Jewel Journal of Scientific Research (JJSR) 9(1&2): 26–33, 2024
©Federal University of Kashere-Nigeria (Print ISSN: 2384 – 6267, Online ISSN 2756-651X)
https://journals.fukashere.edu.ng/index.php/jjsr

Enhancing Image-Based Captcha Scheme Using Cycle-Consistent Generative Adversarial Network

Babantakko, Z. D.[1*] and Zaharaddeen, S. I.[2]
[1]Department of Computer Science, Federal University of Kashere, Gombe.
[2]Department of Computer Science, Federal University Dutse, Jigawa.
***Corresponding Author**: zalihadanazumi31@gmail.com; +2348163175407

Abstract
Researchers have carried out extensive work on improving Completely Automated Public Turing Test to Tell between Computers and Humans Apart (CAPTCHA) to prevent malware or bots from compromising information, this has led to the development of the most secure and user-friendly form of CAPTCHA popularly known as image-based CAPTCHA. It is best used as a shield that protects unauthorized access to information available online. As more sophisticated algorithms emerge, attacks on image-based CAPTCHA have also increased, and using deep learning algorithms attacks on CAPTCHA design have become more vulnerable. Research has proven that Adversarial has a promising direction in overcoming such challenges, hence we proposed an image-based CAPTCHA scheme as an effort to enhance the CAPTCHA design through the use of cycle-consistent generative adversarial network which was minimized with Mean Square Error and Mean Absolute Error and Inception Score was used to evaluate the quality of generated image with an average difference of 0.025 when compared with the existing scheme. That is to say our scheme effectively produces a synthetic image that is indistinguishable from the real image, which can easily fool the DeCAPTCHAs while solving an automated CAPTCHA challenge.

Keywords: Generative Adversarial Network, Inception score, Mean Square Error, Mean Absolute Error
Received: 20th Feb., 2024 *Accepted:* 23th April, 2024 *Published Online:* 20nd June, 2024

Introduction

With the rapid growth of the Internet, Web security has become an important issue, CAPTCHA is a class of automated challenges used to differentiate between legitimate human users and computer programs (`bots') on the Internet, it was first discovered in 2000 in Carnegie Mellon University by a group of scholars, Luis *et. al*., (2003). Various web search engines like Yahoo, Google, Bing, etc. use CAPTCHA to distinguish between an authorized user and a malicious program, The main purpose of CAPTCHA is to disengage bots from unauthorized access to web applications, online fraud, and misconduct on the internet. However, the current CAPTCHA technologies are prone to various forms of attacks such as object recognition, dictionary attack, database attack, e. t. c. Besides that, some of the CAPTCHA lack robustness and they are difficult for human users to use even with multiple attempts by Gutub and Kheshaifaty (2023). CAPTCHA undergoes several stages of development, the earliest and most popularly used CAPTCHA scheme is text-based CAPTCHA, which is designed based on English letters and Arabic numerals. It supports input characters through keyboards. Text-based CAPTCHA as proven by Huang, G. et. al., (2017), is the most easily broken CAPTCHA security with the development of deep learning algorithms such as CNN, KNN, SVM, and more which have segmentation recognition resistance. To overcome the text-based method challenges,

the image-based CAPTCHA was initiated. The next is the image-based CAPTCHAs, which require an understanding of image content by the user and perform some click operation. The newly introduced form of CAPTCHA is the Audio/video-based CAPTCHAs, which possess a special CAPTCHA scheme although they are rarely used in current CAPTCHA systems because they are still under development and improvement, Zhang et al., (2019). To address the aforementioned challenges, this research proposes to develop an image-based CAPTCHA model that takes into account various forms of CAPTCHA challenge. In particular, the Cycle GAN algorithm has been used to design an image-based CAPTCHA scheme that has a qualified attribute of collective style transfer, object transfiguration, and photo enhancement.

Related work

Efforts were made by several scholars to straighten the image-based CAPTCHA design, Tang *et al*. (2019) proposed an image-based CAPTCHA using style transfer learning that increased defense against State of Art attacks, Wang *et al.* (2019), proposed a DeCAPTCHA recognizer using Dense Convolutional Neural Network which achieved 99.9% CAPTCHA recognition accuracy, Kwon *et al*. (2020) developed an enhanced image generation scheme for CAPTCHA design with Fast Gradient Sign and DeepFool methods to improve the security and recognition accuracy by DeCAPTCHA. Hitaj et al. (2020) proposed a scheme called CAPTURE (CAPtcha Technique Uniquely REsistant) using Adversarial examples to enhance the robustness of CAPTCHA against program attacks, Chen et al., (2020) introduced a new CAPTCHA known as StyleCAPTCHA. Users are tasked with classifying stylized images of human and animal faces, the research demonstrates that this CAPTCHA effectively defends against advanced face detection methods and Deep Convolutional Networks (DCNs), in (2022) Jia et al. proposed a text-image-based CAPTCHA that utilizes cognitive processes and semantic reasoning. This approach synthesizes features like sentences, objects, and locations to generate a multi-conditional CAPTCHA that is resistant to CNN classification attacks. Lu et al. (2022) proposed a framework for solving CAPTCHAs using a deep skipping convolutional neural network. The study reveals that certain characters have a higher recognition success rate in both 4 and 5-character CAPTCHAs, making them more vulnerable to being broken. Recently Dinh et al. (2023) presented an enhanced security architecture for CAPTCHA using Neural Style Transfer and Adversarial examples with higher resilience. The results demonstrate a significant improvement in the security of current CAPTCHAs. Ray *et al*. (2023) proposed a new image-based CAPTCHA called Style Matching Captcha (SMC) that utilizes Neural Style Transfer which shows excellent resilience against security attacks but did not take into account visually impaired users, another contribution was made by Jiang et al. (2023) who presented a paper on the enhancement of the security of Diff-CAPTCHA, an image-based CAPTCHA, using the Denoising Diffusion Model. The findings demonstrate an improvement in CAPTCHA security without compromising usability as the major drawback faced by other deep learning-based CAPTCHA due poor quality of synthetic images which reduces the rate of users engaging with the CAPTCHA.

The paper is organized as follows: Section III introduces the proposed method with its implementation details while Section IV presents results and discussion which end with concluding remarks and further research direction in section V.

Methodology

There are few steps followed when developing this scheme which involves the following,

A. Data identification and Preprocessing

The dataset used was reCAPTCHA images and Monet painting images which are publicly available at Kaggle website. We obtained 50% images respectively from reCAPTCHA and Monet painting dataset, were 75% was

classified for training and 25% for testing the scheme.

The preprocessing take place after load the dataset from google drive, where it's converted to numpy array and was resize to 256*256 pixel from its original size (120*120). The processed data was stored in the session memory as program to be run on Google Colab using

T4 GPU.

B. System Process

The system starts operation by loading the dataset from a drive that contains two different domains of heterogenous features, undergoes preprocessing to fit the parameters, and passes to the generator model for generating the best fake sample that looks exactly like the real sample alongside the discriminator whose work is to distinguish between the real and fake image. The composite model was used to update both the generator and discriminator using Mean Absolute Error (MAE) and Mean Square Error (MSE) to minimize and maximize their chance of producing the desired outcome.

MAE is robust to outliers (treat errors as same), easy for interpretation due to same scale of prediction and target values, it's hard to be differentiable while having absolute values, it gives higher predicted values when optimize and also scale-dependent like MSE (Lewinson, 2023).

$$MAE = 1/n * \sum^{n}_{i=1} |Y_i - \hat{Y}_i| \quad (1)$$

$$MSE = 1/n * \sum^{n}_{i=1} (Y_i - \hat{Y}_i)^2 \quad (2)$$

Where **n** is the number of dataset starting from *i* to n

$\mathbf{Y}_i$ is the *i*'th expected value

$\hat{\mathbf{Y}}_i$ is the *i*'th predicted value, their difference is squared which enable the error to have positive value.

Inception Score as widely recognized metric for evaluating a generative model, mostly Generative Adversarial Networks (GANs). It groups the generated images and predicts a single floating output as a score (David, 2019).

$$IS(G) = \exp(E_{x \sim pg} D_{KL}(p(y/x) \| p(y))) \quad (3)$$

Where: $x \sim pg$ x is the generated image from sample pg, $P(y/x)$ is the conditional class distribution, D_{KL} $(p(y/x)$ is the KL-divergence between the distributions p and q, $P(y) = \int_x p(y/x) pg(x)$ is the marginal class distribution, and while the exp is there to make it easier to compare the values

C. *Implementation Algorithm of the Proposed scheme*

Step1: start;

Step2: Load the dataset from drive;

Step3: Preprocessed the data and stored as numpy array;

Step4: Read the generator and discriminator model;

Step5: Read the composite model for updating step4;

Step6: Train and test the model;

Step7: if the minimized error reached a static point, then go to step9;

Step8: else go to step6;

Step9: output sample of generated image

Step10: end

D. System Architecture

The architecture of the scheme was illustrated in figure 1 below, it shows all the structure of the scheme been developed.

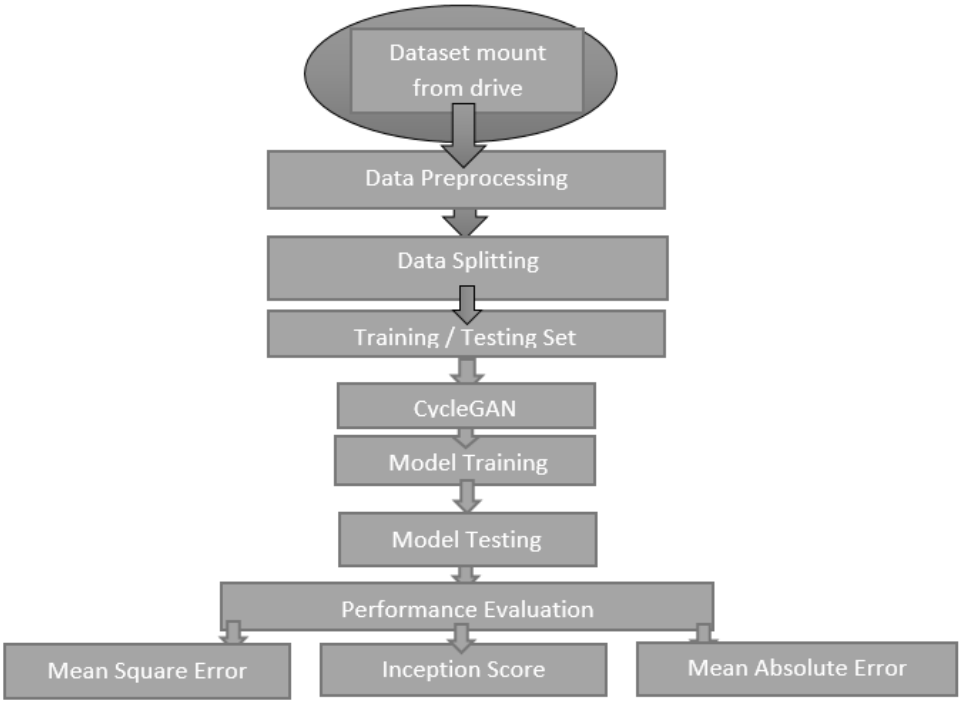

Figure 1. System Architecture

Results and Discussion

The result below illustrates the transformation of images generated by the model and the level of transparency in the resulting images. Both the synthetic and real images demonstrate the outcome of style transfer by incorporating elements from one another. This process yields the desired result at every stage of execution.

Figure 2 demonstrates the transition effect from real to fake images after undergoing data training and testing. The result showcases the transformation obtained and the style transfer implemented in the scheme.

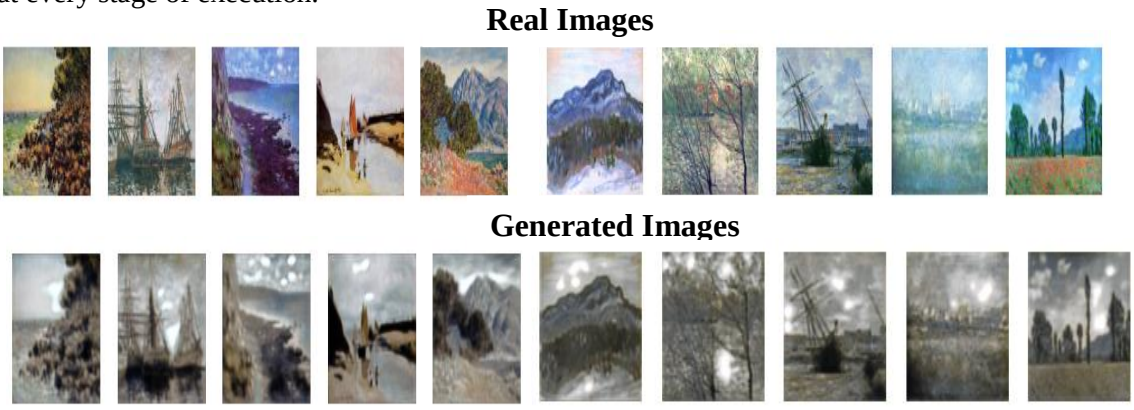

Figure 2. Generated sample A to B

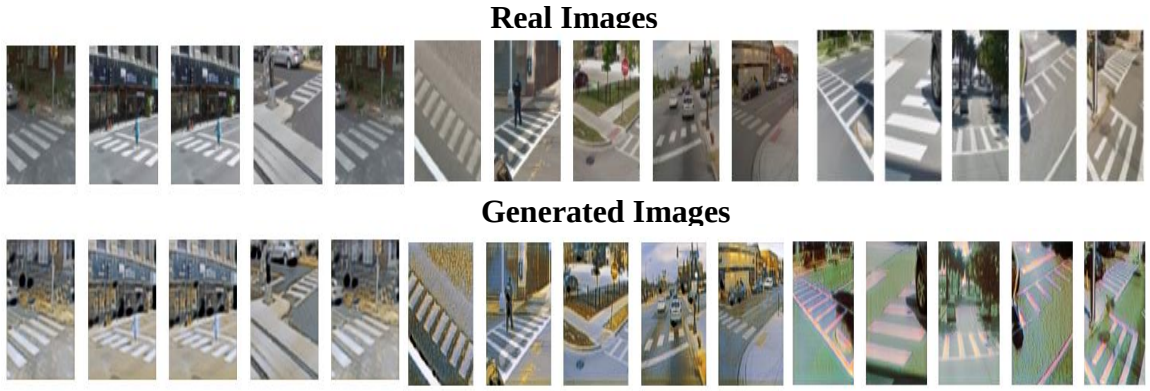

Figure 3. Generated sample of B to A

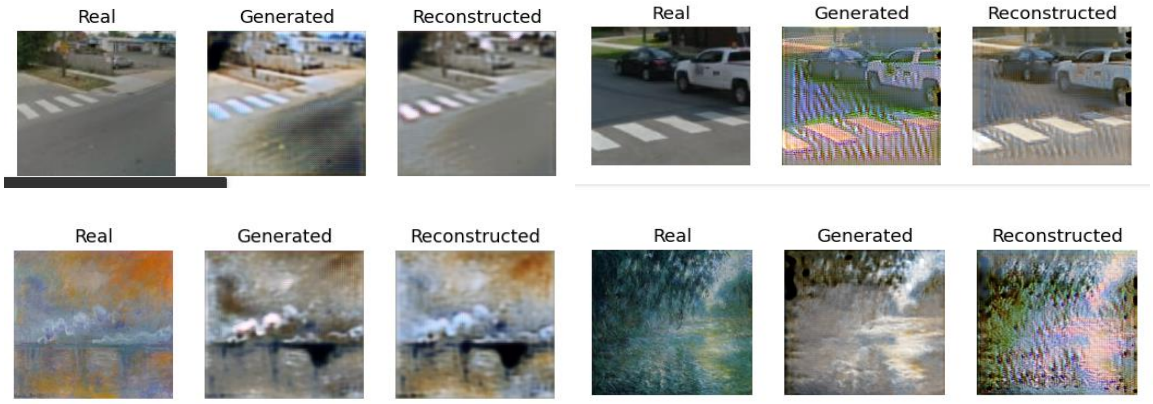

Figure 4. Cycle consistent images of B and A

Figure 3 depicts the sample of images obtained from generator B to A, which applies a style transformation to reCaptcha data using a Monet painting feature. This feature is one of the functionalities offered by CycleGAN.

Figure 4 demonstrates the presence of a radiant transition resulting from the cycle-consistent advantage between the real, generated, and reconstructed images after undergoing appropriate training.

A. Loss Function Evaluation

The loss function is used to determine how well an algorithm's current outcome aligns with the intended outcome. It assesses how effectively the algorithm models the dataset. The loss function can be categorized into regression and classification tasks based on their appropriateness. In our study, the Mean Square Error (MSE) was employed for the adversarial loss, while the Mean Absolute Error (MAE) was utilized for the identity and cycle-consistent loss to minimize errors in the approach. The results, depicted in Figures 5 and 6 below, demonstrated a relatively successful accomplishment of the desired outcome. The losses were combined with the adversarial loss and cycle-consistent loss to reconstruct the images back to their original stage. The losses are minimal that is to say, output can easily fool the DeCAPTCHAs while solving challenges, and centering around a specific point that aligns with the model output.

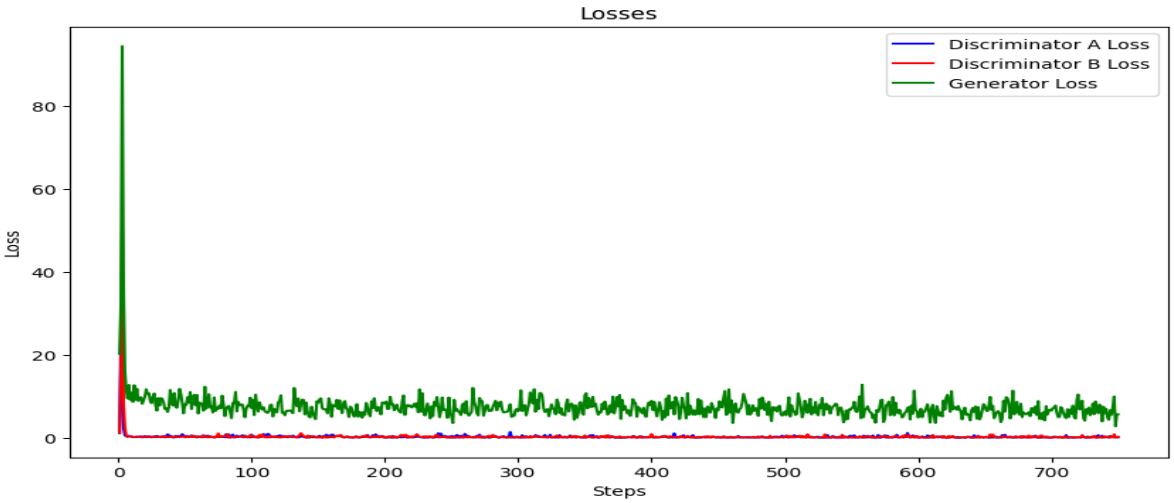

Figure 5. Developed Scheme Loss Function

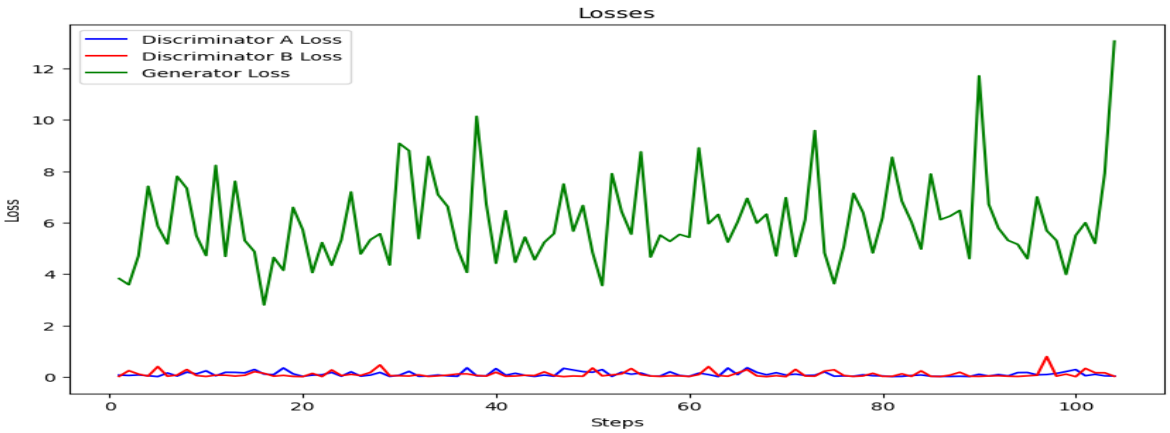

Figure 6. Loss Function of few Images

B. Evaluation using Inception Score

The inception score accomplishes two important tasks in evaluating the generated images: assessing their quality and diversity. A higher score indicates that the scheme has produced clearer and more diverse output. This score is calculated by considering the conditional class distribution p(x/y) as well as the marginal distribution p(y), which represents the probability of the fake or generated images. Table 1 displays the scores obtained using equation (3), with "D" representing the number of datasets. Furthermore, figure 7 provides a graphical representation of these scores.

Table 1: Inception Score of the Developed scheme

Inception Score						
	D_{80}	**D_{100}**	**D_{120}**	**D_{200}**	**D_{500}**	**D_{1000}**
$G_{A\ to\ B}$ ($^{+}$- 0.00013)	1.0097215 0.0043304 32	1.0095241 0.0014390 52	1.007902 0.0012995 09	1.012455 0.0019488 04	1.0127925 0.0021645 42	1.0134413 0.000903011
$G_{B\ to\ A}$ ($^{+}$- 0.00013)	1.0130414 0.0058372 45	1.0629267 0.0273022 19	1.0133126 0.0029597 21	1.0462182 0.0124404 82	1.049478 0.0092027 29	1.049478 0.009202729

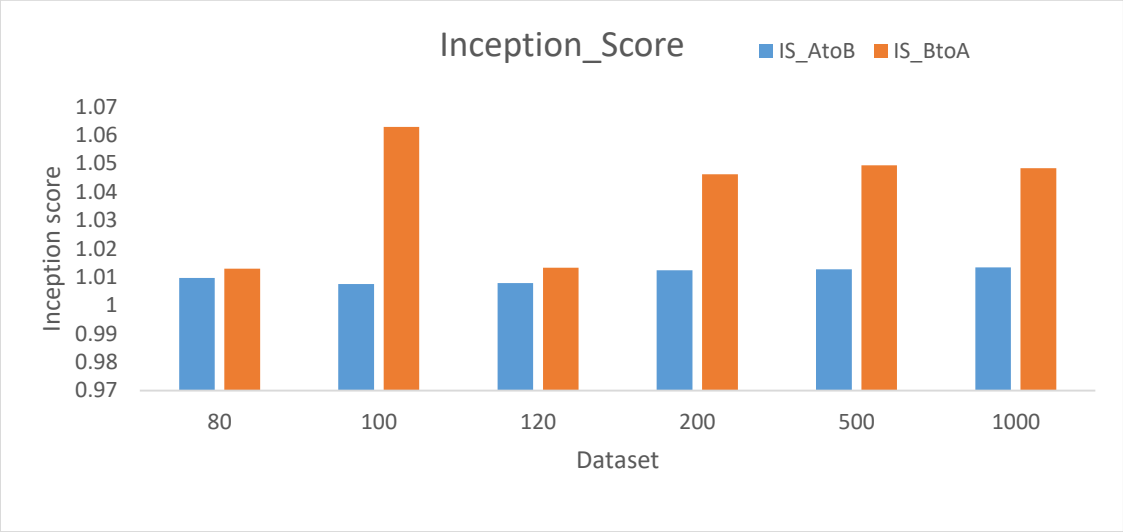

Figure 7. Chart graph of Inception score of developed scheme

Table 2: Comparison based on the IS of the developed scheme with existing work.

Inception Score						
	D $_{no}$	**Generator A to B and B to A**		**D $_{no}$**	**Generator A to B and B to A**	**Difference**
Developed scheme	500	1.01279, 1.04947	Existing work	500	1.02734, 1.03109	-0.014, 0.018
	200	1.01245, 1.04622		200	1.02985, 1.02607	-0.017, 0.020
	100	1.00952, 1.06292		100	1.02649, 1.02676	-0.018, 0.036

Table 2 displays the results of the inception score obtained from equation (3). This score was obtained after assessing the quality and diversity of the images generated by both our scheme and the existing work. To ensure a fair comparison and a more recognizable output, both researchers utilized the same dataset. By calculating the difference between our scheme's scores and those of the existing work's scores, it is clear that our scheme outperforms the other. On average, there is a difference of 0.0246 between the two. This indicates that our scheme is capable of producing more recognizable images and designing a stronger image-based CAPTCHA that can withstand state-of-the-art attacks.

Conclusion and Recommendation

This study contributes to the field of image processing as well as cyber security through the CAPTCHA generation scheme. The major goal was to propose a framework that enhances the design of an image-based CAPTCHA through the use of a Generative Adversarial algorithm, to have better security and maintain greater usability for the users. Therefore, the developed scheme showcased a great possibility for designing an image-based CAPTCHA that is more secure and also maintains remarkable usability based on the outcome showcased above.

References

Dinh, N., Nguyen, T. and Truong, V. (2023). zxCAPTCHA: New Security-Enhanced CAPTCHA. In *2023 15th International Conference on Knowledge and Smart Technology* (KST) (pp. 1-6). IEEE.

Devid Mack, 2019 "A simple explanation of the Inception Score" https://medium.com/octavian-ai/a-simple-explanation-of-the-inception-score-372dff6a8c7a

Gutub, A., & Kheshaifaty, N. (2023). Practicality analysis of utilizing text-based CAPTCHA vs. graphic-based CAPTCHA authentication. Multimedia Tools and Applications, 82(30): 46577-46609.

Hitaj, D., Hitaj, B., Jajodia, S. and Mancini, L. V. (2020). Capture the bot: Using adversarial examples to improve captcha robustness to bot attacks. *IEEE Intelligent Systems*, 36(5): 104-112

Jia, X., Xiao, J. and Wu, C. (2022). TICS: text–image-based semantic CAPTCHA synthesis via multi-condition adversarial learning. *The Visual Computer*, 38(3): 963-975.

Jiang, R., Zhang, S., Liu, L., and Peng, Y. (2023). Diff-CAPTCHA: An Image-based CAPTCHA with Security Enhanced by Denoising Diffusion Model. arXiv preprint arXiv:2308.08367.

Kwon, H., Yoon, H., and Park, K. W. (2020). Robust captcha image generation enhanced with adversarial example methods. *IEICE TRANSACTIONS on Information and Systems*, 103(4): 879-882.

Lu, S., Huang, K., Meraj, T. and Rauf, H. T. (2022). A novel CAPTCHA solver framework using deep skipping Convolutional Neural Networks. *PeerJ Computer Science*, 8: e879.

Lewinson Eryk. (2023) https://towardsdatascience.com/a-comprehensive-overview-of-regression-evaluation-metrics-6264af0926db

Ray, P., Bera, A., Giri, D. and Bhattacharjee, D. (2023). Style matching CAPTCHA: match neural transferred styles to thwart intelligent attacks. Multimedia Systems, 29(4), 1865-1895.

Tang, M., Gao, H., Zhang, Y., Liu, Y., Zhang, P., and Wang, P. (2018). Research on deep learning techniques in breaking text-based captchas and designing image-based captcha. *IEEE Transactions on Information Forensics and Security*, 13(10): 2522-2537.

Von Ahn, L., Blum, M., Hopper, N. J. and Langford, J. (2003). CAPTCHA: Using hard AI problems for security. In International conference on the theory and applications of cryptographic techniques (pp. 294-311). Springer, Berlin, Heidelberg.

Wang, J., Qin, J., Xiang, X., Tan, Y. and Pan, N. (2019). CAPTCHA recognition based on deep convolutional neural network. *Math. Biosci. Eng*, 16(5): 5851-5861.

Zhu, J. Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).