A Decentralized Digital Watermarking Framework for Secure and Auditable Video Data in Smart Vehicular Networks

Xinyun Liu , Ronghua Xu * , Yu Chen

Posted Date: 18 September 2024

doi: 10.20944/preprints202409.1228.v1

Keywords: Intelligent Transportation System (ITS); Internet of Vehicles (IoV); Digital Watermarking; Deep Learning; Blockchain; Security

Article

A Decentralized Digital Watermarking Framework for Secure and Auditable Video Data in Smart Vehicular Networks

Xinyun Liu [1], Ronghua Xu [1,*] and Yu Chen [2]

[1] Department of Applied Computing, Michigan Technological University, Houghton, MI 49931, USA; xinyunl@mtu.edu (X.L.); ronghuax@mtu.edu (R.X.)

[2] Department of Electrical and Computer Engineering, Binghamton University, Binghamton, NY 13902, USA; ychen@binghamton.edu (Y.C.)

* Correspondence: ronghuax@mtu.edu

Abstract: Thanks to the rapid advancements in Connected & Automated Vehicles (CAVs) and vehicular communication, the concept of the Internet of Vehicles (IoV) combined with Artificial Intelligence (AI) and big data technologies promotes the vision of an Intelligent Transportation System (ITS). By enabling a comprehensive data exchange platform, ITS is critical in enhancing road safety, traffic efficiency, and the overall driving experience. However, the open and dynamic nature of IoV networks brings significant performance and security challenges to IoV data acquisition, storage, and usage. To comprehensively tackle these challenges, this paper proposes a **D**ecentralized **D**igital **Wa**termarking framework for smart **Ve**hicular networks (D2WaVe). Specifically, D2WaVe consists of two core components: FIAE-GAN, a novel feature-integrated and attention-enhanced robust image watermarking model based on a Generative Adversarial Network (GAN), and BloVA, a **Blo**ckchain-based **V**ideo frames **A**uthentication scheme. By leveraging an encoder-noise-decoder framework, trained FIAE-GAN watermarking models can achieve the invisibility and robustness of watermarks that can be embedded in video frames to verify the authenticity of video data. Then, BloVA ensures the integrity and auditability of IoV data in the storing and sharing stages. Experimental results based on a proof-of-concept prototype implementation validate the feasibility and effectiveness of our D2WaVe scheme for securing and auditing video data exchange in smart vehicular networks.

Keywords: Intelligent Transportation System (ITS); Internet of Vehicles (IoV); Digital Watermarking; Deep Learning; Blockchain; Security

1. Introduction

With the development of intelligent sensing, the fifth generation (5G) communication, and edge computing technologies, Connected & Automated Vehicles (CAVs) are promising to revolutionize transportation systems. Leveraging onboard smart devices such as wireless sensors, electronic control units (ECUs), GPS antennas, radar, and so on, CAVs can collect and process large volumes of context-aware information (in-vehicle data and environmental data) while enabling information exchange and cooperative tasks between vehicles [1]. Thanks to advancements in CAVs and vehicular networks, the Internet of Vehicles (IoV) concept has become realistic through the seamless integration of CAVs, roadside infrastructures, pedestrian-carry devices, transportation service providers, and transportation management systems. The proliferation of IoV combined with Artificial Intelligence (AI) and big data technologies leads to realizing the vision of intelligent transportation system (ITS) in Smart Cities, which provide ubiquitous, intelligent, and safety applications for communities and society [2].

In the era of ITS, vehicular networks have emerged as a pivotal element for information fusion based on diverse IoV networks and computing paradigms, thereby enhancing road safety, traffic efficiency, and overall driving experience [3]. Figure 1 demonstrates participants and interactions of ITS

 doi:10.20944/preprints202409.1228.v

based on a crossroad scenario. The information interaction between vehicles and other entities refers to vehicle-to-everything (V2X) models, which include vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-roadside Unit (V2R), vehicle-to-grid (V2G), and vehicle-to-pedestrian (V2P) [1]. By enabling a comprehensive data exchange platform supporting the mobility of CAVs and heterogeneity of vehicular services, ITS provides diverse intelligent and safe vehicular applications, like enhanced pedestrian and driving safety, efficient traffic planning, smart parking, and entertainment services [4]. Meanwhile, ITS requires in-vehicle data transmission and traffic information sharing among participants. This inevitably incurs new performance and security challenges as collecting, storing, and transacting large amounts of IoV data.

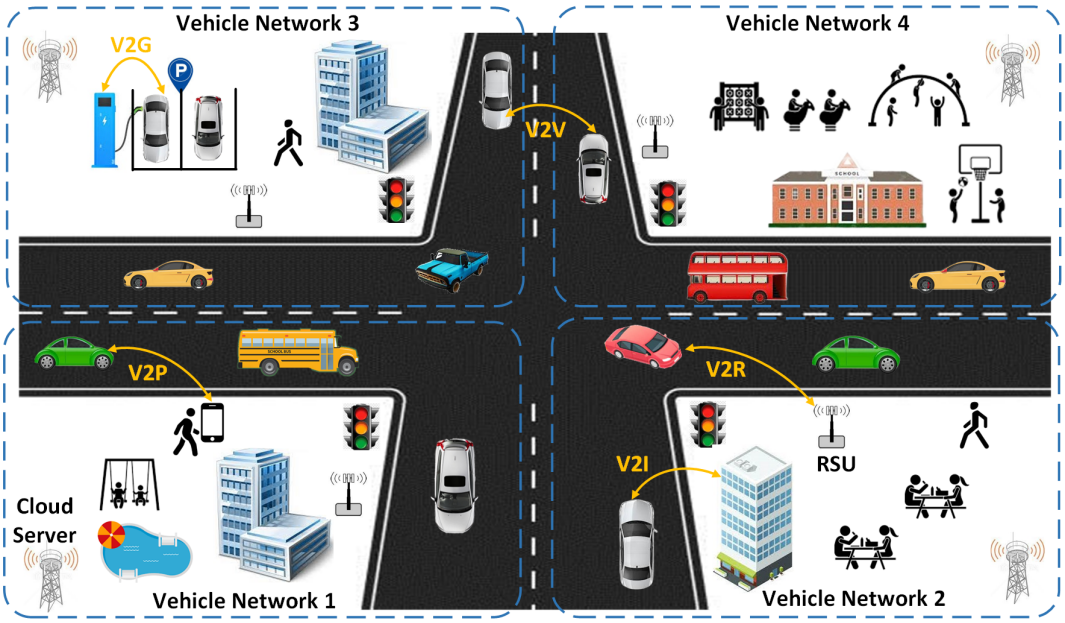

Figure 1. The overview of an ITS consisting of multiple IoV networks.

1.1. Addressing Challenges of Data Source Authenticity

With the growing number of CAVs, tremendous driving-related IoV data, such as video streams from cameras, point cloud data from LiDAR (Light Detection and Ranging), and vehicle control messages are continuously produced. Due to the limited capability of CAVs, generated IoV data are transmitted to surrounding roadside units (RSUs), edge servers, or cloud servers, which process and store raw IoV data. Thus, ensuring the security of IoV data across its lifetime (data acquisition, data in transit, and data at rest) is a significant challenge. Especially for the data source's authenticity and data storage integrity.

By embedding imperceptible information within digital content, digital watermarking has been recognized as an effective technique for data security, such as copyright protection, content identification, and forensics [5]. Thanks to its key features like imperceptibility and robustness, digital image watermarking (DIW) is a prominent solution for safeguarding the security of IoV data throughout its lifecycle, particularly in verifying the authenticity of the data and ensuring the integrity of stored data from multiple source. DIW is especially useful in forensic situations, such as investigating traffic accidents or analyzing vehicle malfunctions, where data is derived from video streams generated by vehicle vision technologies, including in-car cameras and video surveillance systems.

To address these issues and achieve both high invisibility and robustness in digital watermarking, we introduce FIAE-GAN, a novel feature-integrated and attention-enhanced robust image watermarking model based on Generative Adversarial Network (GAN). FIAE-GAN employs an encoder-noise-decoder structure for end-to-end training. A Feature Integration Module (FIM) is

 doi:10.20944/preprints202409.1228.v

incorporated to enhance image feature extraction, capturing both shallow and deep features across multiple layers. Combining the watermark with deep features increases resistance to image noise while using dense connections further boosts robustness. To reduce image distortion from watermarking, the model uses an Attention-Enhanced Module (AEM), which creates an attention mask by analyzing global image features and adjusts the watermark strength for different image regions—suppressing flat or sensitive areas and enhancing inconspicuous or textured ones for better embedding. Additionally, the adversarial relationship between the encoder and discriminator is leveraged to improve the quality of the encoded image. The discriminator helps distinguish the encoded image from the original during iterative training, supported by the AEM.

1.2. Addressing The Issue of Data Storing and Sharing Security

The conventional ITS frameworks rely on cloud servers to collect information and store data. Such a centralized system architecture is prone to single-point failures. For example, successful distributed denial of service (DDoS) attacks to control servers make important ITS services unavailable. In additional, the in-vehicle driving data also contains critical evidence for forensics scenarios, such as traffic accident investigation and vehicle fault analysis [6]. Therefore, it is critical to ensure the data's audibility and traceability in case of disputes about reliable data service.

With attractive characteristics in decentralization, immutability, and transparency, Blockchain has demonstrated great potential to construct a trust-free and secure network infrastructure for complex and heterogeneous IoT ecosystems [7]. Distributed ledger technology (DLT) promises to ensure public auditability and traceability of IoV data. Blockchain promotes the migration from centralized to decentralized IoV frameworks, thereby reducing the risks of single-point failures and improving the availability and resilience of data service in IoV networks.

To improve availability and residence of video data services atop IoV networks with high heterogeneity and dynamics, we propose a Blockchain-based video frames authentication (BloVA) scheme using Blockchain and distributed storage technologies. To address the limitations of traditional monolithic service-oriented architecture (SOA) that lacks flexibility, a microservices-oriented architecture (MOA) is adopted for video data delivery and security schemes, which bring attractive features, such as scalability, fine granularity, loose coupling, continuous delivery [8]. MOA improves system availability from a service architecture aspect. In addition, We also adopt a hybrid on-chain & off-chain storage [9] to build a distributed content delivery fabric for IoV networks. This can improve efficiency of integrating Blockchain and ensure privacy preservation by not directly exposing video data and other sensitive information on transparent distributed ledger.

1.3. Main Contributions

To comprehensively tackle the challenges discussed above, this paper proposes a **D**ecentralized **D**igital **Wa**termarking framework for smart **Ve**hicular networks (D2WaVe) to guarantee video data's authenticity, integrity, and auditability. To ensure the authenticity of video recordings in the IoV network, we explore deep learning-based DIW methods and present a novel feature-integrated and attention-enhanced robust image watermarking model based on GAN (FIAE-GAN). In addition, we integrate BloVA, a blockchain-based security fabric that improves the availability of watermarked video data by storing them in an IPFS (InterPlanetary File System) [10] based storage network. Moreover, we also develop security services on top of the Ethereum Blockchain network that allow authorized users to publicly verify data integrity and audit data exchange without relying on any third-party trust authority.

While many digital watermarking methods have been developed for in-vehicle network security, such as CAN bus authentication [11] and LiDAR data integrity [12], our D2WaVe integrates DIW solutions to ensure authenticity and provenance of video data in IoV networks. Compared to current Blockchain-enabled solutions for IoV data management [13], we introduce DIW technology to ensure the authenticity of raw video data rather than simply using hash values or digital signatures to verify

 doi:10.20944/preprints202409.1228.v

data integrity. To our knowledge, D2WaVe is the first scheme that integrates DIW and Blockchain for video data security in an IoV network.

In summary, the main contributions of this paper are highlighted as follows:

- We propose the system architecture of D2WaVe, a decentralized digital watermarking framework consisting of i) a novel FIAE-GAN model for watermark embedding and extraction and ii) BloVA, a Blockchain-based authentication scheme to support secure video data storage and sharing.
- We introduce a novel FIAE-GAN-based digital watermarking model and provide details on the encoder, decoder, and discriminator.
- We evaluate the performance of the watermarking model and overheads incurred by the video authentication framework. Experimental results demonstrate that both the AEM and FIM are crucial for the watermarking model's performance, surpassing existing models in watermarked image quality and resilience against diverse attacks.

The remainder of the paper is organized as follows: Section 2 provides background knowledge in terms of feature integration in DenseNet and attention module and reviews the state-of-the-art digital watermarking technologies and blockchain-based security solutions to IoV networks. Section 3 introduces the design rationale and system architecture of D2WaVe, as a secure and auditable video data framework for IoV networks. Details of FIAE-GAN model and core components are explained in Section 4. Section 5 reports the prototype implementation and performance evaluation. Finally, Section 6 summarizes this paper with a brief discussion on current limitations and future directions.

2. Background and Related Work

This section describes evolution of digital watermarking technology and state-of-the-arts on deep learning-based watermarking methods. Then we introduce fundamentals of DenseNet and attention-enhance module. Finally, we provide related work on Blockchain-based solutions to IoV networks.

2.1. Digital Image Watermarking for Data Authentication

As a well-established research area, digital image watermarking (DIW) continues to attract significant interest from both academia and industry. DIW offers practical applications in copyright protection, source tracking, and data integrity verification. By embedding imperceptible information within digital content, digital watermarking is widely recognized as a reliable technique for enhancing data security, including forensics, content identification, and copyright protection [5].

Traditional digital watermarking methods can be categorized into spatial-domain and frequency-domain techniques. The spatial-domain techniques involve directly modifying the pixel values of the host image. One common approach is the Least Significant Bit (LSB) method. The frequency-domain techniques embed the watermark in the frequency coefficients of the host image, obtained through transformations such as Discrete Cosine Transform (DCT) [14], Discrete Wavelet Transform (DWT) [15], and Discrete Fourier Transform (DFT) [16]. However, these works normally demand extensive prior experience and involve complex operations, including preprocessing and postprocessing.

In recent years, deep learning has made remarkable strides in computer vision tasks, such as image classification and segmentation, due to its powerful ability to represent complex patterns [17–19]. Unlike traditional machine learning, deep learning models can automatically extract more intricate features from images and offer improved generalization across a wide range of scenarios by training on large-scale datasets. This capability has also been successfully applied to the field of image watermarking.

Existing deep network models for image watermarking are generally classified into two main types: convolutional neural networks (CNNs) and generative adversarial networks (GANs) [20]. A robust image watermarking model based on CNNs incorporated an iterative learning framework to

 doi:10.20944/preprints202409.1228.v

enhance watermark resilience [21], extending the frequency domain commonly utilized in traditional watermarking methods. ReDMark, a deep end-to-end diffusion watermarking framework, can learn watermarking models in any desired transform domain [22]. ReDMark consists of two fully convolutional networks with a residual structure, simulating various attacks through a differentiable network layer, enabling end-to-end training.

GANs offer an alternative approach for deep image watermarking. A GAN comprises a generator and a discriminator, engaging in an adversarial process [23]. This adversarial dynamic makes GANs inherently well-suited for robust image watermarking, balancing capacity, invisibility, and resilience. In recent years, several GAN-based image watermarking models have emerged. Notably, a pioneered model called HiDDeN [24] leverages the adversarial interplay between the generator and the discriminator for robust watermarking. HiDDeN adopts an encoder-noise-decoder architecture, where the encoder generates an imperceptible encoded image, and the decoder effectively retrieves the original watermark.

However, the aforementioned watermarking models, especially the extracted image features used for embedding the watermark, are still insufficient and lack robustness, which diminishes their ability to withstand image noise. More critically, these models fail to emphasize essential image features during the learning process, ultimately reducing the effectiveness of the watermarking.

To tackle these challenges and achieve a balance between invisibility and robustness in digital watermarking, our FIAE-GAN model leverages FIM to capture both shallow and deep image features, enhancing noise resistance by embedding the watermark into deep layers. To minimize distortion, AEM creates an attention mask that adjusts watermark strength, suppressing impact on sensitive regions while reinforcing it in textured areas for optimal embedding.

2.2. Feature Integration Module Using DenseNet

Feature integration is a crucial aspect of image processing, and significant advancements have been seen with the development of CNNs. DenseNet (Densely Connected Convolutional Networks) has been particularly influential in this domain [25]. DenseNet features direct connections between any two layers with the same feature map size, ensuring maximum information flow between layers. The direct connection architecture addresses the vanishing gradient problem, encourages feature reuse, and substantially reduces the number of parameters compared to traditional CNNs. As shown in Figure 2, DenseNet consists of multiple dense blocks. Each dense block includes a sequence of BN–ReLU–Conv1×1 and BN–ReLU–Conv3×3 layers. In this context, BN–ReLU–Convj×j denotes a layer sequence composed of a BatchNorm operation, a rectified linear unit (ReLU) activation, and a convolution with a kernel size of 3×3.

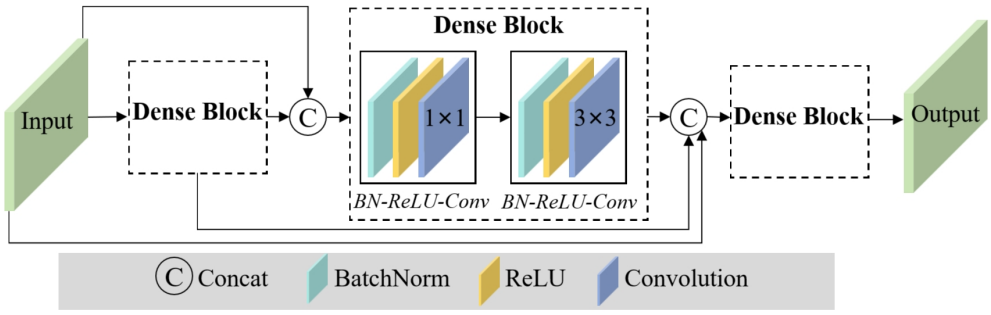

Figure 2. Architecture of DenseNet. The DenseNet extract both shallow and deep features, which are then fused with the watermark to enhance its robustness.

Recent works have leveraged DenseNet's feature integration capabilities to enhance various image-processing tasks. DenseNet was incorporated for image super-resolution [26], demonstrating improved performance due to the effective combination of shallow and deep features. In addition,

 doi:10.20944/preprints202409.1228.v

DenseNet has been employed in medical image analysis for accurate lesion detection by integrating multi-scale features [27].

By extracting multiple features through dense blocks and reusing these features, the network's representational capacity is significantly improved. This feature reuse strategy integrates the benefits of dense connections into the watermarking model, thereby improving resistance to various image distortions and attacks.

2.3. Attention-Enhanced Module

Attention mechanisms have gained significant traction in the deep learning community for their ability to focus on the most relevant parts of the input data, thereby improving the model's performance [28]. Attention mechanisms dynamically weigh the importance of different features, allowing the model to prioritize critical information while suppressing less relevant details. The application of attention mechanisms in image processing tasks has shown promising results. For instance, the introduction of the Squeeze-and-Excitation (SE) network [29] highlighted the importance of channel-wise attention in improving network performance by adaptively recalibrating channel-wise feature responses.

While existing watermarking models utilize attention mechanisms to enhance either invisibility or robustness, achieving both simultaneously remains challenging [30]. Our approach distinguishes from prior methods by employing a spatial attention mechanism that captures the maximum and average values at each spatial position through max and average pooling, as illustrated in Figure 3. Spatial position weights are subsequently learned via a convolutional layer followed by a sigmoid function. These weights are then applied to each spatial position on the feature map, resulting in features with enhanced spatial significance. Additionally, global features are used to create an attention mask, which identifies inconspicuous and texture-rich areas, thereby improving overall watermarking performance.

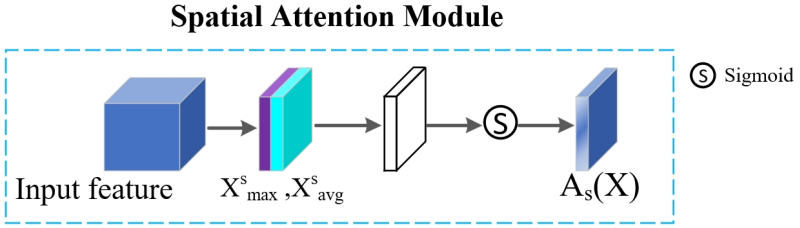

Figure 3. Architecture of Spatial Attention Module. It helps embed the watermark in less noticeable regions.

2.4. Blockchain for IoV Networks

As the underlying technology of Bitcoin [31], Blockchain has demonstrated great potential to revolutionize conventional information technology systems. Essentially, Blockchain is a distributed ledger technology (DLT) that leverages cryptographic functions and consensus protocols to ensure a verifiable, append-only chained data storage of transactions. From a network aspect, Blockchain uses an underlay Peer-to-Peer (P2P) network infrastructure to exchange data (e.g., transactions and blocks) and control messages among participants. All participants (miners or validators) rely a distributed consensus protocol to ensure security and consistency of data on the distributed ledger in a decentralized manner. By implementing user agreement rules into self-executed programs, smart contract brings programmability into Blockchain. The smart contract has become popular for developing decentralized Applications (DApps) that provide efficient and reliable services without depending on any trusted entity.

Blockchain technology and smart contracts have recently been adopted in IoV networks to address diverse security issues. By leveraging consortium blockchain, a decentralized, secure, and distributed data management system is proposed for vehicular edge computing networks [32]. Smart contracts are deployed on a vehicular Blockchain to ensure the security of RSU data storage and data sharing

 doi:10.20944/preprints202409.1228.v

among vehicles. DrivMan [33] is another blockchain-based solution to enable trust management, data provenance, and privacy in IoV networks. Because a physically unclonable function (PUF) can allocate a unique crypto fingerprint to each vehicle, DrivMan achieves data provenance by using PUF to provide the root of trust. A Blockchain-based data-sharing framework is proposed to protect against maliciously tampered or forged multimedia data in vehicular social networks [34]. Unlike the solutions mentioned above, our FIAE-GAN digital watermarking model ensures the authenticity of video data without relying on hardware support like PUF. We also use a Merkel tree to manage on-chain and off-chain data thereby ensuring frame integrity and sequence consistency for video data in IoV networks.

3. D2WaVe: Design Rationale and System Architecture

This section presents details of our proposal, D2WaVe, a decentralized digital watermarking framework atop Blockchain and IPFS storage. We start by discussing system settings and thread models. Then, we present the system architecture along with key components. Finally, we explain the workflows of the video frame authentication scheme.

3.1. System Settings and Adversary Model

The proposed D2WaVe framework assumes permissioned network environments where trustworthy oracles maintain the registration of all participants. Each nodes (e.g., vehicle, camera, edge server and user) uses its public key and privacy key for mutual authentication among peers. In addition, video data generated from source devices (e.g., vehicle camera or networked camera) are sent to edge nodes (RSU or edge server), which acts as trust nodes to perform digital watermarking functions and publish data.

Thus, we also assume i) a secure system environment for video data generation at the source side and processing at edge side; and ii) security communication channels between source and edge. Thus, an adversary cannot tamper with and modify video frames and watermarks by compromising an in-vehicle network, networked camera, and their associated edge computing platforms. Moreover, we assume an adversary cannot break cryptographic hardness, such as computing conflict hash values and compromising digital signature and encryption schemes. We assume the attackers aim to manipulate the video streams in storage and transmission. Here are several possible security attacks on video data in the IoV network.

Forgery attacks: An adversary can use powerful image editing tools and software to modify original video frames and then replace tampered data with real data. Due to advancements in generative AI technology, attacks can leverage deepfake tools to create fake video recording content to disturb normal surveillance system functions [35]. The genuineness of video data is very important in vehicle forensics and accident investigation. Forgery attacks change the meaning of evidence (video frames) without leaving any detectable clues, making it challenging to verify the authenticity of video data in the IoV network.

Integrity attacks: malicious participants in the IoV network can tamper the content of data stored by service providers or modify data transmitted between source and destination. For example, attackers can insert fake frames, replace them with tampered frames, and even delete frames in original video recordings generated by their owners. Therefore, an adversary not only destroys the integrity of each frame but also breaks the time-sequence properties of video recordings that follow a continuous process.

Availability attacks: An adversary can launch Denial of Service (DoS) attacks by sending fake requests to nodes that provide IoV data service. As a result, these service nodes cannot allocate enough resources to serve genuine requests from honest users. In Distributed Denial of Service (DDoS) attacks, an adversary can use bot (zombie) networks to flood network traffic and paralyze a target cloud server. Thus, the participants will not get the required data to support ITS functions.

 doi:10.20944/preprints202409.1228.v

3.2. System Overview

Aiming at a secure-by-design video data service infrastructure for assurance-and resilience-oriented ITS applications, D2WaVe combines the merits of digital watermarks with the security properties of blockchain to resist the threats mentioned above. Figure 4 demonstrates the system architecture consisting of two sub-frameworks: (i) a video frame authentication scheme by using a robust FIAE-GAN based digital watermarking model; (ii) a decentralized security fabric atop Blockchain and distributed data storage.

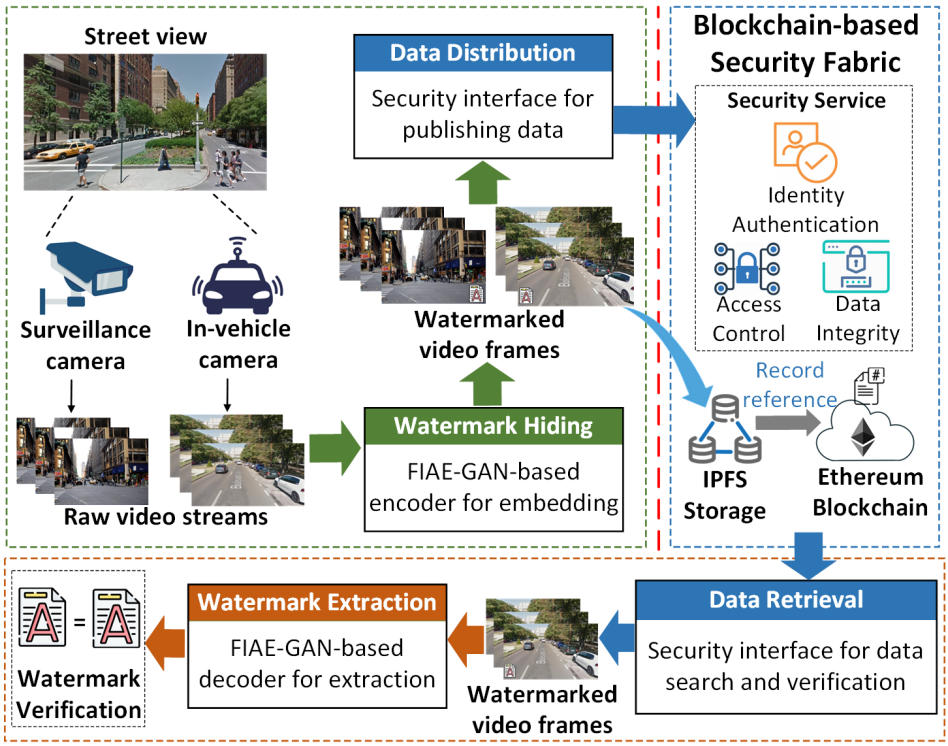

Figure 4. The architecture of D2WaVe.

In the IoV network, in-vehicle and surveillance cameras generate a large volume of video data. The raw video data can be transmitted to nearby edge computing platforms, which provide storage and computing resources. To mitigate video forgery attacks in IoV networks, an FIAE-GAN-based digital watermarking model can generate robust and invisible digital watermarks that data owners embed into specific video frames. Figure 4 shows that edge nodes run FIAE-GAN-based encoder under a secure execution environment to embed proofs into frames of video streams and then create watermarked frames. Data users can utilize the FIAE-GAN decoder to extract digital watermarks from watermarked frames of video streams and then verify whether they are genuine. Given the assumption that attackers cannot access FIAE-GAN-based encoders, it is hard to create valid digital watermarks for fake video frames. Therefore, FIAE-GAN-based digital watermark embedding and verification can improve authenticity during video data generation and use stages. As an essential contribution of this work, the detailed design of the FIAE-GAN-based digital watermarking model is further presented in Section 4.

To address integrity attacks and availability issues in video data distribution and use stages, a blockchain-based security fabric [36] is adopted as the underlying network infrastructure for the IoV network. As the right part of Figure 4 shows, the decentralized fabric contains three sub-systems: (1) security services based on the microservice-oriented architecture (MoA); (2) an Ethereum blockchain network to support security services; and (3) a decentralized data storage (DDS) system atop IPFS network to save off-chain data. By encapsulating security functions into separate containerized microservices loosely coupled with the rest of the system, these security services work independently

 doi:10.20944/preprints202409.1228.v

under a distributed network environment, improving data service availability. In addition, we use the DDS system as an off-chain storage solution to save raw IoV data and meta-information. Compared to centralized data storage solutions, DDS can reduce data loss risks and improve data services' resilience. Moreover, Ethereum Blockchain ensures integrity and traceability of the on-chain data saved by the distributed ledger. Thus, it provides a decentralized and trust-free platform for secure and auditable IoV data.

3.3. Video Frames Authentication Scheme

We use a crossroad scenario to explain how a decentralized digital watermarking framework can secure video data generation, storage, and usage. The system initialization guarantees a secure environment for FIAE-GAN digital watermarking models and algorithms at the data source side (e.g., in-vehicle camera systems, and surveillance camera host machines). In addition, trustworthy system administrators deploy smart contracts supporting security services on the Ethereum Blockchain. Moreover, MoA-based security services work with authorized participants in the IoV network. They expose a set of RESTfull web service APIs to handle user service requests and use local security interfaces to call functions of smart contracts. Figure 4 shows the whole process of video frame authentication can be divided into four stages.

3.3.1. Watermark Generation

For a street view, a surveillance camera S produces video streams for a traffic management system. At the same time, a CAV camera C also continuously produces drive-related video streams that record road situations. Here, we use F_i to represent a frame in a video recording $V = \{F_1, ..., F_n\}$ where n is the total number of frames. Given a frame F_i, FIAE-GAN-based encoder can generate a watermark W_{in} and output a watermarked frame $F_i^{'} = DW\_embed(F_i, W_{in})$. Through the FIAE-GAN watermark hiding process, a set of watermarked frames $F^{'} = \{F_1^{'}, ..., F_m^{'}\}$ can be published along with raw video data V. For efficiency consideration, a data owner embeds W_{in} in some frames for authentication purposes. Thus, the total number of watermarked frames $m \leq n$.

3.3.2. Data Publish

Figure 5-a demonstrate procedures in data publish and retrieve stages. A data owner calls a security interface to save video data into the DDS system and record meta-information on Ethereum Blockchain. After successfully uploading a watermarked frame $F_i^{'}$, the data owner can get a unique hash-based Content Identifier (CID) to address and retrieve data from the DDS system. IPFS publication protocol will distribute data among peers [37]. For each watermarked frame $F_i^{'}$, a meta-data $M_i^{'}$ contains its configuration information (e.g., CID, data format, and size) and verification proofs π (e.g., a hash value of frame and a signature of data owner).

To verify the integrity of watermarked frames $F^{'}$, we use Binary Merkle Tree (MKT) to represent a sequential list of meta-data $M^{'}$, as Figure 5-b shows. First, a security hash function converts $M_i^{'}$ to a digest $D_i^{'}$ located on a BMT leaf. By constructing a hierarchy of hash values growing from Merkle leaves until only a root node is left, a Merkle root $MT\_root = BMT(D(1), D(2), ...D(m))$ can be used to verify the integrity of $M^{'}$ as the whole. Therefore, the data owner calls smart contract functions to store $MT\_root(M^{'})$ and ordered list TX as on-chain data on the Ethereum Blockchain. Each TX_i only save small size of reference information about raw data, like identity information of data owner, D_i, and proof π. While watermarked watermarked frame $F_i^{'}$ along with its metadata $M_i^{'}$ are stored into off-chain storage. Therefore, using hybrid on-chain & off-chain storage for video data distribution promises to improve efficiency of integrating Blockchain without directly recording large size of raw data on the Blockchain. In addition, it also can ensure privacy preservation by not directly exposing video data and other sensitive information on transparent public distributed ledger.

 doi:10.20944/preprints202409.1228.v

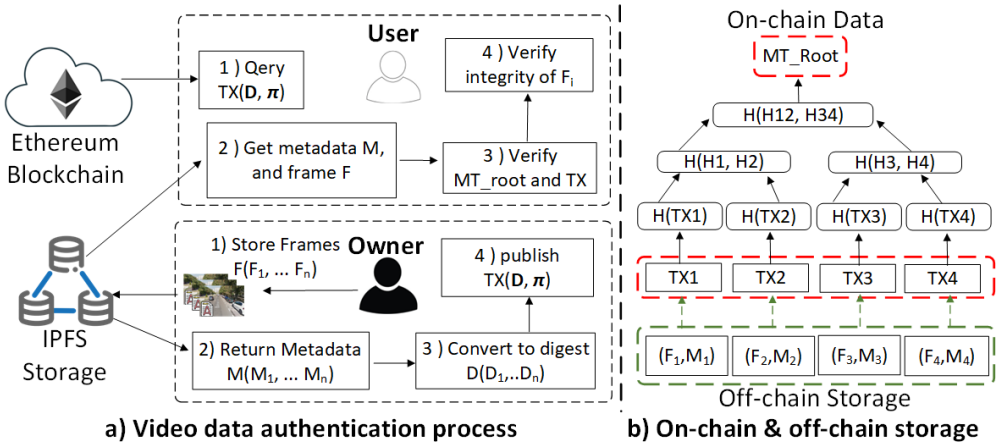

Figure 5. The illustration of video data authentication.

3.3.3. Data Retrieval

This stage focuses on querying video data and meta-information from the DDS system and verifying data integrity. As Figure 5-a shows, a data user calls smart contract functions to query a sequential list of TX and its proof MT_root. Then data user can use information in TX to retrieve list of metadata $M^{'}$ from IPFS storage. To verify integrity of a video segment containing $F^{'}$, he/she can easily reconstruct a Merkle tree of digest $D^{'}$ and calculate its root hash $MT\_root^{'}$. Any Modification on the sequential order of $M^{'}$ or content of $M_i^{'}$ will lead to a different root hash value MT_root of the Merkle tree. Thus, data integrity of the entire $M^{'}$ can efficiently verified by comparing $MT\_root^{'}$ with proof $MT\_root(M^{'})$ recorded on the distributed ledger. If verification on $M^{'}$ passed, data user can retrieve watermarked frames $F^{'}$ from DDS system by using CIDs saved in $M^{'}$. As IPFS relies on a P2P overlay network to discover all available locations from which objects can be retrieved, this can dramatically reduce the impact of availability issues in centralized storage systems (e.g., network congestion or single node failure). Finally, the data owner can verify the integrity of each $F_i^{'}$ by using its meta-information.

3.3.4. Watermark Verification

This stage aims to verify the authenticity of video data that may be used for malicious behavior analysis and vehicle forensics. A data user can extract watermarks from video frames and then verify that the watermarked frames were created by the actual owner. The FIAE-GAN-based decoder ensures that watermarked features can be captured effectively in extraction stage, therefore, watermark W_{out} can be rebuilt blindly without needing any key. Given a watermarked frame F_i, FIAE-GAN-based decoder can reconstruct its watermark $W_{out} = DW\_extract(F_i^{'})$. We assume that the data user uses security communication channels to receive W_{in} from the data owner. Finally, the authenticity of video data can be verified by comparing the similarity of W_{in} and W_{out}.

4. FIAE-GAN based Digital Watermarking Model

To achieve high invisibility and robustness of digital watermarks for video frames, our FIAE-GAN is a feature-integrated and attention-enhanced robust image watermarking model inspired by HiDDeN [24]. This section begins with an overview of FIAE-GAN's general structure and then describes each component in detail.

4.1. FIAE-GAN Watermarking Network Model

A robust watermarking model should ensure that the embedded watermark remains unaffected mainly by various image attacks and that the embedding involves invisible perturbations. To achieve

 doi:10.20944/preprints202409.1228.v

this, the model extracts robust features, integrates them with the watermark, and distributes them over inconspicuous areas to maintain invisibility.

As shown in Figure 6, FIAE-GAN primarily consists of the encoder E_α, the decoder D_β, the noised layers $Noise$, and the discriminator DIS_δ, where α, β, δ represent the trained parameters of the encoder, decoder, and discriminator, respectively. These parameters are continually updated during iterative training to achieve high watermarking invisibility and robustness. Given an original image F_{cover} of size $H \times W \times C$ and an original binary watermark W_{in} of length L, inputting F_{cover} into E_α generates the encoded image F_{en}.

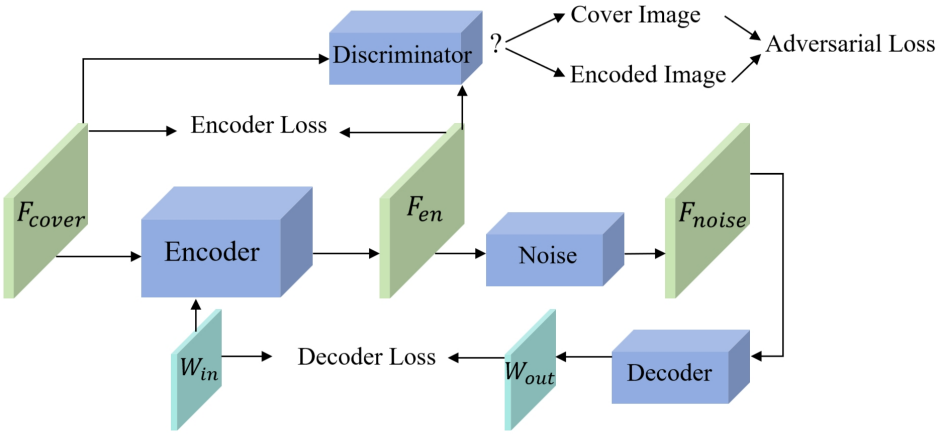

Figure 6. Overall architecture of the proposed FIAE-GAN. The FIAE-GAN is an end-to-end watermarking network designed to automatically generate watermarks with both invisibility and robustness. The key components of the model, indicated in blue boxes, include the encoder, decoder, noise subnetwork, and discriminator.

The discriminator DIS_δ evaluates the likelihood that F_{en} matches the original image by comparing the similarity between F_{cover} and F_{en}. The feedback from DIS_δ assists E_α in generating F_{en}. To ensure watermarking robustness, various types of noise are introduced in $Noise$. Concurrently, adversarial training is conducted where D_β is used to extract the decoded watermark W_{out} from F_{noise} aiming to make W_{out} as identical as possible to W_{in}.

4.2. The structure of encoder

The encoder E_α is designed to embed the watermark into the original image F_{cover} while preserving the image quality and ensuring robustness against various attacks. Additionally, a residual structure enhances training efficiency by incorporating a global residual skip connection. The formula for generating F_{en} is as follows:

$$F_{en} = E(F_{cover}, W_{in}) \tag{1}$$

where $E()$ denotes the process of encoding, as demonstrated in Figure 7. The encoder consists of two key modules: the Feature Integration Module (FIM) and the Attention-Enhanced Module (AEM). The FIM addresses the limitation of extracting only limited image features by capturing shallow and deep image features across multiple layers. Integrating these features enhances the encoder's ability to embed the watermark robustly. The feature X_W is obtained in the following manner:

$$X_W = E_{FIM}(F_{cover}, W_{in}) \tag{2}$$

where E_{FIM} denotes the process of the FIM. The structure of the FIM includes several dense blocks, each containing a sequence of Batch Normalization (BN), Rectified Linear Unit (ReLU) activation, and convolutional layers. The dense connections within the FIM allow for the reuse of these features, significantly boosting the encoder's representational capacity.

 doi:10.20944/preprints202409.1228.v

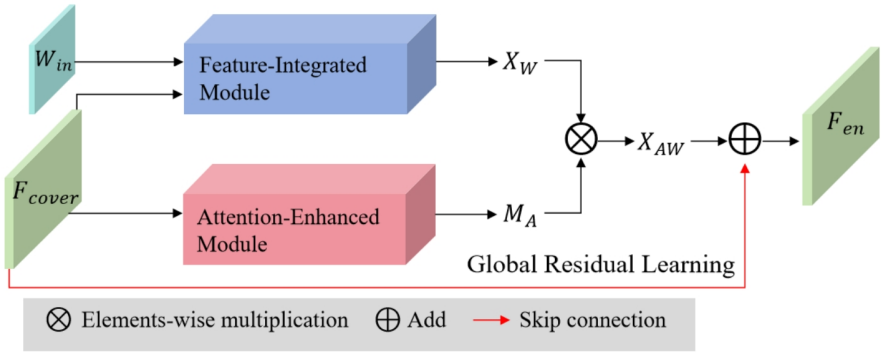

Figure 7. Architecture of encoder. The encoder includes (1) a Feature-Integrated Module (FIM) that utilizes dense connections to extract both shallow and deep features, which are then fused with the watermark to improve its robustness; (2) an Attention-Enhanced Module (AEM) that applies spatial attention to embed the watermark in less noticeable regions of the original image.

Watermarks embedded into various parts of an image impact distortion and robustness differently. Embedding the watermark in inconspicuous regions enhances visual image quality, while placing it in textured regions boosts robustness. We developed a specialized module called attention-enhanced module (AEM) to minimize the distortion from watermark embedding. This module enhances or suppresses specific regions within the spatial domain of the feature map based on their relevance while simultaneously capturing global image features to generate an attention mask, which dictates varying watermarking strengths across different image regions, ensuring a balance between invisibility and robustness. In addition, the attention mask ensures that flat and sensitive areas are suppressed. Meanwhile, inconspicuous and textured regions are intensified, achieving higher embedding strength and maintaining the invisibility of the watermark.

To be specific, based on the global features of F_{cover}, the AEM creates an attention mask M_A to aid in generating F_{en}, allowing it to adjust the features of F_{en} effectively, thereby minimizing image distortion caused by watermark embedding and reducing visibility perturbations. M_A is computed by

$$M_A = E_{AEM}(F_{cover}) \tag{3}$$

where $E_{AEM}()$ denotes the process of the AEM, which consists of a spatial attention mechanism. Then, we utilize the M_A to adjust the distribution of X_W. Finally, we adopt the global residual skip connection to generate F_{en} by

$$F_{en} = F_{cover} + X_W \times M_A \tag{4}$$

The encoding loss function L_E is designed to minimize the distance between F_{cover} and F_{en} by updating the parameter α. L_E is composed of the image reconstruction loss and the visual loss

$$L_E = \epsilon_1 MSE(F_{cover}, F_{en}) + \epsilon_2 SSIM(F_{cover}, F_{en}) \tag{5}$$

where $MSE()$ denotes the mean-square error function, and $SSIM()$ refers to the structural similarity index metric. The parameters ϵ_1 and ϵ_2 denote the weights for the image reconstruction loss and visual loss, respectively.

4.3. Watermark Extraction Decoder

The noise module *Noise* simulates various attacks during iterative training by incorporating differentiable noises to enhance watermarking robustness. This approach helps the network learn to embed watermarks in regions less prone to distortion and to develop a robust watermarking pattern

 doi:10.20944/preprints202409.1228.v

that resists both trained and untrained image noises. The noised image F_{noise} is created by applying various types of noise

$$F_{noise} = Noise(F_{en}, NO_{train}) \tag{6}$$

where NO_{train} is a trained noise. Subsequently, the decoder D_β is employed to extract W_{out} from F_{noise}

$$W_{out} = D_\beta(F_{noise}) \tag{7}$$

Decoder training enhances watermarking robustness by minimizing the difference between W_{out} and W_{in} through updating parameter β. The decoding loss L_D is defined as:

$$L_D = \frac{\sqrt{(W_{in} - W_{out})^2}}{L} \tag{8}$$

Watermark extraction is the inverse of embedding, utilizing a similar structure as the FIM to recover the watermark from encoded image features. The decoder employs dense connections for deep feature extraction to obtain the final watermark. These techniques provide error tolerance, ensuring accurate watermark extraction despite noise-induced distortion. However, excessive distortion can suppress decoding ability. Feedback from the decoder helps the encoder adjust features during noise training to avoid fragile regions, ensuring effective watermark extraction without the original image.

4.4. Discriminator

The discriminator DIS_δ attempts to distinguish the encoded image F_{en} from the original image F_{cover}. At the same time, the encoder aims to deceive DIS_δ by making F_{en} resemble F_{cover}, preventing accurate discrimination. Essentially, DIS_δ checks for the presence of a watermark in the encoded image. This adversarial relationship enhances the encoded image quality through iterative training until Nash equilibrium is achieved.

Training FIAE-GAN by reducing total loss, comprising encoding, decoding, and adversarial losses, involves updating the encoder and decoder parameters by minimizing their losses. The encoder's parameters are influenced by both encoding and decoding losses to embed the watermark in imperceptible and robust areas. If image features are not robust against noise, the decoder's feedback prompts the encoder to adjust. The discriminator updates its parameters using adversarial loss to evaluate image quality and guide the encoder to improve watermark invisibility. Consequently, the encoder, decoder, and discriminator are trained together to achieve optimal performance.

5. Experimental Results and Discussions

This section first introduces experimental configuration for watermarking model training and prototype video data authentication framework implementation. Then, we present a comprehensive evaluation of the performance of the FIAE-GAN watermarking model. We also evaluate latency incurred by different operation stages in the video data authentication process and analyze their financial cost. Finally, we discuss the security features of the proposed solutions that can prevent against diverse attacks on video data service platforms under distributed network environments.

5.1. Prototype Implementation and Experiment Configuration

The proposed watermarking model is trained on the Chinese Traffic Sign Detection Benchmark (CCTSDB) dataset [38], a comprehensive collection designed for evaluating traffic sign detection and recognition algorithms. The CCTSDB consists of 10,000 images for training and the remaining 5,000 for testing to ensure the generalization of the trained model. The model is implemented using PyTorch and executed on an NVIDIA Tesla V100 GPU. All images are resized to $512 \times 512 \times 3$. The performance of the watermarking model is evaluated using the Peak Signal-to-Noise Ratio (PSNR) and Structural

 doi:10.20944/preprints202409.1228.v

Similarity Index Metric (SSIM) to assess watermarking invisibility. In this paper, we utilize both subjective and objective evaluations to thoroughly assess the performance of the watermarking model.

We also implemented a proof-of-concept prototype for video data authentication using Python. Flask [39] is used to develop RESTful APIs for security services. We use a standard Python library cryptography [40] to develop all cryptographic primitives, such as digital signature and hash function (SHA-256). Smart contracts are implemented by using Solidity [41]. All smart contracts are deployed on a private Ethereum network where all participants use Go-Ethereum [42] as client applications to interact with smart contracts. A private Ethereum network consists of 6 miners, and each containerized miner is assigned a single CPU core. The prototype is set up on a ThinkPad P16 equipped with a 13th-generation Intel(R) Core(TM) i7-1370P (20 cores) and 64 GB memory.

5.2. Subjective Evaluation of the proposed Watermarking Model

Subjective evaluation aims to ensure that the watermarked images preserve high visual quality as perceived by human observers while considering factors such as image clarity, overall quality, and the presence of any distortions. Figure 8 demonstrates the subjective invisibility of FIAE-GAN, presenting both the original and the encoded images. As shown in Figure 8 (a) and (b), the original and encoded images are visually indistinguishable, indicating that the watermark has been successfully embedded in imperceptible areas.

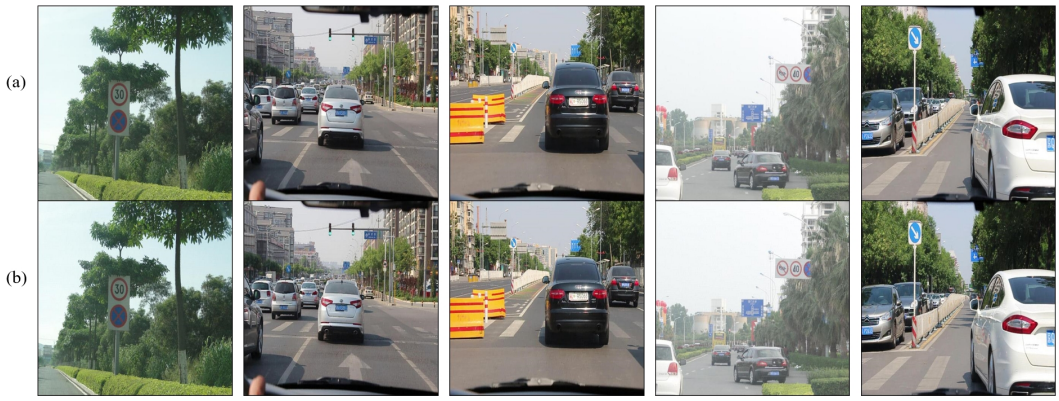

Figure 8. Watermarking performance of FIAE-GAN. (a) Original image. (b) Encoded image.

Additionally, Figure 9 presents a comparison of the histograms of the cover and watermarked images to evaluate the subjective performance of the proposed scheme. This comparison examines the distribution of pixel values in the watermarked image relative to the original image to gauge the extent of distortion caused by the watermarking process. The results show minimal variation in pixel values between the two images, indicating high visual quality with negligible distortion.

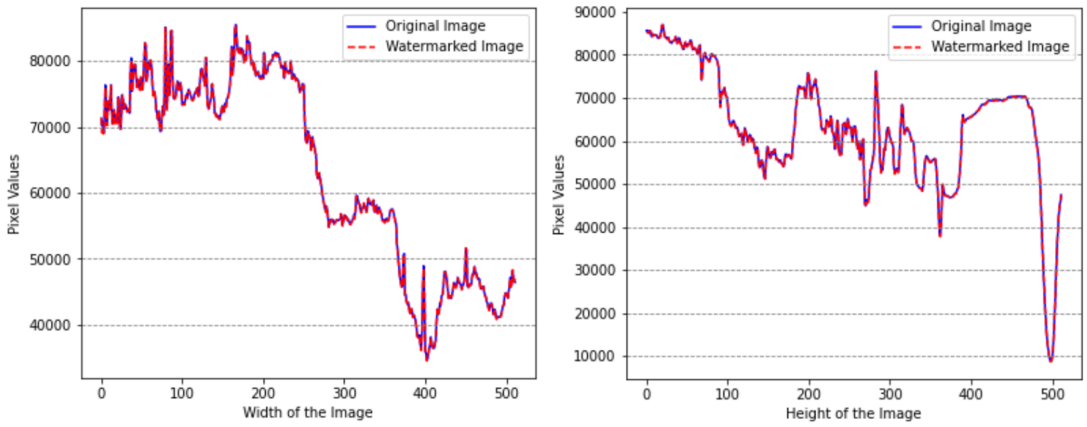

Figure 9. Subjective evaluation through histogram comparison of the original and watermarked images.

 doi:10.20944/preprints202409.1228.v

Furthermore, Scale-Invariant Feature Transform (SIFT) is employed to assess the effectiveness of the watermarking technique by comparing the feature descriptors of the original and watermarked images [43,44] . SIFT is a widely used computer vision algorithm that detects and describes local features in an image by identifying key points such as edges, corners, and blobs. Its ability to generate robust, repeatable descriptors invariant to scale, rotation, and lighting conditions makes it ideal for assessing the impact of watermarking on the original image content. As illustrated in Figure 10, the results demonstrate a significant number of matching descriptors, with minimal distortion observed between the images.

Figure 10. Subjective evaluation through SIFT feature matching between the original and watermarked images.

5.3. Objective Evaluation of the proposed Watermarking Model

Objective evaluation employs quantifiable metrics, such as PSNR, SSIM and analysis of watermarking robustness, to rigorously assess the performance of watermarking models. These metrics offer a standardized approach for evaluating image quality and structural preservation, thereby ensuring consistent and reproducible measurements. Utilizing these measures allows for a precise assessment of the impact of watermarking on image fidelity and facilitates comparisons between different methods. Table 1 presents the PSNR and SSIM values for the watermarking models, which are calculated by averaging the results from 5,000 encoded images in the CCTSDB dataset. Five existing state-of-the-art (SOTA) watermarking models are used for comparison: HiDDeN [24], ReDMark [22], DA [45], TSDL [46], and MBRS [47]. We attempted to replicate the experiments conducted by DA and TSDL, but we were unable to fully reproduce their performance. Therefore, we use the published results from their respective articles for comparison. Additionally, MBRS provides a pre-trained model, and we use their test results for evaluation. To ensure a fair comparison, the binary watermark length is set to $L = 30$ for all the models under consideration.

As shown in Table 1, our FIAE-GAN model outperforms the other five SOTA watermarking models regarding SSIM. Specifically, FIAE-GAN achieves a PSNR of 35.89, 3.75 dB higher than HiDDeN's PSNR of 32.14 and 2.19 dB higher than DA's PSNR of 33.70. Compared to MBRS, which has a PSNR of 35.08, FIAE-GAN demonstrates an improvement of 0.81 dB. Regarding SSIM, FIAE-GAN achieves a value of 0.9679, surpassing HiDDeN's SSIM of 0.9315 by 0.0364. Compared to MBRS, which has an SSIM of 0.8914, FIAE-GAN's SSIM is 0.0765 higher, indicating that FIAE-GAN better preserves image details and overall visual quality. In addition, both ReDMark and FIAE-GAN methods perform almost equally well in terms of image quality retention after watermarking, with ReDMark having a marginally better PSNR. However, FIAE-GAN outperforms RedMark in terms of SSIM, suggesting that our method better preserves the structural similarity of the original image after watermarking.

 doi:10.20944/preprints202409.1228.v

Table 1. Comparison of different models on CCTSDB.

Model	PSNR	SSIM
HiDDeN [24]	32.14	0.9315
TSDL [46]	33.50	-
DA [45]	33.70	-
MBRS [47]	35.08	0.8914
ReDMark [22]	35.93	0.9660
FIAE-GAN	35.89	0.9679

To evaluate the robustness of the proposed watermarking model, the bit accuracy (BA) is calculated by averaging the results across all tested images. The BA results are presented in Figure 11, where a comparative analysis of the proposed method is conducted against HiDDeN [24], TSDL [46], MBRS [47] and ReDMark [22] under various types of noise attacks. The average BA values of HiDDeN, TSDL, MBRS, ReDMark, and our FIAE-GAN are 0.7969, 0.7984, 0.9551, 0.8534, and 0.9794, respectively, which demonstrates our FIAE-GAN significantly outperforms the other four models. Specifically, Compared to HiDDeN, FIAE-GAN demonstrates a 22.9% improvement, while against TSDL, the improvement is approximately 22.6%. In comparison to ReDMark, FIAE-GAN shows a 14.8% increase in performance. Even though MBRS achieves a BA of 0.9551, FIAE-GAN still surpasses it by 2.5%, making it the most robust model among all tested. It is noteworthy that, with regard to JPEG compression and Gaussian noise, our FIAE-GAN demonstrates superior robustness compared to HiDDeN, TSDL, and ReDMark.

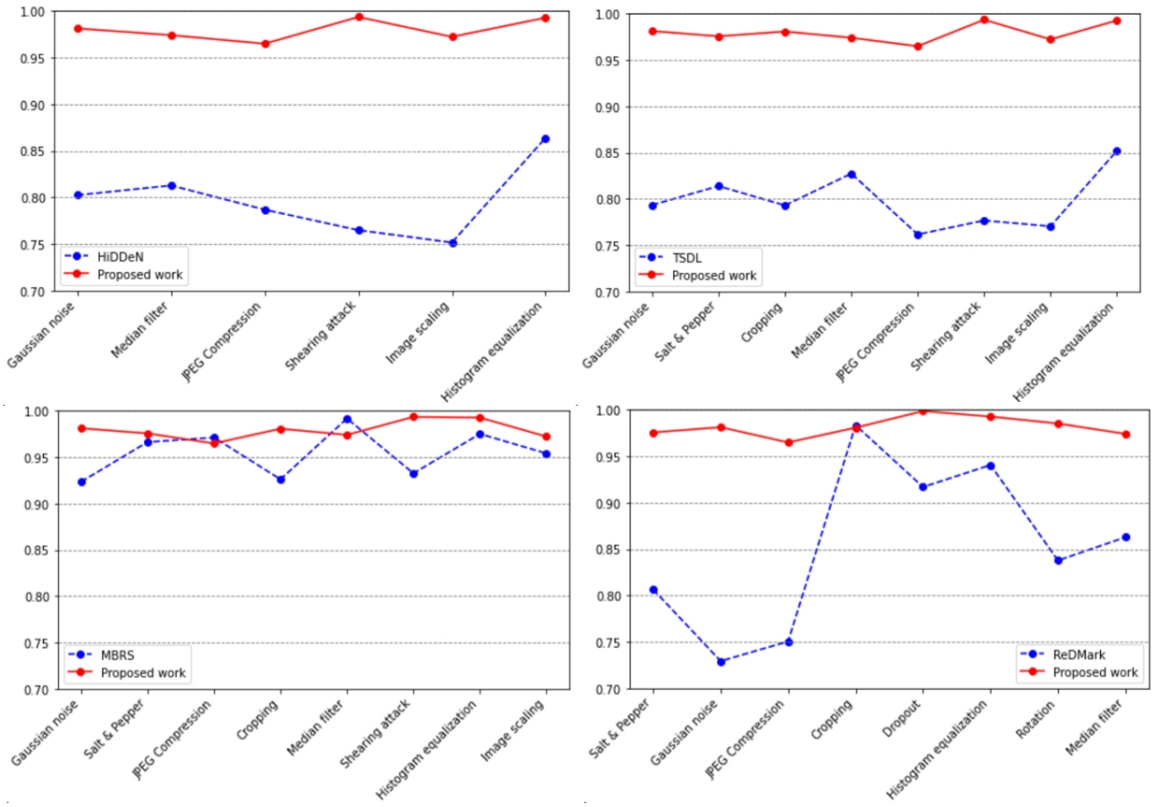

Figure 11. Comparative analysis of proposed work with HiDDeN [24], TSDL [46], MBRS [47], ReDMark [22].

 doi:10.20944/preprints202409.1228.v

For a watermarking system to be practical, it must be robust to a broad spectrum of image noises, extending beyond those encountered during training. To evaluate its generalization capability against diverse image attacks, we test the system using twelve types of untrained noises: Gaussian noise, JPEG Compression, Cropping, Dropout, Salt and Pepper noise, Rotation, Median filter, Adjust Brightness, Adjust Contrast, Image scaling, Shearing attack and Histogram equalization. In the robustness assessment of the suggested work presented in Table 2, our FIAE-GAN model demonstrates superior performance compared to HiDDeN across various noise attacks. Specifically, for Gaussian noise, the FIAE-GAN model exhibits nearly perfect performance with an accuracy of 96.85% at a noise density of 0.10, whereas HiDDeN's accuracy significantly decreases to 76.81%. Correspondingly, in the context of JPEG compression, a notable difference is observed: HiDDeN achieves an accuracy of 73.39% at a quality factor (QF) of 10, while FIAE-GAN performs markedly better with an accuracy of 94.12%. The performance gaps are particularly pronounced in several attack scenarios. Under a Dropout attack with a dropout rate of 0.3, FIAE-GAN achieves an accuracy of 99.87%, compared to HiDDeN's accuracy of 87.91%. Similarly, during a shearing attack with parameters [0.4, 0.4], FIAE-GAN attains an accuracy of 99.36%, whereas HiDDeN's accuracy is considerably lower at 76.47%. In the case of rotation at 90°, HiDDeN's accuracy is only 72.94%, whereas FIAE-GAN achieves a significantly higher accuracy of 98.12%. For Salt and Pepper noise at a density of 0.1, HiDDeN's accuracy is 77.56%, while FIAE-GAN maintains a substantially higher accuracy of 97.57%. Furthermore, with contrast adjustment at a factor of 2.0, HiDDeN's accuracy drops to 72.45%, compared to FIAE-GAN's much higher accuracy of 96.87%. Overall, FIAE-GAN demonstrates robust and reliable performance across various types of noise attacks, showing particularly significant improvements over HiDDeN in handling high noise densities and challenging distortions.

Table 2. Robustness assessment of the suggested work.

Attack	Noise density	HiDDeN [.%]	FIAE-GAN [.%]
Gaussian noise	0.001	85.57	100.00
	0.05	80.23	98.13
	0.10	76.81	96.85
JPEG Compression	QF=10	73.39	94.12
	QF=50	78.67	96.49
	QF=90	95.51	99.21
Cropping	[20, 20, 420, 420]	78.82	98.08
Dropout	0.3	87.91	99.87
Salt & Pepper	0.001	91.34	99.84
	0.05	82.73	99.31
	0.1	77.56	97.57
Rotation	45°	78.81	98.53
	90°	72.94	98.12
Median filter	[2, 2]	88.39	99.64
	[3,3]	81.27	97.41
Adjust Brightness	1.1	92.61	98.94
	1.3	86.38	97.37
Adjust Contrast	1.0	82.92	97.52
	2.0	72.45	96.87
Image scaling	0.5	75.14	97.23
	2	78.45	96.61
Shearing attack	[0.4, 0.4]	76.47	99.36
Histogram equalization	1.0	86.37	99.28

 doi:10.20944/preprints202409.1228.v

5.4. Performance and Cost of Video Data Authentication

We use time latency to evaluate the performance of a data authentication scheme containing six stages. We evaluate the processing time of watermark embedding and extraction. In addition, we also evaluate the end-to-end delays incurred by the data distribution and retrieval stages. The data distribution procedure is divided into two steps: i) save video frames into the IPFS network and prepare meta-data containing CIDs and Merkel root; ii) trigger smart contracts transactions to record them on Ethereum Blockchain. Similarly, the whole data retrieval procedure contains two steps: i) call smart contracts to get meta-data; ii) verify Merkel root and then retrieve frames from the IPFS network. We conducted 50 Monte Carlo test runs for each test scenario and used the averages to measure the results.

Table 3 shows time latency incurred by key steps in the video data authentication process. The watermark embedding loads a cover frame and then hides a binary watermark mixed with features. Thus, the watermark embedding program takes an average of 1.09 sec to create a watermarked frame on the owner's side. The watermark extraction loads a watermarked frame and then rebuilds a binary watermark. It takes an average of 0.62 sec to verify a watermarked frame on the user side. We use 10 sequential watermarked frames (about 110 KB per frame) to evaluate data distribution and retrieval time latency. Data owners (users) simply push (pull) frames through their local peers within an IPFS network. Uploading and downloading data on a DDS system only introduces small delays (about 0.5 s). The latency of recording metadata through smart contracts depends on block confirmation time, which is greatly impacted by the consensus protocol (Proof-of-Work used by our test Blockchain network). Thus, it takes about 5.2 sec to save metadata on Blockchain. In contrast, visiting the local storage of smart contracts and querying meta information on Blockchain takes much less time. Although watermark embedding and data distribution cause more delays (6.69 sec) than verification (0.68 sec), these overheads only occur once during data generation. In addition, our solution demonstrates efficiency in data verification procedures, which is important for scenarios that investigate and analyze large volumes of video data.

Table 3. Latency of data authentication process (Seconds).

Stage	**1**	**2**	**3**	**4**	**5**	**6**
Latency	1.09	0.43	5.17	0.04	0.02	0.62

Authentication Stage: 1: Watermark Embedding; 2: Upload (push) Data onto IPFS; 3: Record Meta on Blockchain; 4: Query Meta on Blockchain; 5: Download (pull) Data from IPFS; 6: Watermark Extraction.

We use gas fees to evaluate cost of executing data authentication operations on Ethereum. Table 4 provide summary about cost incurred by data authentication scheme. In Ethereum, gas is the fee paid by the user who launch transactions to update state of smart contracts. The amount of gas used a smart contract transaction depends on the complexity of a transaction. We only evaluate cost in smart contract deployment and transactions that update smart contract, like storing TX and proofs during data publish stage. Because 1 ETH = 1 billion Gwei in Ethereum, we can present cost in Ether unit by dividing the total gas used by 1,000,000,000. Given the approximate exchange rate at the time of writing (1 Ether = $2,357.76), we convert Ether cost to U.S. dollars.

Table 4. Cost of data authentication scheme.

	Gas Used	**Cost (Ether)**	**Cost ($)**
Deploy Smart Contract	190,573	0.000191	0.45
Update Smart Contract	875,974	0.000876	2.07

 doi:10.20944/preprints202409.1228.v

5.5. Security Analysis

Our solution embeds invisible and robust watermarks in video frames at the data source side. Thus, data users can easily extract watermarks and verify the authenticity of video data at the destination side. Because an adversary cannot control watermark embedding and extraction procedures, he/she cannot create fake frames containing valid watermarks. The security features of the digital watermarking scheme can effectively protect against forgery attacks in IoV networks. In addition, metadata containing audit-proof information is stored on the immutability distributed ledger. An adversary cannot tamper with these on-chain data without controlling the majority (51%) of miners within an Ethereum network. Even data owners cannot insert, replace, and delete watermarked frames that have been published. A Merkel tree structure of metadata can easily detect any content inconsistency. This can prevent integrity attacks to both on-chain and off-chain data. Furthermore, a DDS system atop an IPFS network relies on a distributed hash table (DHT)-based P2P protocol for data distribution and storage among distributed sites. It's difficult for an adversary to disable data service by launching DoS/DDoS attacks on some sites. As a result, our solution can mitigate the impact of availability attacks.

6. Conclusions and Future Work

This paper proposes D2WaVe, a novel decentralized digit watermarking framework to guarantee the security and auditability of video data in IoV networks. Under the umbrella, an FIAE-GAN model, an end-to-end framework for digital watermark embedding and extraction, has been introduced. FIAE-GAN addresses existing deep learning-based watermarking model limitations by incorporating an FIM and an AEM within the encoder. By integrating the DDS system with the Ethereum Blockchain, BloVA, a decentralized video frames authentication scheme, can guarantee the authenticity and integrity of data at rest and in transit. Experimental results demonstrate that the FIAE-GAN model achieves high image quality and outperforms SOTA models in invisibility. The proposed D2WaVe framework incurs small latency, and it is promising for mitigating threats to data integrity and availability in IoV networks.

Meanwhile, open questions still need to be addressed before our D2WaVe framework is applied to real-world IoV scenarios. FIAE-GAN has high computational costs and limited embedding capacity. Future work will focus on developing lightweight network modules to reduce computation time and further improve watermarking performance. In addition, our solution cannot guarantee privacy preservation for video data storage on DDS and protect against identity link attacks in data exchange. In future work, we will explore privacy-enhanced methods (e.g., zero-knowledge proof) for video data sharing.

Author Contributions: Conceptualization, X.L., R.X and Y.C.; methodology, X.L. and R.X.; software, X.L. and R.X.; validation, X.L., R.X and Y.C.; formal analysis, X.L. and R.X.; investigation, X.L., R.X and Y.C.; resources, X.L. and R.X.; data curation, X.L. and R.X.; writing—original draft preparation, X.L. and R.X.; writing—review and editing, X.L., R.X and Y.C.; visualization, X.L. and R.X.; supervision, R.X.; project administration, R.X.; funding acquisition, R.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by Michigan Technological University - ICC Rapid Seeding Awards.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AEM	Attention-Enhanced Module
BN	Batch Normalization
CAV	Connected Automated Vehicles
CCTSDB	Chinese Traffic Sign Detection Benchmark

 doi:10.20944/preprints202409.1228.v

DIW	Digital Image Watermarking
DLT	Distributed Ledger Technology
DFT	Discrete Fourier Transform
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
ECU	Electronic Control Units
FIAE-GAN	Feature-Integrated and Attention-Enhanced Model Based on GAN
FIM	Feature-Integrated Module
GAN	Generative Adversarial Network
GANs	Generative Adversarial Networks
IoV	Internet of Vehicles
ITS	Intelligent Transportation System
MOA	Microservices-Oriented Architecture
P2P	Peer-to-Peer
PSNR	Peak Signal-to-Noise Ratio
PUF	Physically Unclonable Function
ReLU	Rectified Linear Unit
SIFT	Scale-Invariant Feature Transform
SC	Smart Contract
SOA	Service-Oriented Architecture
SSIM	Structural Similarity Index Metric
V2G	Vehicle-to-grid
V2I	Vehicle-to-Infrastructure
V2R	Vehicle-to-Roadside Unit
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything

References

1. Ji, B.; Zhang, X.; Mumtaz, S.; Han, C.; Li, C.; Wen, H.; Wang, D. Survey on the internet of vehicles: Network architectures and applications. *IEEE Communications Standards Magazine* **2020**, *4*, 34–41.
2. Xu, R.; Nikouei, S.Y.; Nagothu, D.; Fitwi, A.; Chen, Y. Blendsps: A blockchain-enabled decentralized smart public safety system. *Smart Cities* **2020**, *3*, 928–951.
3. Xu, R.; Nagothu, D.; Chen, Y. AR-Edge: Autonomous and Resilient Edge Computing Architecture for Smart Cities **2024**. pp. 1–21.
4. Mollah, M.B.; Zhao, J.; Niyato, D.; Guan, Y.L.; Yuen, C.; Sun, S.; Lam, K.Y.; Koh, L.H. Blockchain for the internet of vehicles towards intelligent transportation systems: A survey. *IEEE Internet of Things Journal* **2020**, *8*, 4157–4185.
5. Nin, J.; Ricciardi, S. Digital watermarking techniques and security issues in the information and communication society. In Proceedings of the 2013 27th International Conference on Advanced Information Networking and Applications Workshops. IEEE, 2013, pp. 1553–1558.
6. Li, J.; Song, Z.; Zhang, Z.; Li, Y.; Cao, C. In-Vehicle Digital Forensics for Connected and Automated Vehicles With Public Auditing. *IEEE Internet of Things Journal* **2023**.
7. Xu, R.; Nagothu, D.; Chen, Y.; Aved, A.; Ardiles-Cruz, E.; Blasch, E. A Secure Interconnected Autonomous System Architecture for Multi-Domain IoT Ecosystems. *IEEE Communications Magazine* **2024**, *62*, 52–57.
8. Nagothu, D.; Xu, R.; Nikouei, S.Y.; Chen, Y. A microservice-enabled architecture for smart surveillance using blockchain technology. In Proceedings of the 2018 IEEE international smart cities conference (ISC2). IEEE, 2018, pp. 1–4.
9. Xu, R.; Chen, Y. μDFL: A secure microchained decentralized federated learning fabric atop IoT networks. *IEEE Transactions on Network and Service Management* **2022**, *19*, 2677–2688.
10. Benet, J. Ipfs-content addressed, versioned, p2p file system. *arXiv preprint arXiv:1407.3561* **2014**.
11. Wu, W.; Dai, J.; Huang, H.; Zhao, Q.; Zeng, G.; Li, R. A digital watermark method for in-vehicle network security enhancement. *IEEE Transactions on Vehicular Technology* **2023**, *72*, 8398–8408.

 doi:10.20944/preprints202409.1228.v

12. Changalvala, R.; Malik, H. LiDAR data integrity verification for autonomous vehicle. *IEEE Access* **2019**, *7*, 138018–138031.
13. Wang, C.; Cheng, X.; Li, J.; He, Y.; Xiao, K. A survey: Applications of blockchain in the internet of vehicles. *EURASIP Journal on wireless communications and networking* **2021**, *2021*, 1–16.
14. Ko, H.J.; Huang, C.T.; Horng, G.; Shiuh-Jeng, W. Robust and blind image watermarking in DCT domain using inter-block coefficient correlation. *Information Sciences* **2020**, *517*, 128–147.
15. Liu, J.; Huang, J.; Luo, Y.; Cao, L.; Yang, S.; Wei, D.; Zhou, R. An optimized image watermarking method based on HD and SVD in DWT domain. *IEEE Access* **2019**, *7*, 80849–80860.
16. Jin, T.; Zhang, W. A novel interpolated DFT synchrophasor estimation algorithm with an optimized combined cosine self-convolution window. *IEEE Transactions on Instrumentation and Measurement* **2020**, *70*, 1–10.
17. Ding, S.; Zhang, L.; Pan, M.; Yuan, X. PATROL: Privacy-Oriented Pruning for Collaborative Inference Against Model Inversion Attacks. In Proceedings of the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 4716–4725.
18. Zhou, X.; Lei, X.; Yang, C.; Shi, Y.; Zhang, X.; Shi, J. Handling Data Heterogeneity for IoT Devices in Federated Learning: A Knowledge Fusion Approach. *IEEE Internet of Things Journal* **2023**.
19. Liu, X.; Liu, Z.; Chatterjee, S.; Portfleet, M.; Sun, Y. Understanding human behaviors and injury factors in underground mines using data analytics. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2021, pp. 2459–2462.
20. Chen, B.; Zhang, Z.; Li, Y.; Lu, G.; Zhang, D. Multi-label chest X-ray image classification via semantic similarity graph embedding. *IEEE Transactions on Circuits and Systems for Video Technology* **2021**, *32*, 2455–2468.
21. Mun, S.M.; Nam, S.H.; Jang, H.U.; Kim, D.; Lee, H.K. A robust blind watermarking using convolutional neural network. *arXiv preprint arXiv:1704.03248* **2017**.
22. Ahmadi, M.; Norouzi, A.; Karimi, N.; Samavi, S.; Emami, A. ReDMark: Framework for residual diffusion watermarking based on deep networks. *Expert Systems with Applications* **2020**, *146*, 113157.
23. Huang, Z.; Zhang, J.; Zhang, Y.; Shan, H. DU-GAN: Generative adversarial networks with dual-domain U-Net-based discriminators for low-dose CT denoising. *IEEE Transactions on Instrumentation and Measurement* **2021**, *71*, 1–12.
24. Zhu, J.; Kaplan, R.; Johnson, J.; Fei-Fei, L. Hidden: Hiding data with deep networks. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018, pp. 657–672.
25. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
26. Zhang, K.; Guo, Y.; Wang, X.; Yuan, J.; Ding, Q. Multiple feature reweight densenet for image classification. *IEEE access* **2019**, *7*, 9872–9880.
27. Adegun, A.A.; Viriri, S. FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images. *IEEE Access* **2020**, *8*, 150377–150396.
28. Liu, X.; Zhao, C. AGFA-Net: Attention-Guided and Feature-Aggregated Network for Coronary Artery Segmentation using Computed Tomography Angiography. *arXiv preprint arXiv:2406.08724* **2024**.
29. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
30. Yu, C. Attention based data hiding with generative adversarial networks. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2020, Vol. 34, pp. 1120–1128.
31. Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system. Technical report, Manubot, 2019.
32. Kang, J.; Yu, R.; Huang, X.; Wu, M.; Maharjan, S.; Xie, S.; Zhang, Y. Blockchain for secure and efficient data sharing in vehicular edge computing and networks. *IEEE internet of things journal* **2018**, *6*, 4660–4670.
33. Javaid, U.; Aman, M.N.; Sikdar, B. DrivMan: Driving trust management and data sharing in VANETS with blockchain and smart contracts. In Proceedings of the 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring). IEEE, 2019, pp. 1–5.
34. Shi, K.; Zhu, L.; Zhang, C.; Xu, L.; Gao, F. Blockchain-based multimedia sharing in vehicular social networks with privacy protection. *Multimedia Tools and Applications* **2020**, *79*, 8085–8105.
35. Nagothu, D.; Xu, R.; Chen, Y.; Blasch, E.; Aved, A. Defakepro: Decentralized deepfake attacks detection using enf authentication. *IT Professional* **2022**, *24*, 46–52.

 doi:10.20944/preprints202409.1228.v

36. Xu, R.; Chen, Y.; Chen, G.; Blasch, E. SAUSA: Securing Access, Usage, and Storage of 3D Point CloudData by a Blockchain-Based Authentication Network. *Future Internet* **2022**, *14*, 354.
37. Trautwein, D.; Raman, A.; Tyson, G.; Castro, I.; Scott, W.; Schubotz, M.; Gipp, B.; Psaras, Y. Design and evaluation of IPFS: A storage layer for the decentralized web. In Proceedings of the Proceedings of the ACM SIGCOMM 2022 Conference, 2022, pp. 739–752.
38. Icezero00. CCTSDB-YOLO, 2022.
39. Flask: A Pyhon Microframework. [Online]. Available: https://flask.palletsprojects.com/. Accessed on September 2022.
40. pyca/cryptography documentation. [Online]. Available: https://cryptography.io/. Accessed on September 2022.
41. Solidity. https://docs.soliditylang.org/en/v0.8.13/. Accessed on September 2022.
42. Go-ethereum. https://ethereum.github.io/go-ethereum/. Accessed on September 2022.
43. Bellavia, F.; Colombo, C. Is there anything new to say about SIFT matching? *International journal of computer vision* **2020**, *128*, 1847–1866.
44. Pele, O.; Werman, M. A linear time histogram metric for improved sift matching. In Proceedings of the Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part III 10. Springer, 2008, pp. 495–508.
45. Luo, X.; Zhan, R.; Chang, H.; Yang, F.; Milanfar, P. Distortion agnostic deep watermarking. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 13548–13557.
46. Liu, Y.; Guo, M.; Zhang, J.; Zhu, Y.; Xie, X. A novel two-stage separable deep learning framework for practical blind watermarking. In Proceedings of the Proceedings of the 27th ACM International conference on multimedia, 2019, pp. 1509–1517.
47. Jia, Z.; Fang, H.; Zhang, W. Mbrs: Enhancing robustness of dnn-based watermarking by mini-batch of real and simulated jpeg compression. In Proceedings of the Proceedings of the 29th ACM international conference on multimedia, 2021, pp. 41–49.