 977
Feature 592
−0.4 −0.3 −0.2 −0.1 0.0 0.1
SHAP value (impact on model output)
High
Feature value
Low
High
Feature value
Low
High
Feature value
Low

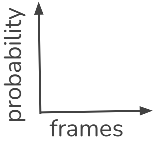

Feature 973
Feature 741
Feature 478
Feature 992
Feature 505
Feature 220
Feature 14
Feature 208
Feature 812
Feature 231
Feature 361
Feature 315
Feature 841
Feature 263
Feature 85
Feature 986
Feature 699
Feature 486
Feature 977
Feature 592
−0.4 −0.3 −0.2 −0.1 0.0 0.1
SHAP value (impact on model output)
High
Feature value
Low

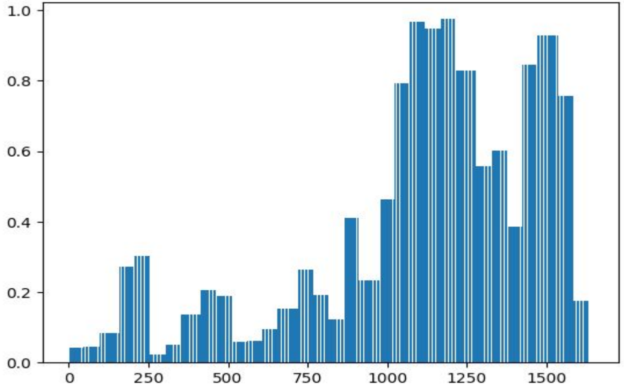

Feature 973
Feature 741
Feature 478
Feature 992
Feature 505
Feature 220
Feature 14
Feature 208
Feature 812
Feature 231
Feature 361
Feature 315
Feature 841
Feature 263
Feature 85
Feature 986
Feature 699
Feature 486
Feature 977
Feature 592
−0.4 −0.3 −0.2 −0.1 0.0 0.1
SHAP value (impact on model output)
High
Feature value
Low

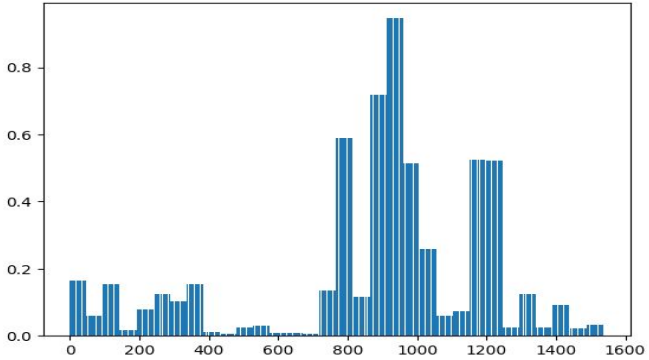
Experimental Results

[Code for model]

80% accuracy achieved



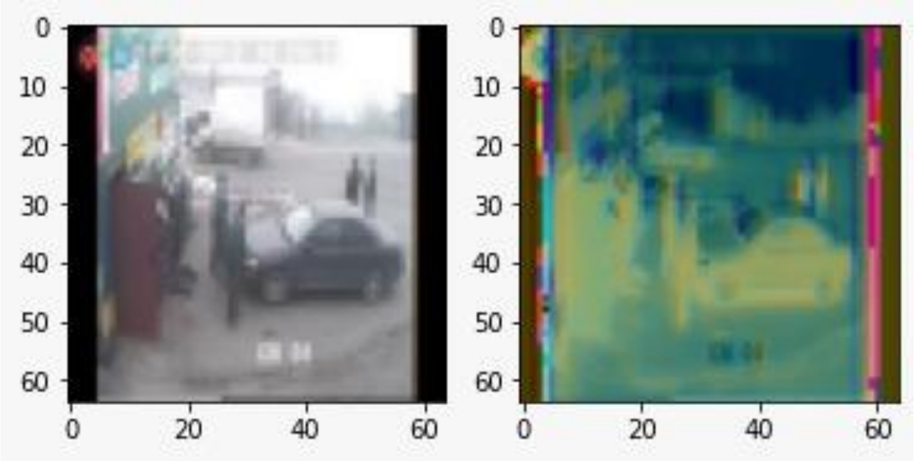


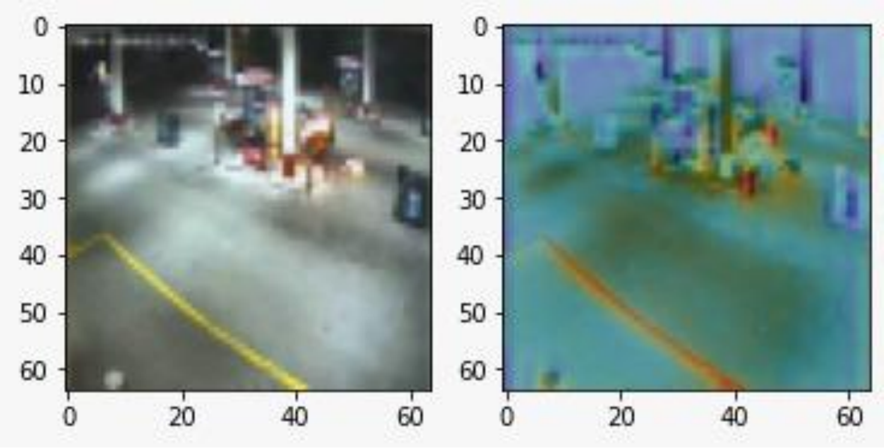


Indian Institute of Information Technology Kottayam

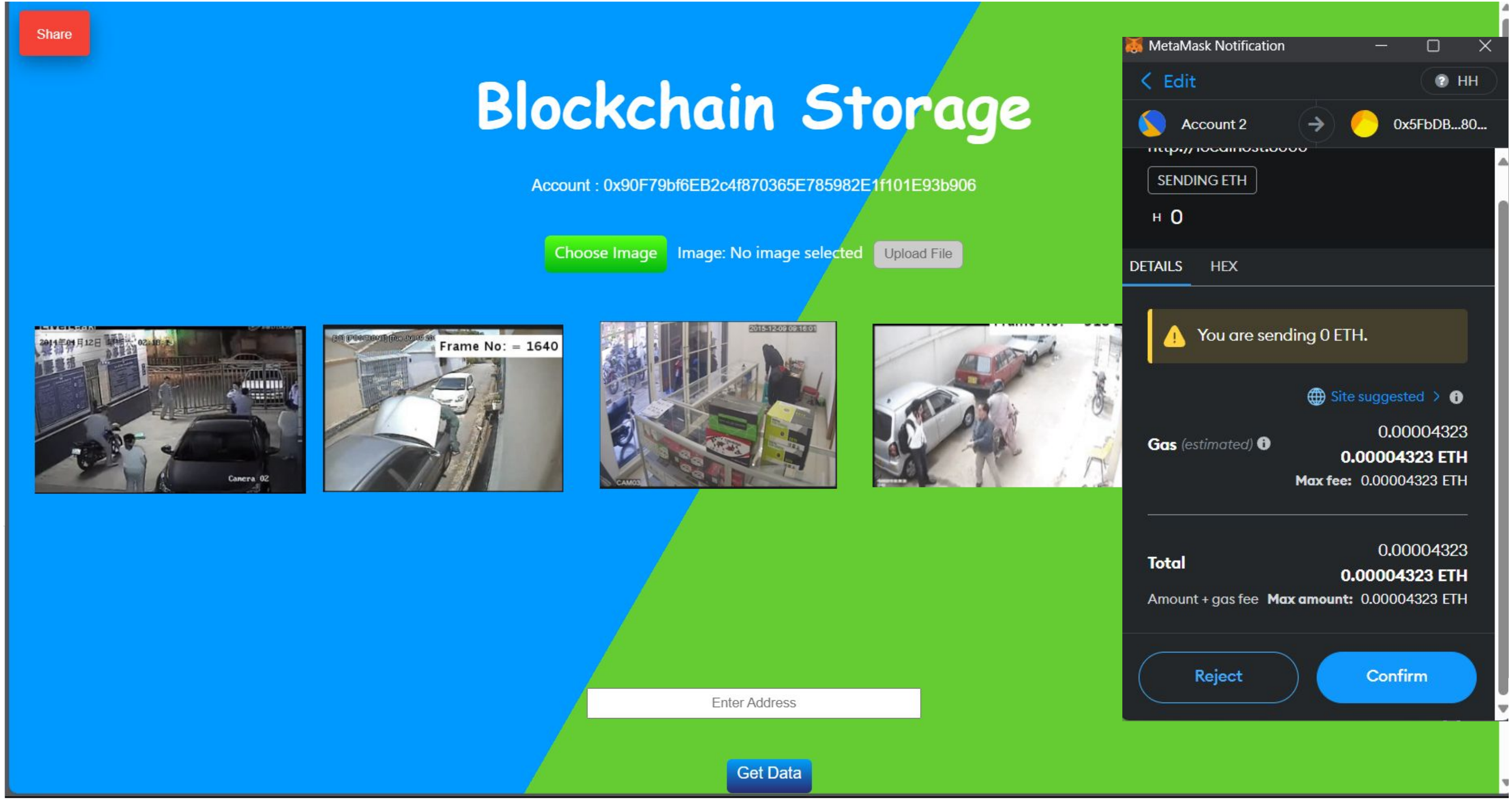
Experimental Results

Car Accident →





← Explosion


Share
Blockchain Storage
Account : 0x90F79bf6EB2c4f870365E785982E1f101E93b906
Choose Image
Image: No image selected
Upload File
Frame No: = 1640
Enter Address
Get Data
MetaMask Notification
Edit
HH
Account 2
0x5FbDB...80...
SENDING ETH
H 0
DETAILS
HEX
You are sending 0 ETH.
Site suggested
Gas (estimated)
0.00004323
0.00004323 ETH
Max fee: 0.00004323 ETH
Total
0.00004323
0.00004323 ETH
Amount + gas fee
Max amount: 0.00004323 ETH
Reject
Confirm

Literature survey

S.No	Authors	Title	Findings	Gap in literature
1	**Muhammad Shafay, Raja Wasim Ahmad, Khaled Salah, Ibrar Yaqoob, Raja Jayaraman & Mohammed Omar**	Blockchain for deep learning: review and open challenges	It reviews existing literature on blockchain integration.	Platform scalability, assurance of secure data exchange, transaction execution latency, platform interoperability support, secure economical models, and computationally expensive consensus protocols
2	**Dev Patel, Harshil Sanghvi, Nilesh Kumar Jadav, Rajesh Gupta, Sudeep Tanwar, Bogdan Cristian Florea, Dragos Daniel Taralunga, Ahmed Altameem, Torki Altameem & Ravi Sharma**	BlockCrime: Blockchain and Deep Learning-Based Collaborative Intelligence Framework to Detect Malicious Activities for Public Safety	a convolutional neural network (CNN)-based Xception model, i.e., BlockCrime is proposed, blockchain technology to securely store the detected crime scene locations, transfer learning has been preferred due to data scarcity	Output report is not being efficiently used. CNN based Model has Uncertainty and bias. Crime Explainability /scene understanding.

contd

S.No	Authors	Title	Findings	Gap in literature
3	**Brestoff, N.E. and Rajagopal, J., Intraspexion Inc**	Using classified text, deep learning algorithms and blockchain to identify risk in low-frequency, high value situations, and provide early warning	Combine usage of Blockchain with Deep learning text classifier to handle the uncertainty.	Explainabity of the decisions mabe by the network. Distributed data for security.
4.	**Kumar, K.K. and Venkateswara Reddy, H.**	Crime activities prediction system in video surveillance by an optimized deep learning framework	Proposes a novel lion-based deep belief neural paradigm	The only purpose is to detect and not what to do with the detected results.

contd

S.No	Authors	Title	Findings	Gap in literature
5.	**Khan, Prince Waqas, Yung-Cheol Byun, and Namje Park.**	A Data Verification System for CCTV Surveillance Cameras Using Blockchain Technology in Smart Cities	CCTV raises privacy concerns, but blockchain offers a modern solution for data integrity and security. It safeguards against data manipulation and provides secure image storage through a distributed ledger.	scalability of the proposed blockchain-based system
6.	**A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian**	Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks	The authors propose Grad-CAM++, an extension of Grad-CAM, to provide generalized visual explanations for deep convolutional neural networks (CNNs). Grad-CAM++ improves object localization and explains occurrences of multiple objects within an image.	The paper addresses the need for more transparent and interpretable deep learning models, particularly in the context of CNNs. Grad-CAM++ specifically focuses on improving visual explanations, which is crucial for understanding how CNNs make predictions.

References

1. Rumesh, Y., Senevirathna, T.T., Porambage, P., Liyanage, M. and Ylianttila, M., 2023, May. Comprehensive Analysis over Centralized and Federated Learning-based Anomaly Detection in Networks with Explainable AI (XAI). In ICC 2023-IEEE International Conference on Communications (pp. 4853-4859). IEEE.

2. Javed, A.R., Ahmed, W., Pandya, S., Maddikunta, P.K.R., Alazab, M. and Gadekallu, T.R., 2023. A survey of explainable artificial intelligence for smart cities. Electronics, 12(4), p.1020.

3. Kobayashi, S., Hizukuri, A. and Nakayama, R., 2023, July. Video Anomaly Detection Using Encoder-Decoder Networks with Video Vision Transformer and Channel Attention Blocks. In 2023 18th International Conference on Machine Vision and Applications (MVA) (pp. 1-4). IEEE.

4. Shafay, M., Ahmad, R.W., Salah, K., Yaqoob, I., Jayaraman, R. and Omar, M., 2023. Blockchain for deep learning: review and open challenges. Cluster Computing, 26(1), pp.197-221.

References

5. Patel, D., Sanghvi, H., Jadav, N.K., Gupta, R., Tanwar, S., Florea, B.C., Taralunga, D.D., Altameem, A., Altameem, T. and Sharma, R., 2022. BlockCrime: Blockchain and Deep Learning-Based Collaborative Intelligence Framework to Detect Malicious Activities for Public Safety. Mathematics, 10(17), p.3195

6. Brestoff, N.E. and Rajagopal, J., Intraspexion Inc, 2018. Using classified text, deep learning algorithms and blockchain to identify risk in low-frequency, high value situations, and provide early warning. U.S. Patent 10,095,992.

7. Kumar, K.K. and Venkateswara Reddy, H., 2022. Crime activities prediction system in video surveillance by an optimized deep learning framework. Concurrency and Computation: Practice and Experience, 34(11), p.e6852.

8. Khan, P.W., Byun, Y.C. and Park, N., 2020. A data verification system for CCTV surveillance cameras using blockchain technology in smart cities. Electronics, 9(3), p.484.

9. A. Chattopadhay, A. Sarkar, P. Howlader and V. N. Balasubramanian, "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 2018, pp. 839-847, doi: 10.1109/WACV.2018.00097.

Thank you