Optically Authenticated Hyperspectral Imaging: Detection and Restoration of Manipulated Data

Pablo Gomez, Roman Jacome *Student Member, IEEE*, Hans Garcia *Member, IEEE*, Iñaki Esnaola and Henry Arguello, *Senior Member, IEEE,*

***Abstract*—Ensuring the authenticity and integrity of hyperspectral images is crucial for remote sensing applications, as unauthorized spectral modifications can compromise data reliability and decision-making processes. Traditional post-acquisition authentication methods leave hyperspectral data vulnerable to manipulation before applying digital protection mechanisms. We present an acquisition-stage optical authentication framework that embeds an optical key directly in the light path of a line-scan hyperspectral imager before digitization. The optical key is designed from scene spectral statistics and jointly optimized for imperceptibility, manipulation detection, and post-attack spectral restoration in a data-driven manner. We initialize and optimize the optical key using distributions parameterized by per-band spectral variance from the training data, which adaptively scales the embedding strength across bands and improves robustness to variance-driven perturbations. To evaluate its resilience, we simulate an optimal attack scenario in which an adversary randomly alters 20%–80% of pixels within a target class. A detection network achieves a maximum accuracy of 97% under light perturbation and maintains above 92% even at 80% manipulation, while maintaining classification performance within 0.6-1.2% of unsigned data. The restoration network brings spectral signatures to within 0.9-1.3° SAM. Experiments on simulated and laboratory datasets show robustness gains of up to 4.7% over traditional methods, with preserved downstream performance.**

***Index Terms*—Hyperspectral image, manipulation detection, optical signing, watermarking.**

I. Introduction

HYPERSPECTRAL images (HSI) are one of the most important modalities in satellite-based remote sensing [1]. After optical sensing, the HSI is digitally compressed and sent to Earth, where critical decisions are made in crucial applications such as environmental monitoring [2], security [3], agriculture [4], weather forecasting [5], and natural disaster management [6]. Two fundamental concerns arise in this area: confidentiality and integrity. Confidentiality focuses on ensuring that unauthorized entities, safeguard sensitive data, do not access hyperspectral images. Integrity addresses the authenticity of the image itself, ensuring that unauthorized entities do not access hyperspectral images, safeguarding sensitive data. In this paper, the principal challenge is ensuring the integrity of hyperspectral images, which is achieved by verifying their authenticity, as any manipulation can significantly affect the accuracy of the decisions based on these images. Ensuring the authenticity of the satellite data is essential. Once the HSI is digitized on the satellite, the data can be immediately compromised, as the satellite is electronically controlled. Therefore, any security algorithm applied to the data after discretization cannot be entirely trusted. To mitigate this issue, an authenticity signature can be applied optically [7]. Several optical cryptographic techniques can be used to ensure data authenticity [8]; however, most of these methods modify the entire image by encoding the phase of the signal, transforming it into a subspace where the compression properties of the original signal are completely lost. Reversing the encoded signal to apply a compression algorithm is not a viable solution, as this process requires the encoding optical key, and it is assumed that the satellite could be digitally compromised. An alternative approach is optical watermarking, which preserves the data structure and its compression properties. However, if the satellite is digitally compromised, an attacker can easily detect and remove the optical key or watermark, once again rendering the data completely vulnerable.

To address these challenges, we propose an authentication framework that embeds an optical key directly into the sensing path before digitization, ensuring that any manipulation of the data can be detected and restored. The embedded optical key is designed in a data-driven manner, ensuring that it is imperceptible to both the attacker and the end user while preserving the fundamental spectral characteristics of the data for compression and transmission and for maintaining performance in downstream tasks such as classification tasks. To optimize the optical key, we consider an optimal attacker who aims not to destroy the data but to modify the hyperspectral cube in a way that remains undetected by the user. This worst-case scenario assumes that the attacker has full knowledge of the cube and its classes, allowing them to transform a specific pixel class to another existing class. Based on this attack model, the optical key is optimized to counteract such manipulations by leveraging two deep neural networks. The first network performs manipulation detection by analyzing both the original and manipulated hyperspectral cubes along with the optical key used for authentication, generating a manipulation map at the acquisition line level. Once the manipulated pixels are identified, the second network takes the manipulation map and the manipulated data as input to restore the altered spectral signatures. Notably, the optimized optical key converges to a distribution directly related to the mean absorbance of the entire dataset, which suggests that the optimization process leads to an optical key energy

Pablo Gomez, Roman Jacome, and Hans Garcia are with the School of Electrical, Electronics, and Telecommunications Engineering of Universidad Industrial de Santander, Colombia. Henry Arguello is with the School of Systems Engineering and Informatics of Universidad Industrial de Santander, Colombia. Iñaki Esnaola is with the School of Electrical and Electronic Engineering, University of Sheffield, UK, and with the Department of Electrical and Computer Engineering, Princeton University, USA.

e-mail: {pablo2228330@correo. rajaccar@correo, hayegaar@ henarfu@}uis.edu.co and esnaola@sheffield.ac.uk

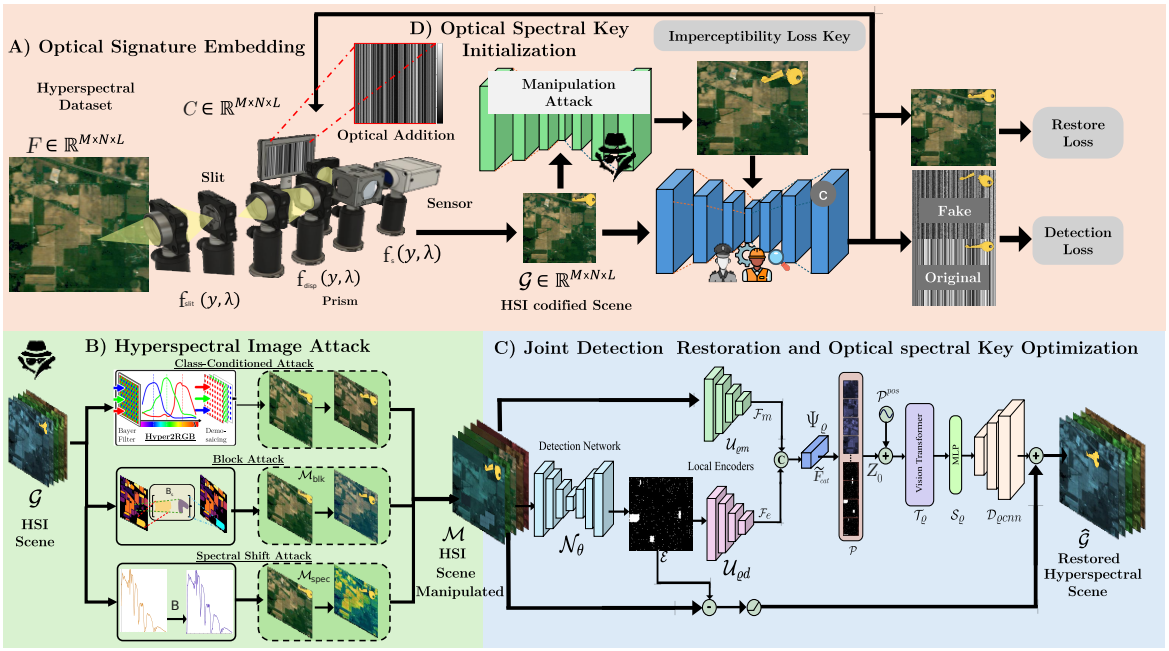

Fig. 1. Overview of the proposed optical authentication framework. A) Optical Signature Embedding: a pushbroom acquisition system integrates an additive spectral key $\mathcal{C}$ into the light path before digitization to ensure pre-acquisition integrity. B) Hyperspectral Image Attack: three types of attacks are applied to signed hyperspectral image. C) Joint Detection, Restoration, and Key Optimization: the detection network $\mathcal{N}_\theta$ estimates the embedded key and localizes tampered pixels, while the restoration network $\mathcal{R}_\varrho$ restore the original spectra using the manipulated image and error map, minimizing detection loss $\mathcal{L}_{det}$, restoration loss $\mathcal{L}_{rec}$, and imperceptibility loss $\mathcal{L}_{sen}$. D) Optical Spectral Key Initialization: the key is derived from the spectral statistics of the scene and optimized for imperceptibility, manipulation detection, and restoration. The spectral key is delivered via a side optical path coupled to the entrance slit, reaching the detector together with the dispersed scene.

distribution inversely proportional to the mean spectral energy of the dataset used. The main contributions of this work are as follows:

- Acquisition-stage optical authentication via light-path embedding of a variance-aware optical key directly in the pushbroom acquisition path, enabling physical authentication at acquisition time and preventing pre-digitization attack while preserving the spectral structure required for compression and analysis.
- Data-driven optical key design guided by mean-spectrum absorbance and per-band spectral variance, adapting the embedding strength across wavelengths to achieve imperceptibility and stable downstream performance.
- Joint optimization of the optical key and a lightweight detection–restoration network, attaining >92% manipulation detection accuracy under severe attacks and up to 4.7% robustness gains over traditional watermarking.
- Dual-arm optical prototype and cross-dataset validation on simulated and laboratory acquisitions, confirming practical feasibility and resilience across manipulation regimes and noise conditions.

II. Related Work

Methods for hyperspectral image authentication and detection can be categorized into post-processing [8]–[16] and acquisition-based approaches [12], [13], [17]–[29]. Most research has focused on post-processing authentication, where security features are embedded or analyzed after the image has been acquired [8]–[11], [14]–[16]. However, such approaches do not prevent modifications during transmission or storage. In contrast, acquisition-based methods integrate authentication mechanisms directly into the optical system, ensuring data integrity from the moment of capture. In the following, we analyze these approaches, their limitations, and how our proposed methodology fills existing gaps. The steganographic [16] and blockchain-integrated watermarking [15] methods provide confidentiality and traceability in the digital domain, but do not address pre-digitization integrity.

A. Post-Processing Authentication Methods

Post-processing authentication techniques operate on the digitalized hyperspectral image $\mathcal{F}(x, y, \lambda)$, where $x \in \mathbb{R}$ and $y \in \mathbb{R}$ represent the spatial dimensions and $\lambda \in \mathbb{R}^+$ represents the spectral dimension. These approaches ensure authenticity through digital modifications, often at the cost of additional storage and processing overhead.

*1) **Cryptographic Authentication**:* Cryptographic methods ensure data integrity by generating external identifiers such as hashes or digital signatures rather than embedding authentication within the hyperspectral image [8]. Hash-based authentication computes a fingerprint of the image, allowing manipulation detection but requiring secure metadata storage, formulated as $H = \text{Hash}(\mathcal{F})$, where $H \in \{0,1\}^n$ is a fixed-length binary digest representing the unique hash value of the hyperspectral cube $\mathcal{F}$. Digital signatures use asymmetric cryptography to verify authenticity. While robust for integrity verification, cryptographic authentication does not embed authentication within the image itself, making it inef-

fective against localized spectral manipulations and dependent on external metadata. In contrast, our approach integrates a spectral key at acquisition, ensuring authentication is intrinsic to the data and resistant to targeted spectral modifications.

*2) **Watermarking-Based Approaches**:* Watermarking techniques embed an imperceptible signature into $\mathcal{F}$, ensuring modification detection [9]–[11]. These methods vary in how the watermark is embedded, with popular approaches including Spatial Domain Watermarking [30], Discrete Cosine Transform (DCT), Wavelet Transform [13], [31], [32], Singular Value Decomposition [13], and Fourier Transform (FT)-based -Watermarking [33]. DCT, DWT, and SVD watermarking operates by transforming the image into a specific domain, where the watermark is embedded in the least significant frequency coefficients. This technique distributes the watermark information across the spectral components, making it robust against spatial manipulations but susceptible to strong compression or filtering. FT-based watermarking, on the other hand, encodes the watermark in the frequency spectrum, leveraging phase alterations or magnitude modulation to increase robustness against geometric transformations. Recent advances include hyperspectral watermarking using discrete wavelet transform combined with forensic-based Archimedes optimization [14], zero-watermarking schemes integrated with multi-channel PCNN and blockchain for secure certification in remote sensing [15], and steganographic data encryption approaches for hyperspectral imaging [16]. While the first two methods enhance robustness and provide traceability in the digital domain, they remain vulnerable to manipulations introduced before acquisition because the signature is not physically embedded at acquisition time. Steganographic encryption, on the other hand, focuses on concealing encrypted information within hyperspectral data for confidentiality, which differs from our goal of continuous integrity verification of the detected scene. A major drawback of frequency-domain watermarking is that the transformation and inverse transformation processes can introduce reconstruction errors, potentially altering the spectral characteristics of hyperspectral images [33]. Additionally, watermarking schemes typically struggle with maintaining imperceptibility while ensuring detectability, as a highly robust watermark may degrade classification accuracy [13], [31], [32]. A general watermarking model can be expressed as:

$$\mathcal{F}' = \mathcal{T}_u(\mathcal{F}) + \alpha \cdot \mathcal{U}, \tag{1}$$

where $\mathcal{T}_u : \mathcal{F} \rightarrow \mathcal{F}'$ represents the transformation (such as DCT or FT), $\mathcal{F}'$ represents the resulting image after the transformation, $\mathcal{U} \in \mathcal{Z}$ is the watermark pattern, and $\alpha \in \mathbb{R}^+$ controls its strength. The effectiveness of watermarking depends on balancing robustness, imperceptibility, and computational efficiency. Beyond generic image watermarking, multi-/hyperspectral schemes operate over the spectral signature as a whole and enable tamper localization in multiband settings, for example, transform- and forensic-based designs for HSI. Zero-watermarking additionally preserves the imagery by binding external signatures to content descriptors and ledgers. All these families remain post-digitization; in contrast, our approach enforces pre-digitization integrity via additive optical injection during capture.

B. Acquisition-Based Authentication Methods

Unlike post-processing methods, authentication at the acquisition stage aims to integrate an optical signature within the imaging system itself [12], [13]. Several approaches have been developed for RGB imaging, where a Coded Aperture (CA) or Color-Coded Aperture (CCA) is introduced in an intermediate optical plane between the scene and the sensor [17]–[20], employing a Fourier Transform to include the encryption [21] and using phase retrieval to encrypt and measure the authenticity of the image [22], [23]. These apertures modulate the incoming light in a structured form, embedding a unique spectral signature into the acquired image. The inserted optical pattern serves as an authentication key, allowing verification of whether an image has been manipulated post-acquisition. A key limitation of these methods is their impact on the performance of primary imaging tasks. Since the CA or CCA modifies the spectral content reaching the sensor, it may introduce distortions that degrade classification or detection accuracy. Additionally, ensuring imperceptibility of the authentication key is challenging, as an overly intrusive signature may interfere with standard image analysis. These challenges have limited the adoption of CA-based authentication techniques in high-precision applications like hyperspectral imaging. The integration of a coded aperture can be represented as $\mathcal{G} = \mathcal{T}_{CA}(\mathcal{F}, \mathcal{K}) + \mathcal{N}$, where $\mathcal{T}_{CA}$ represents the operation that can involve multiplicative encoding, phase modulation, or structured filtering of the spectral content induced by $\mathcal{K} \in \mathcal{Z}$ that is a coded aperture key, and $\mathcal{N} \in \mathcal{R}$ denotes acquisition noise. Depending on the implementation, $\mathcal{T}$ can involve multiplicative encoding, phase modulation, or structured filtering of the spectral content.

A recently proposed methodology [24] follows a different spectral imaging approach by optimizing the spectral key not only for authentication but also for information restoration in manipulated regions. The key insertion process in [24] is based on multiplicative spectral modulation, formulated as

$$\mathcal{G} = \mathcal{F} \odot \mathcal{C} + \mathcal{W}, \tag{2}$$

where $\mathcal{G}$ is the acquired hyperspectral image, $\mathcal{F}$ is the original spectral data, $\mathcal{C}$ is the embedded spectral key and $\odot$ denotes element-wise multiplication. This approach alters the spectral response of the data more aggressively, which impacts classification and detection accuracy. Furthermore, the key in [24] was initialized as random noise, without adapting its distribution to the spectral energy profile of the data set, limiting its robustness and effectiveness. Although not designed for authentication, several recent works have focused on hyperspectral image completion and restoration, aiming to recover missing or corrupted spectral information after acquisition. Notable examples include low-rank tensor convolution-like dictionaries for high-dimensional representation [25], pseudo-side information regularization [26], and parametric tensor sparsity models [27]. Other frameworks jointly address restoration and analysis, such as conditional diffusion models

for underwater HSI restoration and target detection [28], and end-to-end denoising classification pipelines [29]. Although these methods can effectively reconstruct spectral content and suppress noise, they operate entirely in the digital domain and cannot detect or prevent manipulations introduced before digitization. Our approach differs fundamentally by embedding an optical key directly into the sensing path, enabling pre-digitization integrity verification while preserving the spectral structure required for downstream analysis.

III. Proposed Method

In this section, we present the proposed optimization framework and the sensing model of the designed spectral pushbroom optical system. Pushbroom systems are the most widely used architecture for acquiring satellite-based hyperspectral images, as they offer high spatial and spectral resolution while being compatible with the constraints of satellite platforms. Our system integrates an optical key directly during the image acquisition process, enhancing authenticity and ensuring robustness against manipulations, as illustrated in Fig.1. A preliminary benchtop prototype validating the side-path injection and measuring $c(\lambda)$ is provided in the Supplementary.

A. Optical Signature Embedding

First, the proposed pushbroom system acquires hyperspectral image while embedding an optical key to ensure authenticity and manipulation detection as shown in Fig. 1 a). A spatial slit along the x-axis selects a narrow vertical region of the scene, filtering out information from other spatial positions. This operation is modeled by the function $f_{slit}(t, y, \lambda) = f(x_t, y, \lambda)$ where $x_t = v \cdot t$, x_t is the x-axis position at instant time t and v refers to the speed of the camera movement. The transmitted light is then spectrally dispersed by a prism or grating, mapping each wavelength to a specific position along the y-axis according to a dispersion factor $\alpha(\lambda)$:

$$f_{disp}(t, y, \lambda) = f_{slit}(t, y - \alpha(\lambda), \lambda). \tag{3}$$

Here, $\alpha(\lambda)$ is typically modeled with a proportional constant that depends on the dispersive element. At this stage, an optical key $c(\lambda) \in \mathbb{R}^+$ is introduced to modulate the hyperspectral data. Note that the optical key depends only on the wavelength and remains constant for all spatial coordinates. The modulated signal is obtained as:

$$f_s(t, y, \lambda) = f_{disp}(t, y, \lambda) + c(\lambda). \tag{4}$$

In the pushbroom setup, the optical key $c(\lambda)$ is injected through a secondary optical side path that is coaligned with the entrance slit and directed onto the detector contemporaneously with the dispersed scene. This realizes an additive pre-digitization injection; no digital post-processing is used to introduce the optical key. We assume that the pushbroom system operates in continuous platform motion along-track during acquisition. This continuous motion enables the system to capture a full scene by sequentially collecting spectral data line by line. To implement this scanning process, the signal is discretized by sampling at regular time intervals, which correspond to discrete positions along the x-axis. The final discrete HSI is then constructed by stacking the individual 2D sensor acquisitions obtained at each discrete slit position. By concatenating these 2D images along the x-axis, the complete hyperspectral cube is formed:

$$\mathcal{G}_{m,n,\ell} = \int_{m\Delta_t}^{(m+1)\Delta_t} \int_{n\Delta_y}^{(n+1)\Delta_y} \int_{\ell\Delta_\lambda}^{(\ell+1)\Delta_\lambda} f_s(t, y, \lambda) \times \operatorname{rect}\left(\frac{t}{\Delta_t} - m, \frac{y}{\Delta_y} - n, \frac{\lambda}{\Delta_\lambda} - \ell\right) dt dy d\lambda, \tag{5}$$

where Δ_t refers to the resolution of the slit position which is related to the velocity v, Δ_y is the pitch size of the sensor and Δ_λ is the spectral pitch size which depends on the dispersive element $\alpha(\lambda)$ response, $m = 1, \ldots, M$ indexes the discrete slit positions, $n = 1, \ldots, N$ indexes the spatial positions along the y-axis, and $\ell = 1, \ldots, L$ indexes the spectral bands. Here, $\operatorname{rect}(\cdot, \cdot, \cdot)$ is the 3D indicator of the voxel (m, n, ℓ): it is 1 inside the integration ranges of (t, y, λ) in (5) and 0 otherwise. The discrete representation of the optical key is:

$$\mathbf{c}_\ell = \int_{\ell\Delta_\lambda}^{(\ell+1)\Delta_\lambda} c(\lambda) \operatorname{rect}\left(\frac{\lambda}{\Delta_\lambda} - \ell\right) d\lambda, \tag{6}$$

where $\mathbf{c} \in \mathbb{R}^L$. Equation (6) is provided for clarity, showing the equivalent formulation of (5) when the rectangular window is expressed through explicit integration limits. and we formulate the tensor representation of the optical key $\mathcal{C} \in \mathbb{R}^{M \times N \times L}$ as

$$\mathcal{C} = \mathcal{J} \otimes_1 \underbrace{(\mathbf{1}_N \otimes \mathbf{c})}_{\text{replicate on } y\text{-axis}} \tag{7}$$

where $\mathcal{J} \in \mathbb{R}^{M \times N \times L}$ is a tensor of ones and $\otimes$ is the Kronecker product. This formulation replicates the optical key $\mathbf{c}$ along all spatial dimensions. Equation (7) is a compact tensorial notation stating that the spectral key c_ℓ is replicated across the spatial dimensions (m, n) and added to the reflectance cube, yielding a signed hyperspectral cube. Finally, the discrete sensing model

$$\mathcal{G} = \mathcal{C} + \mathcal{F} + \zeta, \tag{8}$$

where ζ accounts for system noise. Structured redundancy of $\mathcal{C}$ ensures unique modulation per spectral band, reinforcing authentication mechanisms. The noise component ζ is naturally introduced during acquisition and influences both the original HSI and the modulated optical key. A photo of the bench, side-path alignment, is provided in supplementary material.

B. Hyperspectral image Attack

We consider three manipulation families: (i) a class-conditioned attack that preserves the RGB appearance while altering the spectra ((10)); (ii) a block attack enforcing spatially coherent $s \times s$ patches ((11)); and (iii) a spectral shift attack perturbing a limited band range ((12)). We start from the class-conditioned attack and then instantiate the block and spectral variants used in our experiments (see Fig.4; and additional data sets are in the Supplement). A sophisticated adversary does not merely inject noise or remove spectral bands but strategically manipulates hyperspectral data while maintaining spatial and spectral consistency. This can be

framed as an optimization problem where the goal is to modify a target class while preserving the overall appearance of the HSI. To evaluate the robustness of the optimized optical key, we simulate the worst-case scenario, assuming that the attacker has full access to the signed hyperspectral cube and its class labels. In this case, we use a training dataset, which follows the spectral distribution of the hyperspectral data defined as $\tilde{\mathcal{F}}$, and after the optical signing process, we obtain $\tilde{\mathcal{G}}$. This manipulation methodology allows the attacker to alter the data in a way that the changes remain undetectable to the end user at a glance, effectively emulating an attack that preserves the visual integrity of the image while modifying its spectral information. This type of attack poses a significant threat to applications such as remote sensing and environmental monitoring, where manipulated data could go unnoticed during visual inspections, yet still affect decisions based on the data.

***Class-conditioned attack**:* To analyze this risk, we define a spectral response function, $\mathcal{S}_r$, which projects the hyperspectral cube into an RGB image:

$$\mathcal{I}_r(\tilde{\mathcal{G}}) = \sum_{l=1}^{L} (\mathcal{S}_r \odot \tilde{\mathcal{G}}_l), \tag{9}$$

where $\mathcal{I}_r$ is the intensity of the projected RGB image for channel r, and $\mathcal{S}_r$ is the spectral response function. This projection enhances visualization for non-expert users while allowing visual assessment of spectral integrity. An optimal manipulation alters a target class $\mathcal{P}_i$ while minimizing changes to a secondary class $\mathcal{P}_j$. The objective ensures that the manipulated HSI $\mathcal{M}$ remains undetectable under standard verification:

$$\mathcal{M}^* = \arg\min_{\mathcal{M}} \Bigg[\sum_{r=1}^{3} \|\mathcal{I}_r(\tilde{\mathcal{G}} - \mathcal{M})\|_2^2 + \|\nabla \mathcal{M}\|_2^2 + \gamma \|\tilde{\mathcal{G}}_{\mathcal{P}_i} - \mathcal{M}_{\mathcal{P}_j}\|_2^2 \Bigg]. \tag{10}$$

Here, $\mathcal{M}$ represents the manipulated HSI, and $\tilde{\mathcal{G}}$ the original signed image. The term $\|\mathcal{I}_r(\tilde{\mathcal{G}} - \mathcal{M})\|_2^2$ ensures minimal perceptual changes in the projected RGB image, $\|\nabla \mathcal{M}\|_2^2$ enforces spatial smoothness, and $\gamma \|\tilde{\mathcal{G}}_{\mathcal{P}_i} - \mathcal{M}_{\mathcal{P}_j}\|_2^2$ minimizes spectral discrepancies to maintain classification plausibility. The optimization is solved iteratively using gradient-based methods, such as Adam or Stochastic Gradient Descent (SGD), which update the manipulated hyperspectral image $\mathcal{M}$ by minimizing the objective function in Equation (10). The gradient with respect to $\mathcal{M}$ is computed to adjust its spectral and spatial components. This ensures convergence to an optimal solution that balances perceptual fidelity, spatial smoothness, and spectral integrity. The formulation highlights the complexity of hyperspectral manipulation, requiring the preservation of optical key data characteristics while minimizing detectable changes. This highlights the need for robust authentication to detect attacks and preserve data authenticity.

***Block attack**:* We constrain (10) to manipulations confined to a block mask $\mathbf{B}_s \in \{0,1\}^{H \times W}$ composed of non-overlapping $s \times s$ patches over the target class $\mathcal{P}_i$. Using a donor field $\hat{\mathcal{G}}_{\mathcal{P}_j}$ with spectra from class $\mathcal{P}_j$, the manipulated cube is described by:

$$\mathcal{M}_{\text{blk}}(x,y,\lambda) = \tilde{\mathcal{G}}(x,y,\lambda)\,\big(1 - \mathbf{B}_s(x,y)\big) + \Big(\alpha\,\tilde{\mathcal{G}}(x,y,\lambda) + (1-\alpha)\,\hat{\mathcal{G}}_{\mathcal{P}_j}(x,y,\lambda)\Big)\,\mathbf{B}_s(x,y). \tag{11}$$

***Spectral-band shift**:* We model an additive shift δ on a contiguous set of bands $\mathcal{B} \subset \{1,\dots,L\}$ applied only to pixels in $\mathcal{P}_i$, while preserving all remaining spectra:

$$\mathcal{M}_{\text{spec},i}(x,y,\lambda) = \tilde{\mathcal{G}}(x,y,\lambda) \;+\; \mathbf{1}_{\mathcal{P}_i}(x,y)\,\mathbf{1}_{\mathcal{B}}(\lambda)\,\delta, \tag{12}$$

with $\delta = 0.03$. We consider three representative band ranges: red–edge $\mathcal{B}_{\text{RE}} = [680; 750]$, NIR $\mathcal{B}_{\text{NIR}} = [1000; 1100]$, and SWIR $\mathcal{B}_{\text{SWIR}} = [2200; 2300]$. Again, (10) is solved under the structural constraint $\mathcal{M} \equiv \mathcal{M}_{\text{spec}}$. In this manuscript, we consider integrity attacks that manipulate the signed cube without attempting to estimate or remove the optical key. Optical Key-extraction attacks (e.g., learning or averaging $c(\lambda)$) are explicitly out of scope for this work and are deferred to future works where time-varying or adaptive optical keys will be assessed. Our present focus is the pre-digitization additive configuration and its detect/restore pipeline under representative content manipulations.

C. Joint Detection, Restoration, and Optical Spectral Key Optimization

With the optical key $\mathcal{C}_{D_i}$ initialized as described in Sec. III-D, we propose a joint framework that simultaneously detects manipulations and restores the original HSI. Unlike conventional methods that address detection and restoration as separate tasks, our approach couples them so that both processes reinforce each other, enhancing robustness against spectral forgeries. Highlighting that the training phase is performed with the forms $\tilde{\mathcal{F}}$ that corresponds to a similar spatio-spectral distribution of the acquired cube. The main challenge is to balance detection accuracy, restoration fidelity, and imperceptibility while preserving hyperspectral data integrity in the optical domain.

Network architectures. The detection network $\mathcal{N}_\theta(\cdot)$ adopts a 3D convolutional encoder–decoder design to jointly exploit spectral and spatial correlations in the hyperspectral cube $\mathcal{F} \in \mathbb{R}^{M \times N \times L}$. It compresses the cube into a compact latent representation and regresses a one-dimensional optical key, which is then spatially replicated to match the scene dimensions. The restoration network $\mathcal{R}_\varrho(\cdot)$ is a hybrid CNN and transformer model. It takes as input the manipulated HSI and the error map produced by $\mathcal{N}_\theta(\cdot)$, using both sources to guide reconstruction, as show in Fig. 1 c)

Optimization objective. We jointly optimize three coupled components, an optical spectral key embedded at acquisition, a detector that estimates the key and localizes manipulations, and a restorer that reconstructs corrupted spectra guided by the detector error map, as shown in Fig. 1 c). The joint loss function is formulated:

$$\mathcal{L} = \lambda_{\text{det}}\mathcal{L}_{\text{det}} + \lambda_{\text{sen}}\mathcal{L}_{\text{sen}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}}, \tag{13}$$

where $\mathcal{L}_{\text{det}}$ constrains optical key estimation and manipulation localization, $\mathcal{L}_{\text{rec}}$ enforces high-fidelity spectral restoration, and $\mathcal{L}_{\text{sen}}$ constrains imperceptibility of the embedded optical key. The overall training objective is:

$$\{\mathcal{C}_D{}^*, \theta^*, \varrho^*\} = \arg\min_{\mathcal{C}_D,\theta,\varrho} \mathbb{E}_{\mathcal{M}} \left[\lambda_{\text{det}}\mathcal{L}_{\text{det}} + \lambda_{\text{sen}}\mathcal{L}_{\text{sen}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}}\right], \tag{14}$$

where $\mathcal{C}_D{}^*$, θ^*, and ϱ^* represent the optimized optical key, detection and restoration model parameters, respectively.

a) Detection Loss: The detection loss $\mathcal{L}_{\text{det}}$ minimizes discrepancy between the estimated and actual optical key for clean HSIs while maximizing mismatch for manipulated HSIs, enabling precise localization of tampered pixels:

$$\mathcal{L}_{\text{det}} = \frac{\|\mathcal{C}_D - \mathcal{N}_\theta(\tilde{\mathcal{G}})\|_2}{\|\mathcal{C}_D\|_2} + \left(1 - \frac{\|\mathcal{C}_D - \mathcal{N}_\theta(\mathcal{M})\|_2}{\|\mathcal{C}_D\|_2}\right)^2, \tag{15}$$

where $\mathcal{N}_\theta : \mathbb{R}^{M\times N\times L} \to \mathbb{R}^{M\times N}$ outputs a 1D optical key repeated along spatial axes.

b) Restoration Loss: The restoration loss $\mathcal{L}_{\text{rec}}$ enforces high-fidelity restoration of the signed HSI from the manipulated image and its error map:

$$\mathcal{L}_{\text{rec}} = \|\mathcal{R}_\varrho(\mathcal{M}, \mathcal{E}) - \tilde{\mathcal{G}}\|_2^2, \tag{16}$$

where $\mathcal{E} = (\mathcal{C}_D - \mathcal{N}_\theta(\mathcal{M}))^2$ is the optical key estimation error map. The error map $\mathcal{E}$ measures the consistency between the embedded optical key and the observed spectral data. For non-manipulated samples, the detector accurately estimates the optical key ($\mathcal{N}_\theta(\tilde{\mathcal{G}}) \approx \mathcal{C}$), yielding near-zero error. Spectral manipulations disrupt the physical relation in 8, producing mismatches in the estimated key and high responses in $\mathcal{E}$ that localize tampered regions. The proposed restoration is a local CNN encoder stage that produces feature maps of the error map and manipulated scene, which are fused and passed through a global Transformer; the resulting token representation is linearly "unpatchified" back to a feature map and finally decoded to band space by a shallow CNN. Formally, the restoration model is defined as $\mathcal{R}_\varrho(\mathcal{M}, \mathcal{E}) = \mathcal{D}_{\varrho_{dec}}(\mathcal{M}, \mathcal{E}) + \text{ReLU}(\mathcal{M} - \mathcal{E})$. The decoding model is modeled as follows:

$$\mathcal{D}_{\varrho_{dec}} := \mathcal{D}_{\varrho_{\text{cnn}}} \circ \mathcal{P}^{-1} \circ \mathcal{S}_\varrho \circ \mathcal{T}_\varrho \circ \mathcal{P}_\varrho \circ \Psi_\varrho \tag{17}$$

and apply it to the concatenated encodings $\text{cat}(\mathcal{U}_{\varrho_m}(\mathcal{M}), \mathcal{U}_{\varrho_d}(\mathcal{E}))$. We employ local encoders and fusion. Let

$$\mathcal{F}_m = \mathcal{U}_{\varrho_m}(\mathcal{M}) \in \mathbb{R}^{C_L\times H\times W}, \tag{18}$$

$$\mathcal{F}_e = \mathcal{U}_{\varrho_d}(\mathcal{E}) \in \mathbb{R}^{C_L\times H\times W}. \tag{19}$$

We fuse them with a 1×1 projection to the Transformer width D:

$$\widetilde{\mathcal{F}}_{cat} = \Psi_\varrho\big(\text{cat}(\mathcal{F}_m, \mathcal{F}_e)\big) \in \mathbb{R}^{D\times H\times W} \tag{20}$$

Here $H'\times W'$ is the Transformer canvas (equal to or a resized version of $H\times W$). We split $\widetilde{\mathcal{F}}_{cat}$ into non-overlapping $P\times P$ patches and map each patch to a token via a strided $D\to D$ projection:

$$\mathcal{Z}_0 = \mathcal{P}_\varrho(\widetilde{\mathcal{F}}_{cat}) \in \mathbb{R}^{N_p\times D}, \qquad N_p = \frac{H'}{P}\cdot\frac{W'}{P}. \tag{21}$$

Learned positional embeddings $\mathcal{P}^{\text{pos}} \in \mathbb{R}^{N_p\times D}$ are added once: $\mathcal{U}_0 = \mathcal{Z}_0 + \mathcal{P}^{\text{pos}}$. We employ transformer blocks with L_T pre-norm blocks, for $t = 1, \ldots, L_T$:

$$\mathcal{U}'_t = \mathcal{U}_{t-1} + \text{MHA}\big(\text{LN}(\mathcal{U}_{t-1})\big), \tag{22}$$

$$\mathcal{U}_t = \mathcal{U}'_t + \text{MLP}\big(\text{LN}(\mathcal{U}'_t)\big), \tag{23}$$

where LN is LayerNorm and the MLP is two linear layers with GELU. Multi-head attention with H_a heads and head width $d_h = D/H_a$ is

$$\text{MHA}(\mathcal{U}) = \left[\text{softmax}\Big(\frac{(\mathcal{U}\mathbf{W}_Q^{(h)})(\mathcal{U}\mathbf{W}_K^{(h)})^\top}{\sqrt{d_h}}\Big)(\mathcal{U}\mathbf{W}_V^{(h)})\right]_{h=1}^{H_a}\mathbf{W}_O. \tag{24}$$

After the last block and a final LayerNorm as $\mathcal{T} = \text{LN}(\mathcal{U}_{L_T}) \in \mathbb{R}^{N_p\times D}$. Linear within-patch synthesis and unpatchify. Each token is linearly expanded to $P\times P$ patch with D feature planes:

$$\mathcal{R}_p = \mathcal{S}_\varrho(\mathcal{T}) \in \mathbb{R}^{N_p\times(P^2D)}. \tag{25}$$

Reassembling the N_p patches gives a feature map as $\mathcal{S} = \mathcal{P}^{-1}(\mathcal{R}_p) \in \mathbb{R}^{D\times H'\times W'}$, which is bilinearly resized to $\mathbb{R}^{D\times H\times W}$ if $H'\times W' \neq H\times W$. Shallow CNN head to band space. A light decoder maps features to band space $\mathcal{D}_{\varrho_{\text{cnn}}}(\mathcal{S}) \in \mathbb{R}^{B\times H\times W}$.

c) Imperceptibility Loss Function: The imperceptibility loss $\mathcal{L}_{\text{sen}}$ ensures that the embedded optical key does not introduce visual or spectral distortions.The Eq 26 represents the preservation of spectral integrity for downstream analysis:

$$\mathcal{L}_{\text{sen}} = \|\tilde{\mathcal{G}} - \tilde{\mathcal{F}}\|_2^2. \tag{26}$$

D. Optical Key Initialization

In (8) and the joint objective in (14), the imperceptibility term in (26) penalizes the total embedding energy, while the detection loss in (15) is explicitly normalized by $\|\mathcal{C}_D\|_2$ and therefore acts scale-invariantly, selecting the spectral direction that maximizes detectability per unit energy. The restoration term steers the reconstructor towards the signed cube while exerting negligible directional bias in the small-perturbation regime. Consequently initialization should provide an optical key direction that is optimal for detection under a fixed energy budget. Let $A(x) = -\log_{10}(x)$ denote the bandwise absorbance operator and let F_ℓ denote the clean-band random variable induced by the training set (expectation taken over the empirical training distribution). A detailed derivation of the optical key initialization is provided in Supplementary (Sec. II). Under (8), a first-order expansion in absorbance for a small additive injection c_ℓ gives:

$$\Delta A_\ell \;=\; A(F_\ell + c_\ell) - A(F_\ell) \;\approx\; -\frac{c_\ell}{F_\ell \ln 10}, \tag{27}$$

the detectable contrast in absorbance scales as $|c_\ell|/F_\ell$. Because $\mathcal{L}_{\text{det}}$ divides by the optical key energy, a natural detection-aligned, scale-free objective for the optical key *direction* c (enforcing $\|c\|_2 = 1$) is:

$$\max_{\|c\|_2=1} \mathbb{E}\left[\sum_{\ell=1}^{L}(\Delta A_\ell)^2\right] \approx \max_{\|c\|_2=1} \sum_{\ell=1}^{L} a_\ell\, c_\ell^2 \text{ with } a_\ell \triangleq \mathbb{E}\big[F_\ell^{-2}\big]. \tag{28}$$

The maximizer of (28) aligns c with the weights per band a_ℓ, that is, it allocates more embedding power where the expected inverse squared reflectance is larger. To connect a_ℓ with a robust pretraining statistic, Jensen's inequality yields:

$$\mathbb{E}\left[\frac{1}{F_\ell^2}\right] \geq \frac{1}{\left(\mathbb{E}[F_\ell]\right)^2} = \frac{1}{\bar{\mathbf{f}}_\ell^2}, \tag{29}$$

where $\bar{\mathbf{f}} \in \mathbb{R}^L$ denotes the average spectrum across the $M \times N$ spatial grid. Hence a stable surrogate for the optimal direction is $c_\ell \propto 1/\bar{\mathbf{f}}_\ell$, which is monotonically equivalent to using the per-band absorbance $-\log_{10} \bar{\mathbf{f}}_\ell$. This motivates initializing the optical key with the absorbance of the mean spectrum and subsequently refining it within the full joint objective. An important aspect of optical key optimization is its initialization. To this end, we devise different random distributions based on the statistical distribution of spectral energy across all spatial samples of the training data $\tilde{\mathcal{F}}$. Specifically, we computed the mean spectral signature as:

$$\bar{\mathbf{f}}_\ell = \frac{1}{M\,N} \sum_{m=1}^{M} \sum_{n=1}^{N} \tilde{\mathcal{F}}_{m,n,\ell}, \qquad \ell = 1, \ldots, L. \tag{30}$$

Since the optical key $\mathbf{c} \in \mathbb{R}^L$ must secure the HSI data with minimal perceptual impact, it is natural to allocate embedding strength inversely to the inherent spectral power. Consequently, we initialize the optical key by computing the absorbance of the mean spectrum as $\mathbf{c} = A(\bar{\mathbf{f}})$, ensuring that wavelengths with higher reflectance receive a lower embedding power. This initialization aligns the optical key structure with the intrinsic properties of the dataset, facilitating robust and imperceptible embedding in downstream optimization. Fig. 2 illustrates this approach by comparing the normalized mean spectrum and absorbance signature for 2 datasets. Shaded regions represent feasible embedding ranges: wider gaps allow greater modulation, whereas narrower gaps constrain the optical key magnitude to maintain fidelity. In addition to absorbance, spectral variance further refines the initialization of the optical key. Let $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_L)^\top \in \mathbb{R}^L$ denote the per-band sample variance of $\tilde{\mathcal{F}}$ (each σ_j computed over all $M \times N$ pixels at band j). We also define the global variance scalar $\overline{\sigma} = \frac{1}{L}\sum_{\ell=1}^{L} \sigma_\ell$. We then sample four variance-aware families around the absorbance:

$$\begin{aligned} \mathbf{c}_{D_1} &= \mathbf{c} + \frac{\boldsymbol{\sigma}\,\epsilon}{\gamma}, & \mathbf{c}_{D_2} &= \mathbf{c} + \frac{\epsilon}{\gamma\,\boldsymbol{\sigma}}, \\ \mathbf{c}_{D_3} &= \mathbf{c} + \overline{\sigma}\,\epsilon, & \mathbf{c}_{D_4} &= \mathbf{c} + \frac{\epsilon}{\overline{\sigma}}, \end{aligned} \tag{31}$$

where $\epsilon \sim \mathcal{N}(0,1)$ and $\gamma > 0$ controls perturbation magnitude; the vector-to-tensor expansion to $\mathcal{C}_{D_i}$ follows (7). The direction of $\mathbf{c}$ is fixed by (29); (31) applies variance-aware re-scalings around this direction. Per-band variance $(\mathbf{c}_{D_1}, \mathbf{c}_{D_2})$ adapts to local spectral variability, while global variance $(\mathbf{c}_{D_3}, \mathbf{c}_{D_4})$ provides a scene adjustment. A extended mathematical process of initialization key is in Supplementary.

IV. Simulated and Acquired Datasets Description

To assess realism beyond standardized benchmarks, we include an acquired laboratory scene captured with our push-broom prototype, where practical acquisition factors stress the signing–detection–restoration pipeline. Additional experiments on Salinas, Pavia University, and synthetic laboratory scenes are provided in the Supplementary. Implementation uses PyTorch on a single NVIDIA RTX 4090 GPU with

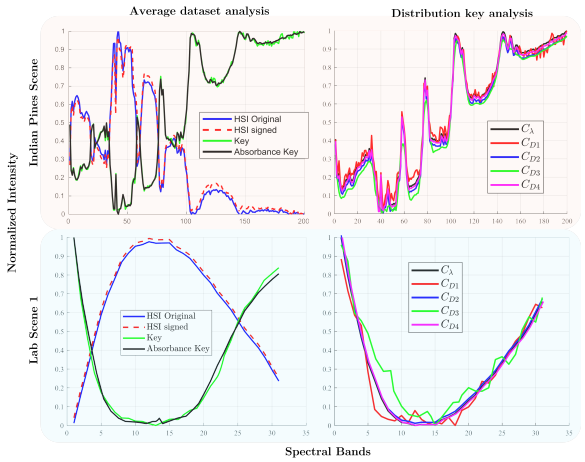

Fig. 2. Optical Key initialization analysis on Indian Pines (top) and Lab Scene 1 (bottom). Left: training-set mean spectrum (HSI Original, blue) vs. signed spectrum (red) with the absorbance seed $A(\bar{\mathbf{f}})$ (black) and the trained optical key (green), showing that signing preserves the spectral shape. Right: optical keys sampled from the initialization families C_λ and C_{D_1}–C_{D_4}, illustrating variability and band-specific emphasis used during optimization.

Algorithm 1 Joint Keying, Manipulation Detection, and restoration for HSI

Require: $\tilde{\mathcal{F}}$, `epochs`, α, $iter$

$\mathbf{c}^0 \sim \mathbf{c}_{\mathbf{D}i}$ ▷ Optical Key Initialization
$\mathcal{C} = \mathcal{J} \otimes (\mathbf{1}_N \otimes \mathbf{c}^0)$
Initialize θ, ϱ
for ℓ = 1:`epochs` **do**
  $\tilde{\mathcal{G}} = \mathcal{C} + \tilde{\mathcal{F}} + \zeta$
  Generate manipulated image $\mathcal{M}$ based on strategies
  for i = 1:$iter$ **do**
    Randomly sample $\overline{\mathcal{M}} \sim \mathcal{M}$ and $\overline{\mathcal{G}} \sim \tilde{\mathcal{G}}$
    $\mathcal{A}_m = \mathcal{N}_\theta(\overline{\mathcal{M}})$
    $\mathcal{A}_s = \mathcal{N}_\theta(\overline{\mathcal{G}})$
    $\mathcal{E} = (\mathcal{A}_m - \mathcal{C})^2$ ▷ Key estimation error map
    $\hat{\mathcal{G}} = \mathcal{R}_\varrho(\overline{\mathcal{M}}, \mathcal{E})$
    $\mathcal{L} = \lambda_{\text{det}}\mathcal{L}_{\text{det}} + \lambda_{\text{sen}}\mathcal{L}_{\text{sen}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}}$ ▷ Compute loss
    $\mathcal{C} = \mathcal{C} - \alpha\nabla_{\mathcal{C}}\mathcal{L}$
    $\theta = \theta - \alpha\nabla_\theta\mathcal{L}$
    $\varrho = \varrho - \alpha\nabla_\varrho\mathcal{L}$
Output $\mathcal{C}, \theta, \varrho$

Adam optimization over 1000 outer epochs and 500 inner iterations per epoch for the detector and restorer networks. Full hyperparameters, additional qualitative results, and reproducibility scripts are available in the Supplementary and will be released at https://github.com/pablogomezt/Spectral_Authenticity_System.git.

V. Results and Discussion

Before reporting results, we clarify the initialization notation used throughout. We denote by D_i the optical key initializations in (31), where D_0 corresponds to the absorbance-based initialization defined in Sec. III-D, D_1 and D_2 denote variance-aware perturbations, and D_3–D_4 use global variance scaling. "Signed D_i" refers to the proposed additive pre-

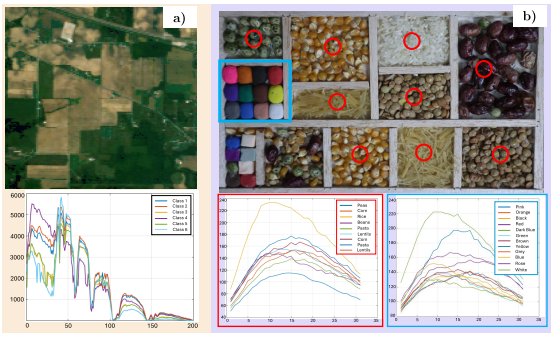

Fig. 3. Datasets and sample spectra. (a) Indian Pines RGB and class mean spectra. (b) Lab Scene 1 RGB (red circles: sampling regions); bottom: mean spectra for materials and color patches.

digitization optical signing with initialization D_i, while "Restored D_i" denotes the output of $\mathcal{R}_\varrho(\cdot)$. Baselines include spatial watermarking [30] and optical multiplicative embedding (Opt. Mult.) [24].

A. Manipulation Attack Performance

We evaluate robustness to 3 attacks types class-fraction (20/40/60/80%), spatially coherent blocks (16/32/64), and band-limited spectral shifts (RE, NIR, SWIR; Δ=0.03) using optical key initializations D_0–D_4 and two baselines (spatial watermarking and optical multiplicative embedding [24]). Across Indian Pines and Lab Scene 1 (Fig. 4), Signed D_1 achieves the best ACC–SAM trade-off: class-fraction shows the expected monotonic trend, block attacks are most damaging due to spatial coherence, and band-limited shifts show mild band-dependent non-monotonicity. Additional failure-case analysis is provided in the Supplementary Material.

From Fig. 4, Signed D_1 is the top performer in all regimes. On Indian Pines, the curve peaks near 0.96–0.97 ACC with a minimum SAM around 2.1°, and across all settings it remains above $\approx$0.90 ACC with SAM below $\approx$4.3°. On Lab Scene 1, it reaches about 0.94–0.95 ACC at best with minimum SAM near 2.0°, and does not fall below $\approx$ 0.86–0.88 ACC nor exceed $\approx$4.7° SAM in the most demanding cases. For class-fraction, block, and band-limited shifts, the D_1 ACC curves stay above both baselines and their SAM curves stay below, confirming the robustness of the method.

B. Classification Performance

We assess the impact of optical signing on downstream classification by training a SpectralFormer [34] solely on the Baseline Not Manipulated data and testing on 4 configurations: Signed D_1, Restored D_1, Opt. Mult. [24], and Baseline Manipulated.We report mean±std over three runs with different random seeds for ACC and Cohen's Kappa in Table I. The results show that Signed D_1 preserves downstream performance within 0.6–1.2% of the baseline ACC on both datasets with similarly small Kappa deltas; after restoration, the Restored D_1 configuration recovers over 80% of the accuracy lost to manipulation while preserving label coherence; and Opt. Mult. remains below both Signed D_1 and Restored D_1 in ACC and Kappa. These trends confirm that the variance-aware initialization D_1 enables secure authentication without sacrificing classification fidelity, and we adopt D_1 as the default in subsequent experiments.

Dataset	Metrics	Baseline Not Manipulated	Signed D_1	Restored D_1	Opt. Mult. [24]	Baseline Manipulated
Indian Pines	ACC ↑	**0.941 ± 0.02**	0.935 ± 0.03	0.921 ± 0.02	0.890 ± 0.01	0.751 ± 0.05
	Kappa ↑	**0.917 ± 0.01**	0.907 ± 0.02	0.891 ± 0.01	0.880 ± 0.03	0.784 ± 0.03
Lab Scene 1	ACC ↑	**0.861 ± 0.03**	0.849 ± 0.02	0.809 ± 0.01	0.830 ± 0.04	0.735 ± 0.06
	Kappa ↑	**0.881 ± 0.01**	0.845 ± 0.02	0.805 ± 0.02	0.840 ± 0.03	0.754 ± 0.06

TABLE I
CLASSIFICATION PERFORMANCE OF THE SPECTRALFORMER TRAINED ON THE BASELINE AND EVALUATED ACROSS OPTICAL KEY CONFIGURATIONS.

C. Imperceptibility of optical key Distribution

Relative to baselines in Fig. 5, Signed D_1 delivers consistent gains, and improves PSNR by roughly 4–5 dB over Opt. Mult. and by about 9–12 dB over watermarking, while lowering spectral error by $\approx$0.8–1.0° vs. Opt. Mult. and $\approx$1.3–1.5° vs. watermarking. The Restored output remains close to Signed, typically within $\approx$0.5° SAM and $\approx$1.5–6 dB PSNR recovering most of the manipulation damage. Qualitatively, Signed preserves textures and color balance, whereas watermarking and Opt. Mult. exhibit blur/tint and larger spectral deviations. These trends are consistent with the variance-aware initialization: concentrating optical key energy along informative spectral ranges increases detectability for a fixed energy budget while keeping the embedding imperceptible.

D. Robustness to Noise

We evaluate robustness under additive corruption by injecting Gaussian, Poisson, and speckle noise at 40, 30, 20, and 10 dB SNR. Detection is measured as ACC and restoration as SAM. We compare Signed D_1 against a spatial watermarking baseline and Opt. Mult. under the same protocol. Across Indian Pines and Lab Scene 1 (Fig. 6), Signed maintains the best balance between ACC and SAM in all SNRs and noise types; at 40 dB it reaches approximately 0.96 ACC with 1.8° SAM in Indian Pines and 0.95 ACC with 2.2° SAM on Lab Scene 1, and even at 10 dB it preserves around 0.85/0.82 ACC with SAM near 3.9° in both datasets. The ordering between methods is stable as SNR decreases and the margin over watermarking and Opt. Mult. widens, indicating slower degradation under severe noise. This behavior is consistent with the additive, pre-digitization optical key propagates linearly through sensor noise, while the optical key-guided detector and the error-aware restoration remain effective at localizing and correcting perturbations. Beyond adversarial manipulations, non-adversarial spectral perturbations such as spectral distortions and compression artifacts are also analyzed. These effects generally preserve the global consistency between the embedded optical key and the spectral signal, yielding low responses in the error map. Additional qualitative and quantitative analysis is provided in the Supplementary.

E. Limitations and Security Considerations

The current system assumes clean training data from the same scene distribution used at the test time; performance

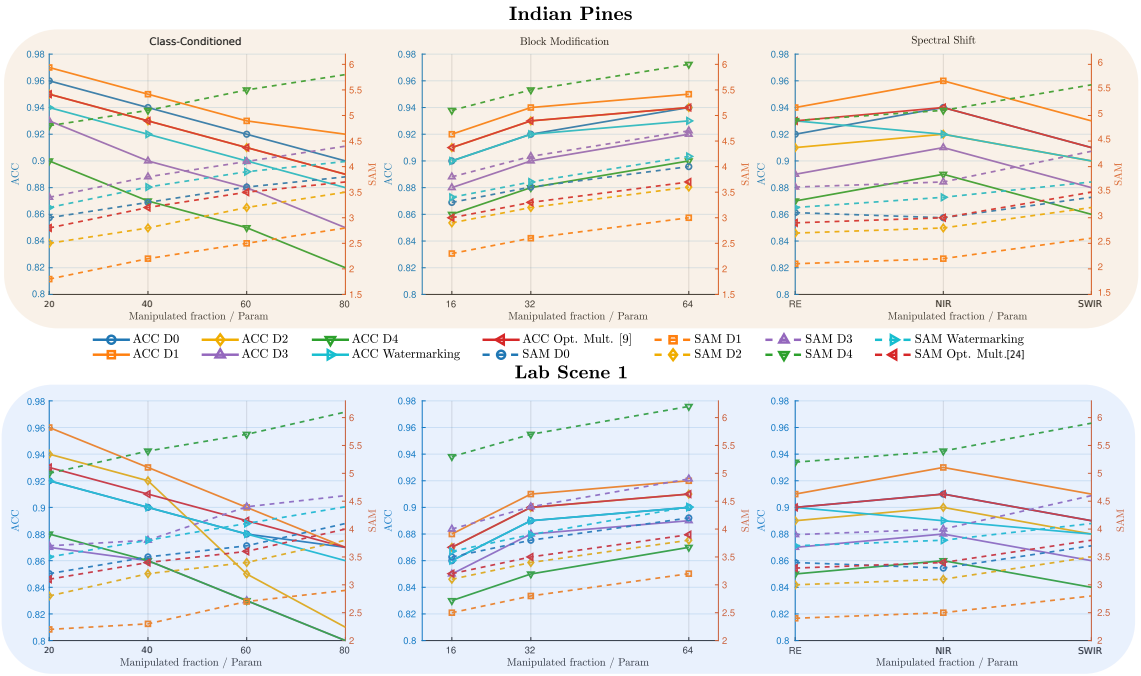

Fig. 4. Detection accuracy (ACC; solid, left axis) and spectral restoration error (SAM; dashed, right axis) under three manipulation regimes: (i) class-fraction (20–80% of a target class), (ii) block (sizes 16/32/64), and (iii) band-limited spectral shifts (RE, NIR, SWIR; $\Delta = 0.03$). Top: Indian Pines; bottom: Lab Scene 1. The Signed configuration with D_1 consistently outperforms watermarking and optical multiplicative embedding [24] across datasets and regimes; block attacks are most harmful, while band-limited shifts show band-dependent non-monotonic trends.

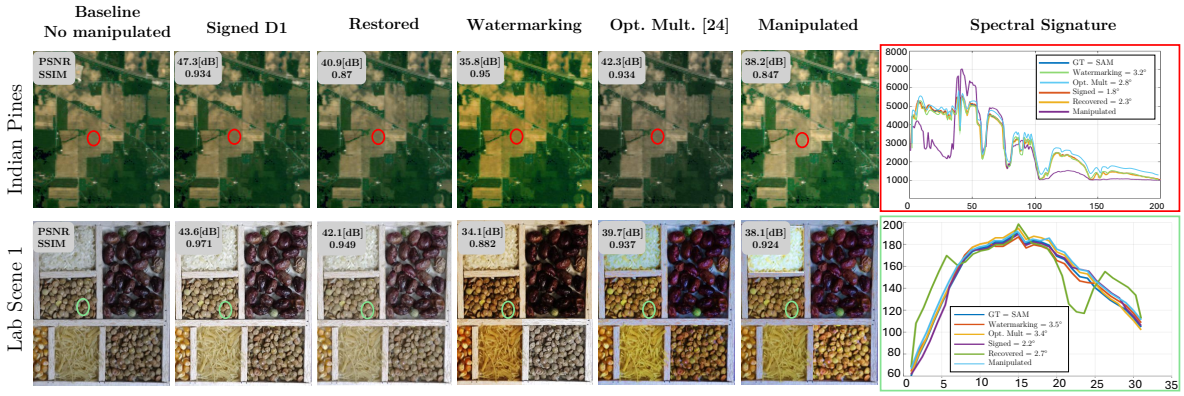

Fig. 5. Visual comparison of imperceptibility and restoration performance for both satellite data (Indian Pines, top row) and laboratory data (Lab Scene, bottom row). Each row displays, from left to right, the original image (Baseline), our signed approach, restored result, watermarking, optical multiplicative embedding [24], and manipulated version, followed by the spectral signature comparison.

can degrade with domain change. In such cases, lightweight adaptation per-scene calibration of detection thresholds and unsupervised fine-tuning driven by the imperceptibility and reconstruction terms can stabilize accuracy. Optical key-extraction attacks are out of scope, yet an informed adversary could approximate the key via white-reference targets; practical mitigation includes time-varying keys and secure scheduling, at the cost of added storage. Using a global key is memory-efficient but spatially predictable. From a sensing perspective, the method requires sufficient SNR to distinguish the key from acquisition noise. Under extreme conditions such as low illumination, both the scene signal and the key may degrade, limiting detection and restoration performance.

VI. Conclusions

This work introduces a variance-sensitive optical key for hyperspectral image authentication jointly optimized for embedding, detection, and restoration. By aligning the optical key with spectral variance, the proposed method achieves up to 4.7% lower spectral error and 4.7% higher detection accuracy than state-of-the-art approaches while consistently outperforming competing methods across all manipulation levels (20–80%). The embedding preserves spectral fidelity with imperceptible deviations below 1.2° and 2.2°, while classification accuracy remains within 0.6–1.2% of the baseline and restoration recovers over 80% of the manipulated accuracy. These results demonstrate robust authenticity verification with minimal spectral degradation. Future work will explore blind

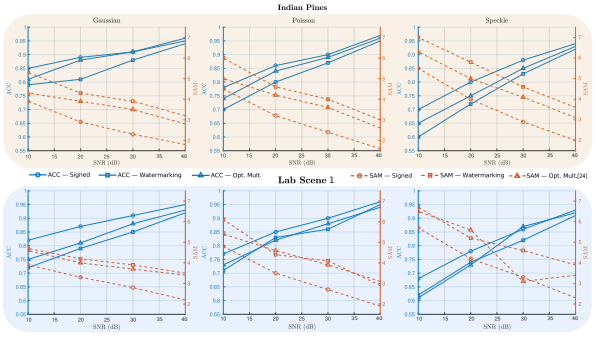

Fig. 6. Detection accuracy and spectral restoration error versus SNR for 3 synthetic noise types-Gaussian, Poisson, and Speckle on Indian Pines (top row) and Lab Scene (bottom row). Curves compare the signed configuration with Watermarking and Opt. Mult. [24]. Across noise types, signed maintains the highest ACC and the lowest SAM, with gaps widening as SNR decreases.

key optimization to further improve security performance.

ACKNOWLEDGMENTS

This work was funded by the Agencia Nacional de Hidrocarburos (ANH) and Ministerio de Ciencia, Tecnología e Innovacion (MINCIENCIAS), under project 110431 and CT 045-2025.

REFERENCES

[1] A. Bhargava, A. Sachdeva, and e. a. Kulbhushan Sharma, "Hyperspectral imaging and its applications: A review," *Heliyon*, vol. 10, no. 12, p. e33208, 2024.

[2] Q. Zhe, W. Gao, C. Zhang, and e. a. Du, "A hyperspectral classification method based on deep learning and dimension reduction for ground environmental monitoring," *IEEE Access*, 2025.

[3] M. Shimoni, R. Haelterman, and C. Perneel, "Hyperspectral imaging for military and security applications: Combining myriad processing and sensing techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 101–117, 2019.

[4] S. Faisal, M. P.-L. Ooi, S. K. Abeysekera, Y.-C. Kuang, and D. Fletcher, "Roadmap for measurement and applications: Uncertainty quantification and visualization for optimal decision-making in hyperspectral imaging-based precision agriculture," *IEEE Instrumentation & Measurement Magazine*, vol. 28, no. 1, pp. 23–32, 2025.

[5] B. Lu and P. D. e. a. Dao, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sensing*, vol. 12, no. 16, 2020.

[6] K. Dasari, S. A. Yadav, and e. a. Kansal, "Fusion of hyperspectral imaging and convolutional neural networks for early detection of crop diseases in precision agriculture," in *in Proc. 2024 Int. Conf. Commun., Comput. Sci. Eng. (IC3SDE)*. IEEE, 2024, pp. 1172–1177.

[7] J. Horvath, S. Baireddy, and e. a. Hao, "Manipulation detection in satellite images using vision transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, June 2021, pp. 1032–1041.

[8] B. Rathor and R. Saharan, "Steganography using bit plane embedding and cryptography," in *in Proc. Int. Conf. Smart Syst., Innov. Comput: SSIC 2017, Jaipur, India*. Springer, 2018, pp. 319–330.

[9] A. Dubey, N. Dixit, and e. a. Arora, "Challenges and opportunities in digital image watermarking," in *Proc. 2024 Int. Conf. Pioneering Develop. Comput. Sci. Digit. Technol. (C2SDT)*, 2024, pp. 29–34.

[10] P. Singh and R. S. Chadha, "A survey of digital watermarking techniques, applications and attacks," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 2, no. 9, pp. 165–175, 2013.

[11] K. M. Hosny, A. Magdi, O. ElKomy, and H. M. Hamza, "Digital image watermarking using deep learning: A survey," *Computer Science Review*, vol. 53, p. 100662, 2024.

[12] H. Chen, C. Tanougast, and e. a. Zhengjun Liu, "Optical hyperspectral image encryption based on improved chirikov mapping and gyrator transform," *Optics and Lasers in Engineering*, vol. 107, pp. 62–70, 2018.

[13] G. Qu, X. Meng, X. Yang, and e. a. Huazheng Wu, "Optical color watermarking based on single-pixel imaging and singular value decomposition in invariant wavelet domain," *Optics and Lasers in Engineering*, vol. 137, p. 106376, 2021.

[14] M. Bodke and S. Chaudhari, "Hyperspectral remote sensing image watermarking using discrete wavelet transform and forensic based investigation archimedes optimization," *Earth Science Informatics*, vol. 17, no. 5, pp. 4297–4313, Jul 2024.

[15] Z. Hou, H. Yan, L. Zhang, and e. a. Ren, "Zero-watermark method based on multichannel pcnn and blockchain for remote sensing image transaction certificate and copyright protection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–15, 2025.

[16] S. Aala, E. Panchakarla, and e. a. Ankam, Rohith Kumar, "Steganographic data encryption technique using hyperspectral imaging: A deceptive approach," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2024, pp. 1–6.

[17] V. Asnani and e. a. Yin, "Malp: Manipulation localization using a proactive scheme," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 12 343–12 352.

[18] Y. Zhao and e. a. Bo Liu, "Proactive image manipulation detection via deep semi-fragile watermark," *Neurocomputing*, vol. 585, p. 127593, 2024.

[19] V. Asnani and e. a. Yin, "Proactive image manipulation detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 15 386–15 395.

[20] M. Shan, J. Guo, and e. a. Zhi Zhong, "Improved multiple-image authentication based on optical interference by wavelength multiplexing," *Appl. Opt.*, vol. 61, no. 23, pp. 6931–6938, Aug 2022.

[21] H. Li, X. Bai, M. Shan, Z. Zhong, L. Liu, and B. Liu, "Optical encryption of hyperspectral images using improved binary tree structure and phase-truncated discrete multiple-parameter fractional fourier transform," *Journal of Optics*, vol. 22, no. 5, p. 055701, apr 2020.

[22] H. Wei and X. Wang, "Optical multiple-image authentication and encryption based on phase retrieval and interference with sparsity constraints," *Optics and Laser Technology*, vol. 142, p. 107257, 2021.

[23] Y. Xiong, J. Du, and C. Quan, "Optical encryption and authentication scheme based on phase-shifting interferometry in a joint transform correlator," *Optics and Laser Technology*, vol. 126, p. 106108, 2020.

[24] P. Gomez, R. Jacome, E. Martinez, H. Garcia, and H. Arguello, "Optical authenticity in pushbroom system for spectral information protection," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[25] J. Xue, Y.-Q. Zhao, T. Wu, and J. C.-W. Chan, "Tensor convolution-like low-rank dictionary for high-dimensional image representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 12, pp. 13 257–13 270, 2024.

[26] P. Liu, Y. Bu, and e. a. Zhao, "Enhancing visual data completion with pseudo side information regularization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 1, pp. 431–444, 2025.

[27] J. Xue and e. a. Zhao, "When laplacian scale mixture meets three-layer transform: A parametric tensor sparsity for tensor completion," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 13 887–13 901, 2022.

[28] Q. Li, J. Li, T. Li, and Y. Feng, "A joint framework for underwater hyperspectral image restoration and target detection with conditional diffusion model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 17 263–17 277, 2024.

[29] X. Li and e. a. Ding, "An end-to-end framework for joint denoising and classification of hyperspectral images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 7, pp. 3269–3283, 2023.

[30] D. P. Mukherjee, S. Maitra, and S. T. Acton, "Spatial domain digital watermarking of multimedia objects for buyer authentication," *IEEE Transactions on multimedia*, vol. 6, no. 1, pp. 1–15, 2004.

[31] X. Wu, J. Hu, and e. a. Gu, "A secure semi-fragile watermarking for image authentication based on integer wavelet transform with parameters," in *Conferences in Research and Practice in Information Technology Series*, vol. 108, 2005, pp. 75–80.

[32] E. Najafi, "A robust embedding and blind extraction of image watermarking based on discrete wavelet transform," *Mathematical Sciences*, vol. 11, pp. 307–318, 2017.

[33] M. Cedillo-Hernandez and e. a. Garcia-Ugalde, "Robust watermarking method in dft domain for effective management of medical imaging," *Signal, Image and Video Processing*, vol. 9, pp. 1163–1178, 2015.

[34] D. Hong, Z. Han, and e. a. Yao, "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Geosci. Remote Sens*, vol. PP, pp. 1–1, 11 2021.