AUSOME: authenticating social media images using frequency analysis

Nihal Poredi, Deearj Nagothu, Yu Chen

Nihal Poredi, Deearj Nagothu, Yu Chen, "AUSOME: authenticating social media images using frequency analysis," Proc. SPIE 12542, Disruptive Technologies in Information Sciences VII, 125420A (15 June 2023); doi: 10.1117/12.2663296

SPIE.

Event: SPIE Defense + Commercial Sensing, 2023, Orlando, Florida, United States

AUSOME: Authenticating Social Media Images using Frequency Analysis

Nihal Poredi[a], **Deeraj Nagothu**[a], **Yu Chen**[a,*]
[a]Dept. of Electrical and Computer Engineering, Binghamton University, Binghamton, NY 13902

ABSTRACT

Ever since human society entered the age of social media, every user has had a considerable amount of visual content stored online and shared in variant virtual communities. As an efficient information circulation measure, disastrous consequences are possible if the contents of images are tampered with by malicious actors. Specifically, we are witnessing the rapid development of machine learning (ML) based tools like DeepFake apps. They are capable of exploiting images on social media platforms to mimic a potential victim without their knowledge or consent. These content manipulation attacks can lead to the rapid spread of misinformation that may not only mislead friends or family members but also has the potential to cause chaos in public domains. Therefore, robust image authentication is critical to detect and filter off manipulated images. In this paper, we introduce a system that accurately **AU**thenticates **SO**cial **ME**dia images (AUSOME) uploaded to online platforms leveraging spectral analysis and ML. Images from DALL-E 2 are compared with genuine images from the Stanford image dataset. Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT) are used to perform a spectral comparison. Additionally, based on the differences in their frequency response, an ML model is proposed to classify social media images as genuine or AI-generated. Using real-world scenarios, the AUSOME system is evaluated on its detection accuracy. The experimental results are encouraging and they verified the potential of the AUSOME scheme in social media image authentications.

Keywords: Digital Media Authentication, Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), DALL-E 2, Deep Neural Networks (DNN).

1 Introduction

The popularity of social media networks has enabled wide demographics of social user interaction along with information broadcast and exchange. Social networks have inadvertently become a primary medium of information gathering, where information is neither curated appropriately nor checked for factual integrity.[1] Even when the media companies attempt to minimize false information like health regulations or political campaigns, the content mutation makes it difficult to create sufficient and appropriate rules in a timely manner.[2] Recently, the evolution of digital media generation using Artificial Intelligence (AI) has gained traction in content generation. Generative technologies like Generative Adversarial Networks (GANs) or Generative Pre-trained Transformer (GPT) along with appropriate training data, a user can generate any media using simple prompts resulting in the widespread dissemination of misinformation.[3]

AI-generated images have been a recent but rapidly growing phenomenon.[4] From advertisement to art created in social media, AI-generated images have found applications in an increasing number of information platforms. However, this success has created its fair share of security risks

Disruptive Technologies in Information Sciences VII, edited by Misty Blowers, James Holt, Bryant T. Wysocki, Proc. of SPIE Vol. 12542, 125420A
 · 0277-786X · doi: 10.1117/12.2663296

to users on and off the Internet. The GAN models can leverage publicly available information like human faces and skeleton structures to automatically regenerate artificial humans in a realistic scenario.[5] Applications with simple user interface masking the complex neural network model to create fake imitations like DeepFakes are easily made available to the public at no cost.[6] The rise in fake artificial media content is tackled by authentication and verification measures.[7] Visual layer attacks primarily focused on altering the perception of events captured,[8,9] spatiotemporal attacks to alter the image-level content and followed by AI-generated attacks[10,11] have become primary research problems in ensuring safe Internet access. Although the advanced detection techniques are primarily focused on media content like audio and video recordings, the current social media applications are driven by overflowing image content, where we lack fast and reliable techniques to detect fake images.

Prior versions of AI/ML generated images often lack image-level content consistency like warped faces and inconsistent background.[12] An increase in training data, advanced ML models, and high-performing computational power led to the generation of images that do not carry any visual traces of inconsistencies. Even mobile applications like Snapchat can generate fake images in an instant using mobile hardware, where a user can swap faces with any celebrity.[13] While the users are provided with a software to generate images, some perpetrators use such technologies to spread fake information throughout social networks resulting in the spread of misinformation.

Since it is not an easy task for common users to tell a conventional image from an AI-generated one, this technology poses threats such as fake news leading to public unrest. As such, there is a growing need to develop the ability of everyday users to detect AI-generated images. Backed with billions of trainable parameters in modern deep learning architectures, text-to-speech and text-to-image models using the GPT framework have made tremendous improvements in accuracy and speed. Such technologies are further improved by providing public access and as a result, a popular text-to-image model called Dall-E gained popularity.[14] Dall-E 2 was made open to the public in September 2022,? the predecessor of which was Dall-E, first introduced in January 2021 by OpenAI. This tool takes text as input and generates a corresponding image. Dall-E 2 has improved upon Dall-E in the level of realism of the output images. This has led to concerns regarding the ability of malicious actors to misuse it to spread fake news.

To minimize fake content, prior efforts have been made to detect artificial images by training a neural network from the available fake data. However, as the generator evolves with more training data, the result keeps getting improvised, and thereby the previous detection schemes aren't effective anymore. By training the models just on the fake images directly, we inadvertently create a never-ending scheme of detectors and generators where both parties never reach an equilibrium. Instead, we focus on the underlying frequency changes that are unique when compared to real images.

In this paper, we examine the Dall-E 2 images closely and compare them with conventionally generated images using various frequency estimation techniques. The differences are leveraged to train a machine learning model such that it can classify social media messages as either real or AI-generated. This tool will empower common users to tell the difference between the two types of pictures. A typical scenario could be a user coming across a tweet about a certain security lapse occurring at a public place with a related picture that was in fact generated using Dall-E 2. In such a case, the user could leverage AUSOME to verify the authenticity of the image. If found to be

AI-generated, the user could seek more information and could alert other users about the fake post. This could avoid people from panicking and causing chaos.

The rest of the paper is organized as follows. The background knowledge and related work are presented in Section 2. Section 3 highlights the features captured using DFT and DCT technologies in AI/ML generated images and genuine images. The proposed AUSOME image classification system is presented in Section 4. Finally, Section 5 presents our conclusion from the proposed work and discusses ongoing efforts.

2 Background Knowledge and Related Work

2.1 AI Generated Images

AI-generated art has come a long way from the classic convolutional neural networks to the modern transformer and diffusion models. Being capable of generating media using a user-defined input, by considering different contexts and media, the AI-generated media has reached a point where it is harder to distinguish the difference.[15] The evolution in creative neural networks like Google's DeepDream,[16] image art translation pix2pix[17] and Nvidia's GauGan[18] has augmented artificial image applications to a certain point where the generated images are sold in auctions.

The introduction of the Transformer models learns context among the given data by tracking the relationships in sequential data. For example, learning the relation between a series of words given in training to understand the given context and the situation they are used in. The image generation model Dall-E[19] was introduced by OpenAI. It used a Transformer which models the text and the image tokens given during training and optimizes the text-to-image token translation. When a text description is given, Dall-E can predict the image tokens and decode them into an image.

By integrating the diffusion pipeline, it is shown that the diffusion models outperform the GANs on image generation. By combining the transformer encoder for text input and the diffusion model for image synthesis, the quality of the generated images is vastly improved compared to that of the traditional Dall-E. As a result, a new model, Dall-E 2[20] is introduced with significantly better output generation along with self-attention in the text-based input. Followed by the introduction of Dall-E 2 models, now several other text-to-image translation models are being created to optimize the performance and create photo-realistic images that are indistinguishable from the real ones.

With the availability of such high-performing image generation tools, where in most cases the tools are available free of cost, the users actively generate more images for personal use. The user interaction also improves the model performance where the data is collected on the type of text input given and the image generated. A sudden influx of artificial images has flooded online media-sharing platforms, leaving consumers unaware of the origins of the media. For the purposes of entertainment, such generated media has gained a lot of popularity and appreciation from the audience but artificially generated images with malicious intent could have adverse consequences. A need for distinguishability between artificial and real images inspired our work in creating the AUSOME model.

2.2 *Spectral Analysis of Images*

Spectral analysis of digital images is a group of image processing techniques that are used to decompose an image into its sine and cosine components. Each of the sine and cosine components belongs to different frequencies which together represent the image in the frequency domain without any information loss. Discrete Fourier Transform (DFT),[21] Discrete Cosine Transform (DCT),[22] and Discrete Wavelet Transform (DWT)[23] are some of the most common frequency analysis methods which are used in image frequency analysis. In this work, we use DFT and DCT to convert digital images from the spatial domain to the frequency domain and analyze the various frequencies present in them to give us an understanding of how they are produced.

2.2.1 *2D Discrete Fourier Transform*

The DFT of an image is essentially a 2D FFT that contains a set of frequencies large enough to completely describe the spatial domain of the image. The FFT along the horizontal axis is first computed, which is then processed along the vertical axis thus producing a bidimensional sequence. The 2D DFT of an $M \times N$ image is calculated as follows:

$$F(a,b) = \Sigma_{x=1}^{N-1}\Sigma_{y=1}^{M-1} f(x,y)e^{-j2\pi(\frac{ax}{N}+\frac{by}{M})}$$

where the exponential component of the formula is the basis function that describes the Fourier domain space, and f(x,y) represents a pixel of the image in the spatial domain. Thus, each point of the DFT is calculated as the product of the image spatial component and the basis function corresponding to the DFT point. The basis functions are nothing but sine and cosine waves at incremental frequencies that describe the complete Fourier domain space. The main purpose of calculating the DFT of an image is to understand its geometric features. The DFT helps us discover the influence of various frequencies on the spatial geometry of the image. The frequencies in the DFT increase as one moves away from the center of the DFT plot. The real part of the DFT describes the influence of each frequency, while the imaginary part describes the phase of each frequency.

2.2.2 *2D Discrete Cosine Transform*

Similar to the 2D DFT, the 2D DCT decomposes an image into sine and cosine components. However, unlike the DFT, the DCT is composed of purely real-valued coefficients. This is an advantage for applications like AUSOME because complex numbers add another level of complexity, which increases the computational overhead. The 2D DCT of an $M \times N$ image is given as follows:

$$F(u,v) = a(u)a(v)\Sigma_{x=0}^{N-1}\Sigma_{y=0}^{M-1} f(x,y)\gamma(x,y,u,v)$$

where

$$\gamma(x,y,u,v) = cos(\frac{\Pi(2x+1)u}{2N})cos(\frac{\Pi(2y+1)v}{2M})$$

and

$$a(u) = \begin{cases} \sqrt{(\frac{1}{N})}, u = 0 \\ \sqrt{(\frac{2}{N})}, u \neq 0 \end{cases} \quad (1)$$

Fig 1 : (a) Example Image and its DCT, (b) 2D DCT basis functions.

Figure 1b depicts the basis functions characteristic of a 2D DCT. Each basis function is a combination of two cosines at different frequencies. The frequencies increase from top to bottom and left to right. A useful feature of the DCT is its energy compaction property which allows for the preservation of low-frequency information as shown in Figure 1a.

2.2.3 Noise Residuals

The digital image acquisition system in modern cameras introduces noise artifacts in the captured samples. As a result, the acquisition device inadvertently creates a digital footprint on each image taken. The signals remaining after the high-level semantic content has been removed using the denoising algorithms, high-pass filters, or transform is called as Noise residual.[24] Exploiting the noise residual to trace the origins of captured images follows a data-driven approach for image analysis since a large number of examples are required to create a reliable Noise residual model.

The noise residual profile is capable of identifying the intra-frame noise consistency and thereby helps with the forensic evaluation of a digital image. The most popular and initial usage of noise residual included the generation of Photo-Response Non-Uniformity (PRNU) for source device identification.[25] Each imaging sensor imprints unique noise residuals, which upon averaging results in PRNU fingerprints. For digital images taken from a camera with a PRNU fingerprint, the image can be traced back to its origin device. However, the main drawbacks of PRNU fingerprint involve the requirement of a large number of images from the camera for good estimates and low signal-to-noise ratio (SNR). Contrary to the dependence on spatial fingerprint, the Noiseprint technique utilizes the data-driven method to extract camera model fingerprint.[26] With the help of deep-learning models, the spatial dependency of the noise residual is exploited by training on images from the same camera and temporally linked image patches. As a result, the model can generate similar noise print even without any information on the source device and can be better for forensic applications against image forgeries.

The images captured using digital devices and the images generated from artificial networks like Dall-E produce different forms of noise residual since both the acquisition systems rely on different noise sources. In physical devices, the noise is introduced from the imaging sensor in the device, but for artificial networks, the produced images lack consistency after the semantic information has been removed. The following section will discuss the noise residual application to compare the images from both domains.

2.3 Related work

The use of frequency analysis to detect AI-generated images has been proven to be successful in the case of GANs.[27,28] The fingerprints discovered after frequency analysis were found to be a function of the GAN architecture and parameter configuration. The fingerprint extraction made use of noise residuals obtained using denoising.[24] They were then processed using DFT to obtain their mean spectra. All GANs exhibited notable peaks in their spectra,[29] suggesting the presence of patterns, which were the fingerprints used for their detection. Certain diffusion models such as GLIDE[30] and Stable Diffusion[31] were also shown to exhibit similar fingerprints. However, the same methods were found to be less effective on newer diffusion models like Dall-E 2.[32]

Artificially modified media recordings like DeepFake consists of additional noise where the static background is modified along with the targeted facial manipulations.[11] However, for continuous recordings like audio and video recordings, environmental fingerprint-based techniques like Electrical Network Frequency (ENF) proved effective for detecting any forgery.[10,33] Since, the underlying fingerprint ENF is randomly fluctuating and unique, so as a result, it is harder for any adversary to recreate the fluctuation patterns.[34] ENF is effective for continuous recordings since the targeted frequency requires a minimum sampling rate, which is challenging for single image-based frequency analysis and renders ENF-based detection a very challenging problem. However, frequency analysis using the Fourier transform has been successful in detecting GAN fingerprints and has been successful in detecting some diffusion model fingerprints.

3 Artifact Analysis

While the Dall-E 2 images in the spatial domain resemble genuine images, they exhibit certain artifacts in the frequency domain. To analyze the differences between genuine and Dall-E 2 images, we performed a frequency analysis of 400 images of cars generated by Dall-E 2, and 400 images of cars obtained from the Stanford cars dataset,[35] which contained real images of cars. Both datasets were analyzed using Discrete Fourier Transform and Discrete Cosine Transform as described in the following subsections.

Algorithm 1 Calculating the mean DFT of a dataset

$R_{total} = 0$
while $n \leq 400$ **do** ▷ 400 images per dataset
  $J_n = I_n + AWGN$ ▷ Corrupt image with AWGN
  $J_n \xrightarrow{filter} K_n$ ▷ Apply gaussian filter
  $R_n = I_n - K_n$ ▷ Noise residual of each image
  $R_{total} = R_{total} + R_n$ ▷ Total Noise residual of the dataset
end while
$R_{mean} = \frac{R_{total}}{400}$ ▷ Mean Noise residual of the dataset
$F = 2DFFT(R_{mean})$ ▷ 2D DFT calculation
$mean\, magnitude\, response = Real(F)$
$mean\, phase\, response = Angle(F)$

3.1 Discrete Fourier Transform based Analysis

As highlighted in Algorithm 1, 400 images from each of the datasets were first corrupted with Additive White Gaussian Noise (AWGN). The resulting noisy image was then passed through a denoising Gaussian filter to obtain the filtered image. The noise residual for each image was then calculated as the difference between the original and filtered image as further illustrated in Figure 2. The noise residuals of all images from each dataset were then averaged to get two mean noise residuals; one corresponding to the 400 Dall-E 2 images, and the other corresponding to the 400 genuine images. The mean frequency response of both datasets was then obtained by applying a 2D FFT to both noise residuals.

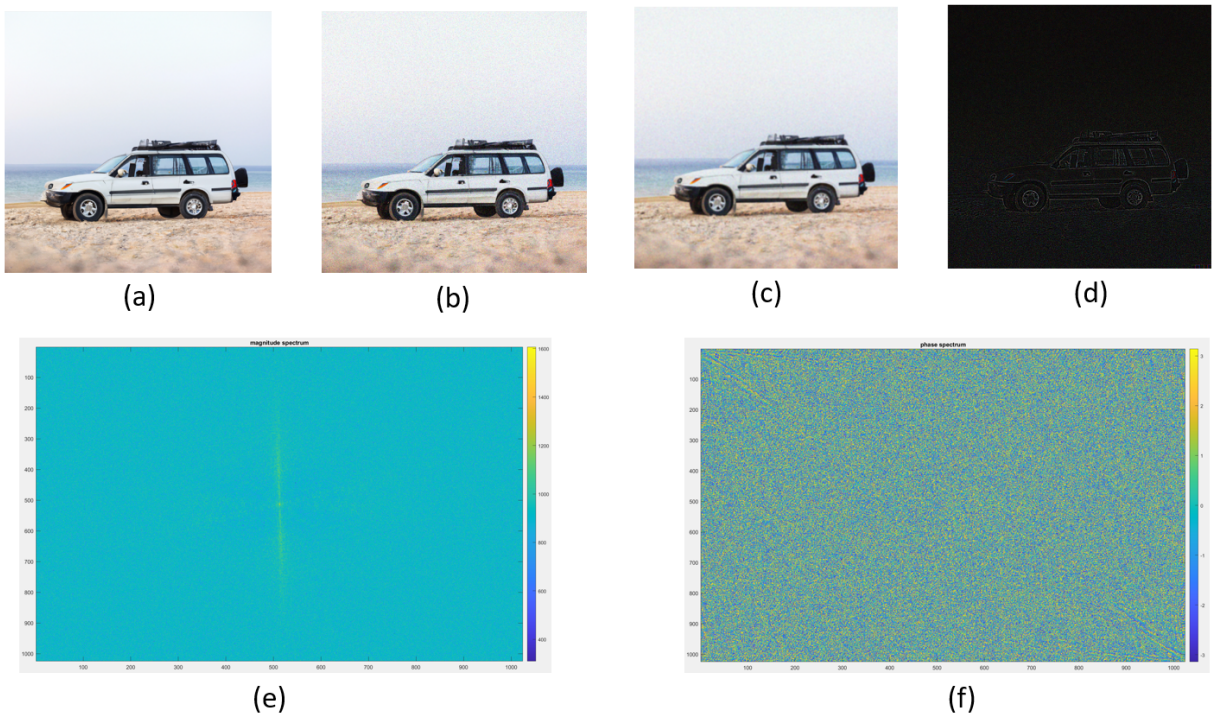


Fig 2 : (a) Dall-E 2 image, (b) Image with noise, (c) Filtered image, (d) Noise residual, (e) Magnitude response, (f) Phase response.

Figure 3 shows the difference between the magnitude responses of both datasets. As is evident, the magnitude response of the Dall-E 2 dataset exhibits characteristic visible patterns in the higher frequency regions which are absent in the Stanford cars dataset. This can be attributed to the interference of periodic cosines in the vertical and horizontal directions of the Dall-E 2 images. We notice that while most of the information is concentrated at low frequencies represented by the white spot at the center of each DFT, there is a pattern of bands at the higher frequencies for the Dall-E 2 dataset. This is clearly an artifact characteristic of the process of generation of a Dall-E 2 image. Since the images from the Stanford dataset are genuine, there is no definite pattern characteristic to their generation, resulting in a pattern-less mean DFT as depicted in Figure 3c.

Also evident from Figures 3(b) and Figure 3 (d) is the difference in the mean phase response of the two datasets. While the Stanford cars dataset yields a plain mean phase response, the Dall-E

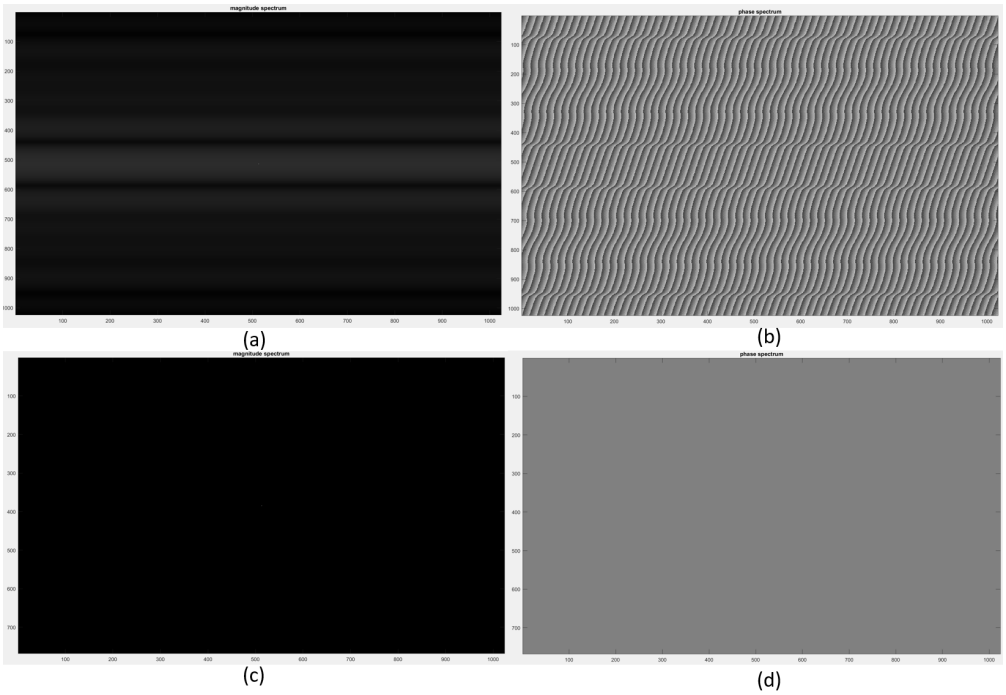


Fig 3 (a) Magnitude Spectrum of Dall-E 2 images, (b) Phase Spectrum of Dall-E 2 images, (c) Magnitude Spectrum of Stanford cars images, (d) Phase Spectrum of Stanford cars images.

2 dataset yields a textured mean phase response with a distinct periodic pattern. This points to a certain structural similarity between images from Dall-E 2 arising from the common underlying processes used in their synthesis.

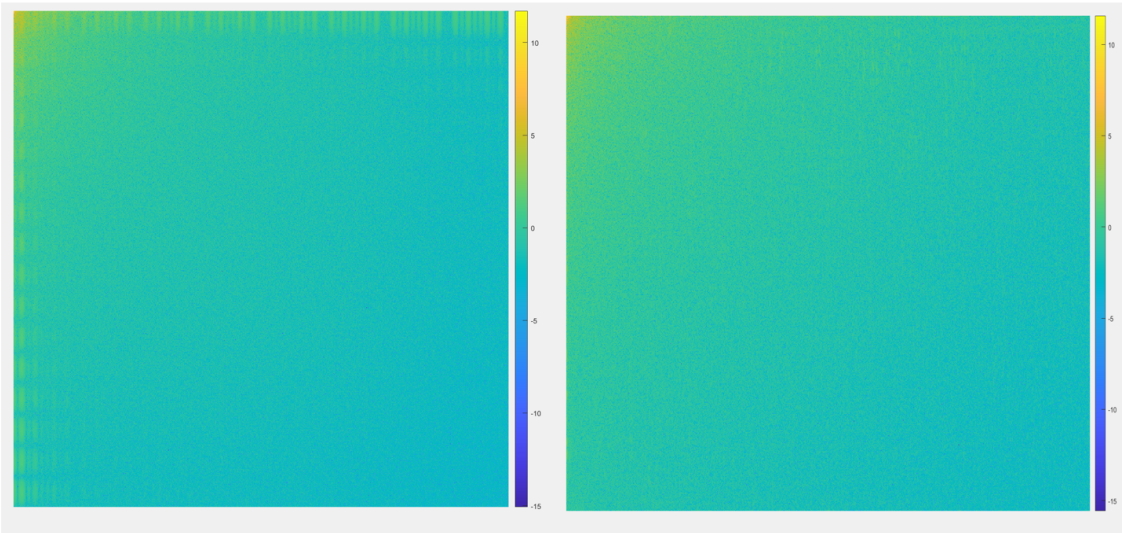
3.2 Discrete Cosine Transform based Analysis

The same datasets of 400, Dall-E 2 images and Stanford cars images were applied in the analysis performed using the DCT. They were first converted to grayscale after which the 2D DCT was applied to each image. The images were analyzed individually as well as using mean values.

Figure 4 shows the DCTs of images from both datasets. Upon close examination, finger-like patterns are observed in the higher frequency regions(top right and bottom left corners) of the Dall-E 2 image, while they are absent from the Stanford cars image. In fact, finger-like patterns were observed in the same regions for all Dall-E 2 images. Some of them had higher intensity along the vertical axis, while others had a higher intensity pattern along the horizontal axis. In complete contrast, no such patterns were characteristic of any of the genuine images from the Stanford cars dataset.

To confirm this phenomenon, the mean DCTs were calculated by averaging the DCTs across each dataset. Figure 5 shows that the mean DCT of the Dall-E 2 dataset shows clear grid-like patterns in the regions of high frequencies. Quite clearly, the mean DCT of the Stanford cars dataset shows no such pattern. This suggests that the grid-like pattern is indeed an artifact unique to images generated artificially by Dall-E 2. This again can be attributed to the similar ways in which different images are synthesized due to common underlying construction processes.

Fig 4 Row 1: Dall-E Image and its DCT. Notice the finger-like patterns at the higher frequencies. Row 2: Real Image and its DCT.



(a) (b)

Fig 5 (a) Mean DCT over 400 Dall-E 2 images, (b) Mean DCT over 400 Stanford cars images.

4 The AUSOME Scheme

The features observed in the frequency domain with DFT and DCT make it possible to detect AI-generated images from the ones taken by a real camera. Inspired by this observation, we introduce a scheme that is able to accurately **AU**thenticate **SO**cial **ME**dia images (AUSOME) uploaded to

online platforms leveraging spectral analysis and ML. Figure 6 illustrates the working flow of the proposed AUSOME scheme that classifies an image as either DALL-E 2 generated, or genuine. A given input image is first processed to obtain its corresponding DCT map, and then the DCT map image is classified by the CNN model as belonging to DALL-E 2 class or the genuine image class. Efforts are ongoing to expand the scale of the training data, as well as the complexity of the CNN model to make it a more robust and versatile authentication tool, which could be used to identify other AI-generated images, in addition to those produced using DALL-E 2.

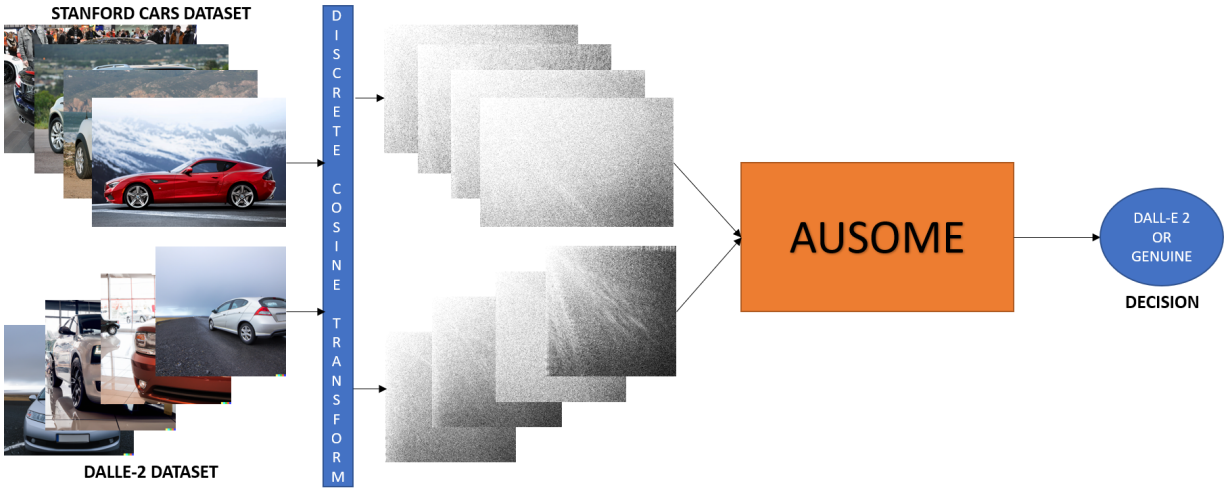


Fig 6 AUSOME system workflow.

The initial dataset comprised 400 images of cars generated using DALL-E 2, and 400 images of cars selected from the Stanford cars dataset. The Stanford cars dataset is a collection of authentic images of cars captured using cameras. From the previous sections, we observe that while the mean DFT shows a clear distinction between the two categories, the individual DFTs do not show an obvious difference. On the other hand, the individual DCTs do show discernible differences between the two classes. Therefore, in this work, the DCTs of all 800 images are used as the training dataset. The training data was then normalized to improve the CNN model.

Layer	Output Shape	# of Params
Flatten	(0, 120000)	0
Dense	(0, 128)	15360128
Dense	(0, 1)	129

Table 1 AUSOME model architecture.

As outlined in Table 1, the AUSOME model begins with a Flatten layer that flattens the input image. The flattened image is then fed to a dense layer (fully connected layer), with 128 hidden units. To end, because the ultimate aim is binary classification, the last layer is a sigmoid, which enables the output of the model to be a single scalar between zero and one; zero being a DALL-E 2 image, and one being a genuine image. The resulting model is then trained on the dataset for several epochs until the validation accuracy is close to 100%. The AUSOME model took 100

epochs to reach a validation accuracy close to 98% which is high enough considering the limited scale of the training data. This is further clarified by the ROC curve depicted in Figure 7, the area under which is very close to one indicating high accuracy.

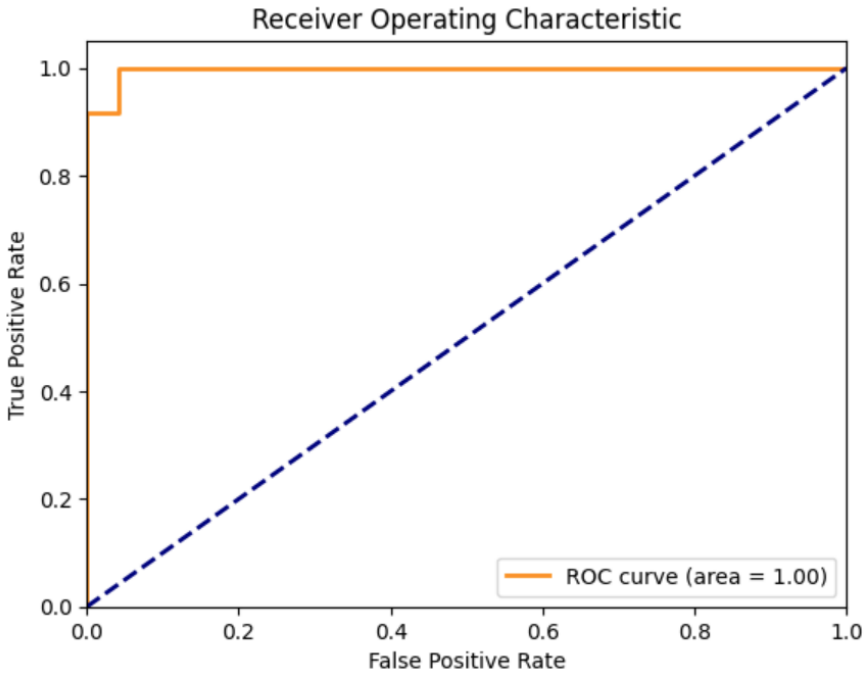


Fig 7 ROC curve for the AUSOME model.

5 Conclusions

This paper underscores the compelling need for image authentication in the era of social media. The threat posed by misinformation is highlighted, and existing techniques for the authentication of well-known AI image models are discussed. A discussion on the latest text-to-image generator DALL-E 2 is included in this paper, which is based on diffusion models. In addition, frequency analysis techniques that bring out the spectral information contained in images are highlighted. Specifically, this work focuses on the Discrete Fourier Transform and Discrete Cosine Transform for image frequency analysis. The Dall-E 2 images are then compared with their genuine counterparts using both techniques, and the artifacts characteristic to them are formalized. The artifacts in the case of DFT analysis are periodic patterns in the magnitude and phase response of the Dall-E 2 dataset. During analysis using DCT, clear grid like patterns are observed in the high frequency regions of the AI generated images. These artifacts provide a strong basis for the formulation of a scheme that differentiates AI generated images from genuine images. Finally, the authentication scheme, AUSOME is introduced, which is based on the Discrete Cosine Transform. Its architecture and workflow are described, and its performance is analyzed. The fact that AUSOME achieves close to 98% accuracy, and a ROC curve area of close to one, makes it a promising tool for detecting Dall-E 2 images. However, efforts are being made to expand the training data to include images from more AI image generators and improve the accuracy and versatility of the model for general use.

References

1 L. Wu, F. Morstatter, K. M. Carley, *et al.*, "Misinformation in social media: definition, manipulation, and detection," *ACM SIGKDD explorations newsletter* **21**(2), 80–90 (2019).

2 J. Shin, L. Jian, K. Driscoll, *et al.*, "The diffusion of misinformation on social media: Temporal pattern, message, and source," *Computers in Human Behavior* **83**, 278–287 (2018).

3 S. Chen, L. Xiao, and A. Kumar, "Spread of misinformation on social media: What contributes to it and how to combat it," *Computers in Human Behavior* , 107643 (2022).

4 Y. Cao, S. Li, Y. Liu, *et al.*, "A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt," *arXiv preprint arXiv:2303.04226* (2023).

5 T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4401–4410 (2019).

6 I. Perov, D. Gao, N. Chervoniy, *et al.*, "Deepfacelab: Integrated, flexible and extensible face-swapping framework," *arXiv preprint arXiv:2005.05535* (2020).

7 L. Verdoliva, "Media forensics and deepfakes: an overview," *IEEE Journal of Selected Topics in Signal Processing* **14**(5), 910–932 (2020).

8 D. Nagothu, Y. Chen, E. Blasch, *et al.*, "Detecting malicious false frame injection attacks on surveillance systems at the edge using electrical network frequency signals," *Sensors* **19**(11), 2424 (2019).

9 D. Nagothu, J. Schwell, Y. Chen, *et al.*, "A study on smart online frame forging attacks against video surveillance system," in *Sensors and systems for space applications XII*, **11017**, 176–188, SPIE (2019).

10 D. Nagothu, R. Xu, Y. Chen, *et al.*, "Defake: Decentralized enf-consensus based deepfake detection in video conferencing," in *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*, 1–6, IEEE (2021).

11 D. Nagothu, R. Xu, Y. Chen, *et al.*, "Defakepro: Decentralized deepfake attacks detection using enf authentication," *IT Professional* **24**(5), 46–52 (2022).

12 W. Zhang and C. Zhao, "Exposing face-swap images based on deep learning and ela detection," in *Proceedings*, **46**(1), 29, MDPI (2019).

13 S.-Y. Wang, O. Wang, A. Owens, *et al.*, "Detecting photoshopped faces by scripting photoshop," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10072–10081 (2019).

14 A. Ramesh, M. Pavlov, G. Goh, *et al.*, "Dall· e: Creating images from text," *OpenAI blog* (2021).

15 A.-S. Maerten and D. Soydaner, "From paintbrush to pixel: A review of deep neural networks in ai-generated art," *arXiv preprint arXiv:2302.10913* (2023).

16 A. Mordvintsev, C. Olah, and M. Tyka, "Inceptionism: Going deeper into neural networks," (2015).

17 P. Isola, J.-Y. Zhu, T. Zhou, *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134 (2017).

18 T. Park, M.-Y. Liu, T.-C. Wang, *et al.*, "Gaugan: semantic image synthesis with spatially adaptive normalization," in *ACM SIGGRAPH 2019 Real-Time Live!*, 1–1 (2019).
19 A. Ramesh, M. Pavlov, G. Goh, *et al.*, "Zero-shot text-to-image generation," in *International Conference on Machine Learning*, 8821–8831, PMLR (2021).
20 A. Ramesh, P. Dhariwal, A. Nichol, *et al.*, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125* (2022).
21 S. Winograd, "On computing the discrete fourier transform," *Mathematics of computation* **32**(141), 175–199 (1978).
22 N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE transactions on Computers* **100**(1), 90–93 (1974).
23 C. E. Heil and D. F. Walnut, "Continuous and discrete wavelet transforms," *SIAM review* **31**(4), 628–666 (1989).
24 K. Zhang, W. Zuo, Y. Chen, *et al.*, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing* **26**(7), 3142–3155 (2017).
25 J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security* **1**(2), 205–214 (2006).
26 D. Cozzolino and L. Verdoliva, "Noiseprint: A cnn-based camera model fingerprint," *IEEE Transactions on Information Forensics and Security* **15**, 144–159 (2019).
27 F. Marra, D. Gragnaniello, L. Verdoliva, *et al.*, "Do gans leave artificial fingerprints?," in *2019 IEEE conference on multimedia information processing and retrieval (MIPR)*, 506–511, IEEE (2019).
28 N. Yu, L. S. Davis, and M. Fritz, "Attributing fake images to gans: Learning and analyzing gan fingerprints," in *Proceedings of the IEEE/CVF international conference on computer vision*, 7556–7566 (2019).
29 X. Zhang, S. Karaman, and S.-F. Chang, "Detecting and simulating artifacts in gan fake images," in *2019 IEEE international workshop on information forensics and security (WIFS)*, 1–6, IEEE (2019).
30 A. Nichol, P. Dhariwal, A. Ramesh, *et al.*, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," *arXiv preprint arXiv:2112.10741* (2021).
31 R. Rombach, A. Blattmann, D. Lorenz, *et al.*, "High-resolution image synthesis with latent diffusion models," (2021).
32 R. Corvi, D. Cozzolino, G. Zingarini, *et al.*, "On the detection of synthetic images generated by diffusion models," *arXiv preprint arXiv:2211.00680* (2022).
33 D. Nagothu, R. Xu, Y. Chen, *et al.*, "Detecting compromised edge smart cameras using lightweight environmental fingerprint consensus," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 505–510 (2021).
34 N. Poredi, D. Nagothu, Y. Chen, *et al.*, "Robustness of electrical network frequency signals as a fingerprint for digital media authentication," in *2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP)*, 1–6, IEEE (2022).
35 J. Krause, M. Stark, J. Deng, *et al.*, "3d object representations for fine-grained categorization," in *4th International IEEE Workshop on 3D Representation and Recognition (3dRR-13)*, (Sydney, Australia) (2013).