*future internet*

*Article*

# Deterring Deepfake Attacks with an Electrical Network Frequency Fingerprints Approach

**Deeraj Nagothu [1], Ronghua Xu [1], Yu Chen [1,*], Erik Blasch [2] and Alexander Aved [2]**

[1] Department of Electrical and Computer Engineering, Binghamton University, Binghamton, NY 13902, USA; dnagoth1@binghamton.edu (D.N.); rxu22@binghamton.edu (R.X.)
[2] The U.S. Air Force Research Laboratory, Rome, NY 13441, USA; erik.blasch.1@us.af.mil (E.B.); alexander.aved@us.af.mil (A.A.)
* Correspondence: ychen@binghamton.edu

**Abstract:** With the fast development of Fifth-/Sixth-Generation (5G/6G) communications and the Internet of Video Things (IoVT), a broad range of mega-scale data applications emerge (e.g., all-weather all-time video). These network-based applications highly depend on reliable, secure, and real-time audio and/or video streams (AVSs), which consequently become a target for attackers. While modern Artificial Intelligence (AI) technology is integrated with many multimedia applications to help enhance its applications, the development of General Adversarial Networks (GANs) also leads to deepfake attacks that enable manipulation of audio or video streams to mimic any targeted person. Deepfake attacks are highly disturbing and can mislead the public, raising further challenges in policy, technology, social, and legal aspects. Instead of engaging in an endless AI arms race "fighting fire with fire", where new Deep Learning (DL) algorithms keep making fake AVS more realistic, this paper proposes a novel approach that tackles the challenging problem of detecting deepfaked AVS data leveraging Electrical Network Frequency (ENF) signals embedded in the AVS data as a fingerprint. Under low Signal-to-Noise Ratio (SNR) conditions, Short-Time Fourier Transform (STFT) and Multiple Signal Classification (MUSIC) spectrum estimation techniques are investigated to detect the Instantaneous Frequency (IF) of interest. For reliable authentication, we enhanced the ENF signal embedded through an artificial power source in a noisy environment using the spectral combination technique and a Robust Filtering Algorithm (RFA). The proposed signal estimation workflow was deployed on a continuous audio/video input for resilience against frame manipulation attacks. A Singular Spectrum Analysis (SSA) approach was selected to minimize the false positive rate of signal correlations. Extensive experimental analysis for a reliable ENF edge-based estimation in deepfaked multimedia recordings is provided to facilitate the need for distinguishing artificially altered media content.

**Keywords:** deepfake attacks; Audio and Video Systems (AVS); Internet of Video Things (IoVT); adversarial machine learning; environmental fingerprint; Electrical Network Frequency (ENF) signals; spectral estimation; Singular Spectrum Analysis (SSA)

**Citation:** Nagothu, D.; Xu, R.; Chen, Y.; Blasch, E.; Aved, A. Deterring Deepfake Attacks with an Electrical Network Frequency Fingerprints Approach. *Future Internet* **2022**, *14*, 125. https://doi.org/10.3390/fi14050125

Academic Editor: Seong Ki Yoo

Received: 30 March 2022
Accepted: 19 April 2022
Published: 21 April 2022

## 1. Introduction

Modern Artificial Intelligence (AI)/Machine Learning (ML) technology is widely integrated with many multimedia applications to help enhance its applications, and General Adversarial Networks (GANs) enable the manipulation of audio or video streams seamlessly based on the probability distribution of each dataset class [1]. Since first introduced in 2015, the development of the generator and the discriminator module of the GAN has led to the generation of deepfaked images that are indistinguishable from real images [2]. Such high-resolution and accurate generation of images had found many applications in modern media. The potential applications of deepfakes include e-health/medical field, commercial applications, and secure privacy in media. With the capability to generate feature characteristics based on a learned probability distribution, a deepfake generation model was proposed to help physically challenged people with entertainment media, where the model extracts motion features from a source subject and generates similar movements using the targeted subject [3]. In medical applications, deepfakes are readily applicable to develop better plastic surgery procedures for facial reconstruction [4]. Along with a guidance-based AI system in surgery, deepfakes are also used to generate training samples for rare medical conditions where the data are limited [4]. Commercial companies develop deepfake techniques to translate text-based messages delivered by artificial or deepfake characters, and similar applications are seen in social media platforms to create online avatars [5]. With the emergence of the metaverse, online deepfake avatars are created to represent virtual presence. Holographic technologies leverage deepfakes to generate 3D historical characters using accurate audio and video data and deliver their story for future generations. Lastly, deepfake applications in privacy preservation stand on a fragile line. One such application includes preserving victims' identities appearing on media platforms by altering their visual and audio characteristics [6].

However, deepfaked video, audio, or photos also can be highly disturbing and able to mislead the public, raising further challenges in policy, technology, social, and legal aspects [7,8]. Currently, there are deepfake tools available in the public domain that allow people to impersonate anyone, from businessmen to music stars, during video chats [9–11]. Deepfake video "attacks" on some public scenarios have raised serious concerns [12,13]. Political leaders' messages are altered to create fake news for the public and lower trust in broadcast messages [14]. Researchers have pointed out that disinformation may actually cause societal disturbance and ruin the foundation of trust [15–18]. For instance, the most recent case was on March 17: a deepfaked video was posted on social media showing that President Zelensky was calling the Ukraine soldiers to lay down their arms [19,20]. Domains such as smart surveillance, which highly depends on the audio and the visual layer input for its functionality, could lose the track of malicious actions when the incoming frames are altered [3]. Government agencies such as the U.S. Defense Advanced Research Projects Agency (DARPA) are concerned about losing the war against deepfake attacks from adversarial hackers that use popular ML techniques to automatically incorporate artificial components into existing video streams [21,22]. Therefore, as a primary cause of misinformation, an imminent need for fast and reliable authentication techniques is of a high priority [14,23].

While the community has been engaging in the endless AI arms race "fighting fire with fire" hoping to have "smarter" ML algorithms [24–26], new ML algorithms keep making fake AVS data more real. Therefore, it is compelling to explore alternative ML deepfake detection solutions. In this paper, we propose to tackle the challenging deepfake attack detection problem leveraging the Electrical Network Frequency (ENF) signal, which is embedded in the recorded AVS data as a fingerprint that is determined by the environmental factors of the recording region. The effectiveness of a fingerprint technique against the deepfake generation model depends on its uniqueness and randomness to avoid forgery and predictions.

The ENF is the instantaneous frequency in the electrical power grid with a nominal value of 50/60 Hz, depending on the geographical location [27,28]. For the rest of this paper, we consider the nominal frequency value as 60 Hz for our testbed in the United States. The Instantaneous Frequency (IF) varies over time due to the varying load balance mechanism and power supply demands, resulting in the fluctuations from the nominal frequency resulting in the ENF signal [29]. The variation in fluctuations is small, and the fluctuations are similar throughout the power grid interconnect. Among the four major power grid interconnects in the USA, the experimental data were collected in the Eastern power grid where the variation of the ENF is in the range of $[-0.02, 0.02]$ Hz from the nominal frequency [30]. While the ENF signal functions as the main power supply, it also gets embedded in the digital multimedia through background hum [31,32] or illumination frequency in audio and video recordings [27,33,34]. Due to the presence of the ENF in audio–video channels, the manipulation in the ENF signal with respect to time is treated as the manipulation or modification of the multimedia recordings [35–37]. The ENF signal is also used for forensic analysis of digital evidence, time of recording estimation [38], media synchronization among multiple channels [39], and geographical tagging of the recording [40].

Although the ENF signal is present in the audio and video channels, some challenges exist when using the ENF as a fingerprint mechanism. Due to the lower frequency range, the Signal-to-Noise Ratio (SNR) for reliable ENF estimation is vital to address. In typical deepfake videos, instantaneous frequency estimation is required for estimation, which depends on the spectral estimation techniques used. In order to adopt the ENF as a fingerprint technique, a solution is also needed to address the redundant ENF reference database for comparing the estimated ENF. This paper analyzes the ENF estimation techniques against deepfake audio and video recordings using different spectral estimation techniques and robust and reliable estimation in low SNR recordings. Our contributions in this paper are as follows:

- Designing of an effective spectral estimation technique using both parametric and non-parametric methods for IF detection.
- Utilizing a Robust Filtering Algorithm (RFA) over a weighted SNR to identify the harmonic ENF embedded in media recordings to enhance the ENF signal estimation in the identified ENF.
- Implementing an effective detection technique against deepfake attacks and an integrated Singular Spectrum Analysis (SSA) based on the correlation coefficient values to reduce the number of false positives in a real-time video broadcasting scenario.
- Demonstrating experimental analysis on the video and audio deepfake attacks' detection using the RFA technique and comparing its effectiveness against traditional spectral estimation techniques.

The rest of the paper is organized as follows. Section 2 discusses the background and related work in deepfake detection technologies and ENF fingerprint applications. Section 3 thoroughly discusses the spectral estimation techniques used in this work for the comparison analysis and the ENF enhancement techniques in low SNR recordings followed by the Singular Spectral Analysis (SSA) approach to minimize the false positive rate caused by correlation outliers in Section 4. Section 5 reports the experimental evaluation of the spectral techniques discussed in Section 3 and the performance evaluation of SSA in edge-based devices. Finally, Section 6 discusses the limitations along with alternate strategies, and we conclude this paper in Section 7.

## 2. Background and Related Work

### *2.1. Deepfake Detection Using Traditional and Trained Models*

Deepfake detection has become a critical problem in digital media authentication. With advanced computational power and the developments in GANs, the resulting media output is very realistic [2]. However, along with its development, many detection techniques were proposed in the early stages to leverage the artifacts introduced in deepfakes. Artifacts such as eye blinking [41], facial distortion, facial symmetry construction [42], and motion artifacts can be visually inspected and identified [43]. Machine-learning-based models were also trained to identify the artifacts. However, the artifacts result from low training data and improvement in the GAN architecture; with more data, the artifacts can be reduced, and more realistic images can be created, leaving the visual-artifact-based detectors redundant.

Hidden features such as GAN fingerprints are unique to the deepfake model architecture [44], and biometric signatures such as heartbeat detection through the skin do not depend on visual artifacts [45]. The signatures can be reliable when the visual artifacts are removed by better training. The GAN also introduces frequency-level artifacts due to the upsampling method in the GAN pipeline [46], and the modified frames can be identified by frequency analysis and studying the compression map [47,48]. Noiseprint is one such fingerprint extracted by suppressing the high-level scene content and leveraging the in-camera processes for unique fingerprints [49]. Noiseprint is applied to reliably localize the frame modification with high performance. Other camera-based fingerprint techniques such as Photo Response Non-Uniformity (PRNU) sensor noise and JPEG compression artifacts were also used in detecting frame-level forgeries due to their dependence on the source device [50,51]. However, these unique artifact-based detectors can also be spoofed using a GAN-based approach where camera traces are inserted into the synthetic images [52]. Along with the reliability of the unique fingerprint for its detection capability, it is also essential that the fingerprint be less prone to forgeries. Hence, we adopted the ENF-based environmental fingerprint where the fluctuations are a random process and signal manipulation in media recordings leaves modification traces.

### *2.2. ENF Applications in Digital Multimedia*

The ENF was initially introduced as a forensic verification technique for law enforcement applications to verify the authenticity of audio recordings [27]. Due to electromagnetic induction, the audio recorders directly connected to the power grid can also embed the ENF fluctuations in the audio recordings [28]. The applications were limited to devices connected directly to the power grid until the presence of the ENF was verified in battery-powered devices through the background hum generated by surrounding electrical appliances connected to the grid and increasing its range of devices [31].

Along with audio, video recordings were also discovered to carry ENF fluctuations in the form of illumination frequency [33,34]. The captured photons from artificial light have similar fluctuations, and the method estimation from the video recordings depended on the imaging sensor used in the capture device. Complementary Metal–Oxide Semiconductors (CMOSs) and Charge-Coupled Devices (CCDs) are the most commonly used imaging sensors with different shutter mechanisms [38]. In the case of CCD sensors, a global shutter mechanism is used where the whole sensor grid is exposed to photon capture at one instant, resulting in capturing the ENF samples equal to the number of frames per second. However, in CMOS, a rolling shutter mechanism captures the ENF sample per row in the sensor grid and vastly increases the captured samples [34]. Due to limited samples in the CCD sensor, an alternative aliasing frequency technique can be used to estimate the ENF fluctuations [33]; however, it is prone to signal noise. Most commercial-grade camera devices use CMOS sensors due to their cost-effective nature, resulting in an effective solution for ENF estimation through video recordings.

The presence of the ENF signal in audio and video recordings has increased its viable applications in identifying the recording time due to its unique fluctuation nature. Although the fluctuations in the ENF are similar throughout the power grid interconnect, the propagation delay can be used to identify the geographical location of the recording within the grid, essentially enabling the ENF technology with the geotagging feature [53]. ENF presence in audio and video recordings can be used to synchronize the media recordings from multiple recorders in commercial applications [39]. Smart grid infrastructure relies on ENF fluctuations to analyze power consumption, create a feedback loop for power outages, and prevent grid-level blackouts [30].

### *2.3. ENF-Based Digital Media Authentication*

The ENF signal can essentially be used for both audio and video forgeries with its forensic capabilities. Modifications such as copy and move, frame replay, spatial modifications, and inserting external recordings can be identified using ENF inconsistencies [36,37]. Many ENF estimation techniques are already proposed using multiple spectrum estimation techniques and phase identifications. In this work, we focus on studying the effects of deepfake generation on the embedded ENF signal, deploy multiple spectral estimation techniques and verify their effectiveness, and analyze the robust and ENF-preserving techniques increasing the likelihood of efficient ENF-based authentication.

## 3. Robust ENF Estimation Techniques

ENF signal estimation primarily depends on a reliable Instantaneous Frequency (IF) estimation from the source recording. Due to the harmonics embedded with the nominal frequency, some harmonic frequencies have a higher SNR and can result in a reliable signal [54,55]. However, the noise interference in some harmonics can also completely interfere with the signal. With deepfake videos, the primary interruptions in extracting a reliable signal from the video are from the moving subjects [56]. The signal estimation is more efficient for a static background, but for a moving subject, there is additional noise embedded due to the pixel intensities' variation [35]. Other challenges include the duration of the audio and video recordings used. The duration is not a problem in a continuous stream of multimedia since the window can be larger. However, in the case of a limited recording length, the spectrum estimation for reliable frequency extraction becomes challenging. We aim to test the effects of parametric and non-parametric spectrum estimation techniques against deepfake videos for this scenario [57]. For non-parametric spectrum estimation, Short-Time Fourier Transform (STFT) was used to estimate the ENF, and in the case of parametric methods [27], we used the Multiple Signal Classification (MUSIC) algorithm [57]. Each method has its own merits in the case of computational power, reliable estimation, and the amount of data sequence required.

### *3.1. Non-Parametric Spectral Estimation Techniques*

Non-parametric approaches do not assume that a model generated the data. The typical approach in this method is to use Fourier analysis, which can result in some higher computational cost. We utilized the Short-Time Fourier-Transform (STFT)-based approach in this work. The ENF signal fluctuations are represented as $f_{ENF} = f_o + f_\Delta$, where $f_o$ is the nominal frequency and $f_\delta$ is the instantaneous signal fluctuation. With the Fourier transform of the input signal $x(n)$, the Power Spectral Density (PSD) is calculated from the spectrogram to estimate the spectral band from the harmonic frequency band ($B$) of interest $f \in k[f_o \pm \frac{B}{2}]$.

From the spectral band, the instantaneous frequencies in each frame window are estimated by the maximum value in the corresponding power density vector for that time instant. To improve the frequency estimation accuracy, quadratic interpolation can be used where the index of the frequency bin numbers is used to obtain the spectral peak. The peak location is given as

$$\Delta = \frac{1}{2} \cdot \frac{\alpha - \gamma}{\alpha - 2 \cdot \beta + \gamma}$$

where $\alpha$ is the previous bin of the max spectral bin, $\beta$ is the max spectral peak, and $\gamma$ is the next bin. If $k^*$ is the bin number of the highest spectral sample, then the resulting frequency estimate bin is adjusted by $\Delta$, and the final interpolated frequency estimate is

$$f_{ENF} = f_o + (k^* + \Delta)\frac{f_s}{N}$$

Here, $f_s$ is the sampling frequency of the input signal and $N$ is the number of FFT bins used. Although the input signal data length is not limited in a continuous input stream, such as a surveillance system audio/video feed, the number $j$ of fast Fourier transforms $FFT_j$ where $j = 1, \ldots, J$ points can be increased for higher accuracy at the cost of increased computational resources. With known nominal frequency bounds, the ENF estimate from this technique can be accurate, but at the same time, if the energy peak is not in the bounds, then it is susceptible to outliers.

### *3.2. Parametric Spectral Estimation Techniques*

The spectrum estimates using parametric methods result in a higher-quality spectrum. It requires less data compared to that of non-parametric methods. However, it is essential that the model parameters be estimated appropriately; otherwise, the estimated model could give wrong or misleading estimates. Among the parametric methods, in this paper, we adopted the Multiple Signal Classification (MUSIC) technique based on the subspace analysis of the signal and noise model [57].

MUSIC is a subspace-based frequency estimation model depending on the eigenanalysis of the observed discrete time signal data. For this algorithm, let the discrete time signal $v(n)$ of finite length $L$ with $K$ sinusoidal components be represented as

$$v(n) = \sum_{k=1}^{K} A_k e^{jn\omega_k} + w(n)$$

where $A_k = |A_k|e^{\phi_k}$ is the complex magnitude of the $K$th harmonic signal component with $\phi_k$ being unknown and assumed to be unknown and uniformly distributed in $[-\pi, \pi]$ and $w(n)$ is the noise.

For a data sequence of length $L = N + M - 1$, the auto-correlation matric $R_v$ of size $M \times M$ is computed. $M$ is the dimension spanned by $v(n)$, and $K$ is the signal subspace, while the $N$-point observed signal,

$$\hat{R_v} = \frac{1}{N}\mathbf{V}^H\mathbf{V}$$

where $(.)^H$ is a Hermitian operator:

$$\mathbf{V} = \begin{pmatrix} \mathbf{v}^T(0) \\ \mathbf{v}^T(1) \\ \dots \\ \mathbf{v}^T(N-1) \end{pmatrix}^T = \begin{pmatrix} v(0) & v(1) & v(2) & \dots & v(M-1) \\ v(1) & v(2) & v(3) & \dots & v(M) \\ \dots & \dots & \dots & \dots & \dots \\ v(N-1) & v(N) & v(N+1) & \dots & v(N+M-2) \end{pmatrix}_{N\times M}$$

With eigen-analysis on $\hat{R_v}$, the eigen vectors corresponding to the $K$ signal subspace $(U_s)$ are $q_1, q_2, \dots, q_K$ and the remaining $q_{K+1}, q_{K+2}, \dots, q_M$ span the noise subspace $(U_n)$. Assuming a signal eigenvector $e$, then it must be orthogonal to the noise subspace eigenvectors $e \perp q_i$ for $\{q_i\}_{i=K+1}^{M}$, where

$$e(\omega_l) = [1, e^{j\omega_l}, e^{j2\omega_l}, \dots, e^{j(M-1)\omega_l}]^T, l = 1, 2, \dots, K$$

The MUSIC algorithm defines a squared norm function:

$$d^2 = \|U_n^H e\|^2 = \sum_{i=K+1}^{M} |e^H q_i|^2$$

If the $e$ vector belongs to the signal subspace, then $d^2 = 0$ due to the orthogonality condition. The reciprocal of the squared norm will result in sharp peaks at desired signal frequencies.

$$P_{MUSIC}(e^{jw}) = \frac{1}{\sum_{i=K+1}^{M} |e^H q_i|^2} \tag{1}$$

### *3.3. Robust ENF Enhancement Techniques*

Multimedia recordings are often susceptible to noise interference. For reliable estimation of the ENF signal from the source recording, robust measures are needed. The following discusses the reliable techniques used for robust ENF estimation in a noisy environment.

#### 3.3.1. Weighted Harmonics Combination

The ENF signal is embedded in multiple harmonics depending on the nominal frequency. For audio recordings, the ENF is present in either even or odd harmonics depending on the type of microphone used. Similarly, in video recordings, the harmonics are the multiple of illumination frequency, which is twice the nominal frequency. Other harmonics can be leveraged to obtain accurate fluctuations for noise interference in targeted frequencies. Therefore, a weighted combination of the harmonics' spectral bins can result in a noise-tolerant spectrum for reliable frequency estimation [54]. The SNR values are computed as a ratio of the PSD ($s(f)$) in the ENF fluctuation range ($f_c$) to the PSD in the spectral band of interest ($f_v$). The optimal ENF fluctuation in the U.S. is ±0.02.

$$w_k = \frac{\sum_{k=1}^{L} s(f_o - f_c, f_o + f_c)}{\sum_{k=1}^{L} s(f_o + f_c, f_o + f_v) + s(f_o - f_c, f_o - f_v)}$$

where $L$ is the maximum number of harmonics carrying ENF fluctuations and can be combined. Using the weights computed for windowed spectral bins, the final spectrum $S(f)$ is evaluated, and quadratic interpolation can be used to estimate the spectral peaks and frequency fluctuations.

$$S(f) = \sum_{k=1}^{L} w_k s(f)$$

In our approach, we used weighted estimation from multiple harmonic bins to identify the frequency with the highest SNR compared to other harmonics. For audio recordings, due to the nature of the microphone used, the ENF is either embedded in the even harmonics or odd harmonics. Similarly, for video recordings, the ENF is embedded on the Frames Per Second (FPS), the illumination frequency, and the type of artificial light used. With the weight matrix, the ENF with the highest SNR is identified, and then, the following filtering algorithm is used to enhance the ENF in that frequency range.

#### 3.3.2. Robust Filtering Algorithm

The RFA was proposed to improve ENF estimation in noisy interference [58]. Instead of reducing the noise after the spectrum is computed, the RFA approach enhances the estimation accuracy by improving the SNR and minimizing in-band noise prior to the ENF estimation.

In the RFA [58], a time-domain preprocessed input signal is encoded into the Instantaneous Frequency (IF) of the Sinusoid-Frequency-Modulated (SFM) signal. A kernel function is utilized to generate the Sinusoidal Time-Frequency Distribution (STFD) of the encoded signal, where the peaks correspond to the denoised ENF. For an optimal selection of the kernel function and the signal encoding, we recommend readers refer [58] for a detailed description of the algorithm. With the help of the RFA, the ENF can be reliably estimated under a −20 dB noise level. For deepfake videos, the underlying ENF, although captured by the imaging sensor from the artificial light, is interfered by the pixel noise and subject movement [56]. Therefore, the RFA technique can be used to minimize the noise, and then, a suitable spectrum estimation technique can be used to estimate the ENF.

### *3.4. ENF Similarity Verification Using the Correlation Coefficient*

Authentication of the ENF carrying multimedia can only be verified when the estimated ENF fluctuations are not tampered with or modified. For this verification, we adopted the correlation coefficient as a measure of similarity to verify the estimated ENF signal from the recording ($P_{ENF}$) with the ground truth ENF ($G_{ENF}$) collected directly from the power grid. The value of the correlation coefficient ($\rho$) varies from $[-1, 1]$, where 1 represents the highest similarity and vice versa.

$$\rho(l) = \frac{\sum_{n=1}^{N} [f_{P_{ENF}}(n) - \mu_{P_{ENF}}][f_{G_{ENF}}(n-l) - \mu_{G_{ENF}}]}{var(P_{ENF}) * var(G_{ENF})} \tag{2}$$

where $l$ represents the lag measure, $N$ is the signal length, and $\mu$ is the mean. Although the reference ENF collected from the power grid is redundant for efficiently deploying this authentication scheme, we later discuss a distributed authentication system that relies on a consensus mechanism designed using the ENF fluctuations. Based on the consensus, the networked multimedia devices broadcast their estimated ENF, and a pseudo ground truth ENF is selected, which is used for authentication.

## 4. ENF-Based Anomaly Detection Using Singular Spectrum Analysis

### *4.1. SSA for Anomaly Detection*

With a reliable ENF signal estimated using the appropriate spectral estimation techniques, we integrated an anomaly detection scheme to analyze the correlation vector and detect major deviations from the historical trend. Figure 1 represents the algorithm flow including the signal estimation process for anomaly detection. The estimated ENF signal and ground truth reference signal were compared using the sliding window algorithm for continuous monitoring of the input media stream. Similarly, the generated correlation coefficients were analyzed for outlier detection and media forgery. The SSA algorithm decomposes the time series vector and performs Singular-Value Decomposition (SVD) for change-point detection [59] or future trend prediction [60]. The following section discusses the algorithm based on the correlation coefficient values for change-point detection analysis.

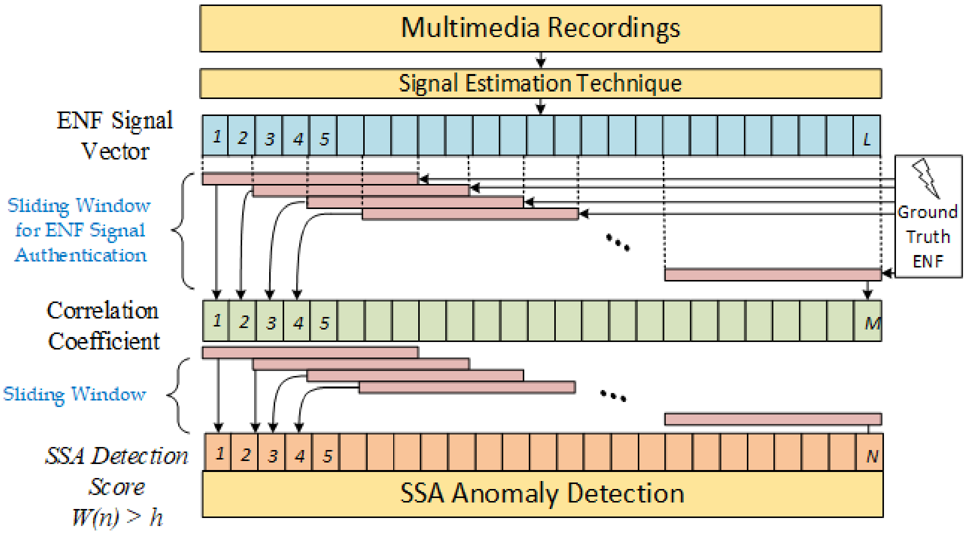

**Figure 1.** Singular Spectrum Analysis (SSA) algorithm using the ENF signal and its correlation coefficient.

### *4.2. SSA Algorithm*

The correlation coefficient as a similarity measure between the ENF estimated from the multimedia recordings and the ground truth ENF can generate false positives due to the signal noise. We integrated a Singular Spectrum Analysis (SSA) technique to regulate the fluctuations in the correlation coefficient and change-point detection [59,60]. The correlation coefficient samples are non-periodic in nature, and in order to integrate the SSA algorithm, a larger window size is required. Let $\rho_{n+1}, \rho_{n+2}, \rho_{n+3}, \cdots, \rho_{n+N}, \rho_{n+N+1}, \cdots, \rho_{n+N+Q}$ be the non-periodic correlation coefficient samples collected from the online ENF comparison and $N$, $M$, $l$, $p$, and $q$ be fixed integers, where $n$ is iterative over new correlation coefficient values, $N$ is the window size for the base matrix, $Q$ is the window size for our test matrix with $Q = q - p$, and $l < M \leq \frac{N}{2}$. For each $n = 0, 1, \ldots$, the following algorithm is executed:

1. Creating the base matrix of size $(M \times K)$ using the initial correlation coefficient values and $K = N - M + 1$,

$$X_B^n = \begin{pmatrix} \rho_{n+1} & \rho_{n+2} & \rho_{n+3} & \cdots & \rho_{n+K} \\ \rho_{n+2} & \rho_{n+3} & \rho_{n+4} & \cdots & \rho_{n+K+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{n+M} & \rho_{n+M+1} & \rho_{n+M+2} & \cdots & \rho_{n+N} \end{pmatrix}_{M\times K}$$

2. Using the base matrix, also known as the Hankel matrix, we compute $R = (X_B^n).(X_B^n)^T$, and the Singular-Value Decomposition (SVD) of the matrix $R$ results in $M$ eigen vectors and eigen values. Among the $M$ eigen vectors, $l < M$ eigen vectors are selected to create a group $I$. The group $I$ consists of $l$-dimensional vectors in subspace $\mathcal{L}_{n,I}$ of $M$-dimensional space $\mathcal{R}^M$. The eigen values computed from the matrix $R$ are arranged in descending order, and the top $l$ values are selected for the matrix $I$, respectively, such that the subspace $\mathcal{L}_{n,I}$ consists of the features of $\mathcal{R}^M$.
3. With the base matrix established, next, a test matrix is constructed of size $(M \times Q)$ with a lag $p$ from the base matrix and $Q = q - p$. The resulting matrix is

$$X_T^n = \begin{pmatrix} \rho_{n+p+1} & \rho_{n+p+2} & \rho_{n+p+3} & \cdots & \rho_{n+q} \\ \rho_{n+p+2} & \rho_{n+p+3} & \rho_{n+p+4} & \cdots & \rho_{n+q+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{n+p+M} & \rho_{n+p+M+1} & \rho_{n+p+M+2} & \cdots & \rho_{n+q+M-1} \end{pmatrix}_{M \times Q}$$

4. With the test matrix $X_T^n$ and the $l$-dimensional subspace $\mathcal{L}_{n,I}$, the detection statistics of abnormal fluctuations in the input values can be calculated with the sum of the squared Euclidean distance between the column vectors of $X_T^n$ and subspace $\mathcal{L}_{n,I}$. The column vectors of $\mathcal{L}_{n,I}$ are represented as $U_{i_1}, U_{i_2}, \ldots, U_{i_l}$. The detection statistics $\mathcal{D}_{n,I,p,q}$ for $n$ iterating over $\{0, 1, \ldots\}$ is given as,

$$\mathcal{D}_{n,I,p,q} = \sum_{j=p+1}^{q} \left( (X_j^{(n)})^T \cdot X_j^{(n)} - (X_j^{(n)})^T \cdot U \cdot U^T . X_j^{(n)} \right)$$

5. With the iterating values, the detection scores are normalized and represented as

$$S_n = \frac{\mathcal{D}_{n,I,p,q}}{\mu_{n,I}}$$

6. The Cumulative Sum of deviations (CUSUM) in the detection statistics are then calculated to eliminate false positives and seek major changes in the input values. A threshold $h$ is used to detect the fluctuations in the correlation coefficient of the ENF values. The detection score is

$$W_1 = S_1$$

$$W_{n+1} = (W_n + S_{n+1} - S_n - \frac{1}{M \cdot Q})^+, n \geq 1$$

$$h = \frac{2t_\alpha}{M \cdot Q} \sqrt{\frac{1}{3} \cdot Q \cdot (3MQ - Q^2 + 1)} \tag{3}$$

where $(a)^+$ represents $max(0, a)$.

## 5. Experimental Study and Performance Analysis

### *5.1. Prototype Implementation*

In our experiments, the DeepFaceLab software was adopted to create video deepfakes [61], and Descript was used to create audio deepfakes [62]. The DeepFaceLab software is capable of generating deepfakes in real-time using face swapping and mapping to the original face by modifying the surrounding pixels. For audio deepfakes, a training time as little as ten minutes of target audio can be used to recreate a deepfake voice to mimic the targeted actor. Software such as this made easily available with almost no usage complexity can only result in more generation of fake media. In this paper, we study the effects of multiple spectrum analysis against deepfake modification and use a signal enhancement technique to estimate the reliable signal to localize the forgery.

Figure 2 presents the overall architecture of the prototype implementation consisting of multiple edge clients and an edge server. The computational complexity and resource allotments are shared across multiple devices for enhanced ENF estimation. Each client collects real-time streams from cameras and then extracts ENF fingerprints, which are used for spectral estimation at the edge devices and SSA detection at the powerful edge server. The deployment of our proposed approach was analyzed on both edge-based devices (Raspberry Pi) and an edge server (desktop), and a detailed performance analysis of computational resource consumption is also presented.

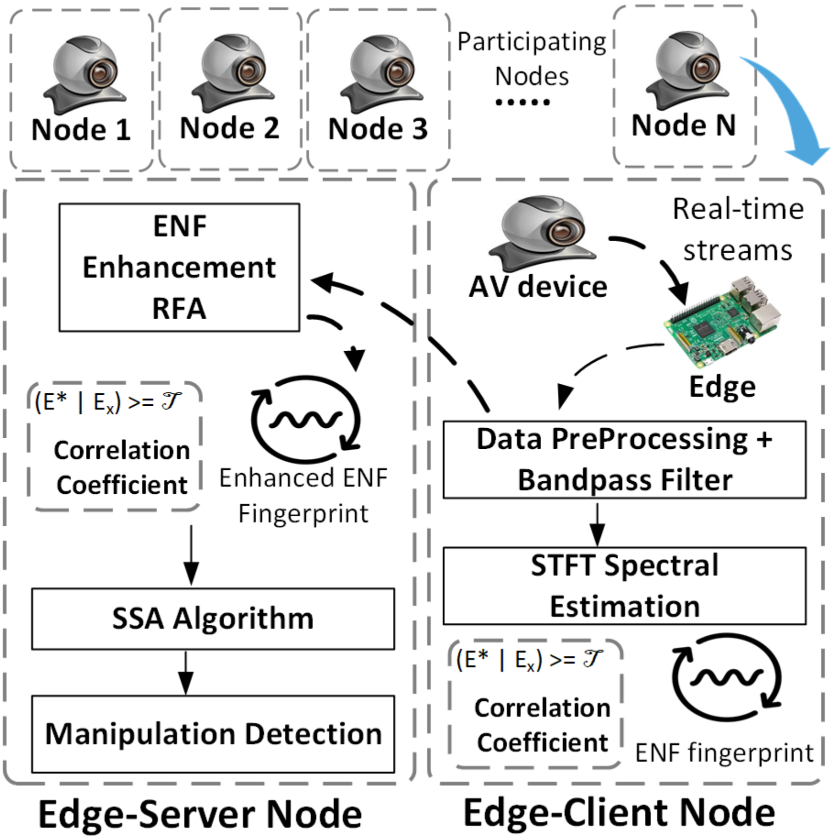

**Figure 2.** System architecture of deterring deepfake attacks at the edge network.

### *5.2. Effects of Spectral Estimation Techniques against Deepfakes*

Spectral estimation techniques have different parameters to control for a reliable estimation. In this work, we used MUSIC- and STFT-based spectral methods. The spectrum is computed from both techniques along with a bandpass filter along the nominal frequency of interest. In the STFT method, the spectral harmonic bands are collected from the spectrogram, whereas the MUSIC method looks for $K$ complex exponential components in the signal. For the ENF signal, the value of $K$ is two. Once the spectrum is computed for both techniques, the maximum frequency bin is identified with the help of quadratic interpolation, and the required frequency fluctuations are estimated. Although each spectrum estimation method has its own advantages, for deepfake videos, it is important that the method be more consistent and fast. MUSIC performs better with a lower signal length, whereas STFT relies on the Fourier transform, which needs more data for its computation. In Figure 3, the performance of each method is measured with respect to the input signal. The ENF was estimated from a static background recording under artificial light along with the ground truth reference signal with a sampling rate of 1000 Hz. The input signal was incremented by six seconds for each round, and the correlation coefficient was measured for similarity with the reference signal. Figure 3 clearly shows that MUSIC performed better with lower-duration recordings, but STFT outperformed the MUSIC method with sufficient input data given.

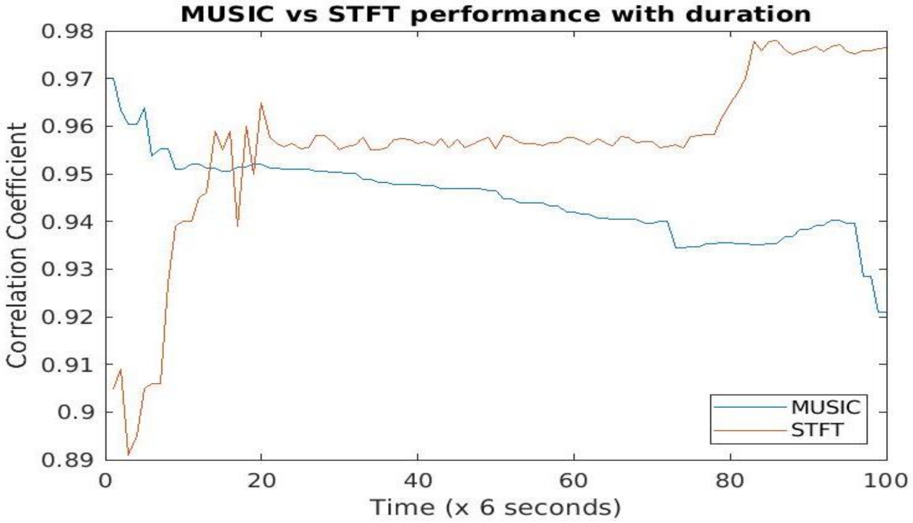

**Figure 3.** Performance of STFT and MUSIC spectrum estimation methods on ENF estimation based on the length of input signal.

With the performance analysis based on the duration of the input signal, next, the STFT and MUSIC algorithms were tested with a fixed-length input signal on its ENF estimation on video recordings. The video recording used includes a talking head subject with movements recorded under artificial light. Figure 4 represents the correlation coefficient for STFT and the MUSIC algorithm. Here, it is clear that STFT performed better in video-based ENF estimation compared to the MUSIC algorithm. For a reliable deepfake manipulation detection, it is vital that the ENF from unmodified recordings be estimated more reliably, so that any significant changes in the ENF can be marked as a potential manipulation. For some audio and video analysis, the harmonic frequencies in the recordings are targeted with external noise interfering with the embedded ENF. For this purpose, we used reliable estimation techniques such as the weighted combination of multiple harmonics [54] along with the robust filtering algorithm proposed in [58]. Table 1 represents the average SNR of the ENF fingerprints in the media recordings. Using the SNR matrix, the targeted frequency range was identified and the RFA algorithm was used to increase the SNR significantly. From Table 1, for power and audio recordings, the ENF signal is stronger in odd harmonics, and in video recordings, it is stronger in even harmonics since the nominal frequency in videos is 120 Hz.

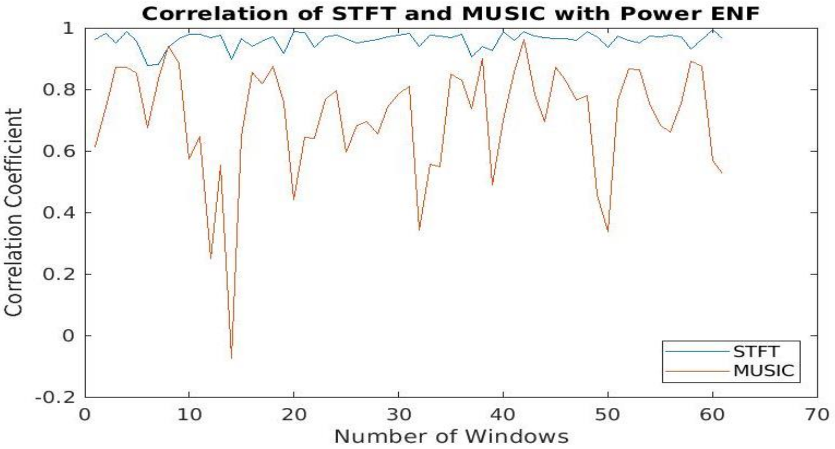

**Figure 4.** Spectrum estimation techniques for ENF estimation in a video recording with a moving subject under artificial light.

**Table 1.** SNR of the ENF fingerprint in media recordings.

| Media | 60 Hz | 120 Hz | 180 Hz | 240 Hz | 300 Hz | 360 Hz |
|---|---|---|---|---|---|---|
| **Power** | 39.88 | 2.456 | 38.647 | 0 | 14.48 | 4.534 |
| **Audio** | 9.761 | 0.888 | 27.94 | 7.106 | 43.717 | 10.585 |
| **Video** | 0 | 8.396 | 0 | 90.163 | 0 | 1.439 |

### *5.3. ENF Enhancement Using the RFA*

The efficiency of the RFA was tested on real-world audio recordings with ENF embedded though background hum. We used the STFT algorithm to estimate the harmonic frequency with the highest SNR and estimate the ENF signal. However, due to external noise, the ENF estimated from a single harmonic frequency had a significantly lower correlation coefficient when compared with its reference ENF. Using the RFA to enhance the ENF in the frequency of interest, the noise was suppressed, and then, the ENF was estimated from the new spectral bins. Figure 5 represents the correlation coefficient histogram of the audio ENF before and after applying the RFA to the recordings. Although it was unclear if the recordings carried any ENF signature before the RFA, it can be clearly seen that the RFA enhanced the embedded ENF and was more reliable for better ENF estimation from recordings with a lower SNR.

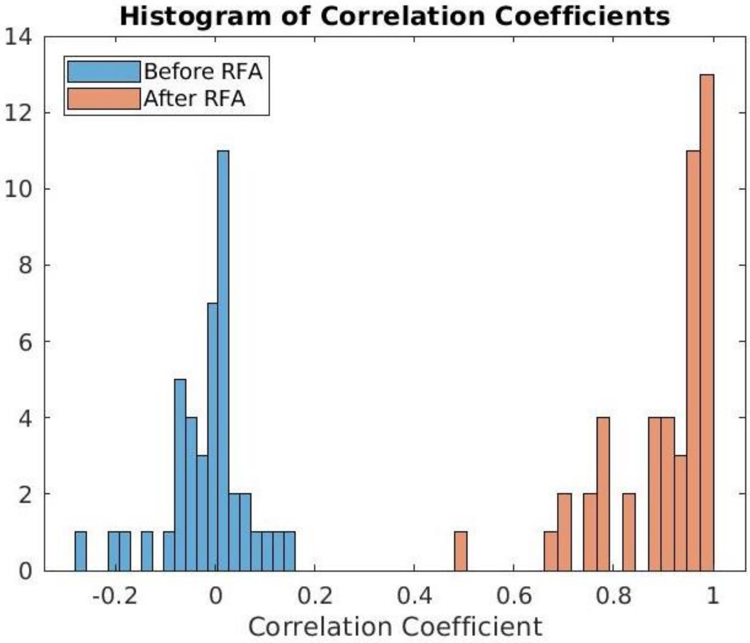

**Figure 5.** Correlation coefficient histogram collected from real-world audio recordings before and after applying the robust filtering algorithm.

ENF estimation from video recordings depends on the presence of artificial light in the recording. With lower-intensity light in the background, the ENF is not reliable due to the interference of other in-camera noise such as ISO sensor noise and other subject-movement-related pixel disturbances. We tested the performance of enhancing the harmonics of the ENF in a video recordings with lower illumination intensity and noise. In Figure 6, we used the STFT method to estimate the ENF from the video recording with and without the RFA to enhance the ENF harmonics. With the improvement in the correlation coefficient of the RFA-enhanced ENF signal, the ENF can be reliably estimated from video recordings with a lower SNR, as long as it carries the embedded artificial light fluctuations.

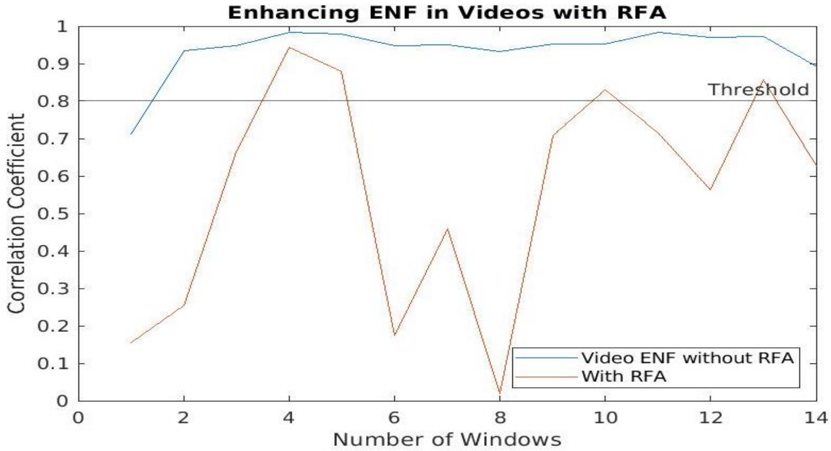

**Figure 6.** Enhancing the embedded ENF signature with the RFA in video recording with a low SNR and external noise.

Attacks on audio and video recordings such as deepfakes alter the original samples of the recordings to create a false perception. Along with the samples, the embedded ENF frequency fluctuations, which are temporal sensitive, are also altered, resulting in interference of the ENF fingerprint. Using a reference signal recorded at the same time instant, the manipulations to the multimedia recordings can be detected and also localized with a reliable ENF estimation method [35,56]. Figure 7 shows the drop in the correlation coefficient of the audio and video deepfake recordings where the ENF was estimated from the RFA-enhanced harmonics. For the video deepfake, the whole recording was swapped with an alternate trained face model, and this resulted in a drop in the overall correlation for the whole video. For the audio recording, a partial deepfake recording was generated and appended to the original recording. The correlation coefficient can also be used to localize the forgery.

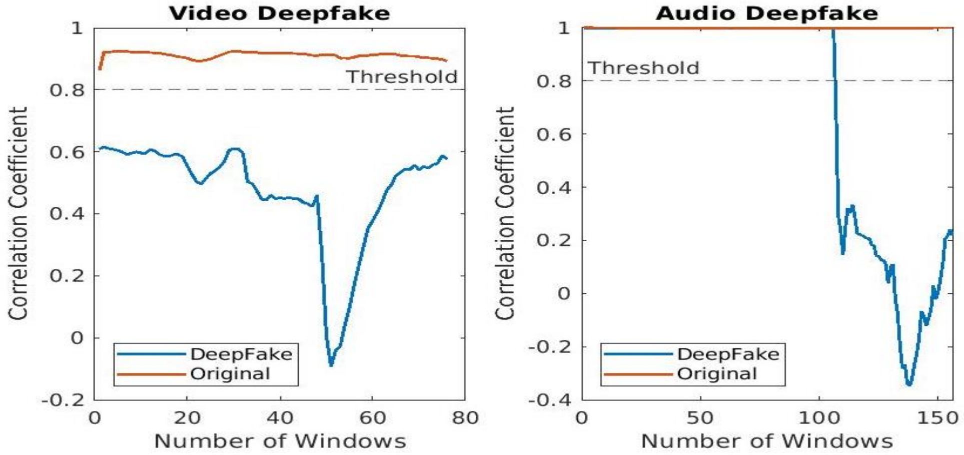

**Figure 7.** Deepfake detection in audio and video recordings by ENF signal comparison with RFA enhancement.

In order to deploy the proposed authentication scheme to reliably authenticate a continuous stream of media input such as surveillance system monitoring, the ENF should be reliably estimated for better correlation. However, sometimes, it is susceptible to outliers that occur due to a frame being skipped due to network delay or frame obfuscation. In order to address the outliers and reduce the false alarm rate, we integrated Singular Spectrum Analysis (SSA) to observe the correlation coefficient vector and suppress the outliers. The performance analysis and the computation overhead were studied from the perspective of edge-based computers.

### 5.4. SSA Performance Analysis

We evaluated SSA performance in terms of processing time and computational resource consumption on the host machine. During our test, we only repeated SSA functions, then evaluated the processing time and resource usages. Thus, deepfake video preprocessing was not considered in this test. Table 2 describes the devices used for the experimental study for SSA detection. The prototype was deployed on a small-scale Local Area Network (LAN) that consisted of multiple desktops and IoT devices. We used the desktop to simulate a fog server, while RPi devices to simulate edge servers.

**Table 2.** Configuration of experimental nodes.

| Device | Redbarn HPC | Raspberry Pi 3 (B) | Raspberry Pi 4 (B) |
|---|---|---|---|
| **CPU** | 3.4 GHz, Core (TM)<br>i7-2600K (8 cores) | 1.2 GHz, Quad core<br>Cortex-A72 (ARM v8) | 1.5 GHz, Quad core<br>Cortex-A72 (ARM v8) |
| **Memory** | 8 GB DDR3 | 1 GB SDRAM | 4 GB SDRAM |
| **Storage** | 350 G HDD | 64 GB (microSD) | 64 GB (microSD) |
| **OS** | Ubuntu 18.04 | Raspbian (Jessie) | Raspbian (Jessie) |

In deepfake attack scenarios, an adversary attempts to use forged or duplicate audio and video streams to fool video surveillance systems. Figure 8 shows how SSA detection identifies suspicious activities. We simulated attack scenarios that inject fake multimedia streams in attack ranges 200–240 and 400–430. Owing to the randomness and unpredictability of the ENF in streams, the injected audio or video streams demonstrate a very low ENF coefficient score by compared with the ground truth ENF. As a result, multimedia streams in attack ranges have $W_n$ higher than threshold $h$, and they will be marked as suspicious. Instead of relying on the experimental threshold of 0.8 used to compare the correlation coefficient factors, here, we rely on change-point detection of the SSA decomposition, where a structural change is detected if $W_n > h$, as mentioned in Equation (3) [59].

Moreover, our SSA detection can also tolerate noise influence to mitigate false alarms. Figure 8 shows that spontaneous detection points with noise influence do not significantly change $W_n$.

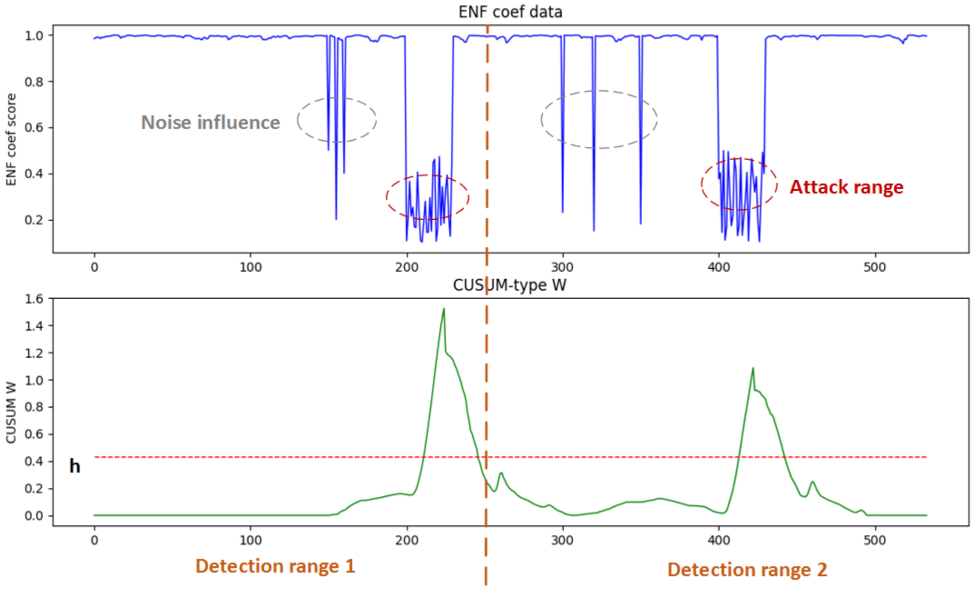

**Figure 8.** The SSA detection on deepfake attack scenarios.

Figure 9a shows the processing time given different stages during the SSA detection. The detection statistics' calculation in Stage D took the longest time, as it needed more computational resources to perform singular-value decomposition on the Hankel matrix and computing the Euclidean distances between the base matrix and test matrix. The detection score Stage S simply converts D into the normalized sum of squired distances Sn, and then, the W stage calculates the CUSUM statistics. Thus, they had less process latency than stage D. As a result, the processing time of the D stage dominated the total latency of executing SSA on all three platforms.

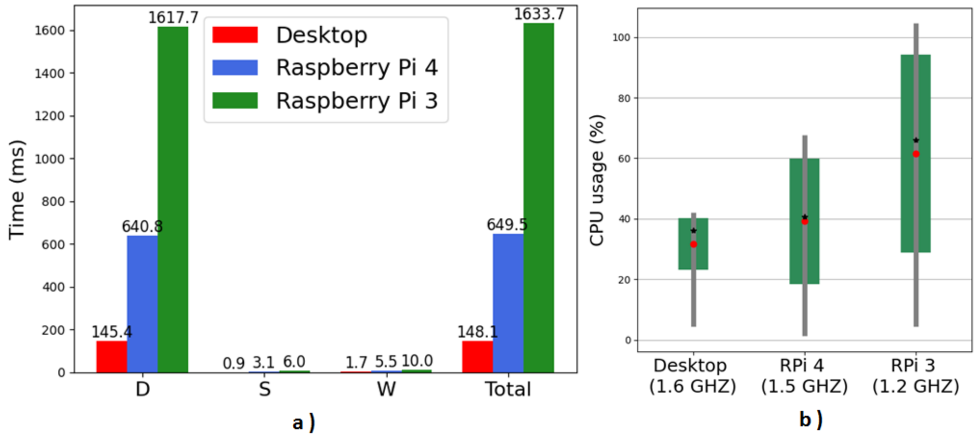

**Figure 9.** (**a**) The latency of executing SSA with different platforms; (**b**) the CPU usage of executing SSA with different platforms.

To evaluate the run time overhead of executing SSA detection on the host machine, only one core was used to run the SSA detection thread. We used the *top* command to monitor the running status of the SSA detection thread and obtained the CPU percentage distribution and average memory usage. Figure 9b shows the CPU usage percentage of executing SSA detection given different devices. Owing to different computing capability, executing SSA detection on the device with a powerful CPU core had a low mean and deviation of CPU usage percentage (desktop < RPi4 < RPi3).

The memory usages during the SSA detection were: desktop: 96 MB, Rpi4: 99 MB, Rpi3: 72 MB. Executing our SSA on different platforms may show different memory usages owing to the heterogeneous CPU architecture (X86 vs. ARM), different OSs (Ubuntu vs. Raspbian), and even various system running statuses. However, the gap was marginal, and both the desktop and Rpi almost demonstrated the same memory cost as executing SSA detection. Moreover, memory usage also included the cost due to OS tasks, such as managing the thread, but the majority of the contribution came from the SSA algorithm's execution. Therefore, it can be used to approximately evaluate the memory cost by SSA detection.

## 6. Discussion

Fake media generation using deepfake technology has raised significant concerns, and we have witnessed multiple "attacks". Although "nice" deepfaked AVS may bring benefits in multiple fields, preventive measures to distinguish fake media from authentic counterparts are necessary to prevent negative impacts. This paper introduces an ENF-fingerprint-based approach to analyze the deepfake-generated media recordings and detect manipulations. While the ENF is verified in indoor audio and video recordings due to the presence of devices connected to the power grid, there are certain limitations. For example, recordings that are generated in outdoor settings or in scenarios where the presence of the ENF is negligible, the proposed ENF authentication is not applicable. Meanwhile, in applications such as conferencing calls that occur in indoor settings under artificial light, the ENF-based detection scheme is an effective countermeasure [63].

The number of samples collected for the ENF is also a key factor for reliable estimation of the embedded signal. Video cameras based on CCD imaging sensors use a global shutter mechanism, where the ENF-carrying samples are restricted to the number of frames collected. Due to the low sampling rate and the higher nominal frequency, the Nyquist criterion is not satisfied [33]. However, the aliasing frequency can be used to identify the signal fluctuations at the cost of reduced accuracy. With the proposed integration of the RFA and a weighted combination of spectral harmonics for ENF enhancement, the algorithms add additional complexity for time-sensitive applications such as surveillance systems. The RFA approach is used to enhance the SNR of the signal prior to the ENF estimation algorithm and thereby is suggested to be applied in scenarios where the external noise interference disrupts the ENF signal noise level down to −20 dB. However, with the increase in signal duration, the time taken for the RFA also exponentially increases compared to the spectral combination method [58]. Figure 10 presents the difference in the execution time for each enhancement algorithm, where the input signal is incremented for 12 s for each round. In order to integrate the RFA with the STFT-based ENF estimation, we used a sliding window approach with a window size of 45–60 s and a shift size of five to ten seconds. For an online detection system, manipulations made to the live feed are detected in less than ten seconds of occurrence provided the ENF enhancement and SSA algorithms are integrated.

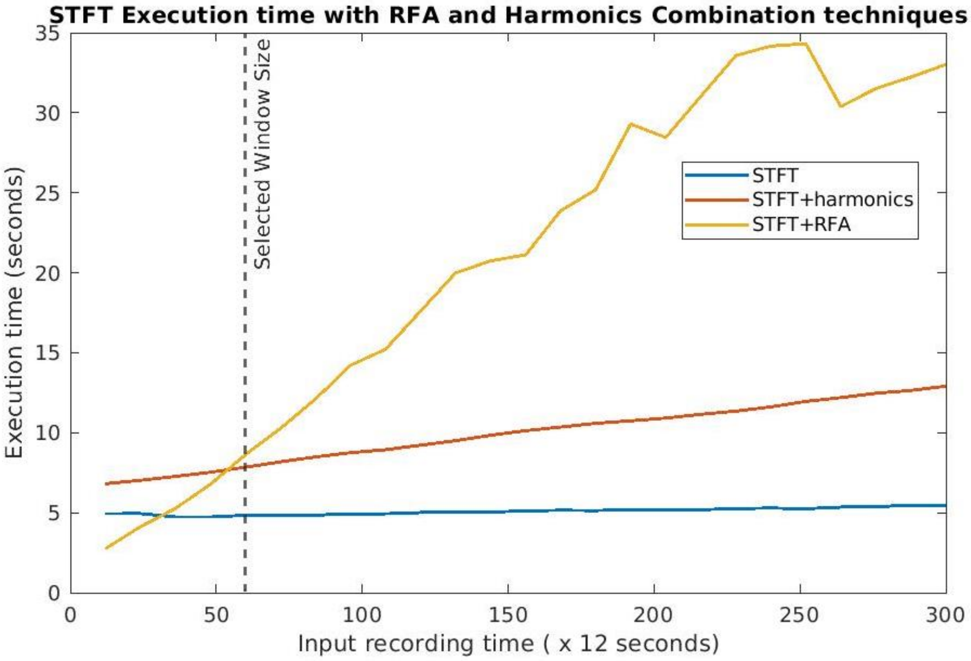

**Figure 10.** Execution time of the STFT algorithm in combination with ENF enhancement techniques.

ENF authentication is not restricted to specific media types, unlike other trained models that depend on input compatibility. It is applicable to audio and video authentication and results in a generalized solution against media manipulation attacks. In our presented work, the media manipulations were detected using an external reference ENF signal, also referred to as the ground truth signal. Deploying external circuity for this purpose could be redundant, and a central reference ENF database would not be effective since the ENF is different for each power grid. Instead, a distributed authentication scheme could be adopted where the ENF estimated from each device can be used to generate a ground truth signal without relying on an external reference signal [64]. Our previous work proposed a consensus mechanism for edge-based devices to estimate the ENF for continuous media input. The broadcast ENF was used to create a mutually agreed ground truth signal, allowing for detecting any faulty nodes. We recommend our prior work on the ENF-based consensus algorithm to detect forgery attacks for further discussion [56,63,64].

## 7. Conclusions

Emerging technologies such as deepfakes have become a common source for generating misinformation to affect trust in online media. Different from existing work on deep-learning-based detection models trained to identify deepfakes, we tackled the problem of identifying frame manipulations such as deepfakes using an environmental fingerprint technique. Using the Electrical Network Frequency (ENF) signal embedded in media recordings through artificial power sources, the integrity of the recording can be verified in both the spatial and temporal domains. In this work, we present a comprehensive analysis of effective spectral estimation techniques such as Short-Time Fourier Transform (STFT) and Multiple Signal Classification (MUSIC) against low Signal-to-Noise Ratio (SNR) media recordings. Our experimental results concluded that STFT is more reliable for ENF estimation. However, according to our findings, for media recordings with a short duration, the MUSIC algorithm has better performance for spectral estimation.

In addition to spectral analysis techniques, we tested signal enhancement algorithms such as the Robust Filtering Algorithm (RFA) and weighted harmonics combinations against deepfake audio and video recordings. From our experiments, the RFA technique significantly improved the SNR of the embedded ENF signal and resulted in reliable verification of signal authenticity. We also integrated the proposed method for online media verification, and based on the experimental results, we adopted STFT with the RFA algorithm considering the execution time complexity in our testbed. Furthermore, to minimize the false positive rate due to outliers, we deployed our ENF-based authentication scheme with the Singular Spectrum Analysis (SSA) method to improve the performance of detecting media manipulations. The results demonstrated a reliable and comprehensive tool against fake media distribution, adaptable to heterogeneous media recordings made under the influence of the power grid.

**Author Contributions:** Conceptualization, D.N., R.X., Y.C., E.B. and A.A.; Data curation, D.N. and R.X.; Formal analysis, D.N. and R.X.; Funding acquisition, Y.C.; Investigation, A.A.; Methodology, D.N., R.X., Y.C. and E.B.; Project administration, Y.C.; Resources, Y.C., E.B. and A.A.; Software, D.N. and R.X.; Supervision, Y.C. and A.A.; Validation, D.N.; Visualization, R.X.; Writing—original draft, D.N., R.X. and Y.C.; Writing—review & editing, E.B. and A.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the U.S. National Science Foundation (NSF) via Grant CNS-2039342 and the U.S. Air Force Office of Scientific Research (AFOSR) Dynamic Data and Information Processing Program (DDIP) via Grant FA9550-21-1-0229. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Air Force.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AI | Artificial Intelligence |
| AVS | Audio and/or Video Stream |
| CMOS | Complementary Metal–Oxide Semiconductor |
| CCD | Charge-Coupled Device |
| CUSUM | Cumulative Sum of deviations |
| DL | Deep Learning |
| ENF | Electrical Network Frequency |
| FFT | Fast Fourier Transform |

| | |
|---|---|
| FPS | Frames Per Second |
| GAN | General Adversarial Network |
| IF | Instantaneous Frequency |
| IoVT | Internet of Video Things |
| LAN | Local Area Network |
| MUSIC | Multiple Signal Classification |
| ML | Machine learning |
| PRNU | Photo-Response Non-Uniformity |
| PSD | Power Spectral Density |
| RFA | Robust Filtering Algorithm |
| SFM | Sinusoid Frequency Modulate |
| SNR | Signal-to-Noise Ratio |
| SSA | Singular Spectrum Analysis |
| STFD | Sinusoidal Time-Frequency Distribution |
| STFT | Short-Time Fourier Transform |
| SVD | Singular-Value Decomposition |

## References

1. Makhzani, A.; Shlens, J.; Jaitly, N.; Goodfellow, I.; Frey, B. Adversarial autoencoders. *arXiv* **2015**, arXiv:1511.05644.
2. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
3. Chan, C.; Ginosar, S.; Zhou, T.; Efros, A.A. Everybody dance now. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 5933–5942.
4. Crystal, D.T.; Cuccolo, N.G.; Ibrahim, A.; Furnas, H.; Lin, S.J. Photographic and video deepfakes have arrived: How machine learning may influence plastic surgery. *Plast. Reconstr. Surg.* **2020**, *145*, 1079–1086. [CrossRef] [PubMed]
5. Pandey, C.K.; Mishra, V.K.; Tiwari, N.K. Deepfakes: When to Use It. In Proceedings of the 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), Virtual, 10–11 December 2021; pp. 80–84.
6. Rothkopf, J. Deepfake Technology Enters the Documentary World. *The New York Times*, 1 July 2020.
7. Palmer, A. Experts Warn Digitally-Altered’Deepfakes’ Videos of Donald Trump, Vladimir Putin, and Other World Leaders Could Be Used to Manipulate Global Politics by 2020. *Daily Mail*, 12 March 2018.
8. Villasenor, J. Artificial Intelligence, Deepfakes, and the Uncertain Future of Truth. Available online: https://www.brookings.edu/blog/techtank/2019/02/14/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/ (accessed on 2 April 2019).
9. Cole, S. This Open-Source Program Deepfakes You during Zoom Meetings, in Real Time. 2020. Available online: https://www.vice.com/enus/article/g5xagy/this-open-source-program-deepfakes-you-during-zoom-meetings-in-real-time (accessed on 18 April 2022).
10. TelanganaToday. Now You Can ‘Deepfake’ Elon Musk in Zoom. 2020. Available online: https://telanganatoday.com/now-you-can-deepfake-elon-musk-in-zoom (accessed on 18 April 2022).
11. Thalen, M. Show up as a Celebrity to Your Next Zoom Meeting with This Deepfake Tool. 2020. Available online: https://www.dailydot.com/debug/live-deepfake-zoom-skype/ (accessed on 18 April 2022).
12. Poulsen, K. We Found the Guy Behind the Viral ‘Drunk Pelosi’ Video. 2019. Available online: https://www.thedailybeast.com/we-found-shawn-brooks-the-guy-behind-the-viral-drunk-pelosi-video (accessed on 18 April 2022).
13. Warner, B. Deepfake Video of Mark Zuckerberg Goes Viral on Eve of House A.I. Hearing. 2019. Available online: http://fortune.com/2019/06/12/deepfake-mark-zuckerberg/ (accessed on 18 April 2022).
14. Verdoliva, L. Media forensics and deepfakes: An overview. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 910–932. [CrossRef]
15. Hall, H.K. Deepfake Videos: When Seeing Isn’t Believing. *Cathol. Univ. J. Law Technol.* **2018**, *27*, 51–76.
16. Manheim, K.M.; Kaplan, L. Artificial Intelligence: Risks to Privacy and Democracy. *Forthcom. Yale J. Law Technol.* **2019**, *21*, 106.
17. Miller, M.J. How Cyberattacks and Disinformation Threaten Democracy. 2018. Available online: https://www.pcmag.com/article/361663/how-cyberattacks-and-disinformation-threaten-democracy (accessed on 18 April 2022).
18. Parkin, S. The Rise of the Deepfake and the Threat to Democracy. 2019. Available online: https://www.theguardian.com/technology/ng-interactive/2019/jun/22/the-rise-of-the-deepfake-and-the-threat-to-democracy (accessed on 18 April 2022).
19. Holroyd, M.; Olorunselu, F. Deepfake Zelenskyy Surrender Video Is the ‘First Intentionally Used’ in Ukraine War. 2022. Available online: https://www.euronews.com/my-europe/2022/03/16/deepfake-zelenskyy-surrender-video-is-the-first-intentionally-used-in-ukraine-war (accessed on 18 April 2022).
20. Wakefield, J. Deepfake Presidents Used in Russia-Ukraine War. 2022. Available online: https://www.bbc.com/news/technology-60780142 (accessed on 18 April 2022).
21. Johnson, T. DARPA Is Racing to Develop Tech that Can Identify Hoax Videos. 2018. Available online: https://taskandpurpose.com/deepfakes-hoax-videos-darpa/ (accessed on 18 April 2022).

22. Knight, W. The US Military Is Funding an Effort to Catch Deepfakes and Other AI Trickery. 2018. Available online: https://www.technologyreview.com/s/611146/the-us-military-is-funding-an-effort-to-catch-deepfakes-and-other-ai-trickery/ (accessed on 18 April 2022).
23. Korshunov, P.; Marcel, S. Deepfakes: A new threat to face recognition? Assessment and detection. *arXiv* **2018**, arXiv:1812.08685.
24. Foster, B. Deepfakes and AI: Fighting Cybersecurity Fire with Fire. 2020. Available online: https://threatpost.com/deepfakes-ai-fighting-cybersecurity-fire/154978/ (accessed on 18 April 2022).
25. Gandhi, A.; Jain, S. Adversarial perturbations fool deepfake detectors. *arXiv* **2020**, arXiv:2003.10596.
26. Neekhara, P.; Hussain, S.; Jere, M.; Koushanfar, F.; McAuley, J. Adversarial Deepfakes: Evaluating Vulnerability of Deepfake Detectors to Adversarial Examples. *arXiv* **2020**, arXiv:2002.12749.
27. Grigoras, C. Applications of ENF analysis in forensic authentication of digital audio and video recordings. *J. Audio Eng. Soc.* **2009**, *57*, 643–661.
28. Cooper, A.J. The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings—An automated approach. In Proceedings of the Audio Engineering Society Conference: 33rd International Conference: Audio Forensics-Theory and Practice, Denver, CO, USA, 5–7 June 2008.
29. Bollen, M.H.; Gu, I.Y. *Signal Processing of Power Quality Disturbances*; John Wiley & Sons: Hoboken, NJ, USA, 2006; Volume 30.
30. Liu, Y.; You, S.; Yao, W.; Cui, Y.; Wu, L.; Zhou, D.; Zhao, J.; Liu, H.; Liu, Y. A distribution level wide area monitoring system for the electric power grid–FNET/GridEye. *IEEE Access* **2017**, *5*, 2329–2338. [CrossRef]
31. Chai, J.; Liu, F.; Yuan, Z.; Conners, R.W.; Liu, Y. Source of ENF in battery-powered digital recordings. In *Audio Engineering Society Convention 135*; Audio Engineering Society: New York, NY, USA, 2013.
32. Fechner, N.; Kirchner, M. The humming hum: Background noise as a carrier of ENF artifacts in mobile device audio recordings. In Proceedings of the 2014 Eighth International Conference on IT Security Incident Management & IT Forensics, Münster, Germany, 12–14 May 2014; pp. 3–13.
33. Garg, R.; Varna, A.L.; Hajj-Ahmad, A.; Wu, M. "Seeing" ENF: Power-signature-based timestamp for digital multimedia via optical sensing and signal processing. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 1417–1432. [CrossRef]
34. Su, H.; Hajj-Ahmad, A.; Garg, R.; Wu, M. Exploiting rolling shutter for ENF signal extraction from video. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 5367–5371.
35. Nagothu, D.; Chen, Y.; Aved, A.; Blasch, E. Authenticating video feeds using electric network frequency estimation at the edge. *EAI Endorsed Trans. Secur. Saf.* **2021**, *7*, e4. [CrossRef]
36. Nagothu, D.; Chen, Y.; Blasch, E.; Aved, A.; Zhu, S. Detecting malicious false frame injection attacks on surveillance systems at the edge using electrical network frequency signals. *Sensors* **2019**, *19*, 2424. [CrossRef]
37. Nagothu, D.; Schwell, J.; Chen, Y.; Blasch, E.; Zhu, S. A study on smart online frame forging attacks against video surveillance system. In Proceedings of the Sensors and Systems for Space Applications XII, Baltimore, MD, USA, 15–16 April 2019; Volume 11017, p. 110170L.
38. Vatansever, S.; Dirik, A.E.; Memon, N. Factors affecting enf based time-of-recording estimation for video. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2497–2501.
39. Su, H.; Hajj-Ahmad, A.; Wu, M.; Oard, D.W. Exploring the use of ENF for multimedia synchronization. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 4613–4617.
40. Hajj-Ahmad, A.; Garg, R.; Wu, M. ENF-based region-of-recording identification for media signals. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 1125–1136. [CrossRef]
41. Jung, T.; Kim, S.; Kim, K. Deepvision: Deepfakes detection using human eye blinking pattern. *IEEE Access* **2020**, *8*, 83144–83154. [CrossRef]
42. Li, Y.; Lyu, S. Exposing deepfake videos by detecting face warping artifacts. *arXiv* **2018**, arXiv:1811.00656.
43. Matern, F.; Riess, C.; Stamminger, M. Exploiting visual artifacts to expose deepfakes and face manipulations. In Proceedings of the 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), Waikoloa Village, HI, USA, 7–11 January 2019; pp. 83–92.
44. Marra, F.; Gragnaniello, D.; Verdoliva, L.; Poggi, G. Do gans leave artificial fingerprints? In Proceedings of the 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 28–30 March 2019; pp. 506–511.
45. Ciftci, U.A.; Demir, I.; Yin, L. How do the hearts of deep fakes beat? deep fake source detection via interpreting residuals with biological signals. In Proceedings of the 2020 IEEE International Joint Conference on Biometrics (IJCB), Houston, TX, USA, 28 September–1 October 2020; pp. 1–10.
46. Jeong, Y.; Kim, D.; Min, S.; Joe, S.; Gwon, Y.; Choi, J. BiHPF: Bilateral High-Pass Filters for Robust Deepfake Detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2022; pp. 48–57.
47. Durall, R.; Keuper, M.; Keuper, J. Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 7890–7899.

48. Frank, J.; Eisenhofer, T.; Schönherr, L.; Fischer, A.; Kolossa, D.; Holz, T. Leveraging frequency analysis for deep fake image recognition. In Proceedings of the International Conference on Machine Learning PMLR, Virtual, 13–18 July 2020; pp. 3247–3258.
49. Cozzolino, D.; Verdoliva, L. Noiseprint: A CNN-based camera model fingerprint. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 144–159. [CrossRef]
50. Lukas, J.; Fridrich, J.; Goljan, M. Digital camera identification from sensor pattern noise. *IEEE Trans. Inf. Forensics Secur.* **2006**, *1*, 205–214. [CrossRef]
51. Li, W.; Yuan, Y.; Yu, N. Passive detection of doctored JPEG image via block artifact grid extraction. *Signal Process.* **2009**, *89*, 1821–1829. [CrossRef]
52. Cozzolino, D.; Thies, J.; Rossler, A.; Nießner, M.; Verdoliva, L. SpoC: Spoofing camera fingerprints. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 990–1000.
53. Garg, R.; Hajj-Ahmad, A.; Wu, M. Feasibility Study on Intra-Grid Location Estimation Using Power ENF Signals. *arXiv* **2021**, arXiv:2105.00668.
54. Hajj-Ahmad, A.; Garg, R.; Wu, M. Spectrum combining for ENF signal estimation. *IEEE Signal Process. Lett.* **2013**, *20*, 885–888. [CrossRef]
55. Chuang, W.H.; Garg, R.; Wu, M. How secure are power network signature based time stamps? In Proceedings of the 2012 ACM Conference on Computer and Communications Security, Raleigh, NC, USA, 16–18 October 2012; pp. 428–438.
56. Nagothu, D.; Xu, R.; Chen, Y.; Blasch, E.; Aved, A. Detecting Compromised Edge Smart Cameras using Lightweight Environmental Fingerprint Consensus. In Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra, Portugal, 15–17 November 2021; pp. 505–510.
57. Hajj-Ahmad, A.; Garg, R.; Wu, M. Instantaneous frequency estimation and localization for ENF signals. In Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, Hollywood, CA, USA, 3–6 December 2012; pp. 1–10.
58. Hua, G.; Zhang, H. ENF signal enhancement in audio recordings. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 1868–1878. [CrossRef]
59. Moskvina, V.; Zhigljavsky, A. An algorithm based on singular spectrum analysis for change-point detection. *Commun. Stat.-Simul. Comput.* **2003**, *32*, 319–352. [CrossRef]
60. Hassani, H. Singular spectrum analysis: Methodology and comparison. *J. Data Sci.* **2007**, *5*, 239–257. [CrossRef]
61. Perov, I.; Gao, D.; Chervoniy, N.; Liu, K.; Marangonda, S.; Umé, C.; Dpfks, M.; Facenheim, C.S.; Rp, L.; Jiang, J.; et al. Deepfacelab: A simple, flexible and extensible face swapping framework. *arXiv* **2020**, arXiv:2005.05535.
62. Descript | Create Podcasts, Videos, and Transcripts. Available online: https://www.descript.com/ (accessed on 18 April 2022).
63. Nagothu, D.; Xu, R.; Chen, Y.; Blasch, E.; Aved, A. DeFake: Decentralized ENF-Consensus Based DeepFake Detection in Video Conferencing. In Proceedings of the 2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP), Tampere, Finland, 6–8 October 2021; pp. 1–6.
64. Xu, R.; Nagothu, D.; Chen, Y. Econledger: A proof-of-enf consensus based lightweight distributed ledger for iovt networks. *Future Internet* **2021**, *13*, 248. [CrossRef]