Journal of Multimedia Processing and Technologies

DLINE JOURNALS

Print ISSN: 0976-4127
Online ISSN: 0976-4135

JMPT 2025: 16 (3)
https://doi.org/10.6025/jmpt/2025/16/3/113-137

Deep Learning based Detection of AI generated Synthetic Images for Digital Forensics

Sonal Lakade[1], Shyam Khairkar[2], Praveen Kokane[3], Amol Kale[1], Rajivkumar Mente[1*]

[1]Dept. of Computer Science, PAH Solapur University
Solapur, India

[2]Directorate of Forensic Science Laboratories
Mumbai. India

[3]Department of Geography, University of Mumbai
Mumbai. India

ABSTRACT

The rapid advancement of developments in artificial intelligence (AI), particularly in Generative Adversarial Networks (GANs), has paved the way for the creation of extremely realistic synthetic images that pose a challenge for digital forensics. Traditional image authentication techniques lack the pace to catch up with the growing sophistication of AI-synthesised images, calling for more innovative detection methods. This review examines the potential application of deep learning technologies, specifically Convolutional Neural Networks (CNNs), in detecting AI-generated synthetic images. The paper discusses several conventional and deep learning based approaches, compares their performance, and indicates the significance of light-weight CNN models in maintaining computational efficiency without sacrificing accuracy. In addition, it elaborates on the implications of explainable AI in bringing transparency to detection models. The review also explores the importance of synthetic image data in computer vision tasks, as well as its challenges, including domain gaps and biases. Lastly, the application of digital forensics in preventing misinformation and malicious acts involving synthetic images is discussed. The results underscore the importance of reliable and interpretable deep learning based detection methods in preserving the integrity of digital forensic examinations.

Keywords: Deep Learning, AI-Generated Synthetic Images, Digital Forensics, Convolutional Neural Networks (*CNNs*), Generative Adversarial Networks (*GANs*), LightWeight *CNNs*, Explainable AI, Image Authentication, Computer Vision, Synthetic Image Detection

Received: 4 March 2025, Revised 8 June 2025, Accepted 26 June 2025

1. Introduction

Ever more realistic, becoming challenging to identify as authentic images. This presents major challenges for digital forensics, in which authenticating images is a critical concern for law enforcement, cybersecurity, and media integrity. Conventional methods of detection struggle to keep pace with the rapidly evolving sophistication of AI-generated imagery. Deep learning based solutions offer a promising approach by leveraging neural networks to identify patterns, textures, and anomalies that are not readily apparent to the naked eye. This research examines how deep learning methods can enhance the detection of AI generated synthetic images, thereby improving the credibility of digital forensic examinations.

With the advent of powerful generative models, such as Generative Adversarial Networks (GANs), creating synth etic images that are nearly indistinguishable from real ones has become trivial. This rapid proliferation of AI generated media introduces new challenges across multiple domains. In digital forensics, manipulated or fully synthetic visual content can compromise evidentiary integrity. In media and journalism, synthetic images can be weaponised to spread misinformation. Likewise, in cyber security, adversaries may exploit these images for purposes such as spoofing, phishing, or identity theft. The scale and sophistication of these threats are growing, making it imperative to develop reliable, interpretable, and scalable detection methods to safeguard digital trust.

1.1 Detection of AI-Generated Synthetic Images with a Lightweight CNN

Recent advances in deep learning, particularly in Generative Adversarial Networks (GANs) [1], have enabled the creation of artificial photographs that are so realistic that they often pass for genuine ones [2]. Although this development holds great promise for several sectors, it also raises serious ethical concerns due to the potential for abuse. The dissemination of misleading information, fraud, identity theft, and the creation of offensive or damaging material are just a few of the significant social effects that may result from such misuses. [3]

Establishing reliable methods to distinguish between actual and artificial intelligence generated information is crucial as deep learning advances at an unprecedented rate and AI-generated visuals become increasingly realistic every day. This is necessary to mitigate the negative consequences and preserve the reliability and accuracy of information found online.

There are now two primary methods for identifying phone photographs, apart from eye examination, which is an inaccurate technique. Among them are deep learning and image processing methods for manually created feature extraction. Identifying a particular tampering

Techniques, such as splicing or copy move, were the goal of early algorithms for identifying fraudulent images caused by image tampering [4]. These methods of interference usually entail modifying specific areas of a picture to add or change elements within it. The majority of early fake detection techniques relied on frequency domain feature extraction. For instance, the method suggested in [5] splits the picture into overlapping blocks and uses the discrete cosine transformation to match the characteristics extracted from these blocks, thereby identifying copy move forgeries. In a similar vein, the discrete wavelet transformation is used in [6] to extract low frequency components for feature extraction from the frequency domain. These components are simultaneously subjected to singular value decomposition in order to produce the feature vectors.

Nevertheless, this approach is laborious and susceptible to blurred, scaled, or twisted picture objects. The same author suggested using the Fourier Mellin transformation in [7] to increase the effectiveness of the earlier approach. While the Bloom filter speeds up the whole detection process, transformation is employed to reduce sensitivity to geometric operations. However, the aforementioned techniques do not perform well when used for picture splicing, which involves combining segments from multiple sources, each with unique textures and characteristics. The color filter array pattern is extracted from the picture using the technique suggested in [8]. To distinguish between real and phone areas, a statistical examination of local discrepancies within these patterns is then carried out. In order to detect the phony pictures, the classification model is used in [9] to detect anomalies that are created by splicing in photos. Since splicing often alters the original image's frequency patterns, feature extraction is accomplished through a discrete cosine transformation, while Markov features are generated from the transform coefficients retrieved.

However, as advanced *GANs* have been developed, images are frequently manipulated using multiple tampering techniques simultaneously, producing more realistic looking images that are difficult to detect as manipulated. This makes it difficult to determine the exact areas of the image that have been tampered with, as well as the type of tampering. As a result, the formerly successful techniques that relied only on feature extraction are no longer able to precisely identify these alterations

Deep learning techniques, such as convolutional neural networks (*CNNs*), can aid in this process. Convolutional neural networks are based on a variety of deep artificial neural network topologies that include several hidden layers of neurons after a convolutional and a pooling layer. Higher level characteristics, like pictures, are gradually extracted from the raw input by these layers. These techniques may approximate more complex decision functions and achieve higher classification accuracy by increasing the number of layers [11]. However, high-Performing *CNNs* nowadays, such as *VGG*Net [12], DenseNet [13], and ResNet [14], involves many layers, which raises the requirement for a large amount of data and increases the difficulty of the training process due to the presence of numerous local optima and numerous hyperparameters. They are also regarded as black-box function approximates, meaning that their judgments cannot be explained [15]. ResNet and Dense Net are more tha n 100 layers deep, but *VGG*Net has 16 layers, including 3 fully connected layers with thousands of neurons and 13 convolutional layers. A lightweight *CNN* for picture fraud detection was presented in [16]. It consists of 17 bottlene ck layers, each comprising 1 × 1 and 3 × 3 convolutions, following a typical 3 × 3 convolutional layer. Higher computing efficiency is achieved by employing tiny kernels, such as 1x1, as opposed to conventional convolutions that operate over large kernels. To identify picture forgeries, a lightweight *CNN* with three convolutional layers each with 32, 64, and 128 filters, respectively and a 3 × 3 kernel size was presented in [17]. However, due to their limited number of layers and kernel sizes, existing lightweight approaches are unable to extract high quality features from input images in comparison to deep *CNNs*, which prevents them from achieving good classificati on performance.

1.2 Synthetic Image Generation

A generator and a discriminator are two separate neural networks that make up the fundamental concept of *GANs*. The generator's job is to use the supplied input data to create new data instances, such as pictures. The discriminator's job is to evaluate these occurrences in the interim to ascertain whether the generator created them or not. In essence, this establishes a contest in which the generator attempts to deceive the discriminator by generating data that is indistinguishable from real data. On the other hand, the discriminator's goal is to precisely distinguish between data generated by the generator and real data. During training, the discriminator is fed both synthetic and actual pictures, and a loss value is then calculated. StyleGAN, first introduced in [18], is among the most widely used GAN structures. An affine transformation is performed after the eight fully linked layers of the mapping network have processed the input characteristics. The goal of this mapping is to provide features that enable the generator to render data efficiently while avoiding feature combinations that are not present in the training dataset. By scaling, rotating, translating, mirroring, and shearing picture objects while maintaining their original connections, an affine transformation simultaneously makes it possible to regulate the creation of images. After adding the random noise, the output of this mapping is further processed by the convolutional layers and adaptive instance normalisation (AdaIN). picture production in this process begins with a low quality picture and progressively adds additional layers to add details and raise the resolution (for example, from 4 × 4 to 1024 × 1024). Blood like artefacts and poor shift invariance, which manifest as an odd alignment of certain items in the picture, are two issues with the images produced by the original StyleGAN version. The enhanced StyleGANv2 architecture, as presented in [19], overcomes this limitation. After noticing that the AdaIN normalisation was linked to the blob artefacts, the authors implemented a weight demodulation procedure that is applied directly to the convolution layer weights. Conversely, progressive expansion gave rise to the problem of shift invariance. StyleGANv2 learns at the ultimate resolution from the start, eliminating the need for incremental development. StyleGANv3 [20] is an additional enhancement of the StyleGAN architecture that incorporates Fourier features and an affine transformation layer to these features, enabling more effective translation and rotation. It also makes other minor adjustments to the architecture, such as introducing new up sampling and down sampling filters, to enhance the quality of the generated images. However, StyleGANv3's superior quality has a cost. Because it struggles with mode collapse, it requires more time, more computing power, and a larger dataset to obtain findings that are adequate. Due to these phenomena, the generator only generates a limited range of pictures, which reduces the diversity of the results. By penalising the generator when it produces images with comparable characteristics, the authors of [21] proposed a module that can be applied to various GAN designs to mitigate mode collapse. Altering the discriminator's behavior [22] using numerous generators [23]are two other tactics that deal with the mode collapse problem.

1.3 Emerging Generative Models for Synthetic Image Creation

While Generative Adversarial Networks (*GANs*) have long dominated the field of synthetic image generation, recent advances have introduced new generative architectures, including Diffusion Models and Variational Autoen coders (*VAEs*). These models have demonstrated impressive capabilities in generating high fidelity images, often surpassing *GANs* in terms of stability, quality, and control. This section highlights the working principles, benefits, and limitations of these emerging models.

1.3.1 Variational Autoencoders

By learning the latent distribution of the data, *VAE* seeks to produce fresh samples. Its suggestion signals the start of deep learning's major advancement in the image production sector. Improvements in Autoencoders, which integrate the idea of variational inference to enhance the model's expressive and probabilistic modelling capabilities, have enabled the development of *VAE*.

The encoder decoder structure is the foundation of *VAEs'* operation; the encoder maps input data to a latent space, and these latent variables are subsequently mapped back to the data space by the decoder. VAEs make the assumption that latent variables adhere to a specific prior distribution, typically the conventional normal distribution, in order to incorporate probabilistic features. The model learns the latent representation of the data and how to produce samples by optimising the Evidence Lower Bound (ELBO) of the observed data. Figure 2

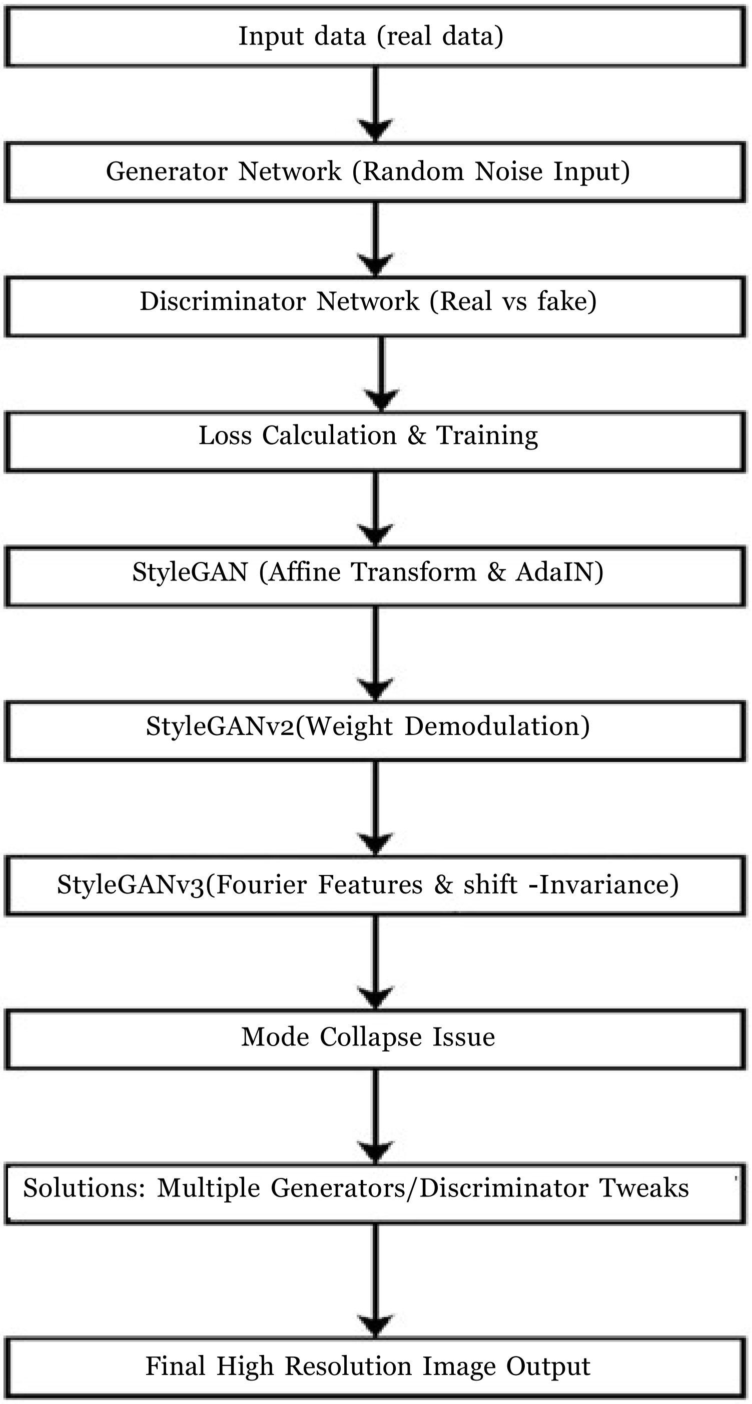
Input data (real data)
Generator Network (Random Noise Input)
Discriminator Network (Real vs fake)
Loss Calculation & Training
StyleGAN (Affine Transform & AdaIN)
StyleGANv2(Weight Demodulation)
StyleGANv3(Fourier Features & shift -Invariance)
Mode Collapse Issue
Solutions: Multiple Generators/Discriminator Tweaks
Final High Resolution Image Output

Figure 1. Flowchart illustrating the evolution of StyleGAN architectures and their improvements, highlighting key components such as affine transformations, adaptive instance normalisation (AdaIN), weight demodulation, Fourier features, and solutions for mode collapse

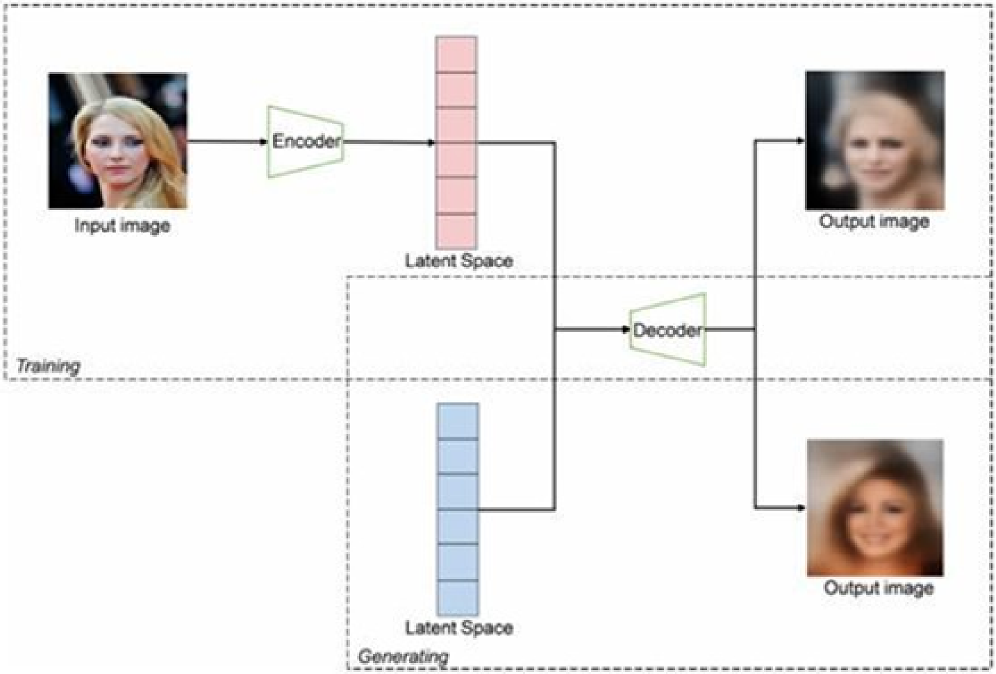

Figure 2. The model structure of *VAE* consists of an encoder, which maps the image to the latent space, and a decoder, which recovers the image from the latent space to the pixel space

Due to the probabilistic modelling of the latent space, *VAE* can generate high quality samples that are comparable to the training data and provide samples with a certain level of continuity. Due to this, the *VAE* performs well on image generation tasks. Perfect reconstruction is not guaranteed, and the reconstructed image produced by *VAE* may be blurrier than the input image due to the probabilistic reconstruction process.

Sohn et al [24] suggested the Conditional Variational Autoencoder (*CVAE*) to regulate the generated material. To control the generated samples under specific conditions, *CVAE*, an extension of the basic *VAE*, incorporates condition information. Its fundamental concept is to take conditional information into account when sampling latent variables, which enables the model to provide samples with particular attributes in response to more data. With significant improvements in tasks such as picture restoration and style transfer, the model is now better suited for generation tasks that require specific conditions. The align Draw concept, the first contemporary text to image model, was presented by Mansimov et al [25]. It can utilise text sequences as input, as it employs a recurrent variational auto encoder with an attention mechanism. Sentences that are not in the training set can be used by this model to create visuals. The vector quantised variational autoencoder (*VQVAE*) was proposed by Van Den Oord et al. [26]. Its design aims to overcome some of the shortcomings of conventional autoencoders, such as the lack of explicit structure and the ambiguity of latent representations, while maintaining the benefits of auto encoders. The fundamental concept of VQ*VAE* is to use a discrete codebook to approximate the latent representation by discretising the continuous encodings in the latent space. A collection of distinct vectors, known as “codebook vectors,”comprises this codebook. The model gains the ability to map the latent representation to the closest codebook vector during training. Huang et al. [27] introduced IntroVAE, which incorporates concepts from *GANs*, enabling it to self evaluate the quality of generated samples and subsequently improve itself.

This nearest vector is mapped back to a continuous latent representation during the generation process. Daniel et al. [28] developed Soft IntroVAE, which is based on IntroVAE and greatly improves training stability by substituting a smooth exponential loss for the hinge-loss term for generated samples.

1.3.2 Diffusion Models

The application of diffusion models to deep learning was initially suggested by Sohl Dickstein et al. Jonathan Hoetal. [29] developed the Denoising Diffusion Probabilistic Model (DDPM) based on this. This model's main concept is to provide high quality data samples by simulating the diffusion and restoration processes in physical systems. By adding noise to pre existing photos and then progressively eliminating it, realistic image production is created. Diffusion models have become the standard model in the field of picture generation, replacing GANs with the release of DDPM.

DDPM's relatively slow sampling speed is one of its drawbacks. Many similar modifications have been proposed to remedy this issue. Improved *DDPM* was proposed by Nichol et al. [30], who added cosinenoise in the forward process and learnable variance in the reverse process. Better likelihood estimates and faster sampling are features of the enhanced DDPM. With the introduction of *DDIM*, Songetal. [31] Substituted a non Markovian process for the original Markov process. With no effect on the quality of the samples produced, *DDIM's* effective sampling strategy significantly increases sampling speed. Zhang et al. [32] developed *DDIM*, an extension of *DDIM* that covers a wider variety of diffusion models. They discovered that the associateds tochastic differential equations may be solved using specific fractional approximations to provide *DDIM*. In terms of quick sampling, the authors also discussed the benefits of deterministic sampling techniques over random sampling techniques. Gaussian mixtures and gamma distributions are two non Gaussian noise distributions that Nach Mani et al. [33] employed in the diffusion process. In terms of both image and audio quality and speed of generation, it outperforms the conventional Gaussian based diffusion approach while retaining the ability to sample any state during the diffusion process. Lu et al. [34] presented *DPM* Solver, a fast ordinary differential equation (*ODE*) solver for Diffusion Probabilistic Models (*DPMs*), aiming to reduce the number of sample steps and thereby increase speed. Both discrete time and continuous time DPMs can benefit from *DPM* Solver, which significantly increases sampling speed while preserving high quality sampling generation. Building upon this framework, the authors presented a higher order solver for *DPM* guided sampling called DPM Solver ++ + + [35]. It can converge in roughly 15 to 20 steps for both pixel space and latent space *DPMs*, producing high-fidelity samples. A novel dynamic programming technique was presented by Watson et al. [36] to determine the best inference schedule for DDPMs that have already been trained. Based on ELBO, the algorithm could determine the best inference schedule, which significantly lowers the number of steps needed to produce high fidelity samples while preserving good sample quality. The research team then developed the Differentiable Diffusion Sampler Search (DDSS), which uses a differential sample quality score to optimise the quick sampler of any diffusion model that has already been trained.

2. Explainable AI

Digital forensics is just one of several industries that have undergone significant changes as a result of artificial intelligence. Although artificial intelligence (AI) has significantly enhanced the effectiveness and scope of forensic investigation, it has also enabled the creation of incredibly lifelike synthetic media, known as "deepfakes," which pose a threat to conventional verification and authentication techniques [37]. Deepfakes utilise artificial intelligence (AI) methods, such as Generative Adversarial Networks (GANs), to create incredibly lifelike audio, video, and image recordings that can deceive even the most discerning viewers [38].

Digital forensics is at a crucial juncture because AI is both a threat and a solution [39]. The gathering, examination, and archiving of digital evidence for use in court cases is known as digital forensics [40]. Synthetic media has emerged as a new threat to the discipline, which has historically concentrated on recovering and evaluating data from devices. Political scandals, corporate disinformation, and personal slander have already resulted from deep fakes, and their ease of production indicates that their use will only increase [41].

Advanced computational methods based on artificial intelligence and deep learning are used to create synthetic media [42]. Generative Adversarial Networks (GANs), which operate by pitting two neural networks against each other, play a crucial role in this process [43]. While the discriminator network assesses the legitimacy of the synthetic content created by the generator network, iterative improvements are made until the generated output is nearly identical to genuine data [44].

Synthetic media is produced using a variety of techniques in addition to GANs [45]. Recurrent neural networks (RNNs) and transformer based architectures are utilised to create deepfake audio and text material, while autoencoders compress and reconstruct data to enable voice cloning and face swapping. For example, image to image translation networks can animate still photos or substitute faces in films with startling realism, and AI models can synthesise a person's voice by analysing hours of speech [46].

There are several sorts of synthetic media, including identity swapping, lip syncing, audio synthesis, and facial reenactment [47]. Each method utilises a distinct set of training datasets and algorithms. Pose estimation and face tracking are used in facial reenactment to alter facial expressions in recordings [48]. Identity switching uses frame by frame mapping to swap the faces of two people. While audio synthesis utilises waveform modelling to simulate vocal patterns, lip syncing aligns new audio with a subject's lip motions [49].

Developing reliable detection systems requires an understanding of these mechanisms [50]. Artificial intelligence (AI) systems can learn to identify the unique digital artefacts left by each synthetic media technique, which include small variations in pixelation, illumination, frame rate, or audio characteristics. However, these artefacts are getting more difficult to detect as generating techniques advance.

2.1 Digital Forensics

According to [51], any digital gadget, including smart phones, tablets, laptops, and desktop computers, may be used for illegal purposes, including fraud, drug trafficking, murder, hacking, forgery, terrorism, etc. Digital forensics is used to aid investigate cybercrimes and to identify the device-assisted crime and its perpetrators in order to combat these illegal activities Although there are numerous definitions of digital forensics, the National Cybersecurity and Communications Integration Center (*NCCIC*) states that "Digital forensics is the discipline that combines elements of law and computer science to collect and analyze data from computer systems, networks, wireless communications, and storage devices in a way that is admissible as evidence in a court of law."

2.1.1 Principles of Digital Forensics

Since software and hardware are always evolving, digital forensics is a relatively new scientific process that requires ongoing study, description, and investigation of its occurrences. Figure 1 illustrates the scientific method using a flowchart [52].

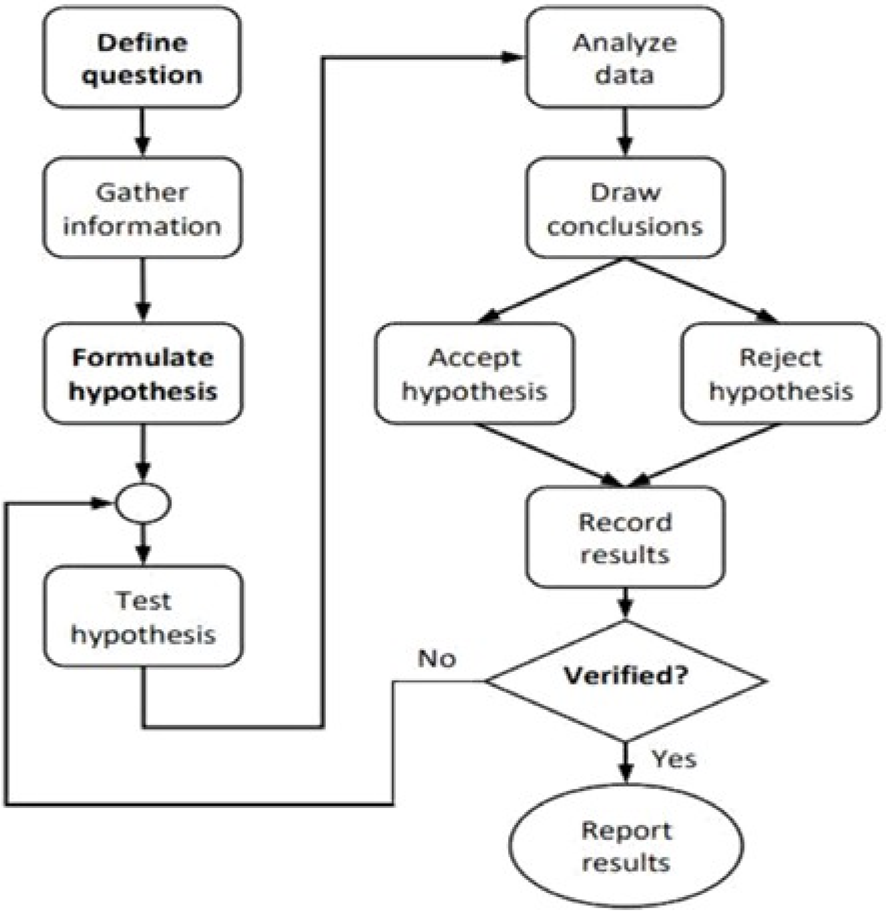

Figure 3. The scientific process of digital forensics

The three primary steps of the scientific method in digital forensics are question definition, hypothesis formulation, and outcome verification.

Clearly defining the query or what we are searching for on that device is the first step. Formulating the hypothe -sis, or determining the purpose of the gadget, is the second step. Verification of the findings or testing them in a controlled setting is the third step.

2.1.2 Characteristics of Digital Forensics

According to the U.S. Department of Justice (https://www.ncjrs.gov), digital forensics may be broken down into five steps:

a) Policy and procedure

b) Evidence evaluation

c) Evidence collection

d) Evidence inspection

e) Evidence documentation and reporting.

• Policy and Procedure

A forensic unit must establish and follow sound rules and processes to ensure accuracy and efficiency. These protocols should outline the precise steps that the forensic unit must follow when conducting a digital forensic investigation. Identification of the issue, potential solutions, testing of the solutions on a control sample, assessment of the outcomes, and method validation must all be part of the procedure creation process.

• Evaluation of the Evidence

To determine what steps to take and how to proceed, digital evidence should be examined from the standpoint of the case. Prioritising, documenting, storing, packing, and transporting the evidence are all essential steps in the evidence evaluation process.

• Obtaining Evidence

Digital evidence is very delicate, and improper handling or examination might alter, contaminate, or even destroy it. To protect and preserve the evidence, the following actions must be followed throughout the evidence collection process: Protecting digital evidence requires following the policies of the forensic unit; disassembling crime computer cases to obtain storage devices; identifying and documenting storage devices that need to be purchased; documenting hardware configuration; and disconnecting storage devices to avoid data damage or alteration.

- A controlled boot is required to get configuration information from the boot procedure.

- To complete the acquisition, storage devices from the suspect's computer must be removed.

- It is necessary to look into geometry storage devices to make sure that all available space, including host-protected data regions, is used.

- It is necessary to get digital evidence with the right technology and software.

- Verification of the acquisition requires comparing the original and the clone sector by sector.

Analysis of the Evidence

Digital examination, which includes data extraction and analysis, shouldn't be done on original evidence. Data recovery from the suspect's media is known as extraction, and the interpretation of the data that has been retrieved is known as analysis. Physical and logical extraction are the two categories. Data from the full hard disk may be recovered by physical extraction. Data is recovered using logical extraction according to the installed program, file system, and operating system.

Computer forensic experts must do the evidence examination, which consists of the following steps:

- It is necessary to set up a working directory on an independent medium in order to extract and/or restore evidence based files and data.

- It is necessary to physically remove the data from the disc using the following techniques: file carving, keyword searching, and removing wasted space from the physical drive.

- Investigating active files, deleted files, file slack, and unallocated file space requires a logical extraction of the data from the disc based on the drive's file system.

- To ascertain the data's relevance to the case, analysis is required. The study includes ownership and possession analysis, application and file analysis, data concealing analysis, and timeline analysis.

Ensuring that every outcome of the extraction and analysis procedures is fully considered is the final stage in the inspection process.

3. Synthetic Image Data and its use in Computer Vision

Although it is possible to train some computer vision models using unlabeled data through unsupervised learning, the resulting performance is usually inferior to that achieved through supervised learning, and in some applications, it can fail to produce a model with meaningful performance. Modern approaches to computer vision primarily revolve around the use of convolutional neural networks (*CNNs*) and deep learning networks to train image processing models [53]. These methods require large amounts of labelled data and significant computational resources for training. Computer vision cannot be used in situations where it is hard to gather huge volumes of data, labelling data is expensive, or both are required due to the need for large quantities of labelled data. Large scale data gathering is challenging for medical applications. Manually annotating crowd counts remains labour intensive, and specialised applications, such as drone vital sign detection, suffer from the same issues. Ensuring that the data gathered is diverse and of high enough quality to train a reliable computer vision model to the application's performance level is even more challenging. Although there are many different kinds of data used in computer.

Vision, this review paper primarily focuses on assessing the use of camera like image data and techniques for creating such data synthetically. Due to the challenges associated with acquiring image data, there has been a growing interest in synthetic image data as a potentially less expensive and more accessible alternative to real data for training. Therefore, the artificial creation of data types used in computer vision, including lidar point clouds, radar scans, and sonar scans, is not taken into consideration.

According to this assessment, any picture data that is either taken from synthetic surroundings or intentionally produced by altering genuine image data is considered synthetic image data. There are several ways to do this, such as digitally altering actual data [54] and capturing images from artificial virtual worlds. Additionally, the visual appearance of synthetic image data varies depending on the application, ranging from photo rea listic computer-generated images [55] to composite imaging [56].

A well configured image synthesis pipeline can automate the data production and labelling process at a relatively low cost compared to human labour, which is the primary benefit of synthesising image data and a key factor in making data generation quicker and less expensive. However, it is crucial to remember that not all image synthesis techniques provide automatic data labelling, and some still need a substantial amount of human labour before the synthesis process can be automated. When referring to synthetic picture data, the adjectives “faster” and “cheaper” are primarily used in contrast to the gathering of actual data for a particular application. Large actual datasets may be difficult to get for many computer vision applications, and the cost of the human labour needed to label and annotate the raw data to the necessary quality standard is high. The quantity of data that can be gathered for training purposes is limited by practical constraints, legal restrictions, and privacy concerns, which can all impact large scale data collection. Additionally, manually annotating data is a time consuming process and is not practicable for large datasets or those requiring numerous labels per data instance, such as crowd counting [57]. While it is possible to train an algorithm to automatically label data in place of a human, doing so requires a large amount of labelled data for training in the first place, creating a somewhat circular problem. Even if manual labelling of large datasets is possible, human error will inevitably lead to a decrease in data quality as the dataset size increases, which will have a negative impact on training [58].

In addition to picture data that would be difficult to get in the actual world, synthesising image data gives the resultant dataset far more control, enabling improved labelling accuracy as well as object and environment information that would otherwise be challenging to gather.

Of course, the limitations of synthetic picture data must also be considered.

Although synthetic data may not be directly related to real objects, environments, or people, real data is frequently still required to serve as a reference for creating synthetic data, and biases in the real data will affect the synthesised output. Data bias is a known problem with synthesised image data, frequently brought on by inherent bias in the input parameters for data generation. Another problem often identified in studies is the domain gap, which is a critical challenge in training detection models with synthetic image data. This term refers to the distributional differences between synthetic images (used during training) and real world images (used during inference or deployment). These differences may include lighting conditions, texture quality, object boundaries, noise, or other visual characteristics. Even photorealistic synthetic images often lack the nuanced complexity of real-world scenes, which causes models trained only on synthetic data to generalise poorly to real images. This phenomenon is well documented in the literature as a key limitation of synthetic data in vision tasks [59]. In the context of digital forensics, this issue is particularly serious, as real world manipulated images are rarely as clean or stylised as synthetic datasets. To address this, recent research has proposed domain adaptation and domain randomisation techniques to bridge the gap and improve generalisation performance [60].

Lastly, the computing costs associated with creating synthetic picture data are a restriction that is seldom mentioned. While many papers have discussed the benefits and limitations of using synthetic data in computer vision, there has been relatively little written about the computational resources and time required to generate the datasets used. This is because large synthetic data sets require a significant amount of computational power to generate within reasonable time frames, and some people may not have access to this computational power. However, these two issues may significantly impact the viability of employing synthetic data in several applications; therefore, this remains a crucial topic of debate.

The use of synthetic data for computer vision has previously been reviewed, with an emphasis on the image generation process [61] or on the application of synthetic image data to specific tasks, such as text to image synthesis [62], pedestrian detection [63], and navigating urban traffic environments. This review paper's objectives are to classify the different kinds of synthetic image data that is currently available by output, examine the techniques used to synthesise such data, talk about how well synthetic data performs in different computer vision tasks, discuss logical extensions to the use of synthetic data today, and pinpoint research gaps that could inspire further study. The challenges and related expenses of acquiring authentic data for computer vision applications are discussed in Section 2, along with the causes of the rise of synthetic picture data.

4. Synthetic Composite Imagery

Real picture data that has been digitally altered or enhanced to include components not present in the original image data is referred to as synthetic composite imagery. This involves splicing together multiple genuine photographs to create a new image, adding artificial elements to the image, and digitally altering the visual surroundings. Figure 4 illustrates how 3D synthetic items or people are projected onto actual backdrop surroundings to create synthetic composite datasets, such as SURREAL [64]. This data type is often used when the background environment has sufficiently important or helpful qualities that the effort required to artificially reproduce the environment or the loss of domain shift is not justified. The main purpose of the SURREAL dataset was to train networks on human component segmentation and depth estimation. Consequently, the artificial humans fail to consider the context in which they are situated. The backdrop serves as a means of bridging the domain gap to actual data by enhancing environmental variety, even if the generated images are readily identifiable as synthetic to the human eye. However, the network must learn properties that are related to the human object.

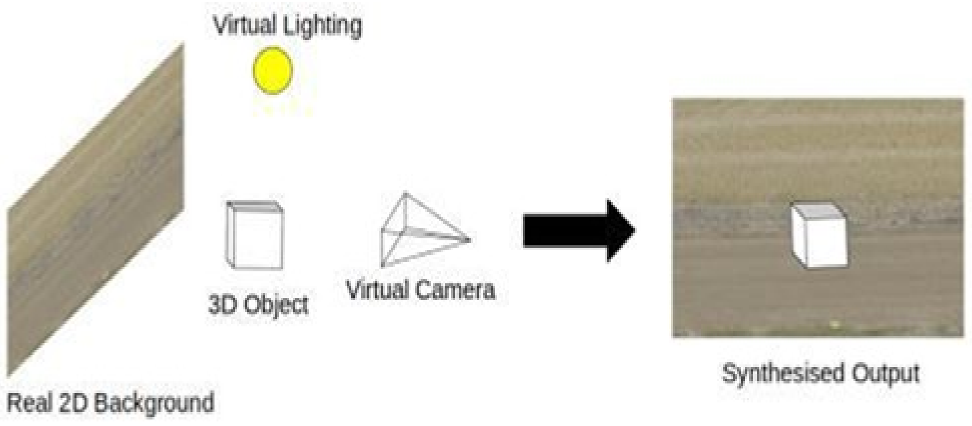

Figure 4. Synthesis via projection of a synthetic 3D object onto a real 2D background

In a similar vein, the Rare Planes dataset [64] provides synthetic composite satellite imagery of planes at various airport locations. However, Figure 5 shows that 2D pictures are immediately superimposed onto backdrops rather than 3D objects being projected onto them. Due to the nature of the image data required, satellite imagery is one of the many computer vision domains where obtaining sizable datasets is challenging. The authors of the research observe that there are no broadly permissively licensed synthetic data sets for such data. While

Figure 2 mentions the use of real 2D backgrounds, in reality, this can also be extended to synthetic 2D backgrounds, which have no bearing on the process of super imposing 2D images on to a background. The Rare Planes dataset comprises a combination of real and synthetic satellite imagery, with aerial images of planes superimposed on top. The AI. The Reverie platform, which uses the Unreal engine to produce realistic synthetic data based on actual airports, was used to generate the synthetic data. Annotating large crowds of data both photos and videos with a high population, in some situations exceeding 1,000 people, requires a significant amount of resources. Additionally, people in crowds are often not totally visible; it's possible that just a portion of their head is visible, with the rest of their body hidden by the surrounding environment. Because manual annotation may be challenging, there may be instances when data is not completely labelled, which introduces bias into the data collection. Two popular techniques exist for synthesising crowd data. The first method involves using 3D human models and either placing them in a 3D virtual environment or projecting them onto a 2D backdrop. In actuality, rendering scenes with over 1,000 models would be incredibly computationally demanding; however, if video footage is required, this remains the simplest means of producing a crowd.

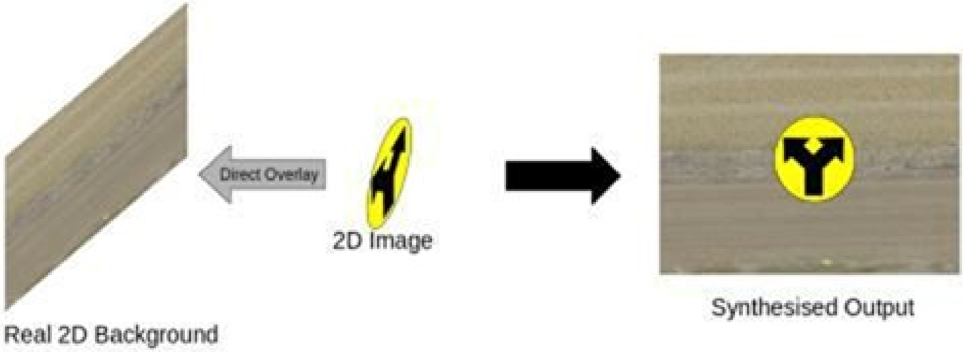

Figure 5. Synthesis via overlaying a 2D image on to a real 2D background

Real data serves as the foundation for datasets, such as foggy scenery, which are then digitally altered to create artificial differences. Such data is produced for applications where certain environmental constraints make data collection challenging, but real world settings and items are still valuable enough that it is not worth the trouble to digitally recreate the complete scenario in order to provide the required data. In actuality, this image synthesis technique may be thought of as an extension of superimposing 2D pictures on a backdrop; however, a filter is used to project the necessary environmental circumstances in place of an image. If necessary, filters are also relatively easier to apply to video data than to 2D pictures.

Some synthetic composites are made without the use of any artificial objects or photographs, even though all synthetic composites are, by definition, image composites. Similar to 2D image overlays, picture compositing involves using tagged 2D objects from one dataset and inserting them into scenes from other datasets. Compared to virtual synthetic data sets, this data synthesis approach often produces data sets with a smaller domain gap; this might be because domain randomisation improves generalisation and increases data variety.

For example, the fish identification data set [65] utilises real fish examples cropped from data gathered using the Deep Vision system [66], which are then placed in random orientations, positions, and sizes on Deep Vision footage backgrounds devoid of any other fish or objects. Since the precise picture was not captured in the actual world, the resulting composite image is still regarded as synthetic data even if it solely contains genuine data. The primary cause of this type of data is the challenge of annotating existing Deep Vision data. Extracting fish from scenarios where the species can be easily recognised by a human takes significantly less time than labelling the original Deep Vision data set by hand. Creating synthetic data with known fish species enables considerably cheaper labelled data.

An extreme kind of picture compositing, image synthesis creates a new item by combining the properties of labelled objects with those of other labelled objects, rather than taking tagged objects and inserting them into other scenarios. From the standpoint of a neural network, the synthesised item still has all the characteristics required to recognise what the object is, even if it may not visually resemble the objects from which the features were taken.

The purpose of the *KITTI*-360 dataset [67] was to enhance training data efficiency by augmenting the *KITTI* dataset [68] with additional items. The study highlighted that while 3D rendered virtual worlds are increasingly being used to generate data on urban environments, creating such an environment requires a significant amount of human intervention before data can be generated automatically. Rather, the study suggested a method for photo realistic integration of artificial elements into natural settings. By producing *KITTI*-360, we were able to insert top notch car models into pre existing *KITTI* scenarios with realistic lighting, thanks to 360 degree environment maps. As the view point on the item varies during the film, the models are created by projecting 2D texture images on to 3D meshes (Figure 6), which are then projected onto backdrops to provide a realistic depiction of the object.

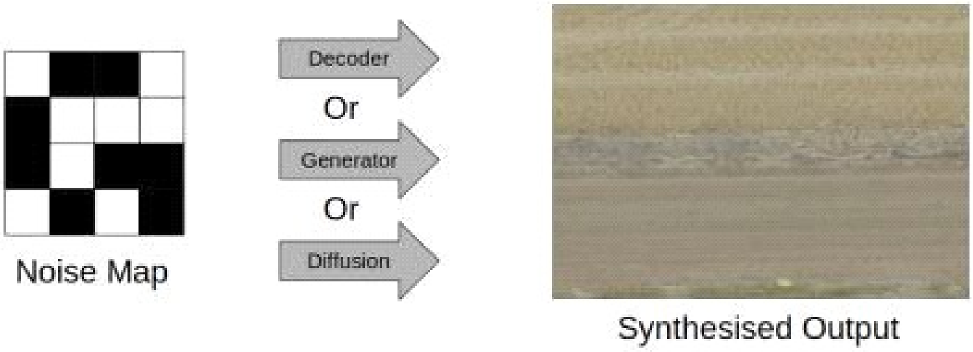

Figure 6. Synthesis via projection of a real 2D image onto a 3D mesh

By projecting actual 2D backgrounds to provide a complete 3D backdrop, the SafeUAV synthetic dataset is a more uncommon extension of mesh projection [28]. Figure 6 shows how Safe UAV reconstructs an urban area using a 3D model in City Engine and then superimposes actual picture data on top of the mesh. Since this dataset was created for semantic segmentation and depth perception tasks from a drone, the outcome ends up greatly distorting the picture data from ground perspectives, but offers a very comparable view from above, which is all that is needed.

The last kind of artificial composite images is more of an extreme extension of digital manipulation than it is an image composite. Noise maps are used as inputs by diffusion models, generative adversarial networks, and variational autoencoders to create synthesised images (Figure 7). These models can employ noise maps to extrapolate features into a whole picture by learning compressed representations of images.

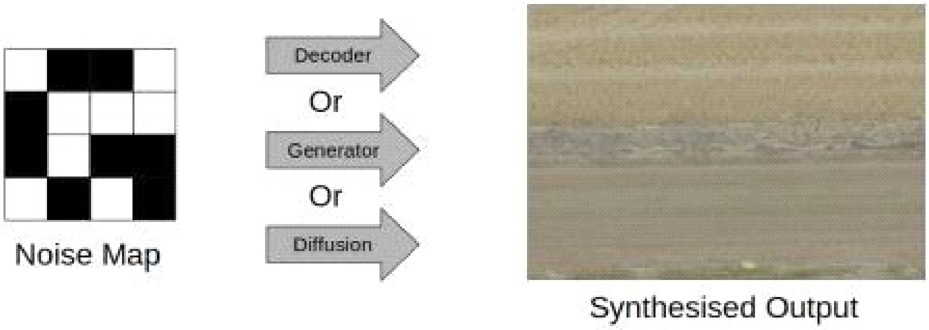

Figure 1.7 Synthesis via processing of noise maps

5. Transfer and Few-Shot Learning Techniques

One of the key obstacles in the field of synthetic picture identification is still the scarcity of labelled training datasets that cover the variety of alteration techniques found in actual forensic cases. Transfer learning has been a popular strategy to lessen this difficulty. Using pre trained deep learning models (such as *VGG*Net, ResNet, or Mobile Net) on massive datasets like ImageNet, it then refines them on datasets of synthetic media tailored to a specific goal. This procedure accelerates model convergence for forensic applications and significantly reduces the need for large amounts of labelled data [69].

Few shot learning, which tries to develop models that can generalise from a small number of labelled samples, is a complementary technique to transfer learning. In identifying new types of synthetic media that are underrepresented in training data, few shot learning has demonstrated promise. The model's robustness in dynamic threat contexts is increased by its ability to quickly adapt to novel manipulations through the use of meta learning approaches and episodic training methodologies.

The domain gap between artificial data used for training and actual altered data encountered during deployment is another pressing issue in this field. The distributional shift between the source (synthetic) and target (actual) domains is minimised via domain adaptation strategies. To align feature distributions and enhance generalisation, techniques such as style transfer, moment matching, and domain adversarial training have been investigated [59]. This is particularly crucial in digital forensics, as synthetic datasets often fall short in real-world detection settings because they fail to accurately represent the intricacies of modified real world images.

Therefore, it is essential to combine domain adaptation, transfer learning, and few shot learning for creating scalable, flexible, and data efficient models for AI-generated picture identification. These methods guarantee that forensic instruments continue to function well despite changing modification strategies and a lack of labelled datasets.

5.1 Broader Synthetic Media Detection: Video, Audio, and Text

As generative artificial intelligence continues to advance, synthetic media has expanded to include text, audio, and video content in addition to static images. Due to the fact that cyber deception today uses multimodal and extremely realistic modalities, these advancements create new challenges for digital forensics. the use of sophisticated identity swapping and facial reenactment techniques, deep fake videos are becoming more and more similar to authentic material. Media credibility, political discourse, and surveillance analysis are all seriously threatened by these manipulations, which utilise models like Face2 Face or Deep Video Portraits that alter facial emotions and synchronise lip movements with external audio streams in real time [70], [71].

Models like Taco Tron, Wave Net, and Voice Loop have made it possible to create incredibly realistic speech from small voice samples in the field of audio synthesis. Voice cloning is a serious issue for fraud detection, forensic phonetics, and authentication systems since these models employ autoregressive deep learning frameworks that may mimic vocal tone, pitch, and emotion [72], [73]. Voice based identity spoofing has already been used in financial scams and social engineering attempts.

Strong, large language models, such as *GPT*, *BERT*, and *T*5, are driving the growth of synthetic textual material. These models can generate news stories, essays, and social media content that is both grammatically correct and contextually coherent. On a global scale, these models are being exploited to disseminate automated propaganda, misinformation, and fake news [74]. Since these algorithms can pick up domain specific terminology, style, and tone that are frequently indistinguishable from human writing, detection becomes difficult.

Digital forensics systems must transition from unimodal to multimodal detection techniques in response to this increasingly complex threat scenario. Multimodal architectures are starting to show promise as solutions, such as audiovisual synchronisation networks and *CLIP* (Contrastive Language Image Pretraining). These models identify cross modal discrepancies that are frequently overlooked in unimodal systems, such as misaligned image text pairs or mismatched lip movement and speech [75].

As a result, creating comprehensive detection frameworks that combine text, audio, and video analysis is not only beneficial but also necessary. Semantic validity across formats, consistency of speaker identity, and spatiotemporal coherence must all be considered by these systems. Furthermore, in applications such as media verification, cybersecurity, and legal evidence analysis, real time performance and forensic explainability remain crucial.

6. Discussion

Recent developments in deep learning have demonstrated that the detection of artificial intelligence (AI) generated synthetic content is significantly enhanced by hybrid architectures that combine Convolutional Neural Networks (*CNNs*) with Transformers and attention mechanisms. *CNNs* are very good at extracting local features, but they may not be as good at modelling long range connections or capturing global context. In contrast, Vision Transformers (*ViTs*) analyse complete image patches in parallel using self attention mechanisms, which allows for better representation of global structural inconsistencies that *CNNs* frequently miss. It has been demonstrated that combining *CNNs* and Transformers into a single framework, such as TransUNet or Swin Transformers, can enhance the accuracy and resilience of synthetic image identification tasks.

Deep fake detection models have been further strengthened by attention mechanisms, particularly self attention and cross modal attention, which enable them to selectively focus on features relevant to space or time. In ViTs or hybrid models, self attention layers help highlight small artefacts that are frequently present in GAN generated images, such as irregular texture patterns, uneven lighting, or boundary discrepancies. Cross modal attention enhances detection accuracy by facilitating collaborative reasoning across visual and aural cues in situations involving films or facial animations.

Accordingly, multimodal deepfake detection techniques that integrate auditory and visual inputs are gaining popularity. These methods process video frames and spectrograms using *CNNs* or recurrent architectures, then correlate lip movements with speech using attention based fusion layers. Manipulation may be indicated by differences between auditory signals and facial articulation. For instance, in lip sync deepfakes, the video may be artificially altered, but the audio may be authentic. Detection frameworks like Lip Forensics and Fake AVCeleb utilise temporal and spatial attention modules to specifically detect audiovisual synchronisation mismatches, thereby significantly enhancing performance in practical forensic scenarios.

One effective strategy for thwarting artificial image and video manipulation is the combination of *CNNs*, Transformers, attention mechanisms, and multimodal learning. These models support explainability and interpretability, which are essential in forensic applications where judgements must be justified in legal and investigative contexts, in addition to improving detection performance. Standardising these hybrid architectures for deepfake detection and improving their scalability and efficiency in operational forensic workflows should be the goals of future research.

7. Conclusion

The review highlights the increasing danger from AI-created synthetic images and the shortcomings of conventional detection techniques in detecting them. Deep learning methods, specifically CNNs, present promising solutions for synthetic image detection using automated feature learning. Although deep CNNs are highly accurate, their computational cost and black boxed nature remain the primary issues. Lightweight CNNs offer a different solution with better efficiency, but have limitations in extracting features. The incorporation of explainable AI methods is important to improving the interpretability of such models for forensic use. Moreover, the training of detection models using synthetic image data also offers opportunities and challenges, with domain gaps and biases potentially influencing model generalisation. Digital forensics must keep pace with these developments to counter the effective abuse of AI-generated images. Future studies should also aim to streamline lightweight CNN designs, enhance interpretability, and create standardised benchmarks to improve the accuracy and credibility of detection for forensic analysis.

8. Future Directions

The identification of artificial intelligence (AI)-generated synthetic content is a dynamic problem that requires reliable and flexible solutions. Future studies must focus on developing comprehensive, explainable, and resource efficient detection frameworks as synthetic media continues to advance in realism and scale. Creating open source, community driven detection toolkits that combine conventional and deep learning based techniques is a potential approach. These toolkits can enhance deployment in practical situations, facilitate reproducibility, and simplify benchmarking. Specifically, using transformer based architectures or lightweight CNNs

[76, 77, 78] with attention mechanisms may offer a compromise between interpretability and performance, which is crucial for forensic validation. Synthetic image detection recently offers new promises in digital image processing. [79, 80, 81, 82]

The efficacy of models in various domains is still hindered by the absence of standardised benchmarks and challenge datasets. Robust performance evaluation will be made possible by the creation of large scale, diversified, labelled datasets with different levels of manipulation. Additionally, the lack of labelled forensic data can be addressed by implementing transfer learning and few shot learning techniques, which improve the generalisation of models trained on synthetic or comparable domains to real world tampered photos. Utilising domain adaptation strategies to bridge the domain gap can help narrow the performance gap between synthetic training data and real world applications.

Integrating blockchain technology and watermarking techniques into media creation pipelines may provide trustworthy provenance monitoring to guarantee the integrity and traceability of digital media. These technologies are essential for forensic and legal use cases since they log transformation history in addition to authenticating. To find discrepancies across modalities, detection systems must eventually develop into cross modal solutions that can collaboratively analyse text, audio, video, and image data. Given the rapid development of generative models, multimodal intelligence, and open decision making, as well as the ongoing co evolution of adversarial and detection models, these are key components of the future of forensic detection.

References

[1] Goodfellow, I., *et al.* (2020). Generative adversarial networks," *Commun.* ACM, V. 63 (11), p. 139–144.

[2] Masood, M., Nawaz, M., Malik, K M., Javed, A., Irtaza, A., Malik, H. (2023). Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward, Appl. Intell., V. 53 (4), p. 3974–4026, doi: 10.1007/s10489-022-03766-z.

[3] Mirsky, Y., Lee, W. (2021). The creation and detection of deep fakes: A survey, *ACMComput. Surv.*, V. 54 (1), p. 1–41.

[4] Thakur, R., Rohilla, R. (2020). Recent advances in digital image manipulation detection techniques: A brief review, Forensic Sci. Int., V. 312, p.110311.

[5] Fridrich, J., Soukal, D., Lukáš, J. Detection of Copy Move Forgery in Digital Images.

[6] Li, G., Wu, Q., Tu, D., Sun, S. (2007). A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD, in *2007 IEEEinternational conference on multimedia and expo*, IEEE, , p. 1750–1753.

[7] Bayram, S., Sencar, H T., Memon, N. (2009). An efficient and robust method for detecting copy-move forgery, in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, p. 1053–1056.

[8] Ferrara, P., Bianchi, T., Rosa, De A., Piva, A. (2012). Image forgery localization via fine-grained analysis of CFA artifacts, *IEEE Trans. Inf. Forensics Secur.*, V. 7 (5), p. 1566–1577.

[9] He, Z., Lu, W., Sun, W., Huang, J. (2012). Digital image splicing detection based on Markov features in DCT and DWT domain, *Pattern Recognit.*, V. 45, (12), p. 4292–4299.

[10] Sharma, D K., Singh, B., Agarwal, S., Garg, L., Kim, C., Jung, K H. (2023). A survey of detection and mitigation for fake images on social media platforms, Appl. Sci., V. 13 (19), p. 10980.

[11] Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y. (2016). Deep Learning Cambridge,” *MA MIT Press http/ /www. Deep. org.*

[12] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large scale image recognition,” *arXiv Prepr. arXiv1409.1556.*

[13] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K Q. (2017). Densely connected convolutional networks,” *In:* Proceedings of the IEEE conference on computer vision and pattern recognition, p. 4700–4708.

[14] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition, *In:* Proceedings of *the IEEE conference on computer vision and pattern recognition*, p. 770–778.

[15] Gu, J., *et al.* (2018). Recent advances in convolutional neural networks, *Pattern Recognit.*, V. 77, p. 354–377.

[16] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., L C. (2018). Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, *In:* Proceedings of the IEEE conference oncomputer vision and pattern recognition, p. 4510–4520.

[17] Abbas, N M., Ansari, M S., Asghar, M N., Kanwal, N., Neill, T O., Lee, B. (2021). Light weight deep learning model for detection of copy-move image forgery with post processed attacks, *In: 2021 IEEE 19th world symposium on applied machineintelligence and informatics (SAMI), IEEE*, p. 125–130.

[18] 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.* P. 4401–4410.

[19] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T. (2020). Analyzing and improving the image quality of stylegan, *In:* Proceedings of the *IEEE/CVF conferenceon computer vision and pattern recognition*, p. 8110–8119.

[20] Karras, T., *et al.* (2021). Alias free generative adversarial networks, *Adv. Neural Inf. Process.Syst*, V. 34, p. 852–863.

[21] Pei, S., Da Xu, R Y., Xiang, S., Meng, G., (2021). Alleviating mode collapse in GAN via diversity penalty module, *arXiv Prepr. arXiv2108.02353.*

[22] Liu, H., *et al.* (2023). Combating mode collapse via offline manifold entropy estimation, *In:* Proceedings of the AAAI Conference on Artificial Intelligence, p. 8834–8842.

[23] Ghosh, A., Kulharia, V., Namboodiri, V P., Torr, P H S., Dokania, P K. (2018). Multi-agent diverse generative adversarial networks, *In:* Proceedings of the IEEE conferenceon computer vision and pattern recognition, pp. 8513–8521.

[24] Sohn, K., Yan, X., Lee, H. (2015-January). Learning structured output representation using deep conditional generative models, Adv. Neural Inf. Process. Syst., p. 3483–3491.

[25] Mansimov, E., Parisotto, E., Ba, J. L., Salakhutdinov, R. (2016). Generating images from captions with attention, *4th* Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc., p. 1–12.

[26] DeepMind, (2017). VQ-VAE: Neural Descrete Representation Learning, *NeurPS*, no. Nips,.

[27] Huang, H., li, zhihang., He, R., Sun, Z., Tan, T. (2018). Intro VAE: Introspective Variational Autoencoders for Photographic Image Synthesis, *In: Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa Bianchi, and R. Garnett, Eds., Curran Associates, Inc.

[28] Daniel, T., Tamar, A. (2021). Soft IntroVAE: Analyzing and Improving the Introspective Variational Autoencoder, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, p. 4389–4398, doi: 10.1109/CVPR46437.2021.00437.

[29] Ho, J., Jain, A., Abbeel, P. (2020). Denoising diffusion probabilistic models, *Adv. NeuralInf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS 2020, p. 1–12,.

[30] Nichol, A. Dhariwal, P. (2021). Improved Denoising Diffusion Probabilistic Models, *Icml*,

[31] Song, J., Meng, C., Ermon, S. (2021). Denoising Diffusion Implicit Models,” *ICLR 2021 -9th Int. Conf. Learn. Represent.*, p. 1–22,.

[32] Zhang, Q., Tao, M., Chen, Y. (2023). Gddim: Generalized Denoising Diffusion Implicit Models, 11th Int. Conf. Learn. Represent. ICLR 2023, no. Ddim, p. 1–31,.

[33] Nachmani, E., Roman, R S., Wolf, L. (2021). Non Gaussian Denoising Diffusion Models.

[34] Lu, C., Zhou, Y., Bao, F., Chen, J C LI, Zhu, J. (2022). DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps, *In: Advances in NeuralInformation Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K.Cho, and A. Oh, Eds., Curran Associates, Inc., p. 5775–5787.

[35] C. Lu, Zhou, Y. Bao, F., Chen, J., Li, C., Zhu, J., (2025). DPM-Solver++: Fast Solver for Guided Sampling of Diffusion Probabilistic Models, *Mach. Intell. Res.*, doi: 10.1007/s11633-025-1562-4.

[36] Watson, D., Ho, J., Norouzi, M., Chan, W. (2021). Learning to Efficiently Sample from Diffusion Probabilistic Models,.

[37] Sai Teja Boppiniti, (2023). Data Ethics in AI: Addressing Challenges in Machine Learning and Data Governance for Responsible Data Science, *Int. Sci. J. Res.*, no. December.

[38] Kolluri, V. (2021). Revolutionary Research On The AI Sentry: An Approach To Overcome Social Engineering Attacks Using Machine Intelligence, *Int. J. Adv. Res. Interdiscip.Sci. Endeav.*, V. 1 (1), p. 2024, doi: 10.61359/11.2206-2405.

[39] Boppiniti, S T., Boppiniti, S T. (2021). Real Time Data Analytics with AI: Leveraging Stream Processing for Dynamic Decision Support, *Int. J. Manag. Educ. Sustain. Dev.*, V. 4, (4).

[40] Yarlagadda, V S T. (2022). AI Machine Learning for Improving Healthcare Predictive Analytics: A Case Study on Heart Disease Risk Assessment, *Trans. Recent Dev. Artif. Intell. Mach. Learn.*, V. 14, (14) SE-Articles, Aug.

[41] Gatla Dtcc, T., Reddy Gatla, T. (2021). a Groundbreaking Research in Breaking Language Barriers: Nlp and Linguistics Development, V. 9 (1), p. 6171, doi: 10.61359/11.2206-2401.

[42] Davuluri, Manaswini., Yarlagadda, Sai Teja,Venkata. (2024). Novel Device for Enhancing Tuberculosis Diagnosis for Faster, More Accurate Screening Results, *Int. J. Innov.Eng. Res. Technol.*, V. 11(11), p. 1–15, doi: 10.26662/ijiert.v11i11.pp1-15.

[43] Kolluri, V. (2015). a Comprehensive Analysis on Explainable and Ethical Machine: Demystifying Advances in Artificial Intelligence," *TIJER* - Int. Res. Journals, V. 2 (7), p. 2349–9249.

[44] Gatla, T R. (2024). AI-driven Regulatory Compliance for Financial Institutions: Examining How AI Can Assist in Monitoring and Complying With Ever-changing Financial Regulations, *SSRN Electron. J.*, V. 12 (3), pp. 607–611, doi: 10.2139/ssrn.4856649.

[45] Srini, G., Global, H T C. (2016). Applications of Big Data Analytics and Applications of Big Data Analytics and *J. Recent Trends.,* March 2015, V. 2, p. 9–13.

[46] Gatla Dtcc, T., Reddy Gatla, T. (2024). A Next Generation Device Utilizing Artificial Intelligence For Detecting Heart Rate Variability And Stress Management, May, V. 12, p. 2320–2882.

[47] Ramesh Kumar, V. (2025). Nanosatellite Constellations in Low Earth Orbit: A Comprehensive Review, *Int. J. Adv. Res. Interdiscip. Sci. Endeav.*, V. 2 (1), p. 420–423, doi: 10.61359/11.2206-2503.

[48] Islam, M M., Hassan, S., Akter, S., Jibon, F A., Sahidullah, M. (2024). A comprehensive review of predictive analytics models for mental illness using machine learning algorithms, *Healthc. Anal.*, V. 6, p. 100350, doi: https://doi.org/10.1016/j.health.2024.100350.

[49] Ghiurãu, D., Popescu, D E. (2025). Distinguishing Reality from AI: Approaches for Detecting Synthetic Content," *Computers*, V. 14 (1), doi: 10.3390/computers14010001.

[50] Kolluri, V. (2024). An Extensive Investigation In to Guardians Of The Digital Realm: Ai- Driven Antivirus And Cyber Threat Intelligence, *Int. J. Adv. Res. Interdiscip. Sci. Endeav.*, V. 1, (2), p. 2024, doi: 10.61359/11.2206-2407.

[51] Sathiyanarayanan, M. (2016). Introduction to Digital Forensics, *Univ. London, UK.*

[52] Gogolin, G. (2021). *Digital forensics explained.* CRC Press.

[53] Atapour Abarghouei, A., Breckon, T P. (2018). Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer, *In:* Proceedings ofthe IEEE conference on computer vision and pattern recognition, p. 2800–2810.

[54] Sakaridis, C. Dai, D., Van Gool, L. (2018). Semantic foggy scene understanding with synthetic data, *Int. J. Comput. Vis.*, V. 126, p. 973–992.

[55] Behl, H S., Baydin, A G., Gal, R., P H S, Torr., Vineet, V. (2020). Autosimulate:(quickly) learning synthetic data generation, *In:* European Conference on Computer Vision, Springer, p. 255–271.

[56] Hattori, H., Lee, N., Boddeti, V N., Beainy, F., Kitani, K M., Kanade, T. (2018). Synthesizing a scene-specific pedestrian detector and pose estimator for static video surveillance: Can we learn pedestrian detectors and pose estimators without real data?," *Int. J. Comput.Vis.*, V. 126, p. 1027–1044.

[57] Wang, Q., Gao, J., Lin, W., Yuan, Y. (2021). Pixel wise crowd understanding via synthetic data, *Int. J. Comput. Vis.*, V. 129 (1), p. 225–245.

[58] Kortylewski, A., Schneider, A., Gerig, T., Egger, B., Morel Forster, A., Vetter, T. (2018). Training deep face recognition systems with synthetic data, *arXiv Prepr.arXiv1802.05891*.

[59] Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B. (2019). Moment matching for multi-source domain adaptation," *Proc. IEEE Int. Conf. Comput. Vis.*, V. (2), p. 1406– 1415, doi: 10.1109/ICCV.2019.00149.

[60] Zhong, L., Fang, Z., Liu, F., Yuan,B., Zhang, Lu, G J. (2023). Bridging the Theoretical Bound and Deep Algorithms for Open Set Domain Adaptation, *IEEE Trans. Neural Networks Learn. Syst.*,V. 34 (8), p. 3859–3873, doi:10.1109/TNNLS.2021.3119965.

[61] Tsirikoglou, A., Eilertsen, G., Unger, J. (2020). A survey of image synthesis methods for visual machine learning, *In: Computer graphics forum*, Wiley Online Library, p. 426–451.

[62] Frolov, S., Hinz, T., Raue, F., Hees, J., Dengel, A. (2021). Adversarial text to image synthesis: A review," *Neural Networks*, vol. 144, pp. 187–209.

[63] Gaidon, A., Lopez, A., Perronnin, F. (2018). The reasonable effectiveness of synthetic visual data, *Int. J. Comput. Vis.*, V. 126 (9), p. 899–901.

[64] Varol, G., *et al.* (2017). Learning from synthetic humans, *In:* Proceedings of the IEEEconference on computer vision and pattern recognition, p. 109–117.

[65] Allken, V., Handegard, N O., Rosen, S., Schreyeck, T., Mahiout, T., Malde, K. (2019). Fish species identification using a convolutional neural network trained on synthetic data, *ICES J. Mar. Sci.*, V. 76 (1), p. 342–349.

[66] Rosen, S., Holst, J C. (2013). DeepVision in trawl imaging: Sampling the water column in four dimensions," *Fish. Res.*, V. 148, p. 64–73.

[67] Abu Alhaija, H., Mustikovela, S K., Mescheder, L., Geiger, Rother, C. (2018). Augmented reality meets computer vision: Efficient data generation for urban driving scenes, *Int. J. Comput. Vis.*, V. 126, p. 961–972.

[68] Geiger, A., Lenz, P., Stiller, C., Urtasun, R. (2013). Vision meets robotics: The kitti dataset, *Int. J. Rob. Res.*, V. 32, (11), p. 1231–1237.

[69] Latif, J., Tu, S., Xiao, C., Rehman, S U., Sadiq, M., Farhan, M. (2021). Digital forensics use case for glaucoma detection using transfer learning based on deep convolutional neural networks, *Secur. Commun. Networks*, V. 2021 (1), p. 4494447.

[70] Kim, H., *et al.* (2018). Deep video portraits," *ACM Trans. Graph.*, V. 37, (4), p. 1–14.

[71] Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., Nießner, M. (2016). Face 2 face: Real time face capture and reenactment of rgb videos, *In:* Proceedings of the IEEEconference on computer vision and pattern recognition, p. 2387–2395.

[72] Wang, Y., *et al.*, (2017). Tacotron: Towards end to end speech synthesis, *arXiv Prepr.arXiv1703.10135.*

[73] Van Den Oord, A., *et al.* (2016). Wavenet: A generative model for raw audio, *arXiv Prepr.arXiv1609.03499,* V. 12, p. 1.

[74] Anirudh, K., Srikanth, M., Shahina, A. (2023). Multilingual fake news detection in low-resource languages: A comparative study using BERT and GPT-3.5, *In:* InternationalConference on Speech and Language Technologies for Low resource Languages, Springer, p. 387–397.

[75] Radford, A., *et al.* (2021). Learning transferable visual models from natural language supervision, *In: International conference on machine learning*, PmLR, p. 8748– 8763.

[76] Hussein, I Haval., Mohammed, O Abdulhakeem., Hassan, M Masoud., Mstafa, J Ramadhan. (2023) Lightweight deep CNN based models for early detection of COVID-19 patients from chest X ray images, Expert Systems with Applications, 223. p. 119900.

[77] Ahad, Md Taimur., Li, Yan., Song, Bo., Bhuiyan, Touhid. (2023). Comparison of CNN based deep learning architectures for rice diseases classification, Artificial Intelligence in Agriculture,V. 9, p. 22-35

[78] Bilous, N., Malko, V., Frohme, M., Nechyporenko, A. (2024). Comparison of CNN-Based Architectures for Detection of Different Object Classes. AI 2024, V. *5*, P. 2300-2320.

[79] Wang, Yuyang., Hao, Yizhi., Cong, Xu Amando. (2024). Harnessing Machine Learning for Discerning AI-Generated Synthetic Images. arXiv.org·

[80] Manisha, C T. Li., Kotegar, K A. (2025). AI-Synthesized Image Detection: Source Camera Fingerprinting to Discern the Authenticity of Digital Images, *In: IEEE Access*, V. 13, p. 29660-29672.

[81] Ghiurãu, D., Popescu, D E., Distinguishing Reality from AI: Approaches for Detecting Synthetic Content. *Computers* 2025, V. 14, P. 1.

[82] Sohail, S., Sajjad, S M., Zafar, A., Iqbal, Z., Muhammad, Z., Kazim, M. (2025). Deepfake Image Forensics for Privacy Protection and Authenticity Using Deep Learning. Information, V. 16, P. 270.