

QATAR UNIVERSITY

COLLEGE OF ENGINEERING

MULTI-ZONAL VEHICLE SURVEILLANCE SYSTEM ENABLED BY A PRIVATE
PERMISSIONED BLOCKCHAIN.

BY

NAJMATH SHARFIYA OTTAKATH

A Thesis Submitted to
the College of Engineering
in Partial Fulfillment of the Requirements for the Degree of
Masters of Science in Computing

January 2023

© 2023. Najmath Sharfiya Ottakath. All Rights Reserved.

COMMITTEE PAGE

The members of the Committee approve the Thesis of
Najmath Sharfiya Ottakath defended on 03/01/2023.

Prof. Dr. Sumaya Al Maadeed
Thesis Supervisor

Dr. Mohammed Abdullah
Committee Member

Dr. Elias Edward Yacoub
Committee Member

Dr. Pilsung choe
Committee Member

Approved:

Khalid Kamal Naji, Dean, College of Engineering

ABSTRACT

OTTAKATH, NAJMATH, SHARFIYA., Masters : January : 2023, Masters of Science
in Computing

Title: MULTI-ZONAL VEHICLE SURVEILLANCE SYSTEM ENABLED BY A PRIVATE PERMISSIONED BLOCKCHAIN.

Supervisor of Thesis: Prof. Dr. Sumaya Al Maadeed.

Privacy, security, accessibility, and reliability are the most essential characteristics of a public security system. Existing surveillance systems provide monitoring and surveillance-based security. However, their inference depends on manual monitoring and action, which may result in a delay in response. Using computer vision techniques, automated surveillance with monitoring through anomaly detection and tracking has been made possible. Nevertheless, they require a centralized storage system, which may result in a delay or security breach, provide a single point of failure, and render the system unavailable and unreliable. Several states of the art have proposed the use of blockchain, a decentralized ledger utilizing a private, permission-based network, to improve the framework's dependability. However, an analysis of its viability in relation to the security standards of confidentiality, integrity, and dependability, as well as its use in an enterprise-grade application such as public security, which may necessitate a highly scalable network, is required. In this regard, the proposed framework is a multi-surveillance system for vehicles that performs a privacy-protected image analysis to re-identify vehicles through images captured by various cameras and analyzed using computer vision techniques, which will be accessible to other nodes and/or surveillance zones via the blockchain ledger. A private permissioned blockchain network, Hyperledger Fabric, is evaluated for improved reliability and reduced latency using fast

and lightweight image analysis tasks, such as combinations of feature extractors and lightweight CNN (Convolutional Neural Network) models. With different approaches in three domains, the accuracy and time required by an edge-based inference tool are measured, resulting in a lightweight tool for surveillance via a permissioned private blockchain network.

DEDICATION

To my parents, grandparents, siblings, aunts, cousins, nieces, and nephews.

ACKNOWLEDGMENTS

In the name of Allah, no man accomplishes a task on his/her own; we are all shaped by the environment we inhabit, the people we interact with, the education we receive from them, and the various forms of guidance we receive. Thus, I would like to acknowledge everyone who have walked with me while I tread my path, provided me inspiration in multitude of forms during research. In particular, my supervisor, Prof. Dr. Somaya Al-Maadeed, whose patience and guidance are exemplary, without which I may not have accomplished this thesis. Adding to that Dr. Amr Mohamed and Dr. Abdullah Khalid Al-Ali who were a constant support in the initial phase of the research. I cannot but mention my family, colleagues, and friends, who were concerned about my success and were an emotional as well as psychological support. In the end nothing is ever accomplished unless it is decreed on you by your creator and thus, I thank the almighty that wrote this as part of my destiny, the provider, the protector, the fair, the most just and the all-knowing.

TABLE OF CONTENTS

DEDICATION	v
ACKNOWLEDGMENTS	vi
LIST OF TABLES	x
LIST OF FIGURES	xii
Chapter 1: INTRODUCTION	1
1.1 Background and issues.....	1
1.2 Research Purpose and questions	4
1.3 Key Contributions	5
Chapter 2: Literature Review.....	7
2.1 Blockchain in video surveillance systems.....	7
2.1.1 Privacy	7
2.1.2 Security	9
2.1.3 Access Control	10
2.2 Vehicle Detection and Recognition	17
Chapter 3: METHODOLOGY	23
3.1 Vehicle ROI	23
3.1.1 Convolutional Neural Networks	24
3.1.2 Deformable convolution neural network.....	26
3.1.3 FPN and Mask R-CNN.....	27
3.1.4 Experimental Setup	30
3.1.4.1 Dataset	30
3.1.4.2 Performance Metrics.....	32

3.2 Vehicle Feature extraction and matching.	33
3.2.1 SIFT.....	34
3.2.2 ORB.....	35
3.2.3 BRIEF	36
3.2.4 Experimental setup	37
3.2.4.1 Datasets.....	37
3.2.4.2 Evaluation Metrics	39
3.3 Feature Clustering and classification.	39
3.3.1 SURF.....	40
3.3.2 MSER	41
3.3.3 KAZE.....	41
3.3.4 Experimental Setup	41
3.4 Light weight CNN classification	42
3.5 Blockchain.....	43
3.5.1 Smart contract.....	45
3.5.2 Experimental setup.....	47
Chapter 4: EXPERIMENTAL RESULT	52
4.1 Vehicle frontal image segmentation and detection	52
4.2 Vehicle ROI and feature descriptor.	53
4.3 Feature clustering and classification approach.....	55
4.4 Light weight CNN based approach.	56
4.5 Blockchain framework	57
Chapter 5: DISCUSSION.....	63
5.1 Vehicle instance segmentation	63

5.2 Vehicle make identification.....	65
5.2.1 Comparison of the results with state of art	65
5.3 Blockchain network	66
5.4 Evaluation of end-to-end framework	70
5.4.1 Reliability of detection.	70
5.4.1.1 Accuracy of detection	70
5.4.1.2 Feature matching.....	71
5.4.1.3 Feature Clustering and matching	71
5.4.2 Blockchain security	72
5.4.3 Privacy	72
5.4.3.1 Blockchain.....	73
5.4.4 Availability	74
5.4.5 Cost	76
5.5 Key findings, limitations, and future works	76
Chapter 6: CONCLUSION.....	80
References	81
Appendix : A	90
Ablation study of CNN.	90
Appendix : B.....	91
Overall Docker statistic graphs	91

LIST OF TABLES

Table 2.1. Applications of blockchain with surveillance.....	13
Table 2.2. Latest Rei-identification techniques.....	20
Table 3.1. Experimental Setup of CNN based model.....	44
Table 4.1. Classification accuracy and detection accuracy using mAP with latency	52
Table 4.2. Ablation study with different backbones and deformable convolution.....	52
Table 4.3. Ablation study based on data augmentation	54
Table 4.4. Feature Matching	55
Table 4.5. Feature Clustering and matching	56
Table 4.6. CNN based approaches.....	57
Table 4.7. Evaluating an image stored on the ledger.	58
Table 5.1. Comparison with existing literature.....	64
Table 5.2. Database1 accuracy comparison.....	67
Table 5.3. Database2 accuracy comparison.....	67
Table 5.4. Database3 accuracy comparison.....	68
Table 5.5. Database4 accuracy comparison.....	68
Table 5.6. End to End latency analysis	74
Table .1. Ablation study of CNN, layerwise analysis.	90

LIST OF FIGURES

Figure 2.1. Surveillance enabled by a consortium blockchain.	15
Figure 2.2. A typical video surveillance block structure where the object of interest information is stored on the chain.	15
Figure 2.3. A Vehicle detection and identification pipeline.	16
Figure 3.1. Base pipeline of blockchain enabled vehicle surveillance.	23
Figure 3.2. Deformed offset filter applied during convolutional operation.	26
Figure 3.3. Instance segmentation model(mask RCNN + R50 + FPN).	29
Figure 3.4. Mosaic tiled image.	29
Figure 3.5. Mosaic tiled image.	30
Figure 3.6. Dataset Distribution per class.	31
Figure 3.7. Data Augmentation techniques utilized to balance the dataset.	31
Figure 3.8. Local feature descriptor and matching.	34
Figure 3.9. SIFT key points with size and without size.	35
Figure 3.10. 874 Matching key points between two makes with label 11 from Compcars dataset.	35
Figure 3.11. ORB feature points.	36
Figure 3.12. 107 matching features	36
Figure 3.13. Dataset samples [54]	39
Figure 3.14. Bag of Visual Words Approach.	42
Figure 3.15. Blockchain network setup with two supernodes with access to two channels.	48
Figure 4.1. Writing 10,000 transactions on single channel network.	59

Figure 4.2. Reading 10,000 transaction on single channel	60
Figure 4.3. Querying, reading, and writing multichannel.	60
Figure 4.4. Read and writing cross channel (reading from channel 1 and writing to channel2).....	60
Figure 4.5. Memory usage for 10,000 writes from a single node to a single channel	61
Figure 4.6. Memory usage for querying on a single node to single channel.....	61
Figure 4.7. Reading/Querying and writing cross channel.	62
Figure 4.8. Querying and writing multi-channel approach.	62
Figure 5.1. Class-wise accuracy.	64
Figure 5.2. mAP vs Accuracy vs Inference time per model	65
Figure 5.3. Time taken in cross channel data transfer.	69
Figure 5.4. Read and write through a complete multi-channel setup.	69
Figure 5.5. End to end latency analysis, The legend represents the configurations set in Table 5.6 for end to end blockchain enabled surveillance.....	75
Figure .1. Docker statistics graphs of a single channel read activity.....	91
Figure .2. Docker statistics graphs of a single channel write activity.....	92
Figure .3. Docker statistics graphs of a cross channel write activity.....	92

CHAPTER 1: INTRODUCTION

Reliability, security, privacy, and availability are quintessential in surveillance due to its prevalence of its use in current times. Surveillance and monitoring are essential components of any community's safety and security. In Qatar, every street corner contains at least one or two cameras that continuously monitor the area. Cameras provide different angles of view, live stream as well as capture videos that can be stored for real time as well as long-term use. An automated intelligent system can enhance the inference from the camera or sensors that are involved. Further, data generated from these are critical to the privacy and security of the people and entities involved in the surveillance system [1].

1.1 Background and issues.

A typical surveillance system is a collection of sensors and cameras that communicate with a control station. Cameras and sensors form a wireless or wired sensor network which monitors, analyses, and stores the content. The sensors sometimes form an edge or fog network for computational efficiency providing a real time advantage [2]. Cloud storage is the most common storage location due to its scalability. This leads to extended vulnerability in the form of advanced cloud-based attacks as it forms a centralized storage system [3], [4]. Security enterprises conforming to the cloud based centralized storage approach like that of Hikvision was easily left vulnerable to attack. This does not just intrude the surveillance system and access the monitoring data but can also modify the content within them. An approach to avoid the centralized cloud server approach and to identify a system where modification is monitored was required [5]. This led to research in blockchain for surveillance which is a decentralized and

distributed system without a single point of vulnerability that can fail the whole system.

Blockchain provides access control, security and privacy of data generated through this system where unauthorized access, tampering and modification of data can be a matter of great concern. It has infiltrated many applications providing security and trust. A typical blockchain system may require at least 50% or more nodes to be compromised for an attack to take place. And so, strengthening against security risk, protecting content and as it is time stamped, provides accountability to the surveillance system [6]. Blockchain can be leveraged for protecting the transfer of images and video, protecting the privacy, integrity of the information shares, as well as enabling authentication and authorization of access of the information generated through surveillance.

With this determined, our research problem entails that the framework enabled for blockchain is efficient in terms of security, reliability, and privacy. For this we model a surveillance application: a vehicle surveillance through blockchain network and evaluate existing techniques with surveillance and blockchain.

To identify a reliable surveillance model for Vehicle Identification. The research objective was to identify an accurate and reliable model that can be used for the purpose of identification. Image processing and computer vision applications are reviewed and an efficient technique for accurate recognition is identified. To evaluate a model for recognition, we modify an existing dataset for instance segmentation where each unique instance of the vehicle is identified with region of interest segmented. The segmentation of the model is performed with well-known state of art method Mask RCNN (Mask Region based Convolutional Neural Network) [7] and an ablation study is conducted to identify the best performing model with the time taken evaluated. The imbalance in existing modified dataset is noted and different types of augmentation techniques and of

the significant mosaic tiled approach is performed to further enhance the generalization of the model [8]. This can be an alternative of the whole pipeline as it performs classification, segmentation as well as detection.

With the region of interest segmented, the objective for a private, reliable, and real-time detection was to identify a lightweight algorithm for vehicle make identification to uniquely identify and re-identify vehicles. Towards this objective, an evaluation of the traditional re-identification approaches using image matching by feature descriptors matching and computer vision-based classification approaches with light weight models were accomplished. With the key point descriptors extracted the data is stored on the chain and then matched with ledger to re-identify the vehicle. This enabled a more private and reliable method for re-identification[9].

However, blockchain is a distributed network and the system scales with increasing number of nodes in the network. With the participants in the node increasing the ledger size, the transaction content increases. In addition, the information from videos and/or images are of large size which adds to the increased computational time of this framework [10]. Thus, the feature descriptor is utilized for re-identification which adds lesser overhead to the network. This re-identification scheme is enabled on the blockchain is evaluated for its real time use, its scalability in the context of a private permissioned network, Hyperledger fabric. Hyperledger Fabric is a blockchain network that enables access control and has an extensive consensus mechanism to validate the transaction [11]. Thus, providing a private and secure network for high security applications. However, the sophisticated consensus mechanism and channel-based data control, with the application of a large-scale surveillance system scaling multiple domains of use such as multiple organizations is a matter of concern [12]. A trade-off in availability for

a real time use case in terms of the number of transactions generated through a single channel network as well as a multichannel network was determined.

Further, in this application we evaluate the writing and query of image and/or its features as an asset in the ledger leading to improved privacy. Thus, leading to light weight protocol for the purpose of vehicle re-identification enabled by Blockchain.

1.2 Research Purpose and questions

The main objective of this research can be thus defined based on the following section.

- Identify an accurate and fast model for segmenting the region of interest for private and reliable detection.
- Evaluate robust feature extraction techniques for the purpose of re-identification and compare them with light-weight CNN (Convolutional Neural Network) based classification models which can be deployed on the edge.
- Evaluate a private, permissioned blockchain network for performing a reliable, private, and secure surveillance system in terms of its scalability and availability.
- State the trade-off in using a private permissioned blockchain system with tiered access control and without access control.

The research question that arises from these objectives are a question of confidentiality, integrity, and availability. How far is the confidentiality is preserved where confidentiality is enabled by a framework that has only the object of interest segmented protecting the privacy.

Furthermore, how reliable is the network in terms of the information shared, which is measured by the accuracy of the classification and segmentation. An image matching scheme is proposed using features for make classification and re-identification as well as image classification approaches as an alternative for make identification. Reliability of surveillance is thus enabled in this approach, and thus the integrity of the approach is secured. Further, with a private permissioned network, access control is enabled using a controlled sharing of the ledger. However, there arises a question of how efficient it is for a real-time surveillance task, which requires an end-to-end analysis of the whole framework with a single channel approach where the access control is equally distributed for all and a multi-channel approach enabling cross-channel communication. This thesis quantifies the reliability and integrity, through accuracy of detection and the availability, through latency of the framework.

1.3 Key Contributions

The key contributions of this thesis are that a reliable, secure, and private blockchain-enabled multi-surveillance framework is proposed. To preserve privacy and improve the accuracy of detection, a pipeline of segmentation and classification was employed. The region of interest segmentation was performed using instance segmentation with mask-RCNN with Resnet-50 backbone, which produced a high accuracy of detection and recognition of each unique instance of vehicle.

Then, several feature sharing approaches for re-identification and classification are evaluated for accuracy and inference time with the segmented region of interest from vehicle frontal images of state-of-the-art datasets. The BoW approaches using feature clustering and matching achieved high accuracy. However, with a trade-off in

time compared to the light-weight CNN approach, with a limited number of convolution layers, produces lower accuracy with a low inference time of 1 ms. With a robust and accurate feature extraction identified, a blockchain network was setup and evaluated for blockchain-based inference. The setup included two scenarios, with the first being a single-channel network with each node provided with the same ledger, and the second being a multi-channel network where a set of nodes were part of a Super node network sharing a ledger with information from other nodes in a private manner. This was tested on the Hyperledger Fabric network, an open-source tool for private permissioned blockchain network orchestration.

The thesis is thus structured based on the following where chapter 2, details the overview of the existing blockchain applications in video surveillance system and proposes a generalized framework for video surveillance on blockchain which enables privacy and security. Further for a vehicle surveillance application, the existing literature of vehicle identification through make classification and re-identification schemes are detailed with requirement for light weight models for the purpose of real time surveillance proposed. Further in chapter 3, the framework discussed in chapter 2, is methodized through vehicle region of interest segmentation for privacy and reliability. Vehicle classification and similarity matching based on make is detailed by three tasks, that is matching through a combination of key-point descriptors and feature matching algorithms, a bag of words approach using feature clustering and a lightweight CNN model. Further, a private permissioned blockchain setup with multichannel feature sharing framework for re-identification. Chapter 4 presents the experimental results, and the key findings are discussed in Chapter 5. Chapter 6 concludes the thesis .

CHAPTER 2: LITERATURE REVIEW

Blockchain-enabled surveillance is a new field. To survey the existing literature, the current state of the art in blockchain in video surveillance is charted. A framework is proposed for common approaches. Vehicle re-identification being the task performed through this framework, existing literature in computer vision for this approach is presented and the problems stated. A solution is then proposed through blockchain in the succeeding chapter based on this review. The following subsection reviews in detail the existing literature and discusses the findings from it.

2.1 Blockchain in video surveillance systems

The aim of this study is to identify current state of art approaches in the domain of vehicle surveillance coupled with blockchain. Research on most recent applications with blockchain was performed by filtering out the most relevant studies to blockchain based surveillance applications. It was noted that blockchain provides one or all the following: privacy, security, and access control for surveillance systems. This is enabled by different types of blockchain setup and the area where the distributed network is utilized. The following section details the current state of art in blockchain based surveillance in terms of privacy preservation, security adherence and access control.

2.1.1 Privacy

One of the most important requirements of any image or video data captured from any device is privacy. Images contain identifying information as well as vulnerable content. These images or video content generated may be restricted under privacy laws for unauthorized access. Individual privacy is a required element of public security.

The current available devices cannot be selective and there's need to monitor a wide coverage area, a scheme to secure privacy is required. Several literatures have dealt with this problem.

By taking limitations of scalability and storage capacity into consideration Fitwi et al. in [13], used lightweight blockchain named (Lib-Pri) for privacy protection where tasks like checking integrity of the videos, blurring keys management, feature sharing and video access sanctioning were performed. Edge computing was performed for real time video analytics where the video was split into frames and a reverse chaotic mask was applied to images which is then stored in an off-blockchain storage. A person of interest is then identified by facial features by applying computer vision techniques on the frames to extract the features and identify a face. A suspected individual's facial features are compared with the original and pushed to the blockchain node to identify the location and recognize the identity. A federated blockchain approach was used in a private permissioned manner with smart contract modified for privacy and confidentiality, with user access control authorized.

Camera identity was another approach used to preserve the privacy of an individual in [14] where Blocksee, a video surveillance system in smart cities was designed in situations where ambiguity of IoT devices posed a threat due to vulnerability of malicious users manipulating the video content. Camera identity enabled validation and immutability to camera settings stored which had camera ownership details enabling privacy through blockchain. A private permissioned architecture was used where access control through ownership was imposed.

2.1.2 Security

A centralized system can be under threat under many circumstances. Blockchain secures the data collected using its function of hashing data which provides data integrity as well as secure storage. Video credibility was the main focus in [15] where they used a blockchain based scheme named Video-Chain, where video integrity evidence was saved on the blockchain. It follows a consortium blockchain, where the entities are given tokens based on privilege of access. The application layer of the video chain updates and verifies the evidence. A new data storage mechanism was built, Trusted Video evidence storage (TVES), which stores both evidence and original data. A high transaction rate protocol, VideoChain, was used for validation. In the video processing part, the evidence was collected by cutting it 10 minutes apart and compressed. A hash of the video was computed as evidence of video integrity. Reliability was improved by adding backup to the original video. Analysis based on security and efficiency of the video chain proved to be a suitable option for implementation.

Dam surveillance was an application which the authors in [16] devised a scheme to secure an IoT solution. A distributed and long-term security solution was accomplished through blockchain technology by providing authentication, data storage, integrity, and traceability of data delivery through the UAV cloud. The performance was measured based on the data delivery ratio.

A combination of convolutional neural network, Interplanetary file system (IPFS), edge computing and permissioned blockchain were utilized for massive data storage, real-time monitoring and large-scale information acquisition in [17]. Content oriented surveillance was accomplished by identifying dangerous individuals with sensors and tracking them. Passive imaging and detecting concealed objects were performed and

secured where data integrity could be kept valid using blockchain.

Multilayered network is usually used with a private blockchain for secure surveillance in addition to a secure storage. The term IBSS (independent blockchain surveillance system) was introduced in [18] where high level layers were used to form the blockchain layer to hash and store the data in an IPFS (Inter planetary file storage) storage which is a peer to peer distributed decentralized storage network. The sensors (cameras) act as nodes to the blockchain where the video content is hashed and saved to the IPFS validated through a consensus mechanism. With the IPFS system, a secure large data storage was achieved with blockchain preserving privacy and security of the video data captured from the sensors.

2.1.3 Access Control

Controlled access to the blockchain transaction secures privacy of the individuals as well as restricts the access to unauthorized individuals or organizations for content related to entities not involved in the surveillance incident. Several modes of access control are identified in state of art where Jeong et al., in [19] and Deepak et al., in [20] presents the state of art in management applications of blockchain with surveillance. Hyperledger fabric, a private blockchain which uses IPFS (inter planetary file system) and CDN (Content defined networking) for storage that are decentralized are detailed in this review.

Access control in terms of content was another application where authorization of access to the decentralized storage was accomplished by Balint in [8] where a blockchain based system for storage of video footage was presented. In this system, data was stored off-chain and a storage platform named storj platform, a decentralized storage system,

was used for storage which has an elaborate encryption method that secures the data. Several types of decentralized storage were also compared in this paper for its use as an off-chain storage for blockchain.

Protecting the index of data stored off-chain was implemented by Nikouei et al., in [21]. They identified the problem of collecting data at the edge along with feature extraction. The closest nodes in the fog network were used to classify the features. Misleading the surveillance system can be easily done at multiple layers where tampering can happen in the cloud system. A blockchain enabled scheme was applied to protect the index of the data at the edge, and fog was used to secure the data. Most decentralized off-chain storage incur high cost in terms of smart contract execution or operational costs and use computationally complex algorithms for encryption. An apt solution for storage that involves better security, lower cost and improved computational efficiency is required.

A lightweight mechanism was further identified on the context of restricted access. It was accomplished by M. Singh et al. in [22] created a lightweight mechanism named one drone one block (ODOB), which was used for surveillance with drones. A modified blockchain structure was envisioned here. ODOB decouples the block ledger from the block header to form a distributed architecture. Here each drone can only access their own block. This makes it simple, trustworthy, and lightweight.

Another lightweight implementation of blockchain with surveillance was achieved in [21], where video metadata was stored on the blockchain to support video integrity. The video metadata such as Frame Rate, Video Position Sequence, Video Frame, and Storage Address was stored as transaction. The video was stored in a distributed storage system called IPFS. The setup of the device included a video camera module connected

to a raspberry pi3 with IPFS storage running on a private instance. Resultant of latency evaluation concluded that only 8 milliseconds was utilized for the whole process, from capturing to storage in IPFS system, thus, achieving video integrity in surveillance.

Apart from the human aspect of suspicious behavior, [23] used surveillance for autonomous detection of stolen car detection and inspections where a blockchain based platform was used. Both remote and local processing of video feeds were done to search for suspicious vehicles. Verification of the suspected vehicles was achieved by blockchain validation system. An open-source license plate recognition model-DeepANPR (Deep learning based Automatic number-plate recognition) using the SUN database, was used for license plate detection. Car recognition was achieved using a ResNet-152 architecture and trained using Cars dataset. Once an anomaly is detected and verified, a local authority can be alerted and published in the blockchain. This system ran at a fault free time of 40 secs producing fast results.

Chain codes or smart contracts were elaborately used for access control by [24], [25] where Hyperledger fabric was adapted for public security. Machine learning and blockchain were combined in this method where real time surveillance was used to identify and track suspected faces. Here the latency was used as an evaluation metric to identify its effectiveness in reporting a suspect. Surveillance events were notified and embedded on to the permissioned blockchain, specifically Hyperledger fabric, which further enabled access control through smart contracts or chain codes in this application. Scalability was achieved due to the federated architecture. Real-time suspect monitoring was achieved successfully with minimal delay. The key points identified from the literature review state that latency is a prime measure of good performance in blockchain based applications for surveillance, hence, this factor must

Table 2.1. Applications of blockchain with surveillance

Method	Appl.	Type	Sim.	Smart Contract
[26]	National security	Consortium	Hyperledger	Yes
[16]	UAV based surveillance	Dam Public	Modified Bit-coin	No
[13]	Suspicious activity detection	Consortium	Modified	-
[27]	Vehicle detection	-	Modified block	No
[24]	Person identification and recognition training	-	-	Yes
[17]	Suspect identification	Permissioned	-	Yes
[22]	Video storage in multi-surveillance	Permissioned	-	Yes
[28]	Person surveillance and forensics	Private	-	-
[14]	Smart city monitoring	Private	Per- missioned	-
[29]	Suspicious person surveillance	Private	Per- missioned	-
[30]	Indoor surveillance IP camera	Public	-	-
[21]	CCTV surveillance	Private	Modified Hash	No
[31]	Smart home surveillance	Public	Per- missioned	Yes

be measured in any blockchain based applications. Privacy is protected on the chain with access control. However, with images and videos on the chain, the privacy remains where participants of the network can still view them. With large scale networks, the complexity of access control limitations increases, thus making a fast real time surveillance a delayed process.

Existing state of art have solved this problem by introducing, artificial intelligence methods such as deep learning and machine learning to extract key frames, or to detect relevant object in a captured scene. However, the issue remains that the object or person of interest is identified visually. A merger of automated surveillance approaches to analyze the scene and detect anomalies to extract specific features, such as abnormal vehicle surveillance and tracking using surveillance applications like dam surveillance, face surveillance etc. have been mentioned in Table ???. The table details the type of surveillance and the type of blockchain simulation used and if there was a smart contract executed. Smart contracts are automated codes that enables interaction between clients and blockchain network. The smart contracts can be utilized for the purpose of automated surveillance reading, writing, and querying of data stored on the chain.

State of art emphasize on lightweight and fast operations since the applications are real time and require very low latency. Scaling the network may further increase the latency as consensus mechanisms are involved this being an assumption requires to be evaluated for different configurations of private permissioned blockchain network where the multi-channel consortium network such as that of the Hyperledger fabric providing controlled access. A large organizational network was evaluated in [4] but an extension of it was required for identifying the performance on multi-channel approach where information was programmed to move from one channel to the other in a hierarchical

fashion.

A blockchain based video surveillance framework is as shown in the Figure 1. A consortium blockchain network with multiple parties having access control managed through policies is shown in 2.1. The transaction data can be the image, information on the image that is the image metadata, time of block creation and/or features of the image etc. such as shown in 2.2 for person re-identification. For the purpose multi-surveillance, the image features are stored on the chain and queried from the blocks.

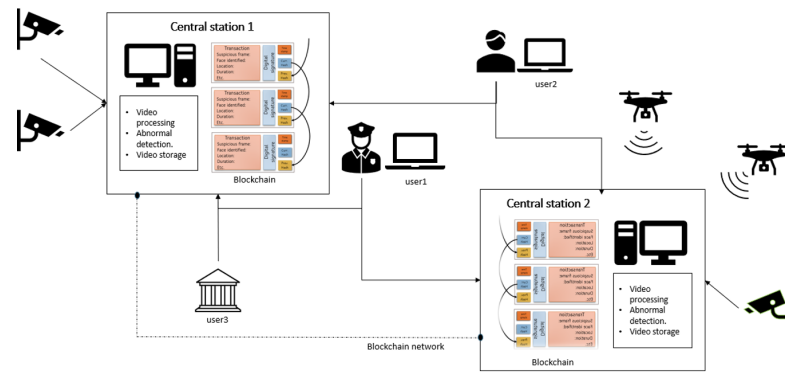


Figure 2.1. Surveillance enabled by a consortium blockchain.

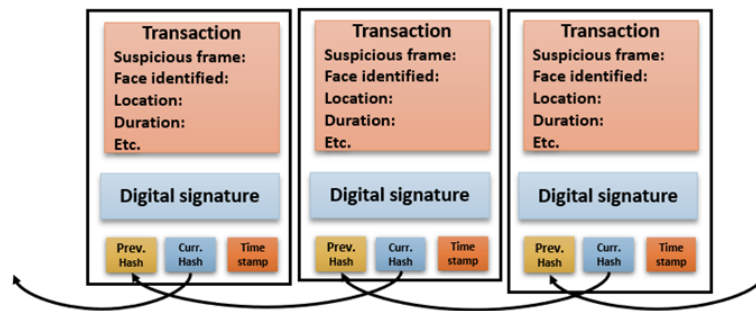


Figure 2.2. A typical video surveillance block structure where the object of interest information is stored on the chain.

Further, fast, and light weight automated surveillance is required to be identified. The research directions from this review details that there is a lack of standardized metrics to quantify the performance of the blockchain based surveillance system. With

the framework presented, the question arises on a real-time surveillance use case where the quantity of latency is of high significance and so there is a need to evaluate the effect of time on availability of resource for surveillance inference. The reliability of that inference is also required to be quantified to evaluate the quality of the framework. The trade-off in terms of availability and reliability is a requirement to be identified.

In the context of vehicle surveillance, vehicle detection and identification are a widely researched area where the detection accuracy and speed of detection provide reliable inference of type of vehicle data. This adds to the reliability of the whole blockchain based surveillance system.

Detection is a broad term used for multiple computer vision tasks that include vehicle re-identification and vehicle recognition which further includes several pre-processing tasks such as vehicle segmentation. For a private and reliable system, the detection should be fast and accurate, leading to this requirement [32]. Although machine learning and deep learning models are accurate while trained on a large dataset of the same class, the amount of complexity of the model as well as the inference time is a question. Current state of art in vehicle re-identification is presented in this section.

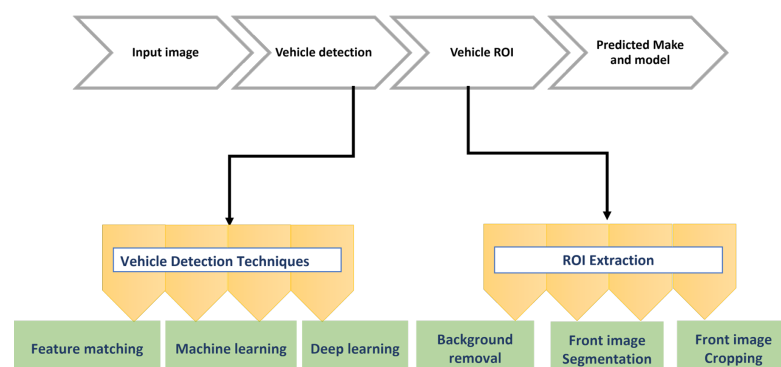


Figure 2.3. A Vehicle detection and identification pipeline.

A typical vehicle detection system is presented below in 2.3, a vehicle's unique

characteristic labels are its make and model. Each vehicle has a distinct structure and logo which is utilized as a pattern for identification. In addition, license plate information can improve the identification of the vehicle. Re-identification is another approach that is required to be performed for tracking the vehicle through multiple cameras which further utilized unique identifying features of a vehicle.

2.2 Vehicle Detection and Recognition

There are several techniques that can be performed in the pipeline using feature matching, machine learning and deep learning. Furthermore, to improve the identification and reduce the noise in them, background removal, region of interest segmentation is performed [33]. Following the research objective to identify a reliable model to identify and re-identify a vehicle for multi-surveillance, a unique representation of the vehicle needs to be identified, and this unique representation is queried and re-identified. This can be performed by classifying a vehicle as make and model and matching the frontal images using key features.

For a unique representation of the vehicle which once detected, the region of interest needs to be cropped for matching. Deep learning models have proven to perform this accurately for the classes that the deep learning model is trained for. The recent literature in this domain solves the challenges of diversity in dataset with multiple large-scale datasets and large number of classes. Further enhancing security several datasets focus on the parts and frontal area of the car enabling more fine-grained classification. In addition, datasets are varied in terms of illumination, exposure, and even environment. This leads to an open research domain for researchers in this field. One such dataset is [25] which considers the vehicle ecosystem of the state of Qatar.

Of the latest in frontal image dataset is a large-scale fine-grained dataset, with diversity in scale from 103 classes [34]. The dataset was annotated for make, model and year of manufacture providing a hierarchical representation of the vehicle. High resolution images with high quality were presented. The dataset was trained on CNN based methods. Several baseline methods have been utilized for vehicle classification including large scale models like Resnet-50. Further baseline analysis with Alexnet, VGG16 and VGG 19 were performed each producing an accuracy above 85% being robust for classification [36]. Apart from CNN, based methods, traditional rule-based approaches are dominant in this field due to the popularity of the problem. Local and global cues were utilized for classification in several approaches. Structural and edge-based features were also a common pick. Further, machine learning was performed with these features to enhance classification. With the feature extraction techniques, edge-based feature extractors like HOG (Histogram of Gradients) and Harris corner detectors performed significantly well for detecting parts of the car like the logo, the grille and the headlights [35]. Robust feature detectors from key points like that of SIFT (Scale invariant Feature Transform) and SURF (Speeded up Robust Features) were employed in several state of art. Adding to these features, corner detectors and line detectors like Hessian matrix and DoG (Derivative of Gaussian) were implemented in [6] producing considerably higher accuracy for smaller number of classes.

With larger number of classes, they fail to produce similar accuracy as with Database 4, Compcars dataset with colored dataset of large classes of cars with more than 162 types of cars, where in the same experiment produced poor results. Adding to the techniques, a bag of features or bag of words approach was implemented with feature detectors for unsupervised clustering, producing a histogram of features for

matching [36]. A typical feature detector algorithm accompanies a matching technique like hamming distance, Euclidean distance, or cosine similarity for identifying similar vehicles for recognition and classification [8]. This is further used for re-identification.

Naïve bayes [37], SVM (Support Vector Machine) [38] , LBP (Local Binary Patterns) [38], and KNN classifier (K- Nearest Neighbor) [37] were common machine learning algorithms used for vehicle make and model classification. CNN (Convolutional Neural Networks) models, where features are automatically engineered are used for classification, were utilized for vehicle make and model classification. They involve transfer learning on prominent pretrained models like that of Alexnet, VGG (Visual Geometry Group) that consists of 16-19 layers of convolutional layers like that of Resnet, and Mobilenet [32], [33]. Adding to this modified CNN networks were introduced such as residual Squeezenet [39] which produced a higher rank-5 accuracy of 99.38. Segmentation was applied as a pre-processing step to remove the background. A compound scaling approach was employed on Efficient net pretrained on ImageNet for classification for the purpose of presenting an app for vehicle make and model classification. Unsupervised deep learning techniques such as auto-encoders were also utilized for this purpose [35]. With each model producing different features automatically generated through CNN based approaches or engineered through edge and geometrical descriptors, for a real time use case, the requirement identified for an efficient model is higher accuracy and faster inference.

Although deep learning models conform to classifying vehicles of types like that of the trained dataset. Challenging and diverse environments can contain vehicle types that are not from the dataset as new types of vehicles are being manufactured every year. Feature extraction and matching techniques without classification do serve the purpose

of re-identification but with an image reference, and for that metric learning is proposed in multiple literature.

Table 2.2. Latest Rei-identification techniques.

Reference	Year	Re-id Method	mAP%	rank 1%	CMC%
[40]	2021	SIFT+ORB-+HSV+S-T+HG	30.46	75.11	
[35]	2021	Double channel symmetric-Resnet-50	49.55	74.36	-
[41]	2021	transformer+ VIT-B/16	79.0	96.5	-
[42]	2021	IPAD (vehicleID)	74.7		87.7
[43]	2021	DPLM+DTL (veri-wild small)	26.7	74.5	-
[44]	2021	deep V (Veri-Wild small)	75.1	92.1	-
[45]	2021	Angular triplet loss + SoftMax (vehicellD)	83.4	77.3	-
[46]	2021	Baseline+WCVL (veri-776)	80.4	-	95.3

The current state of art is summarized in Table 2.2. A typical re-identification uses a trained model that given a query image identifies an object that is best match to the given object in a gallery, that is an image database . Typical techniques include feature matching, image search and image similarity algorithms. Of the latest, are deep learning models like deep metric learning that create a representation of the image and train the model to identify the same representation using positive images and negative images

related to the model. This is employed using loss functions like triplet loss. CNN based model, transformer-based models and attention-based methods are extensively studied for this approach. Table 2.2, shows the latest and most significant results in vehicle re-identification using images.

It is known that computationally complex unsupervised methods and multi-step approaches produce comparable accuracy, but latency in a real time scenario is a question as in [41], and [42]. Adding to this, the use of blockchain may further add to the delay [20]. Hence, a robust feature extraction and matching methodology is required to be identified that is fast and accurate. Vehicle identification can be further improved if segmentation is performed with region of interest extracted. In this work the focus is on the key identifying factor of the vehicle, the logo and bumper design with license plate which infers the make of the vehicles. This inference can be achieved through a trained classification model and image matching through handcrafted features.

Segmentation approaches are often used for removing the background and extracting the vehicle, later classifying the vehicle. In real time uses, the cropped images should be generated from the detected vehicle to localize the frontal part of the car; this requires an added step for vehicle detection which increases time complexity. Thus, required is a single step approach for vehicle make identification. License plate detection also adds up to the vehicle unique features which is further added to the identification system for re-identification for unique id tagging of a vehicle. Thus, a robust model that can detect the region of interest and classify, identifying each instance of the vehicle make is required. An instance segmentation approach for re-identification is an area of interest for vehicle make identification through segmentation and classification. This will be explored in this thesis accomplished by modifying an existing dataset through

polygonal annotations for instance segmentation.

Further, algorithmic approaches for make identification and re-identification are performed on the cropped region using its feature points and descriptors is compared with light weight CNN models for the purpose stating the relevant of both for this task. Thus, a robust, reliable, secure, and private surveillance enabled by a private permissioned blockchain can be achieved. In the following section, the methodology used for the approaches are detailed to evaluate the proposed framework.

CHAPTER 3: METHODOLOGY

To evaluate each pipeline in this system and identify the optimum model for the purpose of a multi-surveillance enabled by blockchain. The pipeline of the methodology is as shown in the 3.1. This chapter follows the methodology used to accomplish the tasks in each section of the pipeline, vehicle region of interest segmentation is performed using CNN based methods explained in 3.1. Section 3.2, details the methodology used to identify the most efficient method for vehicle representation on the chain. In 3.3, the setup of the blockchain framework and the algorithmic approach used for vehicle re-identification through the chain is detailed.

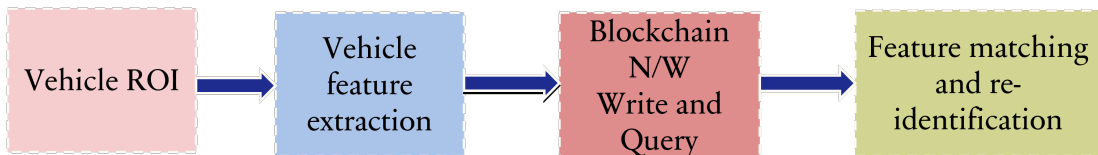


Figure 3.1. Base pipeline of blockchain enabled vehicle surveillance.

3.1 Vehicle ROI

Convolutional neural networks have been the key stone of computer vision applications. They are the most used types of artificial neural networks. Convolutional operations applied to neural networks enable better feature extraction and classification. Convolutional neural networks have evolved based on the requirements of accuracy, generalization, and optimization problems. Requirement of generalization and domain adaptation, lead to rise of several large-scale models trained on large scale data are present. Large scale data is trained on these networks which can be further adapted to other applications. Examples of convolutional neural networks being Alex net, Lenet, Resnet, Google-net, Squeeze-net and so on. In this paper, we utilize Resnet which is a deep residual network consisting of multiple CNN layers. It extracts deep features

and with its residual skip connections, the network is efficient in solving the vanishing gradient descent problem. The following section briefs on the core of the techniques used for instance segmentation.

3.1.1 Convolutional Neural Networks

Convolutional neural networks comprise of four key features which include weight sharing, local connection, pooling and a large number of layers. The layers include the convolutional layer that perform the convolutional operation on small local patches of the input where a given input x with a filter f will produce a feature map of x . The convolution operation for the whole image is computed by Equation 3.1

$$Y_n = \sum_{k=0}^{N-1} (x_k)(f(n - k)) \quad (3.1)$$

where x , f , and N are the input image, filter, and the number of elements in x respectively. The output vector is represented by Y_n .

This is followed by activation function such as tanh, sigmoid and ReLU. The activation functions introduce non-linearity into the network. The subsampling layers that are the pooling layers reduce the feature map resolution leading to reduced complexity and neural network parameters. The extracted features are mapped to the labels in the fully connected layer. All the neurons are transformed into 1D format. The output of convolutional and sampling layers is mapped to each of the neurons producing a fully connected layer. The fully connected layer is spatially aware extracting locational features as well as producing high level complex features. The result of this is linked to the output layer which produces output using a thresholding process. A final dense layer is sometimes used having same number of neurons as classes in case of a multi-class

classification. A SoftMax activation function maps all the dense layer outputs to a vector producing a probability of each class [47]. Accuracy of this prediction is measured by its loss function where the result is compared to that of the ground truth or labelled data. A common loss function used is the categorical cross entropy loss where the following equation describes the loss as L.

$$L = -\sum_{(i=1)}^N y_i \cdot \log(\hat{y}_i) \quad (3.2)$$

As seen in 3.2, y_i is the target prediction, which is the probability that class i occurs, \hat{y}_i is the output prediction or the i th scalar value in the output, N is the output size that is the number of classes to be classified or the number of scalar values in the model output. The minus signifies that the loss gets smaller when distribution comes closer to each other.

This setup is trained through a back-propagation technique. Hyper-parameters such as learning rate, regularization and momentum parameters are set before training process and adjusted according to greedy search. Evolutionary algorithms are further used to automate hyper-parameter tuning such as in [48]. During the back propagation, the biases and weights are updated based on the loss or error rate. The loss function L as in Equation 3.2 is required to be minimum to produce an accurate model. For this purpose, parameters such as kernel (filters), and biases are optimized to achieve the minimum loss. The weights and biases are updated in each network and feed-forward process is iterated with the updated weights. The model converges at the least loss.

Deep residual networks are utilized as the backbone for the framework used in this thesis. Deep residual networks are large networks with skip connections that carry knowledge [49]. In the context of this problem, Instance segmentation is performed

using CNN. Instance segmentation performs detection and delineation of each object in each image or video [50]. Each instance of an object is tagged with an ID enabling unique detection of every object in the scene. Instance segmentation is performed in different stages which include object detection, segmentation, and classification. This is enabled by CNN models as backbones and feature networks with classification heads. Several backbones are proposed for this approach. In this paper, we implement Mask RCNN with a Resnet backbone and Feature pyramid network. The use of this network is justified for its accuracy in object detection and segmentation where pretrained networks trained on several large datasets have superior performance over other models. However, complexity of the model causes the inference time to increase and providing an evident trade-off. We further measure the trade-off of the accuracy vs the time enabling evaluation of a real time use case.

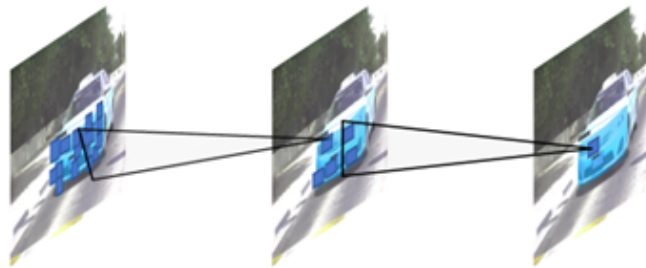


Figure 3.2. Deformed offset filter applied during convolutional operation.

3.1.2 Deformable convolution neural network

With all its advantages of convolutional neural network, the geometric structures of its building modules are fixed. Augmentation is used for transforming the images as a pre-processing step in most convolutional neural networks. Thus, these transformations such as rotation and orientation are fixed by modifying the training data. The structure of the filters in the kernel are also a fixed rectangular window. Pooling mechanisms

produce the same size of the kernels to reduce spatial resolution and thus the objects in the same receptive field are convolved and presented to the activation function, thus only identifying objects in that scale. Deformable convolution enhances geometric transformation and scaling by introducing a 2D offset to the grid sampling locations and thereby the convolution operation offsets from its fixed receptive location to a deformed receptive field. Adding the offset thus augments the spatial sampling locations automatically. The offsets are added after the convolutional operation [51].

3.1.3 FPN and Mask R-CNN

Further to enhance detection at lower levels, image pyramids are computed building a feature pyramid network (FPN). The object or segmentation area is scaled over different position levels in the pyramid. A proportionally sized feature maps at multiple levels are generated from a single input. Cross scale correlation is generated at each block to generate a fusion of these features. FPNs are used with CNNs as a generic solution for building feature maps. A bottom-up approach or top-down approach is used to produce a feature map. In terms of deep residual networks, the feature activation outputs are produced at each stages' last residual block. Figure 6 is a generalized setup of mask RCNN with FPN **7a**.

Mask R-CNN is a region-based CNN that performs object detection and classification with mask generation. The object detection is performed on a region of interest and evaluation was based on this region of interest. A multi-task loss is sampled on the region of interest as the total of classification loss, object detection loss that is the bounding box loss and mask loss.

Complex hierarchical features are extracted from images. Regularization tech-

niques are required to improve overfitting to the dataset. Augmentation techniques are often applied to reduce this overfitting, that includes image transformation such as scaling, translation, rotation and random flipping. It not only increases the data size but also provides a diversity of representation. The augmentation techniques can be divided into pixel level data augmentation, region-based augmentation, and geometric data augmentation. Pixel based augmentation techniques include changes in pixel values[52]. Adding contrast, brightness, or color changes the pixel intensity of the image. Regional augmentation includes that of creating masks of the required region. Motion blur and cutout are common techniques used for region-based augmentation. Geometric transformations are also applied to the data that include flipping, reflection, rotation, cropping etc. In this work the data is setup to augment at different levels that include geometric transformation and region-based transformation. This not only enhances the dataset but also improves the diversity of the same. One approach used in this model is mosaic tiling, where different training images, in this case 4, are taken in different context and stitched into one image creating a mosaic tiled appearance.

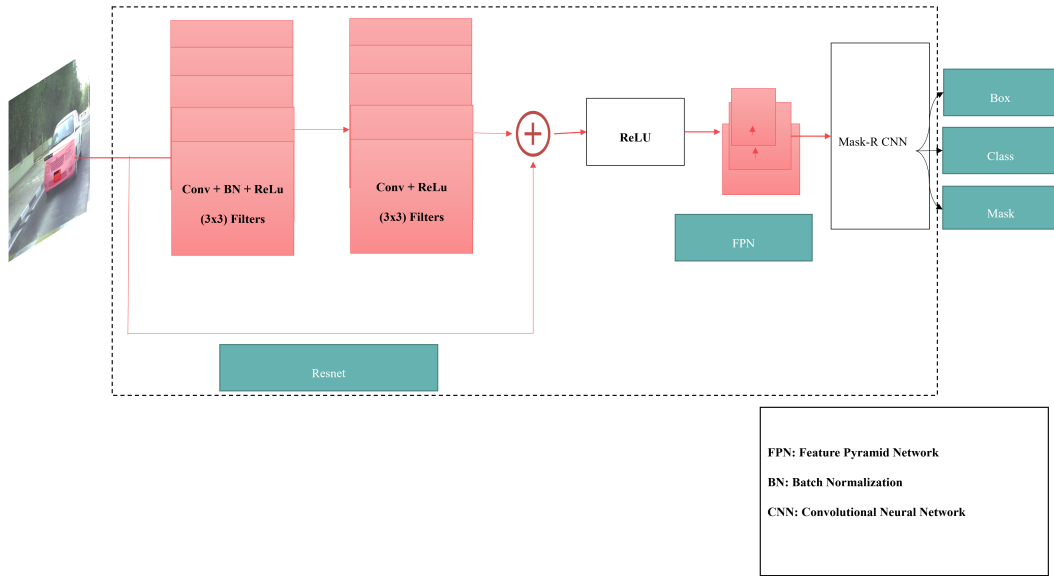


Figure 3.3. Instance segmentation model(mask RCNN + R50 + FPN).



Figure 3.4. Mosaic tiled image.

3.1.4 Experimental Setup

The setup of this network involves three layers. The vehicle with the mask is fed as training data. The data is augmented in three formats separately based on geometric augmentation and pixel-based augmentation. The transformed data is taken as the testing data and is then trained on a Mask RCNN-FPN network. Further, experiment was performed on Mask RCNN- FPN by deforming the convolutional layers. Resnet-101 and Resnet-50 are used as feature extractor backbones for performing baseline assessment on the dataset. The data is further scaled and feature representation extracted at each scale. The setup is as shown in ???. Instance segmentation model. The data is the key to any good model, the following section will detail how the dataset was modified for instance segmentation through new annotations and augmentations used.

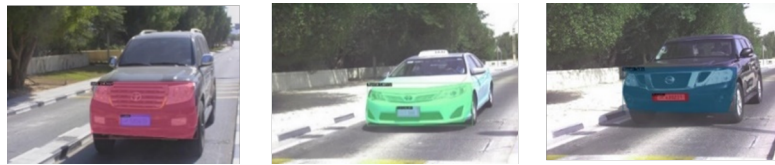


Figure 3.5. Mosaic tiled image.

3.1.4.1 Dataset. Existing dataset was modified for instance segmentation by creating polygonal bounding boxes of the frontal part of the vehicle to capture not just the frontal part of the vehicle but also the curvature of the vehicle as shown in Figure 8. The dataset contains 12 makes of vehicles taken in difference variations of camera exposure during extremely sunny weather to that of evening sunset. Distribution of the dataset is illustrated in Figure 9. The dataset is slightly imbalanced and so augmentation was performed to improve the data count. In addition, license plate is treated as a single class having a rectangular bounding box. A total of 225 images were

split for training, testing and validation with the 157 images for training, 44 images for validation and 24 images for testing with a 70-20-10 for the original format. The classes are very imbalanced and require further augmentation. The image below displays class distribution of the dataset. This dataset contains vehicles that belong to the middle east region specifically Qatar.

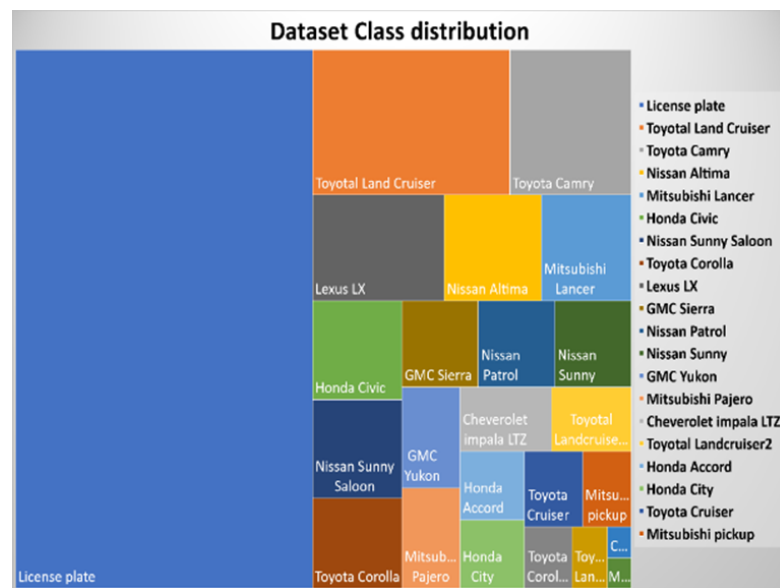


Figure 3.6. Dataset Distribution per class.

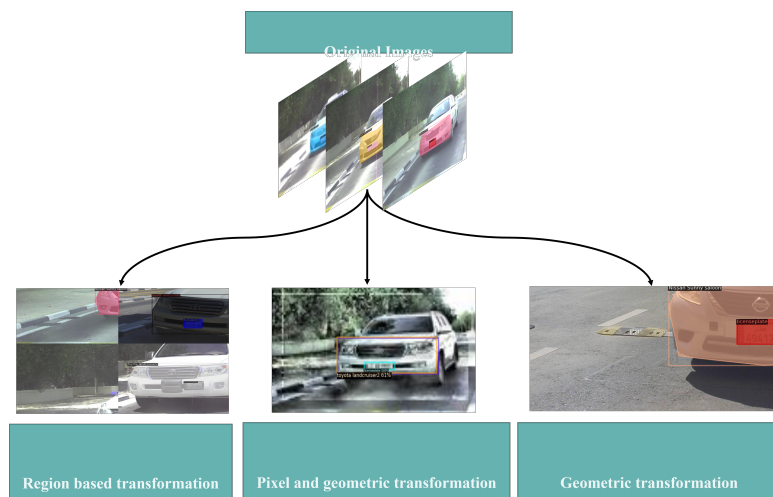


Figure 3.7. Data Augmentation techniques utilized to balance the dataset.

The experiments were conducted by augmenting the dataset to mimic different

camera orientations and noise parameters. An evaluation of both original dataset and partly augmented dataset was performed. Augmentation parameters included in pixel and geometric based include exposure and resizing with auto-orientation, noise, and rotation. In addition, patch-based augmentation which is geometric augmentation was performed where patches of the image were cut out with a certain given size. The third type of augmentation was mosaic tiled approach mimicking a mosaic tiled image [29]. Further three types of augmentation were performed on the dataset as illustrated in 3.7 .

3.1.4.2 Performance Metrics. To calculate the average accuracy, precision and recall must be computed for each image. TP (True positive), FP (False positive), FN (False negative) and TN (True negative) are metrics used for precision and recall.

Accuracy is a ratio of all the true prediction to the total predictions showing the correctness of classification into its respective make as shown in Equation 3.3

$$Accuracy = \frac{Correct\ predictions}{Total\ predictions} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.3)$$

mAP: Average Precision per class Average precision (AP) measures how well the model classifies each class, while mean average precision(mAP) measures how well the model classifies for all the given test dataset. It is a measure of accuracy of identification. It evaluates the performance of the model by averaging the precision values under the IoU (intersection over union) with a threshold of 0.50 to 0.95. AP is calculated in each point with the given threshold [53].

Inference time: The inference time is measured by the time taken to classify and generate a mask for a single input. In the context of this approach, it is the time taken to classify and generate masks for a single frame of a video.

3.2 Vehicle Feature extraction and matching.

Once, the segmentation is performed an image of the frontal part of the vehicle is achieved. This alone can be saved on the blockchain. However, may be prone to privacy issues. Thus, the framework has proposed a feature comparison approach where only the unique feature set is stored on the chain. For this, examined here is feature extraction techniques which are light weight and accurate.

Evaluating existing global and local feature descriptors show that they are time consuming and variant towards scale, rotation, and contrast. Thus, there's a requirement to identify a feature extraction and matching strategy that is relevant for these applications.

The assumption in this stage is that the inference is performed on the blockchain node itself as the stored features are extracted and matched with a given query. A trained model is not required here proving to be light weight. To evaluate this framework a set of key points of the given image and its descriptors were identified and matched. The matching was performed using a Lowes ratio test filtering out the good matches and providing a threshold for scoring the matching. Further the labelled dataset was used to evaluate the matching score determining the accuracy of the model.

A typical methodology involves extracting the keypoints. A detection is succeeded by a description of a feature point which is invariant to illumination, translation, scale and rotation in plane producing a descriptor vector for each feature point. Feature descriptors encapsulate important information into numbers and serve as a kind of numerical "fingerprint" that distinguishes one feature from the other. The descriptors are matched to identify similar features. Common feature matching algorithms are brute-force matcher and FLANN (fast library for approximate nearest neighbors) . A

typical flow is shown in the following flow chart as shown in 3.8. Key point detectors and descriptors used for this approach are as follows:

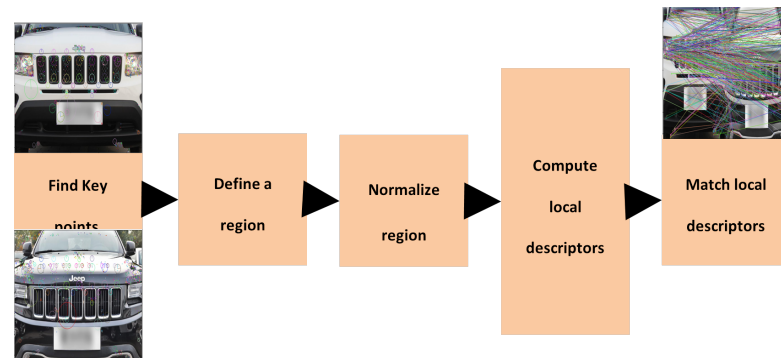


Figure 3.8. Local feature descriptor and matching.

3.2.1 SIFT

Scale invariant feature transform identified objects among clutter, and occlusion. It is invariant to scaling, orientation and illumination changes. Further, it is partially invariant to affine distortion thus suitable for image matching in different domains. A scale space peak selection is performed and then blurred using a gaussian blur operator which then generates a difference of Gaussian Kernel (DoG). Different octaves of the image are represented in the image. Gaussian blur is performed on different octaves of the image. A single pixel in an image is compared with its 8 neighbors as well as 9 pixels in consecutive scale as well as previous scale. A local extremum is considered as a potential key point.

A Taylor series expansion of the scale space is used to get a local extrema and intensity is deciding factor to exclude or include a keypoint. A Hessian matrix is used to compute the principal curvature. With the legitimate keypoints identified the scale orientation assignment is performed with a peak histogram of the neighborhood of the keypoints identified. The orientation is calculated using this assignment with keypoints

created with same location and scale with different directions. The local image region about each keypoint is computed which is highly distinctive and invariant. These features are used for matching between two images by their nearest neighbor. Ratio of the closet distance to the second closet distance should be less than 0.8. Figure 3.9 and 3.10, are the keypoints and features using SIFT local feature descriptor algorithm,

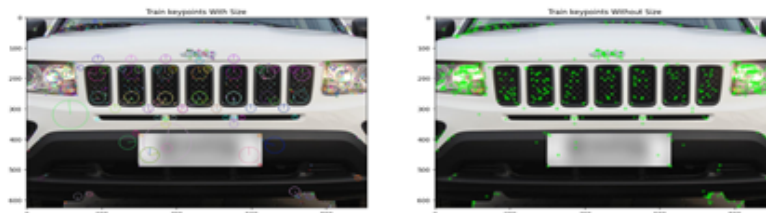


Figure 3.9. SIFT key points with size and without size.

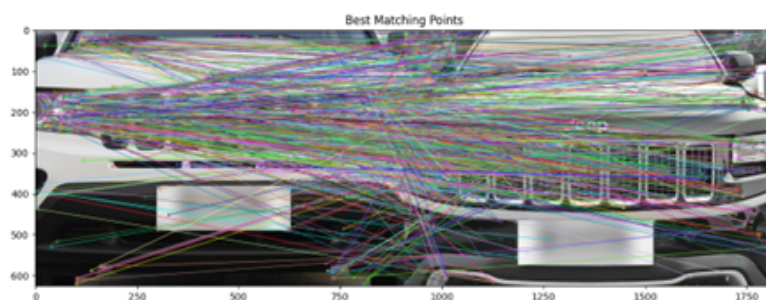


Figure 3.10. 874 Matching key points between two makes with label 11 from Compcars dataset.

3.2.2 ORB

Oriented FAST and Rotated BRIEF (ORB) builds on FAST keypoint detector and BRIEF descriptor. Fast and accurate orientation components are added to FAST with an efficient computation of oriented BRIEF features. The variance and correlation are analyzed of the oriented BRIEF features.

FAST (Features from Accelerated and Segments Test), which computes the brightness of a pixel p to surrounding pixels. A keypoint is chosen when 8 surrounding

pixel is darker or brighter than the given p . Thus, edges of the image can be identified. After the keypoints are identified, an orientation is assigned, and intensity centroid is detected. Corners intensity is assumed to be offset from its center and is attributed as its orientation. Figure 3.11 shows the keypoints extracted from a given vehicle.

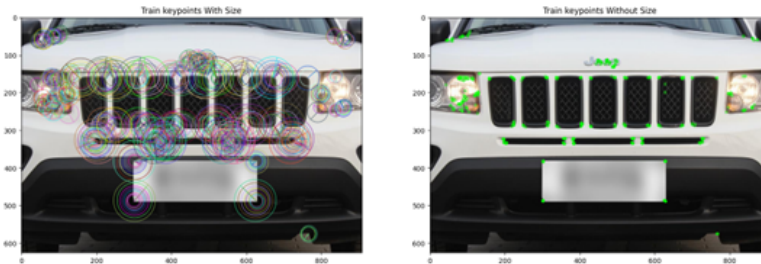


Figure 3.11. ORB feature points

3.2.3 BRIEF

BRIEF (binary robust independent elementary feature) are used to convert the keypoints to a binary feature vector that represents the object. The feature descriptor is typically 128-512 bits string. Patches near the neighborhood are smoothed and its sensitivity reduced which increases the stability of the descriptors. A value of 1 is assigned to the bit of the brighter pixel. Otherwise, it is assigned 0. BRIEF is rotation variant and thus ORB uses rBrief (Rotation-aware BRIEF). Figure 3.12, the matching features between two images.

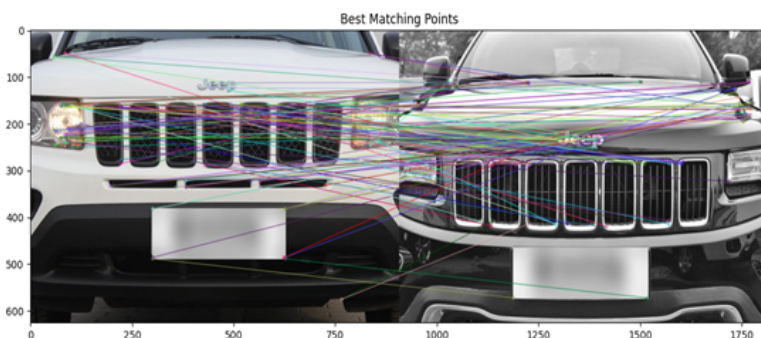


Figure 3.12. 107 matching features

Here, the Euclidian distance between the descriptors is used to match images. The Euclidian distance is calculated for each keypoint from a test picture between its descriptor and all descriptors of the training images, with the best match being correlated to the smallest distance. The barrier between the greatest match and the second-best match is another factor that should be considered, according to Lowe. The match is rejected if the ratio is not higher than the given threshold.

These feature descriptors were combined for the purpose of identifying the most appropriate of them for re-identification in terms of compute time and accuracy of detection. They were benchmarked on four state of art datasets described in the next section.

3.2.4 Experimental setup

An edge node detection framework is also proposed in this framework where the classification of the model is performed on the edge. For this purpose, light weight CNN models are experimented on in the context of inference on the edge. The models were trained on four different vehicle frontal image datasets. An ablation study was conducted to evaluate the effect of number of layers and datasets for the purpose of identifying a model that consumes less space and has low inference time. The datasets used are as follows. The data samples are as shown in 3.13.

3.2.4.1 Datasets. Four state of art datasets are utilized for the purpose of experimentation. Each dataset contains region of interest extracted images, two of them are grayscale and two colored. The following is the detailed description of each.

Database1 consists of 200 images of 25 categories with the most eight cars from same make and model. The resolution of the image is 140 by 70 pixels for all the images

72 images were used for testing and 120 images were used for training.

In Database2, the images were captured at a higher resolution of 150x66 and under various lighting and blurring circumstances. The database includes 8 manufactures and 17 models, some of which vary depending on the year of production. 154 photos were used for training, while 96 images were used for testing.

Database3, consists of 262 frontal vehicle images with 21 car make and model types are presented. 85 images are of 53 unusual vehicle classifications, which typically have one or two samples. The image resolution is that of 2592x1944 pixels and are colored images. Testing consists of 177 photographs and training consists of 85 images. In order to compare with existing literature, image was resized to 128x128.

Database4, consists of 1716 different car models and 163 different car manufactures. Vehicle attributes are extracted and divided into different parts as in headlights, front bumper etc. Of these, 3407 photos of automobile components that correspond to the frontal image are utilized for experimentation. In the experiments of the relevant work, 1374 and 1146 photos are utilized for training and testing, respectively.

All the datasets were modified for size to enhance training parameters as well data augmentation was performed for smaller datasets for better representation. Figure 3.13 are samples of the database.



Examples Databases	1	2	3	4	5
1					
2					
3					
4					

Figure 3.13. Dataset samples [54]

3.2.4.2 Evaluation Metrics. Evaluation of the model is performed using ranked accuracy and matching score. For a single detection, the time consumed is measured identifying the complexity of the approach. This enables availability and reliability measure for the whole framework.

Matching score:

A score is identified based on good matches between ORB features separate and feature combination approaches. Classification accuracy is determined by the true positive detections as image matching requires that the correct image and/or make be re-identified. The distance threshold is engineered to identify the best matching score based on the dataset.

3.3 Feature Clustering and classification.

Assuming the feature selection method has limitations that pertain to the image size and feature availability in terms of keypoints on distorted images, an image clustering and classification-based approach is experimented on to evaluate its performance and its choice as a reliable tool for vehicle identification. Three common and scale invariant

feature detectors are evaluated in this approach. A BoWs is created for the feature on the keypoints, and the strongest feature is identified by measuring the variance. 80 percent of the strongest features are taken from each category. The number of strongest features is limited to the number of features of the image with least feature quantity. This enables a balanced evaluation. Efficient and fair clustering is performed in this approach. The feature sets experimented on are SURF, KAZE and MSER. The following is a brief description of the methodology of each:

3.3.1 SURF

SURF include quicker interest point identification for real-time applications, a decrease in descriptor dimensions, and invariance to changes in geometrical shape, brightness, and size. The detector employs a further approximation of the DoG and is based on the hessian matrix. Additionally, integral image is utilized for this approach. Images that are integral are those in which the input picture at a certain point (x, y) comprises the total number of pixels to the left and above of (x,y) , inclusive.

A distribution of Haar wavelet responses relevant to the neighborhood serves as the descriptor. Only 64 dimensions are present, and the new indexing step is dependent on the Laplacian's sign. A keypoint hessian descriptor is used for the purpose of detection. Using integral images, the value of the Hessian Matrix's determinant is calculated for each individual pixel in constant time where the scale detection and interest point localization is performed. A box filter of Gaussian second order derivative was used for the purpose of identifying pixel intensities in the rectangular areas which are multiplied by the representative coefficient and then the total sum is computed. The existence of gaussian based derivatives improves the speed of SURF and thus speeded

up robust features. SURF is invariant to scale and illumination allowing it to be used for various applications that require fast feature description with high accuracy[55].

3.3.2 MSER

MSER extracts the covariant regions from an image, it is a stable connected component of the images. The methodology of feature extraction is based on identifying regions that do not change through multiple given thresholds. Given that the pixels identified below threshold are considered as white and others are black. A set of extremal regions are generated, which are a set of all connected components [56].

3.3.3 KAZE

KAZE is multiscale feature extractor from keypoints for nonlinear scale spaces which used a non-linear diffusion filtering contrary to gaussian filtering. The KAZE features reduce noise and keep the natural boundaries of the image. Nonlinear scale space is built using the Additive operator splitting techniques and variable conductance filtering [57].

3.3.4 Experimental Setup

To classify, the features are extracted for each image and a local feature set is created. A visual dictionary is generated using a k-means clustering algorithm which is represented by its histogram onto the dictionary using a hard assignment strategy.

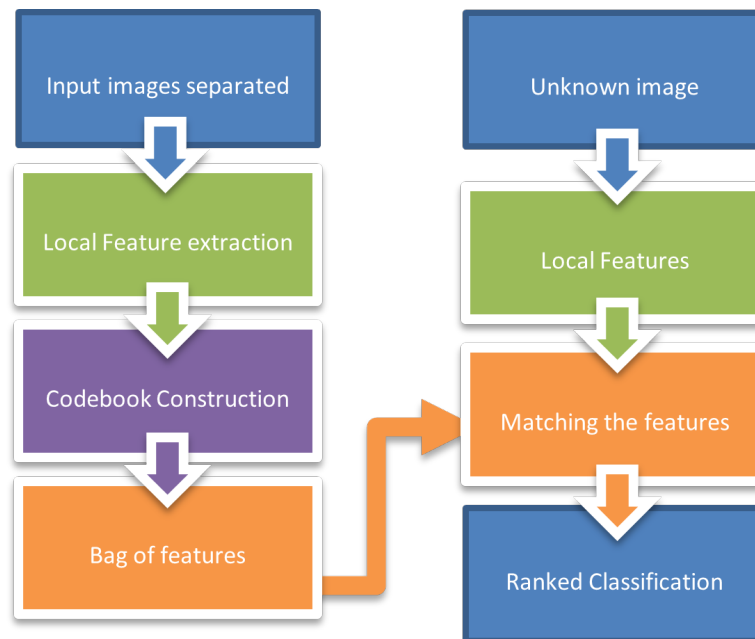


Figure 3.14. Bag of Visual Words Approach.

A Bag of words is created from the training set and indexed using k-means clustering, it is then compared with the query image to identify the most appropriate match for the query image. Evaluation is performed by identifying the matching score and the accuracy of the classification. The matching score is determined by matching the histogram of features with that of extracted features of the query image. Ranked accuracy is determined by identifying the best score for the matches. The methodology is described in Figure 3.14. The training parameters is as in Table 3.1. The smallest features were used for less computational complexity.

3.4 Light weight CNN classification

Deep learning is popular in its accuracy in computer vision tasks using convolutional neural network. However, accuracy is traded with latency and memory requirement. Nevertheless, there are light weight CNN models that can perform classification. The CNN layers are decreased and trained on the given four datasets which

are labelled for make classification. The number of layers is modified, and the most accurate model is identified, and the latency evaluated for real time light-weight deep learning setup. A common baseline, Mobilenet [58] is used for transfer learning with a pretrained model and is evaluated on for classification.

The setup of the CNN is evaluated on a simulated raspberry pi 3 device for inference. The following table indicates the architecture setup. Augmentation approaches are used to improve the imbalanced dataset. With learning rate set to 0.0005 and the data is cross validated to identify the best setup achieving higher accuracy and lesser latency with lower memory usage. Table 3.1 is the setup of each experiment conducted based on number of layers and dataset cross validation with the training parameters. The dataset used are the same as those used for vehicle feature extraction and matching.

3.5 Blockchain

As a cryptocurrency tool it removes the need for third party reducing cost and enhancing mutual trust between two parties. It keeps a ledger of all transaction in each peer-to-peer network thereby forming a decentralized secure ledger system. A blockchain system consists of a set of blocks that contain, data of transactions, a private key, a hash and a nonce secured with a public key. Each block is linked to the previous block which contains the hash value of preceding block and a nonce. Time of transaction is registered in each block. A block is added by validation through a consensus mechanism. A transaction in blockchain can have any kind of data stored, however with limited capacity based on type of blockchain platform used. The fact that blockchain hash cannot be modified or changed and that it is cryptographically secured, enables the property of security and trust [46] . In addition to that, smart contracts can be deployed

Table 3.1. Experimental Setup of CNN based model.

#	Layers	Splits	Epochs
C13	1	90-10	10
C14	1	90-10	100
C2	1	60-40	100
C3	1	80-20	10
C4	1	70-30	100
C5	1	60-40	100
C6	1	50-50	100
C10	1	50-50	10
C11	1	60-40	10
C12	1	70-30	10
C18	2	90-10	10
C7	2	80-20	10
C8	2	60-40	10
C9	2	50-50	10
C1	3	60-40	100
C15	Mobilenet	80-20	20
C16	Mobilenet	80-20	10

to automate, control access, and execute a contract or agreement. These features can be leveraged to protect content and activities in different applications. Surveillance application is modelled in this thesis for wide scale scalable use with blockchain.

Various platforms are available that are employed to develop decentralized apps that are capable of being used as a cryptocurrency transaction ledger but also for both sharing and storing info on the ledger. It is enabled on both private and public platforms, with and without authorization.

3.5.1 Smart contract

A crucial part of the blockchain network's automated processes is played by smart contracts. Smart contracts enable automation through the blockchain, which also boosts processing speed, lowers costs, and creates a non-repudiated network that ensures data integrity is protected. This makes it possible to enforce contracts and manage access, lowers risks and associated costs from third parties, and improves process efficiency [6]. This attribute enables applications like safe and secure data exchange, enforcement of automated contracts, etc. while also providing safety, security, and privacy for a variety of applications [59].

According to its design and the method of access provided for validation, a blockchain can be categorized into public and private, permissioned, and non-permissioned. The public can create and validate blocks on a public blockchain. Data modification is carried out using transactions. As a result, a framework for open access and transparency is created, raising concerns about privacy [60]. On the other hand, private blockchain is restricted, with only approved parties permitted to participate in the blockchain's activities. This can be used for computer vision use cases where it

is important for security and privacy to adhere to especially where unauthorized parties cannot view any transactions or activities that are performed during inference. A consortium blockchain is a blockchain where a predetermined group of nodes oversees the consensus procedure. A consortium or federated blockchain can be created by combining multiple organizations [61].

Hyperledger fabric is a consortium setup enabled blockchain which provides access control through prebuilt smart contracts called chaincodes. Further, chaincodes can be written to interact with this network. The endorser peer and orderer peer, of an organization in the fabric network, play huge role in validating and committing a transaction. Hyperledger Fabric can commit to the ledger without using proof-of-work or other kinds of trustless and hefty consensus, in contrast to permission-less blockchain platforms. It however, takes several steps to approve the transaction proposal before validating the actual transaction [11].

Each channel in the Hyperledger fabric has its own ledger. Furthermore, the main channel was eliminated in Hyperledger Fabric 2.0 in favor of making sure that each channel operated independently of the others. Additionally, each participating node must occupy a distinct place inside the network. As a result, it guarantees that the information in the channel ledger is only accessible to participating nodes. The following is the setup of the Hyperledger fabric for testing the scalability of the network to evaluate the availability of the blockchain network for the purpose of a real-time information transfer for surveillance purposes.

The nodes of the network are considered as operation centers of the Hyperledger fabric having a hierarchical configuration. There are 10 participating organizations in the consortium. Five of the organizations are in the first channel sharing a single ledger.

Node 5 is part of the second channel which constitutes organizations 5 to 10. Node 5 and 6 have their own channel. Evaluation of scalability and setup is performed to analyze the latency involved in a proliferation of surveillance systems through a large network. Figure 3.15 illustrates the organizational setup for the purpose of analyzing the feasibility of blockchain framework for information transfer for real time vehicle re-identification task. Raft consensus was used as consensus mechanism which produces a majority voting scheme to commit a transaction to the ledger.

3.5.2 Experimental setup

The performance of this setup was evaluated using Docker Stats performance measure where its process time and energy used were measured to identify the weight of the process on the edge nodes. This was setup, as each node is simulated on a docker container. Each node is an organization of the consortium, which is part of a traffic zone that captures video feed from sensors. Considering org1, org2, org3, org4 and org5 are part of one zone sharing the same ledger, with org5 being a super-node which also shares a ledger with org6. Nodes org6, org7, org8, org9, and org10 are part of the second zone. Modelled here is a surveillance network which connects several zones of surveillance area with their own ledger, with one node in each network considered as a super-node which shares a completely different ledger with super nodes of other zones, sharing sensitive and required information for target tracking or re-identification. Algorithm 1 is a proposed methodology for re-identification in real time.

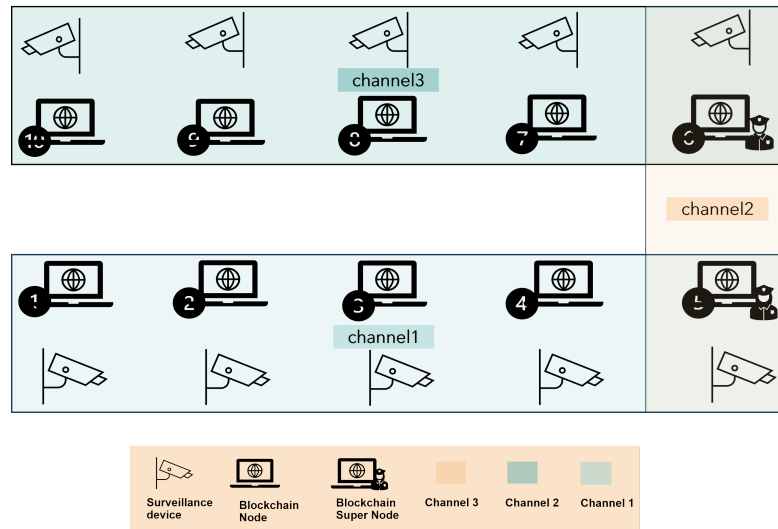


Figure 3.15. Blockchain network setup with two supernodes with access to two channels.

This communication is performed on a condition when an anomalous vehicle is not identified in the current zone and could have shifted to another zone. This, however, requires to be evaluated for its efficiency in terms of latency and energy that may be consumed during the process. Since the network is already established and secured by extensive consensus mechanisms and a transaction flow that is controlled by endorsing, committing, and validating peers in the network the throughput is guaranteed. We assume two cases in this network that a vehicle of interest is identified, then the vehicle region of interest is extracted using the instance segmentation model and the make and model of the vehicle is stored as text transaction data with vehicle license plate data in the ledger. The second setup is that the region of interest is cropped, converted to image, and stored in the ledger. The third setup is that make is identified by classification using light-weight CNN models and is stored with a vehicle license plate detection and recognition model performed in [48] which produced 98.26% accuracy of detection. The fourth setup is performed where the features are only shared on the ledger for a more private transaction.

The setup of the hyperledger fabric for evaluation contains a single orderer performing raft consensus. All peers take part in endorsement. Each organisation has a single peer and a single client for purpose of constraint evaluation of energy consumed. Peers of org1, org5, org6 and org7 are anchor peers which are peer nodes where all other peers can communicate with. Table 5 describes the peer roles in this network. The Batch time out is set at 2s. The Maximum message count is 10 in a batch and the absolute maximum bytes in that batch is set to 99MB considering the device constraints.

The execution time is measured for each transaction in the network scaling to a 10,000 transaction the most. The execution time can be defined by the total time the algorithm, or the process runs in the system, irrespective of its wait time or if its queued process. Hyperledger fabric was setup on an Ubuntu20.0 LTS OS, with dockerized containers for each node, peer and certificate authority based on the configuration. Smart contracts were written to test the cross-channel communication as well as multi-channel communication setup. A whole end to end framework testing was performed assuming simultaneous read, writing and query of transactions take place on the configured multi-zonal network. Experiments was performed using image as data which can be used by feature clustering model and CNN model. Key-points and feature descriptors were the other set of data stored on the chain for re-identification.

The following section states the experimental results on the methodology used. The segmentation model is trained and tested on the newly annotated dataset for instance segmentation. Four datasets are utilized to test the feature descriptors to identify the prominent feature description technique with low latency and high accuracy. Further, the blockchain network is simulated on Hyperledger fabric framework where access control is automated and evaluated in two formats where first the data is shared between

all organizations in the network and the second the data is shared in a hierarchical setup where two consortium frameworks relate to two super-nodes which has access to each consortium. All the defined setup is analyzed for accuracy and the real time performance is evaluated using latency analysis which determines the reliability and availability of the information for surveillance. With each setup evaluated the following is the experimental results elaborated in terms of accuracy of instance segmentation for cropping the region of interest and detection. Classification accuracy of the approaches were utilized for evaluating the make identification through image similarity matching and classification for re-identification. The blockchain setup is evaluated for single channel communication as well as multi-channel communication.

Algorithm 1 Blockchain enabled re-identification using key-point descriptors

```

1: Input : Scene

2:  $f \leftarrow \text{Img}[\text{Key-points, descriptors}]$  refer: Section 3.0.2  $t \leftarrow$  current time

3: function QUERY(Channel_name, f(D), F(t-10...t))

4:   match score  $\leftarrow$  Match (f(D), F(t-10...t)) refer: Section 3.0.2

5:   return match_score, node, time

6: end function

7: function WRITE(Channel_name, data)

8:   Write to Channel_name the data

9: end function

10: Assume vehicle is detected in channeln (D)

11:  $D \leftarrow 1$ 

12:  $\text{key1} \leftarrow \text{kp1}(D), \text{des1}(D)$ 

13: while D do

14:   QUERY(channeln, f(D), F(t-10...t))

15:   if Match score > threshold then

16:     nodef  $\leftarrow$  node,time

17:     WRITE((channel3, nodef)) match score < threshold

18:     WRITE((Channel3, key))

19:     WRITE((Channeln+1, key))

20:     QUERY(channeln+1, (f(D), F(t-10...t))

21:     nodef  $\leftarrow$  node, time

22:     WRITE((channel3, nodef))

23:      $n \leftarrow n+1$ 

24:   end if

```

CHAPTER 4: EXPERIMENTAL RESULT

4.1 Vehicle frontal image segmentation and detection

To evaluate the setup Several experiments were conducted on different augmentation methods on the dataset. Resnet-50 backbone was used for the deformable receptive field-based Mask RCNN. With a batch size of 2, the experiments ran for 1000 iterations and used a pretrained Resnet backbone on COCO dataset. Evaluation was performed using the COCO trainer module. Table 6 is the classification accuracy and segmentation accuracy achieved with the time taken for inference on one image. Table ?? is the average precision in terms of the different thresholds set for IoU.

For a varied analysis, different baselines were experimented on for the purpose of evaluation and identifying the trade-off in reliability and accuracy while performing

Table 4.1. Classification accuracy and detection accuracy using mAP with latency

Model	Lr threshold	fast_rcnn/cls_acc.	mAP	Time
Mask RCNN+R-50 +FPN	3x	0.992	98.772	136 ms
Mask RCNN+R-101	3x	0.996	99.670	310 ms
Mask RCNN+R-50	1x	0.992	99.670	316 ms
Mask RCNN +R-50+FPN (DCONV)	1x	0.984375	90.747	161.81 ms

Table 4.2. Ablation study with different backbones and deformable convolution

Model	Backbone	AP	AP50	AP75
Mask RCNN-DCONV	RESNET-50 + FPN	79.648	96.337	94.350
Mask RCNN-DCONV	RESNET-50 + FPN	74.185	90.747	89.121
Mask RCNN	RESNET-50 + FPN	80.213	98.772	95.950
Mask RCNN	RESNET-101	73.621	88.219	86.265
Mask RCNN	RESNET-50	80.206	99.670	98.730

instance segmentation for the purpose of vehicle recognition. Mask RCNN was used as baseline with a Resnet-50 backbone with Feature pyramid network further modelled with a Resnet-101 backbone with Feature Pyramid Network. The original dataset was augmented in multiple methods to improve the dataset description. The results of the experimentation with original dataset are displayed in Table 4.1. The table describes the classification accuracy of mask RCNN with that of instance segmentation accuracy with the mean average precision metric. The execution time for inference of a single image from the test set is also presented.

The test data is either over-represented or under-represented and thus needs to be balanced for a reliable result. Thus, multiple augmentation techniques are performed to improve data representation. Three types of augmentation approaches are utilized for this task. A large network and smaller network were tested to evaluate the impact of augmentation on data size and the accuracy of the model. The Table 4.3, describes the results of each augmentation type on baseline models ad 4.2, the backbone based ablation study. The inference from the table is clear that mosaic augmentation performs considerably better than any other augmentation type. However, it fails to surpass images with same resolution. The patch-based augmentation has very low inference than expected even though the number of images increases. This could be because of class empty patches in the dataset as each class is represented once in the original image.

4.2 Vehicle ROI and feature descriptor.

Four datasets of varied classes and image distributions are employed for re-identification through image matching. The image matching is performed using key point descriptors mentioned in the methodology section. The following table evaluates

Table 4.3. Ablation study based on data augmentation

Augmentation	Splits#	Model	Backbone	AP	AP50	AP75
(140x70)+exp.+ rot.	471-44-24	MaskRCNN-	R-50 + FPN	65.748	81.708	77.517
		D CONV				
		MaskRCNN	R-101	70.989	88.633	85.148
Full Augmentation	460-44-24	MaskRCNN	R-50	59.502	85.189	67.677
		MaskRCNN-	R-50 + FPN	66.780	83.101	75.029
		D CONV				
Patch input	628-176-96	MaskRCNN	R-101	49.585	66.776	58.586
		MaskRCNN	R-50	60.163	77.906	73.954
		MaskRCNN-	R-50 + FPN	52.475	74.535	64.246
Mosaic Based	471-44-24	D CONV				
		MaskRCNN	R-101	71.569	88.176	84.842
		MaskRCNN	R-50	52.186	74.393	59.095
Mosaic Based	471-44-24	MaskRCNN-	R-50 + FPN	87.698	99.406	98.900
		D CONV				
		MaskRCNN	R-101	83.933	99.568	99.103
		MaskRCNN	R-50	82.463	99.637	98.121

each image matching feature descriptor for its classification accuracy, true positive rate, false positive rate of identification. Table 4.4 are the experimental results of the feature matching algorithms, the time taken was measured for each indicating the complexity of the approaches. Matching based on Euclidean distance was used with two as k value for a fair comparison. The same parameters were used for all the datasets. For Dataset2 the whole image set did not produce feature points for orb detector, and thus evaluation was performed for images that produced those features. True positive rate was used to identify detection accuracy which is most relevant for finding a matching vehicle.

Table 4.4. Feature Matching

Database	Feature Detectors	Success rate (%)	Time taken/image(ms)
Database1	ORB	82.3	39.6
	ORB(KP)+BRIEF(desc)	75.2	25.3
	SIFT (KP)+ BRIEF(Desc)	81.6	48.2
Database2	ORB (KP,DESC)	79.44	3.89
	ORB(KP)+BRIEF (desc)/BRIEF	80.627	2.67
	SIFT (KP)+ BRIEF(Desc)	87.023	8.37
Database3	ORB (KP, DESC)	79.67	148
	ORB(KP)+BRIEF (desc)/BRIEF	85.70	492
	SIFT(KP)+ BRIEF(Desc)	85.65	117.9
Database4	ORB (KP, DESC)	-	-
	ORB(KP)+BRIEF (desc)/BRIEF	-	-
	SIFT(KP)+ BRIEF(Desc)	-	-

Classification accuracy was separately measured identifying true positive rate and true negative rate.

4.3 Feature clustering and classification approach

The bag of feature approach was evaluated on the datasets mentioned in chapter 3. Ranked results were measured with matching score and accuracy of matching or classification. The Table 4.5, are experimental results of each database mentioned in chapter3. Top-1 produced high accuracy and so further experimentation was not performed. Matching score was also measured with average time taken for the whole dataset.

Table 4.5. Feature Clustering and matching

Database	Feature Detectors	Top1(%)^o-score/match Acc.	Time taken in msec
Database1	SURF + Kmeans	99.96 (100%)	282.3
	MSERF + Kmeans	99.96 (100%)	106.41
	KAZE + Kmeans	99.96 (100%)	255.5
Database2	SURF + Kmeans	100 (100%)	394.8
	MSER + Kmeans	100 (100%)	102.69
	KAZE + Kmeans	99.26 (100%)	382.13
Database3	SURF + Kmeans	99.3 (100%)	822.2
	MSER + Kmeans	100 (100%)	102.69
	KAZE + Kmeans	96.9 (100%)	614
Database4	SURF + Kmeans	97.2 (100%)	163.001
	MSER + Kmeans	98.8 (100%)	482.2
	KAZE + Kmeans	92.65 (100%)	576.95

4.4 Light weight CNN based approach.

Reducing the number of layers of the CNN network can considerably improve the inference time. For the objective to identify a lightweight model for vehicle classification. An ablation study was performed to identify models with lesser convolutional layers. The setup as described in Table 3.1 was implemented and the results are as shown in Appendix . The most significant results are presented in Table 16. Database 3 and Database 4 did not perform well in the existing setup. As the aim is to deploy on a raspberry pi3 sensor network, the ram usage was also measured.

Table 4.6. CNN based approaches

#	Dataset	Accuracy (%)	Loss	Time(ms)	RAM usage(k)
C3	Database1	90.6	0.43	1	252.0
C14	Database2	81.8	0.55	1187	767.5
C1	Databse3	57.1	1.58	78	386
Mobilenet	Database3	55.4	3.17	9	733.6
Mobilenet	Database4	38.9	4.84	5	670.3

4.5 Blockchain framework

With Hyperledger fabric network, the following configuration was setup to evaluate the performance in terms of a single channel approach and multi-channel with cross-channel communication. The setup evaluated for maximum scalability for reading and writing on to the blockchain. Appendix 3 is the overall running statistics of the Hyperledger setup for 5 organization network all part of single channel. The NET input, block input and output with memory and CPU usage were measured using the docker status report. The same parameters were evaluated for the multichannel with cross channel communication for reading and writing large transaction through super-nodes such as a surveillance approach where multiple events may occur in an existing scene. The detail graphs are presented in Appendix . The latency is evaluated for the whole transaction process. Further Algorithm 1, is the protocol used for vehicle-based surveillance, where a complete transaction occurs where reading and writing is performed between three channels. Image based blockchain enabled surveillance results are shown in Table 4.7, where the image is encoded using base64 encoding and stored as single entity on

Table 4.7. Evaluating an image stored on the ledger.

Image parameters	Value
Image size	8.18kb
Encoded	10.90kb
height X width	128x128
Time in sec (Single read)	0.165s
Time in sec (multi-channel)	0.920s
Time in sec (multi-channel)	1543.855s
10,000 transactions	

the chain. An image of ‘Audi.png’ from database 3 was used for testing which had a cropped frontal image of an AUDI vehicle of size 128x128px.

The following Figures 4.1-4.7 are resource consumption graphs of the hyperledger fabric based on the docker container statistics. Docker statistics are charted for memory consumption, CPU usage and block input and output. In this thesis, resource consumption is taken into consideration for energy usage, execution time and cost analysis. The figures charted are represented by line graphs where each line indicates the peak usage of CPU and memory at that particular time of protocol usage. In this context, the each line is a docker container consisting of hyperledger fabric nodes, peers in each organization of the network. Further, the chaincodes, cli and the orderer are also containerized in hyperledger fabric. The activity level of each is color coded and represented as a line graph.

Figure 4.1 and Figure 4.2 are CPU resource usage statistics for a single channel network setup. They indicate that the peer0.org5 in orange line is actively using its

CPU resources to commit and validate transaction. Figure 21 is the CPU resource consumption of the multi-channel network setup reading querying and writing data simultaneously. The green line in this figure corresponds to the chaincode for cross-channel transaction initiated by peer0.org5. The active green line indicates the chaincode is processed as well as all other peers are participating in the transaction process. The peak CPU usage was 14% for writing a transaction and 12% for reading from the chain. The peak CPU usage for a multichannel network was 12% consumed by both peer0.org5 and peer0.org6 both the anchor peers of channel3, the consortium channel.

Figure 4.4, represents the graph during channel-to-channel data transfer between org5 and org6. The orange line for peer0.org5 and the golden line for peer0.org6 with the chain-code as green line. The peak CPU consumption is 15% for peer0.org5. Peer0.org6 uses 6% of CPU resources at its peak. The orderer is committing the transaction as the purple line but do not consume as much CPU processing power as org5 which is an anchor peer as well as endorsing peer and is part of the consortium channel.

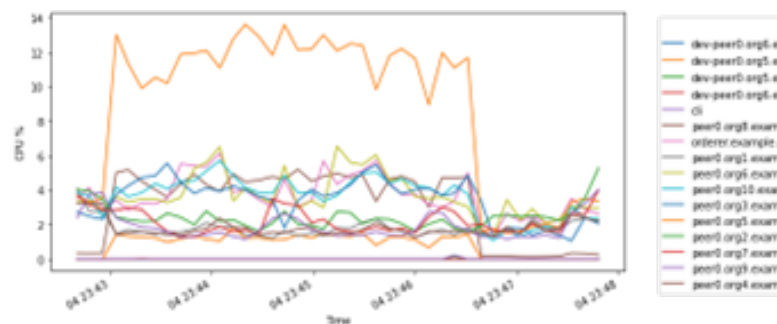


Figure 4.1. Writing 10,000 transactions on single channel network.

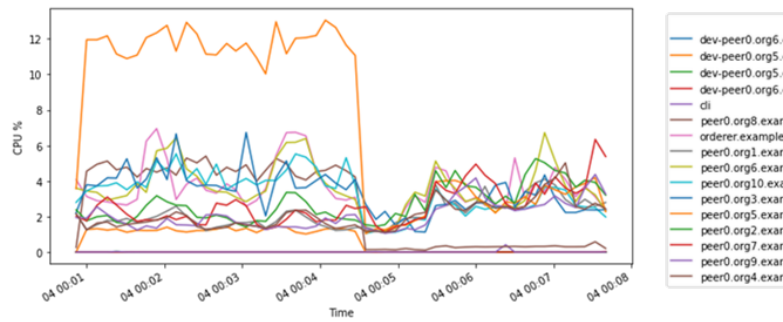


Figure 4.2. Reading 10,000 transaction on single channel



Figure 4.3. Querying, reading, and writing multichannel.

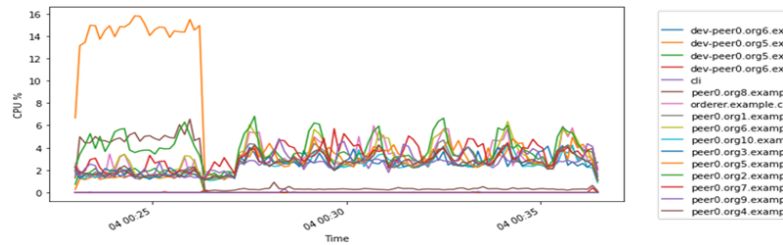


Figure 4.4. Read and writing cross channel (reading from channel 1 and writing to channel2)

Memory usage of the node which is encapsulated in the docker container for a single channel network is graphed in Figure 4.5 and Figure 4.6. Peer0,org8 seems to peak in memory usage which is an anomaly and could be due to other processes in the system. However, it is evident that single channel approach shown in Figure 4.5 and 4.6, has a maximum usage of around 80- 140MB and the peers in channels 2 and 3 are below 80. Further for the multi-channel approach, Figure 4.8 depicts an active process which

involves reading, query and writing through multiple channels, mimicking multiple surveillance zones with restricted access. The memory consumption is steady at 100 MB for the cross-channel peer (dev.peer0.org5) and all other participating peers are at range 45 -100MB in all channels. The active lines indicates that the chaincode is actively used for cross channel transfer. Figure 5.2 identifies the bottleneck zone the cross-channel data transfer nodes that is channel 1 to channel 2 from peer0.org5 in channel1 to peer0.org5 in channel2 ledger. The memory consumption is steady at 80. In the following section, the key findings from the results are discussed and presented with an analysis of each module used for experimentation in this chapter. The relevance of the techniques will be justified by how each phase contributes to the reliability, privacy and integrity of vehicle surveillance using a private permissioned blockchain network.

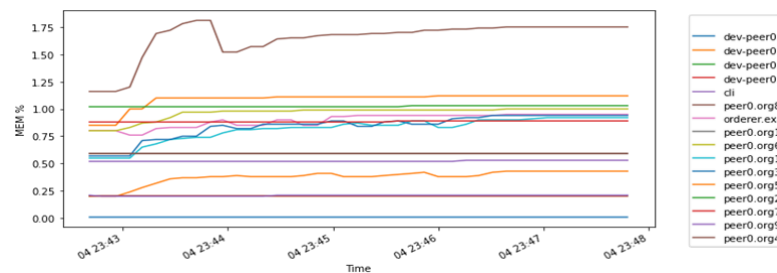


Figure 4.5. Memory usage for 10,000 writes from a single node to a single channel

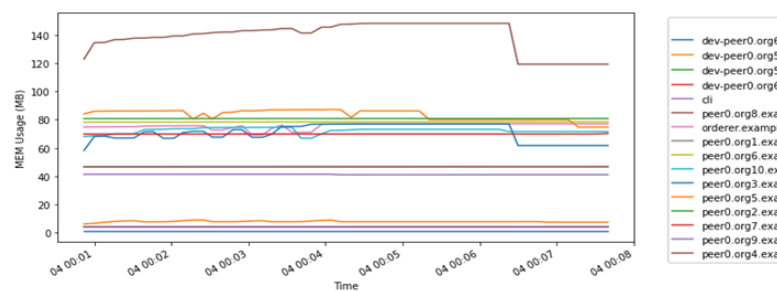


Figure 4.6. Memory usage for querying on a single node to single channel.

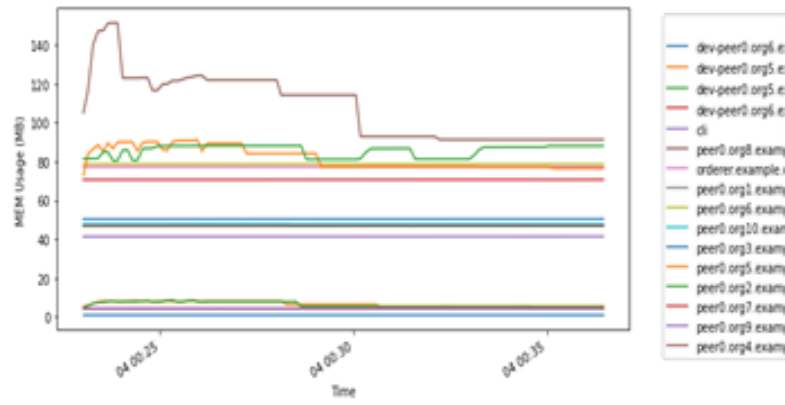


Figure 4.7. Reading/Querying and writing cross channel.

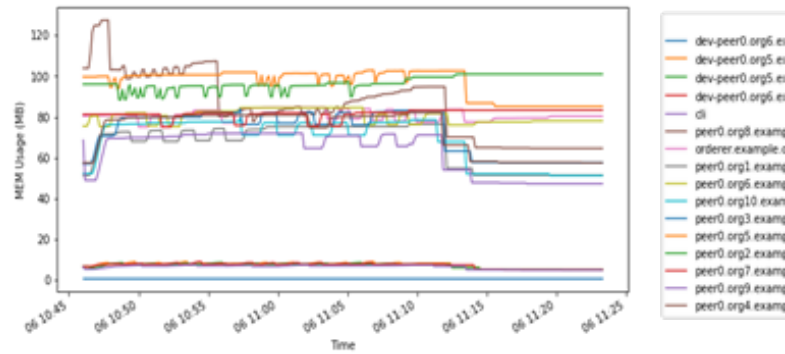


Figure 4.8. Querying and writing multi-channel approach.

CHAPTER 5: DISCUSSION

With the proposed and evaluated framework, this section discusses the parameters that validate or invalidate blockchain enabled surveillance. A private, secure, and reliable surveillance can be achieved through blockchain with some cost. The privacy aspect is accomplished using reliable algorithms that segment the region of interest and the keypoints and descriptors that are shared in the decentralized blockchain ledger that is accessible to restricted parties through access control schemes. A private permissioned ledger can be used for the purpose of security and further the privacy of the framework. To quantify these parameters and elaborate the results, following are the key findings of the approaches used and its significance compared to state of art.

5.1 Vehicle instance segmentation

Mask RCNN is certainly an efficient and reliable method for not just region of interest segmentation but also make recognition with license plate detection. The model presents high accuracy of detection for each class except that of license plate detection which can be due to the varied placement of license plate on some of the vehicle models.

The resnet-50 back bone without FPN with base RCNN produces a high mAP of 99.670. Although Resnet-50 backbone with FPN is hypothesized to produce higher accuracy, it lags 1% but produces faster inference with 174ms faster than base-RCNN. With further experimentation on the CNN module with a deformed convolutional operation the accuracy dropped to 90% which is significantly lesser than expected. This could be due to the added complexity and generalization of the network. It can be noted that the models are inferred on a test set with imbalanced data and thus not reliable for certain classes. With class wise precision, the largest class, the license plate has the least

Table 5.1. Comparison with existing literature.

Method	Model	Classification accuracy
[25]	SIFT + DoG t	74.63%
Ours	MaskRCNN+ FPN + Resnet-50	99.2%

accuracy, license plate covers a smaller area and is similar in geometry to rectangular shapes which can be a reason for the poor performance that is scale variation may affect the performance such as license plate location in a LEXUS car which is at the side of the bumper of the car.

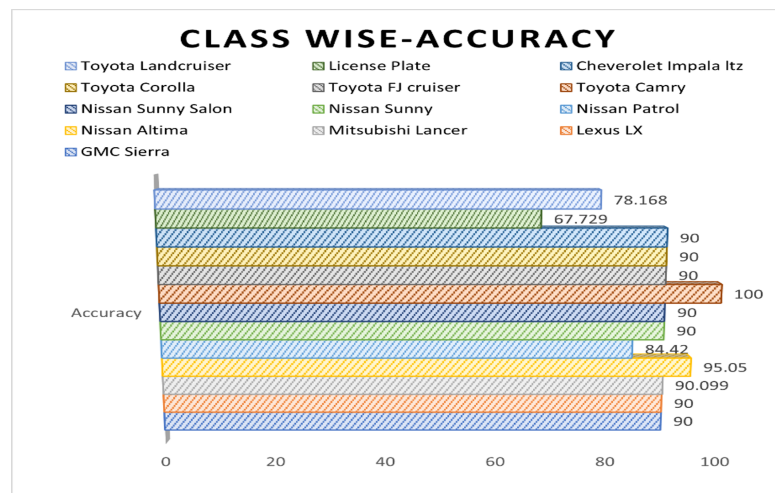


Figure 5.1. Class-wise accuracy.

Compared with existing literature, the existing study with the same dataset produced only 74.63% accuracy with traditional approaches. The classification module in maskRCNN with feature pyramid network and resnet-50 backbone produced higher accuracy, thus achieving higher accuracy of inference compared as seen in Table 5.1. However, this is a complex architecture compared to SIFT and DoG which takes higher time for inference.

Thus, a trade-off in accuracy and time complexity. Light weight modules without

residual architecture like that of Centermask [62] is required to be evaluated for the instance segmentation as future work. The vehicle and license plate classification results based on class are shown in Figure 5.1.

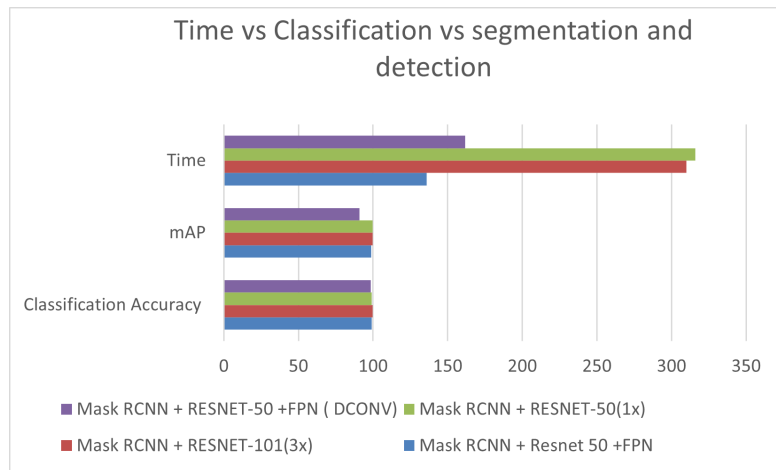


Figure 5.2. mAP vs Accuracy vs Inference time per model

The inference time is independent of the complexity of the model. However, larger learning rate took longer time for inference. It is noted that the same image was used for inference for each model. Mask RCNN with Resnet-50 backbone and feature pyramidal network inferred in the least time as illustrated in 5.2.

5.2 Vehicle make identification.

Vehicle make identification was performed using three methods. Feature matching, feature clustering and light weight CNN models. Dataset 1 and Dataset2 produced considerably better results.

5.2.1 Comparison of the results with state of art

Comparing the data with state of art, the dataset performed well using BoW approach and classification. Light weight models on the other hand produced 10% less

results than the traditional approach. Adding to that colored dataset, database 3 and database 4 produced poor results. Database 3 has a very imbalanced dataset as it was designed for common and uncommon car recognition and so the low accuracy. In case of database 4, large number of images underfit the models. Therefore, it requires a large model for improved performance. However, the trade-off would be large memory usage and high latency. Comparing with existing literature, the Table 5.2, Table 5.3, Table 5.4, and Table 5.5, are results with respect to best performing techniques for accuracy. In Database 1, the C3 with two-layer CNN trained for 100 iterations performed the best. However, Feature classification using bag of words approach produced perfect accuracy, where vehicles with same model were identified accurately. The downfall however would be that the model is trained on a specific dataset and the classification conforms to that datasets vehicle types only.

5.3 Blockchain network

The blockchain network was evaluated on different configurations that included a single channel communication network that mimics a private permissioned blockchain with all organizations having access to the same ledger. The single channel communication are measured in graphs per transaction in Figure 4.2, Figure 4.6, Figure 4.5, and Figure 4.1, show a stark increase in memory usage and CPU for that instance when the transactions are committed. The query organization and the write organization had expectedly increased use of resources. The query process took less time. However, the write process had considerable amount of time due to the validation. Further the lines indicate an active use of cross channel chaincode in dev.peer0.org5, the common node in channel 1 and channel2. The maximum memory consumption of any process was less

Table 5.2. Database1 accuracy comparison

Reference	Method	Classification accuracy
[37]	KNN	96
	Nearest Neighbor	78
[36]	SIFT + Multiscale Harris	93.78
Lightweight CNN	C3	90.6
KP + feature	ORB	82.3
Feature + Kmeans	All	100

Table 5.3. Database2 accuracy comparison

Reference	Method	Classification accuracy
[36]	SIFT + Multi-scale Hessian	98.96
Lightweight CNN	C2	81.8
KP + feature	SIFT + BRIEF	87.023
Feature + Kmeans	All	100

Table 5.4. Database3 accuracy comparison

Reference	Method	Classification accuracy
[36]	SIFT + DoG	98.87
Lightweight CNN	Mobilenet (TL)	55.4
KP, feature	SIFT + BRIEF	85.56
Feature + Kmeans	All	100

Table 5.5. Database4 accuracy comparison

Reference	Method	Classification accuracy
[36]	SIFT + Multi-scale Harris	49.48
	CNN	48.4
Lightweight CNN	Mobilenet	38.9
Feature + KNN	-	-
Feature + Kmeans	All	100%

than 140MB and the CPU usage was less than 14%. This indicates that cost of running a single node in a hyperledger is less in terms of CPU usage and memory consumption in the long run. However, storage requirement is more than 10GB which may add to the cost of the infrastructure.

In terms of the multi-channel approach where multi-zones communicate with each other through the super nodes have higher activity due to cross channel communication in org5 and org6. Scalability of the network through multi-zone approach can confirm that it requires high memory usage and CPU resources. There was also

significant difference in time for a multi-channel and single channel transaction with an image transfer in base 64 code had a difference of 755 ms. Thus, scalability introduces the cost of time.

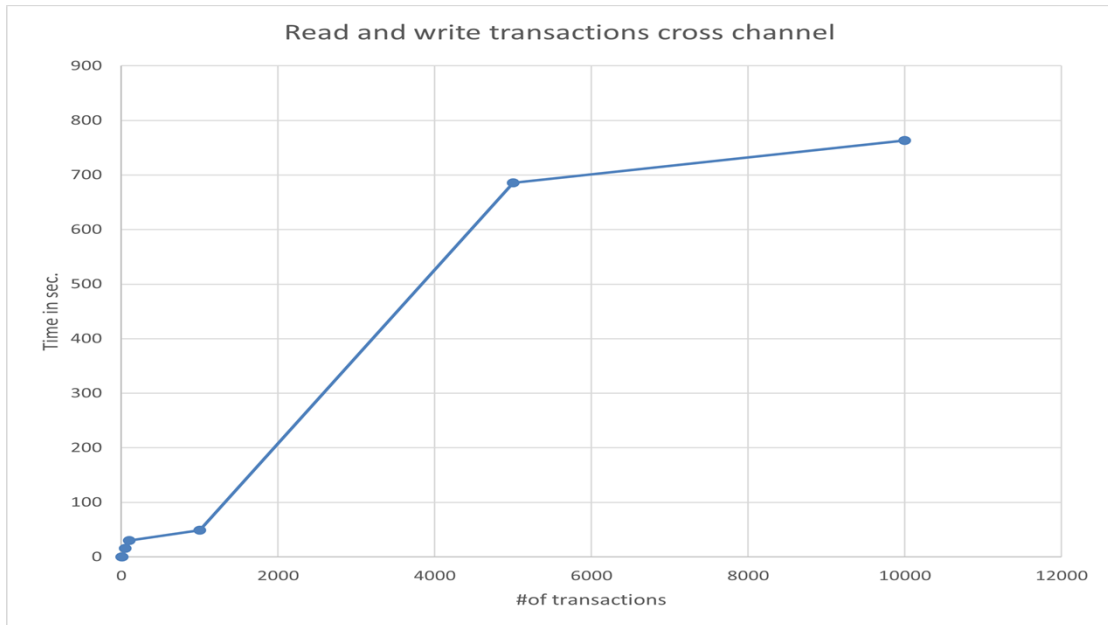


Figure 5.3. Time taken in cross channel data transfer.

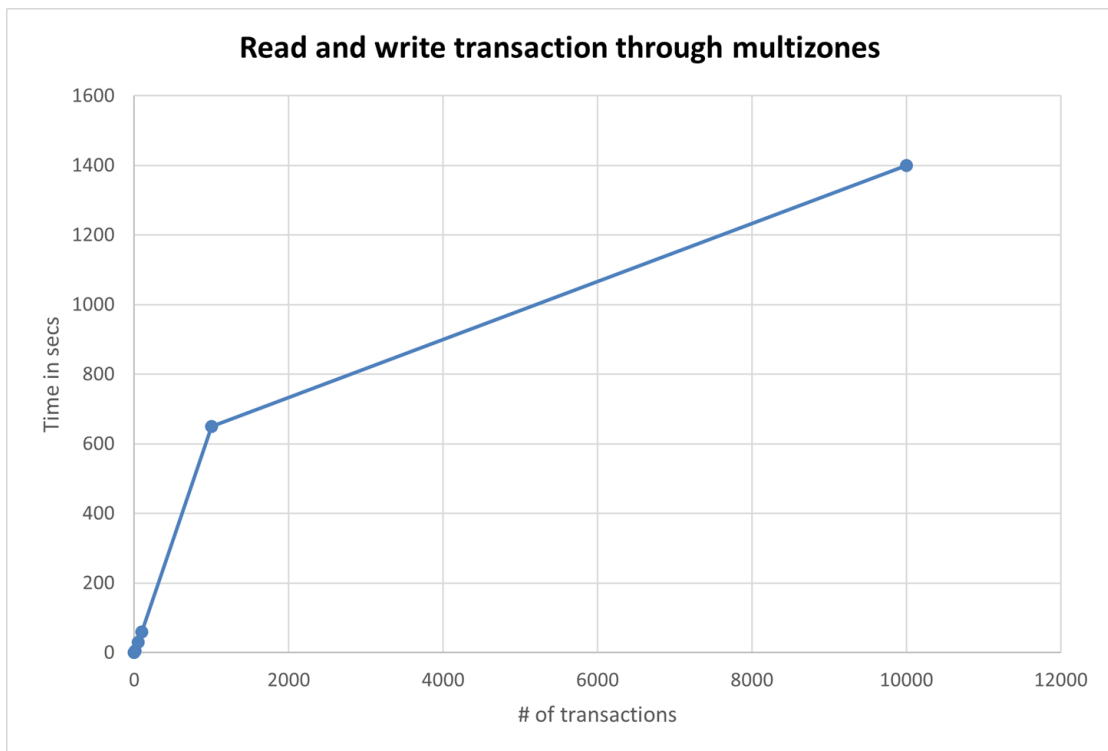


Figure 5.4. Read and write through a complete multi-channel setup.

The time taken with respect to large transactions is illustrate in Figure 29 and Figure 30. It shows a steep rise in time taken with respect to the number of transactions. To identify the effect of the presumed bottleneck the cross-channel communication, represents a similar pattern with respect to multi-zone /multi-channel analysis which confirms that cross channel data transfer produces an evident bottleneck to the private permissioned network. The data analyzed, we can confirm that there's a trade-off in terms of scalability, privacy and availability detailed in the section below.

5.4 Evaluation of end-to-end framework

The evaluation of the proposed framework was expected to be reliable, accurate and private. The indicators of each were quantified by accuracy of detection which supports reliability, the latency of the system that enabled availability in real time. The privacy was enabled by an accurate region of interest segmentation with key point extractors. The following is the detail on how each were accomplished.

5.4.1 Reliability of detection.

The reliability of detection was measured based on the accuracy of information shared and stored on the chain. Accuracy is measured throughout the pipeline, measured by accuracy of segmentation and detection. The following section will discuss key findings in terms of reliability.

5.4.1.1 Accuracy of detection

Accuracy of detection was evaluated on three frameworks, traditional approaches, clustering-based approach and light weight deep learning based approach. Four datasets were evaluated on these and each displayed varied results indication of the quality of the

dataset. The following were inferred for the approaches used.

5.4.1.2 Feature matching

Three feature extractors were used, and matching was performed on a query set. The results indicated that each method performed differently with different dataset. The advantage of this approach is that the algorithms do not require long processing time and the images produce acceptable accuracy. As the algorithms need not be trained, it's a generalized model and does not require continual learning. For the blockchain requirement. Keypoint indicators and descriptors are only required to be stored on the chain and this enables privacy of the information stored. The limitation of this approach is evident in testing that it requires quality images. To enhance accuracy the region of interest needs to be identified. However, this adds to the complexity of the end-to-end framework. The process of feature engineering performed must be based on the dataset and so cannot generalize easily.

5.4.1.3 Feature Clustering and matching

The BoW approach produced the best results in terms of accuracy. However, it takes more time and is more complex than feature matching approach. Positive reliability is confirmed in the perfect accuracy gained for this method.

Exact match is obtained through this process. The time taken for classification is in ms. However, the classification is based on matching the code book. The code book is constructed for this approach is based on the current database. Larger the code book larger the matching time as it must query through all the indexes to find the matching make.

5.4.2 Blockchain security

Blockchain is a decentralized network and provides security through pre-built consensus mechanisms and validation schemes. In this framework, Hyperledger fabric consists of orderers, endorsers and committer peers to validate a transaction. This setup enables a secure network where the participant read/write or query transactions. Each peer in the organizational network is certified by a certificate authority. Encryption mechanisms, hashing of the content and forming a chained network ensures the immutability. Further, the channel wise ledger access controls access to information for those outside the channel. This federated, consortium network strengthens the security of the information stored and thus the reliability of the information retrieved.

However, the limitation comes in the complexity of the model as seen in the results in the previous chapter, the elaborate security mechanisms may add to the delay in the network.

5.4.3 Privacy

Privacy is one of the key factors in a reliable and secure framework. To ensure privacy, the following techniques were evaluated: Their accuracy strengthened their reliability and availability. First, privacy was ensured through region of interest segmentation, in which the driver's face was removed and the background was excluded, so that bystanders or passengers were not included in the identification data shared in the ledger. With a 99.67 percent mean average precision in recognition, the reliability of region of interest segmentation ensured privacy.

Secondly, the cropped or segmented regions' features were extracted at given key-points using state-of-the-art key point detectors, as well as a bag of relevant features,

which further improved privacy. The features, rather than the image itself, are stored on the chain.

The third case is the CNN-based light weight classification, which requires the image to be stored. However, the image is in base64-encoded format. This process reduces the privacy score. The CNN-based approach can be further separated, where the classification layer is separated from the feature extractor layer, ensuring better privacy.

5.4.3.1 Blockchain

Privacy on the blockchain is established and pre-built due to its secure consensus mechanism as well as the data representation proposed in this framework. Three solutions for privacy preservation are exercised and evaluated in this framework. The cropped image was converted to base 64 code and stored on the chain. With the cropped region only saved, the privacy is preserved.

With the classification-based approaches the make and model of the vehicle are identified, and metadata stored. This requires just the information to be stored and not the image itself. This leads to an enhanced privacy scheme but requires further processing to identify unique indicators such as the location, which further adds to the complexity. In addition, duplication of the vehicles can cause confusion. License plate information extracted can improve the re-identification in this case.

Feature sharing is evaluated in this framework as it enables privacy. With features shared, the information is not directly accessible to ledger holders however, the features are matched to re-identify or recognize similar models and are thus private. In addition, the whole feature set is not shared, the significant key points and their features are

Table 5.6. End to End latency analysis

#	Segmentation (ROI)	Vehicle re-id	Blockchain	Total time
1	Mask RCNN + R50 + FPN (136ms)	Feature matching (2.67ms)	Single channel (5 org) (0.4 ms)	139.7
2	Mask RCNN + R 50 + FPN (136ms)	BoW Classification (102.69ms)	Single channel (5 org) (0.4 ms)	239.3
3	Mask RCNN + R50 + FPN (136ms)	Light weight CNN (1ms)	Single channel (5 org) (0.4 ms)	137.4
4	Mask RCNN + R50 + FPN (136ms)	Feature matching (2.67ms)	Multi-channel (5 org - 5 org) (2.5ms)	141.17
5	Mask RCNN + R50 + FPN (136ms)	BoW Matching (102.69ms)	Multi-channel (5 org - 5 org) (2.5ms)	241.19
6	Mask RCNN + R50 + FPN (136ms)	Light weight CNN (1ms)	Multi-channel (5 org - 5 org) (2.5ms)	139.5

extracted and shared on the blockchain network. Thus, privacy is established in every step of the framework.

5.4.4 Availability

The availability of the approach is analyzed based on the real time re-identification capability. The latency is evaluated as follows for each module. The following table lists the least latency approach and most time-consuming approaches.

It is evident from Table 5.6 and Figure 5.5, that the configuration that is most preferable for a real end to end system is the lightweight CNN method through a single channel network. Apart from Database1, the light-weight CNN has less accuracy compared to other approaches. The model with higher accuracy is that of BoW approach but latency is high compared, and thus real time detection may be comparatively delayed with respect to the CNN approach. Availability of the network thus is measured with real time usage. Although the lightweight CNN produces faster inference, the classification is limited to that dataset. And so, the availability for inference on other classes is not possible or may produce false positive re-identification. Feature matching approaches provide a generalized inference however, quality of the data is a concern. Likewise, Bag of features approach produces higher accuracy with feature set of learned class. However, a new vehicle make cannot be matched or classified as its feature set is not included in the code book. Thus, a trade off in terms of accuracy and availability is evident.

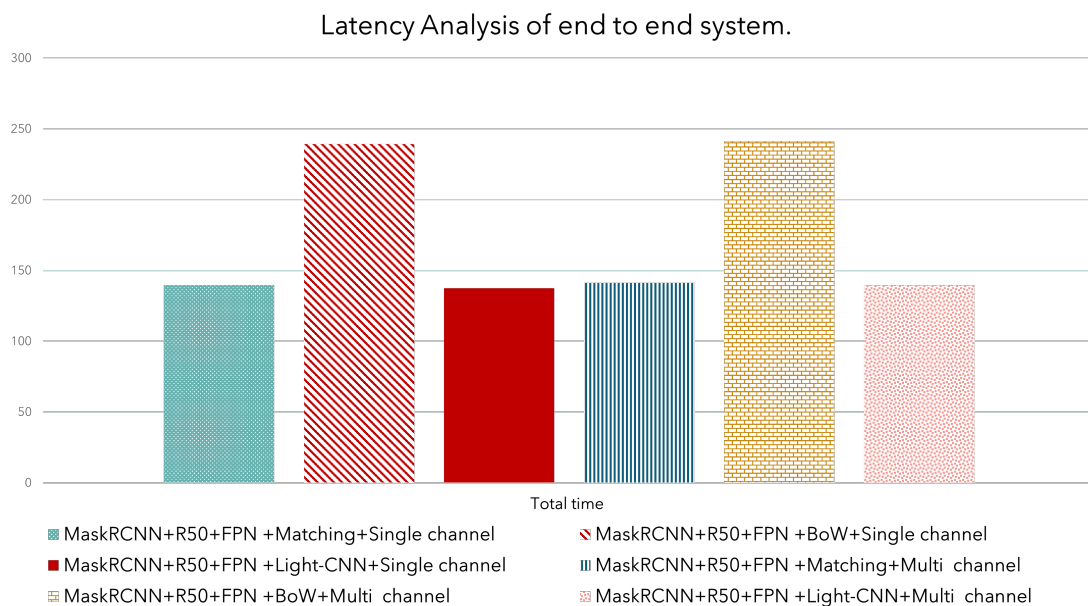


Figure 5.5. End to end latency analysis, The legend represents the configurations set in Table 5.6 for end to end blockchain enabled surveillance.

5.4.5 Cost

As hypothesized, the cost increases in terms of resource consumption and memory usage due to the high information processing rate in a multi-channel setup compared to a single channel setup confirmed from graphs plotted for resource consumption and memory. The spike is evident in a cross-channel data transfer evaluation illustrated in Figure 4.4. It can be established that scaling the network in terms of multi-zone-based approach has a cost in terms of the memory and CPU usage. However, the setup is secure and controlled by super-nodes.

This end-to-end framework is comparable to another similar approach that employs blockchain for anomalous vehicle identification through a cooperative multi-surveillance system [27]. The car recognition accuracy is 87% and takes 40s for the blockchain network to commit a transaction to the ledger. Compared to ours, the reliability is higher as accuracy is higher and the time taken for a single inference is in ms which is comparatively faster than the state of art. Thus, a new framework is proposed and evaluated for robust image feature representation of detected and segmented anomalous vehicles enabled through a secure and access controlled blockchain network, scaling many surveillance zones in shorter time.

5.5 Key findings, limitations, and future works

Within this thesis an accurate and fast model for segmenting the region of interest for private and reliable detection was identified using instance segmentation which accomplishes privacy and reliability as well as blockchain network. The accuracy of the segmentation through classification and detection was presented with comparison to existing literature on the same dataset. Instance segmentation approach presented

in one-of-a-kind approach used here which extracts the region of interest accurately. Existing dataset was annotated and modified for this approach. However, the limitation to this approach is complex architecture and thus increased inference time. As future work dataset requires to be expanded as well as light weight segmentation models evaluated on such as in [62].

Robust feature extraction techniques for the purpose of re-identification through matching and compare them with light-weight CNN based classification models which can be deployed on the edge was identified by evaluating three types of methods: BoW approach, Feature matching approaches and light weight CNN. A clear trade-off between time taken and accuracy was presented. In addition, the problem of generalization in machine learning and deep learning methods were a matter of concern solved by traditional approaches.

The limitation of the approach lies in the dataset used for the purpose of experimentation. The dataset is apt for feature matching approaches however requires increase in image and/or class quantity for deep learning models. Generalization is difficult for a small dataset as new vehicles makes, and models are manufactured every year.

A private, permissioned blockchain network for performing a reliable, private, and secure surveillance system in terms of its scalability and availability is utilized for surveillance. Two frameworks of blockchain were evaluated for which the surveillance scenario was modelled where all the nodes of the network or surveillance centers are part of a single channel that share the same distributed ledger modelling a private network with no centralized node but no access control. In the second setup a cross channel communication model where multiple surveillance zones are modelled is evaluated. This enables cross channel private communication between two channels through super

nodes enabling access control enhancing privacy. Both the frameworks were evaluated for latency and resource usage. The access-controlled network produced larger latency and resource constraints as expected and the bottleneck was at the cross-channel communication side. However, the second setup enhances privacy and improves the network reliability. The time difference in between two networks is in milliseconds.

Images were stored on the chain. Its latency and cost in terms of energy consumption were measured. The approach to safeguarding privacy, where key-point descriptors and feature sets were written, queried, and read for re-identification on the blockchain, was evaluated. Latency and energy consumption in terms of CPU usage and memory usage were measured for 10,000 transactions, showing an increased bottleneck on the cross-channel communication side. Using light-weight CNN and a multi-channel network, the end-to-end framework for a faster blockchain network was identified to be 139 ms in total. For a generalized model using feature matching, the latency was found to be 141.7 ms. Although the highest accuracy was with the BoW approach, the delay was 241 ms. Thus, a trade-off is identified in accuracy with latency, although the difference between each is significantly low.

The application of this approach is in high security surveillance systems where confidentiality and availability are required such as vehicle chase, vehicle anomaly detection and vehicle capture and re-identification. Further this can be used for other domains such as person re-identification or a general re-identification approach. Future work will be to create a single step approach rather than a pipeline as performed now. This would include modifying the existing smart contract. Further, limiting the method for image query in the blockchain network through further constraints. Hyperledger Fabric was modelled here due to its private permissioned setup for availability and

reliability. The setup however was not tested for security which is assumed to be pre-built feature; there is a necessity to verify this as future work. The following chapter concludes the thesis with the most significant contributions highlighted.

CHAPTER 6: CONCLUSION

The key question of this thesis was whether blockchain-based vehicle surveillance was feasible in terms of privacy, availability, latency, and reliability when inferential surveillance was performed using computer vision techniques, in this case vehicle re-identification. A framework was charted based on the literature review, that secured and privatized the vehicle surveillance system using computer vision techniques and a private, permissioned blockchain setup through a multi-zone surveillance representation. This research established a trade-off that was expected in terms of latency and accuracy for this end-to-end system. For a real-time blockchain-based surveillance system, a lightweight vehicle re-identification model through different approaches was evaluated to ensure reliability and privacy in terms of the information stored. In this regard, instance segmentation was performed by modifying an existing dataset. Vehicle re-identification through image matching and make classification was evaluated on four databases using traditional approaches as well as lightweight deep learning approaches. The opportunity to infer at the edge was evaluated concluding that light-weight approaches can be accurate. Lastly, an end-to-end system was evaluated in terms of the accuracy of vehicle surveillance and the overall time taken. The approaches used were comparable to state-of-the-art with lower latency and higher accuracy, with the least time of 137ms and the highest accuracy of detection was 100% for the top-1% rank in the BoW approach. Thus, a multi-vehicle surveillance system is enabled by blockchain with its constraints in a reliable, private, and secure manner.

REFERENCES

- [1] A. S. Elmaghraby and M. M. Losavio, “Cyber security challenges in smart cities: Safety, security and privacy,” *Journal of advanced research*, vol. 5, no. 4, pp. 491–497, 2014.
- [2] O. Elharrouss, N. Almaadeed, and S. Al-Maadeed, “A review of video surveillance systems,” *Journal of Visual Communication and Image Representation*, vol. 77, p. 103 116, 2021.
- [3] R. Kumar and R. Goyal, “On cloud security requirements, threats, vulnerabilities and countermeasures: A survey,” *Computer Science Review*, vol. 33, pp. 1–48, 2019.
- [4] J. Bugeja, D. Jönsson, and A. Jacobsson, “An investigation of vulnerabilities in smart connected cameras,” in *2018 IEEE international conference on pervasive computing and communications workshops (PerCom workshops)*, IEEE, 2018, pp. 537–542.
- [5] N. Kshetri, “Blockchain’s roles in strengthening cybersecurity and protecting privacy,” *Telecommunications policy*, vol. 41, no. 10, pp. 1027–1038, 2017.
- [6] M. Alshaiqli, T. Elfouly, O. Elharrouss, A. Mohamed, and N. Ottakath, “Evolution of internet of things from blockchain to iot: A survey,” *IEEE Access*, vol. 10, pp. 844–866, 2021.
- [7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

- [8] G. A. Murashova and D. Colbry, “Gm fasst: General method for labeling augmented sub-sampled images from a small data set for transfer learning,” *Machine Learning with Applications*, vol. 6, p. 100 168, 2021.
- [9] S. D. Khan and H. Ullah, “A survey of advances in vision-based vehicle re-identification,” *Computer Vision and Image Understanding*, vol. 182, pp. 50–63, 2019.
- [10] H. F. Atlam, A. Alenezi, M. O. Alassafi, and G. Wills, “Blockchain with internet of things: Benefits, challenges, and future directions,” *International Journal of Intelligent Systems and Applications*, vol. 10, no. 6, pp. 40–48, 2018.
- [11] E. Androulaki, A. Barger, V. Bortnikov, *et al.*, “Hyperledger fabric: A distributed operating system for permissioned blockchains,” in *Proceedings of the thirteenth EuroSys conference*, 2018, pp. 1–15.
- [12] S. Shalaby, A. A. Abdellatif, A. Al-Ali, A. Mohamed, A. Erbad, and M. Guizani, “Performance evaluation of hyperledger fabric,” in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, IEEE, 2020, pp. 608–613.
- [13] A. Fitwi, Y. Chen, and S. Zhu, “A lightweight blockchain-based privacy protection for smart surveillance at the edge,” en, in *2019 IEEE International Conference on Blockchain (Blockchain)*, pp., 2019, pp. 552–555.
- [14] P. Gallo, S. Pongnumkul, and U. Nguyen, “Blocksee: Blockchain for iot video surveillance in smart cities,” en, in *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/ICPS Europe)*, 2018, pp. 1–6.

- [15] T. Bui, D. Cooper, J. Collomosse, *et al.*, “Archangel: Tamper-proofing video archives using temporal content hashes on the blockchain,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.
- [16] S. Youssef, S. Rekhis, and N. Boudriga, “A blockchain based secure iot solution for the dam surveillance,” en, *IEEE Wirel. Commun. Netw. Conf. WCNC*, vol. 2019-April, 2019. DOI: [10.1109/WCNC.2019.8885479](https://doi.org/10.1109/WCNC.2019.8885479)..
- [17] R. Wang, W.-T. Tsai, J. He, C. Liu, Q. Li, and E. Deng, “A video surveillance system based on permissioned blockchains and edge computing,” en, in *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2019, pp. 1–6.
- [18] A. Singh, “A multi-layered network model for blockchain based security surveillance system,” en, in *2020 IEEE International Conference for Innovation in Technology (INOCON)*, pp., 2020, pp. 1–5.
- [19] Y. Jeong, D. Hwang, and K.-H. Kim, “Blockchain-based management of video surveillance systems,” en, in *2019 International Conference on Information Networking (ICOIN)*, pp., 2019, pp. 465–468.
- [20] K. Deepak, A. Badiger, J. Akshay, K. Awomi, G. Deepak, and H. Kumar, “Blockchain-based management of video surveillance systems: A survey,” en, in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 1256–1258.

- [21] D. Lee and N. Park, "Blockchain based privacy preserving multimedia intelligent video surveillance using secure merkle tree," en, *Multimed. Tools Appl*, pp. 1–18, 2020.
- [22] M. Singh, G. Aujla, and R. Bali, *Odob: One drone one block-based lightweight blockchain architecture for internet of drones*, en, pp., 2020.
- [23] X. Jiang, F. R. Yu, T. Song, and V. C. Leung, "Intelligent resource allocation for video analytics in blockchain-enabled internet of autonomous vehicles with edge computing," *IEEE Internet of Things Journal*, 2020.
- [24] X. Qi, "Blockchain-based content-oriented surveillance network," en, in *2020 IEEE 91st Vehicular Technology Conference*, pp., vol. VTC2020-Spring, 2020, pp. 1–6.
- [25] G. Saadouli, M. Elburdani, R. Al-Qatouni, S. Kunhoth, and S. Al-Maadeed, "Automatic and secure electronic gate system using fusion of license plate, car make recognition and face detection," en, in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIOT, 2020*, pp. 79–84.
- [26] L. Al-Sahan, F. Al-Jabiri, N. Abdelsalam, A. Mohamed, T. Elfouly, and M. Abdallah, "Public security surveillance system using blockchain technology and advanced image processing techniques," fr, in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIOT, pp., 2020*, pp. 104–111.
- [27] Z. Farr, M. Azab, and E. Samir, "Blockchain-based cooperative autonomous detection of suspicious vehicles," en, in *2020 11th IEEE Annual Information*

- Technology, Electronics and Mobile Communication Conference (IEMCON*, pp., 2020, pp. 188–192.
- [28] R. Michelin, N. Ahmed, S. Kanhere, A. Seneviratne, and S. Jha, *Leveraging lightweight blockchain to establish data integrity for surveillance cameras*, en, arXiv Prepr. arXiv1912.11044, 2019.
- [29] D. Nagothu, R. Xu, S. Nikouei, and Y. Chen, “A microservice-enabled architecture for smart surveillance using blockchain technology,” en, *2018 IEEE International Smart Cities Conference*, no. C2), pp. 1–4, 2018.
- [30] M. Islam and S. Kundu, “Iot security, privacy and trust in home-sharing economy via blockchain,” en, in *Blockchain Cybersecurity, Trust and Privacy*, pp., Springer, 2020, pp. 33–50.
- [31] X. Fan, Z. Zhong, Q. Chai, and D. Guo, “Ucam: A user-centric, blockchain-based and end-to-end secure home ip camera system,” en, in *International Conference on Security and Privacy in Communication Systems*, pp., 2020, pp. 311–323.
- [32] M. Hasan, Z. Wang, M. Hussain, and K. Fatima, “Bangladeshi native vehicle classification based on transfer learning with deep convolutional neural network,” en, *Sensors*, vol. 21, no. 22, p. 7545, 2021.
- [33] I. Fomin, I. Nenahov, and A. Bakhshiev, “Hierarchical system for car make and model recognition on image using neural networks,” en, in *2020 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM*, pp., 2020, pp. 1–6.

- [34] L. Yang, P. Luo, C. Loy, and X. Tang, “A large-scale car dataset for fine-grained categorization and verification,” en, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp., 2015, pp. 3973–3981.
- [35] L. Yang and T. Huang, “A vehicle reidentification algorithm based on double-channel symmetrical cnn,” en, *Adv. Multimed*, vol. 2021, 2021.
- [36] M. Manzoor and Y. Morgan, “Vehicle make and model classification system using bag of sift features,” en, in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC, 2017)*, pp. 1–5. DOI: 10.1109/CCWC.2017.7868475..
- [37] G. Pearce and N. Pears, “Automatic make and model recognition from frontal images of cars,” en, in *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS, 2011)*, pp. 373–378.
- [38] I. Zafar, E. Edirisinghe, S. Acar, and H. Bez, “Two-dimensional statistical linear discriminant analysis for real-time robust vehicle-type recognition,” en, in *Real-Time Image Processing 2007*, vol. 6496, 2007, pp. 9–16.
- [39] H. Lee, I. Ullah, W. Wan, Y. Gao, and Z. Fang, “Real-time vehicle make and model recognition with the residual squeeze-net architecture,” en, *Sensors*, vol. 19, no. 5, p. 982, 2019.
- [40] W. Li, Z. Yong, Y. Wei, and S. Hongxing, “Vehicle reidentification via multifeature hypergraph fusion,” en, *Int. J. Digit. Multimed. Broadcast*, vol. 2021, 2021.
- [41] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, “Transreid: Transformer-based object re-identification,” en, in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15 013–15 022.

- [42] R. Quispe, C. Lan, W. Zeng, and H. Pedrini, "Attributenet: Attribute enhanced vehicle re-identification," en, *Neurocomputing*, vol. 465, pp. 84–92, 2021.
- [43] B. Jiao, X. Tan, J. Zhou, L. Yang, Y. Wang, and P. Wang, *Instance and pair-aware dynamic networks for re-identification*, en, arXiv Prepr. arXiv2103.05395, 2021.
- [44] G. Adaimi, S. Kreiss, and A. Alahi, "Deep visual re-identification with confidence," en, *Transp. Res. part C Emerg. Technol*, vol. 126, p. 103 067, 2021.
- [45] J. Gu, W. Jiang, H. Luo, and H. Yu, "An efficient global representation constrained by angular triplet loss for vehicle re-identification," en, *Pattern Anal. Appl*, vol. 24, no. 1, pp. 367–379, 2021.
- [46] Q. Wang, "Viewpoint adaptation learning with cross-view distance metric for robust vehicle re-identification," en, *Inf. Sci. (Ny)*, vol. 564, pp. 71–84, 2021.
- [47] L. Alzubaidi, J. Zhang, A. J. Humaidi, *et al.*, "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, no. 1, pp. 1–74, 2021.
- [48] N. Ottakath, A. Al-Ali, and S. Al Maadeed, "Vehicle identification using optimised alpr.," 2021.
- [49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [50] A. Ojha, S. P. Sahu, and D. K. Dewangan, "Vehicle detection through instance segmentation using mask r-cnn for intelligent vehicle system," in *2021 5th international conference on intelligent computing and control systems (ICICCS)*, IEEE, 2021, pp. 954–959.

- [51] J. Dai, H. Qi, Y. Xiong, *et al.*, “Deformable convolutional networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 764–773.
- [52] A. Mumuni and F. Mumuni, “Data augmentation: A comprehensive survey of modern approaches,” *Array*, p. 100258, 2022.
- [53] S. D. Khan and H. Ullah, “A survey of advances in vision-based vehicle re-identification,” *Computer Vision and Image Understanding*, vol. 182, pp. 50–63, 2019.
- [54] S. Al-Maadeed, R. Boubezari, S. Kunhoth, and A. Bouridane, “Robust feature point detectors for car make recognition,” *en, Comput. Ind.*, vol. 100, pp. 129–136, 2018.
- [55] H. Bay, A. Ess, T. Tuytelaars, and L. Gool, “Speeded-up robust features (surf,” *en, Comput. Vis. image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
- [56] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide-baseline stereo from maximally stable extremal regions,” *en, Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, 2004.
- [57] P. Alcantarilla, A. Bartoli, and A. Davison, “Kaze features,” *en, in European conference on computer vision*, pp., 2012, pp. 214–227.
- [58] W. Wang, Y. Li, T. Zou, X. Wang, J. You, and Y. Luo, “A novel image classification approach via dense-mobilenet models,” *Mobile Information Systems*, vol. 2020, 2020.
- [59] B. K. Mohanta, S. S. Panda, and D. Jena, “An overview of smart contract and use cases in blockchain technology,” in *2018 9th international conference on*

computing, communication and networking technologies (ICCCNT), IEEE, 2018, pp. 1–4.

- [60] D. Guegan, “Public blockchain versus private blockchain,” 2017.
- [61] R. Lai and D. L. K. Chuen, “Blockchain—from public to private,” in *Handbook of Blockchain, Digital Finance, and Inclusion, Volume 2*, Elsevier, 2018, pp. 145–177.
- [62] Y. Lee and J. Park, “Centermask: Real-time anchor-free instance segmentation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 13 906–13 915.

APPENDIX : A

Ablation study of CNN.

Table .1. Ablation study of CNN, layerwise analysis.

#	Dataset	Image size	Ac- cu- racy(%)	Loss	Infer- ence time(ms)	RAM usage(k)
C1	Database1	140x70	22.9	2.32	4	388.9
C2	Database1	140X70	75.4	3.36	2	768.1
C3	Database1	160x160	90.6	0.43	1	252.0
C4	Database1	160x160	87.2	1.18	1	252.0
C5	Database1	160x160	88.9	0.62	2	252.2
C6	Database1	160x160	88.5	0.85	1	252.0
C7	Database2	128x128	65.1	1.65	15	368.8
C8	Database2	128x128	68.6	1.31	15	368.8
C9	Database2	128x128	61.7	1.45	15	368.8
C10	Database2	128x128	61.7	1.29	1187	193.4
C11	Database2	128x128	66.3	1.19	1187	193.4
C12	Database2	128x128	71.9	0.98	1187	767.5
C3	Database2	128x128	67.4	1.04	1187	767.5
C13	Database2	128x128	72.7	0.97	1187	193.4
C14	Database2	128x128	81.8	0.55	1187	767.5
C4	Database2	128x128	79.7	1.57	1187	193.4
C2	Database2	128x128	77.9	2.20	1187	193.4
C6	Database2	128x128	72.9	1.99	1187	193.4
C6	Database2	160x160	73.8	1.02	1	4
C13	Database2	160x160	77.3	2.28	240	502.0
1 layer	Database3	160x160	57.1	1.58	78	386
2 layers	Database3	160x160	32.1	4.37	887	1004
Mobilenet	Database3	160x160	55.4	3.17	9	733.6
1	Database4	160x160	18.8	5,6	50	502
Mobilenet	Database4	160x160	38.9	4.84	5	670.3

APPENDIX : B

Overall Docker statistic graphs

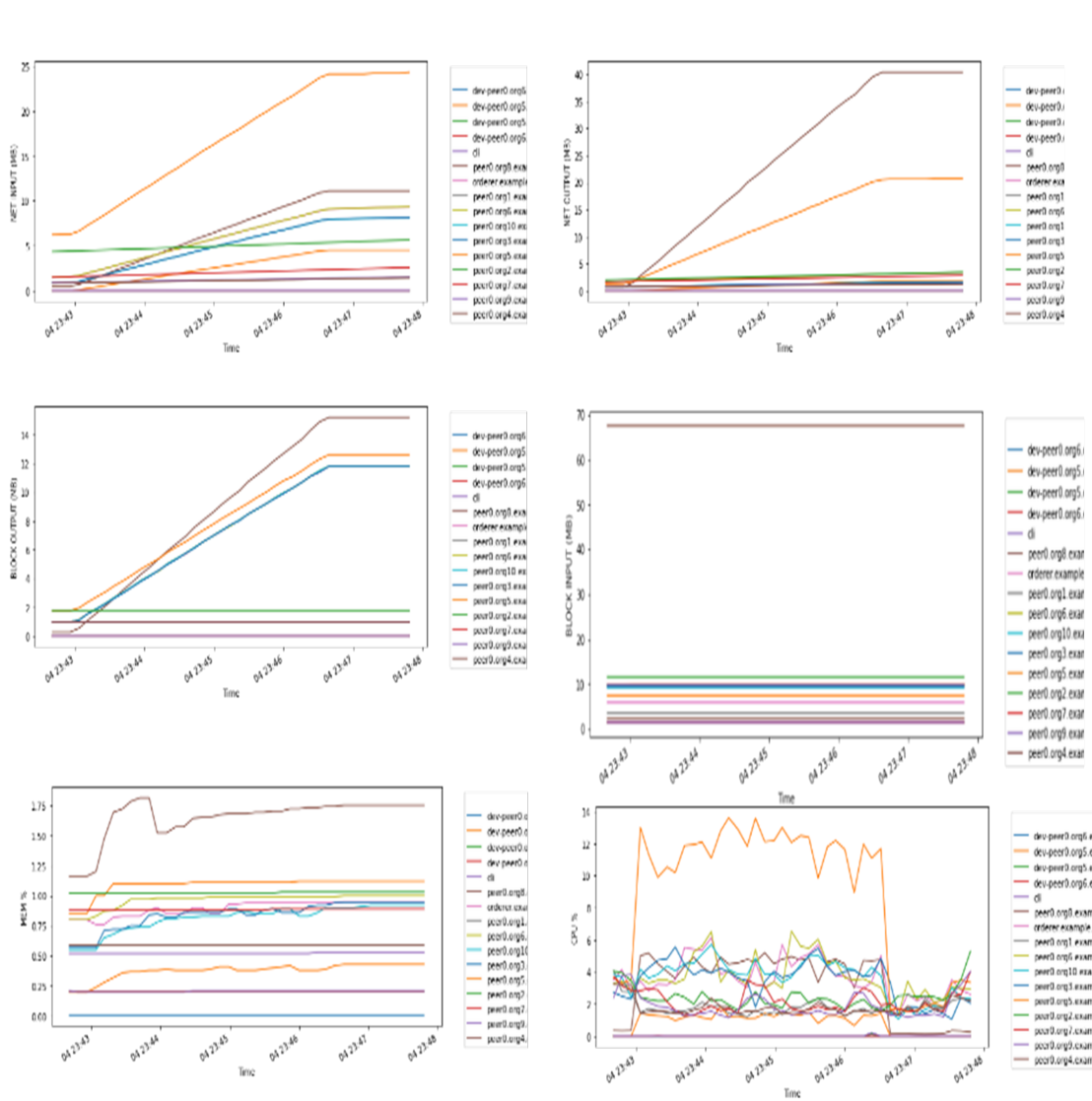


Figure .1. Docker statistics graphs of a single channel read activity.

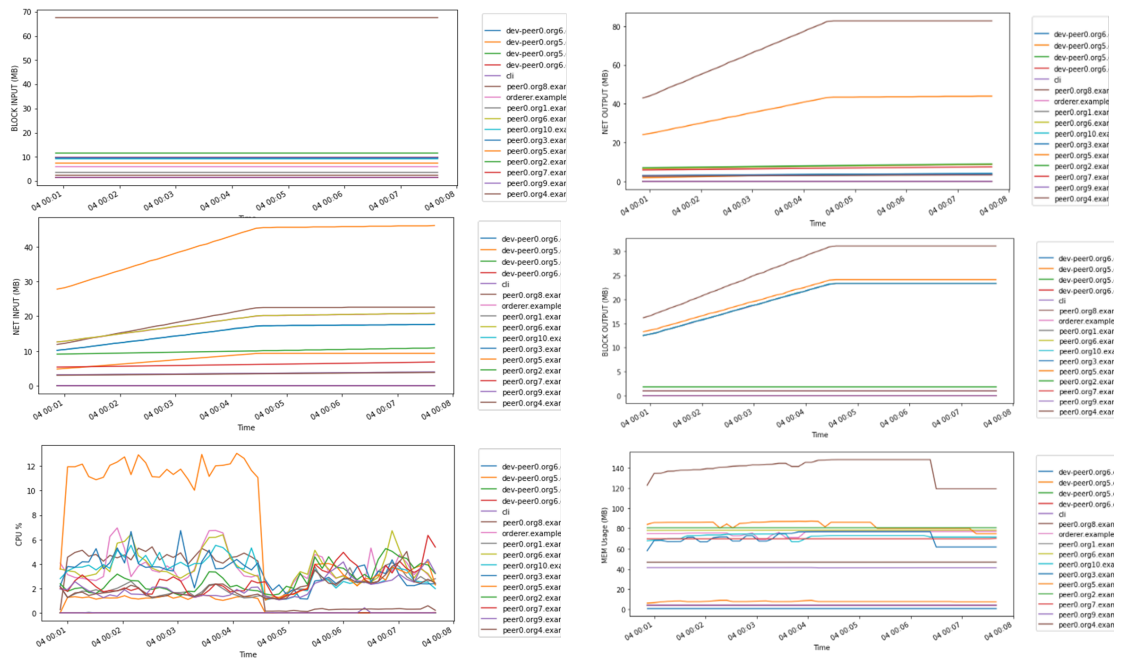


Figure .2. Docker statistics graphs of a single channel write activity.

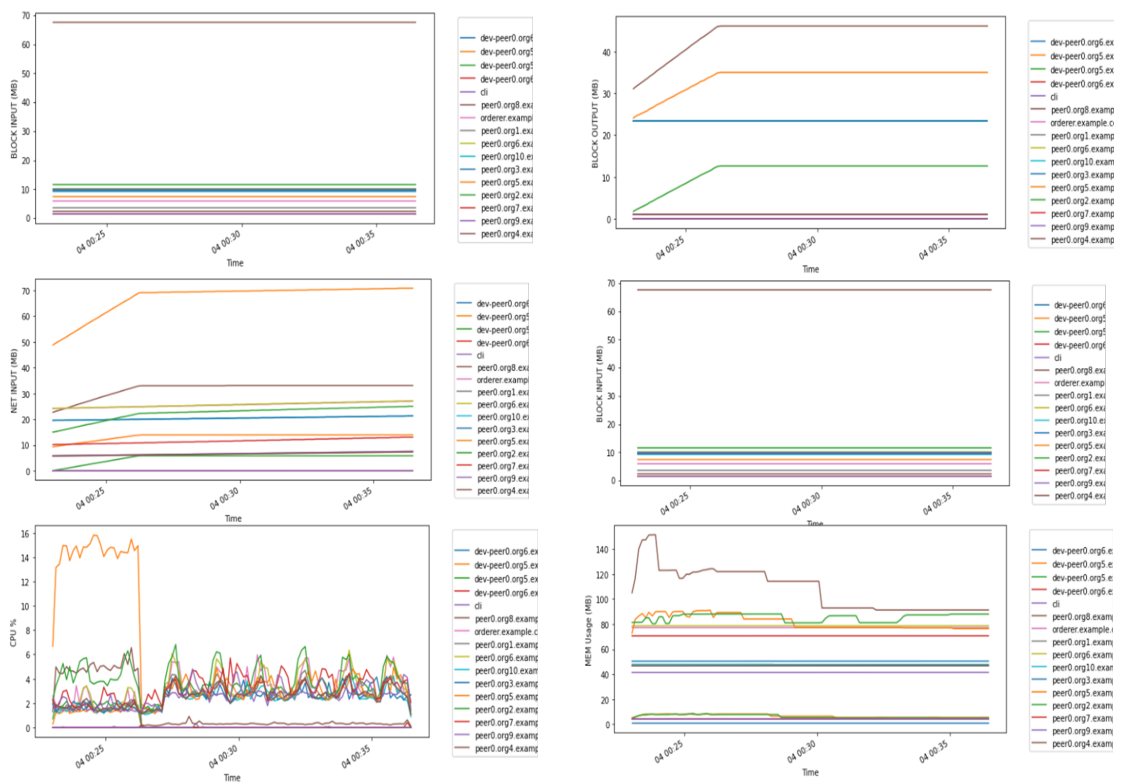


Figure .3. Docker statistics graphs of a cross channel write activity.