Deep Data Hiding for ICAO-Compliant Face Images: A Survey

Jefferson David Rodriguez Chivata[1], Davide Ghiani[1], Simone Maurizio La Cava[1],
Marco Micheletto[1], Giulia Orrú[1], Federico Lama[2], Gian Luca Marcialis[1]
[1]University of Cagliari, Piazza d'Armi I - 09123 Cagliari (Italy),
e-mail: {jeffersond.rodriguez,davide.ghiani,simonem.lac,
marco.micheletto,giulia.orru,marcialis}@unica.it
[2]Dedem S.p.A., Via Cancelleria 59 - 00072 Ariccia (Italy), e-mail: federico.lama@dedem.it

Abstract

ICAO-compliant facial images, initially designed for secure biometric passports, are increasingly becoming central to identity verification in a wide range of application contexts, including border control, digital travel credentials, and financial services. While their standardization enables global interoperability, it also facilitates practices such as morphing and deepfakes, which can be exploited for harmful purposes like identity theft and illegal sharing of identity documents. Traditional countermeasures like Presentation Attack Detection (PAD) are limited to real-time capture and offer no post-capture protection. This survey paper investigates digital watermarking and steganography as complementary solutions that embed tamper-evident signals directly into the image, enabling persistent verification without compromising ICAO compliance. We provide the first comprehensive analysis of state-of-the-art techniques to evaluate the potential and drawbacks of the underlying approaches concerning the applications involving ICAO-compliant images and their suitability under standard constraints. We highlight key trade-offs, offering guidance for secure deployment in real-world identity systems.

1. Introduction

Biometric face recognition is central to identity management systems, particularly when secure and standardized verification is required. The International Civil Aviation Organization (ICAO) defines detailed specifications for facial image acquisition and formatting, which have been adopted globally in Machine Readable Travel Documents (MRTDs) such as biometric passports, and more recently in Digital Travel Credentials (DTCs) [47, 69]. These standards are also increasingly used in remote identity verification systems, including financial services, where face images are employed to meet Know Your Customer (KYC) requirements [26]. While these standards ensure uniformity and interoperability, their predictable structure can be exploited to create manipulated yet compliant images through morphing [13, 16], generative methods [78], and other image-based spoofing strategies [14]. Furthermore, the long validity of ICAO images and the potential exfiltration through data breaches and public dissemination raise concerns about unauthorized reuse and biometric privacy [38, 49].

Presentation Attack Detection (PAD) is a common defense against spoofing, but it operates only at capture time, without any protection once the image has been extracted, stored, or redistributed, and is limited by poor generalisation to novel attacks [48, 52, 55]. Moreover, they often rely on additional sensors or computational modules, which may reduce throughput in real-time scenarios such as border control [7, 25]. To address these limitations, proactive data hiding techniques have been explored for encoding and embedding verifiable information directly into the image in a controlled and application-specific manner [68]. These approaches aim to ensure integrity verification and traceability independently of the acquisition environment [8].

Within this domain, digital watermarking and steganography represent two established paradigms [68]. The former is primarily used to assert integrity or provenance, while the latter was traditionally developed to conceal auxiliary information for covert communication. Both techniques rely on modifying the visual signal to embed data that can be extracted later, typically without compromising the usability or appearance of the host image. In biometric contexts, these methods have been increasingly considered for embedding integrity markers or identifiers that remain robust under common transformations such as compression while preserving recognition performance and visual conformity with standard requirements [57, 86].

Recent advances in deep learning have significantly improved data hiding methods [68], leveraging encoder-decoder networks [87], generative adversarial networks (GANs) [83], transformers [88], invertible neural networks

(INNs) and diffusion models [72] to increase capacity, imperceptibility, and robustness. While some have been explored for biometric ID systems [44], no survey has yet analyzed their applicability to ICAO-constrained biometric images. Existing surveys typically address watermarking and steganography from the perspective of general media security, copyright protection, or covert communication without considering the constraints imposed by biometric standards and operational requirements.

This survey fills that gap by systematically organizing and interpreting modern data-hiding approaches for post-acquisition certification under ICAO constraints. Rather than replicating existing empirical results, we extract and recontextualize key insights from prior work, enabling a new comparative understanding of the strengths, limitations, and suitability of current methods in biometric certification scenarios. Specifically, we provide: (i) a taxonomy of data hiding approaches categorized by robustness, architecture, and learning paradigm and their analysis in function of the risks associated with the real-world applications of ICAO-compliant images; (ii) a comparative analysis of recent watermarking and steganographic models based on such application scenarios; (iii) a critical discussion of their applicability for tamper detection and their potential role for ICAO scenarios. Although the focus is on ICAO-compliant images, the insights extend to other facial recognition scenarios, including, but not limited to, those based on different imagery standards that impose comparable constraints. This analysis provides practical recommendations for designing secure systems by demonstrating that only a subset of existing deep learning techniques meets the combined requirements of imperceptibility, selective robustness, and reliable blind extraction needed for ICAO certification.

The rest of this manuscript is structured as follows. Section 2 revises the key biometric and technical specifications of ICAO-compliant facial images and provides an overview of their real-world applications, as well as the potential threats and existing proactive countermeasures. Section 3 outlines the problem formulation, terminology, and key properties of data hiding techniques to certify biometric images. Section 4 highlights the limitations of traditional data-hiding methods and explores modern learning-based approaches for watermarking and steganography in biometric image certification. Section 5 compares the underlying models, analyzing their potentialities and suitability in real-world applications. Finally, conclusions and future directions are presented in Section 6.

2. ICAO Standards and Associated Threats

This section outlines the core biometric and technical specifications of ICAO-compliant facial images, reviews their deployment across real-world application domains, and examines associated manipulation threats and current proactive countermeasures.

Table 1. Summary of relevant ICAO Portrait Quality parameters for facial images in MRTDs and DTCs [69].

Parameter	Requirement	Section in [69]
Inter-eye distance	$\geq$ 90 px (recommended 120 px)	5.2.4
Background	Uniform, light-colored, without patterns	5.2.5
Lighting	Uniform illumination, no strong shadows	5.2.6
Pose	Full frontal, head vertically aligned	5.3.1
Facial expression	Neutral expression, mouth closed	5.3.2
Saturation (printed)	Non-background pixels with values 0 or 255 each $<$0.1%	6.3
Resolution	35 mm $\times$ 45 mm, scanned portrait $\geq$ 300 dpi	6.4
Compression format	JPEG (printed), JPEG2000 (logical storage)	6.5

2.1. ICAO Requirements for Facial Image Quality

Standardized facial image acquisition and encoding are fundamental to ensure interoperability and security in international identity verification systems. To this end, ICAO determines facial images as the primary biometric in MRTDs and DTCs, as detailed in Document 9303, Part 9 [31]. The standard focuses on organizing biometric data within the Logical Data Structure (LDS) and mandates conformance to ISO/IEC 19794-5 [1], later refined by ISO/IEC 39794-5 [2], for the encoding of facial image data.

However, ICAO Document 9303 does not prescribe specific quality criteria for image acquisition, such as resolution, compression format, or subject pose. These operational aspects are addressed separately in the ICAO Technical Report on Portrait Quality (Version 1.0, 2018) [69], which provides best practice guidelines to ensure that captured portraits meet the operational needs of both automated and human identity verification processes. Table 1 summarizes the key parameters relevant to ICAO-compliant facial images. Within the ICAO documentation, the normative strength of each requirement is explicitly indicated: "shall" designates binding obligations, "should" denotes recommended best practices, and "may" identifies optional elements. This structured terminology balances technical consistency enforcement and operational flexibility.

While the standardization improves interoperability and comparison performance, it also introduces a highly predictable acquisition model. Adversaries can exploit the known constraints to synthesize or manipulate facial images that formally satisfy compliance checks, thereby increasing the difficulty of detecting fraudulent identities in critical verification workflows. The implications of these vulnerabilities and their impact on real-world identity systems are analyzed in the next section.

2.2. Real-World Applications and Associated Risks

ICAO-compliant images, originally developed for passport standardization, are now fundamental in governmental, financial, and commercial identity systems. In the travel sector, ICAO-compliant portraits enable Automated Border Control (ABC) through facial recognition using the biometric template stored in their electronic passports [26]. DTCs

are a digital extension of passports, allowing travelers to store their identity on mobile devices [19, 62], while still requiring ICAO-compliant facial images for global interoperability [50]. Beyond aviation, ICAO-compliant images are widely used in finance for biometric verification in KYC and onboarding, ensuring interoperability and accuracy even in remote authentication [24, 61]. In addition, mobile identity systems and digital wallets increasingly rely on ICAO portraits to support secure user verification in diverse application ecosystems.

Although the operational benefits of standardization are evident, the predictability and stability of ICAO-constrained formats increase the exposure to several attack vectors. In addition to typical concerns associated with facial biometrics, such as aging effects and privacy risks [21, 37], identity verification systems are increasingly challenged by sophisticated presentation attacks [13]. For instance, morphing attacks, in which facial features from two individuals are blended to create a synthetic identity, can evade human inspection and automated recognition if compliance constraints are respected [16]. Similarly, deepfake techniques based on generative models can produce realistic ICAO images that embed fraudulent identities [78].

In parallel, the extension of ICAO standards beyond traditional passport systems further amplifies systemic exposure. In the context of DTCs, biometric data stored on mobile devices become vulnerable to compromise in device breaches [49, 54]. Similarly, financial institutions managing biometric databases for identity verification represent attractive targets for cyberattacks if proper security measures are not enforced [38, 59].

Therefore, this combination of acquisition predictability, mass distribution, and extended operational validity underscores the need for protection mechanisms that persist beyond the moment of capture.

2.3. Need for Proactive Protection Mechanisms

PAD is the primary defense against biometric spoofing, analyzing cues like texture inconsistencies, motion artifacts, or liveness to detect falsified traits at the point of capture [56]. Despite increasing sophistication, its protection is limited to acquisition event. Once the image is stored or transmitted, PAD offers no safeguard against tampering, synthetic alterations, or unauthorized redistribution [7, 55].

In application scenarios involving ICAO-compliant images, where biometric data may circulate across decentralized infrastructures and remain valid for extended periods, the absence of persistent integrity verification becomes a critical vulnerability. Traditional cryptographic techniques can secure the transmission but offer no guarantees once access to the content is obtained. Proactive security strategies have been proposed to address this gap by embedding verifiable integrity markers directly within the biometric content. Unlike capture-time defenses, embedded signals persist across the image lifecycle, enabling post-hoc verification of authenticity and tamper-evidence. Among the candidate approaches, data hiding techniques offer promising solutions to enhance the resilience of ICAO-compliant facial images without compromising their usability for visual inspection and automated recognition [8, 57]. Accordingly, the following section surveys the fundamental data-hiding methods, focusing on their design principles, embedding strategies, and relevance to biometric integrity protection.

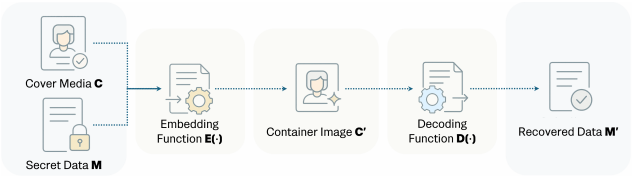

Figure 1. General scheme of a data hiding system.

3. Data Hiding for Biometric Image Certification

This section provides the necessary background on data-hiding techniques for biometric image certification, presenting the problem formulation, key terminology, and fundamental properties that guide the following evaluation of such systems in ICAO-compliant contexts.

3.1. Problem formulation and terminology

The certification of ICAO-compliant biometric images requires embedding security-relevant information directly into the image content, in a manner that preserves its operational usability for recognition and document issuance. In this framework, the data hiding process is modeled by two functions: embedding and extraction (Figure 1). Given a cover image $\mathbf{C}$ and a secret message $\mathbf{M}$, the embedding function $E(\cdot)$ generates a container image $\mathbf{C}'$ according to:

$$\mathbf{C}' = E(\mathbf{C}, \mathbf{M}) \tag{1}$$

where $\mathbf{C}'$ should maintain a high degree of visual and biometric similarity to $\mathbf{C}$. The hidden message is subsequently recovered via a decoding function $D(\cdot)$, producing an estimate $\hat{\mathbf{M}}$:

$$\hat{\mathbf{M}} = D(\mathbf{C}') \tag{2}$$

Throughout this work, we consider data hiding methods whose design and evaluation are driven by the specific needs of biometric certification, rather than by general-purpose communication or copyright protection scenarios. Based on their operational goals, these methods can be categorized into two broad functional classes:

- *Digital watermarking*: these methods embed information to assert properties such as authenticity, integrity,

Figure 2. Comparison between: a) original input image; b) watermarked image obtained with [76] ($PSNR = 45.813$, $SSIM = 0.9861$ for 1 bpp) ; c) steganographic image obtained with [35] ($PSNR = 39.562$, $SSIM = 0.9699$ for 24 bpp).

or provenance of the cover image. The embedded data is expected to survive benign transformations and remain detectable, thus enabling certification even after typical processing such as compression or scaling.

- *Steganography*: these methods aim to conceal the existence of the embedded information, maximizing imperceptibility and minimizing detectability. In the context of biometric certification, steganographic techniques can be reinterpreted to embed fragile integrity signals that, while remaining imperceptible, are disrupted by malicious modifications.

The choice between watermarking and steganography depends on the threat model and operational requirements. Watermarking is typically preferred when persistent verification across benign transformations is desired. Conversely, steganographic embedding may be advantageous when the primary goal is the detection of unauthorized alterations without introducing perceptible changes.

3.2. Key Properties of Data Hiding Systems

Several characteristics and structural properties define data hiding methods in biometric certification contexts, affecting usability, transparency, and extraction requirements. In the following, we describe the most relevant properties [68]: *Embedding visibility*: Data hiding techniques can produce either visible or invisible embeddings. In visible embedding, the presence of the hidden information is perceptible to human observers, serving as an overt signal. Invisible embedding seeks to maintain the perceptual indistinguishability between $\mathbf{C}$ and $\mathbf{C}'$, minimizing the risk of detection.

Blind vs Non-Blind Extraction: In blind data hiding systems, the decoding function $D(\cdot)$ operates exclusively on the container image $\mathbf{C}'$, requiring no access to the original cover $\mathbf{C}$ or any external auxiliary information. Formally: $\hat{\mathbf{M}} = D(\mathbf{C}')$. In non-blind systems, successful extraction depends on additional side information, typically the original cover image $\mathbf{C}$, leading to a decoding function of the form: $\hat{\mathbf{M}} = D(\mathbf{C}', \mathbf{C})$.

Cover-Based vs Coverless Embedding: Cover-based approaches start from a given cover image $\mathbf{C}$ and embed the message $\mathbf{M}$ to produce $\mathbf{C}'$. Coverless approaches, in contrast, generate $\mathbf{C}'$ directly conditioned on $\mathbf{M}$ without relying on an existing cover.

Invertibility: Invertible data hiding methods allow the simultaneous recovery of the embedded message $\mathbf{M}$ and, optionally, the original cover image $\mathbf{C}$ from the container $\mathbf{C}'$.

Security: A secure data hiding method must prevent unauthorized detection or extraction of the embedded data. Adversaries may attempt to detect the presence of embedded data via steganalysis techniques or to decode it without access to the original embedding process or keys.

Fragility and Robustness: The resilience of the embedded message against transformations determines if a method is classified as fragile, semi-fragile, or robust. Fragile methods are highly sensitive to any alteration of the container image $\mathbf{C}'$, leading to significant degradation or loss of embedded information $\mathbf{M}$ even under minor modifications. Semi-fragile methods are designed to withstand benign operations like compression, but fail under malicious semantic manipulations, such as face morphing or swapping. Robust methods aim to maintain the integrity of the embedded data across a broad range of distortions, as signal processing and geometric transformations, and adversarial attacks, ensuring the persistence of hidden data under diverse operational conditions.

Capacity: The capacity of a data-hiding system defines the maximum amount of information that can be embedded and reliably extracted from a container image $\mathbf{C}'$.

The relevance of these properties varies regarding the operational constraints of ICAO-compliant biometric image certification. In this domain, the invisibility of the embedding is mandatory to avoid degradation of recognition performance and comply with visual quality standards. Blind extraction is highly desirable to enable decentralized verification without requiring access to the original cover image. Cover-based embedding is mandatory to ensure the certification process refers to an authentic biometric acquisition. Invertibility is also desired to verify the integrity of the embedded message and the cover image after potential manipulations. Security against unauthorized detection or extraction is desirable to protect the confidentiality of embedded data. Fragile and semi-fragile embedding strategies are mandatory to enable reliable tamper detection, while robust embedding approaches are unsuitable as they may tolerate unacceptable semantic alterations. Finally, capacity must be balanced to carry necessary certification data without compromising invisibility or biometric performance: too little limits effectiveness, while too much may introduce artifacts due to stronger modifications in cover-based embedding. Accordingly, the remainder of this survey focuses on data hiding methods that satisfy the operational requirements identified above. Figure 2 shows examples of cover and container image pairs, illustrating these combined re-

quirements for ICAO-compliant systems.

4. Fragile and Semi-Fragile Data Hiding

4.1. Limitations of Traditional Methods

Early fragile and semi-fragile data hiding systems predominantly relied on traditional embedding techniques operating in the spatial or frequency domain. Spatial domain approaches, such as Least Significant Bit (LSB) substitution or histogram modification [9, 10], offered good imperceptibility but were highly vulnerable to benign transformations like compression, filtering, or even minor noise. Frequency domain methods, relying on transformations such as DFT [51], DWT [23] or DCT [23], improved robustness against certain signal processing operations, such as compression, but often suffered perceptual distortion and lacked fine-grained control over the fragility of embedded information [12]. Although these traditional algorithms have shown effectiveness in specific integrity verification tasks, their applicability is inherently narrow, requiring expert-driven design tailored to specific scenarios. Furthermore, the increasing sophistication of manipulation and removal attacks compromises their long-term reliability [17].

The advent of deep learning introduced a paradigm shift in data hiding. Deep neural networks provide adaptable and generalized frameworks capable of learning complex embedding patterns directly from data. This enables improved resilience against a broader range of attacks, enhanced imperceptibility, and the possibility of dynamically retraining models to prioritize different objectives, such as robustness, invisibility, or payload capacity, without requiring specialized manual engineering [3, 87]. Moreover, the nonlinearity of deep architectures significantly enhances the security of the embedded information against adversarial retrieval attempts. Considering these advantages, learning-based methods have become a promising direction for developing fragile and semi-fragile data hiding systems suitable for modern biometric image certification under ICAO-compliant constraints.

4.2. Fragile and Semi-Fragile Watermarking

Recent deep learning-based watermarking approaches have explored several architectural paradigms that can meet the operational demands of ICAO-compliant biometric certification. Those based on encoder–decoder, GANs, transformers, and INNs have been explored with varying degrees of success and constitute the predominant design strategies. Each architecture presents distinct trade-offs between invisibility, robustness to benign transformations, and sensitivity to semantic manipulations.

Encoder–decoder frameworks represent the foundational architecture adopted in most modern watermarking solutions due to their conceptual simplicity and flexibility. Typical implementations employ convolutional neural networks to embed a carefully controlled payload into cover images, optimizing embedding imperceptibility while preserving extraction accuracy.

A seminal method, *HiDDeN* [87], introduced an end-to-end GAN framework integrating encoder, decoder, and discriminator, using adversarial training to embed resilient watermarks. Despite its innovation, *HiDDeN* showed limited capacity and generalization against common signal attacks, prompting further developments. For instance, *ARW-GAN* [30] incorporated attention-guided feature fusion and dense connections within a full GAN pipeline, enhancing both imperceptibility and robustness. Nonetheless, GAN-based methods still tend to introduce subtle artifacts during generation. Furthermore, their emphasis on synthesis rather than explicit recovery can limit watermark extraction reliability, even under known benign operations.

To mitigate these issues, some methods rely solely on CNN-based architectures, incorporating discriminators and adversarial training while minimizing generative components. For example, *MBRS* [32] proposed a robust end-to-end design against JPEG compression by simulating noise layers during training. Other works like *TDSL* [41] and *Adaptor* [64] employ a two-phase training and embedding strategy to improve resilience against realistic, non-differential transformations like JPEG compression. These methods also allow tunable embedding strength, providing a more flexible trade-off between imperceptibility and robustness. Beyond resilience to benign distortions, approaches such as *FaceSigns* [46] and *WaterLo* [5] expanded their semi-fragile scopes by also including malicious semantic transformations such as face swap within the training process. This enables the watermark to remain robust to expected benign changes (e.g., JPEG compression or filtering) while becoming sensitive to semantic manipulations. Although CNN-based architectures with discriminators have demonstrated strong performance, recent transformer-based models have emerged, leveraging attention mechanisms to embed watermarks in spatially relevant regions adaptively. *StegaFormer* [76] and *WFormer* [43] have shown significant gains in both imperceptibility and extraction accuracy, in exchange for higher computational requirements. Their performance in semi-fragile contexts suggests promising applicability to ICAO standards.

Finally, despite the versatility of encoder–decoder schemes and their synergy with adversarial training, they often suffer from information loss and require precise tuning between robustness and fragility. To address this, INN-based methods like *RIS* [39] offer promising alternatives. Originally explored in steganography, INNs show great potential in ensuring total invertibility and higher imperceptibility, key factors for ICAO-compliant systems.

4.3. Fragile and Semi-Fragile Steganography

Initially developed for covert communication, steganographic methods can be reinterpreted in biometric image certification to act as fragile integrity markers [18]. By embedding sensitive payloads designed to degrade under tampering, steganography offers a complementary strategy for proactively detecting unauthorized modifications in ICAO-compliant scenarios. As in watermarking systems, deep learning-based steganographic methods adopt a variety of architectural paradigms, including encoder–decoder frameworks, GANs, INNs, transformers, and diffusion models.

Encoder–decoder architectures based on fully DNNs and CNNs [4, 79] provide a standard framework for fragile and semi-fragile steganography, embedding entire images (high capacity) into host content while optimizing invisibility and reconstruction accuracy. These systems, despite targeting high visual fidelity, inherently exhibit sensitivity to content alterations, making them suitable for tamper detection. Variants that incorporate additional embedding constraints, such as symmetry preservation [36], further refine the balance between imperceptibility and fragility, strengthening the potential of encoder-decoder architectures for integrity verification tasks.

GAN-based steganography enhances invisibility and undetectability through adversarial training. Early methods like *SteganoGAN* [80] and *ISGAN* [81] offer good capacity and visual quality but lack robustness to perturbations such as compression. Recent models, including *ADBH* [75], *CHAT-GAN* [60], and *Cover-GAN* [40], address this by incorporating attention mechanisms or perturbation simulation, improving robustness. However, they still face decoding challenges under lossy conditions and limited scalability to high-resolution images [80, 81, 75, 60, 40].

INNs have gained attention in steganography for their ability to model concealing and revealing as symmetric, reversible processes [34, 42]. These architectures support high capacity and imperceptibility, often outperforming traditional encoder-decoder schemes in preserving image fidelity. Recent works have introduced explicit robustness mechanisms, such as conditional flows [73] or direct embedding in DCT coefficients [39] to improve robustness under distortions like JPEG compression. While computational demands and invertibility constraints can limit scalability, their capacity to ensure accurate decoding even under moderate image degradation makes them a promising option for scenarios requiring integrity and resilience, such as biometric image certification under ICAO standards.

Transformer-based steganography represents an emerging direction. Models such as *Stegformer* [35] exploit attention mechanisms to distribute the payload across semantically meaningful regions adaptively. This flexibility could facilitate embedding strategies focused on critical biometric features to enhance tamper sensitivity. However, these transformer-based models still prioritize payload capacity and general robustness. Specific adaptations may be necessary to align with fragile or semi-fragile requirements.

Finally, diffusion models have recently been explored for generative steganography [77, 74], synthesizing entire images conditioned on hidden messages. Although techniques like *DERO* [15] achieve state-of-the-art imperceptibility and steganalysis resistance, their generative nature makes them unsuitable for certifying the authenticity of pre-acquired biometric images, as required by ICAO standards.

5. Comparative Evaluation

The evaluation of fragile and semi-fragile data hiding methods is structured along four core dimensions: imperceptibility, robustness, capacity, and security. While a broader set of structural properties has been introduced in Section 3.2 to characterize the behavior of data hiding systems, this section focuses on the quantitative metrics associated with the most critical dimensions for ICAO-compliant biometric certification. Accordingly, each metric must be interpreted in light of the operational objectives of proactive tamper detection. The remainder of this section formally introduces the evaluation criteria and presents a comparative analysis of representative methods along these dimensions.

5.1. Evaluation Metrics

Imperceptibility The visual fidelity between the original image $\mathbf{C}$ and the container image $\mathbf{C}'$ is critical for ICAO-compliant biometric images, where any visible alteration may compromise recognition performance. Standard evaluation metrics include Peak Signal-to-Noise Ratio (PSNR) [27], Structural Similarity Index (SSIM) [27], learning-based perceptual metrics such as LPIPS [82], and Mean Square Error (MSE). Although high imperceptibility is essential across all robustness levels, it becomes particularly critical for fragile and semi-fragile methods targeting biometric certification scenarios [45, 46].

Robustness The evaluation of the ability of a data hiding method to preserve and accurately recover the embedded message $\mathbf{M}$ from the container image $\mathbf{C}'$ after undergoing various transformations must consider both benign operations, such as JPEG compression or resizing [63, 70], and more severe and malicious alterations, such as adversarial perturbations [29] or semantic manipulations [45, 70]. The primary metric for evaluating robustness is the Bit Error Rate (BER), which quantifies the proportion of bits incorrectly recovered between $\mathbf{M}$ and $\hat{\mathbf{M}}$. Lower BER values correspond to greater resilience of the hidden information under distortions. In some cases, Bit Recovery Accuracy (BRA) or normalized cross-correlation (NC) [86] are also used to assess the fidelity of message retrieval, particularly under different types of attacks. For fragile and semi-fragile data hiding tailored to ICAO-compliant biometric certifica-

tion, robustness must be properly tuned: the embedded payload should resist benign signal-level degradations, specifically JPEG compression, but fail when semantic integrity is compromised [46, 5]. Excessive robustness, as in copyright watermarking frameworks [65, 84], would undermine the ability to detect unauthorized biometric modifications.

Capacity The amount of information that can be embedded within an image is typically measured in bits per pixel (BPP) [11]. This metric is particularly used in steganographic approaches, where the objective is to covertly transmit large volumes of data without arousing suspicion. High-capacity methods enable the embedding of complex payloads such as multiple images [20, 35, 42], encryption keys, or metadata, but often at the expense of visual fidelity. In contrast, capacity could not be the primary goal in watermarking contexts. However, maintaining a reasonable embedding capacity can offer operational flexibility, for instance, by allowing individualized watermarking across users or systems, provided that imperceptibility and fragility requirements are not violated.

Security The resistance of a data hiding system against intentional attacks aiming to detect, remove, or corrupt the embedded message $\mathbf{M}$ can be measured in different ways. Evaluation typically includes resilience against adversarial perturbations designed to disrupt decoding [85], detection by steganalysis models [28], watermark removal techniques [45], and generalization to manipulations not encountered during training, such as deepfake generation [67] or presentation attacks [22]. Common metrics include bit recovery accuracy after an attack, steganalysis error rates, attack success rates [65], and deepfake detection performance [86]. In deep learning-based systems, evaluations are often conducted under white-box (i.e., the attacker knows the model) and black-box (i.e., model unknown) scenarios to comprehensively assess vulnerability [33, 66]. In ICAO-compliant applications, security is crucial to ensure that the embedded data resists unauthorized modifications while remaining imperceptible and non-disruptive to recognition systems.

5.2. Comparative Analysis and Discussion

Table 2 offers a comparative overview of state-of-the-art deep learning-based data hiding approaches assessed across core properties relevant to biometric image certification: imperceptibility, robustness, and payload capacity. To ensure fair comparison, all methods are analyzed using metrics reported in their original publications, also considering their applicability to ICAO-compliant scenarios. For payload capacity, a similar value is used to facilitate comparison rather than reporting maximum capacity. The table also reports their architecture (e.g., INNs, transformers), the embedding domain (spatial or frequency), and the hidden content type.

For each method, imperceptibility is quantified using the PSNR between the original image (C) and the container (C'). Values above 40 dB generally denote visually indistinguishable changes [53, 58], crucial in the context of ICAO-compliant facial images, where any visible degradation may affect both human inspection and automated face recognition. Several methods, such as *HiNet*, *RIIS*, *RIS*, and *StegFormer*, consistently exceed this threshold across multiple datasets and resolutions, indicating a strong alignment with ICAO visual quality requirements.

In parallel, recovery fidelity, measuring how accurately the hidden data can be retrieved, is reported as PSNR between secret and recovered images or, in the case of binary string hiding, as the BER. High reconstruction PSNRs ($\geq$35 dB) and low BERs ($\leq$0.3%) are indicative of practical message preservation under distortion-free conditions (e.g., [18]). Approaches such as *StegFormer*, *MBRS*, *RIS*, and *FaceSigns* demonstrate excellent recovery performance, suggesting strong potential for scenarios in which embedded information must be reliably extracted post-verification.

Robustness to JPEG compression is reported as PSNR degradation or BER under compression with quality factors (QF) of 90 or 50. This form of selective robustness is essential for semi-fragile watermarking scenarios. In this context, the "*Extra robustness*" column indicates whether a method also preserves the payload under other operations, such as filtering or geometric changes. While general-purpose watermarking methods typically aim for strong resilience, such robustness may be counterproductive in biometric certification, where fragility to malicious content manipulation is a desired property. Accordingly, methods that are not robust to arbitrary transformations are better aligned with the ICAO-compliant image integrity verification requirements.

Finally, the overall "*Grade*" column provides a high-level qualitative indication of each method's potential suitability in ICAO contexts, reflecting a balanced assessment across imperceptibility, recovery fidelity, robustness behavior, and compliance with minimum image resolution requirements (Table 1). While some approaches, like *SteganoGAN* or *ISGAN*, exhibit good imperceptibility, their lack of robustness and recovery fidelity limits their applicability. Conversely, methods such as *RIS* and *MBRS* strike a more favorable trade-off, suggesting higher compatibility with the goals of biometric image certification. Importantly, INN-based and transformer-based designs appear particularly promising due to their inherent support for invertibility and flexible embedding. While the security property is a critical dimension in our analysis, a detailed comparative assessment is challenging. This is because most of the suitable methods report performance against generic steganalysis benchmarks (e.g., with detectability rates around 55% or less using methods like XuNet [71] or SRNet [6]), which may not reliably reflect the security required against sophisticated, context-aware attacks in real-world ICAO opera-

Table 2. Deep data hiding methods comparison across imperceptibility, robustness, and payload. Underlined values indicate PSNR between cover/stego images; italicized values refer to PSNR between secret/recovered images. BER denotes the bit error rate (%) between embedded/extracted messages (before/after) JPEG compression. Asterisk (*) indicates values under JPEG compression with QF=50.

Model	Venue Year	Framework Type	Domain	Input	Dataset/ Image dimension	Payload (BBP)	Imperceptibility (PSNR dB)	JPG Compression Robustness (QF=90,*=50)		Extra robustness	Grade
								PSNR (dB)	BER (%)		
SteganoGAN [80]	CoRR 2019	Enc-Dec (GAN)	Spatial	Data Hiding	**DIV2K/** $\sim 1024 \times 1024$ **COCO/** $\sim 256 \times 256$	6	38.94 / – 36.33 / –	–	–	✗	Low+
ISGAN [81]	MTAP 2019	Enc-Dec (GAN)	Spatial	Image Hiding	**LFW/** 256×256 **PASCALVOC/** 256×256 **ImageNet/** 256×256	8	34.63 / *33.63* 34.49 / *33.31* 34.89 / *33.42*	–	–	✗	Low
ABDH [75]	AAAI 2020	Enc-Dec (GAN)	Spatial	Image Hiding	**COCO/** 512×512	~ 24	31.91 / *30.66*	– / *32.97*	–	✓	Low
UDH [79]	NeurIPS 2020	Enc-Dec (CNN)	Spatial	Image Hiding	**ImageNet/** 128×128	~ 24	39.13 / *35.0*	–	*0.0* / *0.6**	✓	Medium
MBRS [32]	ACM-MM 2021	Enc-Dec (CNN)	Spatial	Binary String Hiding	**ImageNet/** 400×400 **COCO/** $\sim 400 \times 400$	~ 0.0039	– 39.32 / –	– 42.04 / –	– *0.0012* / *0.00063*	✗	High
ISN [42]	CVPR 2021	INN	Spatial	Image Hiding	**ImageNet/** 144×144 **Paris Street/** 144×144	~ 24	38.05 / *35.38* 40.49 / *43.33*	–	–	✗	Low+
HiNet [34]	CVPR 2021	INN	Frequency (DWT)	Image Hiding	**DIV2K/** 1024×1024 **ImageNet/** 256×256 **COCO/** 256×256	~ 24	48.99 / *52.86* 44.60 / *46.78* 46.52 / *46.98*	–	–	✗	Medium+
FaceSigns [46]	TOMM 2022	Enc-Dec (CNN)	Spatial	Binary String Hiding	**CelebA/** 256×256	0.00065	36.08 / –	–	*0.32* / *0.51*	✓	Medium-
RIIS [73]	CVPR 2022	INN	Spatial	Image Hiding	**DIV2K/** 1024×1024 **ImageNet/** $\sim 256 \times 256$	~ 24	– / *44.19* 43.97 / *46.71*	– / *28.71* 28.17 / *28.53*	–	✓	Medium+
RIS [39]	AAAI 2023	INN	Frequency (DCT)	Binary String Hiding	**MSCOCO/** 256×256	1	48.41 / –	44.13 / –	*0.0* / *0.31*	✗	High+
DeepMIH [20]	PAMI 2022	INN	Frequency (DWT)	Image Hiding	**DIV2K/** 1024×1024 **ImageNet/** 256×256 **COCO/** 256×256	~ 24	43.72 / *41.41* 40.31 / *36.63* 40.30 / *36.55*	–	–	✗	Medium+
Adaptor [64]	TCSVT 2023	Enc-Dec (CNN)	Spatial	Binary String Hiding	**COCO/** 128×128	~ 0.0013	–	38.42* / –	– / *0.016**	✓	Medium-
DIH-OAIN [28]	TCSVT 2023	INN	Spatial	Image Hiding	**COCO/** $\sim 256 \times 256$ **PASCALVOC/** $\sim 256 \times 256$	~ 24	46.56/ *39.73* 54.45/ *48.60*	–	–	✗	Medium-
StegFormer [35]	AAAI 2024	Enc-Dec (Transformer)	Spatial	Image Hiding	**DIV2K/** 1024×1024 **ImageNet/** 256×256 **COCO/** 256×256	~ 24	56.30 / *55.45* 48.79 / *49.18* 48.77 / *49.21*	–	–	✗	Medium+
Stegaformer [76]	BMVA 2024	Enc-Dec (Transformer)	Spatial	Binary String Hiding	**COCO/** 256×256 **DIV2K/** 256×256	3	43.37 / – 47.31 / –	–	*0.32* / – *0.24* / –	✗	Medium-

tional environments. Despite the table's qualitative nature, which limits strict quantitative ranking based on application priorities, it represents the first attempt to systematically assess the suitability of data-hiding approaches for ICAO-compliant scenarios. Specifically, reported values across key properties offer practitioners a valuable basis for selecting appropriate methods for specific operational needs. Notably, the analysis reveals that only a subset of models meet the combined requirements of visual conformity, selective robustness, and reliable decoding necessary for biometric image certification. These findings underscore the need for further development of specialized strategies tailored to biometric identity constraints.

6. Conclusions and Future Directions

This survey explored how steganographic and watermarking techniques can enhance the integrity and traceability of ICAO-compliant facial images, especially where traditional countermeasures like PAD offer limited post-capture protection. These methods are particularly relevant in high-risk scenarios such as border control and KYC, where manipulation attacks like morphing and deepfake attacks pose serious threats. We reviewed core concepts, categorized existing techniques, and analyzed their suitability under ICAO constraints, emphasizing trade-offs between imperceptibility, robustness, capacity, and security. To meet these demands, deep learning-based fragile and semi-fragile methods, particularly those using INNs, emerge as the most promising due to their adaptability and resilience. Based on this analysis, we provided practical guidelines to support the design and deployment of watermarking and steganographic methods in ICAO-compliant systems, identifying key features for effective integrity verification. Despite promising advances, the field remains underexplored and requires further validation. Future research should evaluate the interaction between embedded signals and face recognition pipelines, assess resilience against adversarial and semantic attacks, and develop frameworks to verify compliance with ICAO standards. These steps will help bridge the gap between integrity verification and deployable security solutions in biometric identity management.

Acknowledgment

This work was partially supported by Project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU. Davide Ghiani's PhD grant is partly funded by Dedem SpA under the PNRR program.

References

[1] Information technology — Biometric data interchange formats — Part 5: Face image data, 2011.

[2] Information technology — Extensible biometric data interchange formats — Part 5: Face image data, 2019.

[3] S. Abdelnabi and M. Fritz. Adversarial watermarking transformer: Towards tracing text provenance with data hiding. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 121–140. IEEE, 2021.

[4] S. Baluja. Hiding images in plain sight: Deep steganography. *Advances in neural information processing systems*, 30, 2017.

[5] N. Beuve, W. Hamidouche, and O. Déforges. Waterlo: Protect images from deepfakes using localized semi-fragile watermark. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 393–402, 2023.

[6] M. Boroumand, M. Chen, and J. Fridrich. Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 14(5):1181–1193, 2018.

[7] C. Busch. Challenges for automated face recognition systems. *Nature Reviews Electrical Engineering*, 1(11):748–757, 2024.

[8] P. Capasso, G. Cattaneo, and M. De Marsico. A comprehensive survey on methods for image integrity. *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(11):1–34, 2024.

[9] C.-K. Chan and L.-M. Cheng. Hiding data in images by simple lsb substitution. *Pattern recognition*, 37(3):469–474, 2004.

[10] G. Coatrieux, W. Pan, N. Cuppens-Boulahia, F. Cuppens, and C. Roux. Reversible watermarking based on invariant image classification and dynamic histogram shifting. *IEEE Transactions on information forensics and security*, 8(1):111–120, 2012.

[11] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker. *Digital watermarking and steganography*. Morgan kaufmann, 2007.

[12] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for images, audio and video. In *Proceedings of 3rd IEEE international conference on image processing*, volume 3, pages 243–246. IEEE, 1996.

[13] N. Di Domenico, G. Borghi, A. Franco, and D. Maltoni. Onot: a high-quality icao-compliant synthetic mugshot dataset. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–10. IEEE, 2024.

[14] N. Erdogmus and S. Marcel. Spoofing face recognition with 3d masks. *IEEE transactions on information forensics and security*, 9(7):1084–1097, 2014.

[15] H. Fang, K. Chen, Y. Qiu, Z. Ma, W. Zhang, and E.-C. Chang. Dero: Diffusion-model-erasure robust watermarking. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 2973–2981, 2024.

[16] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *IEEE International Joint Conference on Biometrics*, pages 1–7, 2014.

[17] L. Geng, W. Zhang, H. Chen, H. Fang, and N. Yu. Real-time attacks on robust watermarking tools in the wild by cnn. *Journal of Real-Time Image Processing*, 17:631–641, 2020.

[18] D. Ghiani, J. D. R. Chivata, S. Lilliu, S. M. La Cava, M. Micheletto, G. Orrù, F. Lama, and G. L. Marcialis. Fragile watermarking for image certification using deep steganographic embedding. *arXiv preprint arXiv:2504.13759*, 2025.

[19] Government of the Netherlands. Dutch participation in european dtc pilot. `https://www.government.nl/documents/publications/2023/02/23/dtc`, 2023. Accessed: 2025-04-03.

[20] Z. Guan, J. Jing, X. Deng, M. Xu, L. Jiang, Z. Zhang, and Y. Li. Deepmih: Deep invertible network for multiple image hiding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):372–390, 2022.

[21] G. Guo and N. Zhang. A survey on deep learning based face recognition. *Computer vision and image understanding*, 189:102805, 2019.

[22] Y. Guo and Z. Liu. Coverless steganography for face recognition based on diffusion model. *IEEE Access*, 2024.

[23] M. Hamidi, M. E. Haziti, H. Cherifi, and M. E. Hassouni. Hybrid blind robust image watermarking technique based on dft-dct and arnold transform. *Multimedia Tools and Applications*, 77:27181–27214, 2018.

[24] J. Hayata, K. Nomura, Y. Takata, H. Kumagai, M. Kamizono, T. Kono, Y. Maeda, and N. Fukuda. A trust service model adaptable to various assurance levels by linking digital ids and certificates. In *2024 8th International Conference on Cryptography, Security and Privacy (CSP)*, pages 38–45. IEEE, 2024.

[25] J. Hernandez-Ortega, J. Fierrez, A. Morales, and J. Galbally. Introduction to presentation attack detection in face biometrics and recent advances. *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pages 203–230, 2023.

[26] F. Hidayat, U. Elviani, G. B. G. Situmorang, M. Z. Ramadhan, F. A. Alunjati, and R. F. Sucipto. Face recognition for automatic border control: a systematic literature review. *IEEE Access*, 12:37288–37309, 2024.

[27] A. Hore and D. Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.

[28] X. Hu, Z. Fu, X. Zhang, and Y. Chen. Invisible and steganalysis-resistant deep image hiding based on one-way adversarial invertible networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7):6128–6143, 2023.

[29] H. Huang, Y. Wang, Z. Chen, Y. Zhang, Y. Li, Z. Tang, W. Chu, J. Chen, W. Lin, and K.-K. Ma. Cmua-watermark: A cross-model universal adversarial watermark for combating deepfakes. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 989–997, 2022.

[30] J. Huang, T. Luo, L. Li, G. Yang, H. Xu, and C.-C. Chang. Arwgan: Attention-guided robust image watermarking model based on gan. *IEEE Transactions on Instrumentation and Measurement*, 72:1–17, 2023.

[31] International Civil Aviation Organization. Machine readable travel documents: Part 9 – deployment of biometric identification and electronic storage of data in mrtds. Technical Report Doc 9303, Part 9, International Civil Aviation Organization (ICAO), Montréal, Canada, 2021.

[32] Z. Jia, H. Fang, and W. Zhang. Mbrs: Enhancing robustness of dnn-based watermarking by mini-batch of real and simulated jpeg compression. In *Proceedings of the 29th ACM international conference on multimedia*, pages 41–49, 2021.

[33] Z. Jiang, M. Guo, Y. Hu, J. Jia, and N. Z. Gong. Certifiably robust image watermark. In *European Conference on Computer Vision*, pages 427–443. Springer, 2024.

[34] J. Jing, X. Deng, M. Xu, J. Wang, and Z. Guan. Hinet: Deep image hiding by invertible network. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4733–4742, 2021.

[35] X. Ke, H. Wu, and W. Guo. Stegformer: Rebuilding the glory of autoencoder-based steganography. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 2723–2731, 2024.

[36] A. Khalifa and A. Guzman. Imperceptible image steganography using symmetry-adapted deep learning techniques. *Symmetry*, 14(7):1325, 2022.

[37] S. M. La Cava, G. Orrù, M. Drahansky, G. L. Marcialis, and F. Roli. 3d face reconstruction: the road to forensics. *ACM Computing Surveys*, 56(3):1–38, 2023.

[38] L. Laishram, M. Shaheryar, J. T. Lee, and S. K. Jung. Toward a privacy-preserving face recognition system: A survey of leakages and solutions. *ACM Computing Surveys*, 57(6):1–38, 2025.

[39] Y. Lan, F. Shang, J. Yang, X. Kang, and E. Li. Robust image steganography: hiding messages in frequency coefficients. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 14955–14963, 2023.

[40] G. Li, S. Li, Z. Qian, and X. Zhang. Cover-separable fixed neural network steganography via deep generative models. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 10238–10247, 2024.

[41] Y. Liu, M. Guo, J. Zhang, Y. Zhu, and X. Xie. A novel two-stage separable deep learning framework for practical blind watermarking. In *Proceedings of the 27th ACM International conference on multimedia*, pages 1509–1517, 2019.

[42] S.-P. Lu, R. Wang, T. Zhong, and P. L. Rosin. Large-capacity image steganography based on invertible neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10816–10825, 2021.

[43] T. Luo, J. Wu, Z. He, H. Xu, G. Jiang, and C.-C. Chang. Wformer: A transformer-based soft fusion model for robust image watermarking. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024.

[44] I. McAteer, A. Ibrahim, G. Zheng, W. Yang, and C. Valli. Integration of biometrics and steganography: a comprehensive review. *Technologies*, 7(2):34, 2019.

[45] A. V. Nadimpalli and A. Rattani. Social media authentication and combating deepfakes using semi-fragile invisible image watermarking. *Digital Threats: Research and Practice*, 5(4):1–30, 2024.

[46] P. Neekhara, S. Hussain, X. Zhang, K. Huang, J. McAuley, and F. Koushanfar. Facesigns: semi-fragile neural watermarks for media authentication and countering deepfakes. *arXiv preprint arXiv:2204.01960*, 2022.

[47] N. Palaghias, L. Chondromatidou, J. Calapez, M. Krecek, G. Kouzas, D. Kassimis, J. Zaras, P. Orfanoudakis, I. Crucianu, and D. Oniga. Seamless identity verification at water and land borders. In *2024 5th International Conference in Electronic Engineering, Information Technology & Education (EEITE)*, pages 1–6. IEEE, 2024.

[48] A. Panzino, S. M. La Cava, G. Orrù, and G. L. Marcialis. Evaluating the integration of morph attack detection in automated face recognition systems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3827–3836, 2024.

[49] K. K. Prakasha and U. Sumalatha. Privacy-preserving techniques in biometric systems: Approaches and challenges. *IEEE Access*, 2025.

[50] A.-T. Radutoiu, A. Bassit, R. Veldhuis, and C. Busch. A study on the next generation of digital travel credentials. In *2024 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6. IEEE, 2024.

[51] S. Rawat and B. Raman. A blind watermarking algorithm based on fractional fourier transform and visual cryptography. *Signal Processing*, 92(6):1480–1491, 2012.

[52] M. K. Rusia and D. K. Singh. A comprehensive survey on techniques to handle face identity threats: challenges and opportunities. *Multimedia Tools and Applications*, 82(2):1669–1748, 2023.

[53] M. M. Sadek, A. S. Khalifa, and M. G. Mostafa. Robust video steganography algorithm using adaptive skin-tone detection. *Multimedia Tools and Applications*, 76:3065–3085, 2017.

[54] S. L. Sanna, D. Soi, D. Maiorca, G. Fumera, and G. Giacinto. A risk estimation study of native code vulnerabilities in android applications. *Journal of Cybersecurity*, 10(1):tyae015, 2024.

[55] K. Shaheed, P. Szczuko, M. Kumar, I. Qureshi, Q. Abbas, and I. Ullah. Deep learning techniques for biometric security: A systematic review of presentation attack detection systems. *Engineering Applications of Artificial Intelligence*, 129:107569, 2024.

[56] D. Sharma and A. Selwal. A survey on face presentation attack detection mechanisms: hitherto and future perspectives. *Multimedia Systems*, 29(3):1527–1577, 2023.

[57] U. Sumalatha, K. K. Prakasha, S. Prabhu, and V. C. Nayak. A comprehensive review of unimodal and multimodal fingerprint biometric authentication systems: Fusion, attacks, and template protection. *IEEE Access*, 2024.

[58] J. Sun, Z. Yang, Y. Zhang, T. Li, and S. Wang. High-capacity data hiding method based on two subgroup pixels-value adjustment using encoding function. *Security and Communication Networks*, 2022(1):4336526, 2022.

[59] Z. Tabassum, S. Arsheen, K. Ahmad, and G. Ozdemir. Blockchain-based secure e-kyc solution for banking system. In *2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN)*, pages 565–571. IEEE, 2024.

[60] J. Tan, X. Liao, J. Liu, Y. Cao, and H. Jiang. Channel attention image steganography with generative adversarial networks. *IEEE transactions on network science and engineering*, 9(2):888–903, 2021.

[61] M. Tavares, A. Guerreiro, C. Coutinho, F. Veiga, and A. Campos. Wallid: Secure your id in an ethereum wallet. In *2018 International Conference on Intelligent Systems (IS)*, pages 714–721. IEEE, 2018.

[62] The Finnish Border Guard. Dtc border control faster and smoother. `https://raja.fi/en/dtc`, 2024. Accessed: 2025-04-03.

[63] B. Wang and Y. Wu. Staged adaptive blind watermarking scheme. In *Proceedings of the Asian conference on computer vision*, pages 1812–1827, 2022.

[64] B. Wang, Y. Wu, and G. Wang. Adaptor: Improving the robustness and imperceptibility of watermarking by the adaptive strength factor. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(11):6260–6272, 2023.

[65] J. Wang, H. Wang, J. Zhang, H. Wu, X. Luo, and B. Ma. Invisible adversarial watermarking: A novel security mechanism for enhancing copyright protection. *ACM Transactions on Multimedia Computing, Communications and Applications*, 21(2):1–22, 2024.

[66] R. Wang, F. Juefei-Xu, M. Luo, Y. Liu, and L. Wang. Faketagger: Robust safeguards against deepfake dissemination via provenance tracking. In *Proceedings of the 29th ACM international conference on multimedia*, pages 3546–3555, 2021.

[67] T. Wang, M. Huang, H. Cheng, X. Zhang, and Z. Shen. Lampmark: Proactive deepfake detection via training-free landmark perceptual watermarks. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 10515–10524, 2024.

[68] Z. Wang, O. Byrnes, H. Wang, R. Sun, C. Ma, H. Chen, Q. Wu, and M. Xue. Data hiding with deep learning: A survey unifying digital watermarking and steganography. *IEEE Transactions on Computational Social Systems*, 10(6):2985–2999, 2023.

[69] A. Wolf. Icao: Portrait quality (reference facial images for mrtd), version 1.0. standard. *International Civil Aviation Organization*, 2018.

[70] X. Wu, X. Liao, and B. Ou. Sepmark: Deep separable watermarking for unified source tracing and deepfake detection. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 1190–1201, 2023.

[71] G. Xu, H.-Z. Wu, and Y.-Q. Shi. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters*, 23(5):708–712, 2016.

[72] W. Xu, D. He, L. Qiu, and X. Liang. Stegdiff: Content-adaptive image steganography with denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

[73] Y. Xu, C. Mou, Y. Hu, J. Xie, and J. Zhang. Robust invertible image steganography. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7875–7884, 2022.

[74] Y. Yang, Z. Liu, J. Jia, Z. Gao, Y. Li, W. Sun, X. Liu, and G. Zhai. Diffstega: towards universal training-free coverless image steganography with diffusion models. *arXiv preprint arXiv:2407.10459*, 2024.

[75] C. Yu. Attention based data hiding with generative adversarial networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 1120–1128, 2020.

[76] G. Yu, Q. Xuchong, and Y. Zihan. Effective message hiding with order-preserving mechanisms. *arXiv preprint arXiv:2402.19160*, 2024.

[77] J. Yu, X. Zhang, Y. Xu, and J. Zhang. Cross: Diffusion model makes controllable, robust and secure image steganography. *Advances in Neural Information Processing Systems*, 36:80730–80743, 2023.

[78] P. Yu, Z. Xia, J. Fei, and Y. Lu. A survey on deepfake video detection. *Iet Biometrics*, 10(6):607–624, 2021.

[79] C. Zhang, P. Benz, A. Karjauv, G. Sun, and I. S. Kweon. Udh: Universal deep hiding for steganography, watermarking, and light field messaging. *Advances in Neural Information Processing Systems*, 33:10223–10234, 2020.

[80] K. A. Zhang, A. Cuesta-Infante, L. Xu, and K. Veeramachaneni. Steganogan: High capacity image steganography with gans. *arXiv preprint arXiv:1901.03892*, 2019.

[81] R. Zhang, S. Dong, and J. Liu. Invisible steganography via generative adversarial networks. *Multimedia tools and applications*, 78(7):8559–8575, 2019.

[82] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[83] R. Zhang, K. Zhu, H. Zhang, and X. Zhang. Robust deep image steganography with steganalysis adaptive adversarial learning. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2099–2107, 2019.

[84] X. Zhang, R. Li, J. Yu, Y. Xu, W. Li, and J. Zhang. Editguard: Versatile image watermarking for tamper localization and copyright protection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11964–11974, 2024.

[85] Y. Zhang, D. Ye, C. Xie, L. Tang, X. Liao, Z. Liu, C. Chen, and J. Deng. Dual defense: Adversarial, traceable, and invisible robust watermarking against face swapping. *IEEE Transactions on Information Forensics and Security*, 2024.

[86] Y. Zhao, B. Liu, M. Ding, B. Liu, T. Zhu, and X. Yu. Proactive deepfake defence via identity watermarking. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 4602–4611, 2023.

[87] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fei. Hidden: Hiding data with deep networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 657–672, 2018.

[88] J. Zhu, Y. Zhang, and L. Fei-Fei. Stegformer: Content-adaptive steganography with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20606–20615, 2022.