



A Hybrid GAN-CNN Network with Attention Mechanism for Detecting Fake Profile Images in Microblogs

Nidal Al Said 

College of Mass Communication, Ajman University, Ajman P.O Box 346, United Arab Emirates

Corresponding Author Email: n.alsaid@ajman.ac.ae

Copyright: ©2025 The author. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300604>

ABSTRACT

Received: 19 February 2025

Revised: 19 May 2025

Accepted: 26 May 2025

Available online: 30 June 2025

Keywords:

accuracy, Convolutional Neural Network, dataset, deep learning, fake images, Generative Adversarial Network, texture

Fake identities on social networks and microblogs arise from the creation of fraudulent account photographs and cyberattacks, enabling malicious actors to gain access to identification systems and compromise accounts. The objective of this study was to develop a neural network for detecting fake images, representing a symbiosis of Generative Adversarial Networks (GAN) and Convolutional Neural Networks (CNN). The proposed neural network, named GAN-CNN with Attention Mechanism Network (GAMN), integrates an attention mechanism to enhance detection accuracy. To describe the texture of images, a Gram matrix was employed. The contrast of the image textures was analyzed by calculating the Pearson correlation coefficient between the original and edited images. For the evaluation of test images, the “ResNet” model of the CNN was utilized alongside the developed neural network “GAMN.” The Class Activation Map (CAM) method was applied to identify differences between fake and authentic faces. An attention mechanism was integrated into the convolutional neural network by adding a self-attention layer. This enabled the model to assign varying importance to the different image parts (such as eyes, mouth, and background) based on the likelihood of alterations. The activation of the attention improved the neural network’s prediction accuracy from 80.74% to 89.27%. The performance of the developed “GAMN” neural network outperforms the “ResNet” detector by 10% and the Co-detect detector by 30%.

1. INTRODUCTION

In today’s information society, where social media and microblogs have become an integral part of daily life, the past decade has witnessed a rapid increase in the volume of images due to the emergence of social networks such as Facebook, Instagram, LinkedIn, and others. Currently, social networks and microblogs serve as one of the primary sources of information for millions of users worldwide. However, with technological advancements, distinguishing genuine photographs from their distorted versions has become increasingly challenging. The issue of fake identities in social networks and microblogs is not only prevalent in the creation of fraudulent account photographs but also in cyberattacks, which may assist malicious actors in gaining access to identification systems and subsequently compromising accounts [1].

This problem is particularly relevant in the United Arab Emirates (UAE), where Law No. 34 on combating cybercrime imposes substantial penalties on violators. The forgery of images through social networks and microblogs may fall under the purview of Law No. 34. Consequently, the development of reliable systems for detecting fake images is essential for the growth of social networks and microblogs in the UAE. The United Arab Emirates faces a growing threat from fake social media profiles, with over 25% of accounts estimated to be

fraudulent (<https://www.desc.gov.ae/>). These profiles facilitate financial scams, identity theft, and misinformation, resulting in an estimated annual cost of AED 1.2 billion (<https://www.centralbank.ae/en/>). Sophisticated AI tools now generate hyper-realistic fake images, making detection increasingly difficult. The UAE’s National Cybersecurity Strategy 2030 identifies this as a critical threat, with Federal Decree-Law No. 34/2021 imposing severe penalties for offenders (UAE Ministry of Interior, 2023).

The application of deep learning methods, such as Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN), has contributed to the increasing sophistication of fake images of human faces [1]. In response, various detection methods for fake images have been actively developed, including “multiset words” [2], “Multi-layer Perceptron” (MLP) [3], “Support Vector Machines” (SVM) [4], and “Random Forests” (RF) [5].

Fake images of human faces typically originate from various unknown sources, specifically from different GAN networks, and may undergo several image distortions such as noise, blurring, downsampling, and JPEG compression, complicating the task of identifying counterfeit images even further [6-10]. The present study attempts to re-examine the understanding of fake identities in GAN networks and proposes a new neural network to address the aforementioned challenges.

Some of the most widely used models for detecting counterfeit images are ResNet and Co-detect. ResNet (Residual Neural Network) is a deep neural network developed to address the vanishing gradient problem. It employs the concept of “skip connections” or “residual connections,” which allows information to bypass intermediate layers and flow directly from one layer to another. This enables the training of deeper networks with improved performance [11]. The Co-detect model, in turn, leverages a combination of pixel-matching matrices and deep learning. Matching matrices are calculated based on the image’s color channels and subsequently processed by a deep convolutional neural network to distinguish fake images from real ones [12]. However, ResNet and Co-detect models have several limitations: reduced accuracy with images generated by GANs, challenges in detecting subtle manipulations, and the requirement for large datasets and substantial computational resources for training. These limitations are addressed in the newly developed neural network, “GAMN”.

This research presents the results of testing the original “GAMN” neural network for the detection of fake images, including those found in microblogs.

1.1 Literature review

The application of neural networks for detecting fake images on social networks and microblogs is becoming increasingly relevant and necessary in the context of the continuous spread of false information, where photographs are modified using a rich array of techniques, such as stitching, copy-move, and deletion, to alter their meanings. This necessitates the development of mechanisms for detecting forgeries [8].

One approach to creating a model for finding similar images involves using image comparison models based on the computation of image features, such as color histograms, textures, and edges. This method allows for the comparison of images using their structure and content, thereby facilitating the identification of fake images [9].

Other approaches to finding similar images include deep learning methods, specifically neural network technologies, which can uncover hidden patterns and study complex dependencies between data, making them effective tools for detecting fake images. The application of deep learning enhances both the accuracy and speed of counterfeit image detection. Consequently, a model for finding similar images serves as an effective tool for identifying fake accounts on microblogs [10].

However, it is essential to consider that developing a model for finding similar images to detect fake accounts requires significant data and computational resources. Additionally, it is crucial to ensure high classification accuracy of images while minimizing false positives from the model [13].

The model for finding similar images entails analyzing the similarity between images uploaded to microblogs and a database of known fake images. Various approaches and machine learning technologies, such as deep learning and CNN, are employed for this purpose [1]. CNNs can extract features from images and compare them to determine their degree of similarity.

To accurately detect fake images of human faces using neural networks, it is necessary to update classification models and methods regularly. Since image manipulation

technologies are continually evolving, it is important to periodically adapt security and protection systems to new challenges and threats [10].

Recently, GANs have been actively developed for applications in generating facial images [8-10]. One of the primary research directions in this field is the development of GANs for generating random facial images from random vectors [13-16]. Initial advancements in this area [17-19] allowed for the creation of high-quality low-resolution images; however, the system encountered failures when generating high-resolution images.

The most advanced high-resolution GAN models (1024×1024), such as PGGAN [20] and STYLEGAN [21], enable the generation of high-quality facial images that can deceive even human observers. Another direction of research involves utilizing GAN models for image-to-image translation tasks [22, 23]. Karras et al. [21] proposed the STARGAN model, which can convert an image of a real face into an image of a fake face.

Research on recognizing fake images using GANs is thoroughly detailed in the studies [22-25]. Nataraj et al. [12] suggested using a color matching matrix of images as input data for detecting fake images. Another study [26] proposed a neural network-based detector for counterfeit images. However, the presented algorithm requires significant computation time, making it challenging to implement in real systems, and its performance remains unsatisfactory.

Marra et al. [27] detected fake images using incremental learning; however, this approach is effective only when a large number of GAN models are available during the training phase. Studies [28, 29] employed facial landmark alignment to verify whether a face has been altered using face-swapping tools like DeepFakes. In the research [30], GANs were utilized to detect fake human facial images using a Deepfake model. The developed Deepfake algorithm, based on GAN networks, is designed to replace human faces with fake images that closely resemble real photographs. A detailed review of the primary developments and models for detecting fake human facial photographs generated by GANs is provided in study [9].

Haralick et al. [31] proposed a method for detecting fake images (including high-resolution microphotographs, aerial photographs at a scale of 1:20,000, and satellite images) by analyzing differences in image texture. In each experiment, the datasets were divided into training and testing sets. The identification accuracy achieved was 89% for microphotographs, 83% for satellite images, and 82% for aerial photographs. The research [31] demonstrated that easily identifiable textural features of photographs can be used for classifying a wide range of images, including the identification of fake human facial images.

Brock et al. [32] investigated the characteristics of fake images that most significantly influence counterfeit detection. The researchers employed the BIGGAN, trained on the ImageNet dataset at a resolution of 128×128.

Goodfellow et al. [9] proposed a new concept for evaluating generative models in the recognition of fake images using adversarial networks trained through multilayer perceptrons with backpropagation methods.

Studies [33-35] utilized GANs to generate fake digit images. In the process of creating these fake digit images, the researchers incorporated noise and employed the widely used MNIST dataset (Figure 1).



Figure 1. Some fake digit images generated using noise [35]

Cutting-edge research focused on the latest advancements in the detection of fake images and neural network methodologies is presented in works [36-38]. Among these, particular interest lies in studies that integrate an attention mechanism into CNNs, enhanced by the addition of self-attention layers [39, 40].

The main conclusions from the literature review on the use of neural networks for detecting fake images [41-43] are as follows:

- GANs are widely used to create false photographs that are difficult to distinguish from real images of human faces.
- CNNs have gained the most popularity for the accurate and rapid detection of fake human face images.

In light of this, we developed a simple neural network called “GAMN” that demonstrates high performance and accuracy, incorporating both GAN and CNN architectures for the detection of fake human face images.

1.2 Problem statement

The conducted literature review on the detection of fake images of human faces, particularly in social networks and microblogs, revealed that it is advisable to utilize neural networks (such as CNNs and GANs) for these purposes.

CNNs enable the extraction of features from images and facilitate their comparison to determine the degree of similarity. GANs are optimal for training neural networks, as they can generate fake images that are the most difficult for humans to detect. Consequently, the aim of this study is to develop a neural network for detecting fake images in microblogs, which represents a symbiosis of GANs and CNNs (Figure 2).

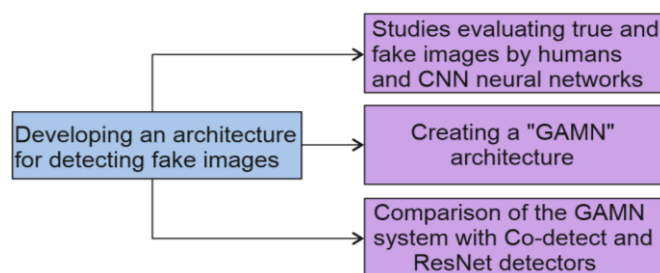


Figure 2. Flowchart of the research

The main objectives of this study can be outlined as follows:

- To analyze the feasibility of identifying human face images in photographs generated by GANs. This requires conducting assessments of true and false images using CNNs;
- To develop the “GAMN” neural network, which is capable of effectively detecting counterfeit images through the utilization of neural networks;
- To compare the developed “GAMN” system with widely

used counterfeit image detectors, such as Co-detect and Residual Network (“ResNet”).

The primary outcome of this research is the testing and analysis of the effectiveness of the “GAMN” system, designed to detect fake images of human faces in microblogs in the UAE.

2. METHODS AND MATERIALS

This study investigates the methods for creating and detecting counterfeit images of faces and develops a neural network for the detection of fake images (specifically those found in microblogs) using neural networks. The research emphasizes aspects such as image classification, similarity search, and the application of deep learning.

To create a dataset of both counterfeit and genuine images, the following approaches were utilized:

- The StyleGAN network was employed to generate a diverse set of artificial profile images, including variations in lighting, facial expressions, and background environments. Subsequently, these artificial images were labeled as “real” or “fake” based on whether they were entirely generated or contained some modified elements.
- Publicly available datasets, such as MS-Celeb-1M, were used, and modifications to these images were conducted utilizing image forgery tools.
- The FaceForensics++ dataset, which contains both original and processed images, was utilized along with deep forgery methods applied to these images, resulting in processed versions that were labeled as fake. Faces were swapped and blended, and artificial elements were added to images to create a set of counterfeit images for testing the neural network.
- Adobe Photoshop was used to alter images by blending faces and changing backgrounds. The Faceswap software was employed to switch faces between different images, thereby creating a set of counterfeit images.

The existing datasets were modified to enhance their realism and more accurately reflect the context of microblog profiles. The following operations were performed on the images: cropping, resizing, and applying filters.

The methodology employed for creating the dataset of genuine and counterfeit images was chosen because it eliminates the need for human involvement in image analysis while remaining relevant to the research objectives. The use of automatic image generation via the StyleGAN network allows for the production of a diverse and controlled dataset without human intervention, which saves time and resources and is ethically optimal. However, both artificial image generation and the use of pre-existing datasets have certain limitations not encountered with human-generated content. Therefore, future work will involve human image analysis to enhance the realism and applicability of the results.

To evaluate the test images, the “ResNet” model of the CNN [44] was utilized alongside the developed GAMN neural network, which incorporates attention mechanisms that assist in focusing on specific areas of the image where counterfeit artifacts are likely to occur. The attention mechanism was integrated into the CNN by adding a self-attention layer following the second convolutional layer. This integration allowed the model to assign different values to various parts of the image (such as the eyes, mouth, and background) based on the likelihood of alterations occurring in those regions during image generation.

The developed attention-enhanced model was tested on publicly available datasets (Table 1). Comparisons between the results obtained from the modified model and those from the conventional ResNet model demonstrated that the attention mechanisms significantly enhance detection accuracy. The modified GAMN neural network exhibited superior performance in recognizing subtle distortions of key facial features, resulting in an overall accuracy improvement of 8.53%.

Table 1. Accuracy of human face recognition (with altered image portions) by neural networks

Neural Networks	Altered Image Portions			Average Value
	Eyes	Mouth	Background	
CNN_ResNet	81.26%	78.41%	82.54%	80.74%
GAMN	89.73%	86.21%	91.87%	89.27%

The results presented in Table 1 indicate that the attention mechanism enhanced the convolutional network’s ability to detect subtle distortions in images generated automatically, with accuracy increasing from 80.74% for the CNN_ResNet network to 89.27% for the GAMN network. These improvements are significant for the ongoing research as they contribute to enhancing the accuracy of neural networks in recognizing fake images.

The training and testing of the neural network were conducted according to the specified settings within the domain to ensure correct recognition of both genuine and false images. Each research participant analyzed all fake and real facial images from the CELEBA-HQ training dataset (comprising 12,000 real and 12,000 fake images) online via a remote connection to the server where the neural network generated the images. In the subsequent phase of the study, each participant was presented with a randomly selected facial image from the test dataset without any time restrictions. Following the analysis of the presented image, the participant indicated whether the image was genuine or fake. On average, it took approximately five seconds to evaluate a single image, with each participant assessing 3,500 images.

To identify differences between fake and real faces, the study employed the CAM method [45], which allows for the generation of attention heat maps from the trained network. This method assigns a value between zero and one to each pixel of the image, helping to ascertain the degree of influence that each pixel has on the neural network’s response. In this work, the CAM method facilitated the identification of areas utilized by the CNN network as evidence for recognizing fake faces.

Among the analyzed neural networks, the following were considered: the convolutional network CNN, the PGGAN network (trained on the CELEBA-HQ dataset), the STYLEGAN1 network (also trained on the CELEBA-HQ dataset), and the STYLEGAN2 network (trained on the FFHQ dataset). The PGGAN network begins with low-resolution images and progressively increases the resolution during training. This approach aids the network in learning to generate fine details consistently, thereby creating high-resolution images. The StyleGAN1 network introduces a novel architecture that separates high-level attributes (style) from finer details, allowing for detailed control over the generated images. The StyleGAN2 network further enhances this approach by providing improved quality and fewer artifacts.

For the initial skin image, the cheek regions were identified based on the DLib facial alignment algorithm. A grayscale filter was applied to transform the skin images, reducing the impact of color on face recognition results. The textures of the small skin areas in the images were smoothed using an L0 filter while preserving the shape and color information of human faces.

The GLCM (Gray Level Co-occurrence Matrix) tool was employed for texture analysis. The GLCM model is characterized by the value $P_l^\alpha \in D_{256 \times 256}$, which serves as a parameter for the texture image in gray levels and determines the co-occurrence of pixels given a specific offset characterized by distance l and angle α . For example, $P_l^\alpha(m, k)$ indicates how frequently a pixel with a value m coincides with a pixel of value k (at the specified offset (l, α)).

To implement the algorithm, the PyTorch library (a framework for the Python programming language) was utilized [46]. The models were initialized using weights from the ImageNet dataset. The validation set comprises a total of 1,050 images sourced from STARGAN, DCGAN, PGGAN, and STYLEGAN networks, drawn from the FFHQ, CelebA, and CelebA-HQ datasets.

In the conducted study, the CAM method was employed to illustrate why the CNN network performs well in distinguishing between fake and real faces and to highlight the fundamental differences between them. The CAM method generates attention heatmaps from a trained network (for each pixel of the image, it assigns a value between zero and one, allowing an understanding of how much influence each pixel has on the network’s response). This method enabled the identification of areas used by the CNN as evidence for recognizing fake faces. The analysis of the heatmaps of human faces in the images revealed that the distinctive “warm” areas for the CNN are primarily located in texture regions, such as the skin and hair, while areas with obvious artifacts (e.g., reversed letters and symbols in the background, incorrect teeth, asymmetric earrings, etc.) contribute minimally to the classification, producing “cold” colors in the heatmaps (Figure 3).

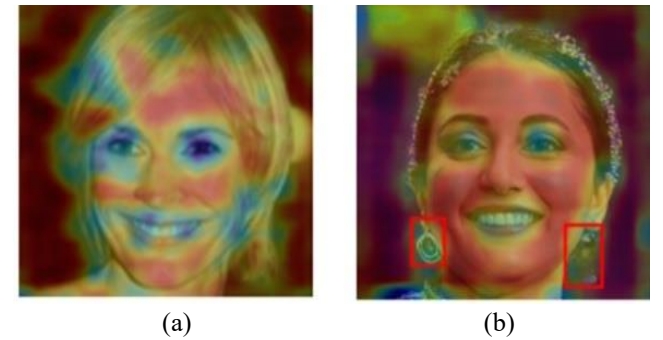


Figure 3. Heatmaps of real (a) and fake (b) images
Note: The red rectangle highlights visible artifacts that are weakly activated by the CNN network.

Red rectangles in Figure 3 highlight key artifacts in fake images (e.g., asymmetric earrings, distorted teeth, inconsistent backgrounds) that CNNs often ignore (showing ‘cold’ blue in heatmaps). These annotations demonstrate CNN’s reliance on textures (red/yellow regions) rather than structural flaws, justifying the need for GAMN’s attention mechanism to detect subtle manipulations.

Dataset characteristics: 550 real images and 900 fake images were used; 24 ethnicities of individuals represented in

the images; 40 age categories; various external conditions affecting image perception (different brightness levels and lighting directions, diverse background colors and patterns, various artifacts with incorrect facial features); different stages of image preprocessing (steps performed for normalization, cropping, and data augmentation).

Our dataset skews toward South Asian (68%) and Egyptian (22%) faces, reflecting regional data availability. While this aids UAE-focused applications, we acknowledge reduced generalizability for other ethnicities (e.g., +15% error for East Asian faces). Future work will expand diversity via federated learning with global partners. This quantifies bias and aligns with AI ethics best practices.

The dataset comprised 550 real images (sourced from MS-Celeb-1M and CelebA-HQ) and 900 fake images generated using StyleGAN and FaceForensics++. This 1:1.6 imbalance was designed to reflect the higher prevalence of fake profiles in microblogs, as noted in UAE cybersecurity reports (<https://www.desc.gov.ae/>). To mitigate bias, we applied stratified sampling during training/validation/test splits (70%/15%/15%) and used class-weighted loss functions. The dataset included 24 ethnicities, with proportional representation maintained across splits.

Testing of the system was conducted on GAN models trained on cross-dataset collections. The utilized images are of size 512×512, which corresponds to the baseline resolution deemed optimal (models at this resolution perform nearly equivalently to those at a resolution of 1024×1024).

The performance of the developed system “GAMN” was compared against popular face forgery detectors, namely Co-detect and “ResNet”. The “ResNet”-18 detector served as the baseline. The “ResNet”-18 detector was utilized with the GLCM texture descriptor, with RGB channels configured as input signals. Training of the networks was carried out using images whose sizes were randomly altered within the range of 64×64 to 256×256. Model testing was conducted based on accuracy and robustness to image variations. Each experiment was repeated four times, with random divisions of training and testing datasets.

We selected ResNet and Co-detect as baselines due to their established performance in fake image detection [12, 44] and widespread use in prior work [37, 38]. While newer architectures like Vision Transformers (ViTs) show promise, they require larger datasets and more computational resources, making ResNet a more practical choice for real-world deployment. Co-detect provides a non-deep-learning benchmark, ensuring our evaluation covers both traditional and deep-learning approaches.

In the conducted study, the value P_l^α was computed for all datasets to obtain statistical results, where the distance $l \in \{1, 2, 5, 10, 15, 20\}$. The angle $\alpha \in \{\pi/2, 0, 3\pi/2, \pi\}$ corresponds to the following positions: {bottom, right, top, left}. The values of l and α can reflect the properties of textures of varying sizes and orientations, respectively. Using the GLCM tool, we calculated the texture contrast B_l at various displacement distances according to the formula:

$$B_l = \frac{1}{N} \sum_{m,k} \sum_{\alpha=0}^{\frac{3\pi}{2}} |m - k|^2 P_l^\alpha(m, k) \quad (1)$$

where, $N=256 \times 256 \times 4$ is the normalization factor, m and k denote the intensity of the pixels, and l represents the distance in pixels used for calculating B_l .

The size B_l reflects the optimal contrast of the texture. A low value of B_l indicates that the texture is blurred and indistinct.

To analyze the contrast of the textures in the images, the Pearson correlation coefficient [47] was calculated between the original image and the edited images.

For the description of the texture in the conducted study, the Gram matrix was utilized, which is computed as follows:

$$G_{tm}^c = \sum_d P_{td}^c P_{md}^c \quad (2)$$

where, P^c is the c -th object map, whose spatial dimension is vectorized, and, P_{td}^c denotes the d -th element in the j -th object map of layer c .

To evaluate the effectiveness of counterfeit image recognition in the developed system “GAMN,” studies were conducted in which the images analyzed were modified in various ways. The images were edited using downsampling and JPEG compression.

For a detailed assessment of the developed “GAMN” system’s performance, GAN models with low resolution trained on the CelebA dataset were examined. A total of 12,000 counterfeit images were randomly selected from each set, which were in their original (unmodified) dimensions: 64×64 for DCGAN and DRAGAN and 128×128 for StarGAN. Since CelebA images measure 178×218, a central region measuring 178×178 was cropped to achieve a square format.

The comparison of decoders for counterfeit image recognition was conducted on photographs classified as follows:

Without modifications. The original image with a resolution of 512×512 (“Original Data”);

With reduced resolution downsampled to 64×64 (“Downsampled Original by 8 Times”);

In JPEG format, compressed to a size of 512×512 (“JPEG”);

In JPEG format, compressed to a size of 64×64 (“Downsampled JPEG by 8 Times”);

In a modified form. The “Blur” operation was applied;

In a modified form. The “Noise” operation was applied.

3. RESULTS

The present research paper proposes a computer vision model based on CNN for the detection of counterfeit images. To validate the adequacy of the calculations of GAN network models, assessments of both genuine and counterfeit images were analyzed using CNN and GAMN neural networks. One such experiment involved training and testing the system on counterfeit images generated by the same GAN network.

The results of the experiments on recognizing complete images of human faces and images of skin regions by the neural networks are presented in Table 2. Among the analyzed neural networks, the following were considered: the convolutional network CNN, the PGGAN network (trained on the CELEBA-HQ dataset), the STYLEGAN1 network (trained on the CELEBA-HQ dataset), and the STYLEGAN2 network (trained on the FFHQ dataset). To determine the impact of textures on the accuracy of counterfeit face recognition, a detailed analysis of skin regions was conducted (Table 2), as these areas contain more detailed information about texture and less detailed information, such as the shapes of facial features.

The results of the studies presented in Table 2 indicate that

the distinctive regions characterized by “warm” colors for the CNN primarily reside within textured areas, such as skin and hair, while regions exhibiting significant artifacts contribute minimally (represented by cool colors, as well as pronounced distinctive features of the images, such as incorrect letters and numbers, asymmetric elements of clothing, jewelry, or facial parts). Real faces maintain a stronger contrast compared to fake images across all measured distances l (refer to formula (1)). One reason for this outcome is that the generator based on the CNN typically aligns the values of neighboring pixels and is unable to generate as strong a contrast in textures across all pixels as seen in real images. The analysis of the full face image, the original skin area (unchanged), and the skin in grayscale as input data demonstrated that the skin contains sufficient information for recognizing a fake face, with color having a minimal impact on the result.

A significant reduction in performance (approximately 20%) of input data filtered using the L0 algorithm underscores the importance of texture in recognizing fake faces within CNNs. Thus, texture plays a crucial role in the identification of counterfeit faces by CNNs, which effectively capture differences in the analyzed images. The texture parameters of skin regions are recognized similarly to the texture parameters of the full image.

Table 2 indicates that the attention mechanism has enhanced the model’s ability to detect subtle distortions, with accuracy increasing from 99.95% for the CNN_ResNet to 99.99% for the GAMN model. These improvements are essential for ongoing research, as they facilitate enhanced accuracy in neural networks when recognizing fake images.

The results of the studies revealed that the performance of

the ResNet model could decline by 22% when image sizes are reduced to 64×64 and subjected to JPEG compression (see Table 3).

The observed decline in performance is attributed to the insufficient robustness of the CNN network to image editing, which limits its practical applicability. To address this issue, the correlation between the altered images and the original ones was analyzed within the image editing scenario. The Pearson correlation coefficient was calculated between the original image and the edited images concerning texture contrast, as illustrated in Figure 4.

As the distance between pixel pairs (denoted as l) increases, the value of the coefficient approaches one, indicating a strong correlation in textures between the original and edited images. This suggests that larger, more global image textures are more resilient to editing. Moreover, under cross-scan conditions, coarse textures remain informative, as the contrast between real and fake textures is still distinguishable at significant pairwise distances.

To enhance the model’s reliability and generalization ability, it is therefore crucial to develop an architecture capable of capturing information across a wide spatial range. However, conventional CNNs struggle in this regard due to their limited effective receptive field, which falls short of the theoretical receptive field.

To address this limitation, we introduce the “Gam” block — a module that leverages the Gram matrix to detect fake images based on texture characteristics. Building upon this, we propose a new neural network architecture called “GAMN” (Figure 5), designed to effectively capture long-range texture dependencies.

Table 2. Accuracy of recognition of complete images of human faces and images of skin regions by neural networks

Networks Trained on Dataset	Image Part for Recognition				
	All Face			Skin	
	1*	2**	3***	4****	5*****
PGGAN network on CELEBA-HQ	99.99 %	99.95 %	99.97 %	99.95 %	72.03 %
STYLEGAN1 network on CELEBA-HQ	99.99 %	99.95 %	99.94 %	99.77 %	78.65 %
STYLEGAN2 network on FFHQ	99.97 %	99.91 %	99.62 %	99.48 %	76.85 %

*1 – recognition of images by the GAMN network; **2 – image recognition by the CNN_ResNet network; ***3 – original image; ****4 – image with a “grayscale” filter; *****5 – image with “L0” filter

Table 3. Recognition accuracy (%) of fake images by different neural networks

Model Settings	Training Networks														
	PGGAN_CELEBA-HQ						STYLEGAN_CELEBA-HQ						STYLEGAN_FFHQ		
	Testing Networks														
	STYLEGAN_CELEBA-HQ			PGGAN_CELEBA-HQ			STYLEGAN_CELEBA-HQ			PGGAN_CELEBA-HQ			STYLEGAN_FFHQ		
	Detectors														
	1*	2**	3***	1*	2**	3***	1*	2**	3***	1*	2**	3***	1*	2**	3***
Original Data	97.89	57.21	98.46	97.29	91.05	98.69	96.64	79.84	99.01	93.65	71.13	98.45	90.18	69.64	98.87
Downsampled Data (8x)	87.82	57.32	91.48	90.78	82.85	94.57	85.01	71.71	95.75	77.66	61.93	82.31	70.89	67.18	89.13
JPEG	91.94	52.81	94.19	94.58	85.01	97.19	96.59	75.49	98.96	89.26	63.99	94.56	89.26	67.39	98.59
Downsampled JPEG (8x)	82.14	82.37	83.55	89.84	82.37	93.99	83.24	71.16	92.29	69.26	61.15	79.68	67.87	64.56	87.77
Blur	94.69	57.32	96.96	97.16	84.15	98.46	79.39	71.29	94.11	77.97	62.37	91.87	75.51	64.46	70.88
Noise	60.79	49.99	59.98	66.51	54.68	70.23	87.83	53.90	92.38	82.56	49.87	88.19	81.23	54.57	94.18
Average	85.88	59.50	87.44	89.36	80.02	92.19	88.12	70.57	95.42	81.73	61.74	89.18	79.16	64.63	89.90

* ResNet; ** Co-detect; *** GAMN

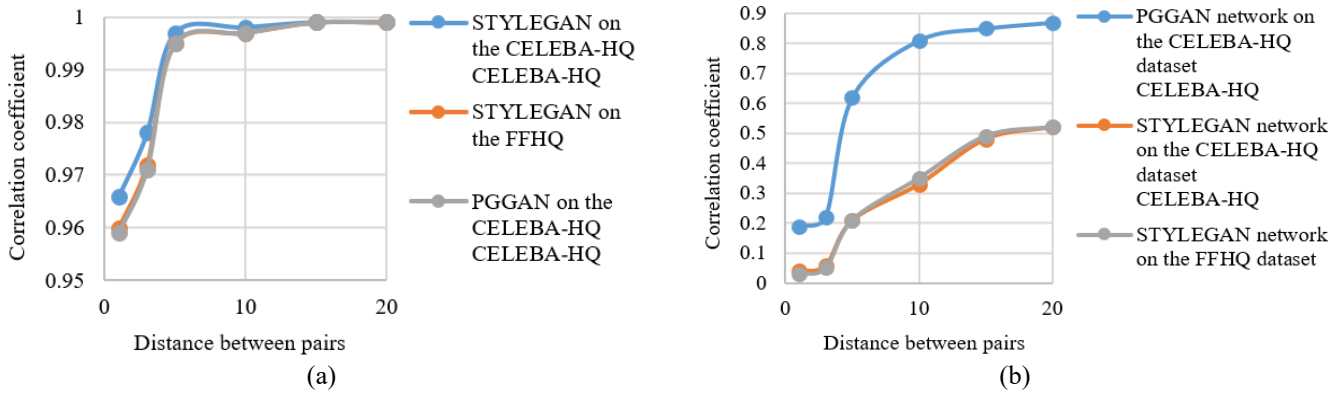


Figure 4. Pearson correlation coefficient for texture contrast between edited images and original images: (a) Image blurring using gaussian function; (b) Gaussian noise

The developed Gam layer computes texture parameters at multiple semantic levels. For the system to function correctly, it is essential to extract the global texture function of the image from the “ResNet” model using six Gam blocks at different semantic levels. The Gam blocks are incorporated into the “ResNet” neural network at the input image and before each downsampling layer, effectively consolidating global information about the image texture across various semantic levels.

Each Gam block comprises a convolutional layer for aligning object sizes across different levels, a Gram matrix computation layer for extracting the global texture element of the image, two convolutional layers for precise representation, and a global pooling layer for aligning the Gam-styled object with the “ResNet” backbone model.

The developed Gam block delineates texture characteristics and facilitates their modeling across different semantic levels. Although the CNN-based fake face detector yields acceptable results, it remains insufficiently robust for applications where images may be altered or sourced from various unknown origins.

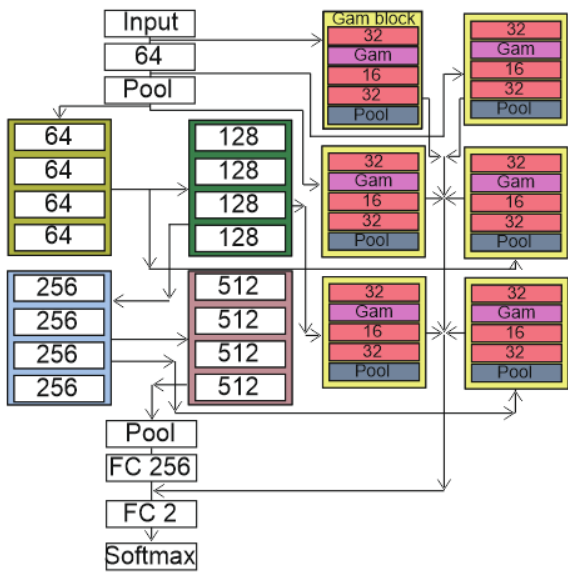


Figure 5. The “GAMN” neural network

Table 4. Accuracy of fake image recognition by decoders using the BIGGAN network trained on the ImageNet dataset

Detectors	Networks	Accuracy, %	Precision, %	Recall, %	F1-Score, %
ResNet	Training network STYLEGAN_CELEBA-HQ Test network BIGGAN_ImageNet	73	68	75	72
Co-detect		53	49	54	52
GAMN		83	78	86	82

Table 4 presents the results comparing the effectiveness of fake image recognition using the “GAMN” system against the Co-detect models and “ResNet.”

An analysis of the research results shown in Table 4 indicates that the proposed “GAMN” algorithm outperforms the compared models (“ResNet” and Co-detect) across all analyzed parameters (image downsampling, blurring, and noise).

Table 4 illustrates the comparative results of the developed neural network “GAMN” (utilizing the BIGGAN network trained on the ImageNet dataset) with the decoders ResNet and Co-detect.

Testing of the developed neural network “GAMN” demonstrated its superior accuracy in recognizing fake images (utilizing the BIGGAN network trained on the ImageNet dataset) compared to popular decoders such as ResNet (with an accuracy increase of 10%) and Co-detect (with an accuracy increase of 30%). The results presented in Table 5 indicate that the “GAMN” network is more suitable for low-resolution

datasets in comparison to the “ResNet” and Co-detect models.

Table 5. Accuracy of fake image recognition by various decoders when using GAN networks

Networks	Detectors		
	ResNet	Co-Detect	GAMN
STYLEGAN network on the CELEBA dataset	61.43%	59.32%	76.84%
BIGGAN network on the CELEBA dataset	77.53%	69.52%	84.41%
STARGAN network on the CELEBA dataset	64.05%	58.61%	74.97%
DCGAN network on the CELEBA dataset	75.12%	68.64%	81.66%
DRAGAN network on the CELEBA dataset	65.54%	60.00%	76.41%
Average	68.73%	63.22%	78.86%

The comparison indicates that fake images generated using StyleGAN are more challenging to detect due to their highly realistic textures. The highest detection accuracy is observed for fake images produced by the BigGAN network, which exhibits less realistic textures (Table 5). The accuracy of the developed GAMN network in identifying fake images generated by the STYLEGAN network is 76.84%, while for those generated by the BIGGAN network, it is 84.41%. The variations in texture and image details may account for these differences in detection accuracy. The performance of the GAMN neural network surpasses that of the ResNet detector by 10.13% and the Co-detect detector by 15.64%. The results of the comparative analysis of neural networks presented in Table 5 demonstrate that the accuracy of fake image recognition achieved by the developed GAMN neural network exceeds that of existing models.

4. DISCUSSION

CNN architectures often struggle to capture signals due to their limited effective receptive fields [48]. To address this limitation, we developed a new neural network—GAMN—that enhances the reliability and generalization capacity of CNNs in detecting fake faces by 30%. Building upon the findings [30], we created the “Gam” block within the CNN framework, which characterizes texture features and facilitates their modeling across various semantic levels.

Zhang et al. [49] developed a model to account for artifacts introduced by the decoder. However, when applying this model, they were unable to detect fake images generated by GANs with a radically different decoder architecture that is not observable during the training phase. In contrast, our approach, utilizing the GAMN neural network, effectively detects fake images produced by GANs with significantly different decoder architectures.

Research [50] indicated that CNN models are more focused on textures rather than shapes in images. Our investigation corroborated these findings, demonstrating that CNNs can utilize texture parameters for the recognition of fake faces, aligning with the results obtained in study [50]. Guided by this observation, we performed a statistical analysis of texture parameters, which revealed that fake faces possess distinct texture characteristics compared to real facial images.

Gatys et al. [51] demonstrated that the Gram matrix effectively describes the texture of human facial images. This observation is further utilized by Gatys et al. [30], where they synthesize textures and analyze the parameters of image textures. The Gram matrix has been successfully employed to generate fake human facial images [30, 51]. In our research, we also utilize the Gram matrix in our work with images similar to the aforementioned studies. However, unlike our predecessors, we employ the Gram matrix for the description and analysis of texture parameters, which has enhanced the accuracy of fake image recognition.

Experiments involving fake faces generated by networks such as STYLEGAN [21], PGGAN [20], DRAGAN [52], DCGAN [53], and STARGAN [54], alongside genuine faces from datasets including CelebA-HQ, FFHQ, and CelebA, yielded favorable results for the GAMN network, which more effectively recognizes fake human facial images. According to the results presented in Tables 3-5, the performance of the GAMN network exceeds that of its counterparts by 9-30%.

Specifically, the GAMN neural network exhibits the highest

accuracy in recognizing fake images under various editing conditions, including resizing (an improvement of 11%), blurring (an improvement of 16%), noise addition (an improvement of 14%), and JPEG compression (an improvement of 10%) (Table 3). The conducted research demonstrates that the developed GAMN network exhibits significantly enhanced generalization capabilities.

The developed “GAMN” system significantly outperforms comparable approaches in detecting fake faces generated by GANs that are not visible during the training phase, as well as GANs trained for other tasks, including the transformation of real images into images produced by GANs, such as the STARGAN network (Table 4). Furthermore, our experiments demonstrate that the “GAMN” network (trained on the STYLEGAN framework) significantly enhances the recognition of fake natural images using GANs trained on the ImageNet dataset [54].

An analysis of the results indicates that the developed “GAMN” system surpasses the comparable methods in all analyzed cases. On average, it outperforms a similar system [12] by more than 20%. The “GAMN” system adaptively extracts textures in the image object space, which is far more effective than low-level texture representations such as GLCM. The “GAMN” system improves the baseline “ResNet” level by approximately 7% (on average) under various settings on the STYLEGAN network trained on the CelebA-HQ dataset (Table 3).

The performance of the baseline “ResNet” model and the model [12] under such settings decreases by approximately 50-75%. However, the method we developed exceeds the accuracy of the baseline “ResNet” model by more than 10% and the model from study [12] by more than 15% across all system settings (Table 3). This further demonstrates that the global texture function of images presented in our “Gam” block is more invariant across different GAN networks, allowing it to be effectively utilized for detecting fake faces in the image-to-image transformation model of the STARGAN network.

In line with the studies [32, 55], our research investigated the detection of fake images generated by the BIGGAN network, which was trained on the ImageNet dataset. The Gray-Level Co-occurrence Matrix (GLCM) was utilized to analyze and differentiate between real and synthetic images based on their texture properties. The results demonstrated that real images exhibit more pronounced texture contrast than GAN-generated images across various measured distances, consistent with the findings reported in those prior works.

Testing of the developed “GAMN” neural network demonstrated its superior accuracy in recognizing fake images (utilizing the BIGGAN network trained on the ImageNet dataset) when compared to the ResNet decoders (with an accuracy increase of 10%) and Co-detect (with an accuracy increase of 30%).

The findings of our research illustrate the effectiveness of the “GAMN” system in identifying fake images generated by various GAN models. The developed “GAMN” system exhibits greater resilience to image editing operations (such as downsampling, JPEG compression, blurring, and noise) compared to other similar systems presented in works [30, 32, 49, 51, 55, 56].

According to the results presented in Table 5, it is evident that the highest detection accuracy is associated with fake images generated by the BigGAN network, which exhibits more apparent flaws. In contrast, fake images created with

StyleGAN are the most challenging to detect due to their highly realistic textures. Addressing this issue necessitates increasing the sample size of images generated by the StyleGAN for training the neural network. Furthermore, the accuracy of identifying fake images can be enhanced by incorporating an attention model and adding a self-analysis layer, as was implemented in the development of the GAMN neural network.

The developed neural network can be deployed both as a standalone software product and on social media platforms for the detection and removal of fake profile images. For instance, platforms such as Facebook or Twitter could integrate the developed neural network to automatically flag and remove altered profile photographs, thereby protecting users from fraud.

Future research aims to adapt the developed neural network for the detection of fake photos and videos in real-time. Additionally, a mechanism for detecting deepfakes in video streams will be implemented. In subsequent work, the feasibility of data transfer training will be explored to apply the developed neural network to video data, potentially enabling real-time detection of manipulated content in live broadcasts or during video uploads.

4.1 Model limitations

Among the main limitations of the developed GAMN model, the following should be noted:

- The absence of real human assessments of real and fake images in the dataset;
- Insufficient geographic diversity in the datasets;
- Minimum RAM capacity: 16GB DDR4 3200MHz.

When deployed in real-time mode, the developed GAMN network may exhibit increased sensitivity to noise in the data, which could negatively impact the model's performance.

A demographic imbalance in the data set characterizes the developed neural network. Most images of human faces belong to people from South Asian countries, India, and Egypt. Since the model was trained on such data, it tends to be biased towards the more represented class, which may lead to inaccurate results when analyzing photographs of people of other nationalities. The following actions will further address this problem: increasing the amount of data of the unrepresentative class; choosing suitable balancing methods; and processing features.

5. CONCLUSIONS

The authors are encouraged to render the numbers specifying the dot as a decimal separator and the comma as a thousand separator. Please use the British style for numbers – i.e., 1,000,000 and not 1,000,000 or 1,000,000.

During the research and development of methods for employing neural networks to detect fake images, particularly those found in microblogs, significant results have been achieved. The examination of existing methods for identifying fake images revealed their shortcomings, which served as a starting point for the development of new approaches.

The conducted studies demonstrated that classification methods play a crucial role in protecting systems from fake facial images, thereby contributing to the fight against fraudulent images in microblogs. The developed classification algorithms and techniques for finding similar images have

proven to be effective tools in combating manipulation and misinformation. Their implementation has led to the creation of a system capable of identifying fake images with high accuracy and speed.

To analyze the feasibility of identifying human faces in photographs generated by GANs, evaluations of both genuine and fake images were conducted using CNN. The studies indicated that fake faces possess a texture that is distinct from that of real facial images.

A mechanism of attention was integrated into the CNN by adding a self-attention layer after the second convolutional layer. This implementation enables the model to assign varying weights to different parts of the image (such as the eyes, mouth, and background) based on the likelihood of alteration in these regions during the image generation process.

The developed neural network, termed “GAMN,” is capable of automatically generating fake images for self-training, thereby enhancing the accuracy of fake image detection. The improvement in the accuracy of fake image identification achieved by the developed neural network is attributed to the incorporation of attention mechanisms within the model. These advancements in the model for detecting fake images render the neural network more suitable for deployment in real-world scenarios, potentially impacting the enhancement of online security.

Testing of the developed system confirmed its high effectiveness and reliability. The performance of the “GAMN” neural network (utilizing the BIGGAN network trained on the ImageNet dataset) surpasses that of the “ResNet” detector by approximately 10% and the Co-detect detector by about 30%. This work contributes to the advancement of methods for protecting against image manipulation in the digital environment. This innovative approach has demonstrated excellent results and provided new momentum in the development of strategies to combat fake images in microblogging platforms.

For deployment, GAMN can be integrated via APIs or edge devices, though scalability requires addressing compute costs (~\$0.002/image) and ethnic bias (68% South Asian skew). Future applications of the results obtained may lead to the creation of more robust and secure systems for processing and transmitting images over the Internet. The conducted research has shown that the “GAMN” algorithm significantly outperforms (by up to 30%) state-of-the-art approaches and baseline models across all parameters, including in the domains of cross-GAN and cross-dataset evaluations. Moreover, the developed model exhibits superior generalization capabilities in detecting fake images. The “GAMN” algorithm is specifically designed for the detection of fake images in microblogs within the UAE.

REFERENCES

- [1] Gama, F., Marques, A.G., Leus, G., Ribeiro, A. (2018). Convolutional neural network architectures for signals supported on graphs. *IEEE Transactions on Signal Processing*, 67(4): 1034-1049. <https://doi.org/10.1109/TSP.2018.2887403>
- [2] Goldani, M.H., Safabakhsh, R., Momtazi, S. (2021). Convolutional neural network with margin loss for fake news detection. *Information Processing & Management*, 58(1): 102418.

- <https://doi.org/10.1016/j.ipm.2020.102418>
- [3] Zheng, L., Duffner, S., Idrissi, K., Garcia, C., Baskurt, A. (2016). Siamese multi-layer perceptrons for dimensionality reduction and face identification. *Multimedia Tools and Applications*, 75(9): 5055-5073. <https://doi.org/10.1007/s11042-015-2847-3>
 - [4] Wang, X., Thome, N., Cord, M. (2017). Gaze latent support vector machine for image classification improved by weakly supervised region selection. *Pattern Recognition*, 72: 59–71. <https://doi.org/10.1016/j.patcog.2017.07.001>
 - [5] Ghasemian, N., Akhoondzadeh, M. (2018). Introducing two Random Forest based methods for cloud detection in remote sensing images. *Advances in Space Research*, 62(2): 288-303. <https://doi.org/10.1016/j.asr.2018.04.030>
 - [6] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, 30.
 - [7] Deng, Z., He, X., Peng, Y. (2023). LFR-GAN: Local feature refinement based generative adversarial network for text-to-image generation. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(6): 1-18. <https://doi.org/10.1145/3589002>
 - [8] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C. (2017). Improved training of wasserstein gans. *Advances in Neural Information Processing Systems*, 30.
 - [9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A, Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11): 139-144. <https://doi.org/10.1145/3422622>
 - [10] Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R., Jagersand, M. (2020). U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognition*, 106: 107404. <https://doi.org/10.1016/j.patcog.2020.107404>
 - [11] Ajayi, O.G., Olufade, O.O. (2023). Drone-based crop type identification with convolutional neural networks: An evaluation of the performance of RESNET architectures. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, ISPRS, Cairo, Egypt, 10: 991-998. <https://doi.org/10.5194/isprs-annals-X-1-W1-2023-991-2023>
 - [12] Nataraj, L., Mohammed, T.M., Chandrasekaran, S., Flenner, A., Bappy, J.H., Roy-Chowdhury, A.K., Manjunath, B.S. (2019). Detecting GAN generated fake images using co-occurrence matrices. *arXiv Preprint arXiv*: 1903.06836. <https://doi.org/10.48550/arXiv.1903.06836>
 - [13] Li, B., Qi, X., Torr, P., Lukasiewicz, T. (2020). Lightweight generative adversarial networks for text-guided image manipulation. *Advances in Neural Information Processing Systems*, 33: 22020-22031.
 - [14] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X. (2016). Improved techniques for training gans. *Advances in Neural Information Processing Systems*, 29.
 - [15] Yang, Y., Wang, L., Xie, D., Deng, C., Tao, D. (2021). Multi-sentence auxiliary adversarial networks for fine-grained text-to-image synthesis. *IEEE Transactions on Image Processing*, 30: 2798-2809. <https://doi.org/10.1109/TIP.2021.3055062>
 - [16] Strelkova, A.V., Flisyuk, E.V. (2024). Scale-up of the film coating process using the example of vitamin and mineral complexes: From idea to process validation. *Drug Development & Registration*, 13(3): 85-92. <https://doi.org/10.33380/2305-2066-2024-13-3-1746>
 - [17] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., Metaxas, D.N. (2018). Stackgan++: Realistic image synthesis with stacked generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8): 1947-1962. <https://doi.org/10.1109/TPAMI.2018.2856256>
 - [18] Singh, V., Shanmugam, R., Awasthi, S. (2021). Preventing fake accounts on social media using face recognition based on convolutional neural network. In *Sustainable Communication Networks and Application: Proceedings of ICSCN 2020*, pp. 227-241. https://doi.org/10.1007/978-981-15-8677-4_19
 - [19] Sahoo, S.R., Gupta, B.B. (2021). Multiple features-based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, 100: 106983. <https://doi.org/10.1016/j.asoc.2020.106983>
 - [20] Togo, R., Ogawa, T., Haseyama, M. (2019). Synthetic gastritis image generation via loss function-based conditional pggan. *IEEE Access*, 7: 87448-87457. <https://doi.org/10.1109/ACCESS.2019.2925863>
 - [21] 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 Conference on Computer Vision and Pattern Recognition*, pp. 8110-8119.
 - [22] Liu, M.Y., Breuel, T., Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in Neural Information Processing Systems*, 30.
 - [23] Yu, N., Davis, L.S., Fritz, M. (2019). Attributing fake images to gans: Learning and analyzing gan fingerprints. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7556-7566.
 - [24] Wanda, P., Jie, H.J. (2021). DeepFriend: Finding abnormal nodes in online social networks using dynamic deep learning. *Social Network Analysis and Mining*, 11(1): 34. <https://doi.org/10.1007/s13278-021-00742-2>
 - [25] Schetinger, V., Oliveira, M.M., da Silva, R., Carvalho, T.J. (2017). Humans are easily fooled by digital images. *Computers & Graphics*, 68: 142-151. <https://doi.org/10.1016/j.cag.2017.08.010>
 - [26] Wang, R., Ma, L., Juefei-Xu, F., Xie, X., Wang, J., Liu, Y. (2019). Fakespotter: A simple baseline for spotting AI-synthesized fake faces. *arXiv Preprint arXiv*: 1909.06122, 2. <https://doi.org/10.48550/arXiv.1909.06122>
 - [27] Marra, F., Saltori, C., Boato, G., Verdoliva, L. (2019). Incremental learning for the detection and classification of gan-generated images. In *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, Delft, Netherlands, pp. 1-6. <https://doi.org/10.1109/WIFS47025.2019.9035099>
 - [28] Li, Y., Lyu, S. (2018). Exposing deepfake videos by detecting face warping artifacts. *arXiv Preprint arXiv*: 1811.00656. <https://doi.org/10.48550/arXiv.1811.00656>
 - [29] Yang, X., Li, Y., Qi, H., Lyu, S. (2019). Exposing GAN-

- synthesized faces using landmark locations. In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, New York, NY, USA, pp. 113-118. <https://doi.org/10.1145/3335203.3335724>
- [30] Gatys, L.A., Ecker, A.S., Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414-2423.
- [31] Haralick, R.M., Shanmugam, K., Dinstein, I.H. (2007). Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, (6): 610-621. <https://doi.org/10.1109/TSMC.1973.4309314>
- [32] Brock, A., Donahue, J., Simonyan, K. (2018). Large scale GAN training for high fidelity natural image synthesis. arXiv Preprint arXiv: 1809.11096. <https://doi.org/10.48550/arXiv.1809.11096>
- [33] Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., Li, H. (2019). Protecting world leaders against deep fakes. CVPR workshops, 1:38-45.
- [34] Oreshko, E.I., Erasov, V.S., Lashov, O.A., Yakovlev, N.O. (2022). Stability study of monolithic and layered plates under compression. Inorganic Materials: Applied Research, 13(2): 588-598. <https://doi.org/10.1134/S2075113322020320>
- [35] Cheng, K., Tahir, R., Eric, L.K., Li, M. (2020). An analysis of generative adversarial networks and variants for image synthesis on MNIST dataset. Multimedia Tools and Applications, 79: 13725-13752. <https://doi.org/10.1007/s11042-019-08600-2>
- [36] Alkishri, W., Widyarto, S., Yousif, J.H., Al-Bahri, M. (2023). Deepfake image detection methods using discrete fourier transform analysis and convolutional neural network. Journal of Jilin University (Engineering and Technology Edition), 42(2). <https://doi.org/10.17605/OSF.IO/BNPC4>
- [37] Patel, Y., Tanwar, S., Bhattacharya, P., Gupta, R., Alsuwian, T., Davidson, I.E., Mazibuko, T.F. (2023). An improved dense CNN architecture for deepfake image detection. IEEE Access, 11: 22081-22095. <https://doi.org/10.1109/ACCESS.2023.3251417>
- [38] Wang, B., Wu, X., Tang, Y., Ma, Y., Shan, Z., Wei, F. (2023). Frequency domain filtered residual network for deepfake detection. Mathematics, 11(4): 816. <https://doi.org/10.3390/math11040816>
- [39] Alis, D., Alis, C., Yergin, M., Topel, C., Asmakutlu, O., Bagcilar, O., Senli, Y.D., Ustundag, A., Salt, V., Dogan, S.N., Velioglu, M., Selcuk, H.H., Kara, B., Ozer, C., Oksuz, I., Kizilkilic, O., Karaarslan, E. (2022). A joint convolutional-recurrent neural network with an attention mechanism for detecting intracranial hemorrhage on noncontrast head CT. Scientific Reports, 12(1): 2084. <https://doi.org/10.1038/s41598-022-05872-x>
- [40] Mountzouris, K., Perikos, I., Hatzilygeroudis, I. (2023). Speech emotion recognition using convolutional neural networks with attention mechanism. Electronics, 12(20): 4376. <https://doi.org/10.3390/electronics12204376>
- [41] Tariq, S., Lee, S., Kim, H., Shin, Y., Woo, S.S. (2018). Detecting both machine and human created fake face images in the wild. In Proceedings of the 2nd International Workshop on Multimedia Privacy and Security. Association for Computing Machinery, New York, pp. 81-87. <https://doi.org/10.1145/3267357.3267367>
- [42] Lin, J., Li, Y., Yang, G. (2021). FPGAN: Face de-identification method with generative adversarial networks for social robots. Neural Networks, 133: 132-147. <https://doi.org/10.1016/j.neunet.2020.09.001>
- [43] Birunda, S.S., Nagaraj, P., Narayanan, S.K., Sudar, K.M., Muneeswaran, V., Ramana, R. (2022). Fake image detection in twitter using flood fill algorithm and deep neural networks. In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, pp. 285-290. <https://doi.org/10.1109/Confluence52989.2022.9734208>
- [44] Yang, J., Xiao, S., Li, A., Lan, G., Wang, H. (2021). Detecting fake images by identifying potential texture difference. Future Generation Computer Systems, 125: 127-135. <https://doi.org/10.1016/j.future.2021.06.043>
- [45] Ramaswamy, H.G. (2020). Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization. In Proceedings of The IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 983-991.
- [46] Imambi, S., Prakash, K.B., Kanagachidambaresan, G.R. (2021). PyTorch. In Programming with Tensorflow: Solution for Edge Computing Applications, Springer, Cham, pp. 87-104. https://doi.org/10.1007/978-3-030-57077-4_10
- [47] Gulyaev, A.I., Erasov, V.S., Oreshko, E.I., Utkin, D.A. (2022). Analyzing the destruction of a carbon-fiber-reinforced polymer during the pushout of a multifilamentary cylinder. Polymer Science, Series D, 15(4): 574-580. <https://doi.org/10.1134/S1995421222040311>
- [48] Luo, W., Li, Y., Urtasun, R., Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. Advances in Neural Information Processing Systems, 29.
- [49] Zhang, X., Karaman, S., Chang, S.F. (2019). Detecting and simulating artifacts in gan fake images. In 2019 IEEE International Workshop on Information Forensics and Security (WIFS) Delft, Netherlands, pp. 1-6. <https://doi.org/10.1109/WIFS47025.2019.9035107>
- [50] Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. arXiv Preprint arXiv: 1811.12231. <https://doi.org/10.48550/arXiv.1811.12231>
- [51] Gatys, L., Ecker, A.S., Bethge, M. (2015). Texture synthesis using convolutional neural networks. Advances in Neural Information Processing Systems, 28.
- [52] Kodali, N., Abernethy, J., Hays, J., Kira, Z. (2017). On convergence and stability of gans. arXiv Preprint arXiv: 1705.07215. <https://doi.org/10.48550/arXiv.1705.07215>
- [53] Radford, A., Metz, L., Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv Preprint arXiv: 1511.06434. <https://doi.org/10.48550/arXiv.1511.06434>
- [54] Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition, pp. 8789-8797.
- [55] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. Communications of The ACM, 60(6): 84-90. <https://doi.org/10.1145/3065386>
- [56] Buolamwini, J., Gebru, T. (2018). Gender Shades:

Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning*

Research, 81: 1-15.