

Image Domain Distinct Native Attribute Fingerprinting for Image Forgery Classification

Jessica McQuagge
Air Force Institute of Technology
jessica.mcquagge@afit.edu

Michael Temple
Air Force Institute of Technology
michael.temple@afit.edu

Christopher Rondeau
Air Force Institute of Technology
christopher.rondeau@afit.edu

Abstract

Image forgery is becoming more difficult to detect due to advances in AI image generation. As such, the usefulness — and even requirement — for detection techniques that are affordable (computationally and monetarily) as well as intuitive and simple are equally increasing. This work demonstrates the first adoption of Distinct Native Attribute (DNA) Fingerprinting to image and forgery detection to achieve similar results while mitigating the cost of implementation. General image classification results with accuracy of $\%C = 98.8\%$ support the overall utility while the ability to detect within-category image forgeries produce an average of $\%C = 81.8\%$. Using an intuitive and small set of features, preliminary results show an approximate average classification accuracy difference of only $\%C_{\Delta} = -9\%$ from more complex solutions. This work demonstrates the ability to adopt DNA Fingerprinting for image classification, and image forgery using Image Domain DNA (ID-DNA) that is holistically less resource intensive while requiring less time, money, and expert knowledge.

Keywords: Image Forgery, Forgery Detection, Classification, Fingerprinting, Perception Dominance

1. Introduction

The concept of perception is defined as “a central means by which we become aware of the surrounding world” (Lande, 2023). This work explores the concept of perception and how it may be impacted by artificially-generated — or forged — images being passed as real images and how to combat that violation of one’s perception. Imagery necessarily utilizes vision which is often the highest valued sense available to form

human perception. Vision is especially trusted in the absence of, or confusion with other sensory data. For centuries courts of law held eyewitness testimony as the standard even as modern forensic techniques were introduced (Nash et al., 2015). The introduction of a visual medium can even change the perception of an event, as was the case in the first televised presidential debates in the United States in 1960 when those who watched the debate on television had the opposite impression of who won compared to those who listened on the radio (Self, 2005). Even in the colloquial realm, the English phrase “seeing is believing” is commonly used to convey the primacy of a visual observation. Given the state of artificially generated image forgeries, the question evolves to whether or not seeing really is still believing.

The potential for confusion between real and artificially generated images has significantly increased with substantial improvements to the field of Artificial Intelligence (AI) image generation, primarily via Generative Adversarial Networks (GAN) and Latent Diffusion Models (LDM) (Rombach et al., 2022; Monkam et al., 2023; Hossain et al., 2023). Using GAN and LDM, AI can alter existing images, create entirely new images from existing images, and create images that appear real from text-based user input.

These advancements in generative AI make it difficult to distinguish real images from AI-generated images. The ability, however, to distinguish real from artificially generated forged images is essential due to its potential impact to news, social media, law enforcement investigations, forensic sciences, and warfare. Realistic forged images create doubts about authenticity and make it difficult to cultivate public trust. Even more detrimental is that an adversary may intentionally exploit this lack of trust. Research

and development of methods to identify forged images are focused on combating this distrust and aiding in legal and moral decisions. It is in this context that this work labels artificially generated images as forgeries. That is, regardless of the intent these artificially generated images are presenting as the real image without qualification.

The current emphasis in research is primarily focused on the use of neural networks to identify real versus artificially generated images (Passos et al., 2024). Neural networks are proficient at identifying artificially generated images but at a high cost; they require complex encoding, extensive computational resources, and are monetarily expensive to implement. Given the current success rates of such classification methods, this work seeks an alternative that could provide a simple and less resource intensive method that could provide comparable results to the neural net-based solutions.

This work introduces the concept of Distinct Native Attribute (DNA) Fingerprinting to image classification. DNA Fingerprinting has previously been used primarily on Radio Frequency (RF) communication signals and exploits physical layer data to classify the identity of the transmitting device regardless of the higher layer credentials it presents (Klein et al., 2009; Lopez et al., 2016; Rondeau et al., 2021). The traditional DNA Fingerprinting process is adopted in this work for image classification because of the relatively low complexity of implementation as compared to neural net-based image classification solutions. Therefore, if successful, the introduction of DNA Fingerprinting as a method for image classification would immediately satisfy the goal of this work to deliver a simple and computationally expedient method. As such, the contributions of this work are to

- Determine if DNA Fingerprinting can be used for image classification and quantify its utility
- Identify AI-generated forged images
- Decrease computational and monetary costs associated with AI-forged image classification

The rest of the paper is organized as follows. Section 2 provides background on current AI image forgery detection methods, an overview of the dataset used in this work, and DNA Fingerprinting. Section 3 summarizes the research methodology and evaluation metrics. Section 4 presents the results of this work, and Section 5 concludes with a discussion of the results and recommendations for future work.

2. Background

This section provides a background on current image forgery techniques and forgery detection methods, an overview of the dataset used in this work, a description of the classifier used in this work, and a description of DNA Fingerprinting.

2.1. AI-Generated Image Detection Methods

Image forgery detection may generally be separated into two categories: active and passive detection (Gill et al., 2017; Mushtaq and Mir, 2014). Active forgery identification is applied when the original information has been processed and a digital watermark or signature has been applied. Passive forgery detection methods search the image itself for statistical changes in the underlying pixel values. Among the most common image forgery techniques are copy-move forgeries and image splicing forgeries (Bayram et al., 2009, Gill et al., 2017, Jain and Goel, 2021, Jwaid and Baraskar, 2017). Generative AI can create realistic images based on text input (Göring et al., 2023). AI is also able to alter images, create new images based on existing images, and create images based on text descriptions, however; these abilities are not currently classified as image forgery techniques. Detecting whether these images are real or generated presents a challenge due to the authentic appearance of the generated images.

As advances in AI have introduced a significant source of image forgeries, researchers have likewise used many of those same advances to combat it. Convolutional Neural Networks (CNN) and Generative Bayesian Optimal detector GANs are among the neural networks used for image classification due to their superiority in detecting intricate patterns within images (Hossain et al., 2023, Monkam et al., 2023).

One such example is AUthenticates SOcial MEDIA images (AUSOME-2), which is a CNN that uses frequency analysis and machine learning algorithms to classify AI generated images from real images (Poredi et al., 2024). AUSOME-2 uses the discrete cosine transform (DCT) of images to identify grid patterns in high-frequency regions of synthetic images, which do not occur in real images. AUSOME-2 classified real versus synthetic images with an accuracy of %C = 96.88%. Bird and Lotfi (2024) used a dataset of 60,000 real and 60,000 fake images, broken into ten categories of images (i.e. birds, planes, etc). CNNs classified the images as real or fake and classified with %C = 92.98% accuracy. Hossain et al. (2023) trained three CNNs under slightly different parameters

to classify real versus fake images using the CIFAKE dataset, which resulted in an average %C = 95.97% accuracy in classification. These data are used in this work since it provides a sufficient dataset for the investigation and ensures baselines of performance exist.

GANs have also shown the propensity to create as well as classify synthetically generated images. Monkam et al. (2023) used Generative Bayesian Optimal detector FAN (G-JOB GAN) to ascertain whether an image was synthetic or real with a %C = 95.7% accuracy.

The advanced coding skill requirements, computational resources, monetary expenses of graphical processing units, and training of the neural networks create a complex problem in itself. Deep learning neural networks require complex parallel calculations which graphics processing units (GPUs) handle more efficiently than central processing units (CPUs) (Zhang, 2021). In 2021, a study showed the GTX3070 GPU took around 12 seconds to complete calculations used in a neural network compared to the CPU's 6 minutes to complete those same calculations. Even as advances are realized, the computational cost remains a significant factor in evaluating the overall utility of the neural networks (Krizhevsky et al., 2017, He et al., 2015).

Methods utilizing grayscale images as opposed to full color images (Bayram et al., 2009, Shailaja Rani and Kumar, 2019) have shown their effectiveness in image forgery detection. Despite this, the first transitional step from RF-based to image-based DNA Fingerprinting that is being addressed here uses full color images. As in earlier RF-based proof-of-concept works, higher-dimensional feature spaces are often considered first given the uncertainty in knowing where the most reliable discrimination information “resides.” When the first-step higher-dimensional results are promising, they are followed by next-step dimensional reduction work aimed at improving overall efficiency. The first-step ID-DNA Fingerprinting results presented in Section 4 are indeed promising and motivate consideration for next-step future activity using grayscale images.

2.2. CIFAR-10 & CIFAKE Dataset

In order to sufficiently establish a useful baseline of performance and to ensure applicability to other published works and methods, the CIFAR-10 and related CIFAKE datasets are used in this work. The CIFAR-10 dataset contains 60,000, 32×32 pixel Red, Green, Blue (RGB) images, divided into 10 categories

such as bird, dog, airplane, etc. (Krizhevsky, 2009). The authors in Bird and Lotfi (2024) used Stable Diffusion version 1.4 to create the CIFAKE dataset as an equivalent to the CIFAR-10 dataset. The CIFAKE dataset contains 60,000, 32×32 pixel artificially generated images, divided into the same ten categories as the CIFAR-10 dataset (Bird and Lotfi, 2024). Therefore, CIFAKE provides a sufficient repository of forged images from the real images in CIFAR-10. A representation of both the CIFAR-10 groups and images alongside the CIFAKE dataset are shown in Figure 1.



Figure 1: Representative Real and Fake Images from CIFAR-10 and CIFAKE Image Datasets.

2.3. Traditional DNA Fingerprinting

DNA Fingerprinting, like the more generic RF fingerprinting, identifies and exploits physical layer features to identify or verify a transmission source. DNA Fingerprinting has been used in a variety of applications such as enhancing security of wireless devices, augmenting physical network security, and classifying devices (Klein et al., 2009; Lukacs et al., 2015; Williams et al., 2010; Reising et al., 2010). While the method of collection and physical layer response generation has changed (wireless, wired,

active/passive), the concept of fundamental end-to-end processing of RF-DNA Fingerprinting has not yet been applied to image classification.

The traditional, physical layer DNA Fingerprinting process begins by selecting a region of a physical layer emission as the region of interest. Most communication signals contain a preamble response in which the same communication symbols are transmitted in a repeatable pattern. This invariant part of a transmitted signal is a commonly used region of interest. Next, features are extracted from various time domain responses, typically the amplitude (AMP), frequency (FRQ), and phase (PHZ) responses. From each region, three statistics (moments) of variance (σ^2), skewness (γ) and kurtosis (κ) are computed. These statistics are used to form the statistic vector

$$\mathbf{F}^{Stat} = \left[\sigma^2 : \gamma : \kappa \right]_{(1 \times 3)}, \quad (1)$$

where $:$ denotes concatenation. Depending on the signal characteristics, the region of interest is divided into N_R subregions which provides the regional statistic vector,

$$\mathbf{F}^{Rgn} = \left[\mathbf{F}_{R1}^{Stat} : \mathbf{F}_{R2}^{Stat} : \dots : \mathbf{F}_{NR+1}^{Stat} \right]_{1 \times [3(N_R+1)]}. \quad (2)$$

The regional vectors are used to form the composite Fingerprint vector given by,

$$\mathbf{F}_{Total} = \left[\mathbf{F}_{AMP}^{Rgn} : \mathbf{F}_{PHZ}^{Rgn} : \mathbf{F}_{FRQ}^{Rgn} \right]_{(1 \times N_F)}, \quad (3)$$

where N_F is the total number of features consisting of the $N_{Stat} = 3$ statistics for the three instantaneous responses across all $N_R + 1$ subregions. This output composite Fingerprint vector serves as the input to the Multiple Discriminant Analysis / Maximum Likelihood (MDA/ML) classifier used in this work in order to achieve the maximum performance while mitigating computational expense.

2.4. MDA/ML Classifier

The MDA/ML processing used here is a readily implementable, computationally efficient process that has provided reliable device discrimination in prior DNA works (Klein et al., 2009; Lukacs et al., 2015; Williams et al., 2010; Reising et al., 2010; Lopez et al., 2016; Rondeau et al., 2019). MDA/ML is selected as the classifier for this work based on successful history with DNA Fingerprinting and its relatively simple implementation compared to a neural network.

The process produces a projection matrix \mathbf{W} with a goal of maximizing between-class separation and minimizing within-class spread. For discriminating N_{Cls} classes using input fingerprints (\mathbf{F}) having N_F features, \mathbf{W} has dimension $N_F \times (N_{Cls} - 1)$ and effectively projects $(1 \times N_F)$ -dimensional \mathbf{F} into the $(N_{Cls} - 1)$ -dimensional decision space.

Given a trained MDA “model” a 1 vs. N_{Cls} declared-class estimate (correct or incorrect) for an input “unknown” testing fingerprint \mathbf{F}_{Tst} is defined as $\mathbf{F}_{Tst}^W = \mathbf{F}_{Tst} \mathbf{W}$, where \mathbf{F}_{Tst}^W is the projection of \mathbf{F}_{Tst} in the Fisher space. Assuming equal probability of class occurrence and equal error costs into the Fisher projection space, the probability relationship becomes a Maximum Likelihood (ML) estimate. The class yielding the highest probability is the declared-class for \mathbf{F}_{Tst}^W .

2.5. Image Domain DNA Fingerprinting

In order to apply traditional DNA Fingerprinting to images, the following areas were modified from the aforementioned DNA Fingerprinting process: statistical measures and subregion anatomy. The process described herein shall be referred to as Image Domain DNA (ID-DNA) Fingerprinting.

Statistical Measures Consistent with the presentation from Equation 1 and previous DNA Fingerprinting works, variance (σ^2), skewness (γ) and kurtosis (κ) are retained as statistical components of \mathbf{F}^{Stat} . In order to remove the assumptions of normality and thereby increase the usefulness of ID-DNA Fingerprints, this work introduces L-moments, a sequence of statistics that have proven useful in various fields. L-moments are linear combinations of ordered statistics with the intent of summarizing distributional shapes (Hosking, 1990). This is in contrast to the traditional statistical moments used in \mathbf{F}^{Stat} which can be influenced by outliers or extreme values. Aside from providing a different quantity for feature generation, L-moments provided added resilience when extreme values are present (Guttman, 1993). L-moments therefore offer a different approach to quantifying the distribution parameters and shapes. They are resilient to peculiar phenomenon likely to be present in images and thus warrant due consideration to include in the ID-DNA \mathbf{F}^{Stat} .

In order to calculate the L-moments (λ_1 through λ_6), let the probability-weighted moment for the r^{th} component be defined as,

$$\beta_r = \int_{-\infty}^{\infty} x(P(x))^r p(x) dx, \quad (4)$$

where $P(x)$ and $p(x)$ are the cumulative distribution

function and probability density function, respectively, of the random variable X . The first six L-moments may then be defined in terms of Equation 4 and are shown in Table 1 along with their 2 corresponding τ -moments.

L-Moment	Definition
λ_1	β_0
λ_2	$\frac{1}{2}(\beta_0 - 2\beta_1 + \beta_2)$
λ_3	$\frac{1}{3}(\beta_0 - 3\beta_1 + 3\beta_2 - \beta_3)$
λ_4	$\frac{1}{4}(\beta_0 - 4\beta_1 + 6\beta_2 - 4\beta_3 + \beta_4)$
λ_5	$\frac{1}{5}(\beta_0 - 5\beta_1 + 10\beta_2 - 10\beta_3 + 5\beta_4 - \beta_5)$
λ_6	$\frac{1}{6}(\beta_0 - 6\beta_1 + 15\beta_2 - 20\beta_3 + 15\beta_4 - 6\beta_5 + \beta_6)$
τ_3	λ_3/λ_2
τ_4	λ_4/λ_2

Table 1: Description of L-Moments and τ -Ratios.

The two τ -moments included in Table 1, τ_3 and τ_4 , are introduced as the normalized forms of the third and fourth L-moments, respectively. These normalized L-moments are similar to the traditional moments of skewness and kurtosis, which are less sensitive to outliers and extreme values compared to traditional moments (Vargo et al., 2010).

The resulting ID-DNA \mathbf{F}^{Stat} contains traditional moments, L-moments, and the τ -ratios in the form,

$$\mathbf{F}^{Stat} = \left[\sigma^2 : \gamma : \kappa : \lambda_1 : \lambda_2 : \lambda_3 : \lambda_4 : \lambda_5 : \lambda_6 : \tau_3 : \tau_4 \right]_{(1 \times 11)}. \quad (5)$$

Subregion Anatomy Traditional DNA Fingerprinting uses subregions (detailed in Equation 2) within the region of interest to compute the composite Fingerprint vector (detailed in Equation 3). This work could implement any number of permutations to translate the concept of subregions from physical layer signals to images. For the sake of a proof-of-concept and minimizing computation, this work pursues easily traceable methods and those consistent with image processing. For a 3-color RGB image, there are a total of 4 different subregions used to generate the equivalent \mathbf{F}_{Total} vector for ID-DNA Fingerprinting. In order to maximize traceability to traditional DNA Fingerprinting, the term *subregion* is retained when describing ID-DNA Fingerprinting.

The first three of the four subregions are a result of a 2-dimensional (2D) convolution filter applied across each $N_{Ch} = 3$ color channels in the spatial domain. Let the correlation kernel be defined as,

$$\mathbf{K} = \frac{\alpha}{1 + d(|a - b|)}, \quad (6)$$

where $d(|a - b|)$ is the Euclidean distance between some center pixel a and a location in the kernel, b , and α is

a scalar used to normalize \mathbf{K} such that the sum of the elements equal 1. This work utilizes \mathbf{K} as a 5×5 square matrix for ease of implementation. Figure 2 displays the output of this correlation process using one of the real CIFAR-10 images from the ‘Truck’ category and shows the original image and corresponding color channels after the 2D correlation filtering.

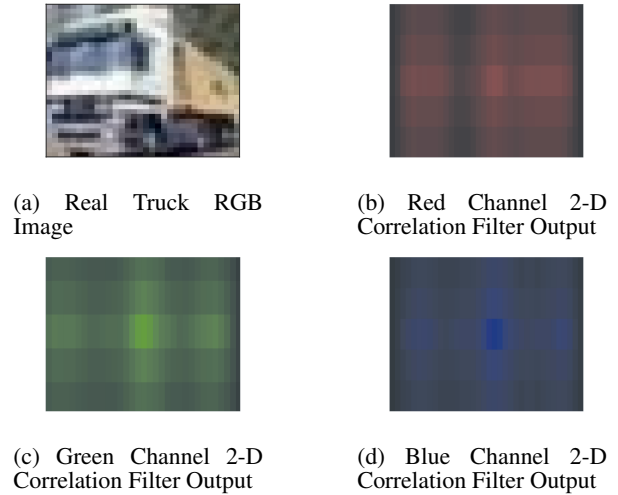


Figure 2: Example of 2D Correlation Color Channel Subregion Creation.

For each image, the statistics from \mathbf{F}^{Stat} shown in Equation 5 form \mathbf{F}_{Ch}^{Stat} for each of the $N_{Ch} = 4$ color channels resulting in the vectors \mathbf{F}_{Red}^{Stat} , \mathbf{F}_{Grn}^{Stat} , and \mathbf{F}_{Blu}^{Stat} for the red, green, and blue channels, respectively.

Whereas the first three subregions observe the statistical distributions within each color channel, the fourth subregion is intended to create statistical observations of the interaction between the color channels. As with the other three subregions, the \mathbf{F}^{Stat} statistics from Equation 5 form \mathbf{F}_{Img}^{Stat} using the entire $m \times n \times 3$ image (for example, the image shown in Figure 2(a)). The final Composite ID-DNA Fingerprint vector is formed by the aggregate of all four image subregions with 11 statistics per subregion for $N_F = 44$ features,

$$\mathbf{F}_{Total} = \left[\mathbf{F}_{Red}^{Stat} : \mathbf{F}_{Grn}^{Stat} : \mathbf{F}_{Blu}^{Stat} : \mathbf{F}_{Img}^{Stat} \right]_{(1 \times N_F)}. \quad (7)$$

3. Methodology

This section resets the overall problem statement and motivates the two-phase methodology used in the adoption of ID-DNA Fingerprinting. The section concludes with a description of the evaluation metrics used.

3.1. Problem Description

The methodology presented here is aimed to support identification of artificially generated forged images in such a way that did not require significant computational resources or algorithmic complexity. Hence, the related goal of this work was to determine if DNA Fingerprinting could be used for image detection in the form of ID-DNA Fingerprinting, described in the previous section, and then to quantify that utility. This naturally led to a two-phase methodology. The first phase began by establishing a determination of whether or not DNA Fingerprinting can successfully be used for image classification. With the established baseline, the second phase ID-DNA Fingerprinting was applied to real and fake images using the CIFAKE dataset. Computation time and algorithmic complexity was noted, and throughout the analysis of the results these elements will be addressed qualitatively vice an explicit quantitative metric.

3.2. ID-DNA Fingerprinting Baseline for Image Classification

The ID-DNA Fingerprint structure described in the preceding sections is implemented using the first 50,000 images from the CIFAR-10 and CIFAKE datasets with 5,000 images from each of the 10 image categories. The baseline is established with the CIFAR-10 dataset. The CIFAR-10 set is entirely real images, so the comparison is a 10-class problem where each class represents a category of the images as defined by CIFAR-10.

3.3. ID-DNA Fingerprinting for Image Forgery Detection

Following the results of ID-DNA Fingerprinting baseline as applied to the CIFAR-10 dataset, the introduction of artificially generated forged images was accomplished using the CIFAKE dataset. This is a consistent progression of difficulty for image classification since the CIFAKE dataset contains real and fake images in the same 10 categories of from CIFAR-10 and the images are the same 32×32 RGB format. This evaluation consisted of 2 different comparisons. The first comparison was a 2-class real vs. fake evaluation to quantify the raw utility of ID-DNA Fingerprinting in general forged image detection. The second comparison reintroduced the knowledge of the CIFAR-10 image categories and performed a series of 10 independent 2-class evaluations within each image category. This is consistent with real-world practices in that the forged images within each category were necessarily presenting as an in-class forgery. Both of

these comparisons allow for an evaluation of general and applied use of ID-DNA Fingerprinting for forged image detection.

3.4. Evaluation Metrics

The metric used for both comparative assessments is the accuracy measure denoted $\%C$. Using the generally less rigorous $\%C$ metric is consistent with previous DNA works (Reising et al., 2010; Lopez et al., 2016; Rondeau et al., 2019) and is motivated by a desire to enhance broader, cross-discipline appreciation for the work. Furthermore, the arbitrary benchmark metric of $\%C > 90\%$ provides this analysis direct comparison to other DNA Fingerprinting works. This same 90% metric is consistent with the performance realized in neural network-based solutions for image forgery detection and will serve as the performance benchmark for all evaluations.

The $\%C$ assessments are augmented here using a generally more rigorous approach based on True Positive (TP), False Positive (FP) Type I error, and False Negative (FN) Type II error metrics commonly used in hypothesis testing. These metrics are generated from confusion matrix results and used to calculate the Precision and Recall measures given in Equations 8 and 9, respectively (James et al., 2017).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

Of particular note is that the confusion matrix $\%C$ equals the average of all individual per-class recall measurements. Given that the classification decisions represent independent Monte Carlo trials, 95% Confidence Interval analysis consistent with Leemis and Park (2005) is used for comparative assessments.

4. Results and Analysis

4.1. ID-DNA Fingerprinting Baseline

The first evaluation established the capability of applying DNA Fingerprinting to image classification using the real images in the CIFAR-10 dataset. The ID-DNA Fingerprints were input into the MDA/ML classifier and the results are shown in Table 2 with an average $\%C = 98.8\%$.

Referencing the arbitrary metric of $\%C = 90\%$, ID-DNA Fingerprinting has successfully adopted

		Declared Class									
		Plane	Auto	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck
True Class	Plane	2500	0	0	0	0	0	0	0	0	0
	Auto	0	2486	0	1	0	13	0	0	0	0
	Bird	0	0	2500	0	0	0	0	0	0	0
	Cat	0	0	0	2500	0	0	0	0	0	0
	Deer	0	0	0	0	2500	0	0	0	0	0
	Dog	0	1	0	0	0	2444	57	0	0	0
	Frog	0	0	0	0	0	62	2383	55	0	0
	Horse	0	5	0	0	0	0	54	2446	0	0
	Ship	0	0	0	0	0	0	0	2	2471	27
	Truck	0	0	0	0	0	0	0	0	39	2461

Table 2: Confusion Matrix for CIFAR-10 Dataset with 2,500 Images per Category with $\%C \approx 98.8\%$.

traditionally RF-based DNA Fingerprinting to images using the CIFAR-10 dataset. Of note, in Table 2 the main sources of confusion are among the image categories having intuitively common features (e.g., 2 eyes, 4 limbs). This includes the most confused categories between the Dog and Frog categories, both of which are animals that may present as visually similar images and may therefore have similar statistical distributions. The second most confused categories were Frog and Horse, both animals. Conversely, out of the total test set of 25,000 comparisons, images of animals and images of vehicles were only confused in 22 instances. The resulting $\%C = 98.8\% > 90\%$ on images as small as 32×32 RGB images also speaks to the robust nature of the ID-DNA Fingerprint to exploit features with relatively limited data. ID-DNA Fingerprinting produced results similar or better to neural-net based solutions using a simple process.

4.2. ID-DNA Fingerprinting Real vs. Fake

Having established sufficient motivation that DNA Fingerprinting may be applied to image classification in general, the next results quantify how well ID-DNA Fingerprinting identifies AI-generated forgeries. The initial investigation was to provide ID-DNA Fingerprints to a 2-class problem of real and fake images without image categories. These results quantify the general case of how well ID-DNA Fingerprinting may be used to detect image forgeries with absolutely no clarifying information about the image. The resulting confusion matrix is shown in Table 3.

The resulting average $\%C = 74.7\% < 90\%$ for this 2-class problem. The majority of the error in these results are from real images incorrectly classifying as fake. The corresponding precision and recall further illuminate the utility of ID-DNA Fingerprinting in this 2-class problem. Using the results from Table 3, the precision and recall values are shown in Table 4

Considering the five measures presented, the model

True	Declared	
	Real	Fake
Real	18041	6959
Fake	5712	19288

Table 3: Confusion Matrix for CIFAKE Image Dataset with 25,000 Images per Image Class with $\%C \approx 74.7\%$.

Measure	Value
Precision _{real}	76.0%
Recall _{real}	72.2%
Precision _{fake}	73.5%
Recall _{fake}	77.2%

Table 4: Precision and Recall for CIFAKE Real and Fake Image Set with 25,000 Images per Image Class.

performs better at identifying false images (indicated by recall of fake). However, the overall accuracy ($\%C$) and precision suggest degraded performance in the ability to successfully discriminate real images. While below the arbitrary $\%C > 90$ benchmark, this was the most complex problem set presented to the ID-DNA Fingerprinting process and classification. For example, unlike the previous case, this model compared real and fake/forged images of animals and vehicles. As was detailed in the previous section, when the knowledge of image categories was included in the process, the output demonstrated the ability to successfully minimize errors between visually distinct categories like animals and vehicles. While the $\%C$ accuracy results are clearly above a random guess, the precision and recall results motivate the need for additional information to better identify image forgeries. Commensurate with the success from real image classification from the baseline results, the next investigation adds back that additional piece of knowledge about the images, i.e., the categories from the CIFAR-10 categorization. This is also consistent with potential real-world applications where real images are always going to be presented as forgeries within some image category.

The results presented here are the aggregate of ten individual 2-class problems (i.e., real vs. fake) within each of the respective image categories from CIFAR-10. Table 5 displays these results and the corresponding precision and recall values, and Figure 3 displays the results graphically with the 95% Confidence Intervals.

With the addition of image categories, the results show a maximum recovery of $\%C_{\Delta} = +18\%C$ (Frog) and an average recovery of $\%C_{\Delta} = +6\%C$. With a relatively simple addition of the image categories,

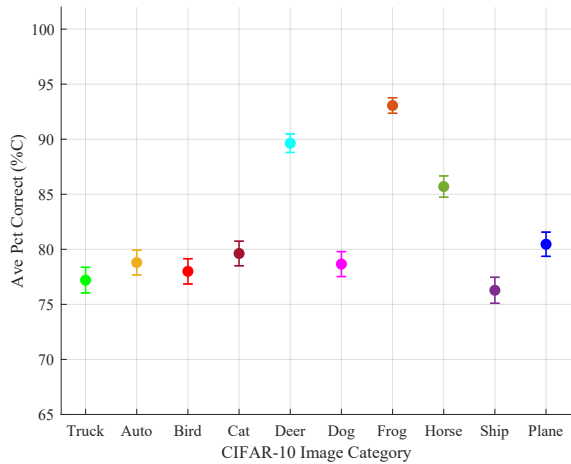


Figure 3: Accuracy (%C) for 10 Individual 2-Class Real/Fake ID-DNA Fingerprint Classification Assessments.

Category	%C	Precision	Recall
Truck	77.2%	81.6%	70.2%
Auto	78.8%	87.1%	67.6%
Bird	78.0%	81.0%	73.2%
Cat	79.6%	82.8%	74.7%
Deer	89.6%	89.8%	89.4%
Dog	78.7%	83.1%	71.9%
Frog	93.1%	92.6%	93.6%
Horse	85.7%	88.1%	72.6%
Ship	76.3%	79.5%	70.9%
Plane	80.5%	85.7%	73.2%

Table 5: Confusion Matrix for CIFAKE Image Dataset with 5,000 Images per Category with %C ≈ 81.8%.

the ID-DNA Fingerprinting successfully identified fake images with a %C = 81.8% when averaging across all 10 image categories using the same $N_F = 44$ features. Consistent with the 2-class problem, the precision values range from 79.5% to 96.6% and correspondingly recall is lower and ranges from 67.6% to 93.6%. This supports the conclusion that ID-DNA Fingerprints can be used successfully for image forgery detection. However, work still remains to balance the aspects of precision and recall to achieve better overall performance. The lower within-category recall values indicate the potential need for either additional features, the removal of confounding features, or a combination of both. Additional or different subregions (e.g., utilization of grayscale images) may be able to make use of the same statistical measures to improve overall performance as well. The lower within-category recall values motivate further investigation in that better recall

performance tended to correspond to higher %C. While not a complete and direct association, this is intuitively consistent that a better ability to identify real images within the image category results in better overall performance. Considering both measures leads to the recommendation to investigate additional methods to improve both precision and recall to boost performance.

While short of the arbitrary %C = 90%, ID-DNA Fingerprinting achieved performance within $\%C_{\Delta} = -9\%$ of the benchmark and some neural net-based solutions using only 44 intuitively traceable features (e.g., 2 eyes, 4 limbs). These results show promise for ID-DNA Fingerprinting for AI forgery detection. Notably, the results presented here were produced using a much more simple process than neural networks require and with similar computation times on the order of seconds without major computational resources (e.g., GPUs). Furthermore, these ID-DNA Fingerprinting results required no hyperparameter tuning, expert knowledge in algorithm development, and utilized a process with direct intuition of how the ID-DNA Fingerprint was created from the corresponding image. Given that forged images will necessarily present within-category, the improved performance realized (%C = 81.8% > 74.7%) with additional information of image categories, strongly support the utility for ID-DNA Fingerprinting in terms of both applicability and scalability. In summary, with easily traceable procedures, an algorithmically simple process, and a modest set of $N_F = 44$ features per image, ID-DNA Fingerprinting achieved results within $\%C_{\Delta} = -9\%$ of an arbitrary %C = 90% accuracy benchmark making it a promising and successful proof-of-concept to compete with neural-net based classification solutions.

5. Summary and Conclusions

Image forgery is becoming progressively more prevalent and difficult to detect with advances in AI image generation. As such, the usefulness — and even requirement — for detection techniques that are affordable (computationally and monetarily) as well as simple from an intuitive stand point are equally increasing. This work demonstrated the ability to adopt DNA Fingerprinting for image classification with %C = 98.8%, and subsequently for AI-forgery with an average %C = 81.8%. The ID-DNA Fingerprinting process is less resource extensive, requiring equal or shorter computation time, money, and knowledge to correctly classify images. Therefore, the preliminary expense of approximately $\%C_{\Delta} = -9\%$ accuracy is counterbalanced by the holistic computational

and resource efficiencies. These initial results of AI-forged image classification demonstrates potential for additional improvements after further refinement of the feature selection process.

The ID-DNA image processing outlined in this paper lays the foundation for future work. The next-step demonstration activity would build upon the demonstrated improvements here and address improved resource efficiency. Recommended future work includes:

- Exploration of additional feature development for ID-DNA Fingerprints using the %C = 90% benchmark real-only classification to determine which features were the most useful in image discrimination
- Investigation into the proper utilization of traditional and L-moments as statistical methods for ID-DNA Fingerprints
- Utilization of additional ID-DNA subregions for additional feature creation and classifier exploitation
- Evaluate the use of additional classifiers beyond MDA/ML such as Random Forest to additionally support feature relevancy assessment
- Replicate the work here with larger images to demonstrate scalability.

The application of DNA Fingerprinting via the introduction of ID-DNA Fingerprinting demonstrated its usefulness and success as a method for image classification and for image forgery detection. With the results presented here, future improvements can secure the holistic resource improvements while further maximizing the classification performance.

6. Acknowledgement

The views in this paper are those of the authors and do not reflect the official policy or position of the Air Force Institute of Technology, the Department of the Air Force, the Department of Defense, or the US Government. This paper is approved for public release, Case #: 88ABW-2024-0502.

References

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

- Bird, J. J., & Lotfi, A. (2024). CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images. *IEEE Access*, *12*, 15642–15650.
- Gill, N. K., Garg, R., & Doegar, E. A. (2017). A Review Paper on Digital Image Forgery Detection Techniques. *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–7.
- Göring, S., Ramachandra Rao, R. R., Merten, R., & Raake, A. (2023). Appeal and Quality Assessment for AI-Generated Images. *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, 115–118.
- Guttman, N. B. (1993). The Use of L-Moments in the Determination of Regional Precipitation Climates. *Journal of Climate*, *6*(12), 2309–2325.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *CoRR*, *abs/1512.03385*.
- Hosking, J. R. M. (1990). L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, *52*(1), 105–124.
- Hossain, M. Z., Uz Zaman, F., & Islam, M. R. (2023). Advancing AI-Generated Image Detection: Enhanced Accuracy through CNN and Vision Transformer Models with Explainable AI Insights. *2023 26th International Conference on Computer and Information Technology (ICCIT)*, 1–6.
- Jain, I., & Goel, N. (2021). Advancements in Image Splicing and Copy-move Forgery Detection Techniques: A Survey. *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 470–475.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning with Applications in R*. Springer.
- Jwaid, M. F., & Baraskar, T. N. (2017). Study and Analysis of Copy-Move & Splicing Image Forgery Detection Techniques. *IEEE International Conference on Computational Intelligence and Computing Research*.
- Klein, R. W., Temple, M. A., & Mendenhall, M. J. (2009). Application of wavelet-based RF fingerprinting to enhance wireless network security. *Journal of Communications and Networks*, *11*(6), 544–555.
- Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images.

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *ACM*, 60(6), 84–90.
- Lande, K. J. (2023). Seeing and visual reference. *Philosophy and Phenomenological Research*, 106(2), 402–433.
- Leemis, L. M., & Park, S. K. (2005). *Discrete-Event Simulation: A First Course*. Prentice-Hall, Inc.
- Lopez, J., Temple, M. A., & Mullins, B. E. (2016). Exploitation of HART WS-DNA Features to Verify Field Device Identity and Infer Operating State. 8985, 24–30.
- Lukacs, M. W., Zeqolari, A. J., Collins, P. J., & Temple, M. A. (2015). “RF-DNA” Fingerprinting for Antenna Classification. *IEEE Antennas and Wireless Propagation Letters*, 14, 1455–1458.
- Monkam, G., Xu, W., & Yan, J. (2023). A GAN-based Approach to Detect AI-Generated Images. 2023 26th ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), 229–232.
- Mushtaq, S., & Mir, A. H. (2014). Digital Image Forgeries and Passive Image Authentication Techniques: A Survey. *International Journal of Advanced Science and Technology*, 73, 15–32.
- Nash, R. A., Hanczakowski, M., & Mazzoni, G. (2015). Eyewitness Testimony. In J. D. Wright (Ed.), *International Encyclopedia of the Social Behavioral Sciences (Second Edition)* (Second Edition, pp. 642–649). Elsevier.
- Passos, L. A., Jodas, D., Costa, K. A. P., Souza Júnior, L. A., Rodrigues, D., Del Ser, J., Camacho, D., & Papa, J. P. (2024). A Review of Deep Learning-Based Approaches for Deepfake Content Detection. *Expert Systems*, n/a(n/a), e13570.
- Poredi, N., Nagothu, D., & Chen, Y. (2024). Authenticating AI-Generated Social Media Images Using Frequency Domain Analysis. 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC), 534–539.
- Reising, D. R., Temple, M. A., & Mendenhall, M. J. (2010). Improving Intra-Cellular Security Using Air Monitoring with RF Fingerprints. 2010 IEEE Wireless Communication and Networking Conference, 1–6.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10674–10685.
- Rondeau, C. M., Temple, M. A., & Betances, J. A. (2019). Dimensional Reduction Analysis for Constellation-Based DNA Fingerprinting to Improve Industrial IoT Wireless Security.
- Rondeau, C. M., Temple, M. A., Betances, J. A., & Schubert Kabban, C. M. (2021). Extending Critical Infrastructure Element Longevity Using Constellation-Based ID Verification. *Computers Security*, 100, 102073.
- Self, J. W. (2005). The First Debate over the Debates: How Kennedy and Nixon Negotiated the 1960 Presidential Debates. *Presidential Studies Quarterly*, 35(2), 361–375.
- Shailaja Rani, P. B., & Kumar, A. (2019). Digital Image Forgery Detection Techniques: A Comprehensive Review. 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 959–963.
- Vargo, E., Pasupathy, R., & Leemis, L. (2010). Moment-Ratio Diagrams for Univariate Distributions. *Journal of Quality Technology*, 42.
- Williams, M. D., Temple, M. A., & Reising, D. R. (2010). Augmenting Bit-Level Network Security Using Physical Layer RF-DNA Fingerprinting. 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, 1–6.
- Zhang, X. (2021). The AlexNet, LeNet-5 and VGG NET applied to CIFAR-10. 2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), 414–419.