eng

MDPI

Article

A New Deepfake Detection Method with No-Reference Image Quality Assessment to Resist Image Degradation

Jiajun Jiang [1,*], Wen-Chao Yang [2], Chung-Hao Chen [1] and Timothy Young [3]

[1] Electrical and Computer Engineering Department, Old Dominion University, Norfolk, VA 23529, USA; cxchen@odu.edu
[2] Department of Forensic Science, Central Police University, Taoyuan City 333322, Taiwan; una135@mail.cpu.edu.tw
[3] Data Science Institute for Machine Learning and AI, The University of Tennessee, Knoxville, TN 37996, USA; tmyoung1@utk.edu
* Correspondence: jjian001@odu.edu

Abstract

Deepfake technology, which utilizes advanced AI models such as Generative Adversarial Networks (GANs), has led to the proliferation of highly convincing manipulated media, posing significant challenges for detection. Existing detection methods often struggle with the low-quality or compressed press, which is prevalent on social media platforms. This paper proposes a novel Deepfake detection framework that leverages No-Reference Image Quality Assessment (NRIQA) techniques, specifically, BRISQUE, NIQE, and PIQUE, to extract quality-related features from facial images. These features are then classified using a Support Vector Machine (SVM) with various kernel functions. We evaluate our method under both intra-dataset and cross-dataset settings. For intra-dataset evaluation, we conduct K-fold cross-validation on two benchmark datasets, DFDC and Celeb-DF (v2), including downsampled versions to simulate real-world degradation. The results show that our method maintains high accuracy even under significant quality loss, achieving up to 98% accuracy on the Celeb-DF (v2) dataset and outperforming several state-of-the-art methods. To improve the transferability of the detection models, we introduce an integrated filtering strategy based on NR-IQA thresholding, which enhances performance in cross-dataset transfer scenarios. This approach yields up to 7% improvement in detection accuracy under challenging cross-domain conditions.

Keywords: Deepfake; No-Reference Image Quality Assessment; image degradation; machine learning

check for updates

Received: 15 August 2025
Revised: 29 September 2025
Accepted: 10 October 2025
Published: 11 October 2025

Citation: Jiang, J.; Yang, W.-C.; Chen, C.-H.; Young, T. A New Deepfake Detection Method with No-Reference Image Quality Assessment to Resist Image Degradation. *Eng* **2025**, *6*, 274. https://doi.org/10.3390/eng6100274

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Deepfake technology is a sophisticated form of AI-based face changing, combining "Deep Learning" and "Fake" to generate highly convincing manipulated videos or images. This technology gained prominence when a Reddit user named "Deepfakes" open-sourced it in 2017, and it has since been widely discussed on forums. While you may not be familiar with the term, you have likely encountered AI face-changing in various forms, from early spoofs of celebrities [1] to more recent video synthesis tools like FakeApp [2], FaceApp [3], and FaceSwap [4]. In addition, GitHub (https://github.com, accessed on 10 October 2025) hosts various open-source projects and related tutorials, including the DeepFaceLab [5]. These applications have reached a level of technical maturity that allows even those without

a computer-related background to easily manipulate human faces, changing facial features, expressions, age, gender, and other facial attributes.

Deepfake video creation has become a concerning issue as it can be used to create convincing fake videos to spread disinformation, manipulate public opinion, and damage reputations. Although Deepfake technology has positive applications, such as enhancing game graphics or replacing actors' faces in film production, its malicious uses are far more prevalent and can lead to severe consequences. The recent release of powerful video generators like OpenAI's Sora, a diffusion model capable of creating realistic videos from text descriptions, further highlights the rapid advancement in this field and its potential impact on media industries [6]. To counter the harm caused by Deepfakes, many research institutions and organizations have initiated research plans, projects, and academic competitions aimed at developing effective detection methods.

A critical challenge for these detection methods is the behavior of social media platforms like TikTok, Twitter, and others. These platforms often compress videos and images to save storage space and bandwidth. While many Deepfake detection techniques perform well on high-quality (HQ) or uncompressed video data, their performance often degrades significantly on the low-quality (LQ) or compressed media prevalent online. Criminals may deliberately spread compressed fake videos or images, making it harder to identify them as fake. Therefore, developing robust detection methods that are resilient to image degradation is a significant and urgent challenge in multimedia forensics.

To address this gap, this paper proposes a novel, lightweight Deepfake detection framework that leverages No-Reference Image Quality Assessment (NR-IQA) techniques. Our approach is specifically designed to maintain high accuracy even when faced with significant quality loss from downsampling and compression.

The main contributions of this work are as follows:

- We systematically apply three distinct NR-IQA models (BRISQUE, NIQE, and PIQUE) to extract robust features for Deepfake detection, demonstrating their effectiveness under severe image degradation (downsampling to 1/4 and 1/16 of the original resolution).
- We propose an integrated, quality-based filtering strategy using NR-IQA thresholds. This pre-classification step improves the generalization and transferability of detection models in cross-dataset scenarios.
- Our framework is computationally efficient and lightweight, making it a practical solution for real-world applications where speed and resource constraints are a concern.
- We conduct a comprehensive evaluation under strict intra-dataset and cross-dataset protocols, providing a reproducible and thorough analysis of our method's performance compared to established deep learning baselines.

The remainder of this paper is organized as follows. Section 2 reviews related work in Deepfake creation and detection. Section 3 details our proposed method. Section 4 presents the experimental results, and Section 5 concludes the paper.

2. Related Work

This section provides an overview of the technologies behind Deepfake creation and the various approaches developed for their detection.

2.1. Deepfake Creation

The foundation of most Deepfake technology is the Generative Adversarial Network (GAN), a framework proposed by Goodfellow et al. [7] that consists of two competing neural networks: a generator and a discriminator. Through an iterative training process, the generator learns to produce increasingly realistic fake images, while the discriminator

becomes better at distinguishing them from real ones. The evolution of GANs has led to several powerful architectures:

DCGAN (Deep Convolutional Generative Adversarial Network): Introduced by Radford et al. [8], DCGAN integrated deep convolutional networks into the GAN architecture, enabling the generation of clearer and more stable high-resolution images. Its strength lies in learning a hierarchy of features, but it can sometimes struggle with capturing the global structure of an image.

StyleGAN: Developed by Karras et al. [9], StyleGAN introduced a style-based generator that allows for more intuitive and scale-specific control over the synthesis process. Its main advantage is the ability to generate highly realistic and high-resolution faces with disentangled attributes (e.g., pose, identity, expression), though it requires significant computational resources for training.

StackGAN: Proposed by Zhang et al. [10], StackGAN specializes in converting text descriptions into photo-realistic images. Its key strength is a two-stage generative process, which first creates a low-resolution image with basic shapes and colors, and then refines it into a high-resolution, detailed image. This progressive approach improves image quality, but it may still face challenges in perfectly aligning complex scenes with intricate text prompts.

CycleGAN & StarGAN: These architectures address the challenge of image-to-image translation. CycleGAN [11] can translate between two domains without paired training data (e.g., converting a horse to a zebra). Its strength is its unsupervised nature, but it can sometimes introduce unwanted artifacts. StarGAN [12], an extension of CycleGAN, enables translation across multiple domains using a single model, offering greater flexibility and efficiency.

These advancements have made Deepfake creation more accessible and the results more convincing, necessitating the development of equally sophisticated detection methods.

2.2. Deepfake Detection

Deepfake video detection has become an important area of research to mitigate the harmful effects of Deepfakes. Various techniques have been proposed for detecting Deepfake videos, including analyzing facial movements, detecting inconsistencies in lighting, and using deep learning algorithms to detect manipulated frames or artifacts. Deepfake video detection is crucial as it enables us to identify and remove malicious content before it can cause harm.

The consequences of not detecting Deepfakes can be significant, leading to political unrest, misinformation, and harm to individuals and organizations. Therefore, developing effective Deepfake detection methods is crucial for safeguarding against the negative impacts of Deepfakes.

2.2.1. Deep Learning-Based Methods

This is the most common approach, where deep neural networks are trained to identify artifacts and inconsistencies in manipulated media.

CNN-based Architectures: Models like MesoNet [13] and XceptionNet [14,15] are designed to capture forgery artifacts at different scales. MesoNet is a lightweight CNN focused on the mesoscopic properties of images, making it computationally efficient but potentially less robust on high-resolution fakes. XceptionNet, a deeper and more powerful architecture, often achieves state-of-the-art performance by learning subtle, low-level artifacts but requires more training data and computational power.

Capsule Networks (CapsuleNet): Introduced as an alternative to traditional CNNs, Capsule Networks aim to better model hierarchical relationships within an image [16].

Instead of individual neurons, they use groups of neurons called "capsules" that can encode spatial information and object properties. Their strength lies in their theoretical robustness to changes in viewpoint and orientation. Nguyen et al. [17] applied this architecture to forensics, demonstrating its potential for detecting forged images. However, Capsule Networks can be more computationally complex and challenging to train than standard CNNs.

Head Pose Analysis: Methods like HeadPose [18] move beyond texture analysis to leverage geometric inconsistencies. This approach identifies discrepancies in 3D head orientations between the synthesized face and the original body, offering a unique detection angle that is robust to some image-level manipulations. Its limitation is that it may fail if the forgery method produces a geometrically consistent result.

Recurrent and 3D CNNs: To capture temporal inconsistencies between video frames, some methods use networks with a temporal dimension. Sabir et al. [19] employed a combination of CNN and RNN, while Zhang et al. [20] used a 3D-CNN. These methods are strong at detecting unnatural motion or flickering but are computationally expensive and may not work on single-image Deepfakes.

Attention and Transformer Models: More recent approaches like ISTVT [21] use Vision Transformers to capture spatial artifacts and temporal inconsistencies simultaneously. Their strength is the ability to model long-range dependencies in data, but they typically require very large datasets for effective training.

While powerful, a common weakness of many deep learning models is their struggle with generalization to unseen datasets and their vulnerability to performance degradation on compressed or low-quality media, which is the primary motivation for our work.

2.2.2. Pattern Recognition-Based Methods

These methods leverage traditional forensic techniques, such as analyzing Photo-Response Non-Uniformity (PRNU), also known as Sensor Pattern Noise (SPN), which is a unique fingerprint left by a camera's sensor on an image. Lukas et al. [22,23] first proposed using SPN for source camera identification, employing techniques like the wavelet-based Mihcak filter [24] to extract the noise pattern. Following this foundational work, researchers like Koopman et al. [25] applied PRNU analysis to the broader task of detecting digital media manipulation.

The main advantage of the PRNU/SPN approach is its strong theoretical grounding and robustness in certain conditions. However, a primary limitation is that the SPN signal might be damaged or removed by common operations like heavy compression and social media reprocessing, limiting its effectiveness for online media. Another challenge is the high computational cost, especially for video analysis. Recent work has sought to address this, such as the method proposed by Yang et al. [26], which accelerates PRNU analysis for videos by focusing on I-frames to make the process more feasible for forensic investigations.

Beyond SPN, other pattern recognition techniques have also been explored. More recently, methods like GrDT [27] have used graph-based approaches to model the geometric relationships of facial landmarks and textures to detect Deepfakes.

2.2.3. Image Quality Assessment-Based Methods

This emerging category operates on the premise that the Deepfake generation process introduces subtle degradations and inconsistencies in image quality that may not be perceptible to humans but can be quantified. These methods often use No-Reference Image Quality Assessment (NR-IQA), which evaluates an image's quality without needing a pristine reference.

Statistical NR-IQA Models: Our work falls into this category, utilizing models like BRISQUE, NIQE, and PIQUE. Mittal et al. [28] first proposed BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator), a model based on natural scene statistics (NSS) in the spatial domain. It computes mean-subtracted contrast-normalized (MSCN) coefficients, which are sensitive to a wide range of distortions. The MSCN of an intensity image $I(i,j)$ is defined as follows [28]:

$$\hat{I}(i,j) = (I(i,j) - \mu(i,j))/(\sigma(i,j) + 1) \quad (1)$$

where $i \in 1,2,\ldots,M$ and $j \in 1,2,\ldots,N$ are spatial indices, and $\mu(i,j)$ and $\sigma(i,j)$ are the local mean and standard deviation, respectively:

$$\mu(i,j) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l} I(i+k, j+l) \quad (2)$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}[I(i+k, j+l)]^2} \quad (3)$$

where w is a 2D circularly symmetric Gaussian weighting function. Following this, the same authors proposed NIQE (Natural Image Quality Evaluator) [29], which operates on a similar statistical basis. Venkatanath et al. [30] later introduced PIQUE (Perception-based Image Quality Evaluator), a block-based method that uses local features to evaluate quality, as illustrated in Figure 1. The primary strength of these models, which we exploit, is their computational efficiency and their robustness when image quality is already low. Yang et al. [31] previously used these models but relied on simple thresholds for classification. Our work extends this by integrating these features with a more powerful SVM classifier and proposing a novel filtering strategy for improved generalization.

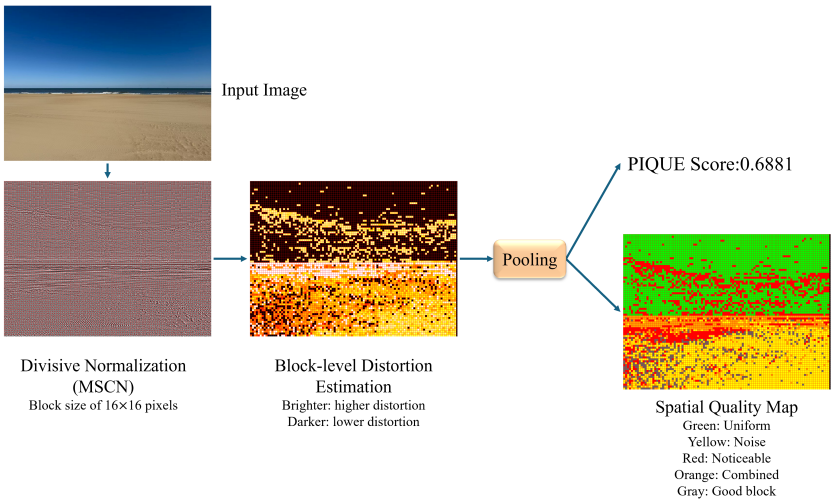


Figure 1. Block diagram of the PIQUE method.

Integrated IQA Systems: Recently, Bouhamed et al. proposed the Facial Image Authenticity Verification (FIAV) method [32], which integrates various learning-based NR-IQA metrics (such as MANIQA and TOPIQ-NR-FACE) with other visual quality features like brightness, sharpness, and blur to distinguish between authentic and AI-generated faces. Their system provides a practical web-based interface for real-time analysis. While such integrated systems show strong performance, the challenge of maintaining detection accuracy on severely degraded or downsampled media remains an important area of investigation. Therefore, our work focuses on the robustness of statistical NR-IQA models (BRISQUE, NIQE, PIQUE) and introduces a quality-based filtering strategy to enhance performance in low-quality and cross-dataset scenarios.

2.3. Discussion on Low-Quality Image Detection

Detecting Deepfakes in low-quality facial images is a significant challenge, as compression and downsampling can obscure or even remove the subtle artifacts that many detectors rely on. Another promising direction is to consider image enhancement and restoration as a potential preprocessing step before detection.

Recent works in image restoration offer powerful tools for this. For instance, the NTIRE 2025 challenge on real-world face restoration [33] has spurred the development of advanced models, many of which are based on Diffusion and Transformer architectures, capable of recovering high-quality facial details from severely degraded inputs. Similarly, Liu et al. proposed VNDHR [34], a variational framework for nighttime image dehazing that effectively removes haze, noise, and color distortions to enhance visibility.

Applying such restoration techniques as a preprocessing step could offer two main benefits: (1) it could restore facial details, making it easier for a subsequent detector to analyze features, and (2) it could standardize the quality of input images, potentially improving a detector's generalization. It is a valuable avenue for future research. However, this approach must be considered carefully. The restoration process might inadvertently "repair" the artifacts that a Deepfake detector is designed to find, or it could introduce new, model-specific artifacts that confuse the detector. A thorough investigation is needed to understand the trade-offs between improved visual quality and preserving forensic traces.

3. Methods

This section presents the proposed Deepfake detection framework, which builds upon the approach introduced in [31]. Our framework is fully automated and modular, enabling flexibility in incorporating different submodules. The method integrates three No-Reference Image Quality Assessment (NR-IQA) techniques—BRISQUE, NIQE, and PIQUE—to extract quality-related features from detected face regions in video frames. These features are subsequently fed into a Support Vector Machine (SVM) classifier with various kernel functions to determine whether a video is genuine or manipulated. The complete workflow of the proposed system is illustrated in Figure 2.

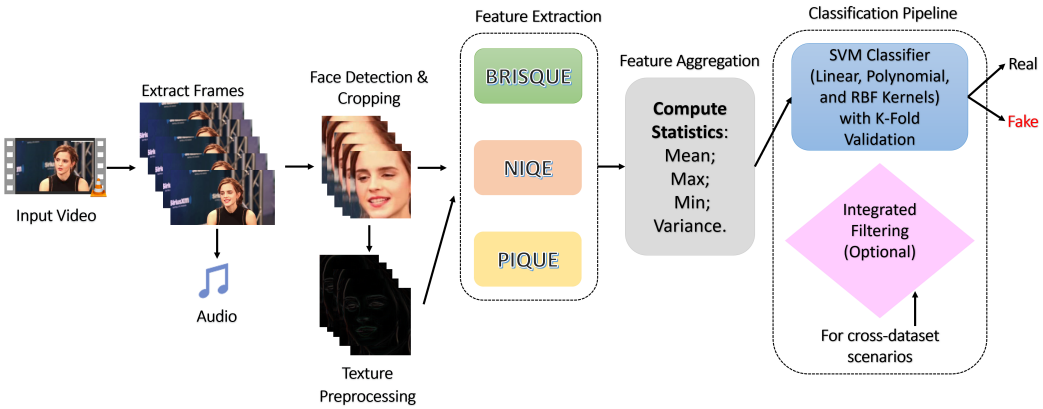


Figure 2. Flowchart of the proposed Deepfake detection framework.

The detection pipeline consists of several sequential stages, each designed to optimize the detection accuracy and computational efficiency. The following subsections describe each component in detail.

In Deepfake detection, we first extract the frames and audio of the input video using the third-party tool FFmpeg. The faces are obtained from the extracted frames. After extracting faces, the NRIQA methods are used to extract the features. Finally, an SVM

(Support Vector Machine) classifier is used to classify the Deepfake videos. The details of the other processes are described as follows:

3.1. Frame and Audio Extraction

We first decompose the input video into individual frames and audio streams using the open-source FFmpeg toolkit. This step is conducted using guidelines from the Scientific Working Group on Digital Evidence (SWGDE) [35], ensuring forensic integrity and reproducibility. All frames are extracted at the original video frame rate to preserve temporal consistency for further processing.

3.2. Face Detection and Extraction

For robust and real-time face localization, we employ the YOLOv4 (You Only Look Once, version 4) object detection framework as proposed by Bochkovskiy et al. [36]. YOLOv4 improves upon previous versions through architectural enhancements and advanced training strategies. We train the model on the WIDER FACE dataset [37], which contains diverse facial images exhibiting wide variations in pose, scale, and occlusion. Detected facial regions are cropped from each frame and passed to the next stage for further analysis.

3.3. Preprocessing Module

To enhance the discriminative power of the extracted facial regions, we incorporate a texture-aware preprocessing module inspired by Xia et al. [38]. This module focuses on emphasizing high-texture regions, as Deepfake manipulations often produce smoother and more homogeneous surfaces compared to authentic videos. The preprocessing is mathematically defined as:

$$p_{i,j} = \sqrt{(m_{i,j} - m_{i+1,j})^2 + (m_{i,j} - m_{i,j+1})^2}, \tag{4}$$

where $m_{i,j}$ denotes the grayscale intensity at pixel location (i, j), and $p_{i,j}$ is the resulting texture-enhanced pixel value. The image dimensions are defined by height H and width W. This operation highlights abrupt local changes in pixel intensity, thereby capturing fine-grained artifacts commonly present in manipulated regions. To further illustrate the effectiveness of the preprocessing step, we present a visual comparison in Figure 3. The figure shows sample face regions extracted from two different datasets, before and after applying the high-texture enhancement described above. As seen in the comparison, the preprocessing module significantly emphasizes texture-rich areas, which are critical for detecting subtle artifacts introduced by Deepfake generation methods. Theoretically, this enhancement helps suppress the smoother, artifact-prone regions that typically appear in manipulated facial regions, thereby improving the discriminative power of subsequent quality assessment features.

3.4. Feature Extraction via NR-IQA

Once the facial regions are preprocessed, we apply three established NR-IQA methods—BRISQUE, NIQE, and PIQUE—to extract perceptual quality scores. The rationale for using multiple methods lies in their complementary sensitivities to different types of distortions.

BRISQUE [28]: As detailed in Section 2.2.3, BRISQUE operates by calculating MSCN coefficients, which quantify local normalized brightness. These coefficients are highly sensitive to a wide range of distortions, including blur, noise, and compression artifacts commonly found in Deepfakes.

NIQE [29]: NIQE works by comparing the statistics of a given image patch to a pre-trained model of "natural" image statistics. It builds a multivariate Gaussian (MVG) model from features of pristine images. The quality score of a test image is then the Mahalanobis distance between the MVG model of its features and the natural MVG model. A smaller distance implies higher quality and closer resemblance to natural scenes.

PIQUE [30]: Unlike the other two methods, PIQUE is a block-based approach. It divides the image into non-overlapping blocks and estimates distortion at the block level. It specifically measures perceptible distortions related to spatial correlation, such as blockiness and noise, before pooling these scores into a final image-wide quality assessment.

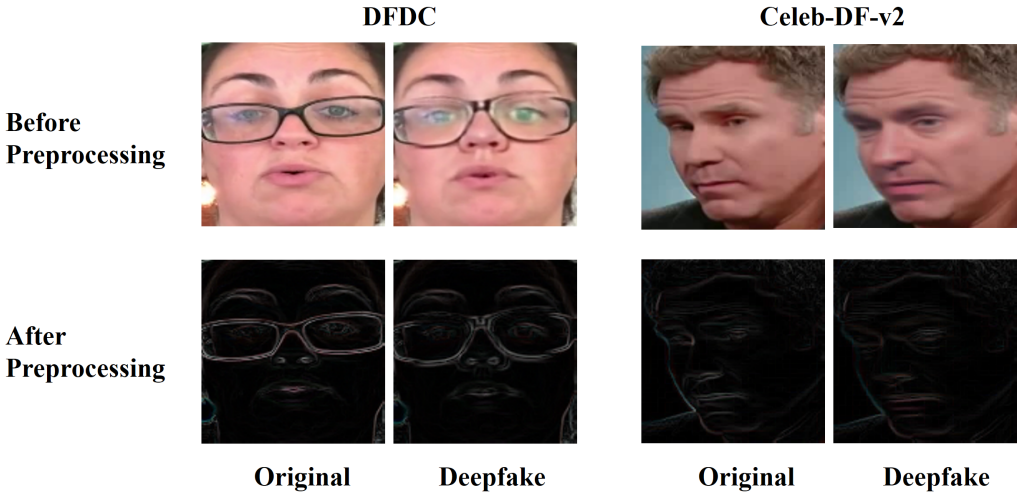


Figure 3. Comparison of two datasets before and after preprocessing.

For each video, we compute these three scores across all detected and preprocessed face regions. To create a concise and robust representation for the entire video, we then calculate four statistical descriptors from the sequence of scores for each NR-IQA method: the mean, maximum, minimum, and variance. This process results in a 12-dimensional feature vector for each video (4 statistics $\times$ 3 NR-IQA methods). This aggregated feature vector serves as the input to our classification module.

3.5. Classification Using SVM

To perform binary classification (i.e., real vs. Deepfake), we adopt the Support Vector Machine (SVM) model due to its effectiveness with small to medium-sized datasets and strong generalization capabilities. We evaluate three kernel functions:

Linear kernel: Suitable for linearly separable data, offering low computational complexity.

Polynomial kernel: Capable of capturing non-linear relationships by introducing polynomial features of the input data.

Radial Basis Function (RBF) kernel: Effective in high-dimensional feature spaces, particularly when the decision boundary is highly non-linear.

Hyperparameters for each kernel are fine-tuned via grid search using cross-validation on the training set.

3.6. Performance Evaluation

The performance of our detection framework is assessed using K-fold cross-validation (with $K = 5, 6, \ldots, 12$) to ensure statistical robustness. We evaluate classification accuracy using different combinations of the NR-IQA feature subsets:

Single-factor: Using one of BRISQUE, NIQE, or PIQUE.

Two-factor combinations: Using any pairwise combination of the three features.

Three-factor fusion: Using all three feature sets (BRISQUE, NIQE, and PIQUE).

The final performance metrics are reported in terms of mean accuracy across the K folds. To ensure a fair evaluation and prevent data leakage, we maintained strict identity separation, meaning that subjects present in the training folds were excluded from the test fold in each cross-validation split. This evaluation strategy allows for a comprehensive understanding of the contributions and limitations of each component in our framework.

3.7. Additional Considerations: Resolution Robustness and Integrated Filtering

To evaluate the robustness of the proposed framework under low-quality conditions commonly found in compressed or streamed video content, we conduct additional experiments on downscaled datasets. Specifically, we generate two lower-resolution versions of each video frame by reducing both height and width by half (1/2 H, 1/2 W) and by one-fourth (1/4 H, 1/4 W), resulting in frames at one-quarter and one-sixteenth the original resolution, respectively. Faces are extracted separately from these downsampled frames and processed through the same pipeline as the original-resolution data.

Furthermore, to enhance detection accuracy and generalizability across domains, we incorporate an integrated filtering strategy based on NR-IQA features. This pre-classification step employs statistically derived quality thresholds to filter out high-confidence samples, allowing the primary classifier to focus on ambiguous or borderline cases. The integration of such a filtering mechanism is theoretically beneficial for cross-dataset transfer and low-resolution inputs, which are known to challenge the generalization ability of existing models.

These two strategies, resolution-based testing and quality-based filtering, are not explicitly illustrated in Figure 2, but are implemented as modular extensions following the face extraction and feature extraction stages. Their effectiveness is empirically evaluated in Section 4.5, which presents results under cross-dataset and low-resolution scenarios.

4. Experimental Results

This section first describes the datasets used for our experiments, then details the implementation specifics and experimental procedures. Finally, we present and discuss the results of our evaluations.

4.1. Deepfake Video Datasets

Due to the high demand for data in deep learning methods, researchers have accumulated large video datasets for training and testing while conducting studies on Deepfake detection. Our study primarily utilizes the DFDC and Celeb-DF (v2) datasets. A summary of these and other common datasets is provided below.

FaceForensics++ (FF++): FaceForensics++ [39] is a widely used dataset for Deepfake detection, extended from the original FaceForensics dataset [40]. It includes four manipulation methods: Deepfakes [1], Face2Face [40], FaceSwap [4], and NeuralTextures [41]. The dataset comprises 1000 original videos hlfrom the YouTube-8M dataset [42] and 4000 manipulated videos, provided in both uncompressed and compressed formats.

Deepfake Detection Challenge (DFDC): The DFDC hldataset [43,44], created for a large-scale public challenge, features approximately 19,000 original videos and over 100,000 fake videos. It was generated using various face-swap methods and includes augmentations like varied frame rates and color transformations, making it a diverse and challenging benchmark.

Celeb-DF (v2): The Celeb-DF (v2) dataset [45,46] is known for its high-quality Deepfake videos with fewer visual artifacts, making it one of the most challenging datasets for detection. It contains 590 original videos from YouTube and 5639 corresponding Deepfake videos, featuring subjects of varying ages, ethnicities, and genders.

4.2. Implementation Details

Due to the generation of the FaceForensics dataset, we primarily used the DFDC and Celeb-DF datasets to validate our method. For the DFDC dataset, we used 303 pairs of already labeled one-to-one correspondence between the original video and the Deepfake video. Since the background lighting in some videos is too dark and the number of extracted faces is too small, these data are removed to avoid affecting the generalizability of the computed means. Finally, 287 pairs of original and Deepfake face datasets are obtained for use in our model. Each video is 10 s (300 frames) with the exact resolution of 1920 × 1080.

We randomly selected 200 pairs of original and Deepfake videos with various resolutions from the Celeb-DF-v2 dataset. Each video contains a varying number of frames, from 230 to 530.

This experiment is performed on a PC with an AMD Ryzen 7 CPU at 2 GHz, 16 GB RAM, Windows 10 64-bit, and MATLAB 2021b. We extract the video frames into BMP format for more detailed image information.

4.3. Comparisons with Different Downscaled Datasets

As mentioned previously, short video platforms often use downscaled or compressed videos due to limitations in network bandwidth and storage capacity. Since existing methods often lack robust validation on such downscaled datasets, we designed our experiments to verify the detection effectiveness of our method on low-quality images.

To simulate this, we downsampled the height (H) and width (W) of the original frames, see Figure 4. The scaling was applied as follows:

1/4 Resolution: The width and height were each scaled by a factor of 1/2 (i.e., 1/2 H, 1/2 W). This results in an image containing one-quarter (1/4) of the original number of pixels.

1/16 Resolution: The width and height were each scaled by a factor of 1/4 (i.e., 1/4 W, 1/4 H). This results in an image containing one-sixteenth (1/16) of the original number of pixels.

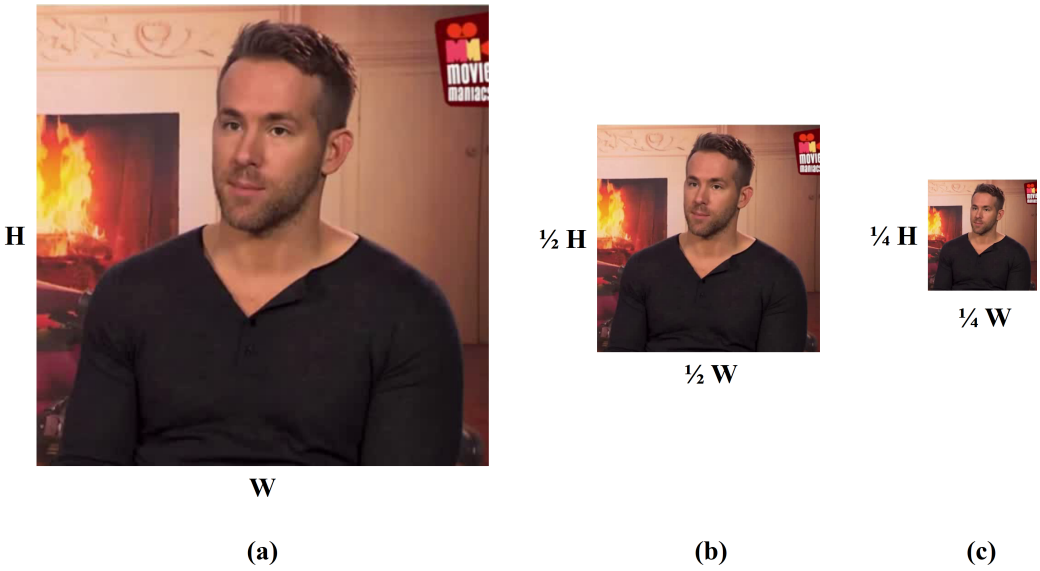


Figure 4. (**a**) Original frame, (**b**) 1/4 downsampling fame, (**c**) 1/16 downsampling frame.

For each sample video, we select the maximum, minimum, average, and variance of the three NRIQA methods as training features for model training. During the training process, considering that the performance of the training data on different kernels will be different, we choose one factor, two factors, or three factors to be trained based on the

linear kernel, polynomial kernel, and RBF kernel, respectively. The error rates of the K-fold verification result are shown in Tables 1 and 2.

Table 1. Comparisons with the original, downsampling of 1/4, and downsampling of 1/16 based on DFDC dataset.

Resolution	Kernel		Linear	Poly	RBF
Original 1920 × 1080	Faces	Single	0.4199	0.4512	0.3484
		Two Factors	0.4181	0.439	0.2387
		Three Factors	0.4181	0.4408	***0.1986***
	Texture	Single	0.453	0.4669	0.392
		Two Factors	0.3519	0.4495	0.2927
		Three Factors	0.3136	0.4425	0.2334
1/4 Downsampling 960 × 540	Faces	Single	0.453	0.4443	0.3833
		Two Factors	0.453	0.453	0.2613
		Three Factors	0.4443	0.4477	***0.2091***
	Texture	Single	0.4146	0.4216	0.3676
		Two Factors	0.4129	0.4146	0.2578
		Three Factors	0.4059	0.4251	***0.2091***
1/16 Downsampling 480 × 270	Faces	Single	0.4592	0.4344	0.3918
		Two Factors	0.4468	0.4291	0.2624
		Three Factors	0.4362	0.422	***0.2199***
	Texture	Single	0.4539	0.4574	0.3706
		Two Factors	0.4539	0.4557	0.266
		Three Factors	0.4415	0.4362	0.2216

Note. **Bold** and *italic* values indicate the best results within each experimental group.

Table 2. Comparisons with the original, downsampling of 1/4, and downsampling of 1/16 based on Celeb-DF v2 dataset.

Resolution	Kernel		Linear	Poly	RBF
Original 1920 × 1080	Face	Single	0.43	0.47	0.1775
		Two Factors	0.38	0.4125	0.02
		Three Factors	0.355	0.4075	***0.0175***
	Texture	Single	0.4075	0.445	0.355
		Two Factors	0.3925	0.4125	0.1125
		Three Factors	0.37	0.3975	***0.0175***
1/4 Downsampling 960 × 540	Face	Single	0.3375	0.4075	0.1525
		Two Factors	0.3275	0.365	0.02
		Three Factors	0.3275	0.3375	***0.0175***
	Texture	Single	0.4525	0.4	0.2975
		Two Factors	0.42	0.37	0.08
		Three Factors	0.4075	0.37	0.02
1/16 Downsampling 480 × 270	Face	Single	0.4275	0.445	0.27
		Two Factors	0.3975	0.3825	0.05
		Three Factors	0.39	0.3725	***0.0175***
	Texture	Single	0.4825	0.46	0.3325
		Two Factors	0.4525	0.455	0.045
		Three Factors	0.4375	0.44	0.02

Note. **Bold** and *italic* values indicate the best results within each experimental group.

The results show that the DFDC dataset had the lowest error rate of 0.1986 for face detection when using the original frames. Moreover, the error rate improved by 5.29% and 10.73% when degrading the frames by one-quarter and one-sixteenth, respectively.

Our method demonstrates significant performance improvements on the Celeb-DF (v2) dataset, particularly on the original frames. Specifically, we utilize the three-factor K-fold method with RBF kernel for face and texture images, resulting in the lowest error rate of 0.0175. Furthermore, even when degrading the original frames by one-quarter and one-sixteenth, the three-factor leveling method with RBF kernel still yields the optimal error rate of 0.0175 for face images, consistent with the results obtained from the original frames. We also note that the minimum error rate for texture images was raised to 0.2 for both degradation effects.

The results of our method indicate that extracting texture features from human facial images does not enhance the accuracy of our method. This may be attributed to losing crucial information to evaluate the image quality after extracting texture features.

4.4. Comparison with Other State-of-Art Methods

In our comparative analysis, we selected Xception [14] and CapsuleNet [17] as representative deep learning-based baselines. Xception is a widely recognized and powerful CNN architecture that has demonstrated strong performance in general Deepfake detection, representing the mainstream approach. CapsuleNet, on the other hand, offers a novel network structure designed to capture spatial hierarchies, providing a different deep learning perspective. By comparing our lightweight framework against these two distinct and high-performing models, our objective is to specifically evaluate its robustness and efficiency, particularly under the challenging conditions of image degradation, rather than to conduct an exhaustive survey of all available detection architectures.

To facilitate a more intuitive comparison with other methods, we convert the average error rates we have calculated into accuracy (%) using the following formula:

$$Accuracy(\%) = (1 - ErrorRate) \times 100 \quad (5)$$

To verify the effectiveness of other existing state-of-the-art models, we selected 40,000 facial images from two datasets, with an equal split between real and fake images. Among them, 80% are used for training and 20% for validation. The experimental results are shown in Table 3.

Table 3. Comparison of the accuracy of our method with existing state-of-art methods on DFDC and Celeb-DF (v2) datasets.

Methods	Dataset	Original	1/4 Downsampling	1/16 Downsampling
Ours	DFDC	80.14	79.09	78.01
	Celeb-DF	98.25	***98.25***	***98.25***
Capsule [17]	DFDC	97.27	96.6	95.65
	Celeb-DF	***98.95***	97.92	97.17
Xception [14]	DFDC	97.1	97.6	95.76
	Celeb-DF	***98.75***	98.16	97.61

Note. **Bold** and *italic* values indicate the best results within each experimental group.

The results presented in Table 3 highlight a key finding of our study. On original, high-quality images, the deep learning models—Xception and CapsuleNet—achieve excellent detection accuracy on both the DFDC and Celeb-DF (v2) datasets. However, their performance shows a noticeable decline as the image resolution is reduced through downsampling.

In sharp contrast, our NR-IQA-based method maintains a remarkably stable accuracy across all tested resolutions, particularly on the challenging Celeb-DF (v2) dataset, where it consistently achieves 98.25% accuracy. This underscores the primary strength of our approach: its robustness to quality degradation. The results suggest that while complex

models may overfit to high-frequency artifacts that are diminished in low-resolution images, our method leverages more fundamental quality-based features that are less sensitive to compression and downsampling. This characteristic makes our framework particularly well suited for real-world scenarios where Deepfake videos are often circulated in a compressed format.

4.5. The Transferability of Integrated Methods

During our experiments, we observed that most existing deep learning-based Deepfake detection models exhibit poor generalizability when evaluated under cross-dataset conditions. For example, models trained on the Celeb-DF (v2) dataset yield significantly lower accuracies when tested on the DFDC dataset, and vice versa. This indicates limited robustness when applying trained models to data from unseen domains or distributions.

To address this issue, we introduce an integrated filtering pathway to enhance model performance in cross-dataset transfer scenarios. Specifically, before classification, we apply a feature-based filtering step using the proposed NR-IQA-based module to screen input images based on predefined quality thresholds. These thresholds are computed from reference features that best separate real and fake images within the training domain.

The threshold selection process uses an NRIQA model trained across datasets at multiple resolution scales. This model is then applied to a subset of the target dataset (referred to as the interactive test set) to determine an optimal reference threshold (TH). The threshold aims to achieve 100% confidence classification for a portion of the samples—namely, high-confidence true positives (TP) and true negatives (TN). The remaining uncertain samples are subsequently passed to the main classifier (e.g., CapsuleNet or Xception) for final prediction.

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \times 100 \quad (6)$$

For example, when testing on DFDC data, the model is trained on Celeb-DF (v2), and a subset of 1212 images (606 real, 606 fake) is used for threshold selection. A similar selection is made when testing in the reverse direction (i.e., Celeb-DF (v2) as test set). The thresholds for each training dataset and resolution are summarized in Table 4, where TH(B) denotes thresholding using BRISQUE scores, and TH(P) refers to thresholds based on PIQUE scores.

Table 4. Threshold settings for different training datasets at different resolutions.

Resolution	Training Dataset	Threshold for Real	Threshold for Fake
Original	Celeb	TH(B) < 37.09337	TH(B) > 53.61025
	DFDC	TH(B) < 33.50054	TH(B) > 55.25435
1/4 Downsampling	Celeb	TH(P) < 61.01035	TH(P) > 85.0005
	DFDC	TH(P) < 50.46148	TH(P) > 89.7685
1/16 Downsampling	Celeb	TH(P) < 50.11057	TH(P) > 83.43516
	DFDC	TH(P) < 60.05618	TH(P) > 84.00795

Table 5 presents the evaluation results of the cross-dataset transfer experiments with and without the integrated filtering method. The column "Accuracy" refers to the performance of the baseline model, while "Accuracy w OUR" represents the performance after applying the integrated pipeline. The results show consistent accuracy improvements, particularly under low-resolution conditions.

Figure 5 provides a detailed illustration of the cross-dataset transfer evaluation results before and after integrating our proposed method (denoted as "Accuracy w OUR"). The horizontal axis of the figure corresponds to specific transfer testing configurations represented by the bar labels. For instance, "C_Raw_CtoD" refers to a scenario where the model

is trained on the Celeb-DF (v2) dataset and tested on the DFDC dataset using the Capsule Net method with raw-resolution images.

Table 5. Comparison of test results across datasets at different resolutions before and after thresholding.

Resolution	Method	Training	Testing	Accuracy	Accuracy w OUR
Original	Capsule Net [17]	Celeb	DFDC	61.88%	62.21%
		DFDC	Celeb	50.17%	51.08%
	Xception Net [14]	Celeb	DFDC	56.68%	56.77%
		DFDC	Celeb	51.92%	53.00%
1/4 Downsampling	Capsule Net [17]	Celeb	DFDC	63.82%	68.35%
		DFDC	Celeb	51.58%	54.17%
	Xception Net [14]	Celeb	DFDC	59.97%	60.72%
		DFDC	Celeb	51.00%	53.25%
1/16 Downsampling	Capsule Net [17]	Celeb	DFDC	64.27%	68.88%
		DFDC	Celeb	52.67%	56.83%
	Xception Net [14]	Celeb	DFDC	58.42%	65.78%
		DFDC	Celeb	48.67%	52.58%

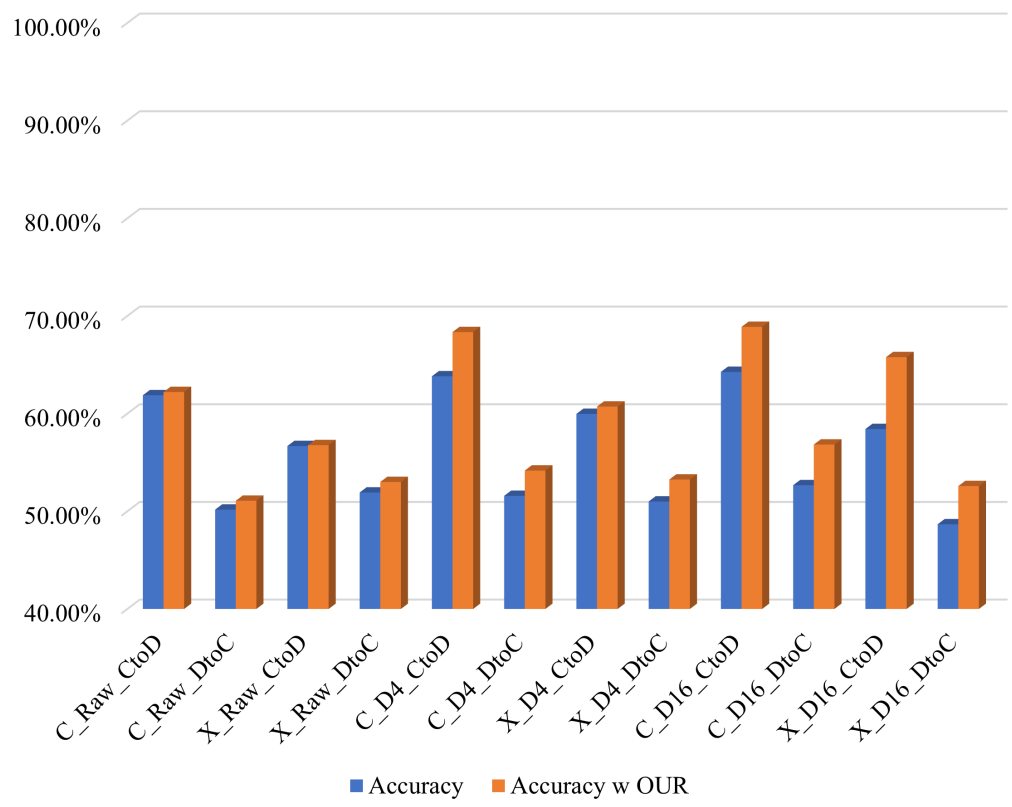


Figure 5. Comparison of test results across datasets at different resolutions with and without integrating.

In this notation, the first character (C or X) denotes the model architecture, Capsule Net and Xception, respectively. The resolution levels are marked as "Raw" for the original resolution, "D4" for images downsampled to one-quarter resolution, and "D16" for one-sixteenth resolution. The suffix "CtoD" indicates a transfer from Celeb-DF (v2) to DFDC, while "DtoC" represents the reverse direction.

The results demonstrate that, across different resolutions, the integration of our proposed method consistently improves accuracy, particularly in low-resolution scenarios where detection performance typically suffers. This highlights the robustness and generalizability of our approach under cross-dataset transfer conditions.

5. Conclusions

In this study, we proposed a lightweight and modular Deepfake detection framework that leverages No-Reference Image Quality Assessment (NR-IQA) methods to improve detection performance. Facial regions are extracted using YOLOv4, followed by a texture-focused preprocessing step to enhance discriminative features. Three NR-IQA techniques—BRISQUE, NIQE, and PIQUE—are employed to extract quality-related features, which are subsequently fed into an SVM classifier for final prediction. The effectiveness of our approach is evaluated through both intra-dataset (via K-fold cross-validation) and cross-dataset transfer testing scenarios.

To assess robustness under real-world conditions such as bandwidth-constrained video streaming, we further evaluate our model on downscaled video frames (1/4 and 1/16 of the original resolution). Our method maintains strong detection capability across resolutions, achieving low error rates where existing methods often degrade. Additionally, we introduce an integrated filtering strategy based on NR-IQA thresholds, allowing high-confidence samples to be filtered before classification. This pre-filtering step improves performance in challenging cross-dataset transfer settings, particularly where direct application of advanced deep learning models yields suboptimal accuracy.

Moreover, our method is computationally efficient, relying on the linear computation of handcrafted features rather than large-scale deep network training. This makes it well suited for scenarios requiring fast, lightweight deployment without sacrificing detection effectiveness. In summary, our proposed framework offers a practical and adaptable solution to Deepfake detection, and it can be readily integrated with existing deep learning-based approaches to enhance their generalizability, especially in cross-domain applications, without incurring significant additional computational cost.

As AI technologies such as Generative Adversarial Networks (GANs) and Large Language Models (LLMs) continue to evolve, the creation of forged media content will become increasingly sophisticated. This poses significant challenges to current detection systems. For future work, we plan to explore hybrid methods that combine spatial and temporal features, as well as multimodal detection strategies. Furthermore, we intend to extend our comparative analysis to include other state-of-the-art methods, especially recent lightweight models such as MesoNet and EfficientNet, to further validate the performance and efficiency of our framework across a wider spectrum of detection architectures. Integrating our NR-IQA-based approach into these broader, more adaptive frameworks will be a key focus of our continued research.

Author Contributions: Methodology, W.-C.Y. and J.J.; Writing—original draft preparation, J.J.; Writing—review and editing, J.J., W.-C.Y., C.-H.C., and T.Y.; Supervision, W.-C.Y. and C.-H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. The experiments were conducted using publicly available datasets, including the DeepFake Detection Challenge (DFDC) and Celeb-DF(v2), both created with informed consent and available on Kaggle. No new data involving human or animal participants were collected in this study.

Informed Consent Statement: The dataset was created by Facebook with paid actors who entered into an agreement to the use and manipulation of their likenesses in our creation of the dataset.

Data Availability Statement: These data were derived from the following resources available in the public domain: https://www.kaggle.com/competitions/deepfake-detection-challenge/data, (accessed on 10 October 2025) and https://www.kaggle.com/datasets/reubensuju/celeb-df-v2/data, (accessed on 10 October 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chesney, R.; Citron, D. Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy. *Lawfare Blog* 2018. Available online: https://www.lawfaremedia.org/article/deepfakes-looming-crisis-national-security-democracy-and-privacy (accessed on 10 October 2025).
2. FakeApp. FakeApp 2.2.0. 2019. Available online: https://www.malavida.com/en/soft/fakeapp/ (accessed on 9 March 2025).
3. FaceApp. FaceApp. 2019. Available online: https://www.faceapp.com/ (accessed on 10 October 2025).
4. Faceswap. Faceswap: Deepfakes Software for All. 2022. Available online: https://github.com/deepfakes/faceswap (accessed on 9 March 2025).
5. Perov, I.; Gao, D.; Chervoniy, N.; Liu, K.; Marangonda, S.; Umé, C.; Dpfks, M.; Facenheim, C.S.; RP, L.; Jiang, J.; et al. DeepFaceLab: Integrated, flexible and extensible face-swapping framework. *arXiv* **2020**, arXiv:2005.05535.
6. Brooks, T.; Peebles, B.; Holmes, C.; DePue, W.; Guo, Y.; Jing, L.; Schnurr, D.; Taylor, J.; Luhman, T.; Luhman, E.; et al. Video Generation Models as World Simulators. 2024; Volume 3. Available online: https://openai.com/research/video-generation-models-as-world-simulators (accessed on 10 October 2025).
7. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
8. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
9. 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.
10. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X.; Metaxas, D.N. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5907–5915.
11. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
12. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8789–8797.
13. Afchar, D.; Nozick, V.; Yamagishi, J.; Echizen, I. Mesonet: A compact facial video forgery detection network. In Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 11–13 December 2018; pp. 1–7.
14. Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Nießner, M. Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1–11.
15. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
16. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3856–3866.
17. Nguyen, H.H.; Yamagishi, J.; Echizen, I. Capsule-forensics: Using capsule networks to detect forged images and videos. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2307–2311.
18. Yang, X.; Li, Y.; Lyu, S. Exposing deep fakes using inconsistent head poses. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8261–8265.
19. Sabir, E.; Cheng, J.; Jaiswal, A.; AbdAlmageed, W.; Masi, I.; Natarajan, P. Recurrent convolutional strategies for face manipulation detection in videos. *Interfaces (GUI)* **2019**, *3*, 80–87.
20. Zhang, D.; Li, C.; Lin, F.; Zeng, D.; Ge, S. Detecting Deepfake Videos with Temporal Dropout 3DCNN. In Proceedings of the IJCAI, Montreal, QC, Canada, 19–27 August 2021; pp. 1288–1294.
21. Zhao, C.; Wang, C.; Hu, G.; Chen, H.; Liu, C.; Tang, J. ISTVT: Interpretable spatial-temporal video transformer for deepfake detection. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 1335–1348. [CrossRef]
22. Lukás, J.; Fridrich, J.; Goljan, M. Digital "bullet scratches" for images. In Proceedings of the IEEE International Conference on Image Processing 2005, Genoa, Italy, 11–14 September 2005; Volume 3, p. III-65.
23. Lukas, J.; Fridrich, J.; Goljan, M. Digital camera identification from sensor pattern noise. *IEEE Trans. Inf. Forensics Secur.* **2006**, *1*, 205–214. [CrossRef]

24. Mihcak, M.K.; Kozintsev, I.; Ramchandran, K. Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising. In Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings, ICASSP99 (Cat. No.99CH36258), Phoenix, AZ, USA, 15–19 March 1999; Volume 6, pp. 3253–3256.
25. Koopman, M.; Rodriguez, A.M.; Geradts, Z. Detection of deepfake video manipulation. In Proceedings of the 20th Irish Machine Vision and Image Processing Conference (IMVIP), Belfast, UK, 29–31 August 2018; pp. 133–136.
26. Yang, W.-C.; Jiang, J.; Chen, C.-H. A fast source camera identification and verification method based on PRNU analysis for use in video forensic investigations. *Multimed. Tools Appl.* **2021**, *80*, 6617–6638. [CrossRef]
27. Xie, H.; He, H.; Fu, B.; Sanchez, V. GrDT: Towards Robust Deepfake Detection using Geometric Representation Distribution and Texture. In Proceedings of the Winter Conference on Applications of Computer Vision, Tucson, AZ, USA, 28 February–4 March 2025; pp. 734–744.
28. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef] [PubMed]
29. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [CrossRef]
30. Venkatanath, N.; Praneeth, D.; Bh, M.C.; Channappayya, S.S.; Medasani, S.S. Blind image quality evaluation using perception based features. In Proceedings of the 2015 Twenty First National Conference on Communications (NCC), Mumbai, India, 27 February–1 March 2015; pp. 1–6.
31. Yang, W.C.; Tsai, J.C. Deepfake detection based on no-reference image quality assessment (nr-iqa). *Forensic Sci. J.* **2020**, *19*, 29–38.
32. Bouhamed, S.A.; Msaddak, Y. Image Quality Assessment for Facial Image Authenticity Verification (FIAV). In *Asian Conference on Intelligent Information and Database Systems*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 205–220.
33. Chen, Z.; Wang, J.; Liu, K.; Gong, J.; Sun, L.; Wu, Z.; Timofte, R.; Zhang, Y.; Zhang, J.; Wu, J.; et al. NTIRE 2025 challenge on real-world face restoration: Methods and results. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 1536–1547.
34. Liu, Y.; Wang, X.; Hu, E.; Wang, A.; Shiri, B.; Lin, W. VNDHR: Variational single nighttime image Dehazing for enhancing visibility in intelligent transportation systems via hybrid regularization. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 10189–10203. [CrossRef]
35. SWGDE. SWGDE Technical Notes on FFmpeg for Forensic Video Examinations (16-V-002-3.0). 2024. Available online: https://www.swgde.org/documents/published-complete-listing/16-v-002-technical-notes-on-ffmpeg-for-forensic-video-examinations/ (accessed on 9 September 2025).
36. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [CrossRef]
37. Yang, S.; Luo, P.; Loy, C.C.; Tang, X. Wider face: A face detection benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5525–5533.
38. Xia, Z.; Qiao, T.; Xu, M.; Wu, X.; Han, L.; Chen, Y. DeepFake video detection based on MesoNet with preprocessing module. *Symmetry* **2022**, *14*, 939. [CrossRef]
39. Rössler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Nießner, M. Faceforensics: A large-scale video dataset for forgery detection in human faces. *arXiv* **2018**, arXiv:1803.09179.
40. Thies, J.; Zollhofer, M.; Stamminger, M.; Theobalt, C.; Nießner, M. Face2face: Real-time face capture and reenactment of rgb videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2387–2395.
41. Thies, J.; Zollhöfer, M.; Nießner, M. Deferred neural rendering: Image synthesis using neural textures. *Acm Trans. Graph. (TOG)* **2019**, *38*, 1–12. [CrossRef]
42. Abu-El-Haija, S.; Kothari, N.; Lee, J.; Natsev, P.; Toderici, G.; Varadarajan, B.; Vijayanarasimhan, S. Youtube-8m: A large-scale video classification benchmark. *arXiv* **2016**, arXiv:1609.08675.
43. Dolhansky, B.; Howes, R.; Pflaum, B.; Baram, N.; Ferrer, C.C. The deepfake detection challenge (dfdc) preview dataset. *arXiv* **2019**, arXiv:1910.08854. [CrossRef]
44. Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; Ferrer, C.C. The deepfake detection challenge (dfdc) dataset. *arXiv* **2020**, arXiv:2006.07397. [CrossRef]
45. Li, Y.; Yang, X.; Sun, P.; Qi, H.; Lyu, S. Celeb-DF (v2): A new dataset for DeepFake Forensics. *arXiv* **2019**, arXiv:1909.12962v3.
46. Li, Y.; Yang, X.; Sun, P.; Qi, H.; Lyu, S. Celeb-df: A large-scale challenging dataset for deepfake forensics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 4–19 June 2020; pp. 3207–3216.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.